Cosmology

Yury Holubeu *

November 12, 2025

This draft is not aimed for distribution. Cosmology is discussed in detail. The note has solved problems and summary of special topics. Readers who have opened this note for the first time are highly recommended to read the preface.

Contents

Preface and main motivation **14**

I —— Typical Cosmology in a Nutshell —— **16**

1 Main concepts of modern cosmology **16**
1.1 Most believed properties of the universe . . . 16
1.1.1 General history of the universe in scope of modern cosmology . . . 16
1.1.2 Properties and objects of the current universe . . . 23
1.1.3 Determination of properties of the universe in a nutshell . . . 23
1.1.4 Concept of shape, expansion of the universe, dark mater, dark energy . . . 24
1.1.5 Main Philosophical Ideas about Universe . . . 24
1.1.6 Validity of modern cosmology, tests, directions of improvement (!?) . . . 25
1.2 Main models of modern cosmology . . . 26
1.2.1 Friedmann universe in a nutshell . . . 26
1.2.2 On distances, expansion, Hubble's law, Hubble Sphere, horizons, redshift . . . 33
1.2.3 Homogeneous inflation in a nutshell . . . 40
1.2.4 Description of equilibrium and non equilibrium processes . . . 46
1.2.5 Nucleosynthesis, baryogenesis, decouplings in a nutshell . . . 51
1.3 Structure formation and anysotropies in a nutshell . . . 58
1.3.1 Overview of fluctuations dynamics, parameters, power spectrum (!!!!!!) . . . 58
1.3.2 Perturbations setup from newtonian and relativisic physics . . . 64
1.3.3 Initial conditions from inflaton fluctuations . . . 72
1.3.4 Derivations of full perturbed Boltzmann and Einstain eq. in a nutshell (!??!) . . . 75
1.3.5 Dynamics of fluctuations, power spectrum and spectral indexes . . . 82
1.3.6 Details of evolution of fluctuations, other approaches in a nutshell . . . 90
1.3.7 Anisotropy predictions, correlations in a nutshell (!!!!!!!!!) . . . 94
1.3.8 Details of anisotropies, other approaches in a nutshell (!!!!!!!!!) . . . 114
1.3.9 Other properties of CMB in a nutshell (!!??) . . . 117
1.4 Mathematics and physical background for applications . . . 118
1.4.1 Physical background for fluctuations, constants, mathematics . . . 118
1.4.2 Basic diffgeom and mathematics for cosmology . . . 120
1.5 Main cosmological parameters and ideas . . . 121
1.5.1 Hubble parameter and tension in a nutshell . . . 121
1.5.2 Dark Matter in a nutshell . . . 121
1.5.3 Cosmological constant (dark energy) in a nutshell . . . 130
1.5.4 Sense of inflanton field . . . 131
1.6 Main experimental facts, their analysis and confirmation of cosmological models . . . 132
1.6.1 Challenges of Λ-CDM . . . 132
1.6.2 Dark Matter detection . . . 133
1.6.3 On experimental data . . . 138
1.6.4 Statistical methods for experiments in cosmology (????) . . . 141

*https://yuriholubeu.github.io/, yuri.holubev@gmail.com

II —— Special Topics in Cosmology in a Nutshell —— 142

2 Other models, theories, properties and data (??) 142
2.1 Main physical background for cosmology . . . 142
2.1.1 Gravity ideas and formalisms for cosmology . . . 142
2.1.2 Gravitational lensing for cosmology in a nutshell (???) . . . 142
2.1.3 On some stars and astrophysics for cosmology . . . 150
2.2 Other models of cosmology in a nutshell . . . 150
2.2.1 Variable Constants models overview . . . 150
2.2.2 Description of formation of Solar System . . . 150
2.2.3 On other models of inflation . . . 151
2.2.4 Primordial phase transitions, quantum bubbles, primordial GWs (????) . . . 151
2.2.5 Cosmic strings, (etc...) . . . 159
2.2.6 On quantum cosmology, wavefunction of universe (????) . . . 160
2.2.7 On bouncing universe model . . . 161
2.2.8 On Conformal cyclic cosmology . . . 161
2.2.9 Other most typical models of cosmology . . . 161
2.2.10 Cosmological models that are considered wrong . . . 163
2.3 On general and more philosophical ideas about physics about cosmology . . . 163
2.3.1 Cosmological paradoxes and philosophical problems . . . 164
2.3.2 What Could be in the Universe if Something Was Different? . . . 164
2.3.3 What is the fate of the universe? . . . 164
2.3.4 What Was in the Beginning of the Universe? . . . 164
2.3.5 Nature of Time, Chronogeometry . . . 164
2.3.6 Some Other Hypothetical Questions . . . 164

3 Some plots from different part of cosmology 165

III Fundamentals of Cosmology 174

4 General Fundamentals of Cosmology 174
4.1 General, Philosophical Approach . . . 174
4.1.1 Why In General We Can Describe a Universe (??) . . . 174
4.1.2 Flatness of the Universe . . . 174
4.1.3 Gravitation on the Scale of the Universe (???) . . . 174
4.1.4 Idea of a Cosmological Constant . . . 174
4.1.5 Gravitational Constant . . . 174
4.1.6 Time in Context of Cosmology (????) . . . 174
4.1.7 Idea of Dark Energy . . . 175
4.1.8 Idea of Dark Matter . . . 175
4.1.9 Redshift . . . 175
4.1.10 Redshift by Landau . . . 177

5 Main Methods of Gravity and Mathematics for Cosmology 183
5.1 Geometry . . . 184
5.1.1 Geodesics . . . 184
5.1.2 Distances . . . 188
5.2 Differential Geometry for Cosmology . . . 190
5.2.1 Killing Vector (??) . . . 190

6 Some Basic Models **190**
6.1 Friedman Universe . . . 190
6.1.1 Cosmic parameters . . . 190
6.1.2 Spacetime Curvature . . . 192
6.1.3 Friedmann Equations . . . 195
6.1.4 Single-Component Universe . . . 200
6.1.5 Two-Component Universe . . . 201
6.1.6 Common misconceptions of horizons and the superluminal expansion . . . 202
6.2 Inflation . . . 217
6.2.1 The Horizon Problem (??) . . . 218
6.2.2 A Shrinking Hubble Sphere . . . 222
6.2.3 The Physics of Inflation . . . 226
6.3 Thermal History . . . 230
6.3.1 The Hot Big Bang . . . 231
6.3.2 Equilibrium . . . 235
6.3.3 Isotropic Space by Landau . . . 238
6.3.4 Closed Isotropic Model by Landau . . . 242
6.3.5 Open Isotropic Model by Landau . . . 246
6.4 Not Popular, but Basic Models . . . 249
6.4.1 Gravitational Stability of Isotropic World . . . 249
6.4.2 Homogeneous Spaces . . . 254
6.4.3 Flat Anisotropic Model . . . 259
6.4.4 Oscillations During Approach to Singularity . . . 263
6.4.5 Singularity in Einstein's Equation for Cosmology (???) . . . 267
6.5 Relics from the Hot Big Bang . . . 270
6.5.1 Introduction . . . 270
6.5.2 Theory . . . 272
6.5.3 Big Bang Cosmology . . . 273
6.5.4 Afterglow of the Big Bang . . . 279
6.5.5 Cosmic Sound Waves . . . 286
6.5.6 Light Relics . . . 291
6.6 Relics from Inflation . . . 300
6.6.1 Inflationary Cosmology . . . 301
6.6.2 Quantum Initial Conditions . . . 306
6.6.3 Primordial Interactions . . . 317
6.6.4 Heavy Relics . . . 325

7 TASI Lectures on Inflation Daniel Baumann **331**
7.1 Classical Dynamics of Inflation . . . 332
7.2 Introduction . . . 333
7.2.1 The Microscopic Origin of Structure . . . 333
7.2.2 Outline of the Lectures . . . 337
7.3 Classical Dynamics of Inflation by Baumann 1th lecture . . . 338
7.3.1 *Review*: The Homogeneous Universe . . . 338
7.3.2 Big Bang Puzzles . . . 345
7.3.3 A First Look at Inflation . . . 348
7.3.4 The Physics of Inflation . . . 353
7.4 Quantum Fluctuations during Inflation by Baumann 2th lecture . . . 360
7.4.1 *Review*: Cosmological Perturbations . . . 361
7.4.2 *Preview*: The Quantum Origin of Structure . . . 366
7.4.3 Quantum Mechanics of the Harmonic Oscillator . . . 368
7.4.4 Quantum Fluctuations in de Sitter Space . . . 371
7.4.5 Primordial Spectra . . . 377
7.5 Contact with Observations by Baumann 3rd lecture . . . 379

7.5.1 Connecting Observations to the Early Universe . . . 379
7.5.2 *Review*: The Cosmic Microwave Background . . . 381
7.5.3 *Review*: Large-Scale Structure . . . 388
7.5.4 Current Evidence for Inflation . . . 390
7.5.5 Future Tests of Inflation . . . 395
7.6 Primordial Non-Gaussianity by Baumann 4th lecture . . . 398
7.6.1 Preliminaries . . . 398
7.6.2 Theoretical Expectations . . . 402
7.6.3 Observational Prospects . . . 405
7.6.4 *Summary*: Lecture 4 . . . 406
7.7 Conclusions . . . 407
7.7.1 *Recap*: TASI Lectures on Inflation . . . 407
7.7.2 Future Prospects and Open Problems . . . 408
7.7.3 Guide to Further Reading . . . 409
7.8 Appendix . . . 411
7.8.1 Cosmological Perturbation Theory . . . 411
7.8.2 Free Field Action for $\mathcal{R}$ . . . 423
7.8.3 A Brief Review of the In-In Formalism . . . 426
7.8.4 Slow-Roll Inflation in the Hamilton-Jacobi Approach . . . 428

8 General Fundamental Questions **431**
8.1 What Cosmology Teaches Us About Our Life? . . . 431
8.2 What Cosmology Teaches Us About Physical Nature of the Word? . . . 432

9 Alternative Fundamental Ideas **432**
9.0.1 Dicke's Explanation of the Redshift . . . 432

10 Cosmology by Weinberg **432**
10.1 The Expansion of the Universe . . . 432
10.1.1 1.1 Spacetime geometry . . . 433
10.1.2 1.1 Spacetime geometry . . . 438
10.1.3 1.2 The cosmological redshift . . . 439
10.1.4 1.2 The cosmological redshift . . . 439
10.1.5 1.3 Distances at small redshift: The Hubble constant . . . 441
10.1.6 1.4 Luminosity distances and angular diameter distances . . . 454
10.1.7 1.5 Dynamics of expansion . . . 456
10.1.8 1.6 Distances at large redshift: Accelerated expansion . . . 464
10.1.9 1.7 Cosmic expansion or tired light? . . . 473
10.1.10 1.8 Ages . . . 475
10.1.11 1.9 Masses . . . 480
10.1.12 1.10 Intergalactic absorption . . . 486
10.1.13 1.11 Number counts . . . 492
10.1.14 1.12 Quintessence . . . 496
10.1.15 1.13 Horizons . . . 503
10.2 The Cosmic Microwave Radiation Background . . . 504
10.2.1 2.1 Expectations and discovery of the microwave background . . . 505
10.2.2 2.2 The equilibrium era . . . 510
10.2.3 2.3 Recombination and last scattering . . . 513
10.2.4 2.4 The dipole anisotropy . . . 524
10.2.5 2.5 The Sunyaev-Zel'dovich effect . . . 526
10.2.6 2.6 Primary fluctuations in the microwave background: A first look . . . 529
10.3 The Early Universe . . . 538
10.3.1 3.1 Thermal history . . . 539
10.3.2 3.2 Cosmological nucleosynthesis . . . 547

10.3.3 3.3 Baryonsynthesis and leptonsynthesis . . . 557
10.3.4 3.4 Cold dark matter . . . 566
10.4 Inflation . . . 576
10.4.1 4.1 Three puzzles . . . 577
10.4.2 4.2 Slow-roll inflation . . . 581
10.4.3 4.3 Chaotic inflation, eternal inflation . . . 587
10.5 General Theory of Small Fluctuations . . . 588
10.5.1 5.1 Field equations . . . 589
10.5.2 5.2 Fourier decomposition and stochastic initial conditions . . . 596
10.5.3 5.3 Choosing a gauge . . . 601
10.5.4 5.4 Conservation outside the horizon . . . 608
10.6 Evolution of Cosmological Fluctuations . . . 616
10.6.1 6.1 Scalar perturbations - kinetic theory . . . 617
10.6.2 6.1 Scalar perturbations - kinetic theory . . . 620
10.6.3 6.2 Scalar perturbations - the hydrodynamic limit . . . 629
10.6.4 6.3 Scalar perturbations - long wavelengths . . . 635
10.6.5 6.4 Scalar perturbations - short wavelengths . . . 640
10.6.6 6.5 Scalar perturbations - interpolation & transfer functions . . . 651
10.6.7 6.5 Scalar perturbations - interpolation & transfer functions . . . 653
10.6.8 6.6 Tensor perturbations . . . 658
10.7 Anisotropies in the Microwave Sky . . . 671
10.7.1 7.1 General formulas for the temperature fluctuation . . . 671
10.7.2 7.2 Temperature multipole coefficients: Scalar modes . . . 682
10.7.3 7.3 Temperature multipole coefficients: Tensor modes . . . 696
10.7.4 7.4 Polarization . . . 702
10.8 The Growth of Structure . . . 725
10.8.1 8.1 Linear perturbations after recombination . . . 726
10.8.2 8.2 Nonlinear growth . . . 739
10.8.3 8.3 Collapse of baryonic matter . . . 744
10.9 Gravitational Lenses . . . 746
10.9.1 9.1 Lens equation for point masses . . . 747
10.9.2 9.2 Magnification: Strong lensing and microlensing . . . 749
10.9.3 9.3 Extended lenses . . . 754
10.9.4 9.3 Extended lenses . . . 754
10.9.5 9.3 Extended lenses . . . 756
10.9.6 9.4 Time delay . . . 757
10.9.7 9.5 Weak lensing . . . 761
10.9.8 9.6 Cosmic strings . . . 773
10.10 Inflation as the Origin of Cosmological Fluctuations . . . 774
10.10.1 10.1 Scalar fluctuations during inflation . . . 774
10.10.2 10.2 Tensor fluctuations during inflation . . . 786
10.10.3 10.3 Fluctuations during inflation: The slow-roll approximation . . . 788
10.10.4 10.3 Fluctuations during inflation: The slow-roll approximation . . . 793
10.10.5 10.4 Multifield inflation . . . 795
10.11 Appendix A Some Useful Numbers . . . 803
10.12 Appendix B Review of General Relativity . . . 804
10.13 Appendix C Energy Transfer between Radiation and Electrons . . . 819
10.14 Appendix D The Ergodic Theorem . . . 823
10.15 Appendix E Gaussian Distributions . . . 825
10.16 Appendix F Newtonian Cosmology . . . 826
10.17 Appendix G Photon Polarization . . . 828
10.18 Appendix H The Relativistic Boltzmann Equation . . . 830
10.19 Glossary of Symbols . . . 840

10.20 Assorted Problems . . . 843

11 Challenges for ΛCDM: An update by Perivolaropoulos, Skara 846
11.1 Introduction . . . 847
11.2 Hubble tension . . . 851
11.2.1 Methods for measuring H_0 and data . . . 851
11.2.2 Theoretical models . . . 882
11.3 Other Tensions - Curiosities . . . 902
11.3.1 Growth tension . . . 902
11.3.2 CMB anisotropy anomalies . . . 911
11.3.3 Cosmic dipoles . . . 918
11.3.4 BAO curiosities . . . 921
11.3.5 Parity violating rotation of CMB linear polarization (Cosmic Birefringence) . . . 922
11.3.6 Small-scale curiosities . . . 923
11.3.7 Age of the Universe . . . 924
11.3.8 The Lithium problem . . . 925
11.3.9 Quasars Hubble diagram . . . 926
11.3.10 Oscillating signals in short range gravity experiments . . . 927
11.3.11 Anomalously low baryon temperature . . . 928
11.3.12 Colliding clusters with high velocity . . . 928
11.4 Conclusions-Discussion-Outlook . . . 930
11.5 Appendix . . . 936

IV Properties of the Universe in Detail 940

12 Theory by Dodelson 940
12.1 The Standard Model and Beyond . . . 940
12.1.1 The Expanding Universe . . . 940
12.1.2 The Hubble Diagram . . . 944
12.1.3 Big Bang Nucleosynthesis . . . 946
12.1.4 The Cosmic Microwave Background . . . 949
12.1.5 Beyond the Standard Model . . . 950
12.2 The Smooth, Expanding Universe . . . 955
12.2.1 General Relativity . . . 956
12.2.2 Distances . . . 964
12.2.3 Evolution of Energy . . . 966
12.2.4 Cosmic Inventory . . . 969
12.3 Beyond Equilibrium . . . 983
12.3.1 Boltzmann Equation for Annihilation . . . 983
12.3.2 Big Bang Nucleosynthesis . . . 986
12.3.3 Recombination . . . 992
12.3.4 Dark Matter . . . 995
12.4 The Boltzmann Equations . . . 1003
12.4.1 The Boltzmann Equation for the Harmonic Oscillator . . . 1004
12.4.2 The Collisionless Boltzmann Equation for Photons . . . 1006
12.4.3 Collision Terms: Compton Scattering . . . 1012
12.4.4 The Boltzmann Equation for Photons . . . 1016
12.4.5 The Boltzmann Equation for Cold Dark Matter . . . 1018
12.4.6 The Boltzmann Equation for Baryons . . . 1022
12.5 Einstein Equations . . . 1028
12.5.1 The Perturbed Ricci Tensor and Scalar . . . 1028
12.5.2 Two Components of the Einstein Equations . . . 1031
12.5.3 Tensor Perturbations . . . 1034

12.5.4 The Decomposition Theorem . . . 1039
12.5.5 from Gauge to Gauge . . . 1040
12.6 Initial Conditions . . . 1042
12.6.1 The Einstein-Boltzmann Equations at Early Times . . . 1042
12.6.2 The Horizon . . . 1045
12.6.3 Inflation . . . 1047
12.6.4 Gravity Wave Production . . . 1055
12.6.5 Scalar Perturbations . . . 1060
12.7 Inhomogeneities . . . 1068
12.7.1 Prelude . . . 1068
12.7.2 Large Scales . . . 1074
12.7.3 Small Scales . . . 1079
12.7.4 Numerical Results and Fits . . . 1086
12.7.5 Growth Function . . . 1088
12.7.6 Beyond Cold Dark Matter . . . 1090
12.8 Anisotropies . . . 1093
12.8.1 Overview . . . 1093
12.8.2 Large-Scale Anisotropies . . . 1098
12.8.3 Acoustic Oscillations . . . 1099
12.8.4 Diffusion Damping . . . 1104
12.8.5 Inhomogeneities to Anisotropies . . . 1107
12.8.6 The Anisotropy Spectrum Today . . . 1114
12.8.7 Cosmological Parameters . . . 1119
12.9 Cosmological perturbations by J. Lesgorgues . . . 1125
12.9.1 Linear cosmological perturbations . . . 1126
12.9.2 CMB temperature anisotropies by Lesgourgues . . . 1135
12.9.3 Matter power spectrum by Lesgourgues . . . 1159
12.10 Probes of Inhomogeneities . . . 1171
12.10.1 Angular Correlations . . . 1172
12.10.2 Peculiar Velocities . . . 1178
12.10.3 Direct Measurements of Peculiar Velocities . . . 1180
12.10.4 Redshift Space Distortions . . . 1183
12.10.5 Galaxy Clusters . . . 1188
12.10.6 Lyth bound (??) . . . 1193
12.11 Weak Lensing and Polarization . . . 1193
12.11.1 Gravitational Distortion of Images . . . 1194
12.11.2 Geodesics and Shear . . . 1197
12.11.3 Ellipticity as an Estimator of Shear . . . 1200
12.11.4 Weak Lensing Power Spectrum . . . 1203
12.11.5 Polarization: The Quadrupole and the Q/U Decomposition . . . 1209
12.11.6 Polarization from a Single Plane Wave . . . 1212
12.11.7 Boltzmann Solution . . . 1217
12.11.8 Polarization Power Spectra . . . 1220
12.11.9 Detecting Gravity Waves . . . 1223

V Problems 1227

13 Typical problems of modern cosmology 1227
13.1 Problems about Homogeneous Dynamics and Thermal History . . . 1227
13.1.1 Mathematical Problems for Basic Preparation to Cosmology . . . 1227
13.1.2 Typical Problems about Friedman Universe (?!??!) . . . 1227
13.1.3 Other Problems about Friedman Universe (?!??!) . . . 1231
13.1.4 Solutions by Dodelson to Chapter 1 . . . 1232

13.1.5 Problems on thermal history (!!?) . . . 1233
13.1.6 Problems on homogeneous inflation 1 . . . 1238
13.1.7 Problems on homogeneous inflation 2 . . . 1241
13.1.8 Problems on Einstein equations . . . 1243
13.2 Problems about Inhomogeneities and Anisotropies . . . 1245
13.2.1 Problems about initial conditions . . . 1245
13.2.2 Problems on inhomogeneities . . . 1251
13.2.3 Problems on Anisotropies . . . 1264
13.2.4 Problems from probes of inhomogeneities by Dodelson (??????) . . . 1270
13.2.5 *Problem Sets* by Baumann's TASI lectures . . . 1272
13.2.6 Solution some problems by Dodelson chapter 2 . . . 1275
13.2.7 Solution some problems by Dodelson chapter 3 . . . 1279
13.2.8 Solution some problems by Dodelson chapter 4 . . . 1280
13.2.9 Solution some problems by Dodelson chapter 5 . . . 1282
13.2.10 Solution some problems by Dodelson chapter 11 . . . 1282

14 Extra Problems about Basic Models 1285
14.1 General questions . . . 1285
14.1.1 Short problems for checking basics of modern cosmology . . . 1285
14.1.2 Short questions and problems for knowledge checking . . . 1286
14.1.3 Problems to Check Deep Understanding of Cosmology . . . 1290
14.1.4 Typical general questions from curious people . . . 1290

15 Problems about Special Modesl 1291
15.0.1 Problems about Weak Lensing and Polarization 1 . . . 1291
15.0.2 Problems about Weak Lensing and Polarization 2 . . . 1293
15.0.3 Problems about Quantum Cosmology (???) . . . 1295

VI Other Topics in Cosmology 1298

16 Stochastic backgrounds of cosmological origin 1298
16.1 22.1 Characteristic frequency of relic GWs . . . 1298
16.1.1 22.2 GW production by classical fields . . . 1300
16.1.2 22.2.1 General formalism . . . 1300
16.1.3 22.2.2 GW generation by a stochastic scalar field . . . 1304
16.1.4 22.3 GWs from preheating after inflation . . . 1307
16.1.5 22.3.1 Parametric resonance in single-field inflation . . . 1307
16.1.6 22.3.2 Tachyonic preheating in hybrid inflation . . . 1309
16.2 22.4 GWs from first-order phase transitions . . . 1310
16.2.1 22.4.1 Crossovers and phase transitions . . . 1310
16.2.2 22.4.2 First-order phase transitions in cosmology . . . 1313
16.2.3 22.4.3 Thermal tunneling theory . . . 1315
16.2.4 22.4.4 Bubble dynamics and GW production . . . 1329
16.3 22.5 Cosmic strings . . . 1337
16.3.1 22.5.1 Global and local strings . . . 1337
16.3.2 22.5.2 Effective description and Nambu-Goto action . . . 1341
16.3.3 22.5.3 String dynamics. Cusps and kinks . . . 1345
16.3.4 22.5.4 Gravitational radiation from cosmic strings . . . 1350
16.4 22.6 Alternatives to inflation . . . 1357
16.5 22.7 Bounds on primordial GW backgrounds . . . 1362
16.5.1 22.7.1 The nucleosynthesis bound . . . 1362
16.5.2 22.7.2 Bounds on extra radiation from the CMB . . . 1368
16.5.3 22.7.3 Bounds from the CMB at large angles . . . 1370

16.5.4 22.7.4 Limits on stochastic backgrounds from interferometers . . . 1370
16.6 Further reading . . . 1371

17 Special Models of Cosmology 1374
17.1 Cosmic Web . . . 1374
17.2 Special Models of Cosmology from Landau's Field Theory (??) . . . 1375
17.3 Other Special Models . . . 1375
17.3.1 Press–Schechter formalism (???) . . . 1375
17.3.2 Lyman-alpha forest (???) . . . 1376

18 Dark Matter Candidates 1376

19 Primordial Gravitational Waves 1376

20 Black Holes in Cosmology 1376

21 Quantum Cosmology: Basics 1376

22 Cosmology from Philosophical Point of View 1376
22.0.1 Other, Exotic Questions . . . 1376

23 Unfinished and Wrong Cosmological Theories 1377
23.1 Theories that Living Creatures Influence Universe (???) . . . 1377
23.2 Astrology (???) . . . 1377
23.3 What Could be in the Universe if Something Was Different? . . . 1377

24 Cosmological constant by Jérôme Martin 1377
24.1 Introduction . . . 1377
24.2 The CC . . . 1380
24.3 The Classical Cosmological Constant Problem . . . 1382
24.3.1 Phase Transition . . . 1382
24.3.2 The Electro-weak Phase Transition . . . 1383
24.4 The Quantum-Mechanical Cosmological Constant Problem . . . 1390
24.4.1 The Zero-Point Energy Density . . . 1390
24.4.2 The Vacuum and the Feynman Diagrams . . . 1395
24.4.3 The Vacuum in Presence of Interactions . . . 1397
24.5 The Bubble Diagrams . . . 1398
24.5.1 Quantum Mechanics as a Field Theory . . . 1399
24.5.2 Perturbations Theory in Quantum Mechanics . . . 1402
24.5.3 Quantum Mechanics and Bubble Diagrams . . . 1406
24.5.4 Quantum Mechanics and Vacuum Energy . . . 1408
24.6 The Gaussian Effective Potential . . . 1410
24.6.1 The Gaussian Effective Potential in Quantum Mechanics . . . 1410
24.6.2 Comparison with the One Loop Effective Potential . . . 1412
24.6.3 The Gaussian Effective Potential in Field Theory and the Vacuum Energy . . . 1416
24.6.4 Vacuum Energy and Gaussian Effective Potential . . . 1422
24.7 Contribution from Other Fields . . . 1423
24.7.1 Fermion Fields . . . 1423
24.7.2 Proca Fields . . . 1427
24.7.3 Mass-less Vector Field . . . 1430
24.7.4 Application: Super-Symmetry . . . 1431
24.8 The Vacuum Energy Density in Curved Space-time . . . 1440
24.9 The Value of the Cosmological Constant . . . 1445
24.10 Measuring the Cosmological Constant in Cosmology . . . 1446
24.10.1 The Accelerating Universe . . . 1447

24.10.2 The Hubble Diagram . . . 1448
24.11 Measuring the Cosmological Constant Elsewhere Than In Cosmology? . . . 1452
24.11.1 The Static and Spherically Symmetric Gravitational Field in Presence of a Cosmological Constant . . . 1452
24.11.2 Planets Orbits . . . 1454
24.11.3 Modifications of the Atomic Levels . . . 1458
24.12 Do the Vacuum Fluctuations Really Exist? . . . 1463
24.12.1 The Lamb Shift . . . 1463
24.12.2 The Casimir Effect . . . 1466
24.13 Do the Quantum Fluctuations Gravitate? . . . 1477
24.14 The Weak Equivalence Principle in Quantum Mechanics . . . 1481
24.14.1 The Collela, Overhausser and Werner (COW) Experiment . . . 1481
24.14.2 The Quantum Galileo Experiment . . . 1486
24.15 Conclusions . . . 1497

VII Experimental Tests 1499

25 Tests of Well-Known Properties of the Universe 1499
25.1 Test that Universe is Expanding . . . 1499
25.2 Test that Universe is Approximately Flat . . . 1499

26 Some Mathematical Methods for Analysis of Experiments 1499
26.1 Analysis . . . 1499
26.1.1 the Likelihood Function . . . 1499
26.1.2 Signal Covariance Matrix . . . 1505
26.1.3 Estimating the Likelihood Function . . . 1514
26.1.4 the Fisher Matrix: Limits and Applications . . . 1524
26.1.5 Mapmaking and Inversion . . . 1529
26.1.6 Systematics . . . 1532

27 Planck 2018 results. VI. Cosmological parameters 1541
27.1 Introduction . . . 1542
27.2 Methodology and likelihoods . . . 1544
27.2.1 Theoretical model . . . 1544
27.2.2 Power spectra and likelihoods . . . 1545
27.2.3 The CMB lensing likelihood . . . 1556
27.3 Constraints on base ΛCDM . . . 1559
27.3.1 Acoustic scale . . . 1560
27.3.2 Hubble constant and dark-energy density . . . 1563
27.3.3 Optical depth and the fluctuation amplitude . . . 1563
27.3.4 Scalar spectral index . . . 1565
27.3.5 Matter densities . . . 1565
27.3.6 Changes in the base-ΛCDM parameters between the 2015 and 2018 data releases 1567
27.4 Comparison with high-resolution experiments . . . 1568
27.5 Comparison with other astrophysical data sets . . . 1570
27.5.1 Baryon acoustic oscillations . . . 1570
27.5.2 Type Ia supernovae . . . 1571
27.5.3 Redshift-space distortions . . . 1572
27.5.4 The Hubble constant . . . 1573
27.5.5 Weak gravitational lensing of galaxies . . . 1575
27.5.6 Galaxy clustering and cross-correlation . . . 1576
27.5.7 Cluster counts . . . 1578
27.6 Internal consistency of ΛCDM model parameters . . . 1579

27.6.1 Consistency of high and low multipoles . . . 1579
27.6.2 Lensing smoothing and $A_{\rm L}$ . . . 1581

28 Planck 2018 results. X. Constraints on inflation 1599
28.1 Introduction . . . 1600
28.2 Methodology and data . . . 1602
28.2.1 Cosmological models and inference . . . 1602
28.2.2 Data . . . 1604
28.3 *Planck* 2018 results for the main inflationary observables . . . 1608
28.3.1 Results for the scalar spectral index . . . 1609
28.3.2 Ruling out $n_{\rm s} = 1$ . . . 1611
28.3.3 Constraints on the scale dependence of the scalar spectral index . . . 1612
28.3.4 Constraints on spatial curvature . . . 1613
28.3.5 Constraints on the tensor-to-scalar ratio . . . 1615
28.3.6 Beyond the tensor-to-scalar ratio consistency condition . . . 1616
28.4 Implications for single-field slow-roll inflationary models . . . 1618
28.4.1 Constraints on slow-roll parameters . . . 1619
28.4.2 Implications for selected slow-roll inflationary models . . . 1621
28.5 Reconstruction of the inflaton potential . . . 1625
28.5.1 Taylor expansion of $V(\phi)$ in the observable region . . . 1625
28.5.2 Taylor expansion of $H(\phi)$ in the observable region . . . 1628
28.5.3 Taylor expansion of full $V(\phi)$ . . . 1630
28.5.4 Free-form potential reconstruction . . . 1632
28.6 Primordial power spectrum reconstruction . . . 1635
28.6.1 Penalized likelihood . . . 1635
28.6.2 Bayesian reconstruction . . . 1639
28.6.3 Cubic spline reconstruction . . . 1644
28.7 Search for primordial features in the *Planck* power spectrum . . . 1649
28.7.1 Power spectrum templates with features . . . 1649
28.7.2 Data analysis . . . 1651
28.7.3 Feature candidates and their evidence . . . 1651
28.7.4 Axion monodromy . . . 1656
28.8 Combined power spectrum and bispectrum analysis for oscillatory features . . . 1658
28.8.1 Approach . . . 1658
28.8.2 Models . . . 1660
28.8.3 Data analysis . . . 1661
28.8.4 Estimating significance . . . 1661
28.9 Constraints on isocurvature fluctuations . . . 1665
28.9.1 Background and modeling . . . 1666
28.9.2 Results for generally correlated adiabatic and isocurvature modes . . . 1669
28.9.3 Role of lensing parameter $A_{\rm L}$ and likelihood choices . . . 1671
28.9.4 Specific CDI models . . . 1674
28.9.5 Compensated BDI–CDI mode . . . 1677
28.10 Constraints on anisotropic models of inflation . . . 1682
28.10.1 Dipolar asymmetry . . . 1682
28.10.2 Quadrupolar asymmetry . . . 1685
28.11 Conclusions . . . 1688

VIII Consequences of High Energy Physics on Cosmology 1691

29 Inflation in String Theory by Baumann's 5th lecture **1691**
29.1 Why Combine Two Speculative Ideas? . . . 1691
29.2 UV Sensitivity of Inflation . . . 1691
29.2.1 Effective Field Theory and Inflation . . . 1691
29.2.2 The Eta Problem . . . 1693
29.2.3 The Lyth Bound . . . 1694
29.2.4 Super-Planckian Fields and Flat Potentials . . . 1695
29.3 Inflation in String Theory . . . 1696
29.3.1 From String Compactifications to the Inflaton Action . . . 1696
29.3.2 *Case Study:* Warped D-brane Inflation . . . 1697
29.3.3 *Case Study:* Axion Monodromy Inflation . . . 1703
29.4 Outlook . . . 1706
29.4.1 Theoretical Prospects . . . 1706
29.4.2 Observational Signatures? . . . 1707

30 Scalar Speed Limits and Cosmology: Acceleration from D-cceleration by Silverstein, Tong **1707**
30.1 Introduction . . . 1708
30.2 The Basic Setup . . . 1710
30.2.1 The System at Weak Coupling . . . 1710
30.2.2 The System at Strong Coupling and the Speed Limit . . . 1711
30.3 The Global CFT . . . 1712
30.3.1 Effective Action and Approximation Scheme . . . 1712
30.3.2 Dynamics . . . 1713
30.3.3 Antibranes . . . 1715
30.4 Coupling to Gravity . . . 1716
30.4.1 CFT Coupled to Gravity . . . 1717
30.4.2 QFT with Mass Gap Coupled to Gravity . . . 1722
30.5 Cosmological Solutions . . . 1723
30.5.1 Cosmology in the AdS Throat . . . 1724
30.5.2 Cosmology in the Cut-Off Throat . . . 1728
30.5.3 Curvature Corrections . . . 1730
30.5.4 Relation to other Works . . . 1731
30.6 Perturbations and Particle Production . . . 1732
30.6.1 W boson and String Production . . . 1732
30.6.2 Perturbations of ϕ . . . 1733

IX Appendix **1735**

A Introduction and Overview of Cosmology **1735**
A.1 Another Motivation for Cosmology . . . 1735
A.1.1 Other General Motivation . . . 1735
A.1.2 Motivation for Exotic Theories in Cosmology . . . 1736
A.1.3 Amazing Facts . . . 1736
A.2 Mindset of a Professional in Cosmology . . . 1736
A.2.1 Isn't 21st Century Cosmology Bullshit? (????) . . . 1736
A.2.2 A Cosmologist's View of the World (???) . . . 1736
A.2.3 How to Derive and Solve Equations in Cosmology? . . . 1737
A.2.4 Equations in Cosmology: Important Features of Particular Cases . . . 1737
A.2.5 Ways to Earn Money by Knowing Cosmology and Most Used Topics (??) . . . 1738
A.2.6 Overview of Misconceptions About Cosmology (??) . . . 1738
A.2.7 General Approach to a Research in Cosmology . . . 1740
A.2.8 Ways to Guess the Main Ideas (???) . . . 1741

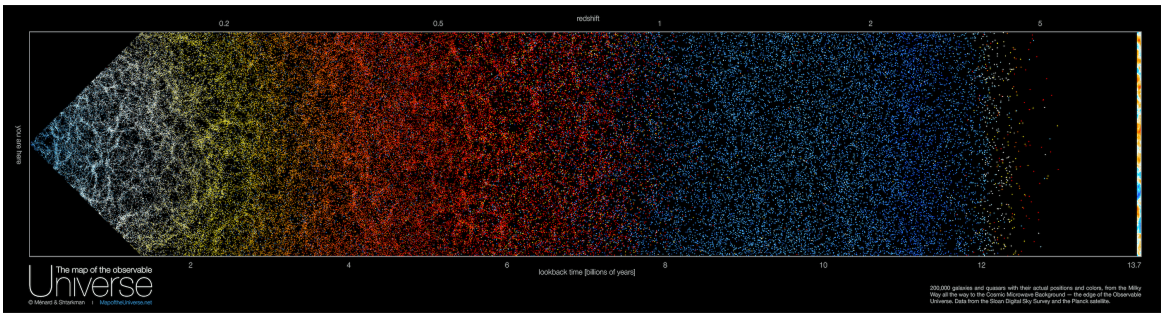
A.2.9 Effective Studying of Cosmology . . . 1741
A.3 Acknowledgements . . . 1741
A.4 Literature on Cosmology . . . 1741
A.4.1 Main Literature . . . 1741
A.4.2 Articles on Different Effects . . . 1742
A.4.3 Additional, Popular, Literature for Help . . . 1744
A.4.4 Suggested Readings from Good Books . . . 1745
A.5 History of Cosmology in a Nutshell (???) . . . 1747
A.6 Overview of Cosmology . . . 1748
A.6.1 Popular Overview of Cosmology . . . 1748
A.6.2 Results of Cosmology . . . 1750
A.6.3 Applications of Cosmology in Other Branches of Physics . . . 1750
A.6.4 Review of Further Developments in Cosmology . . . 1750
A.6.5 What Is Going on in Typical Sections of Cosmology . . . 1750
A.6.6 Some Potential Applications . . . 1750
A.7 Description of the Note . . . 1750
A.7.1 Description of Chapters and Sections . . . 1750
A.7.2 Designations and Constants . . . 1751
A.8 Cosmological Puzzles for Different Cases . . . 1751
A.8.1 Puzzles about Construction of Cosmology . . . 1751
A.8.2 General Questions about the Universe . . . 1751
A.8.3 General Questions about What Cosmologists are Doing . . . 1752
A.8.4 Questions for Those Who Say They Have Mastered Cosmology . . . 1752

B Some Physics for Cosmology 1752
B.0.1 Some General Theory of Relativity . . . 1752
B.0.2 Some Statistical Physics . . . 1752
B.0.3 Some Nuclear Physics . . . 1752
B.0.4 Some Particle Physics . . . 1753
B.0.5 Some Other Physics . . . 1753

C Some Mathematics for Cosmology 1753

D Bibliography 1754
D.0.1 From Lectures . . . 1754
D.0.2 From Famous Reviews . . . 1786
D.0.3 From Experimental Articles . . . 1869

Preface and main motivation



Special Topics in a Nutshell

Fundamentals of Cosmology

Properties of the Universe in Detail

Problems

Other Topics in Cosmology

Experimental Tests

Consequences of High Energy Physics on Cosmology

This note contains a discussion of the cosmological constant, detailed CMB overview, description of quantum bubbles during cosmological phase transitions. Also, there is a popular introduction. I used the following literature.

Let's discuss some minimum knowledge and motivation that would be good to understand for studying the subject.

А как на самом деле устроена Вселенная?

(раскрою мысль, что интересно узнать природу Вселенной)

Что на самом деле было давно в прошлом?

(раскрою мысль, что интересно узнать природу Вселенной)

Квантовая космология - фундаментально очень интересная и возможно верная модель устройства мира в самом его "начале"

Особенно в 2020х годах можно многое узнать про это без особых усилий, потому что это типичная тема многих статей. Квантовая космология - то, где наше интуитивное представление о мире перестает работать и там происходит казалось бы абсолютно непонятно что. Однако есть много исследований про это, и все-таки интересно понять, что же там было. Конечно, ожидать, что эти вопросы будут решены или за пару месяцев станут понятными, абсолютно не разумно, тем не менее, подумать про это особо интересно. В целом, это прекрасное хобби любого физика - подумать про квантовую космологию на досуге. Ну и плюс если что-то там откроешь - то это очень будет актуальным открытием.

Рассказывая о большинстве предметах, можно найти их важные связи и применения в космологии, что очень круто дополнит рассказ, потому что будет казаться, что ты знаешь что-то о глобальном

(распишу потом, что очень круто о таких приложениях говорить и например, о чем можно упоминать! вот например - тут применяется, такой эффект. И такого в космологии полно, редкий предмет в физике не применяется в ней. Классическая механика может быть, и то философский подход такой же)

Головоломки для мотивации

(??? что-то пока не знаю даже, что придумать тут такое для начала? потом напишу, когда нарешаюсь задач.)

С чего начался мир?

(типичный вопрос, который очень мотивирует ответить)

(тут абзац будет, типа мы не знаем, но вот там в доли секунд мы еще знаем что-то. только вот по факту экспериментальное подтверждение - тема на 100 страниц и пары месяцев раздумий про это.)

Какой возраст Вселенной и почему он такой? (????????)

(я не знаю,почему он такой, но это мотивирует учить космологию.)

Derive the age of Universe (??????)

(??? still I have no idea)

Что было 10 миллиардов лет назад?

(я не знаю, что было, но это мотивирует учить космологию.)

Что будет со Вселенной в будущем?

Part I
Typical Cosmology in a Nutshell

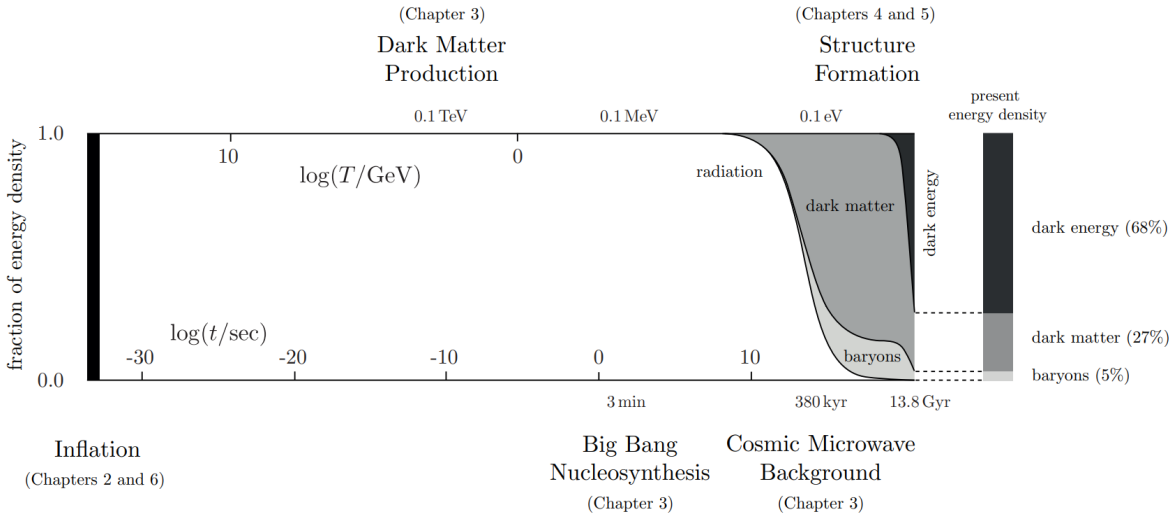
1 Main concepts of modern cosmology

1.1 Most believed properties of the universe

(очень тоже важные указания на то, какой вообще моделью пользоваться и почему? много мнений, много философских соображений, когда-то и я напишу, пока что не спрашивают меня, ну и не до этого.)

1.1.1 General history of the universe in scope of modern cosmology

общая хронология



t	$\rho^{1/4}$	Event
10^{-42} s	10^{18} GeV	Inflation begins?
$10^{-32\pm6}$ s	$10^{13\pm3}$ GeV	Inflation ends, Cold Big Bang begins?
$10^{-18\pm6}$ s	$10^{6\pm3}$ GeV	Hot Big Bang begins?
10^{-10} s	100 GeV	Electroweak phase transition?
10^{-4} s	100 MeV	Quark-hadron phase transition?
10^{-2} s	10 MeV	γ, ν, e, $\bar{e}$, n, and p in thermal equilibrium
1 s	1 MeV	ν decoupling, $e\bar{e}$ annihilation.
100 s	0.1 MeV	Nucleosynthesis (BBN)
10^4 yr	1 eV	Matter-radiation equality
10^5 yr	0.1 eV	Atom formation, photon decoupling (CMB)
$\sim 10^9$ yr	10^{-3} eV	First bound structures form
Now	10^{-4} eV (2.73 K)	The present.

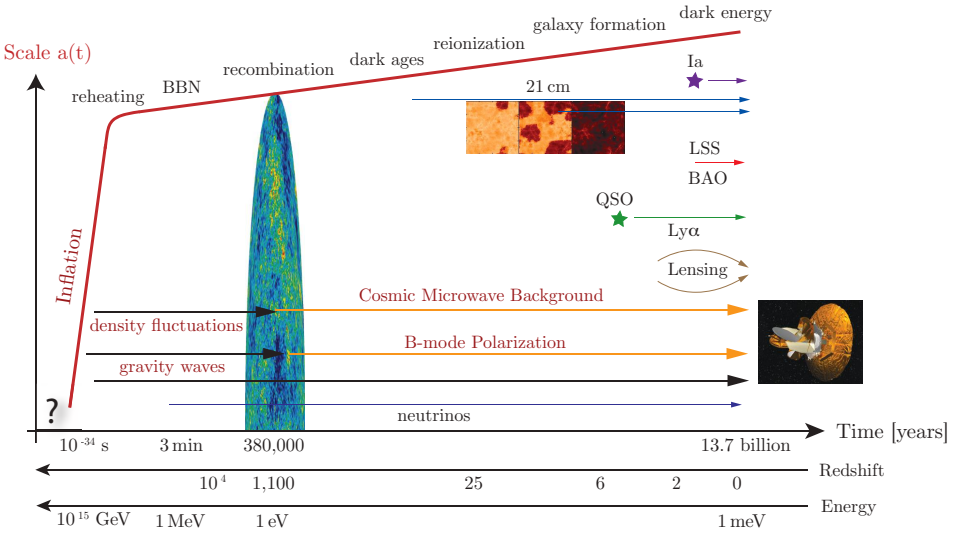
Event	time t	redshift z	temperature T
Inflation	10^{-34} s(?)	–	–
Baryogenesis	?	?	?
EW phase transition	20ps	10^{15}	100GeV
QCD phase transition	20μs	10^{12}	150MeV
Dark matter freeze-out	?	?	?
Neutrino decoupling	1 s	$6 \cdot 10^9$	1MeV
Electron-positron annihilation	6 s	$2 \cdot 10^9$	500keV
Big Bang nucleosynthesis	3 min	$4 \cdot 10^8$	100keV
Matter-radiation equality	60kyr	3400	0.75eV
Recombination	$260 - 380$kyr	$1100 - 1400$	$0.26 - 0.33$eV
Photon decoupling	380kyr	$1000 - 1200$	$0.23 - 0.28$eV
Reionization	$100 - 400$Myr	$11 - 30$	$2.6 - 7.0$meV
Dark energy-matter equality	9 Gyr	0.4	0.33meV
Present	13.8Gyr	0	0.24meV

Epoch	Temperature	Characteristics
Big Bang 0 s	∞K ∞eV	singularity (vacuum fluctuation?)
Planck 0 s $< t \leq 10^{-43}$ s	$> 10^{40}$ K $> 10^{36}$eV	quantum gravity
Grand Unification 10^{-43} s $\leq t \leq 10^{-36}$ s	$10^{36} - 10^{40}$ K $10^{26} - 10^{32}$eV	gravity freezes out the "grand unified force" (GUT)
Inflationary 10^{-36} s $\leq t \leq 10^{-32}$ s	$10^{33} - 10^{36}$ K $10^{29} - 10^{32}$eV	inflation begins strong force freezes out
Electroweak 10^{-32} s $\leq t \leq 10^{-12}$ S	$10^{20} - 10^{33}$ K $10^{16} - 10^{29}$eV	weak force freezes out 4 distinct forces(EM dominates) baryogenesis: baryons and antibaryons annihilate
Quark 10^{-12} s $\leq t \leq 10^{-6}$ s	$10^{16} - 10^{20}$ K $10^{12} - 10^{16}$eV	universe contains hot quark-gluon plasma: quarks, gluons and leptons
Hadron 10^{-6} s $\leq t \leq 1$ s	$10^{12} - 10^{16}$ K $10^{8} - 10^{12}$eV	quarks and gluons bind into hadrons
Lepton 1 s $\leq t \leq 3$ min	$10^{10} - 10^{12}$ K $10^{6} - 10^{8}$eV	universe contains photons (γ), muons ($\mu^{\pm}$), electrons/positrons ($e^{\pm}$), and neutrinos ($\nu, \bar{\nu}$); nucleons n and p in equal numbers
1 s	$\leq 10^{12}$ K $\leq 10^{8}$eV	$\mu^+$and μ^-annihilate; ν and $\bar{\nu}$ decouple; $e^{\pm}, \gamma$ and nucleons remain. Reactions: $e^+ + n \rightleftharpoons p + \nu_e$ $e^- + p \rightleftharpoons n + \nu_e$ $n \rightarrow p + e^- + \bar{\nu}_e$
100 s	10^{10} K, 10^{6}eV	e^+ and e^-annihilate

(!!!!!!!!!!!!! тут напишу важное пояснение, что означает каждый важный процесс из таблицы!)

Reionization is the process that caused electrically neutral atoms in the universe to reionize after the lapse of the "dark ages". It is the second of two major phase transitions of gas in the universe[citation needed] (the first is recombination). While the majority of baryonic matter

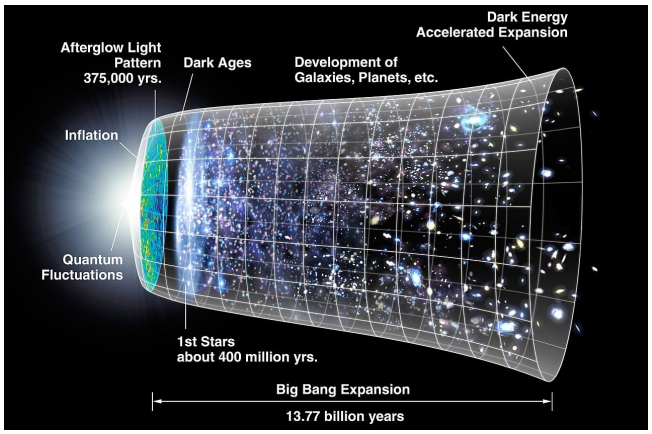
in the universe is in the form of hydrogen and helium, reionization usually refers strictly to the reionization of hydrogen, the element.



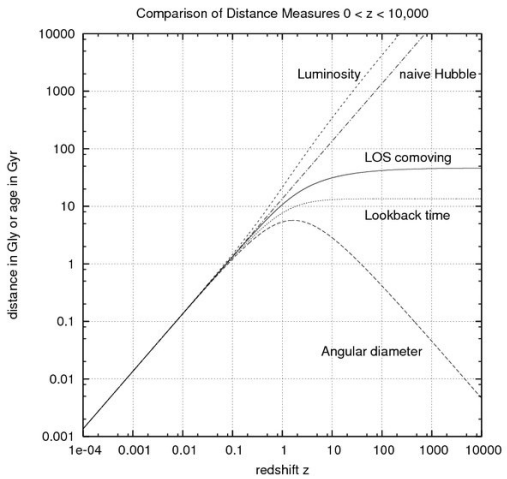
Отдельно подчеркну, что:

• Вначале у Вселенной было очень много энергии, и много света. Как-то появились протоны и электроны, возможно, от инфлятонов. Дальше они собирались в другую материю.

• CMB образовался в именно эпоху доминации материи, не на границе перехода к ней.



Время Вселенной по красоному смещению:



(потом еще додумаю этот график, в целом, можно им и пользоваться)

A Thermal-Particle History of the universe

- Baryogenesis. Particles and antiparticles annihilate through $e^+ + e^- \to \gamma + \gamma$. If initially the universe was filled with equal amounts of matter and anti-matter then we expect these annihilations to lead to a universe dominated by radiation. However, we do observe an overabundance of matter (mostly baryons) over anti-matter in the universe today. Models of baryogenesis try to derive the observed baryon-to-photon ratio

$$\eta := \frac{n_b}{n_\gamma} \sim 10^{-9}$$

from some dynamical mechanism, i.e. without assuming a primordial matter-antimatter asymmetry as an initial condition. Although many ideas for baryogenesis exist, none is singled out by experimental tests.

(!?!?! добавлю ключевую формулу или график!)

- Electroweak phase transition. At 100GeV particles receive their masses through the Higgs mechanism. (????)

(!?!?! добавлю ключевую формулу или график!)

- QCD phase transition to baryons. Below 150MeV the strong interactions between the quarks and the gluons become important (before was only weak interaction). Quarks and gluons then form bound three-quark systems (baryons), and quark-antiquark pairs (mesons). These baryons and mesons are the relevant degrees of freedom below the scale of the QCD phase transition.

(!?!?! добавлю ключевую формулу или график!)

- Dark matter freeze-out. Since dark matter is very weakly interacting with ordinary matter we expect it to decouple relatively early on. In §3.3.2, we will study the example of WIMPs - weakly interacting massive particles that freeze out around 1MeV. We will show that choosing natural values for the mass of the dark matter particles and their interaction cross section with ordinary matter reproduces the observed relic dark matter density surprisingly well.

(!?!?! добавлю ключевую формулу или график!)

- Neutrino decoupling. Neutrinos only interact with the rest of the primordial plasma through the weak interaction. The estimate in (3.1.10) therefore applies and neutrinos decouple at 0.8MeV.

(!?!?! добавлю ключевую формулу или график!)

- Electron-positron annihilation. Electrons and positrons annihilate shortly after neutrino decoupling. The energies of the electrons and positrons gets transferred to the photons, but not the neutrinos. In §3.2.4, we will explain that this is the reason why the photon temperature today is greater than the neutrino temperature.

(!?!?! добавлю ключевую формулу или график!)

- Big Bang nucleosynthesis. Around 3 minutes after the Big Bang, the light elements were formed. In §3.3.4, we will study this process of Big Bang nucleosynthesis (BBN).

(!?!?! добавлю ключевую формулу или график!)

- Recombination. Neutral hydrogen forms through the reaction $e^- + p^+ \to \mathrm{H} + \gamma$ when the temperature has become low enough that the reverse reaction is energetically disfavoured.

At temperatures above about 1 eV, the universe still consisted of a plasma of free electrons and nuclei. Photons were tightly coupled to the electrons via Compton scattering, which in turn strongly interacted with protons via Coulomb scattering. There was very little neutral hydrogen. When the temperature became low enough, the electrons and nuclei combined to form neutral atoms (recombination20), and the density of free electrons fell sharply. The photon mean free path grew rapidly and became longer than the horizon distance. The photons

decoupled from the matter and the universe became transparent. Today, these photons are the cosmic microwave background.

(!?!?! добавлю ключевую формулу или график!)

- Photon decoupling. Before recombination the strongest coupling between the photons and the rest of the plasma is through Thomson scattering, $e^- + \gamma \to e^- + \gamma$. The sharp drop in the free electron density after recombination means that this process becomes inefficient and the photons decouple. They have since streamed freely through the universe and are today observed as the cosmic microwave background (CMB).

(!?!?! добавлю ключевую формулу или график!)

Main temperature properties of the universe

The temperature of the universe 1 second after the Big Bang was about 1MeV, and evolved as $t^{-1/2}$ before that. Because $\frac{T}{1\text{MeV}} \simeq 1.5 g_\star^{-1/4} \left(\frac{1\text{sec}}{t}\right)^{1/2}$. (?? допишу, почему идейно??)

Formation of Solar System

There is evidence that the formation of the Solar System began about 4.6 billion years ago with the gravitational collapse of a small part of a giant molecular cloud.[1] Most of the collapsing mass collected in the center, forming the Sun, while the rest flattened into a protoplanetary disk out of which the planets, moons, asteroids, and other small Solar System bodies formed.

(очень коротко напишу, нужно будет - вдали уже есть раздел про это)

Empirical Evidence for main properties of the universe (???!!!)

(очень хорошая идея их написать тут, позже сделаю)

- The CMBR: probes the recombination era, $t \sim 10^5$yr, $z \sim 1100$, based on a well understood atomic and macroscopic physics

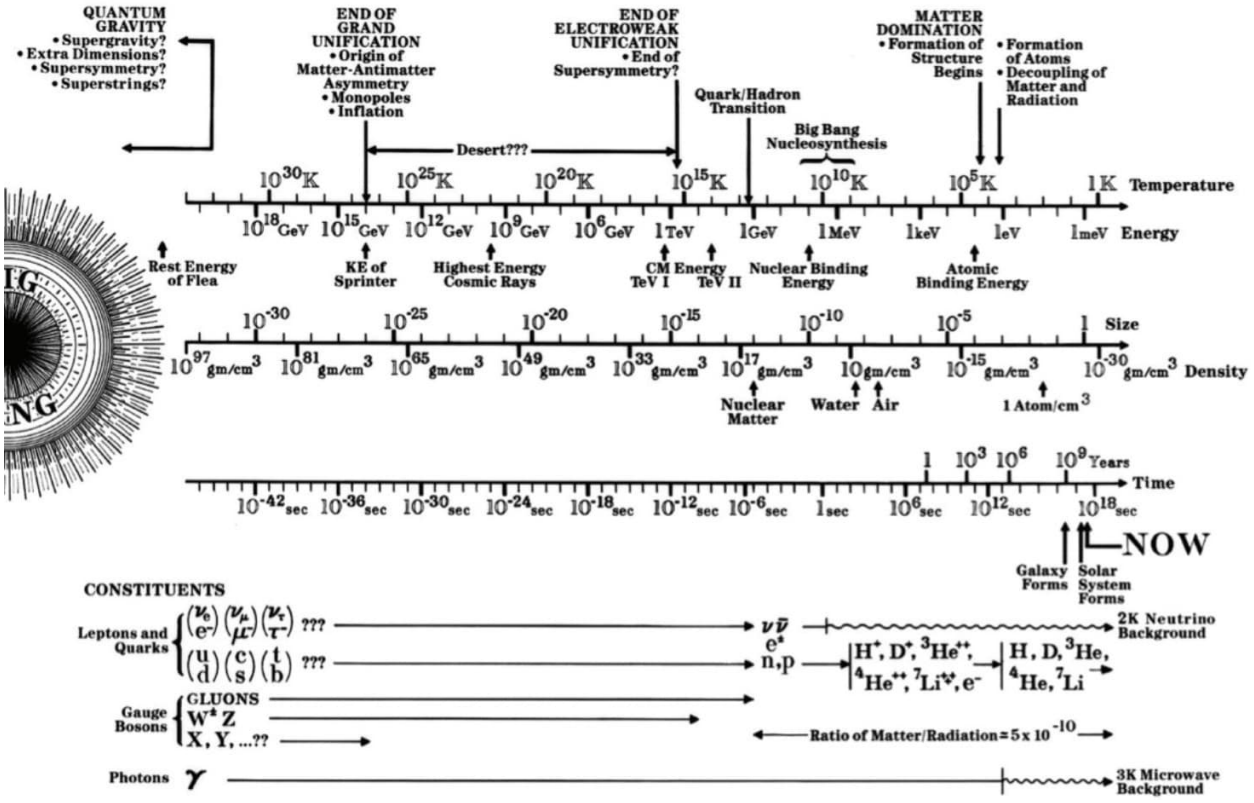
- Nucleosynthesis: probes the $t \sim 10^{-3} - 10^2$ sec era, $z \sim 10^9$, compare the model predictions with observed abundances of the lightest elements, based on a well understood nuclear physics

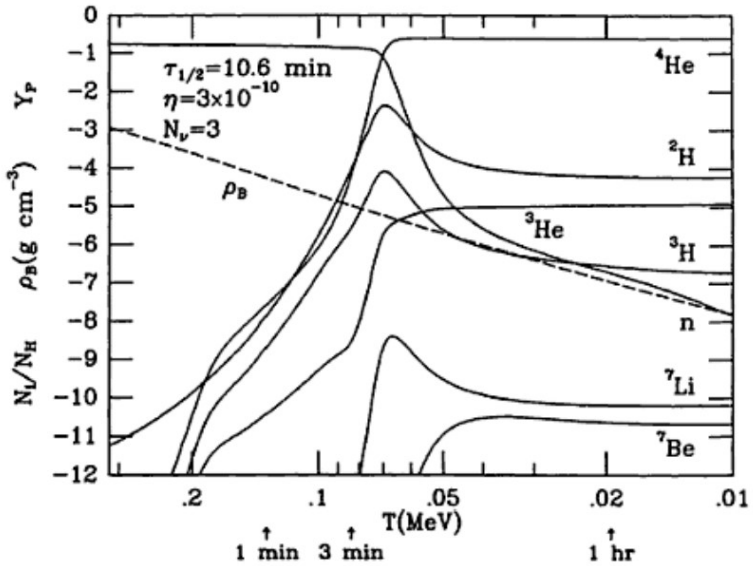
- Matter-antimatter asymmetry: probes the baryogenesis era, $t \sim 10^{-6}$sec, $z \sim 10^{12}$, but only in suggesting that some symmetry breaking did occur

- Predictions of the inflationary scenario: flatness, uniformity of CMBR, absence of monopoles, the right type of density fluctuation spectrum - it all supports the idea that inflation did happen, but does not say a lot about its detailed physics

> Cosmological observations can indicate or constrain physics well outside the reach of laboratory experiments

History of nucleosynthesis





- We will use this to derive the evolution of T as different particle species annihilate in thermal equilibrium

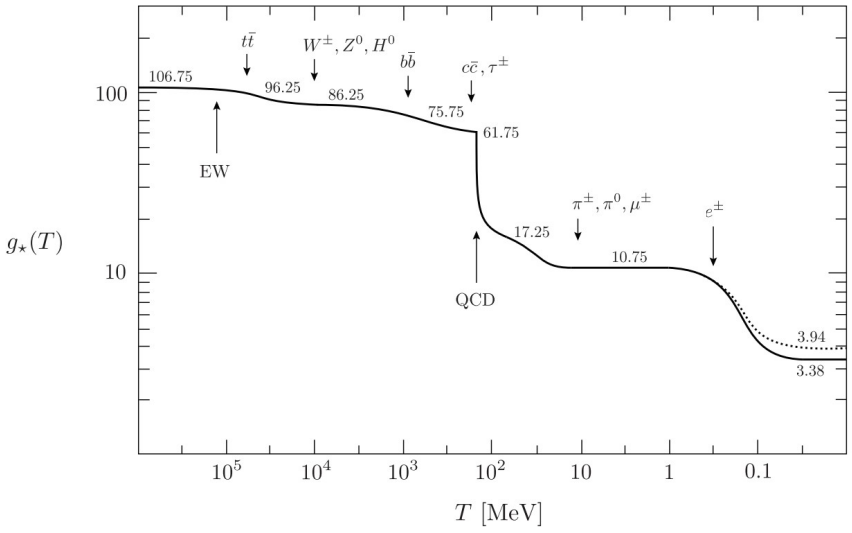
- Setting the entropy density before and after is equivalent to setting

$$g_* T^3\big|_{\text{initial}} = g_* T^3\big|_{\text{final}}$$

- When particle species disappear through annihilation, they dump their entropy into the remaining species and hence raise the temperature

By way of summarizing the features of an expanding universe that I have outlined above and that we will explore in great detail in the coming chapters, let's construct a time line. We

can characterize any epoch in the universe by the time since the Big Bang; by the value of the scale factor at that time; or by the temperature of the cosmic plasma. For example, today, $a = 1; t \simeq 14$ billion years; and $T = 2.725\ \mathrm{K} = 2.35 \cdot 10^{-4}\mathrm{eV}/k_B$.

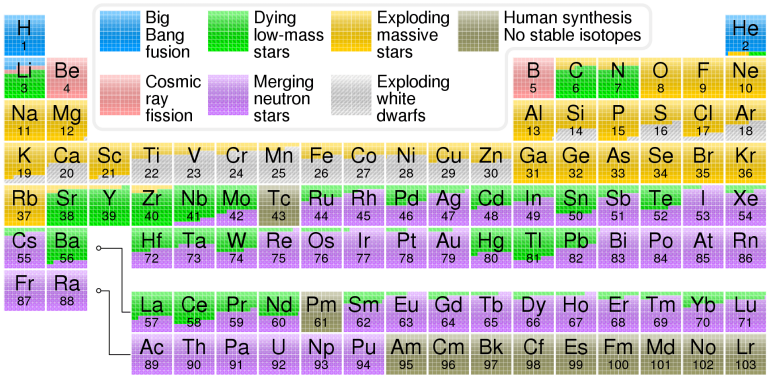


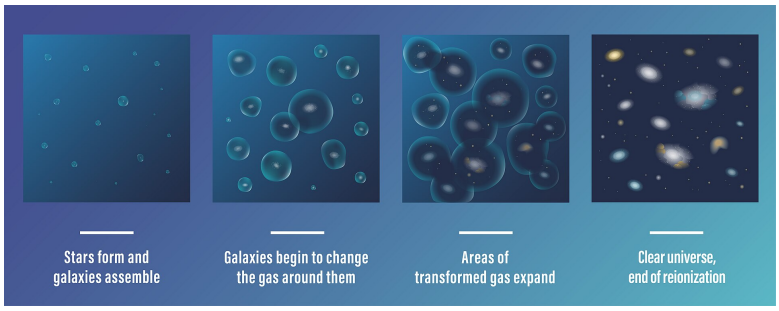
(??? почему график такой - видимо, будет неотвеченным вопросом еще долго, не актуальный вопрос.)

Early on, most of the energy in the universe was in the form of radiation. Eventually, since the energy of a relativistic particle falls as $1/a$ while that of nonrelativistic matter remains constant at m, matter overtook radiation. At relatively recent times, the universe appears to have become dominated not by matter, but by some dark energy, whose density remains relatively constant with time. The evidence for this unexplained form of energy is new and certainly not conclusive, but it is very suggestive.

Light elements formed when the universe was several minutes old, and the CMB decoupled from matter at a temperature of order $k_B T \sim 1/4\mathrm{eV}$. Heavy elementary particles may make up the dark matter in the universe; if they do, their abundance was fixed at very high temperatures of order $k_B T \sim 100\mathrm{GeV}$.

The early end of the time line allows for a brief period of inflation, during which primordial perturbations were produced. These small perturbations began to grow when the universe became dominated by matter. The dark matter grew more and more clumpy, simply because of the attractive nature of gravity. An overdensity of dark matter of 1 part in 1000 when the temperature was 1eV grew to 1 part in 100 by the time the temperature dropped to 0.1eV. Eventually, at relatively recent times, perturbations in the matter ceased to be small; they became the nonlinear structure we see today. Anisotropies in the CMB today tell us what the universe looked like when it was several hundred thousand years old, so they are wonderful probes of the perturbations.



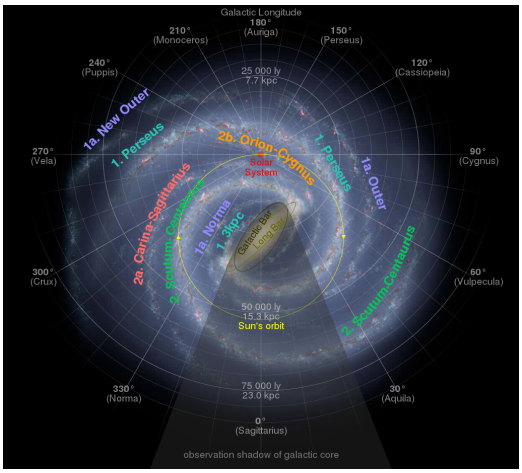


1.1.2 Properties and objects of the current universe

(пока это не пишу, отдельный курс Калтеха про это, нужно будет - недели просижу и доучу)

(об этом пока в записи про астрофизику!)

Overview



Large Scales structures

(см. лекцию Калтеха или мою запись про галлактики, там полно всякого, пока не до этого)

Idea of correlation between galaxies (??)

(вот это важно дописать когда-то будет, пока примерно это понимаю, но уже читал про это)

Voids (????)

(напишу про результаты наблюдений и на чем остановилось объяснение этого???)

1.1.3 Determination of properties of the universe in a nutshell

What is the age of the universe and why? (!??!)

(подробно про то, почему это так?)

There are at least 3 ways that the age of the universe can be estimated:

The age of the chemical elements.

The age of the oldest star clusters.
The age of the oldest white dwarf stars.
(see https://www.astro.ucla.edu/~wright/age.html)

1.1.4 Concept of shape, expansion of the universe, dark mater, dark energy

(самая суть, про все это подробнее ниже будет еще больше указаний)

What is a shape and curvature of universe?

What is a dark energy (Λ)?

What is a dark matter?

Concept of expansion of the universe

Почему космологи считают, что Вселенная расширяется, и почему?

(??? напишу список фактов, приведу также графики из экспериментов для них!)

что сейчас заставляет Вселенную расширяться?

(??? а что сейчас заставляет Вселенную расширяться???? хз, что-то не знаю. сам вакуум - то есть космо постоянная что ли???)

Как это, расширение со сверхсветовой скоростью? (????)

(допишу потом, странно ведь это, да???)

Other important general questions about the universe

Why universe don't fall into one point?

(отвечу потом нормально)

1.1.5 Main Philosophical Ideas about Universe

(main ideas and examples)

What is gravity in a context of universe?

(main statements, formulas and experimental facts)

О границах нашего познания (???)

Возможна ли рациональная космология как наука (???)

(хз, мб и нет, просто тоже важный вопрос)

Принципы метрологии

(услышал, что они важные, иначе херней будешь заниматься, думая, что что-то научное делаешь)

On anthropic principle

(напишу что-то, интересно, но вряд ли это на самом деле важно)
(вики пока)

The anthropic principle (aka the "observation selection effect") is the hypothesis, first proposed in 1957 by Robert Dicke, that the range of possible observations that could be made about the universe is limited by the fact that observations could happen only in a universe capable of developing intelligent life. Proponents of the anthropic principle argue that it explains why the universe has the age and the fundamental physical constants necessary to accommodate conscious life, since if either had been different, no one would have been around to make observations. Anthropic reasoning is often used to deal with the idea that the universe seems to be finely tuned for the existence of life.

There are many different formulations of the anthropic principle. Philosopher Nick Bostrom counts them at thirty, but the underlying principles can be divided into "weak" and "strong" forms, depending on the types of cosmological claims they entail. The weak anthropic principle (WAP), as defined by Brandon Carter, states that the universe's ostensible fine tuning is the result of selection bias (specifically survivorship bias). Most such arguments draw upon some notion of the multiverse for there to be a statistical population of universes from which to select. However, a single vast universe is sufficient for most forms of the WAP that do not specifically deal with fine tuning. Carter distinguished the WAP from the strong anthropic principle (SAP), which considers the universe in some sense compelled to eventually have conscious and sapient life emerge within it. A form of the latter known as the participatory anthropic principle, articulated by John Archibald Wheeler, suggests on the basis of quantum mechanics that the universe, as a condition of its existence, must be observed, thus implying one or more observers. Stronger yet is the final anthropic principle (FAP), proposed by John D. Barrow and Frank Tipler, which views the universe's structure as expressible by bits of information in such a way that information processing is inevitable and eternal.

On metaphysic ideas in cosmology

(их пока не знаю, но очень много наберется, уверен.)

1.1.6 Validity of modern cosmology, tests, directions of improvement (!?)

(Про то, "что и почему не могло бы быть?")

(это подробно в разных разделах будет у меня. подробно ответ на вопрос, не занимаемся ли мы фигней? а то очень даже можем и заниматься! может, нет ничего такого, о чем теоретики говорят, какой-то темной материи, еще чего-то, а просто в принципе мы оторваны от реальности? про это много я буду рассуждать в теории, но тут тоже когда-то укажу основные выводы. пока во все верю, не до сомнений.)

(тут типа такой-то график получили, как его описать? а вот так вот, по сорвеменной космологии. а в такой модификации он тоже описывается, но таки-то особенности тогда. Обзор критический всего будет)

Насколько можно верить в расширение Вселенной и почему?

(потом допишу, вообще, там разные крутые результаты есть, которые отлично подтвердились, кстати...)

Насколько можно верить в термодинамику Вселенной и почему?

Насколько можно верить в инфляцию и почему?

Насколько можно верить в темную материю и темную энергию и почему?

О некоторых альтернативных теориях

О теории Дикке

(указания, что такая есть, и почему она неверная, если так. по не до этого)

1.2 Main models of modern cosmology

1.2.1 Friedmann universe in a nutshell

Main equations and parameters

Robertson-Walker Metric

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\gamma_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j$$

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\,\mathrm{d}\Omega^2\right]$$

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\left[\mathrm{d}\chi^2 + S_k^2(\chi)\mathrm{d}\Omega^2\right]$$

$$\mathrm{d}s^2 = a^2(\tau)\left[\mathrm{d}\tau^2 - \left(\mathrm{d}\chi^2 + S_k^2(\chi)\mathrm{d}\Omega^2\right)\right] \quad \text{by} \quad \mathrm{d}\tau := \frac{\mathrm{d}t}{a(t)}$$

$$\mathrm{d}s^2 = \mathrm{d}t^2 - e^{2Ht}\,\mathrm{d}\boldsymbol{x}^2 \quad \text{if} \quad H \equiv \partial_t \ln a = \text{const} \quad \text{(de Sitter)}$$

$$\Gamma^{\mu}_{00} = \Gamma^0_{0\beta} = 0 \qquad \Gamma^0_{ij} = a\dot{a}\gamma_{ij}, \qquad \Gamma^i_{0j} = \frac{\dot{a}}{a}\delta^i_j, \qquad \Gamma^i_{jk} = \frac{1}{2}\gamma^{il}\left(\partial_j\gamma_{kl} + \partial_k\gamma_{jl} - \partial_l\gamma_{jk}\right) \quad \textit{(realt.)}$$

$$R_{00} = -3\frac{\ddot{a}}{a}, \qquad R_{ij} = -\left[\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 + 2\frac{k}{a^2}\right]g_{ij}. \qquad R = -6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2}\right]. \quad \textit{(realt.)}$$

Symmetry

$$a \to \lambda a, \quad r \to r/\lambda, \quad k \to \lambda^2 k.$$

Friedmann equations

Считаем, что у нас есть во Вселенной некотрые ρ and P за счет составляющих Вселенную, и просто за счет них и происходит динамика метрики Фридмана по уравнениям Эйнштейна

$G_{\mu\nu} = 8\pi G T_{\mu\nu}$.

$$\left(\frac{\dot a}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}, \qquad \frac{\ddot a}{a} = -\frac{4\pi G}{3}(\rho + 3P). \qquad \text{(real time, } c=1, \Lambda{=}0) \qquad \left[\frac{1}{\text{time}^2}\right]$$

$$(a')^2 + ka^2 = \frac{8\pi G}{3}\rho a^4, \qquad a'' + ka = \frac{4\pi G}{3}(\rho - 3p)a^3. \qquad \text{(conf. time)} \qquad \left(\tfrac{\dot a}{a}\right)^2 = \tfrac{1}{a^2}\left(\tfrac{da}{d\eta}\tfrac{d\eta}{dt}\right)^2 = \tfrac{(a')^2}{a^4}$$

$$\frac{\dot a^2 + kc^2}{a^2} = \frac{8\pi G\rho + \Lambda c^2}{3}, \qquad \frac{\ddot a}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} \qquad \text{(realt time, } c\neq 1, \Lambda \neq 1)$$

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}. \qquad \text{(real time, by } H := \dot a/a)$$

$$H^2 = \frac{\rho}{3M_{\rm pl}^2} \qquad \text{(real time, flat space)}$$

$$H^2(a) = H_0^2\left[\Omega_{r,0}\left(\frac{a_0}{a}\right)^4 + \Omega_{m,0}\left(\frac{a_0}{a}\right)^3 + \Omega_{k,0}\left(\frac{a_0}{a}\right)^2 + \Omega_{\Lambda,0}\right],$$

$$\frac{H^2}{H_0^2} = \Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda.$$

$$\Omega_{I,0} := \frac{\rho_{I,0}}{\rho_{\text{crit},0}} \qquad \Omega_{k,0} := -\frac{k}{(a_0 H_0)^2} \qquad \Omega_\Lambda = \frac{\rho_\Lambda}{\rho_{\rm crit}}, \quad \rho_{\rm crit} = \frac{3c^2H_0^2}{8\pi G}, \quad \rho_\Lambda = \frac{c^2\Lambda}{8\pi G},$$

(??? выведу еще раз!) Два уравнения Фридмана зависимые: второе получается из первого, однако уравнение второго порядка требует больше начальных условий, так что если решили 1-е, то следует проверить, выполняется ли 2-е (?? встречу в задаче - напишу точнее.)

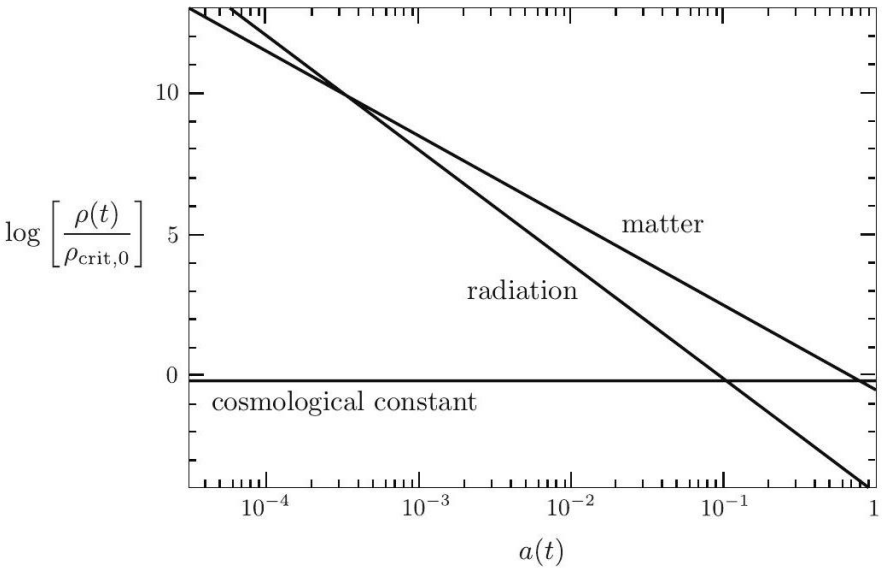
$$\rho_{\text{crit},0} := \frac{3H_0^2}{8\pi G} = 1.9\cdot 10^{-29} h^2 \frac{\text{gr}}{\text{cm}^3} = 2.8\cdot 10^{11} h^2 M_\odot \text{Mpc}^{-3} = 1.1\cdot 10^{-5} h^2 \frac{\text{protons}}{\text{cm}^3}.$$

$$\frac{\rho_0}{\rho_\Lambda} = \frac{\Omega_m}{\Omega_\Lambda} = \frac{\Omega_m}{1-\Omega_m}, \qquad \Omega_\Lambda = \frac{c^2\Lambda}{8\pi G}\frac{8\pi G}{3c^2H_0^2} = \frac{\Lambda}{3H_0^2}, \qquad \sqrt{3\Lambda} = 3H_0\sqrt{\Omega_\Lambda} = 3H_0\sqrt{1-\Omega_m},$$

(?!!! додумаю смысл новых определений!!) we will have

$$|\Omega_k| \le 0.01, \quad \Omega_r = 9.4\cdot 10^{-5}, \quad \Omega_m = 0.32, \quad \Omega_\Lambda = 0.68$$

Идея в том, что константой можно положить параметр в один момент времени в процессе их изменения. (?? где-то в диффурах проверю такой ход, прямо вот такого не встречал там (???))



Качественно: в момент начала Вселенной было много излучения, потом оно собралось в материю, а сейчас Вселенная так расширилась, что вакуума стало более всего. График именно такой из-за логарифмического масштаба. (???? там вроде как-то темная материя из ничего получается еще, а не через свет, но это допишу потом)

(абзац про то, могла ли Вселенная после расширения начать сужаться?)

(абзац про то, могла ли Вселенная иметь два больших взрыва?)

Continuity equation and its implementation to Friedmann equation*

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + P) = 0.$$

Оно получается из $\nabla_\mu T^\mu_\nu = 0$, взяв компоненту $\nu = 0$, putting $T^i{}_0 = 0$ from isotropy, а $\Gamma_{\text{frw}} = \frac{\dot{a}}{a}$ (? еще раз выведу?). Именно через него и входит давление в космологические модели.

Если в ур. Фридмана $H^2 = \rho/(3M_{pl}^2)$ после разных его модификаций вставить ур. непрерывности, то можно получить тождества:

$$\dot{H} = -\frac{\rho_\phi + P_\phi}{2M_{\text{pl}}^2} \qquad \text{(real time, take der of 1 Fr. eq.)}$$

$$\dot{H} + H^2 = -\frac{1}{6M_{\text{pl}}^2}(\rho + 3P) = -\frac{H^2}{2}\left(1 + \frac{3P}{\rho}\right). \qquad \text{(real time, put expr-s to the lhs.)}$$

(про приложения мб укажу)

Conformal time

$$\eta = -\frac{1}{H}e^{-Ht},$$

$$\eta := \int dt/a = \int \frac{da}{a}\frac{1}{H}\frac{a_0}{a}.$$

$$t = \int a d\eta.$$

We just need a preparation for formulas in conf. time:
$\frac{d}{dt} = \frac{d\eta}{dt}\frac{d}{d\eta} = \frac{1}{a}\frac{d}{d\eta}$
prime - the derivative w.r.t. conf. time η.

$$\ddot{a} = \frac{d}{dt}\frac{da}{dt} = \frac{d}{dt}\frac{da}{d\eta}\frac{d\eta}{dt} = \frac{d}{dt}\frac{a'}{a} = \frac{d\eta}{dt}\frac{d}{d\eta}\left(\frac{a'}{a}\right) = \frac{1}{a}\left(-\frac{a'^2}{a^2} + \frac{a''}{a}\right) = -\frac{a'^2}{a^3} + \frac{a''}{a^2}.$$

(?????????) Note that in these $c = 1$ units, the conformal time doubles as the comoving size of the horizon. In an open universe, it is also related to the development angle by

$$\chi = \sqrt{-K}\eta$$

Asymptotic relations are often useful for converting values. Before curvature or Λ domination, the conformal time

$$\eta = \frac{2\sqrt{2}}{k_{eq}}\left(\sqrt{1 + a/a_{eq}} - 1\right) = 2\left(\Omega_0 H_0^2\right)^{-1/2}(a_{eq}/a_0)^{1/2}\left(\sqrt{1 + a/a_{eq}} - 1\right)$$

It also takes on simple asymptotic forms, e.g.

$$t = \frac{2}{3}\left(\Omega_0 H_0^2\right)^{-1/2} a_0^{-3/2}\left[(a + a_{eq})^{1/2}(a - 2a_{eq}) + 2a_{eq}^{3/2}\right]. \quad \text{RD/MD}$$

(!!! абзац про то, почему оно такое удобное. Что оно бежит от 0 до времени Вселенной, т.п. много дописать хорошего можно!)

Величины констант и параметров в FRW

The expansion rate is given by the Hubble parameter

$$H^2 \equiv \left(\frac{1}{a}\frac{da}{dt}\right)^2 = \left(\frac{\dot{a}}{a}\frac{a_0}{a}\right)^2 = \left(\frac{a_0}{a}\right)^4 \frac{a_{eq}+a}{a_{eq}+a_0}\Omega_0 H_0^2 - \left(\frac{a_0}{a}\right)^2 K + \Omega_\Lambda H_0^2,$$

where the curvature is $K = -H_0^2\left(1-\Omega_0-\Omega_\Lambda\right)$.

$H_0 = 100h\mathrm{kms}^{-1}\mathrm{Mpc}^{-1} = 2.13\times10^{-42}h\mathrm{GeV} = (2997.9)^{-1}h\mathrm{Mpc}^{-1} = (3.08\times10^{17})^{-1}\,h\ \mathrm{s}^{-1} = (9.77)^{-1}h\mathrm{Gyr}^{-1}.$

Present day densities in a given particle species X are measured in units of the critical density $\rho_X\left(a_0\right) = \Omega_X\rho_{\mathrm{crit}}$, where

$\rho_{\mathrm{crit}} = \frac{3H_0^2}{8\pi G} = 1.87\cdot10^{-29}h^2\ \mathrm{g\ cm}^{-3} = 8.09\cdot10^{-47}h^2\mathrm{GeV}^4 = 1.05\cdot10^4h^2\mathrm{eVcm}^{-3} = 1.12\cdot10^{-5}h^2\ \mathrm{protons\ cm}^{-3} = 2.77\cdot10^{11}h^2\mathrm{M}_\odot\mathrm{Mpc}^{-3}.$

For the CMB,

$$n_{\gamma0} = 399.3\Theta_{2.7}^3\ \mathrm{cm}^{-3},$$
$$\rho_{\gamma0} = 4.4738\times10^{-34}\Theta_{2.7}^4\ \mathrm{g\ cm}^{-3}$$
$$\Omega_\gamma = 2.3812\times10^{-5}h^{-2}\Theta_{2.7}^4$$

and for the neutrinos

$$\rho_{\nu0} = \left[\left(1-f_\nu\right)^{-1}-1\right]\rho_{\gamma0},$$
$$\Omega_\nu = \left[\left(1-f_\nu\right)^{-1}-1\right]\Omega_\gamma,$$

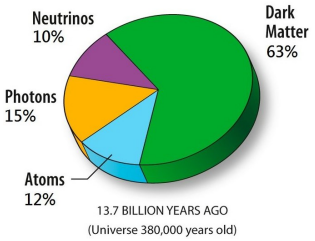
with $\left(1-f_\nu\right)^{-1} = 1.68$ for the standard model, or for the total radiation

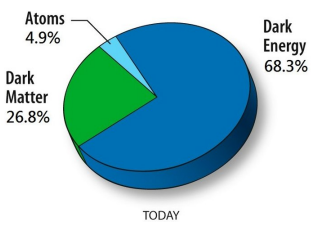
$$\rho_{r0} = \left(1-f_\nu\right)^{-1}\rho_{\gamma0}$$
$$\Omega_r = \left(1-f_\nu\right)^{-1}\Omega_\gamma.$$

Основные свойства компонентов Вселенной (???)

In FRW with flat single-component universe:

	$w := P/\rho$	$\rho(a)\sim a^{-3(1+w)}$	$a(t)$	$a(\tau)$
RD	$\frac{1}{3}$	a^{-4}	$t^{1/2}$	τ
MD	0	a^{-3}	$t^{2/3}$	τ^2
ΛD	-1	a^0	e^{Ht}	$-\tau^{-1}$





(?? тут очень коротко, почему так? вдали обоснование! что на СМБ смотрим - не информативно, нужно более содержательно.)

Matter (including dark matter)

$$|P| \ll \rho.$$

From $\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + P) = 0$

$$\rho \propto a^{-3}$$

Most of the matter in the universe is in the form of invisible dark matter. This is usually thought to be a new heavy particle species, but what it really is, we don't know.

Baryons

(Абзац Что такое темная материя? все равно не место про это рассуждать много, где-то дальше отдельно пр оэто будет)

(важные указания, пусть даже на 1 страницу, но к этому пониманию все и сводится!)

Radiation

Излучение - это фотоны, нейтрино, гравитоны. Нейтрино почти всегда были с почти скоростью света.

$$P = \frac{\rho}{3}$$

$$\rho \propto a^{-4} \qquad \text{or} \qquad E \propto a^{-1}$$

(??? короткий абзац, как доказать формулы? статфиз короткий нужен.) (-4)-я степень ожидаема, потому что (-3) - расширение Вселенной, и еще есть изменеие энергии от изменения длины волны (?? да??)

Dark energy (Λ)

$$P = -\rho$$

$$\rho \propto a^{0}$$

Vacuum energy.

(очень важный абзац Что такое темная энергия?!!!!!! многое напишу как будет время!!!!)

(важные указания, пусть даже на 1 страницу, но к этому пониманию все и сводится!)

О моделях с компонентами Вселенной

Методы анализа вселенной с разными компонентами (!!??!!)

Часто можно ввести a_{eq} в момент совпадения компонент, и через него выражать остальное.

Иногда можно переписать уравнение через конформное время.

(общие важные указания про такие задачи, важный параграф)

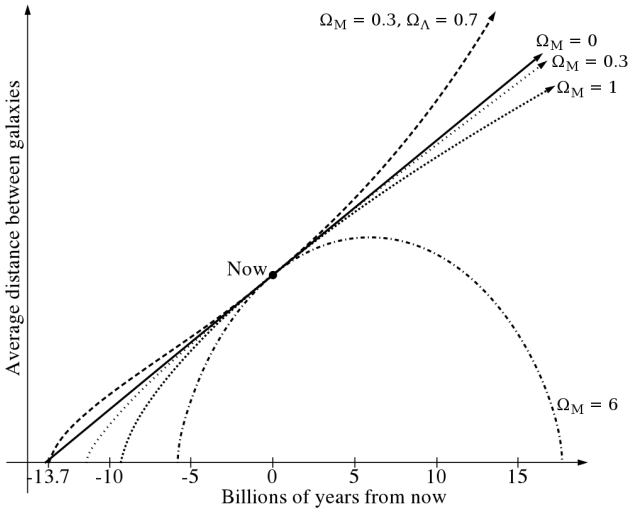
О Вселенной Фридмана с доминацией одной из компонент (???)

$$\frac{\dot{a}}{a} = H_0\sqrt{\Omega_I}a^{-\frac{3}{2}(1+w_I)}.$$

so

$$\text{real t.:}\quad a(t) \propto \begin{cases} t^{2/3(1+w_I)} & w_I \neq -1 & t^{2/3} & \text{MD} \\ & & t^{1/2} & \text{RD} \\ e^{Ht} & w_I = -1 & & \Lambda\text{D} \end{cases}, \quad \text{conf t.:}\quad a(\tau) \propto \begin{cases} \tau^{2/(1+3w_I)} & w_I \neq -1 & \tau^2 & \text{MD} \\ & & \tau & \text{RD} \\ (-\tau)^{-1} & w_I = -1 & & \Lambda\text{D} \end{cases}$$

(потом допишу, тут есть опыт изучения этого уже.)



О Вселенной Фридмана с доминацией двух компонент (???)

(порешаю потом, там разные можно модели придумать.)

$$(a')^2 = \frac{8\pi G}{3}\rho a^4 \qquad a'' = \frac{4\pi G}{3}(\rho - 3P)a^3 \qquad \text{(conf time!)}$$

$$\rho := \rho_m + \rho_r \equiv \frac{\rho_{\rm eq}}{2}\left[\left(\frac{a_{\rm eq}}{a}\right)^3 + \left(\frac{a_{\rm eq}}{a}\right)^4\right].$$

$$C = \left(\frac{4\pi G}{3}\rho_{\rm eq} a_{\rm eq}^4\right)^{1/2}$$

$$a(\tau) = a_{\rm eq}\left[\left(\frac{\tau}{\tau_\star}\right)^2 + 2\left(\frac{\tau}{\tau_\star}\right)\right]$$

$$\tau_\star := \left(\frac{\pi G}{3}\rho_{\rm eq} a_{\rm eq}^2\right)^{-1/2} = \frac{\tau_{\rm eq}}{\sqrt{2}-1}.$$

Другие свойства и уравнения

Perfect fluid

$$T_{\mu\nu} = (\rho + P)U_\mu U_\nu - P g_{\mu\nu},$$

where ρ and P are the energy density and the pressure of the fluid and U^μ is its four-velocity (relative to the observer).

Number Density:

$$\frac{\dot n}{n} = -3\frac{\dot a}{a} \Rightarrow n(t) \propto a^{-3}.$$

because of (????)

$$N^0 = n(t), \quad N^i = 0$$

$$\nabla_\mu N^\mu = \partial_\mu N^\mu + \Gamma^\mu_{\mu\lambda} N^\lambda = 0$$

$$\frac{dn}{dt} + \Gamma^i_{i0} n = 0$$

$$T^\mu{}_\nu = g^{\mu\lambda}T_{\lambda\nu} = \mathrm{diag}\Big[\rho, -P, -P, -P\Big] = (\rho + P)U^\mu U_\nu - P\delta^\mu_\nu,$$

Geodesics in FRW and a change of a momenta and energy of a particle in it

$$\frac{dU^\mu}{ds} + \Gamma^\mu_{\alpha\beta} U^\alpha U^\beta = 0, \qquad \Gamma^\mu_{\alpha\beta} := \frac{g^{\mu\lambda}}{2}\left(\partial_\alpha g_{\beta\lambda} + \partial_\beta g_{\alpha\lambda} - \partial_\lambda g_{\alpha\beta}\right)$$

$$U^\alpha \left(\frac{\partial U^\mu}{\partial X^\alpha} + \Gamma^\mu_{\alpha\beta} U^\beta\right) = 0. \qquad \frac{d}{ds} U^\mu\left(X^\alpha(s)\right) = \frac{dX^\alpha}{ds}\frac{\partial U^\mu}{\partial X^\alpha} = U^\alpha \frac{\partial U^\mu}{\partial X^\alpha}$$

$$P^\alpha \frac{\partial P^\mu}{\partial X^\alpha} = -\Gamma^\mu_{\alpha\beta} P^\alpha P^\beta. \qquad P^\mu := mU^\mu$$

$$\frac{\dot{p}}{p} = -\frac{\dot{a}}{a} \quad \Rightarrow \quad p \propto \frac{1}{a}.$$

$$p = E \propto \frac{1}{a} \quad \text{(massless particles)} \ ,$$

$$p = \frac{mv}{\sqrt{1-v^2}} \propto \frac{1}{a} \quad \text{(massive particles)}$$

The homogeneity of the FRW background implies $\partial_i P^\mu = 0$, so

$$P^0 \frac{dP^\mu}{dt} = -\Gamma^\mu_{\alpha\beta} P^\alpha P^\beta, = -\left(2\Gamma^\mu_{0j} P^0 + \Gamma^\mu_{ij} P^i\right) P^j,$$

so

$$P^j = 0 \Rightarrow \frac{dP^i}{dt} = 0.$$

for $\mu = 0$ because of $\Gamma^0_{0j} = 0$, we have $E\frac{dE}{dt} = -\Gamma^0_{ij} P^i P^j = -\frac{\dot{a}}{a} p^2$, we have that obvious properties.

SUMMARY Smooth Expanding universe

Comoving distance between us and a source at redshift z:

$$\chi(z) = \int_0^z \frac{dz'}{H\left(z'\right)}.$$

How much light could have traveled since $t = 0$, it is conformal time:

$$\eta = \int_0^t \frac{dt'}{a\left(t'\right)} = \int_z^\infty \frac{dz'}{H\left(z'\right)}.$$

Photons in the universe have a Bose-Einstein distribution with zero chemical potential, so their energy density can be determined by measuring their temperature. Neutrinos have a Fermi-Dirac distribution, also probably with zero chemical potential, but there is some ambiguity in their energy density because of our ignorance of the neutrino masses. Early on, this ambiguity is irrelevant since the temperatures are so much larger than the masses and neutrinos behave relativistically. Thus, the uncertainty in neutrino mass does not affect Big Bang nucleosynthesis at temperatures of order 1MeV and probably not even the epoch of matter radiation equality at temperatures of order 1eV. The neutrino temperature is a factor of $(4/11)^{1/3}$ smaller than the photon temperature. This, and the difference in statistics, implies that a species of massless neutrinos has an energy density equal to 0.23 times that of photons. A single neutrino generation with mass m_ν contributes $\Omega_\nu = 0.01\left(m_\nu/0.94\text{eVh}^2\right)$. In addition to photons and neutrinos, the universe consists of baryons, best determined by nucleosynthesis to have $\Omega_b h^2 = 0.0205 \pm 0.0018$; dark matter ($\Omega_m \simeq 0.3$); and dark energy (with $\Omega_{\text{de}} \simeq 0.7$), a new form of energy with negative pressure.

$$a_{\text{eq mt-rad}} = 4.15 \cdot 10^{-5}/\Omega_m h^2.$$

1.2.2 On distances, expansion, Hubble's law, Hubble Sphere, horizons, redshift

Metrics, distances, times

$x^i := \{x^1, x^2, x^3\}$ - comoving coordinates
$x^i_{\text{phys}} = a(t)x^i$ - physical coordinates

$$v^i_{\text{phys}} := \frac{dx^i_{\text{phys}}}{dt} = a(t)\frac{dx^i}{dt} + \frac{da}{dt}x^i := v^i_{\text{pec}} + Hx^i_{\text{phys}} \qquad \text{(physical velocity)}$$
$$v^i_{\text{pec}} := a(t)\dot{x}^i \qquad \text{(peculiar velocity)}$$

$$H := \frac{\dot{a}}{a}$$

О метриках в космологии

Negatively (-) and positively (+) curved space:

$$\mathrm{d}\ell^2 = a^2\left[\ \mathrm{d}\boldsymbol{x}^2 + k\frac{(\boldsymbol{x}\cdot\mathrm{d}\boldsymbol{x})^2}{1-k\boldsymbol{x}^2}\right] := a^2\gamma_{ij}\ \mathrm{d}x^i\ \mathrm{d}x^j$$

$$\gamma_{ij} := \delta_{ij} + k\frac{x_i x_j}{1 - k\cdot x_k x^k}, \qquad k := \begin{cases} 0 & \text{Euclidean} \\ +1 & \text{spherical} \\ -1 & \text{hyperbolic} \end{cases}$$

this is from $\mathrm{d}\ell^2 = a^2\left[\ \mathrm{d}\boldsymbol{x}^2 \pm \mathrm{d}u^2\right], \quad \boldsymbol{x}^2 \pm u^2 = \pm 1.$
Or in polar coordinates:

$$\mathrm{d}\boldsymbol{x}^2 = \mathrm{d}r^2 + r^2\left(\ \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right),$$

$$\mathrm{d}\ell^2 = a^2\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\ \mathrm{d}\Omega^2\right],$$

where $\mathrm{d}\Omega^2 := \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2$.
Or by χ, which simplifies radial part

$$\mathrm{d}\chi := \mathrm{d}r/\sqrt{1-kr^2}$$

$$\mathrm{d}\ell^2 = a^2\left[\ \mathrm{d}\chi^2 + S_k^2(\chi)\mathrm{d}\Omega^2\right], \qquad S_k(\chi) := \begin{cases} \mathrm{sh}\,\chi & k=-1 \\ \chi & k=0 \\ \sin\chi & k=+1 \end{cases}.$$

Metric, luminosity, angular diameter distances

(потом додумаю, не самые важные понятия)
On metric distance $d_m = S_k(\chi)$

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\left[\mathrm{d}\chi^2 + S_k^2(\chi)\mathrm{d}\Omega^2\right],$$
$$S_k(\chi) := \begin{cases} R_0\,\mathrm{sh}\,(\chi/R_0) & k=-1 \\ \chi & k=0 \\ R_0\sin(\chi/R_0) & k=+1 \end{cases}.$$

$$\chi(z) = \int_{t_1}^{t_0}\frac{\mathrm{d}t}{a(t)} = \int_0^z\frac{\mathrm{d}z}{H(z)}$$

On luminosity distance d_L (??????)

$$d_L = d_m(1+z)$$

$$F = \frac{L}{4\pi\chi^2}.$$

$$F = \frac{L}{4\pi d_m^2(1+z)^2} := \frac{L}{4\pi d_L^2},$$

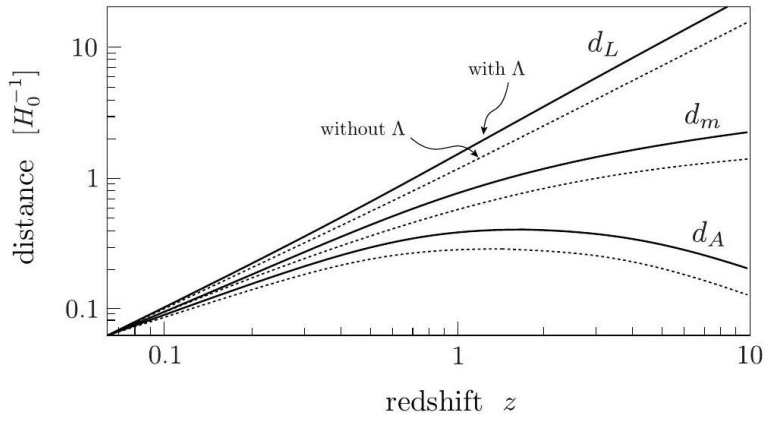
On angular diameter distance

$$d_A = \frac{D}{\delta\theta} = \frac{d_L}{(1+z)^2}$$

$$d_A = \frac{d_m}{1+z}.$$

because $D = a\,(t_1)\,S_k(\chi)\delta\theta = \frac{d_m}{1+z}\delta\theta$

(вставлю мб картинку потом)



(??? этот график я не понимаю, ну и не актуально пока что)

Смысл и общие свойства времен в космологии

Время - проще говоря это просто параметр, который увеличивается. Можно считать, что настоящее время конформное, а то, к которому при привыкли - лишь его приближение. Так или иначе в этих моделях, это просто возрастающий параметр.

(наверняка, что-то потом допишу поглубже)

Standard Candle

- Standard Candle: objects of same luminosity L, measured flux F
- Apply rules again: at emission in conformal-comoving coordinates
- L is the energy per unit time at emission
- Since $E \propto \lambda^{-1}$ and comoving wavelength $\Lambda \propto \lambda/a$ so comoving energy $\mathcal{E} \propto \Lambda^{-1} \propto aE$
- Per unit time at emission $\Delta t = a\Delta\eta$ in conformal time
- So observed luminosity today is $\mathcal{L} = \mathcal{E}/\Delta\eta = a^2 L$
- All photons must pass through the sphere at a given distance, so the comoving surface area is $4\pi D_A^2$
- Put this together to the observed flux at $a = 1$

$$F = \frac{\mathcal{L}}{4\pi D_A^2} = \frac{L}{4\pi D_A^2}\frac{1}{(1+z)^2}$$

Notice the flux is diminished by two powers of $(1+z)$

Standard Ruler

- Standard Ruler: object of known physical size λ
- Let's apply our rule of thumb: at emission the comoving size is Λ:

$$\lambda = a(t)\Lambda$$

Now everything about light is normal: the object of comoving size Λ subtends an observed angle α on the sky α

$$\alpha = \frac{\Lambda}{D_A(z)}$$

- This is the easiest way of thinking about it. But we could also define an effective physical distance $d_A(z)$ which corresponds to what we would infer in a non expanding spacetime

$$\alpha \equiv \frac{\lambda}{d_A(z)} = \frac{a\Lambda}{d_A(z)} = \frac{\Lambda}{D_A(z)} \rightarrow d_A(z) = aD_A(z) = \frac{D_A(z)}{1+z}$$

- Since $D_A \rightarrow D_A\left(D_{\text{horizon}}\right)$ whereas $(1+z)$ unbounded, angular size of a fixed physical scale at high redshift actually increases with radial distance
- Paradox: the further away something is, the bigger it appears - Easily resolved by thinking about comoving coordinates - a fixed physical scale λ as the universe shrinks and $a \rightarrow 0$ will eventually encompass the whole observable universe out to the horizon in comoving coordinates so of course it subtends a big angle on the sky!
- But there are no such bound objects in the early universe there is no causal way such bigger-than-the-horizon objects could form
- Knowing λ or Λ and measuring α and z allows us to infer the comoving angular diameter distance $D_A(z)$

Изотропное пространство

У пространства с положительной кривизной

$$V = \int_0^{2\pi}\int_0^{\pi}\int_0^{\pi} a^3 \sin^2\chi \sin\theta d\chi d\theta d\varphi = 2\pi^2 a^3.$$

Таким образом, пространство положительной кривизны оказывается «замкнутым само в себе» - конечным по объему, но, разумеется, не имеющим границ.
(допишу пл ЛЛ2, интересные соображения)

Закрытая изотропная модель

(допишу пл ЛЛ2, интересные соображения)

Открытая изотропная модель

(допишу пл ЛЛ2, интересные соображения)
(пока идея просто подробно указать эти основные причины таких моделей космологии)

Redshift z, Hubble's law $z \simeq H_0 d$ (!??!)

Красное смещение - увеличение длины волны (уменьшение энерги) для движущегося наблюдателя.

$$\lambda_0 = \frac{a\left(t_0\right)}{a\left(t_1\right)}\lambda_1.$$

Since $a\left(t_0\right) > a\left(t_1\right)$, $\lambda_0 > \lambda_1$.

For classical waves

$$\Delta t_1 = a(\tau_1)\,\Delta\tau \quad \text{and} \quad \Delta t_0 = a(\tau_0)\,\Delta\tau$$

If Δt is the period of the light wave, the light is emitted with wavelength $\lambda_1 = \Delta t_1$ (in units where $c = 1$), but is observed with wavelength $\lambda_0 = \Delta t_0$, so that

$$\frac{\lambda_0}{\lambda_1} = \frac{a(\tau_0)}{a(\tau_1)}$$

$$z := \frac{\lambda_0 - \lambda_1}{\lambda_1} \qquad \text{(Redshift parameter)}$$

λ_0 - wavelenght today, λ_1 - wavelenght at some t_1,

$$1 + z = \frac{1}{a(t_1)}.$$

$$a(t_1) = a(t_0)\left[1 + (t_1 - t_0)\,H_0 + \cdots\right] \qquad H_0 := \frac{\dot{a}(t_0)}{a(t_0)}.$$

parameter h is defined such that

$$H_0 := 100h\mathrm{kms}^{-1}\mathrm{Mpc}^{-1},$$

it is used to keep track of how uncertainties in H_0 propagate into other cosmological parameters.

$$h \approx 0.67 \pm 0.01.$$

Наиболее надёжная оценка постоянной Хаббла на 2013 год составляла

$$H_0 = 67,80 \pm 0,77(\text{км/с})/\text{Мпк}.$$

$$1\text{ пк} = \frac{1}{\operatorname{tg} 1''}\text{ a.e.} \approx \frac{360\cdot 60\cdot 60}{2\pi}\text{ a.e.} \approx 206264,8\text{ a.e.} = 3,08\cdot 10^{16}\text{ м}$$
$$= 30,85\text{ трлн км (петаметров)} = 3,26\text{ св. года.}$$

$$h := H_0/\left(100\text{ km s}^{-1}\mathrm{Mpc}^{-1}\right)$$

$$t_H \equiv \frac{1}{H_0} = \frac{1}{67.8(\text{ km/s})/\text{Mpc}} = 4.55\times 10^{17}\text{ s} = 14.4\text{ billion years (Hubble time).}$$

Вообще, разные есть результаты измерений, так что это отдельное обсуждение - есть ли вообще такой параметр?

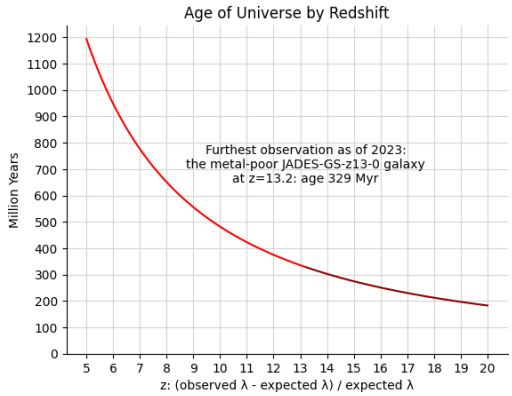
(??? короче говоря, даже тут какие-то проблемы, что разные измерения дают разные значения, так что все это может быть просто неверным. додумаю потом.)

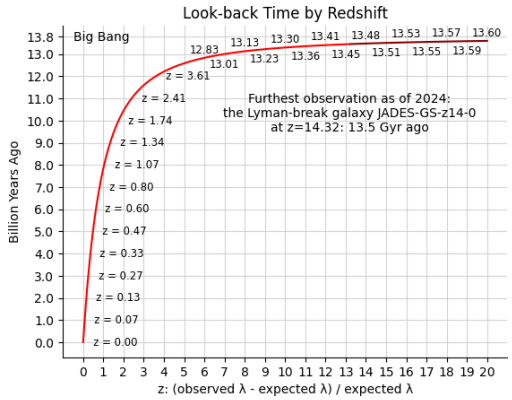
he largest-observed redshift, corresponding to the greatest distance and furthest back in time, is that of the cosmic microwave background radiation; the numerical value of its redshift is about $z = 1089$ ($z = 0$ corresponds to present time), and it shows the state of the universe about 13.7 billion years ago, and 379,000 years after the initial moments of the Big Bang.

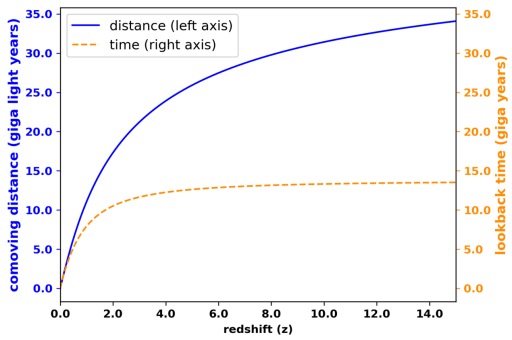
(!?!?!!? абзац, как вообще интерпретируем кр смещ??? почему именно так, почему не через другие модели, через непостоянную скорость света нпр. в теории много про это обсуждения.)

Observations of redshift

(see wiki)







О другом про и других способах объяснения красного смещения z и закона Хабла $z \simeq H_0 d$ (!??!)

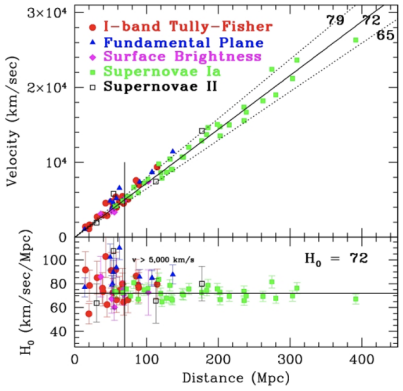
(!!! важный абзац про теорию Дикке, который объяснил с помощью задержки гравитационного замедления света. такое вообще может быть???? у меня в плане оптики неуверенность.)

Redshift surveys

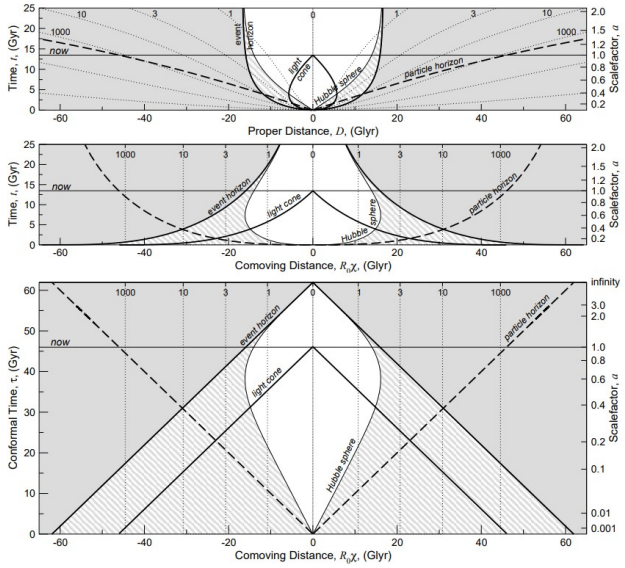
(см лекции Калтеха Physics 217bc: The Standard Model - Cosmology, потом допишу, пока не до этого)

Main about Hubble sphere and horizons (????)

$$\frac{d}{dt}(aH)^{-1} < 0.$$



Light and Horizons



Горизонт частиц - максимальное расстояние, с которого было влияние на настоящий момент пространства-времени. Горизонт событий - максимальное расстояние объектов, на которые мы могли бы повлиять. Если бы не было расширения, то это было бы бесконечностью, а так - там интеграл сходится

(?? я пока остановился на том, что вообще разные есть определения, потому что в статье про это как будто наоборот они. когда-то еще подумаю про это)

The comoving horizon is finite,

$$\chi_{\rm ph}(t) = \frac{2H_0^{-1}}{(1+3w)} a(t)^{\frac{1}{2}(1+3w)} = \frac{2}{(1+3w)}(aH)^{-1}.$$

$$\chi_{\rm eh}(\tau) = \tau_f - \tau = \int_t^{t_f} \frac{{\rm d}t}{a(t)}.$$

Particle horizon.-Eq. (2.1.2) tells us that the maximal comoving distance that light can travel between two times τ_1 and $\tau_2 > \tau_1$ is simply $\Delta\tau = \tau_2 - \tau_1$ (recall that $c := 1$). Hence, if the Big Bang 'started' with the singularity at $t_i := 0$, [2] then the greatest comoving distance

from which an observer at time t will be able to receive signals travelling at the speed of light is given by

$$\chi_{\rm ph}(\tau) = \tau - \tau_i = \int_{t_i}^{t} \frac{{\rm d}t}{a(t)}$$

- Event horizon.-Just as there are past events that we cannot see now, there may be future events that we will never be able to see (and distant regions that we will never be able to influence). In comoving coordinates, the greatest distance from which an observer at time t_f will receive signals emitted at any time later than t is given by

$$\chi_{\rm eh}(\tau) = \tau_f - \tau = \int_{t}^{t_f} \frac{{\rm d}t}{a(t)}$$

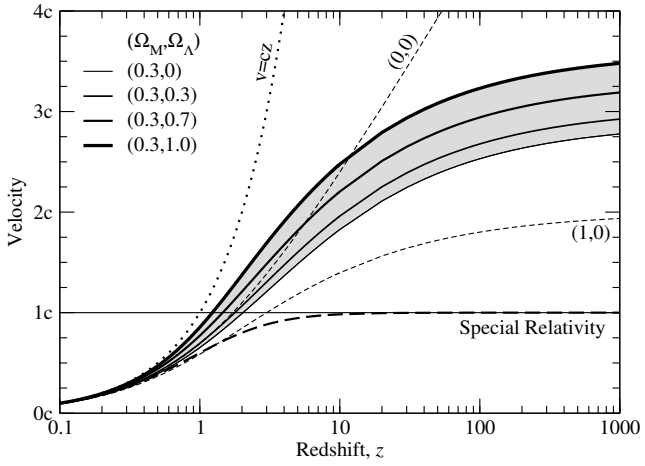
Growing Hubble Sphere

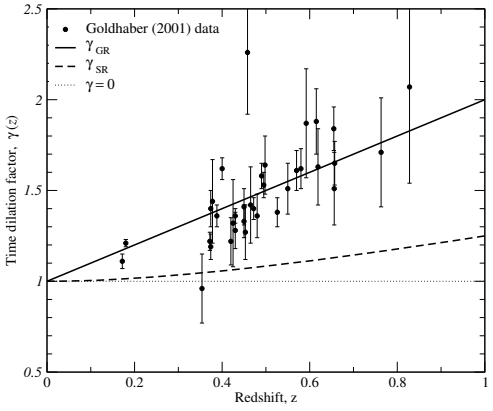
comoving Hubble radius $(aH)^{-1}$.
(????????????????? откуда это??????)

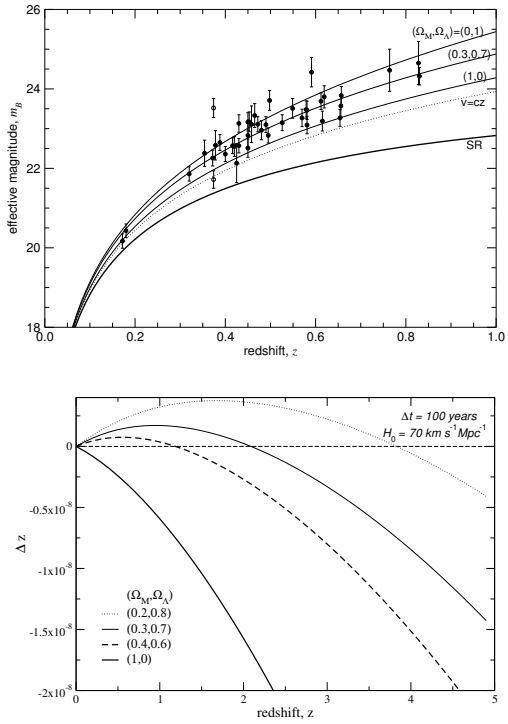
$$(aH)^{-1} = H_0^{-1} a^{\frac{1}{2}(1+3w)}$$

$$\chi_{\rm ph}(a) = \frac{2H_0^{-1}}{(1+3w)} \left[a^{\frac{1}{2}(1+3w)} - a_i^{\frac{1}{2}(1+3w)} \right] := \tau - \tau_i$$

$$\chi_{\rm ph}(\tau) = \int_{t_i}^{t} \frac{{\rm d}t}{a} = \int_{a_i}^{a} \frac{{\rm d}a}{a\dot{a}} = \int_{\ln a_i}^{\ln a} (aH)^{-1} \, {\rm d}\ln a$$







Why is the CMB so uniform?

(??????????????????)
If there wasn't enough time for these regions to communicate, why do they look so similar? This is the horizon problem

Inflation - Solution of the Horizon Problem

$$\tau_i = \frac{2H_0^{-1}}{(1+3w)} a_i^{\frac{1}{2}(1+3w)} \overset{a_i \to 0, w < -\frac{1}{3}}{\longrightarrow} -\infty$$

Hubble Radius vs. Particle Horizon

1.2.3 Homogeneous inflation in a nutshell

Essene and main formulas of inflation

Both the horizon problem and the flatness problem can be traced back to the fact that during the standard Big Bang evolution the comoving Hubble radius, $(aH)^{-1}$, grows monotonically with time. During inflation on the other hand the comoving Hubble radius is temporarily decreasing.

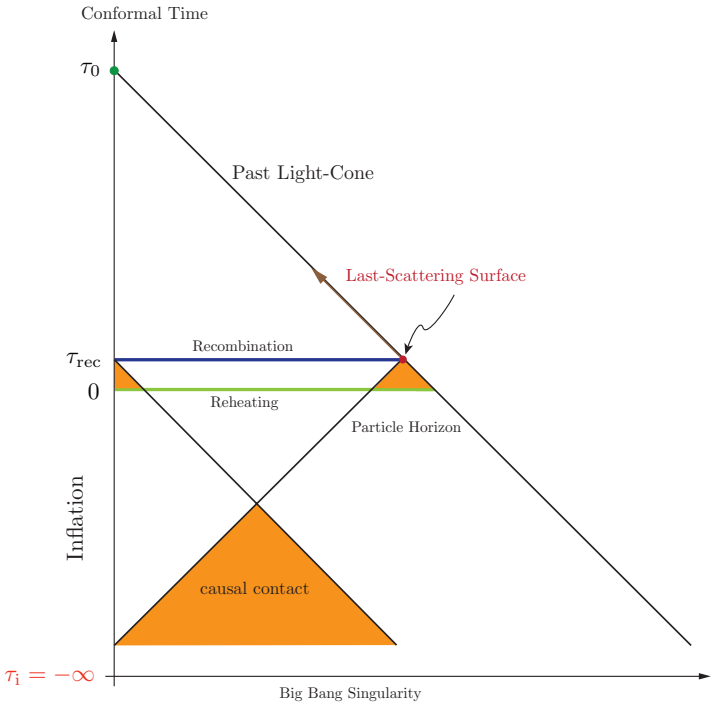


Figure 1: Conformal diagram of inflationary cosmology. Inflation extends conformal time to negative values! The end of inflation creates an "apparent" Big Bang at $\tau = 0$. There is, however, no singularity at $\tau = 0$ and the light cones intersect at an earlier time if inflation lasts for at least 60 e-folds.

This changes the causal structure of the early universe making the horizon problem a fiction of extrapolating the conventional FRW expansion back to arbitrarily early times. From the Einstein Equations one may show that a shrinking Hubble radius corresponds to accelerated expansion as it occurs if the universe is filled with a negative pressure component. The three equivalent conditions for inflation therefore are

$$\frac{d(aH)^{-1}}{dt} < 0 \qquad \Rightarrow \qquad \frac{d^2a}{dt^2} > 0 \qquad \Rightarrow \qquad p < -\frac{\rho}{3} \, .$$

A negative pressure fluid can be modeled by scalar field ϕ, the inflaton, with the following action

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2}R + \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right] \, .$$

This will lead to inflation if the slow-roll conditions are satisfied

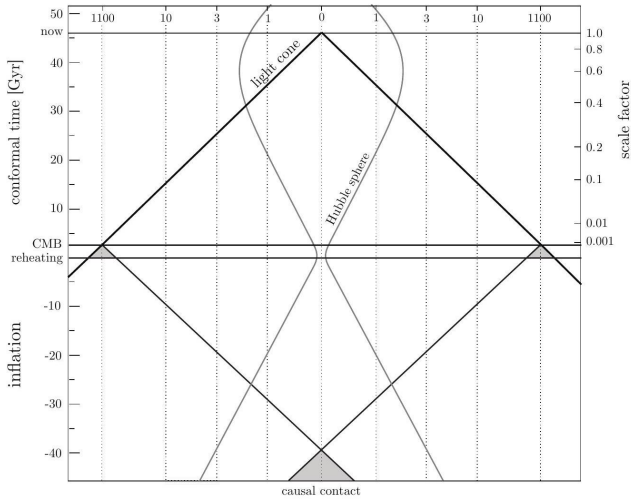
$$\epsilon_{\rm v} = \frac{M_{\rm pl}^2}{2}\left(\frac{V_{,\phi}}{V}\right)^2 , \qquad \eta_{\rm v} = M_{\rm pl}^2\frac{V_{,\phi\phi}}{V} , \qquad \epsilon_{\rm v}, |\eta_{\rm v}| < 1 \, .$$

The number of e-folds of inflationary expansion $N(\phi) = \int_{\phi_{\rm end}}^{\phi} \frac{\mathrm{d}\phi}{\sqrt{2\epsilon_{\rm v}}}$ needs to be at least 60 to solve the horizon problem.

CMB fluctuations are created during four e-folds about 60 e-folds before the end of inflation. Even in the restricted framework for single-field slow-roll inflation described by the above action, there are a multitude of inflationary models characterized by different choices for the inflationary potential $V(\phi)$.

На данный момент нет четкого понимания, каков конкретно смысл модели скалярного поля. Возможно, есть не видимое никем "инфляторное поле", которое и расширяло Вселенную, а к настоящему моменту исчезло. Возможно, у нас такой вакуум сам по себе, что в себе имел потенциал, и поэтому расширял пространство (то есть себя же). Возможно, это вообще что-то другое или вообще не было инфляции. Однако, есть известная модель со

скалярным полем, которая неплохо описывает расширение, которое объясняет экспериментальные данные (CMB или что???)



A defining characteristic of inflation is the behavior of the comoving Hubble radius, $1/(aH)$, which shrinks quasi-exponentially. A mode with comoving wavenumber k is called super-horizon when $k < aH$, and sub-horizon when $k > aH$. The inflaton is taken to be in a vacuum state, defined such that sub-horizon modes approach the Minkowski vacuum for $k \gg aH$. After a mode exits the horizon, it is described by a classical probability distribution with variance given by the power spectrum evaluated at horizon crossing

$$P_{\rm s}(k) = \left.\frac{H^2}{2k^3}\frac{H^2}{\dot\phi^2}\right|_{k=aH}.$$

Inflation also produces fluctuations in the tensor part of the spatial metric. This corresponds to a spectrum of gravitational waves with power spectrum

$$P_{\rm t}(k) = \left.\frac{4}{k^3}\frac{H^2}{M_{\rm pl}^2}\right|_{k=aH}.$$

For slow-roll models the scalar and tensor spectra are expressed purely in terms of $V(\phi)$ and $\epsilon_{\rm v}$ (or $V_{,\phi}$)

$$\Delta^2_{\rm s}(k) \approx \left.\frac{1}{24\pi^2}\frac{V}{M_{\rm pl}^4}\frac{1}{\epsilon_{\rm v}}\right|_{k=aH}, \quad \Delta^2_{\rm t}(k) \approx \left.\frac{2}{3\pi^2}\frac{V}{M_{\rm pl}^4}\right|_{k=aH},$$

where $\Delta^2(k) \equiv \frac{k^3}{2\pi^2}P(k)$. The scale dependence is given by

$$n_{\rm s} - 1 \equiv \frac{d\ln\Delta^2_{\rm s}}{d\ln k} = 2\eta_{\rm v} - 6\epsilon_{\rm v},$$
$$n_{\rm t} \equiv \frac{d\ln\Delta^2_{\rm t}}{d\ln k} = -2\epsilon_{\rm v}.$$

The tensor-to-scalar ratio is

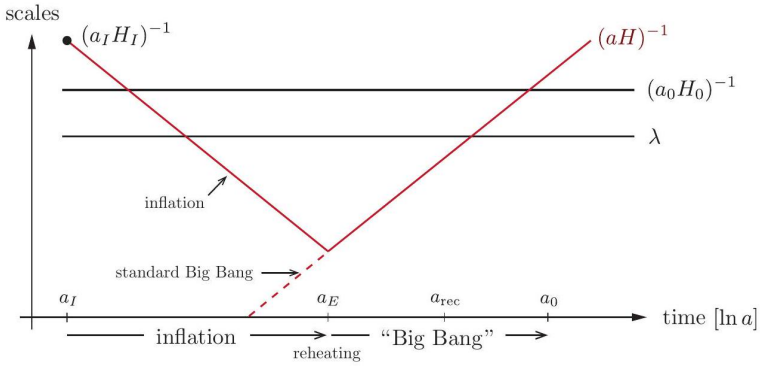
$$r \equiv \frac{\Delta^2_{\rm t}}{\Delta^2_{\rm s}} = 16\epsilon_{\rm v}.$$

By the Lyth bound, r relates directly to total field excursion during inflation

$$\frac{\Delta\phi}{M_{\rm pl}} \approx \left(\frac{r}{0.01}\right)^{1/2}$$

A large value for r therefore correlates both with a high scale for the inflationary energy and a super-Planckian field evolution.

Conditions for Inflation



(!!??? описание очень важного явления: выхода за границы горизонта и заход в него!)

- Accelerated expansion. From $\frac{d}{dt}(aH)^{-1} = \frac{d}{dt}(\dot{a})^{-1} = -\frac{\ddot{a}}{(\dot{a})^2}$, we see that a shrinking comoving Hubble radius implies accelerated expansion

$$\ddot{a} > 0 \qquad \text{(real time)}$$

- Slowly-varying Hubble parameter.

$$\frac{d}{dt}(aH)^{-1} = -\frac{\dot{a}H + a\dot{H}}{(aH)^2} = -\frac{1}{a}(1-\varepsilon), \quad \text{where} \quad \varepsilon := -\frac{\dot{H}}{H^2}. \qquad \text{(real time)}$$

The shrinking Hubble sphere therefore also corresponds to

$$\varepsilon = -\frac{\dot{H}}{H^2} < 1 \qquad \text{(real time)}$$

- Quasi-de Sitter expansion.-For perfect inflation, $\varepsilon = 0$, the spacetime becomes de Sitter space $\mathrm{d}s^2 = \mathrm{d}t^2 - e^{2Ht}\,\mathrm{d}\boldsymbol{x}^2$, where $H = \partial_t \ln a = \text{const}$.

- Negative pressure. Because consider a perfect fluid with pressure P and density ρ. From Friedmann equation, $H^2 = \rho/\left(3M_{\rm pl}^2\right)$, and the continuity equation, $\dot{\rho} = -3H(\rho + P)$, we have

$$\dot{H} + H^2 = -\frac{1}{6M_{\rm pl}^2}(\rho + 3P) = -\frac{H^2}{2}\left(1 + \frac{3P}{\rho}\right). \qquad \text{(real time)}$$

$$-\frac{\dot{H}}{H^2} \equiv \varepsilon = \frac{3}{2}\left(1 + \frac{P}{\rho}\right) < 1 \quad \Leftrightarrow \quad w := \frac{P}{\rho} < -\frac{1}{3}, \qquad \text{(real time)}$$

- Constant density. The continuity equation $\dot{\rho} = -3H(\rho + P)$ with $\dot{H} + H^2 = -\frac{H^2}{2}\left(1 + \frac{3P}{\rho}\right)$ gives (??? check)

$$\left|\frac{d\ln\rho}{d\ln a}\right| = 2\varepsilon < 1.$$

For $\varepsilon \ll 1$, the energy density is therefore nearly constant.

Typical model of scalar field-driven inflation

General Scalar Inflanton Field Dynamics

Пусть есть скалярное поле, которое задает инфляцию, с некоторым потенциалом $V(\phi)$:

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi). \qquad P_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi), \qquad H^2 \overset{\text{1F eq}}{=} \frac{1}{3M_{\rm pl}^2}\left[\frac{1}{2}\dot{\phi}^2 + V\right] \qquad \text{(real time)}$$

From here $2H\dot{H} = \frac{1}{3M_{\rm pl}^2}\left[\dot\phi\ddot\phi + V'\dot\phi\right]$, where $V' := dV/d\phi$, and we add from $\dot H \overset{\text{cont+1F}}{=} -\frac{\rho_\phi + P_\phi}{2M_{\rm pl}^2}$ the equation $\dot H = -\frac{1}{2}\frac{\dot\phi^2}{M_{\rm pl}^2}$:

$$\ddot\phi + 3H\dot\phi + V' = 0. \qquad \text{(real time)}$$

The same can be obtained from a Lagrangian $L = a^3\left[\frac{1}{2}\dot\phi^2 - V(\phi)\right]$ which corresponds to metric $\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\mathrm{d}\boldsymbol{x}^2$, (for which $g = -a^6$), if we just use $\frac{d}{dt}\left(\frac{\partial L}{\partial\dot\phi}\right) - \frac{\partial L}{\partial\phi} = 0$.

In particular, if $V = \frac{1}{2}m^2\phi^2$ in terms of $\phi(t) \equiv a^{-3/2}(t)\chi(t)$ we have

$$\ddot\chi + \left(m^2 - \frac{3}{2}\dot H - \frac{9}{4}H^2\right)\chi = 0. \qquad \text{(real time)}$$

Parameters for general inflation

Parameters can be defined differently! The main difference can be in real/confirmal time (so always write them explicitly) and in signs (so before any derivations start in the beginning write, which parameters a certain book uses and wrt which time the derivatives!). Never start without been sure wrt to which time the derivative (write it!) and think that you are in a dead-end, if you haven't written, wrt to what the derivatives! Usually problems can arise only because of definitions of symbols.

$$M_{\rm pl}^2 := \frac{\hbar c}{8\pi G} = 2.4\cdot 10^{18}\ \text{GeV}$$

$$\varepsilon := -\frac{\dot H}{H^2} \equiv -\frac{d\ln H}{dN}$$

$$\eta := \frac{d\ln\varepsilon}{dN} = \frac{\dot\varepsilon}{H\varepsilon} \qquad \mathrm{d}N \equiv \mathrm{d}\ln a = H\ \mathrm{d}t$$

Conditions of SR:

$$\varepsilon < 1, \qquad |\eta| < 1$$

$$\epsilon \equiv \frac{d}{dt}\left(\frac{1}{H}\right) \qquad \text{(from Craps)}$$

$$\delta \equiv \frac{1}{H}\frac{d^2\phi/dt^2}{d\phi/dt} \qquad \text{(from Craps)}$$

$$\epsilon = \frac{1}{16\pi G}\left(\frac{V'}{V}\right)^2, \quad \delta = \epsilon - \frac{1}{8\pi G}\frac{V''}{V}. \qquad \text{(from Craps)}$$

(абзац про то, почему именно такие параметры??? мне так и не объясняли это.)
For a scalar field inflation model we have:

$$\delta := -\frac{\ddot\phi}{H\dot\phi} \qquad \text{(conf.(?????) t.)}$$

$$\varepsilon := \frac{\frac{1}{2}\dot\phi^2}{M_{\rm pl}^2H^2} \qquad \text{(conf.(?????) t.)} \qquad \left(\text{so } \dot\varepsilon = \frac{\dot\phi\ddot\phi}{M_{\rm pl}^2H^2} - \frac{\dot\phi^2\dot H}{M_{\rm pl}^2H^3}\right)$$

$$\eta = 2\frac{\ddot\phi}{H\dot\phi} - 2\frac{\dot H}{H^2} = 2(\varepsilon - \delta). \qquad \text{(conf.(?????) t.)}$$

$$\{\varepsilon, |\delta|\} \ll 1 \qquad \text{implies} \qquad \{\varepsilon, |\eta|\} \ll 1.$$

(!! еще раз решу задачи - пропишу заготовки для них! многое еще указать можно очень нужного)

Slow-Roll Inflation

Slow-roll condition in terms of a potential:

$$\epsilon_{\rm v} := \frac{M_{\rm pl}^2}{2}\left(\frac{V'}{V}\right)^2, \qquad |\eta_{\rm v}| := M_{\rm pl}^2\frac{|V''|}{V}$$

$$\epsilon_{\rm v} \ll 1 \qquad |\eta_{\rm v}| \ll 1.$$

Definition of SR is that:

$$\left(\frac{d\phi}{dt}\right)^2 \ll V(\phi)$$

$$\varepsilon \ll 1$$

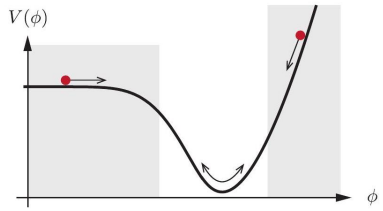
or $\frac{1}{2}\dot\phi^2 \ll V$ and hence

$$H^2 \overset{\text{sl r}}{\approx} \frac{V}{3M_{\rm pl}^2}.$$

and from $|\delta| \ll 1$

$$3H\dot\phi \overset{\text{sl r}}{\approx} -V'.$$

Below two gray regions satisfy slow-roll inflation:



Properties of SR inflation:

$$\frac{V_{,\phi\phi}}{H^2} \approx \frac{3M_{\rm pl}^2 V_{,\phi\phi}}{V} = 3\eta_{\rm v} \ll 1$$

$$\delta + \varepsilon := -\frac{\ddot\phi}{H\dot\phi} - \frac{\dot H}{H^2} \approx M_{\rm pl}^2\frac{V''}{V}.$$

Since $a' = a^2 H$, with $H \approx$ const,

$$\frac{a''}{a} \approx 2a'H = 2a^2H^2 \gg a^2 V_{,\phi\phi}$$

Slow-roll parameters in terms of a potential:

$$\epsilon_{\rm v} := \frac{M_{\rm pl}^2}{2}\left(\frac{V'}{V}\right)^2,$$

$$|\eta_{\rm v}| := M_{\rm pl}^2\frac{|V''|}{V}.$$

Properties of inflation

(collection of useful formulas, see exercises for proofs!?!?!)

$$\frac{d\epsilon}{d\eta} = 2aH\epsilon(\epsilon+\delta)$$

$$4\pi G\left(\dot\phi^{(0)}\right)^2 = \epsilon a^2 H^2,$$

Reheating (giving energy from inflaton to matter fields)

During inflation most of the energy density in the universe is in the form of the inflaton potential $V(\phi)$. Inflation ends when the potential steepens and the inflaton field picks up kinetic energy. The energy in the inflaton sector then has to be transferred to the particles of the Standard Model.

$$\ddot{\phi} + 3H\dot{\phi} = -m^2\phi.$$

$$\dot{\rho}_\phi + 3H\rho_\phi = -3HP_\phi = -\frac{3}{2}H\left(m^2\phi^2 - \dot{\phi}^2\right)$$

(не важно это пока)

О других свойствах типичной инфляции

Amount of inflation

The total number of ' e-folds' of accelerated expansion are

$$N_{\text{tot}} := \int_{a_I}^{a_E} \mathrm{d}\ln a = \int_{t_I}^{t_E} H(t)\mathrm{d}t,$$

(????)

$$N_{\text{tot}} = \int_{\phi_I}^{\phi_E} \frac{1}{\sqrt{2\epsilon_{\text{v}}}} \frac{|\mathrm{d}\phi|}{M_{\text{pl}}},$$

$$N_{\text{cmb}} = \int_{\phi_{\text{cmb}}}^{\phi_E} \frac{1}{\sqrt{2\epsilon_{\text{v}}}} \frac{|\mathrm{d}\phi|}{M_{\text{pl}}} \approx 60$$

Аргументы, почему инфляции не может быть и ответы на них (??!?!?)

(типа там ранее чем 10^{-42}с у нас что-то было кроме большого взрыва? хз, не до этого вопроса.)

1.2.4 Description of equilibrium and non equilibrium processes

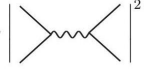
(???? статфиз подтяну параллельно, по статфизу уже пробелы, потому что эти темы не очень понятны сразу.)

Laws of local Thermal Equilibrium (??????????????)

We compare the rate of interactions Γ and the rate of expansion H.

$$\text{If } \Gamma \gg H \qquad \text{then} \qquad t_c := \frac{1}{\Gamma} \ll t_H := \frac{1}{H}.$$

Below $T \sim 10^{16}$GeV, but above 100 GeV, it is therefore satisfied.

Из $\sigma \sim$  $\sim \frac{\alpha^2}{T^2}$, ($\alpha := g_A^2/4\pi$ is a structure constant for gauge boson A)

$$\Gamma = n\sigma v \quad \sim \quad T^3 \cdot \frac{\alpha^2}{T^2} = \alpha^2 T.$$

We wish to compare this to the Hubble rate $H \sim \sqrt{\rho}/M_{\rm pl}$, where by dimensional argument (??????) $\rho \sim T^4$, so

$$H \sim \frac{T^2}{M_{\rm pl}^2}.$$

$$\frac{\Gamma}{H} \sim \frac{\alpha^2 M_{\rm pl}}{T} \sim \frac{10^{16}\text{GeV}}{T} \qquad \alpha \sim 0.01.$$

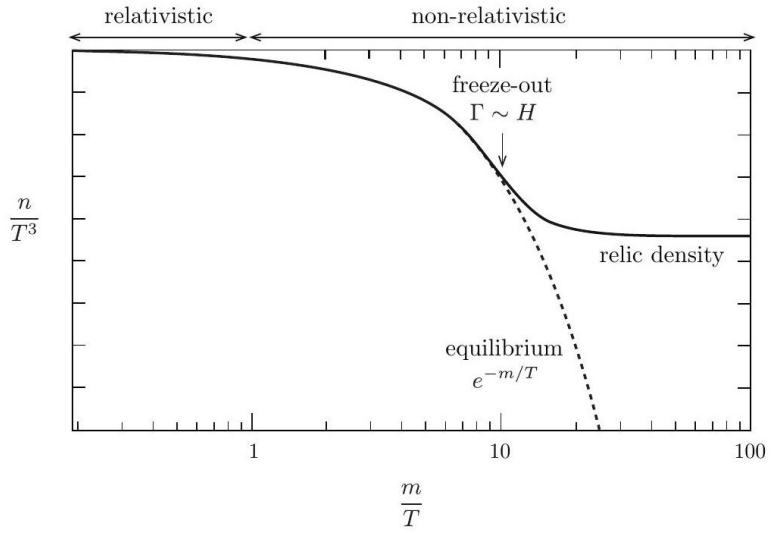
$$f(E) = \frac{1}{e^{E/T} \pm 1},$$

$$\rho_r \propto \sum_i \int \mathrm{d}^3 p f_i(p) E_i(p) = \frac{\pi^2}{30} g_\star(T) T^4,$$

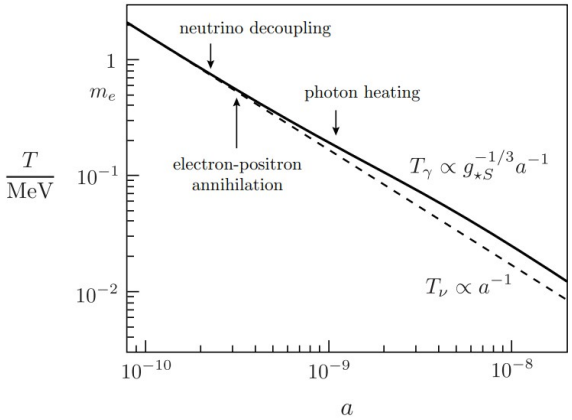
What is an entropy of the universe and what are others entropies, that are used in cosmology? (?!??!)

(отдельно напишу, важный вопрос)

Decoupling and Freeze-Out



(???? выведу еще этот график, пока просто верю, что это так)



Neutrinos are decoupled and their temperature redshifts simply as

$$T_\nu \propto a^{-1}.$$

The energy density of the electron-positron pairs is transferred to the photon gas whose temperature therefore redshifts more slowly,

$$T_\gamma \propto g_{\star S}^{-1/3} a^{-1}.$$

By the weak force $\sigma \sim$ $\sim G_F^2 T^2$,; $G_F \sim \alpha/M_W^2 \sim 1.17 \cdot 10^{-5}\mathrm{GeV}^{-2}$.

$$\frac{\Gamma}{H} \sim \frac{\alpha^2 M_{\rm pl} T^3}{M_W^4} \sim \left(\frac{T}{1\mathrm{MeV}}\right)^3,$$

which drops below unity at $T_{\rm dec} \sim 1\mathrm{MeV}$.

Counting relativistic gases

$$n := \frac{g}{(2\pi)^3}\int \mathrm{d}^3 p f(p) \qquad \text{(particles density)}$$

$$\rho = \frac{g}{(2\pi)^3}\int \mathrm{d}^3 p f(p) E(p) \qquad \text{(energy density)}$$

$$P = \frac{g}{(2\pi)^3}\int \mathrm{d}^3 p f(p)\frac{p^2}{3E} \qquad \text{(pressure)} \qquad f(p) := \frac{1}{e^{(E(p)-\mu)/T} \pm 1}$$

The factor of $p^2/3E$ is the usual $\langle p_x v_x\rangle = \langle \mathbf{p}\cdot\mathbf{v}\rangle/3$ factor from kinetic theory, with $\mathbf{v} = \mathbf{p}/E$.
(?? кст, а там хим пот от T не зависит в этих моделях???)
Chemical equilibrium: $\mu_1 + \mu_2 = \mu_3 + \mu_4$. Thermal equilibrium: T is the same.
For photon: $\mu_\gamma = 0$. For antiparticle: $\mu_{\bar{X}} = -\mu_X$.
At early times all $\mu_i \approx 0$ (???? додумаю, почему, пока верю, мб это и не доказано?)

$$x \equiv m/T \qquad \xi \equiv p/T$$

$$n = \frac{g}{2\pi^2}T^3 I_\pm(x), \qquad I_\pm(x) := \int_0^\infty \mathrm{d}\xi \frac{\xi^2}{\exp\left[\sqrt{\xi^2+x^2}\right] \pm 1}$$

$$\rho = \frac{g}{2\pi^2}T^4 J_\pm(x), \qquad J_\pm(x) := \int_0^\infty \mathrm{d}\xi \frac{\xi^2\sqrt{\xi^2+x^2}}{\exp\left[\sqrt{\xi^2+x^2}\right] \pm 1}$$

$$\int_0^\infty \mathrm{d}\xi \frac{\xi^n}{e^\xi - 1} = \zeta(n+1)\Gamma(n+1), \qquad \zeta(z) \text{ - Riemann zeta-function}$$

$$\int_0^\infty \mathrm{d}\xi \xi^n e^{-\xi^2} = \frac{1}{2}\Gamma\left(\frac{1}{2}(n+1)\right),$$

Relativistic limit

$x \to 0$ ($m \ll T$), so $I_\pm(0) = \int_0^\infty \mathrm{d}\xi \frac{\xi^2}{e^\xi \pm 1}$ and for bosons $I_-(0) = 2\zeta(3) \approx 2 \cdot 1.2$. For fermions, $I_+(0) = I_-(0) - 2 \times \left(\frac{1}{2}\right)^3 I_-(0) = \frac{3}{4} I_-(0)$ (?? ещё раз выведу!)

$$n = \frac{\zeta(3)}{\pi^2} g T^3 \begin{cases} 1 & \text{bosons} \\ \frac{3}{4} & \text{fermions} \end{cases}$$

$$\rho = \frac{\pi^2}{30} g T^4 \begin{cases} 1 & \text{bosons} \\ \frac{7}{8} & \text{fermions} \end{cases}$$

If $\mu \neq 0$ and $T \gg m$, for fermions $n \approx \frac{gT^3}{2\pi^2} \int\limits_0^\infty \mathrm{d}\xi \frac{\xi^2}{e^{\xi-\chi}+1}$ ($\xi := p/T$ and $\chi := \mu/T$), so

$$n - \bar{n} = \frac{g}{2\pi^2} \int\limits_0^\infty \mathrm{d}p p^2 \left(\frac{1}{e^{(p-\mu)/T}+1} - \frac{1}{e^{(p+\mu)/T}+1} \right) = \frac{gT^3}{6\pi^2} \left[\pi^2 \left(\frac{\mu}{T}\right) + \left(\frac{\mu}{T}\right)^3 \right] \quad \mu_{\bar{X}} = -\mu_X$$

We just need to carefully evaluate two approximated integrals.

Non-relativistic Limit

$x \gg 1 (m \gg T)$, so $I_\pm(x) \approx \int\limits_0^\infty \mathrm{d}\xi \frac{\xi^2}{e\sqrt{\xi^2+x^2}} \approx \sqrt{\frac{\pi}{2}} x^{3/2} e^{-x}$. (??? перевыеду еще раз интеграл)

$$n = g \left(\frac{mT}{2\pi} \right)^{3/2} e^{-m/T}$$

$$\rho = mn + \frac{3}{2} nT.$$

$$P = nT \ll \rho = mn.$$

If $\mu \neq 0$, $n = g \left(\frac{mT}{2\pi}\right)^{3/2} e^{-(m-\mu)/T}$,

$$n - \bar{n} = 2g \left(\frac{mT}{2\pi} \right)^{3/2} e^{-m/T} \operatorname{sh}\left(\frac{\mu}{T}\right)$$

(??? строчка, почему? забыл, типичная задача)

Effective Number of Relativistic Species

$$\rho_r = \sum_i \rho_i = \frac{\pi^2}{30} g_\star(T) T^4, \qquad \text{(r - radiation)}$$

$g_\star(T)$ - the effective number of relativistic degrees of freedom at the temperature T. The sum over particle species may receive two types of contributions:

- Relativistic species in thermal equilibrium with the photons, $T_i = T \gg m_i$,

$$g_\star^{th}(T) = \sum_{i=b} g_i + \frac{7}{8} \sum_{i=f} g_i.$$

When the temperature drops below the mass m_i of a particle species, it becomes nonrelativistic and is removed from the sum in (3.2.55). Away from mass thresholds, the thermal contribution is independent of temperature.

- Relativistic species that are not in thermal equilibrium with the photons, $T_i \neq T \gg m_i$,

$$g_\star^{dec}(T) = \sum_{i=b} g_i \left(\frac{T_i}{T} \right)^4 + \frac{7}{8} \sum_{i=f} g_i \left(\frac{T_i}{T} \right)^4.$$

We have allowed for the decoupled species to have different temperatures T_i. This will be relevant for neutrinos after e^+e^-annihilation (see §3.2.4).

$g_b = 28$ photons (2), $W^\pm$and $Z^0(3 \cdot 3)$, gluons $(8 \cdot 2)$, and Higgs (1)
$g_f = 90$ quarks $(6 \cdot 12)$, charged leptons $(3 \cdot 4)$, and neutrinos $(3 \cdot 2)$

$$g_\star = g_b + \frac{7}{8} g_f = 106.75$$

Conservation of Entropy (????????)

(here is a very crazy idea about the change of entropy of the universe. later I'll add a discussion of sense of it!)

Entropy in equilibrium pf particles and acceleration is conserved. It follows from $T\,\mathrm{d}S = \mathrm{d}U + P\,\mathrm{d}V$, $U = \rho V$ and the continuity equation $\dot{\rho} + 3H(\rho + P) = 0$. (because $V \propto a^3$.)

entropy density: $$s := S/V = \frac{\rho + P}{T} = \sum_i \frac{\rho_i + P_i}{T_i} \equiv \frac{2\pi^2}{45} g_{\star S}(T) T^3,$$

Effective degrees of freedom in entropy:

$$g_{\star S}(T) := g^{th}_{\star S}(T) + g^{dec}_{\star S}(T)$$

Note that for species in thermal equilibrium $g^{th}_{\star S}(T) = g^{th}_{\star}(T)$. However, given that $s_i \propto T_i^3$, for decoupled species we get

$$g^{dec}_{\star S}(T) := \sum_{i=b} g_i \left(\frac{T_i}{T}\right)^3 + \frac{7}{8} \sum_{i=f} g_i \left(\frac{T_i}{T}\right)^3 \neq g^{dec}_{\star}(T)$$

The conservation of entropy has two important consequences:

- It implies that $s \propto a^{-3}$. The number of particles in a comoving volume is therefore proportional to the number density n_i divided by the entropy density

$$N_i := \frac{n_i}{s}.$$

If particles are neither produced nor destroyed, then $n_i \propto a^{-3}$ and N_i is constant. This is case, for example, for the total baryon number after baryogenesis, $n_B/s := (n_b - n_{\bar{b}})/s$.

$$T \propto g_{\star S}^{-1/3} a^{-1}.$$

(???????)

Away from particle mass thresholds $g_{\star S}$ is approximately constant and $T \propto a^{-1}$, as expected. The factor of $g_{\star S}^{-1/3}$ accounts for the fact that whenever a particle species becomes non-relativistic and disappears, its entropy is transferred to the other relativistic species still present in the thermal plasma, causing T to decrease slightly less slowly than a^{-1}. We will see an example in the next section (cf. fig. 3.5).

Substituting $T \propto g_{\star S}^{-1/3} a^{-1}$ into the Friedmann equation

$$H = \frac{1}{a}\frac{da}{dt} \simeq \left(\frac{\rho_r}{3M_{\mathrm{pl}}^2}\right)^{1/2} \simeq \frac{\pi}{3}\left(\frac{g_\star}{10}\right)^{1/2} \frac{T^2}{M_{\mathrm{pl}}},$$

we reproduce the usual result for a radiation dominated universe, $a \propto t^{1/2}$, except that there is a change in the scaling every time $g_{\star S}$ changes. For $T \propto t^{-1/2}$, we can integrate the Friedmann equation and get the temperature as a function of time

$$\frac{T}{1\mathrm{MeV}} \simeq 1.5 g_\star^{-1/4} \left(\frac{1\mathrm{sec}}{t}\right)^{1/2}.$$

Boltzmann equation*

$$\frac{dn_i}{dt} + 3\frac{\dot{a}}{a} n_i = 0$$

$$\frac{1}{a^3}\frac{d\left(n_i a^3\right)}{dt} = C_i\left[\{n_j\}\right]$$

$$1+2 \rightleftarrows 3+4,$$

$$\frac{1}{a^3}\frac{d\left(n_1 a^3\right)}{dt} = -\alpha n_1 n_2 + \beta n_3 n_4$$

$$\frac{1}{a^3}\frac{d\left(n_1 a^3\right)}{dt} = -\langle\sigma v\rangle\left[n_1 n_2 - \left(\frac{n_1 n_2}{n_3 n_4}\right)_{\rm eq} n_3 n_4\right]$$

$$\frac{d\ln N_1}{d\ln a} = -\frac{\Gamma_1}{H}\left[1 - \left(\frac{N_1 N_2}{N_3 N_4}\right)_{\rm eq}\frac{N_3 N_4}{N_1 N_2}\right] \qquad \Gamma_1 := n_2\langle\sigma v\rangle \qquad N_i := n_i/s$$

(?????) For $\Gamma_1 \gg H$, the natural state of the system is chemical equilibrium. Imagine that we start with $N_1 \gg N_1^{\rm eq}$ (while $N_i \sim N_i^{\rm eq}, i = 2,3,4$). The r.h.s. of (3.3.86) then is negative, particles of type 1 are destroyed and N_1 is reduced towards the equilibrium value $N_1^{\rm eq}$. Similarly, if $N_1 \ll N_1^{\rm eq}$, the r.h.s. of (3.3.86) is positive and N_1 is driven towards $N_1^{\rm eq}$. The same conclusion applies if several species deviate from their equilibrium values. As long as the interaction rates are large, the system quickly relaxes to a steady state where the r.h.s. of (3.3.86) vanishes and the particles assume their equilibrium abundances.

1.2.5 Nucleosynthesis, baryogenesis, decouplings in a nutshell

Equilibrium properties

Neutrino Decoupling

$$\nu_e + \bar\nu_e \leftrightarrow e^+ + e^-,$$
$$e^- + \bar\nu_e \leftrightarrow e^- + \bar\nu_e$$

The cross section for these interactions was estimated in (3.1.9), $\sigma \sim G_F^2 T^2$, and hence it was found that $\Gamma \sim G_F^2 T^5$. As the temperature decreases, the interaction rate drops much more rapidly that the Hubble rate $H \sim T^2/M_{\rm pl}$:

$$\frac{\Gamma}{H} \sim \left(\frac{T}{1{\rm MeV}}\right)^3.$$

physical momentum of a particle scales as $p \propto a^{-1}$. It is therefore convenient to define the time-independent combination $q := ap$, so that the neutrino number density is

$$n_\nu \propto a^{-3}\int {\rm d}^3 q \frac{1}{\exp\left(q/aT_\nu\right)+1}.$$

Electron-Positron Annihilation

$$e^+ + e^- \leftrightarrow \gamma + \gamma$$

$$g_{\star S}^{th} = \begin{cases} 2 + \frac{7}{8}\cdot 4 = 5.5 & T \gtrsim m_e \\ 2 & T < m_e \end{cases}$$

Since, in equilibrium, $g_{\star S}^{\rm th}\left(aT_\gamma\right)^3$ remains constant, we find that aT_γ increases after electron-positron annihilation, $T < m_e$, by a factor $(11/4)^{1/3}$, while aT_ν remains the same.

that the temperature of neutrinos is slightly lower than the photon temperature after e^+e^- annihilation,

$$T_\nu = \left(\frac{4}{11}\right)^{1/3} T_\gamma.$$

For $T \ll m_e$, the effective number of relativistic species (in energy density and entropy) therefore is

$$g_\star = 2 + \frac{7}{8} \cdot 2N_{\rm eff} \left(\frac{4}{11}\right)^{4/3} = 3.36,$$
$$g_{\star S} = 2 + \frac{7}{8} \cdot 2N_{\rm eff} \left(\frac{4}{11}\right) = 3.94,$$

where we have introduced the parameter $N_{\rm eff}$ as the effective number of neutrino species in the universe. If neutrinos decoupling was instantaneous then we have $N_{\rm eff} = 3$.

Cosmic Neutrino Background

$$n_\nu = \frac{3}{4} N_{\rm eff} \cdot \frac{4}{11} n_\gamma$$

$$\rho_\nu = \frac{7}{8} N_{\rm eff} \left(\frac{4}{11}\right)^{4/3} \rho_\gamma \Rightarrow \Omega_\nu h^2 \approx 1.7 \cdot 10^{-5} \quad (m_\nu = 0)$$

$$\Omega_\nu h^2 \approx \frac{\sum m_{\nu,i}}{94\text{eV}}$$

Dark Matter particle dynamics (!!??)

(скоро это решу!)

Dark Matter Freeze-Out

$$X + \bar{X} \leftrightarrow \ell + \bar{\ell}$$

$$\frac{dN_X}{dt} = -s\langle\sigma v\rangle \left[N_X^2 - (N_X^{\rm eq})^2\right] \qquad N_X^{\rm eq} := n_X^{\rm eq}/s$$

$$x \equiv \frac{M_X}{T}; \qquad \frac{dx}{dt} = \frac{d}{dt}\left(\frac{M_X}{T}\right) = -\frac{1}{T}\frac{{\rm d}T}{dt} x \simeq Hx; \qquad H = \frac{H(M_X)}{x^2} \quad (T \propto a^{-1})$$

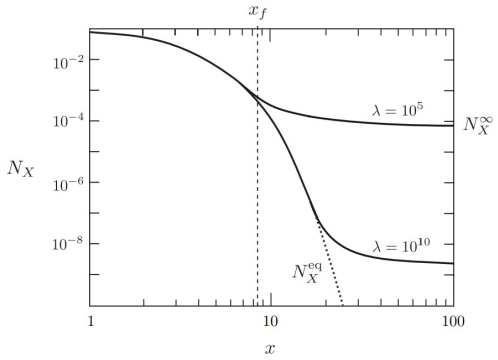
Riccati equation: $$\frac{dN_X}{dx} = -\frac{\lambda}{x^2}\left[N_X^2 - (N_X^{\rm eq})^2\right], \qquad \lambda := \frac{2\pi^2}{45} g_{\star S} \frac{M_X^3 \langle\sigma v\rangle}{H(M_X)}.$$

The final relic abundance, $N_X^\infty := N_X(x = \infty)$, determines the freeze-out density of dark matter.

$$N_X \gg N_X^{\rm eq} \qquad \frac{dN_X}{dx} \simeq -\frac{\lambda N_X^2}{x^2} \quad (x > x_f).$$

$$\frac{1}{N_X^\infty} - \frac{1}{N_X^f} = \frac{\lambda}{x_f} \qquad (N_X^f := N_X(x_f), \quad x \in \{x_f, \infty\})$$

$$N_X^f \gg N_X^\infty \qquad N_X^\infty \simeq \frac{x_f}{\lambda}$$



WIMP Miracle (??????)

We relate the freeze-out abundance of dark matter relics to the dark matter density today:

$$\Omega_X := \frac{\rho_{X,0}}{\rho_{\text{crit },0}} = \frac{M_X n_{X,0}}{3M_{\rm pl}^2 H_0^2} = \frac{M_X N_{X,0} s_0}{3M_{\rm pl}^2 H_0^2} = M_X N_X^\infty \frac{s_0}{3M_{\rm pl}^2 H_0^2}.$$

where we have used that the number of WIMPs is conserved after freeze-out, i.e. $N_{X,0} = N_X^\infty$. Substituting $N_X^\infty = x_f/\lambda$ and $s_0 := s(T_0)$, we get

$$\Omega_X = \frac{H(M_X)}{M_X^2} \frac{x_f}{\langle\sigma v\rangle} \frac{g_{\star S}(T_0)}{g_{\star S}(M_X)} \frac{T_0^3}{3M_{\rm pl}^2 H_0^2}$$

where we have used (3.3.92) and (3.2.62). Using (3.2.67) for $H(M_X)$, gives

$$\Omega_X = \frac{\pi}{9} \frac{x_f}{\langle\sigma v\rangle} \left(\frac{g_\star(M_X)}{10}\right)^{1/2} \frac{g_{\star S}(T_0)}{g_{\star S}(M_X)} \frac{T_0^3}{M_{\rm pl}^3 H_0^2}.$$

Finally, we substitute the measured values of T_0 and H_0 and use $g_{\star S}(T_0) = 3.91$ and $g_{\star S}(M_X) = g_\star(M_X)$:

$$\Omega_X h^2 \sim 0.1 \left(\frac{x_f}{10}\right) \left(\frac{10}{g_\star(M_X)}\right)^{1/2} \frac{10^{-8}\text{GeV}^{-2}}{\langle\sigma v\rangle}.$$

This reproduces the observed dark matter density if

$$\sqrt{\langle\sigma v\rangle} \sim 10^{-4}\text{GeV}^{-1} \sim 0.1\sqrt{G_F}.$$

The fact that a thermal relic with a cross section characteristic of the weak interaction gives the right dark matter abundance is called the WIMP miracle.

(читал, но нужно додумать!!)

Recombination (formation of atoms from electrons and nuclei)

Saha Equilibrium

$$e^- + p^+ \leftrightarrow \text{H} + \gamma$$

Free electron fraction $X_c := \frac{n_e}{n_b}$ n_b is the baryon density, $\eta := 5.5 \cdot 10^{-10} (\Omega_b h^2/0.020)$ is the baryon-to-photon ratio, $B_{\rm H} := m_p + m_c - m_{\rm H} = 13.6\text{eV}$.

The proof is as follows. Since $T < m_i, i = \{e, p, \text{H}\}$,

$$n_i^{\rm eq} = g_i \left(\frac{m_i T}{2\pi}\right)^{3/2} \exp\left(\frac{\mu_i - m_i}{T}\right)$$

$$\mu_p + \mu_c = \mu_{\rm H} \qquad (\mu_\gamma = 0)$$

$$\left(\frac{n_{\rm H}}{n_e^2}\right)_{\rm eq} = \left(\frac{2\pi}{m_e T}\right)^{3/2} e^{B_{\rm H}/T}$$

here we have taken into account $m_{\rm H} \approx m_p$, $B_{\rm H} := m_p + m_c - m_{\rm H} = 13.6{\rm eV}$ and $g_p = g_e = 2$, $g_{\rm H} = 4$.

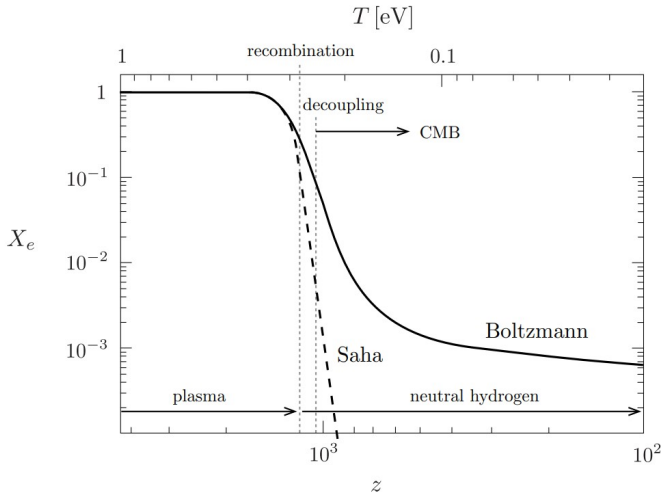
$$n_b = \eta n_\gamma = \eta \cdot \frac{2\zeta(3)}{\pi^2} T^3$$

where $\eta = 5.5 \cdot 10^{-10}\,(\Omega_b h^2/0.020)$ is the baryon-to-photon ratio.

$n_b \approx n_p + n_{\rm H} = n_c + n_{\rm H}$ and hence

$$\frac{1 - X_e}{X_c^2} = \frac{n_H}{n_e^2} n_b$$

$$\left(\frac{1 - X_e}{X_e^2}\right)_{\rm eq} = \frac{2\zeta(3)}{\pi^2}\eta \cdot \left(\frac{2\pi T}{m_e}\right)^{3/2} e^{B_{\rm H}/T} \quad \text{(Saha equation)}$$



(?? как решать ур Саха? Пока просто берем численный график как решение и все.)

Hydrogen Recombination

Let's estimate temperature when a lot of hydrogen is formed.

$$T_{\rm rec}\ \text{is the temperature where } X_e = 10^{-1}$$

From Saha eq.: (? this is the exercise, I'll do it later)

$$T_{\rm rec}\ \approx 0.3{\rm eV} \simeq 3600\ {\rm K}.$$

$T_{\rm rec} \ll B_{\rm H} = 13.6{\rm eV}$ because since there are very many photons for each hydrogen atom ($\eta \sim 10^{-9} \ll 1$), Even when $T < B_{\rm H}$, the high-energy tail of the photon Plank distribution contains photons with energy $E > B_{\rm H}$ so that they are still produces from hydrogen recombination.

(??? потом мб глубже пойму)

From $T_{\rm rec}\ = T_0\,(1 + z_{\rm rec}\,)$, with $T_0 = 2.7$ K

$$z_{rec} \approx 1320.$$

Since matter-radiation equality is at $z_{eq} \simeq 3500$, we conclude that recombination occurred in the matter-dominated era.

From $a(t) = (t/t_0)^{2/3}$,

$$t_{rec} = \frac{t_0}{(1 + z_{rec})^{3/2}} \sim 290000{\rm yrs}$$

Photon Decoupling

Let's estimate temperature when photons will be no more coupled with electrons.

$$e^- + \gamma \leftrightarrow e^- + \gamma$$
$$\Gamma_\gamma \approx n_e \sigma_T \qquad \sigma_T \approx 2 \cdot 10^{-3} \text{MeV}^{-2} \quad \text{(Thomson)}$$

Since $\Gamma_\gamma \propto n_e$, the interaction rate decreases as the density of free electrons drops.

Photons and electrons decouple roughly when the interaction rate becomes smaller than the expansion rate, (???)

$$\Gamma_\gamma (T_{dec}) \sim H (T_{dec})$$
$$\Gamma_\gamma (T_{dec}) = n_b X_e (T_{dec}) \sigma_T = \frac{2\zeta(3)}{\pi^2} \eta \sigma_T X_e (T_{dec}) T_{dec}^3$$
$$H (T_{dec}) = H_0 \sqrt{\Omega_m} \left(\frac{T_{dec}}{T_0} \right)^{3/2}.$$

we have

$$X_e (T_{dec}) T_{dec}^{3/2} \sim \frac{\pi^2}{2\zeta(3)} \frac{H_0 \sqrt{\Omega_m}}{\eta \sigma_T T_0^{3/2}}.$$

Using the Saha equation for $X_e (T_{dec})$, we find

$$T_{dec} \sim 0.27 \text{eV}$$
$$z_{dec} \sim 1100$$
$$t_{dec} \sim 380000 \text{yrs}$$

Although T_{dec} isn't far from T_{rec}, the ionization fraction decreases significantly between recombination and decoupling, $X_e (T_{\text{rec}}) \simeq 0.1 \to X_e (T_{dec}) \simeq 0.01$. This shows that a large degree of neutrality is necessary for the universe to become transparent to photon propagation.

(это в самое начало мб тоже напишу)

Electron Freeze-Out

We will not have Saha's behavior, since there will be no more equilibrium later.

$$\frac{1}{a^3} \frac{d (n_e a^3)}{dt} = -\langle \sigma v \rangle \left[n_e^2 - (n_e^{\text{eq}})^2 \right]$$
$$\langle \sigma v \rangle \simeq \sigma_T \left(\frac{B_{\text{H}}}{T} \right)^{1/2} \qquad \text{(belive in this)}$$

Writing $n_e = n_b X_e$ and using that $n_b a^3 = \text{const.}$, we find

$$\frac{dX_e}{dx} = -\frac{\lambda}{x^2} \left[X_e^2 - (X_e^{\text{eq}})^2 \right]; \qquad x := B_{\text{H}}/T$$
$$\lambda := \left[\frac{n_b \langle \sigma v \rangle}{xH} \right]_{x=1} = 3.9 \cdot 10^3 \left(\frac{\Omega_b h}{0.03} \right)$$

(??? need to derive!!! use matter-domination??)

We use a solution for DM, since there was the same equation

$$X_e^\infty \simeq \frac{x_f}{\lambda} = 0.9 \cdot 10^{-3} \left(\frac{x_f}{x_{rec}} \right) \left(\frac{0.03}{\Omega_b h} \right).$$

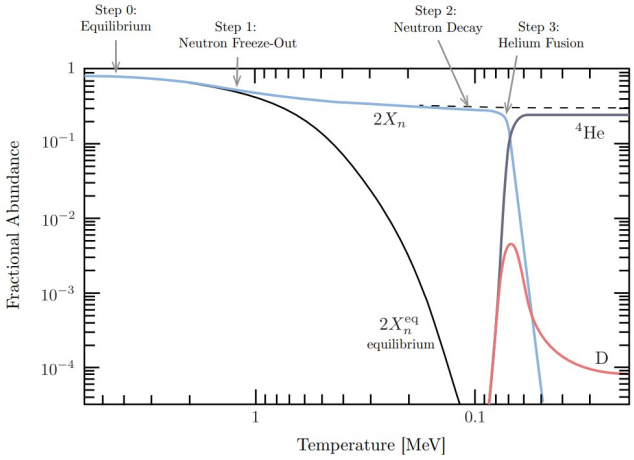
Assuming that freeze-out occurs close to the time of recombination, $x_{rec} \approx 45$, we capture the relic electron abundance pretty well (see fig. 3.8).

From $\Gamma_e (T_f) \sim H (T_f)$,

$$T_f : \qquad X_e (T_f) T_f = \frac{\pi^2}{2\zeta(3)} \frac{H_0 \sqrt{\Omega_m}}{\eta \sigma_T T_0^{3/2} B_{\text{H}}^{1/2}}$$

From Saha equation $T_f \sim 0.25 \text{eV}$ and hence $x_f \sim 54$.

Big Bang Nucleosynthesis (formation of hydrogen, helium, lithium) (????)



Step 0: Equilibrium Abundances

$$\frac{n_{\rm He}}{n_{\rm H}} \sim \frac{1}{16}$$

$$\frac{n_{\rm He}}{n_{\rm H}} \sim \frac{1}{16}$$

1. No elements heavier than helium. Essentially no elements heavier than helium are produced at appreciable levels. So the only nuclei that we need to track are hydrogen and helium, and their isotopes: deuterium, tritium, and ^{3}He.

2. Only neutrons and protons above 0.1MeV. Above $T \approx 0.1$MeV only free protons and neutrons exist, while other light nuclei haven't been formed yet. Therefore, we can first solve for the neutron/proton ratio and then use this abundance as input for the synthesis of deuterium, helium, etc.

Let us demonstrate that we can indeed restrict our attention to neutrons and protons above 0.1MeV. In order to do this, we compare the equilibrium abundances of the different nuclei:

- First, we determine the relative abundances of neutrons and protons. In the early universe, neutrons and protons are coupled by weak interactions, e.g. β-decay and inverse β-decay

$$n + \nu_e \leftrightarrow p^+ + e^-,$$
$$n + e^+ \leftrightarrow p^+ + \bar{\nu}_e.$$

$$\mu_n = \mu_p \qquad \mu_e \approx 0 \qquad \mu_\nu \approx 0$$

$$n_i^{\rm eq} = g_i \left(\frac{m_i T}{2\pi}\right)^{3/2} \exp\left(\frac{\mu_i - m_i}{T}\right)$$

$$\left(\frac{n_n}{n_p}\right)_{\rm eq} = \left(\frac{m_n}{m_p}\right)^{3/2} e^{-(m_n - m_p)/T}$$

$$\left(\frac{n_n}{n_p}\right)_{\rm eq} = e^{-\mathcal{Q}/T} \qquad \mathcal{Q} := m_n - m_p = 1.30\text{MeV}$$

For $T \gg 1$MeV, there are therefore as many neutrons as protons. However, for $T < 1$MeV, the neutron fraction gets smaller. If the weak interactions would operate efficiently enough to maintain equilibrium indefinitely, then the neutron abundance would drop to zero. Luckily, in the real world the weak interactions are not so efficient.

- Next, for deuterium (an isotope of hydrogen (pn)):

$$n + p^+ \leftrightarrow \mathrm{D} + \gamma.$$

$$\mu_n + \mu_p = \mu_{\rm D} \qquad \mu_\gamma = 0.$$

$$\left(\frac{n_{\rm D}}{n_n n_p}\right)_{\rm eq} = \frac{3}{4}\left(\frac{m_{\rm D}}{m_n m_p}\frac{2\pi}{T}\right)^{3/2} e^{-(m_{\rm D}-m_n-m_p)/T},$$

($g_{\rm D} = 3$ and $g_p = g_n = 2$).

$$m_{\rm D} \approx 2m_n \approx 2m_p \approx 1.9{\rm GeV}$$
$$B_{\rm D} := m_n + m_p - m_{\rm D} = 2.22{\rm MeV}$$
$$\left(\frac{n_{\rm D}}{n_p}\right)_{\rm eq} = \frac{3}{4} n_n^{\rm eq}\left(\frac{4\pi}{m_p T}\right)^{3/2} e^{B_{\rm D}/T}.$$
$$n_n \sim n_b = \eta n_\gamma = \eta \cdot \frac{2\zeta(3)}{\pi^2} T^3.$$
$$\left(\frac{n_{\rm D}}{n_p}\right)_{\rm eq} \approx \eta\left(\frac{T}{m_p}\right)^{3/2} e^{B_{\rm D}/T}.$$

The smallness of the baryon-to-photon ratio η inhibits the production of deuterium until the temperature drops well beneath the binding energy $B_{\rm D}$, so that $e^{B_{\rm D}/T}$ can compete with $\eta \sim 10^{-9}$.

The same applies to all other nuclei. At temperatures above 0.1MeV, then, virtually all baryons are in the form of neutrons and protons. Around this time, deuterium and helium are produced, but the reaction rates are by now too low to produce any heavier elements.

Neutron nonequilibrium dynamics and Helium fussion

$$X_n := \frac{n_n}{n_n + n_p}$$

$$\left(\frac{n_n}{n_p}\right)_{\rm eq} = e^{-\mathcal{Q}/T} \qquad X_n^{\rm eq}(T) = \frac{e^{-\mathcal{Q}/T}}{1 + e^{-\mathcal{Q}/T}}$$

$$X_n^{\rm eq}(0.8{\rm MeV}) = 0.17. \qquad (T_{\rm final} \sim T_{\rm dec} \sim 0.8{\rm MeV})$$

$$X_n^\infty \sim X_n^{\rm eq}(0.8{\rm MeV}) \sim \frac{1}{6}$$

(there are more complicated ways for calculation of this in theory, I don't think they are very important)

We need to account for a finite lifetime of the neutron:

$$X_n(t) = X_n^\infty e^{-t/\tau_n} = \frac{1}{6} e^{-t/\tau_n} \qquad \tau_n = 886.7 \pm 0.8{\rm sec}$$

Helium is produced by:

$$n + p^+ \leftrightarrow {\rm D} + \gamma.$$
$${\rm D} + p^+ \leftrightarrow {}^3{\rm He} + \gamma,$$
$${\rm D} + {}^3{\rm He} \leftrightarrow {}^4{\rm He} + p^+.$$

Since the deuterium binding energy is rather small, the deuterium abundance becomes large rather late (at $T < 100{\rm keV}$). So although heavier nuclei have larger binding energies and hence would have larger equilibrium abundances, they cannot be formed until sufficient deuterium has become available. This is the deuterium bottleneck. Only when there is enough deuterium, can helium be produced.

$$T_{\rm nuc}: \qquad (n_{\rm D}/n_p)_{\rm eq} \sim 1$$

$$T_{\rm nuc} \sim 0.06{\rm MeV} \qquad \left(\left(\frac{n_{\rm D}}{n_p}\right)_{\rm eq} \approx \eta \left(\frac{T}{m_p}\right)^{3/2} e^{B_{\rm D}/T}.\right)$$

from $\frac{T}{1{\rm MeV}} \simeq 1.5 g_\star^{-1/4} \left(\frac{1{\rm sec}}{t}\right)^{1/2}$ with $g_\star = 3.38$

$$t_{\rm nuc} = 120{\rm sec} \left(\frac{0.1{\rm MeV}}{T_{\rm nuc}}\right)^2 \sim 330{\rm sec}$$

$$X_n\left(t_{\rm nuc}\right) \sim \frac{1}{8}$$

(??? а ниже - потом додумаю, пока не продумывал)

$$\frac{n_{\rm He}}{n_{\rm H}} = \frac{n_{\rm He}}{n_p} \simeq \frac{\frac{1}{2} X_n\left(t_{\rm nuc}\right)}{1 - X_n\left(t_{\rm nuc}\right)} \sim \frac{1}{2} X_n\left(t_{\rm nuc}\right) \sim \frac{1}{16},$$

Influence of parameters on nucleosynthesis

Changing the input, e.g. by new physics beyond the Standard Model (BSM) in the early universe, would change the predictions of BBN. In this way BBN is a probe of fundamental physics.

- $g_\star$: the number of relativistic degrees of freedom determines the Hubble parameter during the radiation era, $H \propto g_\star^{1/2}$, and hence affects the freeze-out temperature

$$G_F^2 T_f^5 \sim \sqrt{G_N g_\star} T_f^2 \quad \rightarrow \quad T_f \propto g_\star^{1/6}$$

Increasing $g_\star$ increases T_f, which increases the n/p ratio at freeze-out and hence increases the final helium abundance.

- τ_n : a large neutron lifetime would reduce the amount of neutron decay after freeze-out and therefore would increase the final helium abundance.

- $\mathcal{Q}$: a larger mass difference between neutrons and protons would decrease the n/p ratio at freeze-out and therefore would decrease the final helium abundance.

- η: the amount of helium increases with increasing η as nucleosythesis starts earlier for larger baryon density.

- G_N: increasing the strength of gravity would increase the freeze-out temperature, $T_f \propto G_N^{1/6}$, and hence would increase the final helium abundance.

- G_F: increasing the weak force would decrease the freeze-out temperature, $T_f \propto G_F^{-2/3}$, and hence would decrease the final helium abundance.

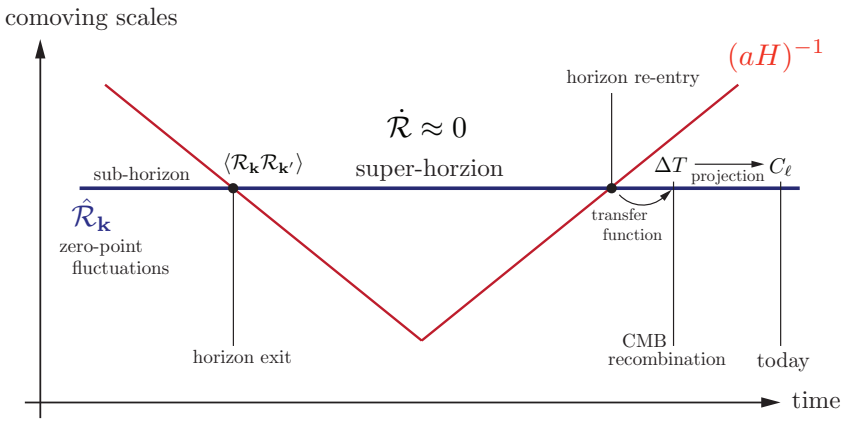
1.3 Structure formation and anysotropies in a nutshell

(тут очень большая и жесткая теория...)

1.3.1 Overview of fluctuations dynamics, parameters, power spectrum (!!!!!!)

Main properties and plots of fluctuation dynamics (??????!!!!)

(!!! только графики, ключевые формулы и ключевые слова!!!)



$$k^2\Phi + 3\frac{\dot{a}}{a}\left(\dot{\Phi} - \Psi\frac{\dot{a}}{a}\right) = 4\pi G a^2\left[\rho_{\rm m}\delta_m + 4\rho_r\Theta_{r,0}\right]$$
$$k^2(\Phi+\Psi) = -32\pi G a^2\rho_r\Theta_{r,2}$$

$m \in \{$ baryons, dark matter$\}$, $r \in \{$ neutrinos, photons$\}$

$$\rho_{\rm m}\delta_m \equiv \rho_{\rm dm}\delta + \rho_b\delta_{\rm b} \quad ; \quad \rho_r\Theta_{r,0} \equiv \rho_\gamma\Theta_0 + \rho_\nu\mathcal{N}_0$$
$$\rho_{\rm m}v_m \equiv \rho_{\rm dm}v + \rho_b v_{\rm b} \quad ; \quad \rho_r\Theta_{r,1} \equiv \rho_\gamma\Theta_1 + \rho_\nu\mathcal{N}_1.$$

A combination of B and E equations leads to an algebraic equation for the potential,

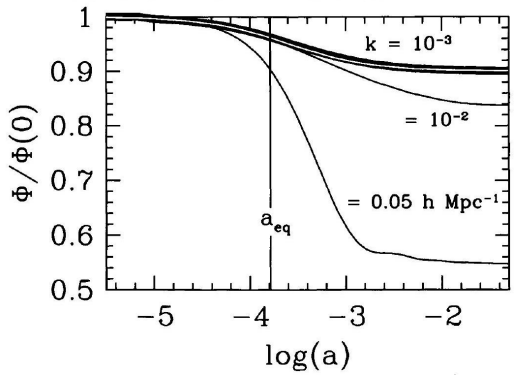
$$k^2\Phi = 4\pi G a^2\left[\rho_{\rm m}\delta_m + 4\rho_r\Theta_{r,0} + \frac{3aH}{k}\left(i\rho_{\rm m}v_m + 4\rho_r\Theta_{r,1}\right)\right].$$

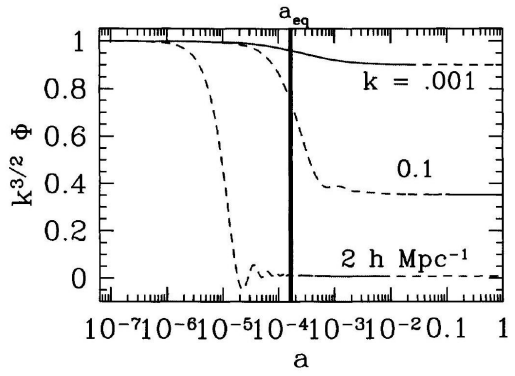
For tensor perturbations, $h_+$and $h_\times$; each of these evolves independently and satisfies

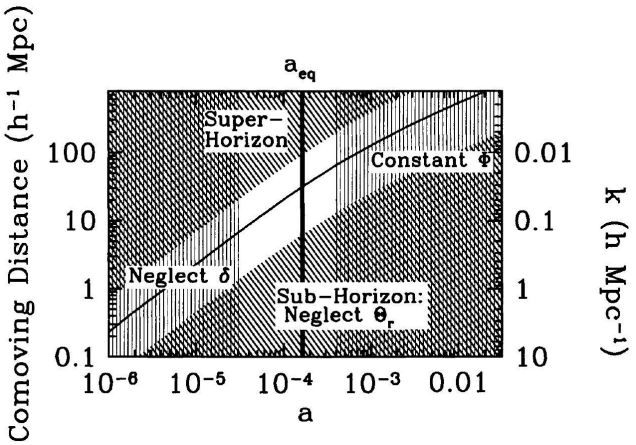
$$\ddot{h}_\alpha + 2\frac{\dot{a}}{a}\dot{h}_\alpha + k^2 h_\alpha = 0 \qquad \alpha \equiv +\times$$

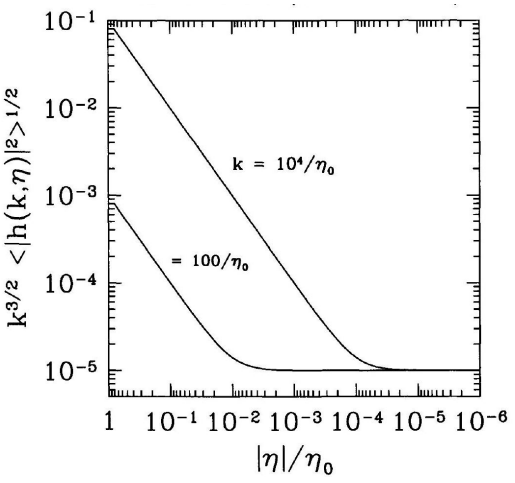
		RD		MD	
		Φ	δ_m (Δ_m)	Φ	δ_m (Δ_m)
$k \gg k_{\rm eq}$:	superhorizon	*const.*	*const.* (a^2)	–	–
	subhorizon	a^{-2}	$\ln a$	*const.*	a
$k \ll k_{\rm eq}$:	superhorizon	*const.*	*const.* (a^2)	*const.*	*const.* (a)
	subhorizon	–	–	*const.*	a

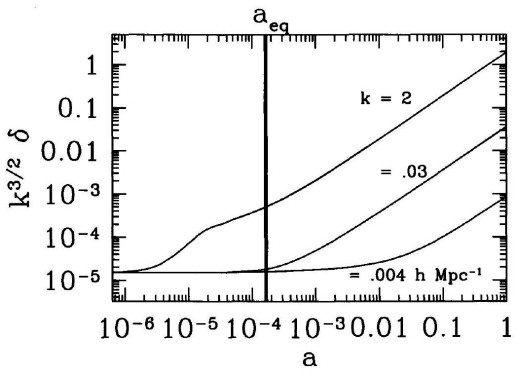
Gravitation potential perturbations: They are constant when the modes are outside the horizon. Two of the modes enter the horizon during the radiation era. While they are inside the horizon during the radiation era their amplitudes decrease as a^{-2}. The resulting amplitudes in the matter era are therefore strongly suppressed. During the matter era the potential is constant on all scales. The longest wavelength mode in the figure enters the horizon during the matter era, so its amplitude is only suppressed by the factor of $\frac{9}{10}$ coming from the radiation-to-matter transition.


1
0.9
0.8
0.7
0.6
0.5
Φ/Φ(0)
k = 10⁻³
= 10⁻²
= 0.05 h Mpc⁻¹
a_eq
−5
−4
−3
−2
log(a)


a_eq
1
0.8
0.6
0.4
0.2
0
k^3/2 Φ
k = .001
0.1
2 h Mpc⁻¹
10⁻⁷10⁻⁶10⁻⁵10⁻⁴10⁻³10⁻²0.1 1
a


Comoving Distance (h⁻¹ Mpc)
a_eq
Super-
Horizon
Constant Φ
Neglect δ
Sub-Horizon:
Neglect θ_r
100
10
1
0.1
0.01
0.1
1
10
k (h Mpc⁻¹)
10⁻⁶ 10⁻⁵ 10⁻⁴ 10⁻³ 0.01
a

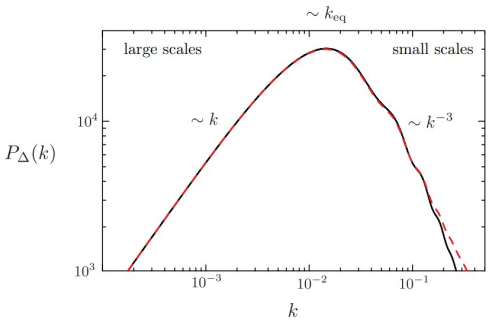

10⁻¹
10⁻²
10⁻³
10⁻⁴
10⁻⁵
k^3/2 <|h(k,η)|²>^1/2
k = 10⁴/η₀
= 100/η₀
1 10⁻¹ 10⁻² 10⁻³ 10⁻⁴ 10⁻⁵ 10⁻⁶
|η|/η₀



$$\Delta_m := \delta_m - 3\mathcal{H}v_m$$

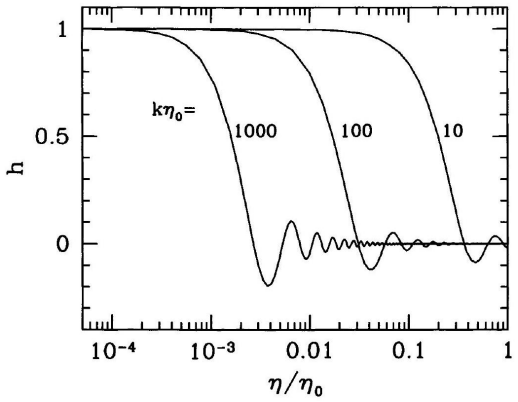
$$\Delta_{m,k}(z) = T(k,z)\mathcal{R}_k.$$

$$P_\Delta(k,z) \equiv |\Delta_{m,k}(z)|^2 = T^2(k,z)\,|\mathcal{R}_k|^2\,.$$



The matter power spectrum $P_\Delta(k)$ at $z = 0$ in linear theory (solid) and with non-linear corrections (dashed). On large scales, $P_\Delta(k)$ grows as k. The power spectrum turns over around $k_{\rm eq} \sim 0.01{\rm Mpc}^{-1}$ corresponding to the horizon size at matter-radiation equality. Beyond the peak, the power falls as k^{-3}. Visible are small amplitude baryon acoustic oscillations in the spectrum.

For gravitational waves:



Summary of evolution of fluctuations during Inflation

A defining characteristic of inflation is the behavior of the comoving Hubble radius, $1/(aH)$, which shrinks quasi-exponentially. A mode with comoving wavenumber k is called super-horizon when $k < aH$, and sub-horizon when $k > aH$. The inflaton is taken to be in a vacuum state, defined such that sub-horizon modes approach the Minkowski vacuum for $k \gg aH$. After a

mode exits the horizon, it is described by a classical probability distribution with variance given by the power spectrum evaluated at horizon crossing

$$P_{\rm s}(k) = \left.\frac{H^2}{2k^3}\frac{H^2}{\dot\phi^2}\right|_{k=aH} .$$

Inflation also produces fluctuations in the tensor part of the spatial metric. This corresponds to a spectrum of gravitational waves with power spectrum

$$P_{\rm t}(k) = \left.\frac{4}{k^3}\frac{H^2}{M_{\rm pl}^2}\right|_{k=aH} .$$

For slow-roll models the scalar and tensor spectra are expressed purely in terms of $V(\phi)$ and $\epsilon_{\rm v}$ (or $V_{,\phi}$)

$$\Delta^2_{\rm s}(k) \approx \left.\frac{1}{24\pi^2}\frac{V}{M_{\rm pl}^4}\frac{1}{\epsilon_{\rm v}}\right|_{k=aH} , \qquad \Delta^2_{\rm t}(k) \approx \left.\frac{2}{3\pi^2}\frac{V}{M_{\rm pl}^4}\right|_{k=aH} ,$$

where $\Delta^2(k) \equiv \frac{k^3}{2\pi^2}P(k)$. The scale dependence is given by

$$\begin{aligned} n_{\rm s} - 1 &\equiv \frac{d\ln\Delta^2_{\rm s}}{d\ln k} = 2\eta_{\rm v} - 6\epsilon_{\rm v} , \\ n_{\rm t} &\equiv \frac{d\ln\Delta^2_{\rm t}}{d\ln k} = -2\epsilon_{\rm v} . \end{aligned}$$

The tensor-to-scalar ratio is

$$r \equiv \frac{\Delta^2_{\rm t}}{\Delta^2_{\rm s}} = 16\epsilon_{\rm v} .$$

By the Lyth bound, r relates directly to total field excursion during inflation

$$\frac{\Delta\phi}{M_{\rm pl}} \approx \left(\frac{r}{0.01}\right)^{1/2} .$$

A large value for r therefore correlates both with a high scale for the inflationary energy and a super-Planckian field evolution.

below SUMMARY AND SPECTRAL INDICES of Initial conditions by Dodelson

In order to understand how scales which should be uncorrelated today are observed to have almost identical temperatures, we are virtually forced into the theory of inflation. In addition to explaining away the nagging fine-tuning problems of the standard cosmology, inflation is also a mechanism for generating primordial perturbations over the smooth universe.

Inflation predicts that quantum-mechanical perturbations in the very early universe are first produced when the relevant scales are causally connected. Then these scales are whisked outside the horizon by inflation, only to reenter much later to serve as initial conditions for the growth of structure and anisotropy in the universe. The perturbations are best described in terms of the Fourier modes. The mean of a given Fourier mode, for example for the gravitational potential, is zero:

$$\langle\Phi(\vec{k})\rangle = 0$$

Further, the perturbations to one Fourier mode are uncorrelated with those to another. However, a given mode has nonzero variance, so

$$\left\langle \Phi(\vec{k})\Phi^*(\vec{k}')\right\rangle = (2\pi)^3 P_\Phi(k)\delta^3\left(\vec{k}-\vec{k}'\right)$$

In the case of scalar perturbations, the ones of most importance for us, the power spectrum is given by Eq. (6.82). Perturbations to the tensor part of the metric are also produced and are also Gaussian with mean zero; the power spectrum of tensor modes is given by Eq. (6.59). The scalar spectrum depends on the slow roll parameter ϵ, defined in Eq. (6.35), which is

proportional to the derivative of the Hubble rate. Since the Hubble rate is close to constant during inflation - because of the dominance of potential energy- ϵ is typically small.

A spectrum in which $k^3 P_\Phi(k)$ is constant (i.e., does not depend on k) is called a scale-invariant (scale-free spectrum). Apart from small deviations encoded in the slow-roll parameters, both the scalar and the tensor perturbations are scale free. This is both a blessing and a curse. It is good because it is a fairly definite prediction, easy to test. It is unfortunate because a scale-free spectrum is what one might have expected even without the complex machinery of inflation. Indeed, a scale-free spectrum is also referred to as a Harrison-Zel'dovich-Peebles spectrum, crediting the smart people who first proposed it as the appropriate distribution for the initial conditions, a proposal that predates inflation by many years. This really is too bad, because if we observe a scale-free spectrum, and most present observations are consistent with this, then inflation cannot fairly claim all the credit. However, if we observe a small mixture of tensor modes and/or a small deviation from a scale-free spectrum, then this will go a long way toward convincing skeptics that inflation is responsible for the primordial perturbations.

To quantify the deviations from scale invariance, it is conventional to write the primordial power spectra as

$$P_\Phi(k) = \frac{8\pi}{9k^3}\frac{H^2}{\epsilon m_{\rm Pl}^2}\bigg|_{aH=k} := \frac{50\pi^2}{9k^3}\left(\frac{k}{H_0}\right)^{n-1}\delta_H^2\left(\frac{\Omega_m}{D_1(a=1)}\right)^2$$
$$P_h(k) = \frac{8\pi}{k^3}\frac{H^2}{m_{\rm Pl}^2}\bigg|_{aH=k} := A_T k^{n_T-3}$$

These equations serve to define the scalar and tensor amplitudes, δ_H (subscript H for amplitude at horizon crossing) and A_T, and the spectral indices, n and n_T. Note that this convention - which has become common - says that a scale-free scalar spectrum corresponds to $n = 1$, while $n_T = 0$ for a scale-free tensor spectrum. The factor of $\Omega_m/D_1(a=1)$, where Ω_m is the fraction of the critical density in matter today and D_1 is the growth function which will be defined in Chapter 7 (Eqs. (7.4) and (7.77)), is part of this convention. It is inconvenient at this stage because we have not even encountered the growth function yet, but it has become standard to include in the definition of δ_H (Liddle and Lyth, 1993; Bunn and White, 1997). The resulting expression for the matter power spectrum today looks much simpler when these factors are included here. We pay the price of complexity now for the benefit of simplicity later.

We can relate the primordial spectral indices n and n_T to the slow-roll parameters ϵ and δ. Consider first the tensor spectrum. By virtue of the definition in Eq. (6.100),

$$\frac{d\ln[P_h]}{d\ln k} = n_T - 3$$

The logarithmic derivative has two terms, first the trivial one $d\ln(k^{-3})/d\ln(k)$ which cancels the -3 here, leaving $n_T = 2d\ln H/d\ln(k)$. The logarithmic derivative of the Hubble rate at horizon crossing is a bit subtle:

$$\frac{d\ln H}{d\ln k}\bigg|_{aH=k} = \frac{k}{H}\frac{dH}{d\eta}\cdot\frac{d\eta}{dk}\bigg|_{aH=k}.$$

By definition (Eq. (6.35)), $\dot H = -aH^2\epsilon$, and $d\eta|_{aH=k}/dk = -\,d(aH)^{-1}|_{aH=k}/dk = 1/k^2$, so

$$\frac{d\ln H}{d\ln k}\bigg|_{aH=k} = -\frac{k}{H}\frac{aH^2\epsilon}{k^2}\bigg|_{aH=k} = -\epsilon.$$

Therefore, the primordial spectral index of tensor perturbations produced by inflation is

$$n_T = -2\epsilon$$

The scalar spectral index follows from a similar argument. Taking the logarithmic derivative of P_Φ leads to

$$n - 1 = \frac{d}{d\ln(k)}\left[\ln\left(H^2\right) - \ln(\epsilon)\right]$$

The derivative of H again gives -2ϵ while the logarithmic derivative of ϵ is $-2(\epsilon+\delta)$ (Exercise 12). So,

$$n = 1 - 4\epsilon - 2\delta.$$

The fact that the tensor index n_T is proportional to ϵ leads to one of the robust predictions of inflation. Many inflationary models have been proposed which offer different predictions for ϵ and δ. Almost all of these, however, maintain the feature that the ratio of tensor to scalar modes (which we saw earlier was proportional to ϵ) is directly related to the tensor spectral index (here also seen to be directly proportional to ϵ).

below Scale invariance and spectral indexes by Craps

A scale-invariant spectrum is one for which $k^3P(k)$ is k-independent. The spectra P_h and P_Ψ are not scale-invariant because the quantities H^2 and ϵ are to be evaluated at the time of horizon crossing, $aH = k$, which depends on k.

Scalar perturbations spectral index (tilt) $n_s \equiv n$: $\quad k^3P_\Psi(k) \sim k^{n-1}$.

The tensor spectral index n_T : $\quad k^3P_h(k) \sim k^{n_T}$.

In these conventions scale-invariant spectra of scalar and tensor perturbations would correspond to $n = 1$ and $n_T = 0$, respectively.

The deviations from scale invariance can be expressed in terms of the slow-roll parameters:

$$n_T = \frac{d\ln H^2}{d\ln k} = \frac{2k}{H}\frac{dH}{d\eta}\frac{d\eta}{dk}\bigg|_{aH=k}. \qquad \left(P_h(k) = \frac{8\pi G H^2}{k^3}\bigg|_{aH=k}\right)$$

$$n_T = \frac{2k}{H}\left(-aH^2\epsilon\right)\eta^2\bigg|_{aH=k} = -2\epsilon \qquad \epsilon = -\dot{H}/\left(aH^2\right); \quad aH \approx -1/\eta$$

So, ϵ controls both the tensor spectral index and the tensor-to-scalar ratio $P_h(k)/P_\Psi(k)$; this is a robust prediction of inflation.

$$n - 1 = \frac{d}{d\ln k}\left(\ln H^2 - \ln\epsilon\right)\bigg|_{aH=k},$$

the first term will again give -2ϵ and because of $\frac{d\epsilon}{d\eta} = 2aH\epsilon(\epsilon+\delta)$ the second term in parentheses will give $-2(\epsilon+\delta)$, so

$$n = 1 - 4\epsilon - 2\delta.$$

1.3.2 Perturbations setup from newtonian and relativisic physics

(суммирую Баумана и круто, минимум флуктуаций сделан)

Perturbed Newtonian fluid equations (continuity, Euler's eq.)

$$\partial_t\rho = -\nabla_{\boldsymbol{r}}\cdot(\rho\boldsymbol{u}), \qquad \text{(Continuity eq.)}$$

$$\left(\partial_t + \boldsymbol{u}\cdot\boldsymbol{\nabla}_{\boldsymbol{r}}\right)\boldsymbol{u} = -\frac{\nabla_{\boldsymbol{r}}P}{\rho} - \nabla_{\boldsymbol{r}}\Phi. \qquad \text{(Euler eq., basically } F = ma)$$

$$\nabla_r^2\Phi = 4\pi G\rho \qquad \text{(Poisson eq.)}$$

$$\rho(t,\boldsymbol{r}) = \bar{\rho}(t) + \delta\rho(t,\boldsymbol{r}).$$

Non-relativistic fluid in static space without gravity:

$$\partial_t\delta\rho = -\nabla_r\cdot(\bar{\rho}\boldsymbol{u}),$$
$$\bar{\rho}\partial_t\boldsymbol{u} = -\nabla_r\delta P.$$

$\partial_t^2\delta\rho - \nabla_r^2\delta P = 0; \qquad \delta P = c_s^2\delta\rho$ (adiab.); $\qquad \left(\partial_t^2 - c_s^2\nabla^2\right)\delta\rho = 0.$ c_s - speed of sound

$\delta\rho(t,\boldsymbol{r}) = \int \frac{\mathrm{d}^3k}{(2\pi)^3} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\delta\rho_{\boldsymbol{k}}(t) \qquad \left(\partial_t^2 + c_s^2k^2\right)\delta\rho_{\boldsymbol{k}} = 0, \qquad \delta\rho_{\boldsymbol{k}} = A_{\boldsymbol{k}}e^{i\omega_k t} + B_{\boldsymbol{k}}e^{-i\omega_k t}, \quad \omega_k := c_s k.$

Non-relativistic fluid in static space with gravity:

$$\left(\partial_t^2 - c_s^2\nabla_{\boldsymbol{r}}^2\right)\delta\rho = 4\pi G\bar{\rho}\delta\rho \qquad \nabla^2\delta\Phi = 4\pi G\delta\rho$$

$$\delta\rho = A\exp[i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})] \qquad \omega^2 = c_s^2k^2 - 4\pi G\bar{\rho}$$

$$\omega = 0 \quad \text{at} \quad k_{\mathrm{J}} := \frac{\sqrt{4\pi G\bar{\rho}}}{c_s}.$$

For small scales (large wavenumber), $k > k_{\mathrm{J}}$, the pressure dominates and we find the same oscillations as before. On large scales, $k < k_{\mathrm{J}}$, gravity dominates, the frequency ω becomes imaginary and the fluctuations grow exponentially.

$$\lambda_{\mathrm{J}} = \frac{2\pi}{k_{\mathrm{J}}} = c_s\sqrt{\frac{\pi}{G\bar{\rho}}}. \qquad \text{(Jeans' lengt)}$$

Non-relativistic fluid in expanding space:

$$\boldsymbol{r}(t) = a(t)\boldsymbol{x} \qquad \boldsymbol{\nabla}_{\boldsymbol{r}} = a^{-1}\nabla_{\boldsymbol{x}} \qquad \left(\frac{\partial}{\partial t}\right)_r \equiv \left(\frac{\partial}{\partial t}\right)_x + \left(\frac{\partial \boldsymbol{x}}{\partial t}\right)_r \cdot \boldsymbol{\nabla}_x = \left(\frac{\partial}{\partial t}\right)_x - H\boldsymbol{x}\cdot\boldsymbol{\nabla}_x$$

Continuity equation:

$$\left[\frac{\partial}{\partial t} - H\boldsymbol{x}\cdot\boldsymbol{\nabla}\right][\bar{\rho}(1+\delta)] + \frac{1}{a}\boldsymbol{\nabla}\cdot[\bar{\rho}(1+\delta)(Ha\boldsymbol{x} + \boldsymbol{v})] = 0$$

$$\delta := \frac{\delta\rho}{\bar{\rho}}$$

$$\frac{\partial\bar{\rho}}{\partial t} + 3H\bar{\rho} = 0 \qquad \text{(0-th order, } \bar{\rho}\propto a^{-3}, \quad \nabla_{\boldsymbol{x}}\cdot\boldsymbol{x} = 3)$$

$$\left[\frac{\partial}{\partial t} - H\boldsymbol{x}\cdot\boldsymbol{\nabla}\right][\bar{\rho}\delta] + \frac{1}{a}\boldsymbol{\nabla}\cdot[\bar{\rho}Ha\boldsymbol{x}\delta + \bar{\rho}\boldsymbol{v}] = 0 \qquad \text{(1-st order)}$$

or $\left[\frac{\partial\bar{\rho}}{\partial t} + 3H\bar{\rho}\right]\delta + \bar{\rho}\frac{\partial\delta}{\partial t} + \frac{\bar{\rho}}{a}\nabla\cdot\boldsymbol{v} = 0$, so

$$\dot{\delta} = -\frac{1}{a}\nabla\cdot\boldsymbol{v}$$

Euler equation:

$$\dot{\boldsymbol{v}} + H\boldsymbol{v} = -\frac{1}{a\bar{\rho}}\boldsymbol{\nabla}\delta P - \frac{1}{a}\boldsymbol{\nabla}\delta\Phi.$$

In the absence of pressure and gravitational perturbations, this equation simply says that $\boldsymbol{v}\propto a^{-1}$, which is something we already discovered in Chapter 1.

Poisson equation:

$$\nabla^2\delta\Phi = 4\pi Ga^2\bar{\rho}\delta$$

Totally we have:

$$\ddot{\delta} + 2H\dot{\delta} - \frac{c_s^2}{a^2}\nabla^2\delta = 4\pi G\bar{\rho}\delta$$

$2H\dot{\delta}$ - friction term, which wasn't in static space without gravity ($(\partial_t^2 - c_s^2\nabla_r^2)\,\delta\rho = 4\pi G\bar{\rho}\delta\rho,$). Below the Jeans' length $\lambda_{\mathrm{J}}(t) = \frac{2\pi}{k_{\mathrm{J}}} = c_s(t)\sqrt{\frac{\pi}{G\bar{\rho}(t)}}$ fluctuations oscillate with decreasing amplitude. Above the Jeans' length, the fluctuations experience power-law growth, rather than the exponential growth we found for static space.

Perturbations from relativistic physics

Relativistic perturbed metric variables and gauges

$$A, B, h_{ij}: \qquad \mathrm{d}s^2 = a^2(\tau)\left[(1+2A)\mathrm{d}\tau^2 - 2B_i\,\mathrm{d}x^i\,\mathrm{d}\tau - (\delta_{ij}+h_{ij})\,\mathrm{d}x^i\,\mathrm{d}x^j\right],$$

$$C, E, E_j, E_{ij}: \qquad h_{ij} = \underbrace{2C\delta_{ij} + 2\partial_{\langle i}\partial_{j\rangle}E}_{\text{scalar}} + \underbrace{2\partial_{(i}\hat{E}_{j)}}_{\text{vector}} + \underbrace{2\hat{E}_{ij}}_{\text{tensor}},$$

$\mathcal{H} := a'/a$ - the Hubble parameter in conformal time.
Gauge-Invariant perturbations - the Bardeen variables:

$$\begin{aligned}
&\Psi := A + \mathcal{H}(B - E') + (B - E')',\\
&\hat{\Phi}_i := \hat{E}_i' - \hat{B}_i,\\
&\hat{E}_{ij},\\
&\Phi := -C - \mathcal{H}(B - E') + \frac{1}{3}\nabla^2 E.
\end{aligned}$$

Because by

$$X^\mu \mapsto \tilde{X}^\mu \equiv X^\mu + \xi^\mu(\tau, x), \quad \text{where} \quad \xi^0 \equiv T, \quad \xi^i \equiv L^i = \partial^i L + \hat{L}^i$$

$$\begin{aligned}
&A \mapsto A - T' - \mathcal{H}T,\\
&B \mapsto B + T - L', \qquad \hat{B}_i \mapsto \hat{B}_i - \hat{L}_i',\\
&C \mapsto C - \mathcal{H}T - \frac{1}{3}\nabla^2 L,\\
&E \mapsto E - L, \qquad \hat{E}_i \mapsto \hat{E}_i - \hat{L}_i, \qquad \hat{E}_{ij} \mapsto \hat{E}_{ij}
\end{aligned}$$

(something about gauge, прямо лень смотреть это, что-то про калибровки... отдельно нужно дочитывать)
(!!! скоро все другое допишу!!!)
10 degrees of freedom of the metric is decomposed into $4 + 4 + 2$ SVT degrees of freedom:
- scalars: A, B, C, E - vectors: $\hat{B}_i, \hat{E}_i$ - tensors: $\hat{E}_{ij}$

$\mathcal{H} := a'/a$ - the Hubble parameter in conformal time.

Newtonian gauge: $B = E = 0$

$$g_{\mu\nu} = a^2 \begin{pmatrix} 1+2\Psi & 0 \\ 0 & -(1-2\Phi)\delta_{ij} \end{pmatrix}$$

Spatially-flat gauge: $C = E = 0$
Clever people say that Π_{ij} is not important, so from now on $\Pi_{ij} = 0$, this enforces $\Phi = \Psi$.

Relativistic Matter perturbation, adiabatic and isocurvature terms

(допишу скоро очень из Бумана тоже формулы про это, с них я и начал)
(у Лесгоргуеса тоже хорошая теория про это, ее тоже тут напишу!!!)
Modes of observational interest were superhorizon when inflation ended, and inflation provides initial conditions for their subsequent evolution. Superhorizon modes undergo very simple evolution, unaffected by local interactions. In single-field inflation, the time-shifting $\delta\eta(\mathbf{x}) \sim \frac{\delta\phi(\mathbf{x})}{\dot{\phi}^{(0)}}$ is the only degree of freedom that determines superhorizon modes. Initial conditions of this type are called adiabatic.

If the universe contains a number of uncoupled perfect fluids (labeled by an index x) with sound speeds $c_x^2 = \delta P_x/\delta\rho_x$, we find for superhorizon modes

$$\rho_x(\eta,\mathbf{x}) = \rho_x^{(0)}(\eta+\delta\eta(\mathbf{x})) = \rho_x^{(0)}(\eta) + \dot\rho_x^{(0)}(\eta)\delta\eta(\mathbf{x}),$$
$$P_x(\eta,\mathbf{x}) = P_x^{(0)}(\eta+\delta\eta(\mathbf{x})) = P_x^{(0)}(\eta) + \dot P_x^{(0)}(\eta)\delta\eta(\mathbf{x}).$$

$$\dot\rho_x^{(0)} = -3\frac{\dot a}{a}\left(\rho_x^{(0)} + P_x^{(0)}\right), \qquad \text{(energy conserv-n.)}$$
$$\frac{\delta\rho_x}{\rho_x^{(0)}+P_x^{(0)}} = -3\frac{\dot a}{a}\frac{\delta\rho_x}{\dot\rho_x^{(0)}} = -3\frac{\dot a}{a}\delta\eta(\mathbf{x}),$$

Using the equations of state for matter and radiation, this implies (still for superhorizon modes)

$$\delta = \delta_b = \frac{3}{4}\delta_\nu = \frac{3}{4}\delta_\gamma.$$

(!!! выведу, это не очевидно вообще!)

Einstein's equations can be used to show that, for superhorizon modes during radiation domination,

$$\Psi = -\frac{1}{2}\delta_\gamma = \text{ constant}.$$

(??? how???)

In general, superhorizon metric and density perturbations are constant except when the equation of state changes.

Perturbed relativistic continuity, Euler and Einstein equations

From the $\nu = 0$ of $\nabla_\mu T^\mu{}_\nu = 0$ we get relativistic continuity eq.:

$$\bar\rho' = -3\mathcal{H}(\bar\rho+\bar P) \qquad \text{0-th or-er cont eq}$$
$$\delta\rho' = -3\mathcal{H}(\delta\rho+\delta P) + 3\Phi'(\bar\rho+\bar P) - \nabla\cdot\boldsymbol{q}. \qquad \text{1-st or-er cont eq}$$

$$\delta \equiv \frac{\delta\rho}{\bar\rho} \quad \text{and} \quad \boldsymbol{v} = \frac{\boldsymbol{q}}{\bar\rho+\bar P}$$
$$\delta' + \left(1+\frac{\bar P}{\bar\rho}\right)(\boldsymbol{\nabla}\cdot\boldsymbol{v} - 3\Phi') + 3\mathcal{H}\left(\frac{\delta P}{\delta\rho} - \frac{\bar P}{\bar\rho}\right)\delta = 0.$$

In the limit $P \ll \rho$, we have corrected Newtonian continuity equation in conformal time $\delta' + \boldsymbol{\nabla}\cdot\boldsymbol{v} - 3\Phi' = 0$.

From $\nabla_\mu T^\mu{}_\nu = 0$ the $\nu = i$ component we get relativistic Euler eq.:

$$\boldsymbol{v}' + \mathcal{H}\boldsymbol{v} - 3\mathcal{H}\frac{\bar P'}{\bar\rho'}\boldsymbol{v} = -\frac{\boldsymbol{\nabla}\delta P}{\bar\rho+\bar P} - \nabla\Psi.$$

Perturbed relativistic Einstein equations

The trace-free part of the ij equation, $G_{ij} = 8\pi G T_{ij}$:

$$\partial_{\langle i}\partial_{j\rangle}(\Phi-\Psi) = 0.$$

(if we have drop anisotropic stress at r.h.s.)

$G_{00} = 8\pi G T_{00}$:

$$\nabla^2\Phi = 4\pi G a^2\bar\rho\delta + 3\mathcal{H}\left(\Phi' + \mathcal{H}\Phi\right)$$

$G_{0i} = 8\pi G T_{0i}$:

$$\Phi' + \mathcal{H}\Phi = -4\pi G a^2(\bar\rho+\bar P)v$$

00-th equation:

$$\nabla^2\Phi = 4\pi G a^2 \bar{\rho}\Delta,$$
$$\bar{\rho}\Delta := \bar{\rho}\delta - 3\mathcal{H}(\bar{\rho}+\bar{P})v$$

trace-part of the ij equation, i.e. $G^i_i = 8\pi G T^i{}_i$.

$$\Phi'' + 3\mathcal{H}\Phi' + \left(2\mathcal{H}' + \mathcal{H}^2\right)\Phi = 4\pi G a^2 \delta P$$

Because

$$\Gamma^0_{00} = \mathcal{H} + \Psi', \qquad \Gamma^0_{0i} = \partial_i\Psi, \qquad \Gamma^i_{00} = \delta^{ij}\partial_j\Psi,$$
$$\Gamma^i_{j0} = \mathcal{H}\delta^i_j - \Phi'\delta^i_j, \qquad \Gamma^i_{jk} = -2\delta^i_{(j}\partial_{k)}\Phi + \delta_{jk}\delta^{il}\partial_l\Phi. \qquad \Gamma^0_{ij} = \mathcal{H}\delta_{ij} - \left[\Phi' + 2\mathcal{H}(\Phi+\Psi)\right]\delta_{ij}$$

$$\begin{aligned}
R_{00} =& -3\mathcal{H}' + \nabla^2\Psi + 3\mathcal{H}\left(\Phi' + \Psi'\right) + 3\Phi'' \\
R_{0i} =& 2\partial_i\Phi' + 2\mathcal{H}\partial_i\Psi \\
R_{ij} =& \mathcal{H}' + 2\mathcal{H}^2 - \Phi'' + \nabla^2\Phi - 2\left(\mathcal{H}' + 2\mathcal{H}^2\right)(\Phi+\Psi) - \mathcal{H}\Psi' - 5\mathcal{H}\Phi'\big]\,\delta_{ij} + \partial_i\partial_j(\Phi - \Psi).
\end{aligned}$$

$$a^2 R = -6\left(\mathcal{H}' + \mathcal{H}^2\right) + 2\nabla^2\Psi - 4\nabla^2\Phi + 12\left(\mathcal{H}' + \mathcal{H}^2\right)\Psi + 6\Phi'' + 6\mathcal{H}\left(\Psi' + 3\Phi'\right)$$

$$\begin{aligned}
G_{00} &= 3\mathcal{H}^2 + 2\nabla^2\Phi - 6\mathcal{H}\Phi' \\
G_{0i} &= 2\partial_i\left(\Phi' + \mathcal{H}\Psi\right). \\
G_{ij} &= -\left(2\mathcal{H}' + \mathcal{H}^2\right)\delta_{ij} + \big[\nabla^2(\Psi - \Phi) + 2\Phi'' + 2\left(2\mathcal{H}' + \mathcal{H}^2\right)(\Phi + \Psi) + 2\mathcal{H}\Psi' + 4\mathcal{H}\Phi'\big]\delta_{ij} + \partial_i\partial_j(\Phi - \Psi)
\end{aligned}$$

Also

$$T_{0i} = g_{0\mu}T^\mu_i = g_{00}T^0_i = \bar{g}_{00}T^0_i = -a^2 q_i$$

$$T^i_i = -3(\bar{P} + \delta P)$$

Summary of Einstein eq. by Dodelson

In this definitions of scalar metric perturbations

$$g_{00}(\vec{x},t) \equiv -1 - 2\Psi(\vec{x},t) \qquad g_{ij}(\vec{x},t) \equiv a^2\delta_{ij}(1 + 2\Phi(\vec{x},t)). \qquad g_{0i}(\vec{x},t) \equiv 0$$

In Fourier space $G_{\mu\nu} = 8\pi G T_{\mu\nu}$ relate perturbations in the metric to perturbations in the matter and radiation by:

$$k^2\Phi + 3\frac{\dot{a}}{a}\left(\dot{\Phi} - \Psi\frac{\dot{a}}{a}\right) = 4\pi G a^2\left[\rho_{\rm m}\delta_m + 4\rho_r\Theta_{r,0}\right]$$
$$k^2(\Phi + \Psi) = -32\pi G a^2\rho_r\Theta_{r,2}$$

"m" - all matter (baryons and dark matter), "r" - all radiation (neutrinos and photons):

$$\rho_{\rm m}\delta_m := \rho_{\rm dm}\delta + \rho_b\delta_{\rm b}; \qquad \rho_r\Theta_{r,0} := \rho_\gamma\Theta_0 + \rho_\nu\mathcal{N}_0$$
$$\rho_{\rm m}v_m := \rho_{\rm dm}v + \rho_b v_{\rm b}; \qquad \rho_r\Theta_{r,1} := \rho_\gamma\Theta_1 + \rho_\nu\mathcal{N}_1$$

Some of the other components of Einstein's equation are redundant, but sometimes one is more useful than another.

For example, one combination of these equations leads to an algebraic equation for the potential,

$$k^2\Phi = 4\pi G a^2\left[\rho_{\rm m}\delta_m + 4\rho_r\Theta_{r,0} + \frac{3aH}{k}\left(i\rho_{\rm m}v_m + 4\rho_r\Theta_{r,1}\right)\right]$$

Other components of Einstein's equation contain information about vector and tensor perturbations. Scalar, vector, and tensor perturbations are decoupled: each evolves independently of the others.

Tensor perturbations are characterized by h_+ and $h_\cdot$; each of these evolves independently and satisfies

$$\ddot{h}_\alpha + 2\frac{\dot{a}}{a}\dot{h}_\alpha + k^2 h_\alpha = 0 \qquad \alpha := +, \times$$

In an expanding universe, this equation tells that amplitude of a gravitational wave falls off once the mode enters the horizon.

On Conserved Curvature Perturbation and its conservation

$$\gamma_{ij} \equiv a^2\left[(1+2C)\delta_{ij} + 2E_{ij}\right] \qquad E_{ij} \equiv \partial_{\langle i}\partial_{j\rangle}E$$

$$\text{curvature perturbation} = C - \frac{1}{3}\nabla^2 E.$$

$${}^{(3)}\Gamma^i_{jk},\ R_{(3)},\ \partial_i\partial_j E^{ij} == \frac{2}{3}\nabla^4 E \Rightarrow \qquad a^2 R_{(3)} = -4\nabla^2\left(C - \frac{1}{3}\nabla^2 E\right).$$

Gauge-invariant comoving curvature perturbation:

$$\mathcal{R} := C - \frac{1}{3}\nabla^2 E + \mathcal{H}(B+v) \quad \text{comoving gauge } \left(B_i = 0 = q^i\right)$$

Because (!!?! see exercises??!??)

$$\mathcal{R} = -\Phi + \mathcal{H}v \qquad (\text{Newtonian gauge } B = E = 0, C \equiv -\Phi)$$

we put v from $0i$ Einstein eq. $\Phi' + \mathcal{H}\Phi = -4\pi G a^2(\bar{\rho} + \bar{P})v$:

$$\mathcal{R} = -\Phi - \frac{\mathcal{H}\left(\Phi' + \mathcal{H}\Phi\right)}{4\pi G a^2(\bar{\rho} + \bar{P})},$$

from it by $\frac{d}{d\eta}$ and not trivial transformations and input of other equations:

$$-4\pi G a^2(\bar{\rho} + \bar{P})\mathcal{R}' = \mathcal{H}\frac{\bar{P}'}{\bar{\rho}'}\nabla^2\Phi + 4\pi G a^2 \mathcal{H}\delta P_{\text{nad}}$$

$$\delta P_{\text{nad}} := \delta P - \frac{\bar{P}'}{\bar{\rho}'}\delta\rho. \quad \text{(non-adiabat pressure perturb-n)}$$

δP_{nad} vanishes, if $P = P(\rho)$ (barotropic liquid). In that case, $\mathcal{H}\frac{\bar{P}'}{\bar{\rho}'}\nabla^2\Phi \sim \mathcal{H}k^2\Phi \sim \mathcal{H}k^2\mathcal{R}$,

$$\frac{d\ln\mathcal{R}}{d\ln a} \sim \left(\frac{k}{\mathcal{H}}\right)^2.$$

Thus $\mathcal{R}$ doesn't evolve on super-Hubble scales, $k \ll \mathcal{H}$ and we use it for connection perturbations in different times.

Setup for Scalar Perturbations from inflation from energy-momenta tensor by Dodelson

Scalar Field Perturbations around a Smooth Background

$$\phi(\vec{x},t) = \phi^{(0)}(t) + \delta\phi(\vec{x},t),$$

$$T^{\mu}_{\nu;\mu} = \frac{\partial T^{\mu}{}_{\nu}}{\partial x^{\mu}} + \Gamma^{\mu}{}_{\alpha\mu}T^{\alpha}{}_{\nu} - \Gamma^{\alpha}_{\nu\mu}T^{\mu}_{\alpha} = 0.$$

$$0 = \frac{\partial \delta T^{0}_{0}}{\partial t} + ik_i \delta T^{i}{}_{0} + 3H\delta T^{0}{}_{0} - H\delta T^{i}{}_{i}.$$

$$\delta T^{i}{}_{0} = \frac{ik_i}{a^3}\dot{\phi}^{(0)}\delta\phi.$$

$$T^{0}{}_{0} = \frac{-1}{2}\left(\phi^{(0)}_{,0} + \delta\phi_{.0}\right)^2 - \frac{1}{2a^2}\delta\phi_{.i}\delta\phi_{.i} - V\left(\phi^{(0)} + \delta\phi\right).$$

$$\delta T^{0}_{0} = \frac{-\dot{\phi}^{(0)}\dot{\delta}\phi}{a^2} - V'\delta\phi.$$

$$\delta T^{i}{}_{j} = \delta_{ij}\left(\frac{\dot{\phi}^{(0)}\dot{\delta}\phi}{a^2} - V'\delta\phi\right).$$

The conservation equation:

$$\left(\frac{1}{a}\frac{\partial}{\partial\eta} + 3H\right)\left(\frac{-\dot{\phi}^{(0)}\dot{\delta}\phi}{a^2} - V'\delta\phi\right) - \frac{k^2}{a^3}\dot{\phi}^{(0)}\delta\phi - 3H\left(\frac{\dot{\phi}^{(0)}\dot{\delta}\phi}{a^2} - V'\delta\phi\right) = 0.$$

Carrying out the time derivatives (the only subtle one is $\partial V'/\partial\eta = V''\dot{\phi}^{(0)}$), multiplying by a^3, and collecting terms leads to

$$-\dot{\phi}^{(0)}\ddot{\delta}\phi + \dot{\delta}\phi\left(-\ddot{\phi}^{(0)} - 4aH\dot{\phi}^{(0)} - a^2V'\right) + \delta\phi\left(-a^2V''\dot{\phi}^{(0)} - k^2\dot{\phi}^{(0)}\right) = 0$$

$$\ddot{\delta\phi} + 2aH\dot{\delta}\phi + k^2\delta\phi = 0.$$

ations to $\delta\phi$ is identical to Eq. is we can trivially copy our rest e power spectrum of fluctuatio

$$P_{\delta\phi} = \frac{H^2}{2k^3}.$$

Super-Horizon Perturbations

$$\frac{\partial \delta T^{0}{}_{0}}{\partial t} + ik_i\delta T^{i}{}_{0} + 3H\delta T^{0}{}_{0} - H\delta T^{i}{}_{i} = -3(\mathcal{P}+\rho)\frac{\partial\Psi}{\partial t}$$

and ρ are the zero-order pressure and energy density.

$$\Psi \ll \frac{\delta T^{0}{}_{0}}{\mathcal{P}+\rho}.$$

$$k^2\Psi + 3aH(\dot{\Psi} + aH\Psi) = 4\pi G a^2 \delta T^{0}{}_{0}.$$

$$\Psi \sim G\frac{\delta T^0{}_0}{H^2} \sim \frac{\delta T^0{}_0}{\rho} = \frac{\mathcal{P}+\rho}{\rho}\left(\frac{\delta T^0{}_0}{\mathcal{P}+\rho}\right).$$

One way to deal with the coupling between the metric perturbations and those to the energy density is to define

$$\zeta \equiv -\frac{ik_i\delta T^0{}_i H}{k^2(\rho+\mathcal{P})} - \Psi.$$

For sub-horizon modes and those which have just left the horizon, Ψ is negligible; $P+\rho = \left(\dot{\phi}^{(0)}/a\right)^2$ from Eqs. (6.26) and (6.27); and Eq. (6.64) fixes the numerator of the first term in ζ. We are left with

$$\zeta = -aH\delta\phi/\dot{\phi}^{(0)}$$

around the time of horizon crossing. After inflation ends, $ik_i\delta T^0{}_i = 4ak\rho_r\Theta_1$, proportional to the dipole of the radiation. Since the pressure of radiation is equal to a third of the energy density,

$$\begin{aligned}\zeta &= -\frac{3aH\Theta_1}{k} - \Psi \\ &= -\frac{3}{2}\Psi \quad \text{(post inflation)}.\end{aligned}$$

The second equality follows from the initial conditions relating the dipole to the potential (Eq. (6.16)).

The variable ζ is so important because it is conserved when the perturbation moves outside the horizon (Figure 6.8). We will show that ζ is conserved shortly, but first let's appreciate the importance of this conservation. Since we know that, after inflation, $\zeta = -3\Psi/2$, we can immediately relate Ψ coming out of inflation to the $\delta\phi$ at horizon crossing,

$$\Psi|_{\text{post inflation}} = \frac{2}{3}aH\frac{\delta\phi}{\dot{\phi}^{(0)}}\bigg|_{\text{horizon crossing}}.$$

$$P_\Psi = \frac{4}{9}\left(\frac{aH}{\dot{\phi}^{(0)}}\right)^2 P_{\delta\phi}\bigg|_{aH=k}$$

It remains to prove that ζ is conserved on super-horizon scales. To see this, let's turn to the conservation equation, Eq. (6.73). On large scales, $k_i\delta T^i{}_0$ is proportional to k^2 and so can be ignored, leaving

$$\frac{\partial\delta T^0{}_0}{\partial t} + 3H\delta T^0{}_0 - H\delta T^i{}_i = -3(\mathcal{P}+\rho)\frac{\partial\Psi}{\partial t}.$$

On large scales, you will show (Exercise 13) that the energy-momentum tensor satisfies

$$\frac{ik_i\delta T^0{}_i H}{k^2} = \frac{\delta T^0{}_0}{3}.$$

Therefore, on large scales

$$\zeta = -\Psi - \frac{1}{3}\frac{\delta T^0{}_0}{\rho+\mathcal{P}}.$$

Eliminating Ψ in favor of ζ in the conservation equation leads to

$$\frac{\partial\delta T^0{}_0}{\partial t} + 3H\delta T^0{}_0 - H\delta T^i{}_i = 3(\mathcal{P}+\rho)\frac{\partial\zeta}{\partial t} + (\rho+\mathcal{P})\frac{\partial}{\partial t}\left[\frac{\delta T^0{}_0}{\rho+\mathcal{P}}\right].$$

The partial derivative on the right acting on $\delta T^0{}_0$ cancels the first term on the left, leaving

$$\delta T^0{}_0 \left[3H + \frac{1}{\rho+\mathcal{P}}\left(\frac{d\rho}{dt} + \frac{d\mathcal{P}}{dt}\right)\right] - H\delta T^i{}_i = 3(\mathcal{P}+\rho)\frac{\partial\zeta}{\partial t}.$$

Recall from Eq. (2.55) that $d\rho/dt = -3H(\rho+\mathcal{P})$, so the first term in brackets cancels the second, and

$$\frac{\partial\zeta}{\partial t} = \frac{-1}{3(\rho+\mathcal{P})^2}\left[H(\rho+\mathcal{P})\delta T^i{}_i - \delta T^0{}_0\frac{d\mathcal{P}}{dt}\right]$$

Spatially Flat Slicing

$$ds^2 = -(1+2A)dt^2 - 2aB_{,i}dx^i dt + \delta_{ij}a^2 dx^i dx^j$$

$$v = ikB - \frac{ik\dot{\phi}^{(0)}\delta\phi}{(\rho+\mathcal{P})a^2} \quad \text{(spatially flat slicing)}$$

where I have evaluated $\delta T^0{}_i$ with Eq. (6.64). Thus, we can create a gauge-invariant variable proportional to $\delta\phi$ in a spatially flat slicing if we subtract off the kB term. Bardeen's Φ_H (Eq. (5.76)) is simply equal to aHB, so the combination

$$\zeta \equiv -\Phi_H - \frac{iaH}{k}v$$

is gauge invariant and in spatially flat slicing is equal to

$$\zeta = -\frac{aH}{\dot{\phi}^{(0)}}\delta\phi \quad \text{(spatially flat slicing).}$$

1.3.3 Initial conditions from inflaton fluctuations

Classical inhomogeneous inflaton field Oscillators and Mukhanov-Sasaki equation

$$\phi(\tau,\boldsymbol{x}) \equiv \bar{\phi}(\tau) + \frac{f(\tau,\boldsymbol{x})}{a(\tau)}$$

$$S = \int \mathrm{d}\tau\mathrm{d}^3x\sqrt{-g}\left[\tfrac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right] \overset{\text{FRLW}}{=} \int \mathrm{d}\tau\mathrm{d}^3x\left[\tfrac{1}{2}a^2\left((\phi')^2 - (\nabla\phi)^2\right) - a^4V(\phi)\right] \qquad \text{(real t.)}$$

$$S^{(1)} = -\int \mathrm{d}\tau\mathrm{d}^3x a\left[\bar{\phi}'' + 2\mathcal{H}\bar{\phi}' + a^2V_{,\phi}\right]f \qquad \Rightarrow \quad \bar{\phi}'' + 2\mathcal{H}\bar{\phi}' + a^2V_{,\phi} = 0 \qquad \text{(real t.)}$$

$$S^{(2)} \overset{\text{int b p}}{=} \frac{1}{2}\int \mathrm{d}\tau\mathrm{d}^3x\left[(f')^2 - (\nabla f)^2 + \left(\frac{a''}{a} - a^2V_{,\phi\phi}\right)f^2\right]. \qquad \text{(real t.)}$$

Also we neglect the term $a^2V_{,\phi\phi}$, because during slow-roll inflation $\frac{V_{,\phi\phi}}{H^2} \approx \frac{3M_{\rm pl}^2V_{,\phi\phi}}{V} = 3\eta_{\rm v} \ll 1$. Also, since $a' = a^2H$, with $H \approx$ const., we also have $\frac{a''}{a} \approx 2a'H = 2a^2H^2 \gg a^2V_{,\phi\phi}$. For this action the Euler-Lagrange equation gives the Mukhanov-Sasaki equation $f'' - \nabla^2 f - \frac{a''}{a}f = 0$, or for each mode:

$$f_k'' + \left(k^2 - \frac{a''}{a}\right)f_k = 0$$

On subhorizon scales, $k^2 \gg a''/a \approx 2\mathcal{H}^2$, the Mukhanov-Sasaki equation reduces to

$$f_k'' + k^2 f_k \approx 0.$$

We see that each Fourier mode satisfies the equation of motion of a simple harmonic oscillator, with frequency $\omega_k = k$. Quantum zero-point fluctuations of these oscillators provide the origin of structure in the universe.

(?? может, что-то еще допишу, просто вывели уравнения и все.)

Quantum Fluctuations in de Sitter Space

Canonical Quantisation

$$\left[\hat{f}(\tau,\boldsymbol{x}),\hat{\pi}\left(\tau,\boldsymbol{x}'\right)\right]=i\delta\left(\boldsymbol{x}-\boldsymbol{x}'\right)\qquad\left[\hat{f}_k(\tau),\hat{\pi}_{\boldsymbol{k}'}(\tau)\right]=i\delta\left(\boldsymbol{k}+\boldsymbol{k}'\right)$$

$$\hat{f}_k(\tau)=f_k(\tau)\hat{a}_{\boldsymbol{k}}+f_k^*(\tau)a_{\boldsymbol{k}}^\dagger,$$

$$f_k''+\omega_k^2(\tau)f_k=0,\qquad\omega_k^2(\tau):=k^2-\frac{a''}{a}$$

$$\left[\hat{a}_{\boldsymbol{k}},\hat{a}_{\boldsymbol{k}'}^\dagger\right]=\delta\left(\boldsymbol{k}+\boldsymbol{k}'\right)$$

$$\hat{a}_k|0\rangle=0,$$

$$|m_{\boldsymbol{k}_1},n_{\boldsymbol{k}_2},\cdots\rangle=\frac{1}{\sqrt{m!n!\cdots}}\left[\left(a_{\boldsymbol{k}_1}^\dagger\right)^m\left(a_{\boldsymbol{k}_2}^\dagger\right)^n\cdots\right]|0\rangle.$$

Choice of Vacuum

At sufficiently early times (large negative conformal time τ) all modes of cosmological interest were deep inside the horizon, $k/\mathcal{H}\sim|k\tau|\gg1$,

$$\omega_k^2=k^2-\frac{a''}{a}\approx k^2-\frac{2}{\tau^2}\overset{\tau\to-\infty}{\longrightarrow}k^2,$$

and the Mukhanov-Sasaki equation reduces to

$$f_k''+k^2f_k\approx0$$

$$f_k\propto e^{\pm ik\tau}$$

$$\lim_{\tau\to-\infty}f_k(\tau)=\frac{1}{\sqrt{2k}}e^{-ik\tau}\qquad\text{(init. cond, unique physical Bunch-Davies vacuum.)}$$

(???? main words, what is it????)

$$f_k(\tau)=\frac{e^{-ik\tau}}{\sqrt{2k}}\left(1-\frac{i}{k\tau}\right).$$

Zero-Point Fluctuations

$$f_k(\tau)=\frac{e^{-ik\tau}}{\sqrt{2k}}\left(1-\frac{i}{k\tau}\right)=\left\langle|\hat{f}|^2\right\rangle:=\left\langle0\left|\hat{f}^\dagger(\tau,\boldsymbol{0})\hat{f}(\tau,\boldsymbol{0})\right|0\right\rangle$$

$$\Delta_f^2(k,\tau):=\frac{k^3}{2\pi^2}\left|f_k(\tau)\right|^2$$

$$\Delta_{\delta\phi}^2(k)\approx\left.\left(\frac{H}{2\pi}\right)^2\right|_{k=aH}$$

Quantum-to-Classical Transition (??????)

(??? how does it work, in 2 line???)

$$\hat{f}(\tau,\boldsymbol{x}) = \int \frac{\mathrm{d}^3k}{(2\pi)^{3/2}} \left[f_k(\tau)\hat{a}_{\boldsymbol{k}} + f_k^*(\tau)a_{\boldsymbol{k}}^\dagger \right] e^{i\boldsymbol{k}\cdot\boldsymbol{x}}$$
$$\hat{\pi}(\tau,\boldsymbol{x}) = \int \frac{\mathrm{d}^3k}{(2\pi)^{3/2}} \left[f_k'(\tau)\hat{a}_{\boldsymbol{k}} + f_k^{*\prime}(\tau)a_{\boldsymbol{k}}^\dagger \right] e^{i\boldsymbol{k}\cdot\boldsymbol{x}}$$

In the superhorizon limit, $k\tau \to 0$, we have

$$f_k(\tau) \approx -\frac{1}{\sqrt{2}k^{3/2}}\frac{i}{\tau} \qquad \text{and} \qquad f_k'(\tau) \approx \frac{1}{\sqrt{2}k^{3/2}}\frac{i}{\tau^2}$$

$$\hat{f}(\tau,\boldsymbol{x}) = -\frac{i}{\sqrt{2}\tau}\int \frac{\mathrm{d}^3k}{(2\pi)^{3/2}}\frac{1}{k^{3/2}} \left[\hat{a}_{\boldsymbol{k}} - a_{\boldsymbol{k}}^\dagger \right] e^{i\boldsymbol{k}\cdot\boldsymbol{x}},$$
$$\hat{\pi}(\tau,\boldsymbol{x}) = \frac{i}{\sqrt{2}\tau^2}\int \frac{\mathrm{d}^3k}{(2\pi)^{3/2}}\frac{1}{k^{3/2}} \left[\hat{a}_{\boldsymbol{k}} - a_{\boldsymbol{k}}^\dagger \right] e^{i\boldsymbol{k}\cdot\boldsymbol{x}} = -\frac{1}{\tau}\hat{f}(\tau,\boldsymbol{x}).$$

Now,

$$\begin{aligned}\langle 0|\delta\phi(\eta,\mathbf{x})\delta\phi(\eta,\mathbf{y})|0\rangle &= \frac{1}{a^2(\eta)}\int\frac{d^3k}{(2\pi)^{3/2}}\int\frac{d^3k'}{(2\pi)^{3/2}}\chi_{\mathbf{k}}(\eta)\chi_{\mathbf{k}'}^*(\eta)e^{i\mathbf{k}\cdot\mathbf{x}}e^{-i\mathbf{k}'\cdot\mathbf{y}}\delta\left(\mathbf{k}-\mathbf{k}'\right)\\ &= \int\frac{d^3k}{(2\pi)^3}e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}\frac{|\chi_{\mathbf{k}}(\eta)|^2}{a^2(\eta)}.\end{aligned}$$

Comparing with $\xi_f(\mathbf{x},\mathbf{y}) = \int \frac{dk}{k}\frac{\sin(kr)}{kr}\frac{k^3\sigma_k^2}{2\pi^2}$, and using $\langle f_{\mathbf{k}}f_{\mathbf{k}'}\rangle \equiv (2\pi)^3 P_f(k)\delta\left(\mathbf{k}+\mathbf{k}'\right)$, $\chi_{\mathbf{k}} = \frac{1}{\sqrt{2k}}\left(1-\frac{i}{k\eta}\right)e^{-ik\eta}$ and $a(\eta) = -\frac{1}{H\eta}$, we get

$$P_{\delta\phi}(k) = \sigma_k^2 = \frac{1}{2k}\left(H^2\eta^2 + \frac{H^2}{k^2}\right) \approx \frac{H^2}{2k^3},$$

where in the last step we ignored the first term. To justify this, note that given a distance $r \equiv |\mathbf{x}-\mathbf{y}|$, the integral (48) will mainly get contributions from $k \leq 1/r$ (the high k contributions tend to cancel each other because of the oscillatory integrand).

From scalar field perturbations to density perturbations

When connecting the theory of inflation to observations, we will be interested in metric perturbations long after inflation has ended, for instance at the time of recombination, when the CMB photons were emitted. At the end of inflation, the energy stored in the potential of the inflaton is converted into radiation and matter (reheating). "Initial conditions" produced by inflation can be reliably propagated to later times by identifying a quantity that is conserved for superhorizon modes, independently of the details of reheating.

For a mode deep inside the horizon during SR inflation:

$$-\frac{k^2}{a^2}\Psi = 4\pi G\delta\rho, \qquad \text{(Poiss. eq for gr pot } \Psi\text{)}$$

and to estimate orders of magnitude we can continue to use this equation up to the time of horizon crossing, $k^2 \sim a^2H^2$. For a mode crossing the horizon:

$$\Psi \sim G\frac{\delta\rho}{H^2} \sim \frac{\delta\rho}{\rho},$$

and before horizon crossing Ψ was even smaller.

For a slowly rolling scalar field, the energy density is dominated by the potential, $\rho = V(\phi)$ (????? this is an important formula, why it is so???), this how the scalar field fluctuations give rise to density perturbations

$$\delta\rho = V'\left(\phi^{(0)}\right)\delta\phi, \qquad \frac{\delta\rho}{\rho} = \frac{V'\left(\phi^{(0)}\right)}{V\left(\phi^{(0)}\right)}\delta\phi \sim \sqrt{G\epsilon}\delta\phi,$$

which thus also have a scale invariant power spectrum.

It can be checked (?? how?) that the resulting metric perturbation Ψ can be ignored in the equation of motion of the scalar field, which justifies a posteriori our computation, where we considered a scalar field in an exponentially expanding homogeneous spacetime, ignoring the backreaction of the scalar on the metric.

Space-dependent fluctuations $\delta\phi$ of the inflaton cause different spatial regions to reach this critical value of the inflaton at different times, with time delay

$$\delta\eta(\mathbf{x}) \sim \frac{\delta\phi(\mathbf{x})}{\dot{\phi}^{(0)}}.$$

More longer inflation in some regions causes space to curve, this curvature will lead to perturbations in the gravitational potential Ψ after inflation has ended.

Общие важные вопросы про квантовые флуктуации в начале Вселенной

Не бред ли, что квантовые осцилляции были настолько большими, что так сильно действовали? (?????)

(тут ответ на этот важный вопрос. наверняка есть большое обсуждение этого)

1.3.4 Derivations of full perturbed Boltzmann and Einstain eq. in a nutshell (!??!)

(это пока не актуально, тут пока углубления Додельсона, и без них можно многое важное понять по упрощенным лекциям. потом тут будут указания про более глубокое решение этой же задачи.)

Method

$$\begin{aligned}
\dot{\Theta}_{r.0} + k\Theta_{r.1} &= -\dot{\Phi} \\
\dot{\Theta}_{r.1} - \frac{k}{3}\Theta_{r,0} &= \frac{-k}{3}\Phi \\
\dot{\delta} + ikv &= -3\dot{\Phi} \\
\dot{v} + \frac{\dot{a}}{a}v &= ik\Phi.
\end{aligned}$$

$$k^2\Phi + 3\frac{\dot{a}}{a}\left(\dot{\Phi} + \frac{\dot{a}}{a}\Phi\right) = 4\pi G a^2\left[\rho_{\mathrm{dm}}\delta + 4\rho_r\Theta_{r,0}\right].$$

$$k^2\Phi = 4\pi G a^2\left[\rho_{\mathrm{dm}}\delta + 4\rho_r\Theta_{r,0} + \frac{3aH}{k}\left(i\rho_{\mathrm{dm}}v + 4\rho_r\Theta_{r,1}\right)\right]$$

Idea of expansion in modes (!!!???)

(общее важное соображение, которое там потом везде и используется, допишу потом)

Boltzmann equations for fluctuations by Dodelson

(это же в доп. кинетике укажу!)

SUMMARY BOLTZMANN EQUATIONS by Dodelson

$$\Theta_0(\vec{x},t) \equiv \frac{1}{4\pi}\int d\Omega' \Theta\left(\hat{p}',\vec{x},t\right). \qquad \text{(monopole)}$$

$$\Theta_1 \equiv i\int_{-1}^{1}\frac{d\mu}{2}\mu\Theta(\mu) \qquad \text{(dipole)}$$

$$\Theta_l := \frac{1}{(-i)^l}\int_{-1}^{1}\frac{d\mu}{2}\mathcal{P}_l(\mu)\Theta(\mu)$$

Photon perturbations can be described either by $\Theta(k,\mu,\eta)$ or by a whole hierarchy of moments, $\Theta_l(k,\eta)$. Same is for neutrino distribution.

strength of the polarization is Θ_P.

$$\dot{\Theta}+ik\mu\Theta = -\dot{\Phi}-ik\mu\Psi-\dot{\tau}\left[\Theta_0-\Theta+\mu v_{\rm b}-\frac{1}{2}\mathcal{P}_2(\mu)\Pi\right] \qquad \text{(Boltz. eq. for photons)}$$

$$\Pi = \Theta_2+\Theta_{P2}+\Theta_{P0}$$

$$\dot{\Theta}_P+ik\mu\Theta_P = -\dot{\tau}\left[-\Theta_P+\frac{1}{2}\left(1-\mathcal{P}_2(\mu)\right)\Pi\right]$$

$$\dot{\delta}+ikv = -3\dot{\Phi}$$

$$\dot{v}+\frac{\dot{a}}{a}v = -ik\Psi$$

$$\dot{\delta}_{\rm b}+ikv_{\rm b} = -3\dot{\Phi}$$

$$\dot{v}_{\rm b}+\frac{\dot{a}}{a}v_{\rm b} = -ik\Psi+\frac{\dot{\tau}}{R}\left[v_{\rm b}+3i\Theta_1\right] \qquad \frac{1}{R} := \frac{4\rho_\gamma^{(0)}}{3\rho_b^{(0)}}$$

$$\dot{\mathcal{N}}+ik\mu\mathcal{N} = -\dot{\Phi}-ik\mu\Psi. \qquad (\mathcal{N}\text{ - neutrinos distr. pert., no scatt.})$$

(!! допишу лучше потом, в чем смысл каждого ур-я???)

THE BOLTZMANN EQUATION FOR THE HARMONIC OSCILLATOR

$$\frac{\partial f}{\partial t}+\frac{p}{m}\frac{\partial f}{\partial x}-kx\frac{\partial f}{\partial p}=0.$$

$$\frac{df(t,x,p)}{dt} = \frac{\partial f}{\partial t}+\frac{\partial f}{\partial x}\frac{dx}{dt}+\frac{\partial f}{\partial p}\frac{dp}{dt}.$$

$$E=\frac{p^2}{2m}+\frac{1}{2}kx^2. \qquad \frac{dx}{dt}\equiv\frac{p}{m}. \qquad \frac{dp}{dt}=-kx$$

$$f(p,x) = f_{\rm EQ}(E)$$

$$\frac{p}{m}\frac{\partial f(E)}{\partial x}-kx\frac{\partial f(E)}{\partial p} = \frac{df}{dE}\left[\frac{p}{m}\frac{\partial E}{\partial x}-kx\frac{\partial E}{\partial p}\right]$$
$$= 0.$$

THE COLLISIONLESS BOLTZMANN EQUATION FOR PHOTONS

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x^i}\cdot\frac{dx^i}{dt} + \frac{\partial f}{\partial p}\frac{dp}{dt} + \frac{\partial f}{\partial \hat{p}^i}\cdot\frac{d\hat{p}^i}{dt}$$

$$P^i \equiv C\hat{p}^i.$$

$$\begin{aligned} p^2 &= g_{ij}\hat{p}^i\hat{p}^j C^2 \\ &= a^2(1+2\Phi)\delta_{ij}\hat{p}^i\hat{p}^j C^2 \\ &= a^2(1+2\Phi)C^2 \end{aligned}$$

$C = p(1-\Phi)/a$

$$P^i = p\hat{p}^i\frac{1-\Phi}{a}.$$

$$\frac{dx^i}{dt} = \frac{\hat{p}^i}{a}(1+\Psi-\Phi).$$

$$\frac{d}{dt}[p(1-\Psi)] = -\Gamma^0{}_{\alpha\beta}\frac{P^\alpha P^\beta}{p}(1+\Psi).$$

$$\frac{dp}{dt}(1-\Psi) = p\frac{d\Psi}{dt} - \Gamma^0{}_{\alpha\beta}\frac{P^\alpha P^\beta}{p}(1+\Psi).$$

multiply both sides by $(1+\Psi)$; drop all terms quadratic in Ψ; and reexpress the total time derivative of Ψ in terms of partial derivatives so that

$$\frac{dp}{dt} = p\left\{\frac{\partial\Psi}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}\right\} - \Gamma^0{}_{\alpha\beta}\frac{P^\alpha P^\beta}{p}(1+2\Psi).$$

(дальше еще какие-то преобразования, словами напишу потом)

$$\frac{1}{p}\frac{dp}{dt} = -H - \frac{\partial\Phi}{\partial t} - \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}.$$

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial f}{\partial x^i} - p\frac{\partial f}{\partial p}\left[H + \frac{\partial\Phi}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}\right] \qquad \text{(Boltzmann eq. for photons)}$$

$$f(\vec{x},p,\hat{p},t) = \left[\exp\left\{\frac{p}{T(t)[1+\Theta(\vec{x},\hat{p},t)]}\right\} - 1\right]^{-1}$$

$$\begin{aligned} f &\simeq \frac{1}{e^{p/T}-1} + \left(\frac{\partial}{\partial T}\left[\exp\left\{\frac{p}{T}\right\} - 1\right]^{-1}\right)T\Theta \\ &= f^{(0)} - p\frac{\partial f^{(0)}}{\partial p}\Theta. \qquad f^{(0)} \equiv \left[\exp\left\{\frac{p}{T}\right\} - 1\right]^{-1} \end{aligned}$$

Now,

$$\dot{\tilde{v}}_{\rm b} + \frac{\dot{a}}{a}\tilde{v}_{\rm b} + ik\tilde{\Psi} = \dot{\tau}\frac{4\rho_\gamma}{3\rho_b}\left[3i\tilde{\Theta}_1 + \tilde{v}_{\rm b}\right] \qquad \text{(conf time)}$$

$$\left.\frac{df}{dt}\right|_{\text{zero order}} = \frac{\partial f^{(0)}}{\partial t} - Hp\frac{\partial f^{(0)}}{\partial p} = 0$$

$$\left.\frac{df}{dt}\right|_{\text{first order}} = -p\frac{\partial f^{(0)}}{\partial p}\left[\frac{\partial\Theta}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial\Theta}{\partial x^i} + \frac{\partial\Phi}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}\right].$$

(with words I'll here explain a derivation in Dodelson)

COLLISION TERMS: COMPTON SCATTERING

$$C[f(\vec{p})] = \sum_{\vec{q},\vec{q}',\vec{p}'} \mid \text{Amplitude } \mid^2 \{f_e(\vec{q}')\, f(\vec{p}') - f_e(\vec{q}) f(\vec{p})\}$$

$$C[f(\vec{p})] = \frac{1}{p}\int \frac{d^3q}{(2\pi)^3 2E_e(q)} \int \frac{d^3q'}{(2\pi)^3 2E_e(q')} \int \frac{d^3p'}{(2\pi)^3 2E(p')} |\mathcal{M}|^2 (2\pi)^4 \times$$
$$\times \delta^3[\vec{p}+\vec{q}-\vec{p}'-\vec{q}']\, \delta[E(p)+E_e(q)-E(p')-E_e]\, (q') \left\{f_e(\vec{q}')\, f(\vec{p}') - f_e(\vec{q}) f(\vec{p})\right\}.$$

$$C[f(\vec{p})] = \frac{\pi}{4m_e^2 p}\int \frac{d^3q}{(2\pi)^3}\frac{d^3p'}{(2\pi)^3 p'} \delta\left[p + \frac{q^2}{2m_e} - p' - \frac{(\vec{q}+\vec{p}-\vec{p}')^2}{2m_e}\right] |\mathcal{M}|^2 \Big\{ f_e(\vec{q}+\vec{p}-\vec{p}')\, f(\vec{p}') - f_e(\vec{q}) f(\vec{p}) \Big\}$$

(снова, словами напишу, что происходит, всякие преобразования)

$$C[f(\vec{p})] = \frac{n_e \sigma_T}{p} \int_0^\infty dp' p' \left[\delta(p-p')\left(-p' \frac{\partial f^{(0)}}{\partial p'}\Theta_0 + p \frac{\partial f^{(0)}}{\partial p}\Theta(\hat{p})\right) + \vec{p}\cdot\vec{v}_{\rm b} \frac{\partial \delta(p-p')}{\partial p'}\left(f^{(0)}(p') - f^{(0)}(p)\right)\right]$$

$$C[f(\vec{p})] = -p\frac{\partial f^{(0)}}{\partial p} n_e \sigma_T \left[\Theta_0 - \Theta(\hat{p}) + \hat{p}\cdot\vec{v}_{\rm b}\right]$$

(??? тут картинка есть про монополь, которую я не понимаю.)

THE BOLTZMANN EQUATION FOR PHOTONS

$$\dot{\Theta} + \hat{p}^i \frac{\partial \Theta}{\partial x^i} + \dot{\Phi} + \hat{p}^i \frac{\partial \Psi}{\partial x^i} = n_e \sigma_T a \left[\Theta_0 - \Theta + \hat{p}\cdot\vec{v}_{\rm b}\right]. \qquad \text{(conf. t.)}$$

$$\dot{\tilde{\Theta}} + ik\mu\tilde{\Theta} + \dot{\tilde{\Phi}} + ik\mu\tilde{\Psi} = -\dot{\tau}\left[\tilde{\Theta}_0 - \tilde{\Theta} + \mu\tilde{v}_{\rm b}\right].$$

$$\dot{\tau} \equiv \frac{d\tau}{d\eta} = -n_e\sigma_T a.$$

$$\tau(\eta) \equiv \int_\eta^{\eta_0} d\eta' n_e \sigma_T a$$

(допишу слова про смысл происходящего)

THE BOLTZMANN EQUATION FOR COLD DARK MATTER

$$P^\mu = \left[E(1-\Psi), p\hat{p}^i \frac{1-\Phi}{a}\right].$$

$$\frac{df_{\rm dm}}{dt} = \frac{\partial f_{\rm dm}}{\partial t} + \frac{\partial f_{\rm dm}}{\partial x^i}\frac{dx^i}{dt} + \frac{\partial f_{\rm dm}}{\partial E}\frac{dE}{dt} + \frac{\partial f_{\rm dm}}{\partial \hat{p}^i}\frac{d\hat{p}^i}{dt}$$

$$\frac{\partial f_{\rm dm}}{\partial t} + \frac{\hat{p}^i}{a}\frac{p}{E}\frac{\partial f_{\rm dm}}{\partial x^i} - \frac{\partial f_{\rm dm}}{\partial E}\left[\frac{da/dt}{a}\frac{p^2}{E} + \frac{p^2}{E}\frac{\partial\Phi}{\partial t} + \frac{\hat{p}^i p}{a}\frac{\partial\Psi}{\partial x^i}\right] = 0$$

$$n_{\rm dm} = \int \frac{d^3p}{(2\pi)^3} f_{\rm dm}$$

$$v^i \equiv \frac{1}{n_{\rm dm}}\int \frac{d^3p}{(2\pi)^3} f_{\rm dm}\frac{p\hat{p}^i}{E}$$

$$\frac{\partial n_{\rm dm}}{\partial t}+\frac{1}{a}\frac{\partial\left(n_{\rm dm}v^i\right)}{\partial x^i}+3\left[\frac{da/dt}{a}+\frac{\partial\Phi}{\partial t}\right]n_{\rm dm}=0$$

$$\frac{\partial n^{(0)}_{\rm dm}}{\partial t}+3\frac{da/dt}{a}n^{(0)}_{\rm dm}=0$$

$$n_{\rm dm}=n^{(0)}_{\rm dm}[1+\delta(\vec{x},t)]$$

$$\frac{\partial\delta}{\partial t}+\frac{1}{a}\frac{\partial v^i}{\partial x^i}+3\frac{\partial\Phi}{\partial t}=0$$

(дальше словами)

$$\frac{\partial\left(n_{\rm dm}v^j\right)}{\partial t}+4\frac{da/dt}{a}n_{\rm dm}v^j+\frac{n_{\rm dm}}{a}\frac{\partial\Psi}{\partial x^j}=0.$$

$$\frac{\partial v^j}{\partial t}+\frac{da/dt}{a}v^j+\frac{1}{a}\frac{\partial\Psi}{\partial x^j}=0$$

$$\dot{\tilde{\delta}}+ik\tilde{v}+3\dot{\tilde{\Phi}}=0$$

$$\dot{\tilde{v}}+\frac{\dot{a}}{a}\tilde{v}+ik\tilde{\Psi}=0.$$

THE BOLTZMANN EQUATION FOR BARYONS

Electrons and protons are coupled by Coulomb scattering ($e+p\rightarrow e+p$).

$$\frac{\rho_e-\rho_e^{(0)}}{\rho_e^{(0)}}=\frac{\rho_p-\rho_p^{(0)}}{\rho_p^{(0)}}\equiv\delta_b$$

(словами про вывод)

$$\frac{\partial n_e}{\partial t}+\frac{1}{a}\frac{\partial\left(n_ev^i_{\rm b}\right)}{\partial x^i}+3\left[\frac{da/dt}{a}+\frac{\partial\Phi}{\partial t}\right]n_e=\left\langle c_{ep}\right\rangle_{QQ'q'q}+\left\langle c_{e\gamma}\right\rangle_{pp'q'q}$$

$$\dot{\tilde{\delta}}_b+ik\tilde{v}_{\rm b}+3\dot{\tilde{\Phi}}=0$$

(словами про вывод)

$$\frac{\partial v^j_{\rm b}}{\partial t}+\frac{da/dt}{a}v^j_{\rm b}+\frac{1}{a}\frac{\partial\Psi}{\partial x^j}=\frac{1}{\rho_b}\left\langle c_{e\gamma}q^j\right\rangle_{pp'q'q}$$

$$\dot{v}_{\rm b}+\frac{\dot{a}}{a}\tilde{v}_{\rm b}+ik\tilde{\Psi}=\dot{\tau}\frac{4\rho_\gamma}{3\rho_b}\left[3i\tilde{\Theta}_1+\tilde{v}_{\rm b}\right]$$

Perturbed Einstain equations by Dodelson

The Perturbed Ricci Tensor and Scalar

Two Components of the Einstein Equations

Tensor Perturbations

5.3.1 Christoffel Symbols for Tensor Perturbations
5.3.2 Ricci Tensor for Tensor Perturbations
5.3.3 Einstein Equations for Tensor Perturbations

The Decomposition Theorem

From Gauge to Gauge

THE EINSTEIN-BOLTZMANN EQUATIONS AT EARLY TIMES

(следующая часть Додельсона, просто относится к этой теме)

$$\dot{\Theta}_0 + \dot{\Phi} = 0$$
$$\dot{\mathcal{N}}_0 + \dot{\Phi} = 0.$$
$$\dot{\delta} = -3\dot{\Phi}$$
$$\dot{\delta}_{\rm b} = -3\dot{\Phi}.$$

$$3\frac{\dot{a}}{a}\left(\dot{\Phi} - \frac{\dot{a}}{a}\Psi\right) = 16\pi G a^2 \left(\rho_\gamma \Theta_0 + \rho_\nu \mathcal{N}_0\right)$$

But since radiation dominates, $a \propto \eta$ (recall Eq. (2.100) and the discussion immediately afterward) so $\dot{a}/a = 1/\eta$. Therefore,

$$\frac{\dot{\Phi}}{\eta} - \frac{\Psi}{\eta^2} = \frac{16\pi G\rho a^2}{3}\left(\frac{\rho_\gamma}{\rho}\Theta_0 + \frac{\rho_\nu}{\rho}\mathcal{N}_0\right) = \frac{2}{\eta^2}\left(\frac{\rho_\gamma}{\rho}\Theta_0 + \frac{\rho_\nu}{\rho}\mathcal{N}_0\right)$$

where the last equality follows by virtue of the zero-order Einstein equation.

To simplify further, we can define the ratio of neutrino energy density to the total radiation density as

$$f_\nu \equiv \frac{\rho_\nu}{\rho_\gamma + \rho_\nu}.$$

Then multiplying Eq. (6.4) by η^2 leads to

$$\dot{\Phi}\eta - \Psi = 2\left(\left[1 - f_\nu\right]\Theta_0 + f_\nu \mathcal{N}_0\right).$$

Recall that Eq. (6.1) relates the derivative of the monopoles to the derivative of the potential. We can therefore eliminate both monopoles from Eq. (6.6) by differentiating both right- and left-hand sides. Then,

$$\ddot{\Phi}\eta + \dot{\Phi} - \dot{\Psi} = -2\dot{\Phi}$$

where the right-hand side follows since both $\dot{\Theta}_0$ and $\dot{\mathcal{N}}_0$ are equal to $-\dot{\Phi}$ for these large scale modes.

So far we have used only one Einstein equation. The second, Eq. (5.33), describes how the higher moments of the photon and neutrino distributions cause $\Psi + \Phi$ to be nonzero. Let us here neglect these higher order moments, which cause the sum of the gravitational potentials to be slightly nonzero. [1] Under this approximation, we can eliminate Ψ everywhere by simply setting it to $-\Phi$. Then,

$$\ddot{\Phi}\eta + 4\dot{\Phi} = 0.$$

Setting $\Phi = \eta^p$ leads to the algebraic equation

$$p(p-1) + 4p = 0$$

$$\Theta_1 = \mathcal{N}_1 = \frac{iv_{\rm b}}{3} = \frac{iv}{3}$$
$$= -\frac{k\Phi}{6aH}.$$

ζ conservation for superhorizon modes and power spectrum $P_\Psi(k)$

Fluctuations generated during inflation are related to the metric perturbations that they generate at later times by

$$\zeta := -\Psi + \frac{1}{3}\frac{\delta\rho}{\rho + P} \qquad \text{constant for sup-hor modes.}$$

It doesn't depend on the eq. of state.

During slow-roll inflation, for modes inside the horizon and those which have just left the horizon, $\Psi \approx 0$. From $\delta\rho = V'\left(\phi^{(0)}\right)\delta\phi$ and $d\rho/dt = -3H(\rho + P)$

$$\zeta = -aH\frac{\delta\phi}{\dot{\phi}^{(0)}} \qquad \text{(during inflation).}$$

$$\zeta = -\frac{3}{2}\Psi \qquad \text{(after inflation).} \qquad \Psi = -\frac{1}{2}\delta_\gamma = \text{ constant}$$

(!!! еще раз решу это скоро!!!)

The conservation of ζ for superhorizon modes gives $\Psi(\delta\phi)$, with it from definitions we get a relation between the post-inflation power spectrum of Ψ and the power spectrum of $\delta\phi$ at horizon crossing during inflation:

$$P_\Psi(k) = \frac{4}{9}\left(\frac{aH}{\dot{\phi}^{(0)}}\right)^2 P_{\delta\phi}\bigg|_{aH=k}.$$

$$4\pi G\left(\dot{\phi}^{(0)}\right)^2 = \epsilon a^2 H^2 \qquad \Rightarrow \qquad P_\Psi(k) = \frac{16\pi G}{9\epsilon}P_{\delta\phi}\bigg|_{aH=k} = \frac{8\pi G H^2}{9\epsilon k^3}\bigg|_{aH=k}.$$

Here little ϵ is in the denominator, thus the gravitational potential induced by the time delay much larger than the metric perturbations during inflation. The flatter the potential, the longer it takes to cover an additional distance in ϕ space and the more curvature is generated by additional inflation.

(relations of power spectrum from Dodelson)

We can immediately relate the power in ζ to the power in $\delta\phi$,

$$P_\zeta = \left(\frac{aH}{\dot{\phi}^{(0)}}\right)^2 P_{\delta\phi}.$$

We know $P_{\delta\phi}$ from Eq. (6.72) and the prefactor is $4\pi G/\epsilon$, so

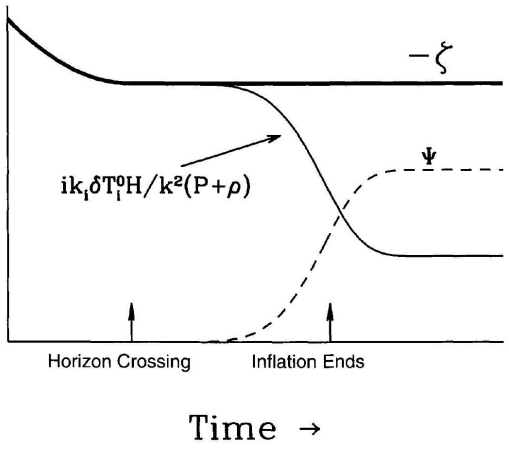
$$P_\zeta = \frac{2\pi G H^2}{\epsilon k^3}\bigg|_{aH=k}.$$

Equation (6.96) is very useful, for it expresses the power spectrum of a gauge invariant quantity. Although we computed it in a gauge of the form in Eq. (6.91), once we have this answer, we can compute ζ in any gauge and then relate the power in the perturbation variables of that gauge to P_ζ.

Throughout this book, we have been working in conformal Newtonian gauge. In this gauge, $\Phi_H = -\Phi$, so ζ as defined in Eq. (6.93) is indeed given by Eq. (6.77). We argued in Section 6.5.2 that in conformal Newtonian gauge, after inflation, $\zeta = 3\Phi/2$, so $P_\Phi = 4P_\zeta/9$, or using Eq. (6.96),

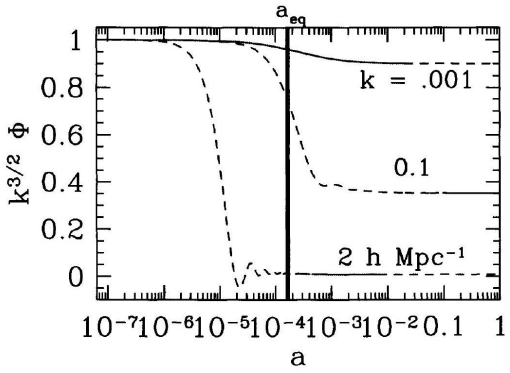
$$P_\Phi = \frac{8\pi G H^2}{9\epsilon k^3}\bigg|_{aH=k}$$

in exact agreement with our earlier calculation finalized in Eq. (6.82).



1.3.5 Dynamics of fluctuations, power spectrum and spectral indexes

Evolution of gravitational potential fluctuations



Inflation sets the initial condition for these superhorizon modes Φ in Newtonian gauge

$$\mathcal{R} = -\Phi - \frac{2}{3(1+w)}\left(\frac{\Phi'}{\mathcal{H}} + \Phi\right)$$

$$\Phi'' + 3(1+w)\mathcal{H}\Phi' + wk^2\Phi = 0$$

Superhorizon Limit

$k \ll \mathcal{H}$,

$$\Phi = \text{ const.} \qquad \text{(superhorizon)}$$

$$\delta = -\frac{2}{3}\frac{k^2}{\mathcal{H}^2}\Phi - \frac{2}{\mathcal{H}}\Phi' - 2\Phi \qquad \text{(Poisson eq)}$$

$$\delta \approx -2\Phi = \text{ const.}$$

$$\delta_m = \frac{3}{4}\delta_r \approx -\frac{3}{2}\Phi_{\rm RD},$$

Radiation-to-Matter Transition

$$\mathcal{R} = -\frac{5+3w}{3+3w}\Phi \quad \text{(superhorizon)} \qquad \mathcal{R} := -\Phi - \frac{\mathcal{H}\left(\Phi' + \mathcal{H}\Phi\right)}{4\pi G a^2(\bar{\rho}+\bar{P})}$$

$$\mathcal{R} = -\frac{3}{2}\Phi_{\rm RD} = -\frac{5}{3}\Phi_{\rm MD} \Rightarrow \Phi_{\rm MD} = \frac{9}{10}\Phi_{\rm RD},$$

Gravitational potential fluctuations in Radiation Era

$w = \frac{1}{3}$, we get

$$\Phi'' + \frac{4}{\tau}\Phi' + \frac{k^2}{3}\Phi = 0$$

$$\Phi_k(\tau) = A_k\frac{j_1(x)}{x} + B_k\frac{n_1(x)}{x}, \quad x := \frac{1}{\sqrt{3}}k\tau$$

here $j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x} = \frac{x}{3} + \mathcal{O}\left(x^3\right)$, $n_1(x) = -\frac{\cos x}{x^2} - \frac{\sin x}{x} = -\frac{1}{x^2} + \mathcal{O}\left(x^0\right)$.

$$\Phi_k(\tau) = -2\mathcal{R}_k(0)\left(\frac{\sin x - x\cos x}{x^3}\right) \quad \text{(all scales)}$$

$$\Phi_{\boldsymbol{k}}(\tau) \approx -6\mathcal{R}_{\boldsymbol{k}}(0)\frac{\cos\left(\frac{1}{\sqrt{3}}k\tau\right)}{(k\tau)^2} \quad \text{(subhorizon)}$$

During the radiation era, subhorizon modes of Φ therefore oscillate with frequency $\frac{1}{\sqrt{3}}k$ and an amplitude that decays as $\tau^{-2} \propto a^{-2}$ (see fig. 5.2). Remember this.

Gravitational potential fluctuations in Matter Era

$w = 0$,

$$\Phi'' + \frac{6}{\tau}\Phi' = 0$$

$$\Phi \propto \begin{cases} \text{const.} \\ \tau^{-5} \propto a^{-5/2} \end{cases}$$

Evolution of radiation fluctuations

Acoustic oscillations in the perturbed radiation density are what gives rise to the peaks in the spectrum of CMB anisotropies
(??? plot???])

radiation fluctuations in Radiation Era

In radiation era for all scales $\Phi_k(\tau) = -2\mathcal{R}_k(0)\left(\frac{\sin x - x\cos x}{x^3}\right)$, so from the Poisson equation

$$\delta_r = -\frac{2}{3}(k\tau)^2\Phi - 2\tau\Phi' - 2\Phi,$$
$$\Delta_r = -\frac{2}{3}(k\tau)^2\Phi.$$

(??? why?? what is Δ_r???)
Outside the horizon δ_r is constant, Δ_r grows as $\tau^2 \propto a^2$.

Inside the horizon,

$$\delta_r \approx \Delta_r = -\frac{2}{3}(k\tau)^2\Phi = 4\mathcal{R}(0)\cos\left(\frac{1}{\sqrt{3}}k\tau\right),$$

which is the solution to

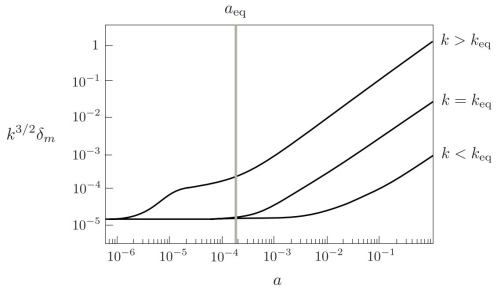
$$\delta_r'' - \frac{1}{3}\nabla^2\delta_r = 0$$

We see that subhorizon fluctuations in the radiation density oscillate with constant amplitude around $\delta_r = 0$.

radiation fluctuations in Matter Era

$$\left.\begin{array}{l}\text{(C) } \delta_r' = -\frac{4}{3}\boldsymbol{\nabla}\cdot\boldsymbol{v}_r \\ \text{(E) } \boldsymbol{v}_r' = -\frac{1}{4}\boldsymbol{\nabla}\delta_r - \boldsymbol{\nabla}\Phi\end{array}\right\} \quad \delta_r'' - \frac{1}{3}\nabla^2\delta_r = \frac{4}{3}\nabla^2\Phi \quad = \text{ const.}$$

Evolution of dark matter fluctuations



The effects we discussed above lead to a post-processing of the primordial perturbations. This evolution is often encoded in the so-called transfer function. For example, the value of the matter perturbation at redshift z is related to the primordial perturbation $\mathcal{R}_{\boldsymbol{k}}$ by

$$\Delta_{m,k}(z) = T(k,z)\mathcal{R}_k$$

$$P_\Delta(k,z) := |\Delta_{m,k}(z)|^2 = T^2(k,z)\,|\mathcal{R}_k|^2$$

Harrison-Zel'dovich plot shows predicted matter power spectrum for scale-invariant initial conditions, $k^3\,|\mathcal{R}_k|^2 = \text{const}$:

$$P_\Delta(k) = \begin{cases} k & k < k_{\rm eq} \\ k^{-3} & k > k_{\rm eq}\end{cases}$$

(??? why sc inv is important??? why inflation gives exactly this spectrum?????)

dark matter fluctuations at Early Times

$$\mathcal{H}^2 = \frac{\mathcal{H}_0^2\Omega_m^2}{\Omega_r}\left(\frac{1}{y}+\frac{1}{y^2}\right), \quad y := \frac{a}{a_{\rm eq}}$$

$$\left.\begin{array}{l}\text{(C) } \delta_m' = -\boldsymbol{\nabla}\cdot\boldsymbol{v}_m \\ \text{(E) } \boldsymbol{v}_m' = -\mathcal{H}\boldsymbol{v}_m - \boldsymbol{\nabla}\Phi\end{array}\right\} \quad \delta_m'' + \mathcal{H}\delta_m' = \nabla^2\Phi$$

$$\delta_m'' + \mathcal{H}\delta_m' - 4\pi G a^2\bar{\rho}_m\delta_m \approx 0$$

or in Mészáros (Meszaros) equation

$$\frac{d^2\delta_m}{dy^2} + \frac{2+3y}{2y(1+y)}\frac{d\delta_m}{dy} - \frac{3}{2y(1+y)}\delta_m = 0.$$

$$\delta_m \propto \begin{cases} 2+3y \\ (2+3y)\ln\left(\frac{\sqrt{1+y}+1}{\sqrt{1+y}-1}\right) - 6\sqrt{1+y} \end{cases}$$

dark matter fluctuations at Intermediate Times

$$\Delta_m = \frac{\nabla^2\Phi}{4\pi G a^2\bar{\rho}} \propto \begin{cases} a \\ a^{-3/2} \end{cases}$$

dark matter fluctuations at Late Times

$$\nabla^2\Phi = 4\pi G a^2 \bar{\rho}_m \Delta_m$$

Pressure fluctuations are negligible, so the Einstein equations give

$$\Phi'' + 3\mathcal{H}\Phi' + \left(2\mathcal{H}' + \mathcal{H}^2\right)\Phi = 0$$

To get an evolution equation for Δ_m, we use a neat trick. Since $a^2\bar{\rho}_m \propto a^{-1}$, we have $\Phi \propto \Delta_m/a$. Hence, eq. (5.2.28) implies

$$\partial_\tau^2\left(\Delta_m/a\right) + 3\mathcal{H}\partial_\tau\left(\Delta_m/a\right) + \left(2\mathcal{H}' + \mathcal{H}^2\right)\left(\Delta_m/a\right) = 0$$

which rearranges to

$$\Delta_m'' + \mathcal{H}\Delta_m' + \left(\mathcal{H}' - \mathcal{H}^2\right)\Delta_m = 0$$

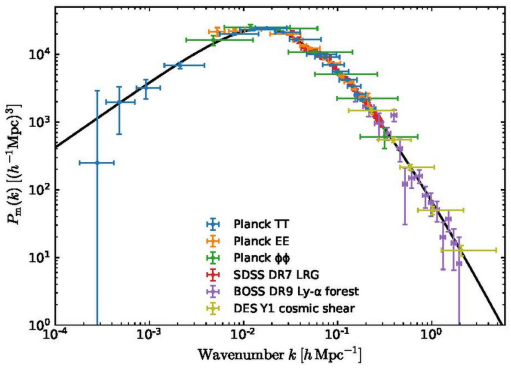
Exercise. - Show that (5.2.30) follows from (5.2.29). Use the Friedmann and conservation equations to show that

$$\mathcal{H}' - \mathcal{H}^2 = -4\pi G a^2(\bar{\rho} + \bar{P}) = -4\pi G a^2\bar{\rho}_m$$

Using (5.2.31), eq. (5.2.30) becomes

$$\Delta_m'' + \mathcal{H}\Delta_m' - 4\pi G a^2\bar{\rho}_m\Delta_m = 0.$$

Power Spectrum from DM



The matter power spectrum $P_\Delta(k)$ at $z = 0$ in linear theory (solid) and with non-linear corrections (dashed). On large scales, $P_\Delta(k)$ grows as k. The power spectrum turns over around

$k_{\rm eq} \sim 0.01{\rm Mpc}^{-1}$ corresponding to the horizon size at matter-radiation equality. Beyond the peak, the power falls as k^{-3}. Visible are small amplitude baryon acoustic oscillations in the spectrum.

(??? описание и качественное объяснение графика!)

From

$$\ddot{h} + 2\frac{\dot{a}}{a}\dot{h} + k^2 h = 0.$$

$$\tilde{h} := \frac{ah}{\sqrt{16\pi G}}$$

$$\hat{\tilde{h}}(\vec{k},\eta) = v(k,\eta)\hat{a}_{\vec{k}} + v^*(k,\eta)a^\dagger_{\vec{k}},$$

where the coefficients of the creation and annihilation operators satisfy the equation

$$\ddot{v} + \left(k^2 - \frac{\ddot{a}}{a}\right) v = 0.$$

$$v = \frac{e^{-ik\eta}}{\sqrt{2k}}\left[1 - \frac{i}{k\eta}\right]$$

$$\left\langle \hat{\tilde{h}}^\dagger(\vec{k},\eta)\hat{\tilde{h}}(\vec{k}',\eta)\right\rangle = |v(\vec{k},\eta)|^2(2\pi)^3\delta^3(\vec{k}-\vec{k}') = \frac{16\pi G}{a^2}|v(k,\eta)|^2(2\pi)^3\delta^3(\vec{k}-\vec{k}')$$
$$:= (2\pi)^3 P_h(k)\delta^3(\vec{k}-\vec{k}')$$

(тут у нас $k\eta \ll 1$ что ли, поэтому числитель 1????)

$$P_h(k) = \frac{16\pi G}{a^2}\frac{1}{2k^3\eta^2}$$
$$= \frac{8\pi G H^2}{k^3}.$$

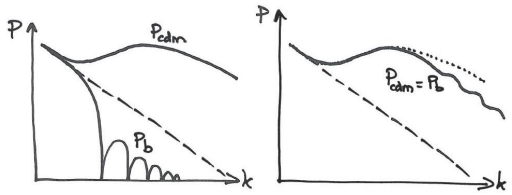
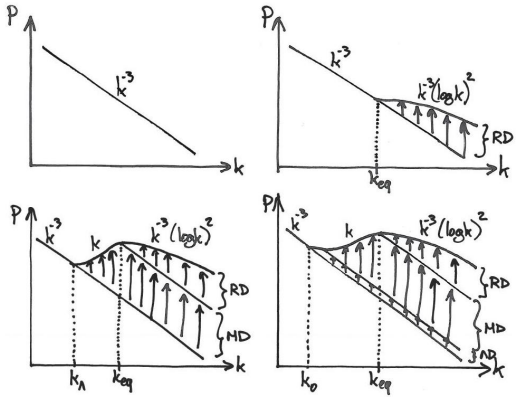


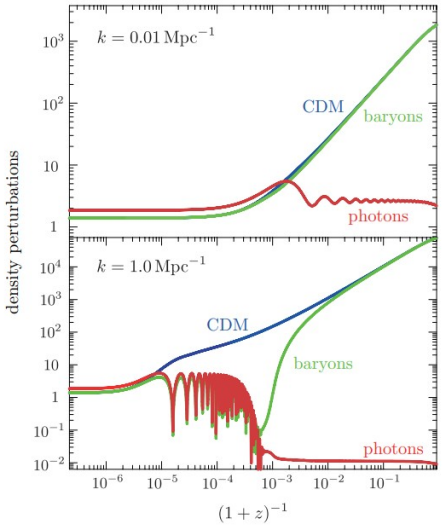
Fig. 14. Shape of the matter power spectrum $P(k)$ (log-log scale) at four different times: (upper left) when initial conditions are imposed (and all wavenumbers are super-Hubble); (upper right) at radiation/matter equality (arrows show the logarithmic growth during radiation domination); (lower left) at matter /Λ equality (lower set of arrows show the growth during matter domination); (lower right) today (lower set of arrows show the growth during Λ domination).



(I'll add description from Lesgourgues later, now it is not needed)

On other topics about fluctuation

Evolution of Baryons fluctuation (?)



(это не нужно, потом мб почитаю, пока не планирую)

$$D'' + \frac{2}{\tau} D' = 0 \Rightarrow D \propto \begin{cases} \text{const.} \\ \tau^{-1} \end{cases},$$

of δ_m is governed

$$\delta_m'' + \frac{2}{\tau}\delta_m' - \frac{6}{\tau^2}\delta_m = 0 \quad \Rightarrow \quad \delta_m \propto \begin{cases} \tau^2 \\ \tau^{-3} \end{cases}.$$

$$\frac{\delta_b}{\delta_c} = \frac{\bar{\rho}_m \delta_m + \bar{\rho}_c D}{\bar{\rho}_m \delta_m - \bar{\rho}_b D} \to \frac{\delta_m}{\delta_m} = 1,$$

Tensors and Lyth bound properties

(here will be a good preparation for section of problems)

The Lyth argument places a theoretical upper bound on the amount of gravitational waves produced during inflation, given the amount of departure from the homogeneity of the cosmic microwave background (CMB).

- During slow-roll inflation, the ratio of gravitational waves to inhomogeneities of the CMB is correlated to the inflationary potential steepness.
- Temperature inhomogeneities of the CMB were successfully and accurately measured in the CMB.
- There are current CMB polarization experiments (see this article for instance for an overview of gravitational wave observatories) aimed at measuring the primordial gravitational wave signature in the CMB.
- However, to date, a significant signal of primordial gravitational waves was not detected. Thus the ratio cannot exceed a certain value.
- Thus the steepness of the inflationary potential is bounded.

$$r \equiv \frac{A_t}{A_s} \overset{\text{SR}}{=} \frac{8\dot{\phi}^2}{M_{\text{pl}}^2 H^2}.$$

The inflaton field travels during (observable) inflation a "distance"

$$\Delta\phi \equiv |\phi_E - \phi_\star| : \qquad \frac{\Delta\phi}{M_{\text{pl}}} = \frac{N_\star}{60}\sqrt{\frac{r}{0.002}},$$

($N_\star$ is the total number of e-folds between the time when the CMB scales exited the horizon and the end of inflation).

$$V^{1/4} = \left(\frac{3\pi^2}{2} r A_s\right)^{1/4} M_{\rm pl.} \approx 3.3 \cdot r^{1/4} \cdot 10^{16}\text{GeV}.$$

This exceeds the highest energy scales that will be probed at the LHC by a factor $\mathcal{O}(10^{13})$!

Dynamics of dark matter on sub-Hubble scales from Newtonian perturbation theory

DM during the radiation-dominated era: from $\ddot{\delta} + 2H\dot{\delta} - \frac{c_s^2}{a^2}\nabla^2\delta = 4\pi G\bar{\rho}\delta$

$$\ddot{\delta}_m + 2H\dot{\delta}_m - 4\pi G \sum_I \bar{\rho}_I \delta_I = 0 \qquad I \in \text{matter, radiation.}$$

$$\ddot{\delta}_m + \frac{1}{t}\dot{\delta}_m - 4\pi G\bar{\rho}_m\delta_m \approx 0.$$

$$\ddot{\delta}_m \sim H^2\delta_m \sim \frac{8\pi G}{3}\bar{\rho}_r\delta_m \gg 4\pi G\bar{\rho}_m\delta_m$$

where we have used that $\bar{\rho}_r \gg \bar{\rho}_m$. We can therefore ignore the last term in (4.1.32) compared to the others. We then find

$$\delta_m \propto \begin{cases} \text{const.} \\ \ln t \propto \ln a \end{cases}.$$

The rapid expansion due to the effectively unclustered radiation reduces the growth of δ_m to only logarithmic. We need to wait until the universe becomes matter dominated in order for the dark matter density fluctuations to grow significantly.

DM during the matter-dominated era: from $\ddot{\delta} + 2H\dot{\delta} - \frac{c_s^2}{a^2}\nabla^2\delta = 4\pi G\bar{\rho}\delta$

$$\ddot{\delta}_m + 2H\dot{\delta}_m - 4\pi G\bar{\rho}_m\delta_m = 0$$

Here $a \propto t^{2/3}$, so $H = 2/3t$ and by $4\pi G\bar{\rho}_m = \frac{3}{2}H^2$

$$\ddot{\delta}_m + \frac{4}{3t}\dot{\delta}_m - \frac{2}{3t^2}\delta_m = 0,$$

$\delta_m \propto t^p$

$$\delta_m \propto \begin{cases} t^{-1} \propto a^{-3/2} \\ t^{2/3} \propto a \end{cases}.$$

Hence, the growing mode of dark matter fluctuations grows like the scale factor during the MD era.

DM during the Λ-dominated era: from $\ddot{\delta} + 2H\dot{\delta} - \frac{c_s^2}{a^2}\nabla^2\delta = 4\pi G\bar{\rho}\delta$

$$\ddot{\delta}_m + 2H\dot{\delta}_m - 4\pi G \sum_i \bar{\rho}_I \delta_I = 0$$

where $I = m, \Lambda$. As far as we can tell, dark energy doesn't cluster (almost by definition), so we can write

$$\ddot{\delta}_m + 2H\dot{\delta}_m - 4\pi G\bar{\rho}_m\delta_m = 0$$

In the Λ dominated regime $H^2 \approx \text{const} \gg 4\pi G\bar{\rho}_m$.

$$\delta_m \propto \begin{cases} \text{const.} \\ e^{-2Ht} \propto a^{-2} \end{cases}$$

Matter fluctuations stop growing once dark energy comes to dominate.

Primordial Perturbations of Curvature

At horizon crossing, we switch from the inflaton fluctuation $\delta\phi$ to the conserved curvature perturbation $\mathcal{R}$. The power spectra of $\mathcal{R}$ and $\delta\phi$ are related via eq. (6.1.7),

$$\Delta^2_{\mathcal{R}} = \frac{1}{2\varepsilon}\frac{\Delta^2_{\delta\phi}}{M^2_{\rm pl}}, \quad \text{where} \quad \varepsilon = \frac{\frac{1}{2}\dot{\phi}^2}{M^2_{\rm pl}H^2}.$$

Substituting (6.4.50), we get

$$\Delta^2_{\mathcal{R}}(k) = \frac{1}{8\pi^2}\frac{1}{\varepsilon}\frac{H^2}{M^2_{\rm pl}}\bigg|_{k=aH}$$

Exercise. - Show that for slow-roll inflation, eq. (6.5.59) can be written as

$$\Delta^2_{\mathcal{R}} = \frac{1}{12\pi^2}\frac{V^3}{M^6_{\rm pl}\left(V'\right)^2}$$

Because the right-hand side of (6.5.59) is evaluated at $k = aH$, the power spectrum is purely a function of k. If $\Delta^2_{\mathcal{R}}(k)$ is k-independent, then we call the spectrum scale-invariant. However, since H and possibly ε are (slowly-varying) functions of time, we predict that the power spectrum will deviate slightly from the scale-invariant form $\Delta^2_{\mathcal{R}} \sim k^0$. Near a reference scale $k_\star$, the k dependence of the spectrum takes a power-law form

$$\Delta^2_{\mathcal{R}}(k) := A_s\left(\frac{k}{k_\star}\right)^{n_s-1}$$

The measured amplitude of the scalar spectrum at $k_\star = 0.05\text{Mpc}^{-1}$ is

$$A_s = (2.196 \pm 0.060)\cdot 10^{-9}$$

To quantify the deviation from scale-invariance we have introduced the scalar spectral index

$$n_s - 1 := \frac{d\ln\Delta^2_{\mathcal{R}}}{d\ln k}$$

where the right-hand side is evaluated at $k = k_\star$ and $n_s = 1$ corresponds to perfect scaleinvariance. We can split (6.5.61) into two factors

$$\frac{d\ln\Delta^2_{\mathcal{R}}}{d\ln k} = \frac{d\ln\Delta^2_{\mathcal{R}}}{dN}\cdot\frac{dN}{d\ln k}$$

The derivative with respect to e-folds is

$$\frac{d\ln\Delta^2_{\mathcal{R}}}{dN} = 2\frac{d\ln H}{dN} - \frac{d\ln\varepsilon}{dN}$$

The first term is just -2ε and the second term is $-\eta$ (see Chapter 2). The second factor in (6.5.62) is evaluated by recalling the horizon crossing condition $k = aH$, or

$$\ln k = N + \ln H$$

Hence, we have

$$\frac{dN}{d\ln k} = \left[\frac{d\ln k}{dN}\right]^{-1} = \left[1 + \frac{d\ln H}{dN}\right]^{-1} \approx 1 + \varepsilon.$$

To first order in the Hubble slow-roll parameters, we therefore find

$$n_s - 1 = -2\varepsilon - \eta.$$

Primordial Perturbations of Gravitational Waves and tensor spectral indexes

$$\mathrm{d}s^2 = a^2(\tau)\left[\mathrm{d}\tau^2 - \left(\delta_{ij} + 2\hat{E}_{ij}\right)\mathrm{d}x^i\ \mathrm{d}x^j\right]$$

$$S = \frac{M_{\rm pl}^2}{2}\int \mathrm{d}^4x\sqrt{-g}R \quad \Rightarrow \quad S^{(2)} = \frac{M_{\rm pl}^2}{8}\int \mathrm{d}\tau\mathrm{d}^3x a^2\left[\left(\hat{E}'_{ij}\right)^2 - \left(\nabla\hat{E}_{ij}\right)^2\right].$$

It is convenient to define

$$\frac{M_{\rm pl}}{2}a\hat{E}_{ij} := \frac{1}{\sqrt{2}}\begin{pmatrix} f_+ & f_\cdot & 0 \\ f_\cdot & -f_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

so that

$$S^{(2)} = \frac{1}{2}\sum_{I=+,\cdot}\int \mathrm{d}\tau\mathrm{d}^3x\left[\left(f'_I\right)^2 - \left(\nabla f_I\right)^2 + \frac{a''}{a}f_I^2\right]$$

$$\Delta_t^2 := 2\cdot\Delta_{\hat{E}}^2 = 2\cdot\left(\frac{2}{aM_{\rm pl}}\right)^2\cdot\Delta_f^2$$

Using (6.4.50), we get

$$\Delta_t^2(k) = \left.\frac{2}{\pi^2}\frac{H^2}{M_{\rm pl}^2}\right|_{k=aH}.$$

This result is the most robust and model-independent prediction of inflation. Notice that the tensor amplitude is a direct measure of the expansion rate H during inflation. This is in contrast to the scalar amplitude which depends on both H and ε.

The scale-dependence of the tensor spectrum is defined by n_t - tensor spectral index:

$$\Delta_t^2(k) := A_t\left(\frac{k}{k_\star}\right)^{n_t}$$

Scale-invariance corresponds to $n_t = 0$.

Tensor-to-scalar ratio is

$$r := \frac{A_t}{A_s}$$

$$r = 16\varepsilon$$
$$n_t = -2\varepsilon$$

Thus always $n_t = -r/8$.

Correlation and concept of power spectrum

(просто мысли Лесгоргуеса впишу и все.)

1.3.6 Details of evolution of fluctuations, other approaches in a nutshell

(now just by Dodelson. это пока не актуально, тут пока углубления Додельсона, и без них можно многое важное понять по упрощенным лекциям. потом тут будут указания про более глубокое решение этой же задачи.)

Transfer Function (!!??!?!?!!?????)

The transfer function for gravity wave evolution as

$$T(k) := \frac{h_\alpha(k,\eta)}{h_\alpha(k,\eta=0)} \left(\frac{k\eta}{3j_1(k\eta)}\right).$$

You should recognize the term in parentheses as the inverse of the matter-dominated solution you derived in Exercise 12. Solve Eq. (5.63) numerically and compute the transfer function. Compare your solution with the fit of Turner, White, and Lidsey (1993),

$$T(y) = \left[1 + 1.34y + 2.5y^2\right]^{1/2}$$

where $y := (k\eta_0/370h)$ (with h parametrizing the Hubble constant). Assume the universe today is flat and matter dominated, but account for transition from matter to radiation.

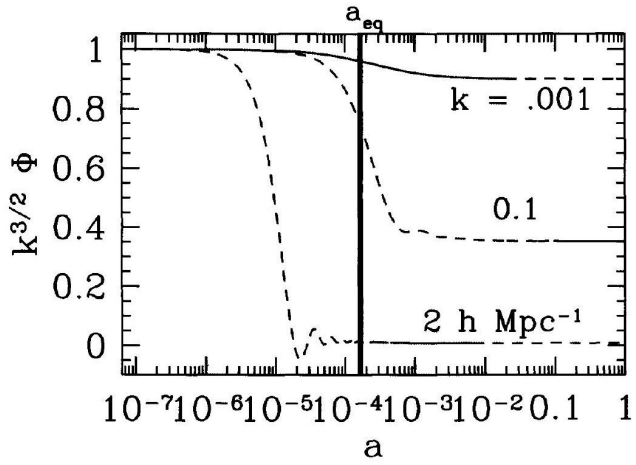


Figure 2: Figure 7.2. The linear evolution of the gravitional potential Φ. Dashed line denotes that the mode has entered the horizon. Evolution through the shaded region is described by the transfer function. The potential is unnormalized, but the relative normalization of the three modes is as it would be for scale-invariant perturbations. Here baryons have been neglected, $\Omega_m = 1$, and $h = 0.5$.

$$\Phi(\vec{k}, a) = \Phi_{\rm p}(\vec{k}) \cdot \{\text{Transfer Function } (k)\} \cdot \{\text{ Growth Function } (a)\}$$

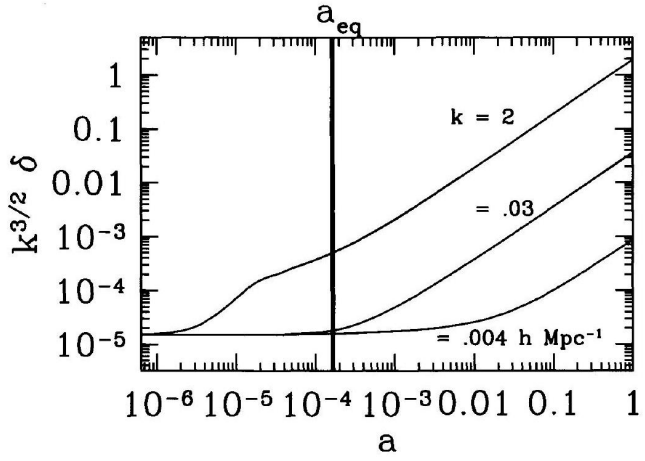
$$T(k) := \frac{\Phi\,(k, a_{\rm late}\,)}{\Phi_{\text{Large-Scale}}\ (k, a_{\rm late}\,)} \qquad \text{(transfer function)}$$

where $a_{\rm late}$ denotes an epoch well after the transfer function regime and the LargeScale solution is the primordial Φ decreased by a small amount. We will derive in Section 7.2 that - neglecting anisotropic stresses - this factor is equal to (9/10). The second caveat concerns the growth function. The ratio of the potential to its value right after the transfer function regime is defined to be

$$\frac{\Phi(a)}{\Phi\,(a_{\rm late}\,)} := \frac{D_1(a)}{a} \quad (a > a_{\rm late}\,),$$

where D_1 is called the growth function. In the flat, matter-dominated case depicted in Figure 7.2, then, the potential is constant so $D_1(a) = a$. With these conventions, we have

$$\Phi(\vec{k}, a) = \frac{9}{10}\Phi_{\rm p}(\vec{k})T(k)\frac{D_1(a)}{a} \quad (a > a_{\rm late}\,)$$



potential at late times is to use Poisson's equation (the large- k, no-radiation limit of Eq. (5.81))

$$\Phi = \frac{4\pi G\rho_{\rm m}a^2\delta}{k^2} \quad (a > a_{\rm late}\,).$$

The background density of matter is $\rho_{\rm m} = \Omega_m\rho_{\rm cr}/a^3$, and $4\pi G\rho_{\rm cr} = (3/2)H_0^2$, so

$$\delta(\vec{k},a) = \frac{k^2\Phi(\vec{k},a)a}{(3/2)\Omega_m H_0^2} \quad (a > a_{\rm late}\,).$$

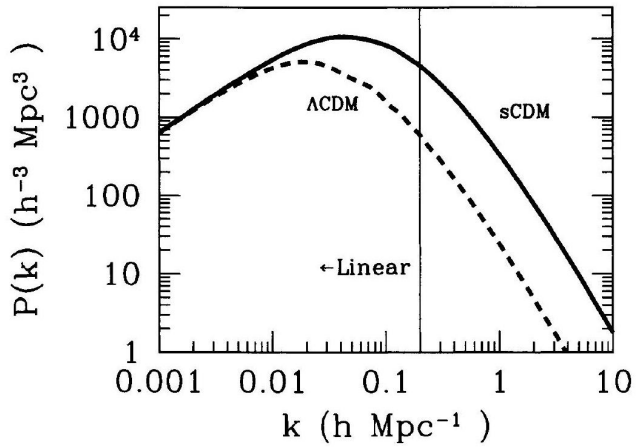
This, together with Eq. (7.5), allows us to relate the overdensity today to the primordial potential

$$\delta(\vec{k},a) = \frac{3}{5}\frac{k^2}{\Omega_m H_0^2}\Phi_{\rm p}(\vec{k})T(k)D_1(a) \quad (a > a_{\rm late}\,).$$

Equation (7.8) holds regardless of how the initial perturbation $\Phi_{\rm p}$ was generated. In the context of inflation, $\Phi_{\rm p}(\vec{k})$ is drawn from a Gaussian distribution with mean zero and variance (Eq. (6.100)) $P_\Phi = (50\pi^2/9k^3)\,(k/H_0)^{n-1}\,\delta_H^2\,(\Omega_m/D_1(a=1))^2$. So the power spectrum of matter at late times is

$$P(k,a) = 2\pi^2\delta_H^2\frac{k^n}{H_0^{n+3}}T^2(k)\left(\frac{D_1(a)}{D_1(a=1)}\right)^2 \quad (a > a_{\rm late}\,).$$

Matter-power spectrum:



(!!!!!! допишу потом асиптотики, так и не понял их!!!)

The power spectrum in two Cold Dark Matter models, with (ACDM) and without (sCDM) a cosmological constant. The spectra have been normalized to agree on large scales. The spectrum in the cosmological constant model turns over on larger scales because of a later aeq. Scales to the left of the vertical line are still evolving linearly.

Evolution of Large Scales perturbations

7.2.1

Super-horizon Solution (!!!)

(тут много важных формул, скоро пропишу их!!!)

$$\dot{\Theta}_{r,0} = -\dot{\Phi}$$
$$\dot{\delta} = -3\dot{\Phi}$$
$$3\frac{\dot{a}}{a}\left(\dot{\Phi} + \frac{\dot{a}}{a}\Phi\right) = 4\pi G a^2 \left[\rho_{\rm dm}\delta + 4\rho_r\Theta_{r,0}\right].$$

(словами про математику тут)

$$\Phi'' + \frac{21y^2 + 54y + 32}{2y(y+1)(3y+4)}\Phi' + \frac{\Phi}{y(y+1)(3y+4)} = 0.$$

7.2.2

Through Horizon Crossing

Evolution of Small Scales perturbations

7.3.1

Horizon Crossing

7.3.2

Sub-horizon Evolution

Numerical Results and Fits (????)

Growth Function (???)

Since radiation can be ignored, the coefficient multiplying δ in Eq. (7.56) is now $4\pi G\rho_{\rm dm} = (3/2)H_0^2\Omega_m a^{-3}$. Also when differentiating Eq. (7.54) previously, we set $(1/aHy)' = -(1+y)^{-1}(2aHy)^{-1}$; here we need to account for other contributions to H' so Eq. (7.57) becomes

$$\delta'' + ikv\left(\frac{d(aHy)^{-1}}{dy} - \frac{1}{aHy^2}\right) = \frac{3\Omega_m H_0^2}{2y^3a^2H^2a_{\rm eq}}\delta.$$

Replacing the velocity term using the continuity equation as before leads to

$$\frac{d^2\delta}{da^2} + \left(\frac{d\ln(H)}{da} + \frac{3}{a}\right)\frac{d\delta}{da} - \frac{3\Omega_m H_0^2}{2a^5H^2}\delta = 0.$$

Here I have divided by $a_{\rm eq}^2$ and we will now use a as the variable instead of y. In this large y limit, all factors of $a_{\rm eq}$ disappear.

$$D_1(a) \propto H(a)\int^a \frac{da'}{\left(a'H\left(a'\right)\right)^3}$$

Beyond Cold Dark Matter (?!!!!)

(конец 7го раздела Додельсона, спишу позже)

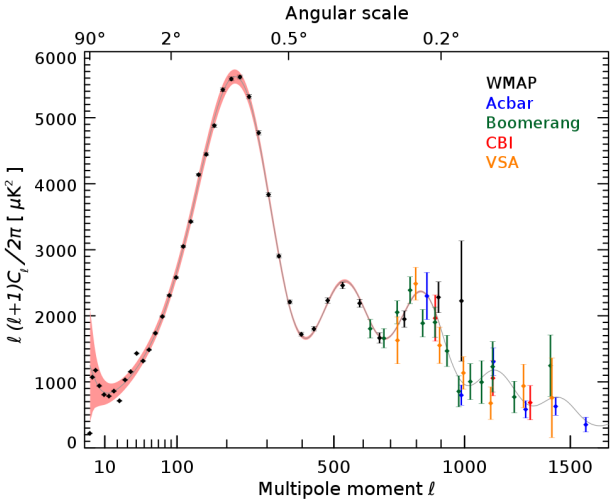
Baryons

Massive Neutrinos

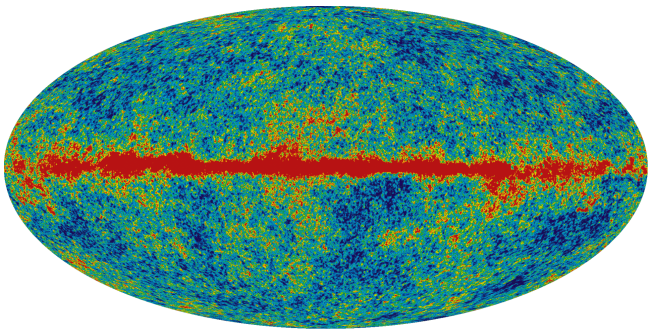
Dark Energy

1.3.7 Anisotropy predictions, correlations in a nutshell (!!!!!!!!!)

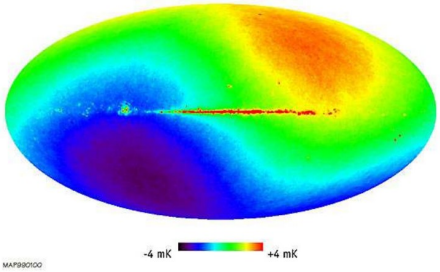
Main properties of CMB and parameters for it



(???? в чем смысл этих параметров тут???)

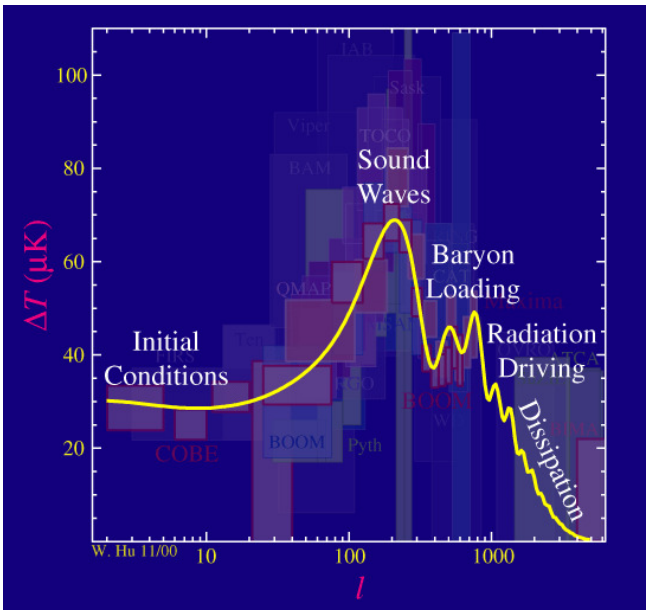


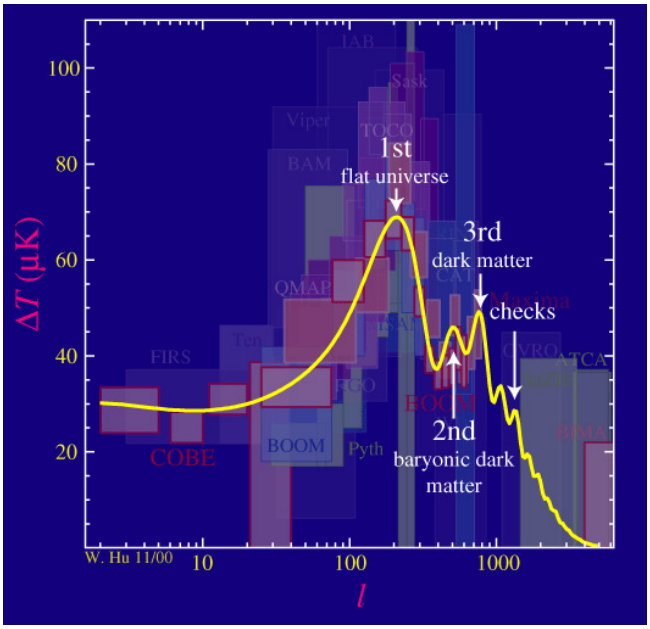
Тут у нас изображена карта звездного неба, и на ней выбраны фотоны с очень малой температурой. Цветом показана их температура.



Можно размыть его, выделить большие круги, посмотреть, насколько они скоррелированы (звучит как вообще бред). Корреляции же можно теоретически посчитать.

$$\Theta_l := \frac{1}{(-i)^l} \int_{-1}^{1} \frac{d\mu}{2} \mathcal{P}_l(\mu)\Theta(\mu)$$





Independent leading effects controlling the shape of the matter power spectrum $P(k)$ in the minimal ΛCDM model.

	Effect	**Relevant quantity**	**Parameter**
(P1)	scale of the maximum	$k_{\rm eq}$	$\omega_{\rm m}, \Omega_\Lambda$
(P2)	slope for $k \gg k_{\rm eq}$ and BAO amplitude	$\Omega_{\rm b}/\Omega_{\rm cdm}$	$\omega_{\rm b}, \omega_{\rm m}$
(P3)	BAO scale	$r_{\rm s}(\eta_{\rm drag})$	$\omega_{\rm b}, \omega_{\rm m}$
(P4)	Global amplitude	amplitude of primordial spectrum and duration of ΛD	$A_{\rm s}, \Omega_\Lambda$
(P5)	Global tilt	tilt of primordial spectrum	$n_{\rm s}$

(!!! here later I'll add explanation!!!)

Independent leading effects controlling the shape of the CMB temperature power spectrum C_l in the minimal ΛCDM model.

	Effect	**Relevant quantity**	**Parameter**
(C1)	Peak scale	$\theta_{\rm peak} = \dfrac{\pi}{l_{\rm peak}} \sim \dfrac{d_{\rm s}\|_{\rm dec}}{d_{\rm a}\|_{\rm dec}}$	$\leftarrow \omega_{\rm m}, \omega_{\rm b}$ $\leftarrow \Omega_\Lambda, \omega_{\rm m}$
(C2)	Odd/even peak amplitude ratio	$R\|_{\rm dec}$	$\omega_{\rm b}$
(C3)	Overall peak amplitude	$\frac{a_{\rm dec}}{a_0}$	$\omega_{\rm m}$
(C4)	Damping enveloppe	$\theta_{\rm d} = \dfrac{\pi}{l_{\rm d}} = \dfrac{a_{\rm dec} r_{\rm d}\|_{\rm dec}}{d_{\rm a}\|_{\rm dec}}$	$\leftarrow \omega_{\rm m}, \omega_{\rm b}$ $\leftarrow \Omega_\Lambda, \omega_{\rm m}$
(C5)	Global amplitude	$\mathcal{P}_\mathcal{R}(k_*)$	$A_{\rm s}$
(C6)	Global tilt	$\frac{d \log \mathcal{P}_\mathcal{R}}{d \log k}$	$n_{\rm s}$
(C7)	Additional plateau tilting (LISW)	$\frac{a_\Lambda}{a_0}$	Ω_Λ
(C8)	Amplitude for $l \geq 40$ only	$\tau_{\rm reio}$	$\tau_{\rm reio}$

(C1) Since all peaks in multipole space correspond to the harmonics of a single correlation length in real space, the scale of the peak is controlled (in good approximation) by a single number $l_{\rm peak}$. It depends on the ratio of the sound horizon at decoupling by the angular diameter distance to decoupling. The first quantity depends on the evolution prior to decoupling, and in particular on the expansion history and sound speed. Hence it depends on $\omega_{\rm m}$ (governing the time of equality) and $\omega_{\rm b}$ (governing $c_{\rm s}^2$ as a function of a). The second quantity depends on the expansion and geometry of the universe after decoupling, i.e. on Ω_Λ and H_0, or in our parameter basis Ω_Λ and $\omega_{\rm m}$.

(C2) When studying the SW contribution to the C_l's, we have seen that the asymmetry between the amplitude of odd and even peaks depends on the shift of the zero-point of acoustic oscillations by a term $-R\psi$. The value of the ratio R at decoupling is governed in our parameter basis by $\omega_{\rm b}$.

(C3) A shift in the time of radiation/matter equality affects the amplitude of all peaks for two reasons: it controls the duration of the intermediate stage between equality and decoupling, during which acoustic oscillations are damped by baryonic effects, and the EISW. Both effects go in the same direction. If equality takes place earlier, there is less time for damping (hence all peaks are increased). The metric fluctuations are also less stabilized at decoupling, and the EISW is larger (hence the first peak is increased even more).

(C4) Diffusion damping near the time of recombination controls the envelope of the peaks (the function $\exp[-(l/l_{\rm d})^2]$ suppresses them, starting essentially from the third one). It depends on the ratio of the damping scale at decoupling by the angular diameter distance to decoupling. The first quantity depends on the Thomson scattering rate prior to decoupling, and in particular on $\omega_{\rm m}$ (governing the value of conformal time at equality) and $\omega_{\rm b}$ (governing the ionization fraction as a function of a). The second quantity depends on the expansion and geometry of the universe after decoupling, i.e. on Ω_Λ and H_0, or in our parameter basis Ω_Λ and $\omega_{\rm m}$. The parameter dependence of effects (C1) and (C4) could be thought to be similar: in fact, it is not, because the sound horizon and the diffusion scale depend on very different combinations of $\omega_{\rm m}$ and $\omega_{\rm b}$.

(C5) The global amplitude of the C_l's depends trivially on that of the primordial spectrum, fixed by $A_{\rm s}$.

(C6) The global slope of the C_l's depends trivially on the tilt of the primordial spectrum, fixed by n_s.

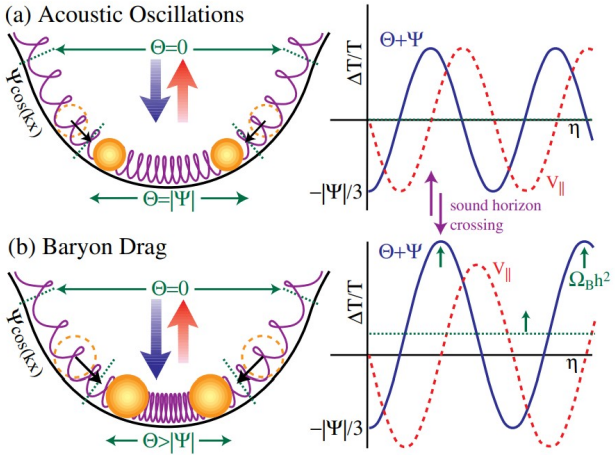
(C7) The LISW effect tilts the Sachs-Wolfe plateau (on top of the effect of n_s). The plateau is more lifted at small l's if Λ domination is longer, i.e. if metric fluctuations decay during a larger amount of time. Hence this effect is enhanced by large values of Ω_Λ.

(C8) The C_l amplitude is suppressed if reionization takes place early, i.e. if $\tau_{\rm reio}$ is large, but without affecting the largest scales (small l's).

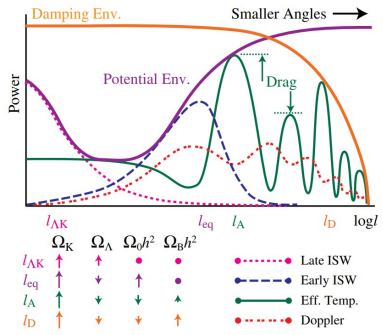
$$D_l := l(l+1)C_l/(2\pi)$$

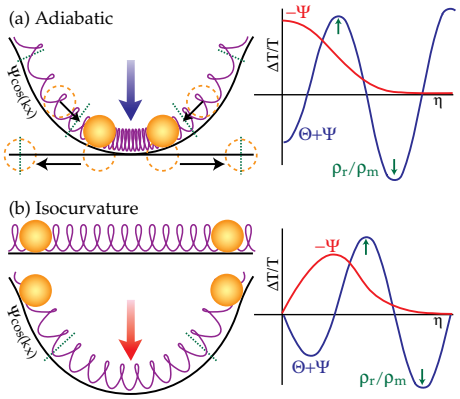
This angular power spectrum is a plot of how much the temperature varies from point to point on the sky (the y-axis variable) vs. the angular frequency ell (the x-axis variable). Ell=10 means that there are ten cycles in the fluctuation around the whole sky, while ell=100 means that there are 100 cycles around the sky.

The acoustic oscillations in the perturbed radiation density are what gives rise to the peaks in the spectrum of CMB anisotropies. (!!! ad eq!!!)



(a) Photon-dominated system. Fluid compression through gravitational infall is resisted by photon pressure setting up acoustic oscillations. Displayed here is a potential well in real space $-\pi/2 \lesssim kx \lesssim \pi/2$. Gravity displaces the zero point so that at the bottom of the well, the temperature is $\Theta_0 = |\Psi| = -\Psi$ at equilibrium with $\Psi/3$ excursions. This displacement is exactly cancelled by the redshift Ψ a photon experiences climbing out from the bottom of the potential well. Velocity oscillations lead to a Doppler effect 90 degrees phase shifted from the temperature perturbation. **(b) Photon-baryon system.** Baryons increase the gravitating mass, causing more infall and a net zero point displacement, even after redshift. Temperature crests (compression) are enhanced over troughs (rarefaction) and velocity contributions. (here from Hu $g_{00} = -[1+2\Psi(\mathbf{x},t)], g_{ij} = (a/a_0)^2\,[1+2\Phi(\mathbf{x},t)]\gamma_{ij}$)





Define the Fourier components of $f(\mathbf{x})$ by

$$f(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}.$$

They satisfy $f_{-\mathbf{k}} = f_{\mathbf{k}}^*$. The simplest inflationary models predict Gaussian perturbations, with probability density formally given by

$$p\left(f_{\mathbf{k}}\right) = \frac{1}{\pi V \sigma_k^2} \exp\left(-\frac{|f_{\mathbf{k}}|^2}{V\sigma_k^2}\right)$$

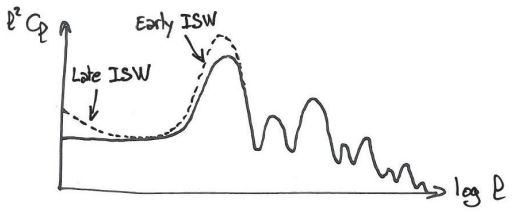
for every Fourier coefficient, where V is the volume of space (which we take to be infinite) and where σ_k only depends on $k = |\mathbf{k}|$ (not to be confused with the parameter k in the spatial curvature!). For the expectation value of the product of two Fourier coefficients, we find

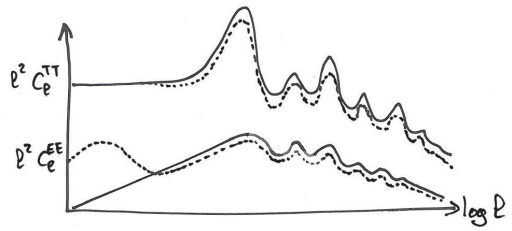
$$\langle f_{\mathbf{k}} f_{\mathbf{k}'} \rangle = V\sigma_k^2 \delta_{\mathbf{k},-\mathbf{k}'} = (2\pi)^3 \sigma_k^2 \delta\left(\mathbf{k}+\mathbf{k}'\right) \equiv (2\pi)^3 P_f(k) \delta\left(\mathbf{k}+\mathbf{k}'\right),$$

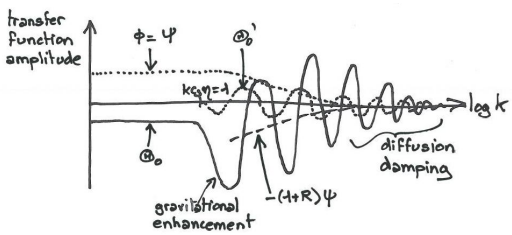
where $P_f(k)$ is the power spectrum.

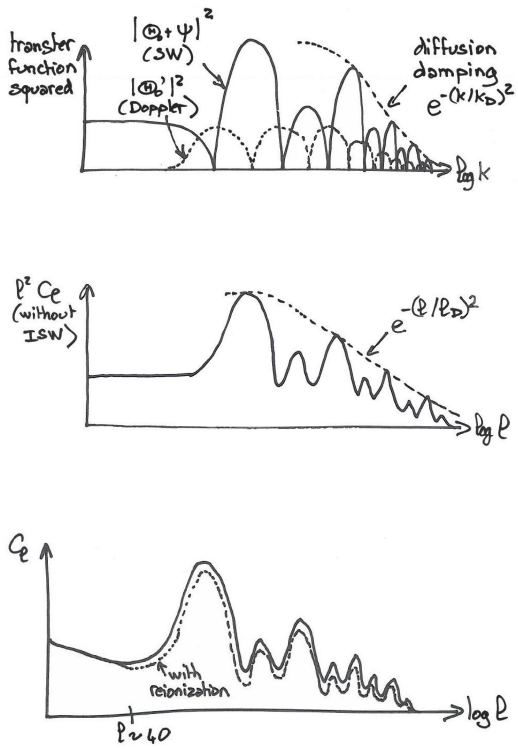
Another way to characterize a Gaussian random field is by its two-point correlation function

$$\xi_f(\mathbf{x},\mathbf{y}) \equiv \langle f(\mathbf{x}) f(\mathbf{y}) \rangle.$$









• It is difficult to measure the full CMB spectrum on the ground due to absorption by water molecules in the atmosphere. The first accurate measurement over all wavelengths was performed by the COBE satellite in 1989, which found that the CMB's spectrum in all directions was very close to that of an ideal blackbody.

• The CMB's temperature was found to be anisotropic due to the Earth's peculiar velocity. By subtracting off Earth's velocity about the Sun, the Sun's velocity in the galaxy, and the galaxy's motion velocity relative to the Local Group, we can measure the peculiar velocity of the Local Group; it is moving quickly towards the nearest supercluster.

• Subtracting off this dipole distortion gives much smaller fluctuations, with

$$\langle T \rangle = 2.725 \text{ K}, \quad (\delta T/T)_{\text{rms}} = 1.1 \times 10^{-5}$$

Thus the CMB is very nearly isotropic, providing strong evidence for the Big Bang.

• The temperature variations can be used to infer the matter distribution on the surface of last scattering, and hence the matter power spectrum at that time. However, this is somewhat subtle. Matter and radiation perturbations are related, since we have adiabatic perturbations, $\delta_r = \frac{4}{3}\delta_m$, and from $\rho \sim t^4$ $\delta_r = 4\frac{\delta T}{T}$, so

$$\frac{\delta T}{T} = \frac{1}{3}\delta_m$$

• We must account for the fact that the photons redshift when they climb out of the gravitational potential. Naively, we would have $\delta T/T = \delta\Phi/c^2$, where Φ is the potential (why????). However, part of this is cancelled by the simultaneous expansion of the universe, and

$$\frac{\delta T}{T} = \frac{\delta\Phi}{3c^2} \text{(Sachs-Wolfe effect)}$$

$$\delta\Phi(\mathbf{k}) = -\frac{4\pi G}{k^2}\bar{\rho}a^2\delta_m(\mathbf{k}) \qquad \text{(Poisson eq, } c = 1\text{)}$$

which means the Sachs-Wolfe effect dominates for modes with $k \lesssim aH$ at last scattering, i.e. superhorizon modes.

$$\frac{\delta T(\hat{\mathbf{n}})}{T} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} a_{lm}Y_{lm}(\theta,\phi), \quad Y_{lm}(\theta,\phi) = N_{lm}e^{im\phi}P_l^m(\cos\theta)$$

N_{lm} - normalization on the sphere.

The spherical harmonics have $O(1)$ values on the sphere, but vary on angular scales $\theta \sim \pi/l$. (???? what also do they quantitatively do???!!!!? how exactly do they work???)

The Sachs-Wolfe effect dominates for $l \lesssim 50$.

For Gaussian perturbations the coefficients a_{lm} are uncorrelated,

$$\langle a_{lm} a^*_{l'm'} \rangle = C_l \delta_{ll'} \delta_{mm'}$$

(the multipole moments C_l have no m-dependence by isotropy).

(????? how do we average????? do we fix some initial point? or where do we fix a radius???)

For two directions with $\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}' = \cos\theta$,

$$\frac{\langle \delta T(\hat{\mathbf{n}}) \delta T(\hat{\mathbf{n}}') \rangle}{T^2} = \sum_{lml'm'} \langle a_{lm} a_{l'm'} \rangle Y_{lm}(0,0) Y_{l'm'}(\theta, 0) = \sum_l \frac{2l+1}{4\pi} C_l P_l(\cos\theta)$$

(we used $P_l^m(1) = \delta_{m0}$ and $N_{l0} = \sqrt{(2l+1)/4\pi}$).

• For some intuition, the functions $P_l(\cos\theta)$ are about 1 for $\theta \lesssim 1/l$, and oscillate rapidly with lower magnitude for $\theta \gtrsim 1/l$. Thus, the temperature correlation function above is parametrically $l_0^2 C_{l_0}$ where $l_0 \sim 1/\theta$. Alternatively, the variance of the temperature goes roughly as $\int l^2 C_l d\log l$.

• In the plot of $l(l+1)C_l$ the error bars are larger at small l - "cosmic variance" effect. It is because there are fewer of the coefficients a_{lm} available to average over.

• For small l, the Sachs-Wolfe effect dominates. By decomposing $\delta\Phi(\mathbf{k})$ into spherical harmonics, one can show that (!!! this is an important relation, I'll derive it soon!!!!!)

$$C_l = \frac{16\pi T^2}{9} \int dk k^2 P_\Phi(k) j_l^2(kr) \qquad (j_l(kr) \text{ - sph. Bessel f.})$$

For the Harrison-Zel'dovich spectrum, this implies that $C_l \sim 1/l(l+1)$, which is indeed observed in the left part of the plot above, where the curve is flat. The slight tilt of the curve can be used to establish $n_s \approx 0.97$, in agreement with direct measurements of the matter power spectrum.

(!!! maybe I'll write what below in section with illustrations!!!)

• At higher l, we see a pattern of peaks and troughs, associated with baryon acoustic oscillations. The first peak at $l \approx 200$ represents an acoustic wave that had time to undergo a single compression before decoupling. Its l value gives us information about the curvature of the universe, and hence tells us that the universe is nearly flat (???? why???).

• Historically, this was known by the 1990s, and at the time, it was regarded as a puzzle because the amounts of matter and radiation did not add up to the critical density. (???? why????)

Thus, CMB observations lent early evidence to dark energy, which was regarded as established by supernova redshift measurements in 1998, though the CMB itself doesn't directly say anything about it.

• The next two peaks represent oscillations that decoupled at the first rarefaction, and the second compression, respectively. (Oscillations that decoupled in between a rarefaction and a compression, i.e. when at uniform density, correspond to the troughs.) These give information about the amount of baryonic and dark matter in the universe. (why??????)

More dark matter lowers all of the peaks, while more baryonic matter enhances the odd-numbered peaks (why??????).

• For high $l \gtrsim 10^3$, the power spectrum is suppressed. This is because photon decoupling takes a finite time, and during this process, photon diffusion wipes out small-scale fluctuations. (why??????)

• Additional information can be extracted from the polarization pattern of the CMB, which is decomposed into E-modes and B-modes, as mentioned earlier. The E-modes arise automatically from Thomson scattering in an inhomogeneous plasma, and are in accordance with the standard theory of inflationary perturbations. The B-modes are expected to be much weaker, and haven't been observed yet, but would give information about the scale of inflation.

Boltzmann equation

$$\frac{d}{d\eta} f = C\left[f, f_e\right]$$

$$f(\eta, \vec{x}, \vec{p}) = \frac{1}{e^{\frac{p}{T(\eta,\vec{x})}} - 1} \ .$$

$$\bar{f}(\eta, p) = \frac{1}{e^{\frac{p}{\bar{T}(\eta)}} - 1} \ , \qquad \delta f(\eta, \vec{x}, \vec{p}) = \frac{d\bar{f}}{d\log p} \ \frac{\delta T(\eta, \vec{x})}{\bar{T}(\eta)} \ .$$

$\Theta(\eta, \vec{x}) \equiv [\delta T(\eta, \vec{x})/\bar{T}(\eta)]$.

$$\frac{d(a\,p)}{d\eta} = -a\,p\,\phi' - a\,\epsilon\,\hat{n}\cdot\vec{\nabla}\psi \ . \qquad \text{(geodesic eq in Newt g)}$$

$$\frac{d\ln(a\,p)}{d\eta} = -\phi' - \hat{n}\cdot\vec{\nabla}\psi \ .$$

$$f(\eta, \vec{x}, \vec{p}) = \frac{1}{e^{\frac{p}{T(\eta,\vec{x},\hat{n})}} - 1} \ . \qquad (\hat{n} \equiv \vec{p}/p)$$

$(\vec{k}\cdot\hat{n}) = k\cos\theta$.

$$\Theta(\eta, \vec{k}, \theta) = \sum_l (-i)^l (2l+1)\Theta_l(\eta, \vec{k}) P_l(\cos\theta) \ .$$

$$\Theta' + \hat{n}\cdot\vec{\nabla}\Theta - \phi' + \hat{n}\cdot\vec{\nabla}\psi = -\Gamma\left(\Theta - \Theta_0 - \hat{n}\cdot\vec{v}_e\right) \ .$$

Γ is the conformal Thomson scattering rate, Θ_0 the temperature monopole (i.e. the average of $\Theta(\eta, \vec{x}, \hat{n})$ over all directions $\hat{n}$), and $\vec{v}_e$ the bulk velocity of electrons, equal to that of baryons due to tight Coulomb interactions.

This equation is illuminating, since it shows that at early times, when photons, electrons and baryons form a tightly-coupled fluid, the fact that Γ is huge forces Θ to evolve in such way that the parenthesis on the right-hand side vanishes. When this is the case, Θ can only have two non-zero component: a monopole Θ_0, and a dipole equal to $\hat{n}\cdot\vec{v}_\mathrm{b}$. The condition on the dipole can be written equivalently as $\theta_\gamma = \theta_\mathrm{b}$.

In the tightly-coupled regime, the dipole component of $\Theta(\eta, \vec{x}, \hat{n})$ is given by $\hat{n}\cdot\vec{v}_\mathrm{b}$, with $\theta_\mathrm{b} \equiv \vec{\nabla}\cdot\vec{v}_\mathrm{b}$. Hence in Fourier space this component reads $i(\hat{n}\cdot\vec{k})k^{-2}\theta_\mathrm{b} = i(\cos\theta)k^{-1}\theta_\mathrm{b}$. This justifies the fact that initial conditions for Θ in Fourier space depend on θ, but not on the two degrees of freedom of $\hat{n}$. As mentioned above, this property is preserved by the isotropic equations of motion, so that at all times we can study Θ as a function of the arguments $(\eta, \vec{k}, \theta)$ instead of $(\eta, \vec{k}, \hat{n})$.

Photon scattering rate

- The free electron fraction $x_e(\eta)$ (????) starts from one at high redshift. It decreases sharply at the recombination time (corresponding to $z \sim 1080$ or $T \sim 0.3$ eV), and freezes at a very small value (due to departure from thermal equilibrium). Around $z \sim 10$, star formation causes a reionization of the universe, and x_e goes up again to one.

- The Thomson scattering rate $\Gamma = \sigma_T a n_e x_e$ evolves like $a^{-2}x_e$. Before recombination, $\Gamma \gg \frac{a'}{a}$, and the universe is opaque. The sudden drop of x_e at recombination renders the universe transparent: $\Gamma \ll \frac{a'}{a}$. Due to the dilution factor coming from n_e, Γ remains much smaller than $\frac{a'}{a}$ even during reionization, and the universe keeps being transparent (this is why despite of reionization, most photons emitted at recombination do not interact anymore, and allow us to observe anisotropies on the last scattering surface).

- The optical depth $\tau(\eta) \equiv \int_{\eta}^{\eta_0} d\eta \Gamma(\eta)$ represents the opacity of the universe at a given time, when seen from today (when $\eta = \eta_0$). It tends to infinity when $\eta \longrightarrow 0$, falls below one at recombination and stabilize at a value of the order of 0.1 between recombination and reionization. After reionization it decreases smoothly and reaches zero today by definition.

- The visibility function $g(\eta) \equiv -\tau' e^{-\tau}$ gives the probability that a CMB photon seen today experienced its last scattering at time η. It starts from negligible values at high redshift (suppressed by the $e^{-\tau}$ factor). It has a narrow spike around the time of recombination, and then it falls again to negligible values due to the smallness of τ' between recombination and reionization. It develops a second smaller and wider spike around reionization. This function shows that most CMB photons did not interact between the last scattering surface and today, while a minority rescattered at reionization. The width of the recombination spike gives an indication on the thickness of the last scattering "surface".

- The diffusion length $\lambda_{\rm d}(\eta)$ is an important quantity for understanding the damping of temperature anisotropies on small scales. At any given time, the comoving mean free-path (mfp) of photons is given by $r_{\rm mfp} = 1/\Gamma(\eta)$ (still in units where $c = 1$). If the photons experience a random walk analogous to Brownian motion in a gas, the comoving distance over which they travel between time $\eta_{\rm ini}$ and η can be approximated by

$$r_{\rm d} \sim \left[\int_{\eta_{\rm ini}}^{\eta_0} d\eta \; \Gamma \; r_{\rm mfp}^2\right]^{1/2} = \left[\int_{\eta_{\rm ini}}^{\eta_0} d\eta \; \Gamma^{-1}\right]^{1/2} .$$

  This physical diffusion length of photons is given in this approximation by $\lambda_{\rm d} \simeq a r_{\rm d}$. It grows quickly from very small to very large scales (comparable to the Hubble scale) around the time of recombination.

Temperature anisotropy in a given direction

$$ds^2 = -(1+2\psi)dt^2 + (1-2\phi)a^2 d\vec{l}^{\,2} ,$$

$$\frac{\delta T}{\bar{T}}(\hat{n}) = \Theta(\eta, \vec{o}, -\hat{n})$$

(since in a direction $\hat{n}$, we see photons traveling towards $-\hat{n}$).

$$\forall \mathcal{F}(\eta, \vec{x}, \hat{n}) \qquad \frac{d}{d\eta}\mathcal{F}(\eta, \vec{x}, \hat{n}) = \mathcal{F}' + \frac{dx_i}{d\eta}\frac{\partial \mathcal{F}}{\partial x_i} + \frac{dn_i}{d\eta}\frac{\partial \mathcal{F}}{\partial n_i} = \mathcal{F}' + \hat{n}\cdot\vec{\nabla}\mathcal{F} . \qquad)O(1), \frac{dx_i}{d\eta} = \hat{n})$$

$$\mathcal{F}(\eta, \vec{x}, \hat{n}) = e^{-\tau(\eta)} \left(\Theta(\eta, \vec{x}, \hat{n}) + \psi(\eta, \vec{x})\right) .$$

$$\frac{d}{d\eta}\left[e^{-\tau}(\Theta + \psi)\right] = e^{-\tau}\left(\Theta' + \psi' + \hat{n}\cdot\vec{\nabla}(\Theta + \psi)\right) - \tau' e^{-\tau}(\Theta + \psi) .$$

Use linearized Boltzmann equation with (??) $\tau' = -\Gamma$

$$\frac{d}{d\eta}\left[e^{-\tau}(\Theta + \psi)\right] = -e^{-\tau}\tau'(\Theta_0 + \psi + \hat{n}\cdot\vec{v}_{\rm b}) + e^{-\tau}(\phi' + \psi') .$$

$$\frac{d}{d\eta}\left[e^{-\tau}(\Theta + \psi)\right] = g\,(\Theta_0 + \psi + \hat{n}\cdot\vec{v}_{\rm b}) + e^{-\tau}(\phi' + \psi') .$$

$$(\Theta+\psi)|_{\rm obs} = \int_{\eta_{\rm ini}}^{\eta_0} d\eta \left[g\left(\Theta_0+\psi+\hat{n}\cdot\vec{v}_{\rm b}\right) + e^{-\tau}(\phi'+\psi') \right] \ ,$$

$|_{\rm obs} := (\eta_0, \vec{o}, \hat{n})$.

In instantaneous decoupling approximation, in which all photons are assumed to decouple precisely at the time $\eta_{\rm dec}$: we can replace the visibility function g by the Dirac function $\delta_D(\eta - \eta_{\rm dec})$, and $e^{-\tau}$ by the Heaviside function $H(\eta - \eta_{\rm dec})$.

$$(\Theta+\psi)|_{\rm obs} = \left(\Theta_0+\psi+\hat{n}\cdot\vec{v}_{\rm b})\right)|_{\rm dec} + \int_{\eta_{\rm dec}}^{\eta_0} d\eta \left(\phi'+\psi'\right) \ ,$$

$|_{\rm dec} := (\eta_{\rm dec}, -r_{\rm dec}\hat{n}, \hat{n})$

$\Theta|_{\rm obs}$ is the temperature anisotropy measured by the observer in the direction $-\hat{n}$, while $\Theta_0|_{\rm dec}$ is the temperature anisotropy in the point of the last scattering surface seen in the same direction. If only these two terms where present, this relation would simply tell us that that the temperature anisotropy seen today in a given direction is equal to the intrinsic anisotropy in the point where the observed photons last scattered.

$\hat{n}\cdot\vec{v}_{\rm b}|_{\rm dec}$ stands for the correction to this temperature coming from by the usual Doppler. Indeed, this correction is caused by the velocity of the baryon-photon fluid (that we assumed to be tightly coupled until $\eta_{\rm dec}$) projected along the line of sight.

Next, we expect a correction from gravitational effects. The redshifting and blueshifting of the photons traveling along gravitational potential fluctuations should affect the observed temperature anisotropy relatively to the intrinsic one. It turns out that if the gravitational potential was constant in time (but, of course, not in space), this effect would be conservative, and would only depend on $\psi|_{\rm obs} - \psi|_{\rm dec}$. This explains the second term on the left-hand and right-hand sides. But if ψ varies in time, the effect is not conservative anymore: intuitively, the amount of blueshifting and redshifting experienced by photons traveling across a potential well do not compensate each other if the gravitational well gets deeper between the time at which the photon enters and leaves the well. This explains the addition term $\int d\eta\, \psi$. A similar effect is caused by dilation effects along the line-of-sight, and contributes like $\int d\eta\, \phi$.

Finally, we can drop the second term on the left-hand side, because this term represents only a tiny *isotropic* correction to the observed anisotropies. It is impossible to measure it with the CMB map only, because it is formally equivalent to a redefinition of the average temperature $\bar{T}$, but only by a tiny amount of the order of $10^{-5}\bar{T}$.

$$\Theta|_{\rm obs} = \underbrace{(\Theta_0+\psi)|_{\rm dec}}_{\rm SW} + \underbrace{\hat{n}\cdot\vec{v}_{\rm b}|_{\rm dec}}_{\rm Doppler} + \underbrace{\int_{\eta_{\rm dec}}^{\eta_0} d\eta \left(\phi'+\psi'\right)}_{\rm ISW} \ .$$

Sachs-Wolfe (SW) term, and includes the intrinsic temperature term Θ_0 and the "gravitational Doppler shift" term ψ at one point on the last scattering surface. The second term is the conventional Doppler term. The ISW term and contains all non-conservative gravitational effects occurring in a universe with non-static metric fluctuations.

Let us first focus on the term $\Theta_0|_{\rm dec}$. We have seen that on super-Hubble scales, temperature anisotropies only have a monopole and a dipole component, related respectively to δ_γ and θ_γ. We can be more precise now. We know from thermodynamics that the local value of the photon density is proportional to the temperature to the power four. Taking the derivative of $\log\rho_\gamma = \log T^4$, we get $\delta_\gamma = 4\,\delta T/\bar{T}$, where on the right-hand side the temperature anisotropy is averaged over all directions $\hat{n}$: hence $\delta_\gamma = 4\Theta_0$.

Let us now focus on the term $\psi|_{\rm dec}$. on super-Hubble scales for adiabatic initial conditions, $\delta_\gamma = \frac{4}{3}\delta_{\rm b}$, decoupling takes place at the beginning of matter domination, for super-Hubble scales and during matter domination,

$-2\phi = -2\psi = \delta_{\rm tot} = \delta_{\rm b} = \delta_{\rm cdm}$.

Hence, $-2\psi = \frac{3}{4}\delta_\gamma$.

So on the last scattering surface,

$$\Theta_0 + \psi = \frac{1}{4}\delta_\gamma + \psi = (-\frac{1}{4}\,2\,\frac{4}{3} + 1)\,\psi = (-\frac{2}{3} + 1)\,\psi = \frac{1}{3}\psi \ .$$

Moreover, on super-Hubble scales, we can neglect the Doppler term, which can be shown to be important only on sub-Hubble scales. Integrated Sachs-Wolfe term plays a small role on large scales, because ϕ and ψ are static during most of the evolution after decoupling. Hence we get an approximation for CMB anisotropies smoothed over small scales:

$$\Theta|_{\rm obs,\,large\,scales} \simeq \frac{1}{3}\psi|_{\rm dec} = -\frac{1}{8}\delta_\gamma|_{\rm dec} \ .$$

The term ψ wins over the term Θ_0, leading to a minus sign in front of δ_γ in the above relation. This means that an over-density on the last scattering surface ($\delta_\gamma > 0$), corresponding to a potential well ($\psi < 0$), leads to a cold spot in the observed map ($\Theta < 0$). Conversely, a hot spot corresponds to an under-density. Hence, due to the "gravitational Doppler shift" effect accounted by the term ψ (often called the Sachs-Wolfe effect), the patterns that we observe on CMB map are inverted with respect to intrinsic fluctuations on the last scattering surface.

Spectrum of temperature anisotropies

$$\frac{\delta T}{\bar{T}}(\hat{n}) = \Theta(\eta_0, \vec{o}, -\hat{n}) = \sum_{lm} a_{lm} Y_{lm}(\hat{n}) \ .$$

$$a_{lm} = (-i)^l \int \frac{d^3\vec{k}}{2\pi^2} Y_{lm}(\hat{k}) \Theta_l(\eta_0, \vec{k}) \ ,$$

$\hat{k} \equiv \vec{k}/k$. In linear perturbation theory and assuming Gaussian initial conditions, both Θ_l and a_{lm} are Gaussian random variables.

$$\langle a_{lm} a^*_{l'm'} \rangle = \delta^K_{ll'} \delta^K_{mm'} \left[\frac{1}{2\pi^2} \int \frac{dk}{k} \Theta_l^2(\eta_0, k) \mathcal{P}_\mathcal{R}(k) \right] \ .$$

$$C_l \equiv \frac{1}{2\pi^2} \int \frac{dk}{k} \Theta_l^2(\eta_0, k) \mathcal{P}_\mathcal{R}(k) \ .$$

In a universe with linear and Gaussian perturbations, the C_l's encode all the information concerning the cosmological model describing our universe that is contained in the CMB temperature map.

Since the theory of cosmological perturbation is stochastic, the a_{lm}'s should be seen as random numbers, and the averaging symbols $\langle ... \rangle$ is meant over many realizations of the theory. In a sense, "a given realization" means "a given universe", and the average holds over many universes, all obeying to the same cosmological model, which is encoded in the spectrum C_l.

However, we observe CMB anisotropies in our universe, i.e. in only one realization. CMB maps allow us to measure a definite value of each a_{lm}. Hence the squares $|a_{lm}|^2$ are not expected to be equal to C_l, even if we postulated the right model: there should be some scattering around C_l. This scattering limits the possibility to find the best theory matching the observations. However we can reduce considerably the scattering by noticing that for any fixed l, the statistical distribution of $|a_{lm}|^2$ is independent of m. Hence, for an ideal full-sky CMB experiment, the best estimator of the underlying C_l's is the average between all observed coefficients $|a_{lm}|^2$ with fixed l,

$$C_l^{obs} := \frac{1}{2l+1} \sum_{-l \le m \le l} |a_{lm}^{obs}|^2 \ .$$

$$\hat{C}_l := \frac{1}{2l+1} \sum_{-l \le m \le l} |a_{lm}|^2 \ .$$

$$\text{Gaussian } a_{lm}\text{'s} \Rightarrow \qquad \langle \hat{C}_l \rangle = C_l; \qquad\qquad \langle (\hat{C}_l - C_l)^2 \rangle = \frac{2}{2l+1} C_l^2 \ .$$

Since for larger l we perform an average over more values of m, the relative scattering decreases. Cosmic variance can be seen as a theoretical error: because of it, we cannot reconstruct the underlying model with infinite precision, even if we have infinitely precise observations.

(???? still don't fully understand the averaging)

Line-of-sight integral in Fourier space

$$\begin{aligned} \Theta_l(\eta_0, k) &= \int_{\eta_{\rm ini}}^{\eta_0} d\eta \ S_T(\eta, k) \ j_l(k(\eta_0 - \eta)) \ , \\ S_T(\eta, k) &\equiv \underbrace{g \, (\Theta_0 + \psi)}_{\rm SW} + \underbrace{\left(g \, k^{-2} \theta_{\rm b} \right)'}_{\rm Doppler} + \underbrace{e^{-\tau} (\phi' + \psi')}_{\rm ISW} \ . \quad \text{(temperature source f-n)} \end{aligned}$$

In instantaneous decoupling approximation ($e^{-\tau} = H \, (\eta - \eta_{\rm dec}) \, ; g(\eta) = \delta_D \, (\eta - \eta_{\rm dec})$) we integrate the Doppler term by part, and get:

$$\begin{aligned} \Theta_l(\eta_0, k) \ \simeq \ & \left[\Theta_0(\eta_{\rm dec}, k) + \psi(\eta_{\rm dec}, k) \right] j_l(k(\eta_0 - \eta_{\rm dec})) \\ & + k^{-1} \theta_{\rm b}(\eta_{\rm dec}, k) \ j_l'(k(\eta_0 - \eta_{\rm dec})) \\ & + \int_{\eta_{\rm dec}}^{\eta_0} d\eta \ \left[\phi'(\eta, k) + \psi'(\eta, k) \right] j_l(k(\eta_0 - \eta)) \end{aligned}$$

(note that in the second line, i.e. in the Doppler term, the prime stands for the derivative of the function $j_l(x)$ with respect to its argument, not with respect to conformal time). This approximate result can be plugged into Eq. (234) to obtain the final spectrum C_l. We see that each C_l can be decomposed into six terms: the power spectrum $C_l^{\rm SW}$ of the SW term, coming from the first line of Eq. (239) squared, that of the Doppler term, coming from the second line of Eq.(239) squared, that of the Integrated Sachs-Wolfe term, coming from the third line of Eq. (239) squared, and finally the three cross-spectra involving each pair of terms.

For large values of l, the spherical Bessel functions $j_l(x)$ and $j_l'(x)$ are very peaked near $x \simeq l$. Hence, for the Sachs-Wolfe and Doppler contributions to the spectrum C_l, the integral over k in Eq. (234) will pick up mainly modes with $k(\eta_0 - \eta_{\rm dec}) \simeq l$. This shows that the SW contribution to C_l is given by the product of the primordial spectrum with the squared transfer function $(\Theta_0 + \psi)$ at a given value of η and k:

$$C_l^{\rm SW} \sim \left[\Theta_0(\eta_{\rm dec}, k) + \psi(\eta_{\rm dec}, k) \right]^2 \mathcal{P}_{\mathcal{R}}(k) \ , \qquad k \simeq \frac{l}{(\eta_0 - \eta_{\rm dec})}$$

(for simplicity, we did not write numerical factors and powers of l or k in front of this expression). In other words, $C_l^{\rm SW}$ depends on the power spectrum of the perturbation $(\Theta_0 + \psi)$, evaluated at the time of decoupling, and for wavenumbers in the vicinity of $k = \frac{l}{(\eta_0 - \eta_{\rm dec})}$,

$$C_l^{\rm SW} \sim \langle |\Theta_0 + \psi|^2 \rangle_{(\eta, k) \simeq (\eta_{\rm dec}, \, l/(\eta_0 - \eta_{\rm dec}))} \ .$$

We reached this result with mathematical arguments, but it has a very simple geometrical interpretation, illustrated in Fig. 139. The spectrum C_l encodes the correlation between structures on CMB maps seen under an angle $\theta = \pi/l$. This angle subtends a given physical scale on the last scattering surface, namely $\theta \times d_{\rm a}(z_{\rm dec})$, where $d_{\rm a}(z)$ is the angular diameter distance to objects of redshift z. Since the Sachs-Wolfe term $(\Theta_0 + \psi)$ contributes to the temperature map only at the time $\eta \simeq \eta_{\rm dec}$, the spectrum $C_l^{\rm SW}$ should depend on the power spectrum of the Sachs-Wolfe term $\langle |\Theta_0 + \psi|^2 \rangle$ at that time, and for a wavenumber such that

$$\frac{\lambda}{2} = \frac{\pi a(\eta_{\rm dec})}{k} = \theta \, d_{\rm a}(z_{\rm dec}) \ .$$

The reason for which λ has been divided by two is that for spherical harmonics, $\theta = \pi/l$ is the angle between a maximum and a minimum, while for a Fourier mode the distance between a maximum and a minimum is one half of the wavelength, $\frac{\lambda}{2} = \frac{\pi a}{k}$. Eq. (242) leads to

$$\frac{a(\eta_{\rm dec})}{k} = \frac{d_{\rm a}(z_{\rm dec})}{l} \ .$$

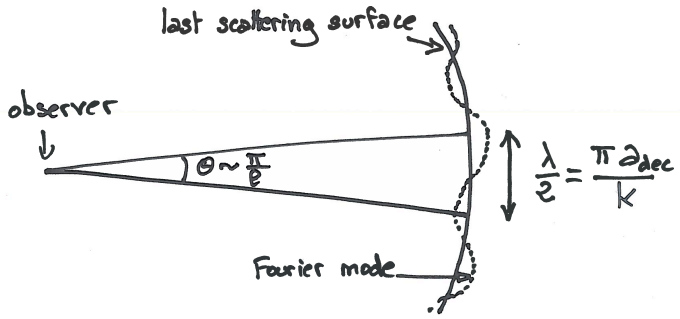


Figure 3: A multipole l refers to structures seen on the last scattering surface under an angle $\theta = \pi/l$. These structures are seeded by Fourier modes with a half wavelength $\frac{\lambda}{2} = \frac{\pi a_{\rm dec}}{k}$.

$$d_{\rm a}(z) := a(t(z)) \int_{t(z)}^{t_0} \frac{dt}{a} \ ,$$

$t(z)$ being the proper time at which an object seen today with a redshift z emitted light. The conformal time $\eta(z)$ is defined similarly. In terms of conformal time,

$$d_{\rm a}(z) = a(\eta(z)) \int_{\eta(z)}^{\eta_0} d\eta = a(\eta(z)) \ [\eta_0 - \eta(z)] \ ,$$

and for the case of a point located on the last scattering surface,

$$d_{\rm a}(z_{\rm dec}) = a(\eta_{\rm dec}) \ (\eta_0 - \eta_{\rm dec}) \ .$$

Hence, Eq. (243) can be written as

$$\frac{1}{k} = \frac{(\eta_0 - \eta_{\rm dec})}{l} \ ,$$

which is the same relation between k and l as in Eqs. (240, 241). In those equations, we implicitly performed a small-angle approximation. For large angles (small l's), it is inaccurate to say that a given angle/mutipole corresponds to a single Fourier mode on the last-scattering surface, and it is important to keep the spherical Bessel function of Eq. (238) and the integral over k of Eq.(234). In summary, Eqs. (240, 241) represent the instantaneous decoupling *and* small-angle limit of the true power spectrum $C_l^{\rm SW}$.

$$C_l^{\rm Doppler} \sim \langle |\theta_{\rm b}|^2 \rangle_{(\eta,k) \simeq (\eta_{\rm dec},\, l/(\eta_0 - \eta_{\rm dec}))} \ ,$$

$$C_l^{\rm ISW} \sim \int_{\eta_{\rm dec}}^{\eta_0} d\eta \, (\eta_0 - \eta) \ \langle |\phi' + \psi'|^2 \rangle_{(\eta,k) \simeq (\eta,\, l/(\eta_0 - \eta))} \ .$$

$$\begin{array}{lcll} \text{SW} & : & \langle|\Theta_0+\psi|^2\rangle & \text{at } (\eta,k)\simeq(\eta_{\rm dec},\ l/(\eta_0-\eta_{\rm dec}))\ , \\ \text{Doppler} & : & \langle|\theta_{\rm b}|^2\rangle & \text{at } (\eta,k)\simeq(\eta_{\rm dec},\ l/(\eta_0-\eta_{\rm dec}))\ , \\ \text{ISW} & : & \langle|\phi'+\psi'|^2\rangle & \text{for all } (\eta,k)\simeq(\eta,\ l/(\eta_0-\eta))\ . \end{array}$$

Before entering into details, we can make a guess. Even if the primordial spectrum $\mathcal{P}_{\mathcal{R}}(k)$ is smooth, the temperature spectrum C_l should contain some structure.

$$d_{\rm s}\equiv a\int_{t_{\rm ini}}^{t}\frac{c_{\rm s}dt}{a}=a\int_{\eta_{\rm ini}}^{\eta}c_{\rm s}d\eta\ , \qquad \text{(sound horison)}$$

$\theta\sim d_{\rm s}(\eta_{\rm dec})/d_{\rm a}(z_{\rm dec})$.
$l\sim\pi/\theta\sim\pi\ d_{\rm a}(z_{\rm dec})/d_{\rm s}(\eta_{\rm dec})$,

Acoustic photon oscillations

An overdense region attracts matter towards it due to gravitational forces, while photon-matter interactions produce a pressure in the opposite direction. Once the overdense region attracts enough matter, the photon-matter interactions become dominant. After a while, after this pressure pushes away enough matter, the gravitational forces kick in again and become dominant. We see that we are in an oscillatory state, in where the counteracting gravity and pressure forces produce sound waves in the early universe.

Consider now a spherical sound wave propagating out from this overdensity. The dark matter stays at the origin of the overdensity since it interacts only through gravitational forces, but the baryons and photons push outwards at some sound speed velocity near half the speed of light (peebles). Once the photons decoupled from matter (at the era of last scattering), the pressure force essentially went to 0 due to the lack of photon-matter interactions. This lack of pressure stopped the baryons from propagating any further outwards, and this distance between the original overdensity site and the effective ring of baryons is called the 'sound horizon'.

During the maximal compression (or rarefaction) in the potential wells, the modes reach an extrema. After recombination, this picture is frozen and those modes will correspond to the peaks of the CMB power spectrum, since those modes will have enhanced temperature.

As long as electrons, baryons and photons are tightly coupled, they form an effective single fluid in which density waves propagate at the sound speed

$$c_{\rm s}^2=\frac{\delta p_\gamma+\delta p_{\rm b}}{\delta\rho_\gamma+\delta\rho_{\rm b}}$$

(the density and pressure of electrons is always negligible with respect to that of photons). The density fluctuation δ_x of each species $x=\gamma,b$ can be inferred from the local value of the equilibrium temperature.

The fact that $\rho_{\rm b}\propto T^3$ and $\rho_\gamma\propto T^4$ implies (by def-s) $\delta_\gamma=\frac{4}{3}\delta_{\rm b}$, and tight coupling imposes $\theta_\gamma=\theta_{\rm b}$.

Using the fact that $|\delta p_{\rm b}|\ll|\delta p_\gamma|$,

$$c_{\rm s}^2=\frac{1}{3(1+R)}\ , \quad R\equiv\frac{4\bar\rho_{\rm b}}{3\bar\rho_\gamma}\propto a\ .$$

In the tightly-coupled regime from Boltzmann and Einstein eq.:

$$\Theta_0''+\frac{R'}{1+R}\Theta_0'+k^2c_{\rm s}^2\Theta_0=-\frac{k^2}{3}\psi+\frac{R'}{1+R}\phi'+\phi''\ .$$

This equation follows from the combination of the continuity and Euler equations for photons and baryons. Given that R is proportional to the scale factor, we could replace R' by $(a'/a)R=aHR$. The second term on the left-hand side is a damping term, increasing with the contribution of baryons to the total energy of the fluid. The third term accounts for pressure forces in the effective fluid. The first term on the right-hand side accounts for the gravitational force, and the last two terms for dilation effects.

$$R = 0, \psi = 0, \phi = 0 \quad \Rightarrow \quad \text{(harm osc)} \quad \Theta_0 = \Theta_{\rm ini}\, \cos(k c_{\rm s} \eta + \varphi) \ ,$$

with two constants of integration $(\Theta_{\rm ini}, \varphi)$. We know that for adiabatic initial conditions and in the Newtonian gauge, photon density/temperature fluctuations should be constant in the super-Hubble limit, $k\eta \ll 1$: this fixes the phase to $\varphi = 0$. In the opposite limit, this solution corresponds to the propagation of acoustic oscillations. Actually, the limit between the constant and oscillatory regime is not set by the value of $k\eta$, but by that of $kc_{\rm s}\eta$. In fact, the condition $kc_{\rm s}\eta \ll 1$ is equivalent to $\lambda \gg d_{\rm s}$, where λ is a physical wavelength ($\lambda = 2\pi a/k$), and $d_{\rm s}$ is the physical sound horizon, given in the case of a constant sound speed by:

$$d_{\rm s} = a \int_{\eta_{\rm ini}}^{\eta} c_{\rm s} d\eta \simeq a c_{\rm s} \eta$$

(assuming $\eta \gg \eta_{\rm ini}$). Hence, the phase $kc_{\rm s}\eta$ of the cosine stands for the ratio $2\pi d_{\rm s}/\lambda$. Modes start oscillating when their wavelength becomes smaller than the sound horizon, and later on, the number of oscillations is given by the ratio between these two scales.

In reality, R grows with time (and crosses one roughly around the time of decoupling). In addition, the gravitational source terms in Eq. (253) can play a role in some regimes. Let us describe qualitatively the evolution of Θ_0 in the different regions in (k, η) space shown in Fig. 141. In this figure, the horizontal axis corresponds to wavenumbers k (large wavelengths are on the left), and the vertical axis to conformal time, flowing from top to bottom. The super-Hubble and sub-Hubble regions are separated by the solid diagonal line corresponding to $k = aH$ (equivalent to $k\eta = 1$ during radiation domination). From top to bottom, the horizontal lines correspond to the time of equality between radiation and matter, to the time of photon decoupling, and to the time today. The upper dashed line separates wavelengths bigger/smaller than the sound horizon in the baryon-photon fluid before decoupling (after decoupling, this notion does not make sense anymore). As we have just seen, this limit corresponds to $\lambda = d_{\rm s}$, or (up to a factor 2π) to $k = (a/d_{\rm s})$. At early times, $c_{\rm s} = 1/\sqrt{3}$, and this condition reads $k\eta = \sqrt{3}$. Just before decoupling, R becomes large, $c_{\rm s}$ goes to zero, and the comoving sound horizon $(d_{\rm s}/a)$ becomes asymptotically constant, explaining the shape of the upper dashed line. Finally, the lower dashed line separates wavelengths bigger/smaller than the diffusion length defined in section 12.9.2: this line corresponds to $k = 1/r_{\rm d}$, where $r_{\rm d}$ is given in first approximation by Eq. (209).

The evolution of Θ_0 in the super-Hubble region is trivial: as long as $k\eta \ll 1$ — and *a fortiori* $kc_{\rm s}\eta \ll 1$ — the fluctuation Θ_0 is frozen, and remains approximately equal to its initial value.

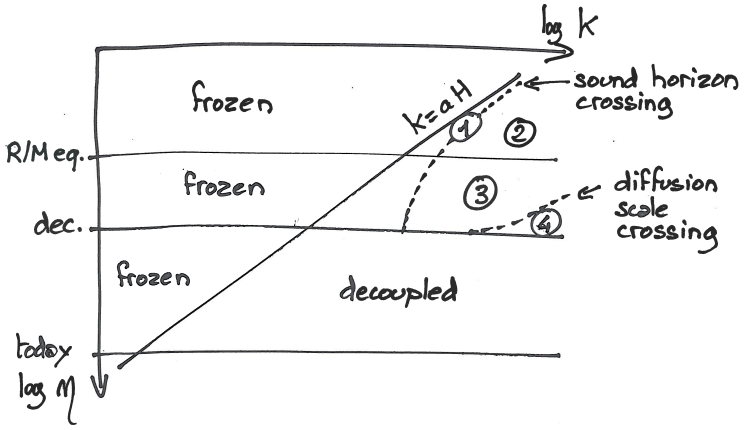


Figure 4: Different regions in (k, η) space, corresponding to qualitatively different behaviors for photon (and baryon) perturbations.

①: in the figure corresponds to modes that are crossing the sound horizon before decoupling. This is precisely the region in which gravitational source terms are important. They shift the

zero point of oscillations, and boost their amplitude (due to gravitational forces and dilution effects). This happens during a limited amount of time, because the metric fluctuations quickly decay inside the sound horizon during radiation domination, making the gravitational source terms negligible. An approximation for the zero-point of oscillations can be found by setting Θ_0'' and Θ_0' to zero in equation (253), and by keeping only the first gravitational term:

$$\Theta_0^{\text{equilibrium}} = -\frac{1}{3c_\text{s}^2}\psi = -(1+R)\psi \ .$$

Since the gravitational potential is non-zero on super-Hubble scales (we have seen that for adiabatic initial conditions $-2\psi = \delta_{\text{tot}}$), the equilibrium point is shifted away from zero on those scales. It reaches asymptotically zero on sub-sound-horizon scales.

②: wavelengths are smaller than the sound horizon during radiation domination. In this regime, the metric fluctuations have decayed, so the source term in Eq. (253) can be neglected. The friction term can also be neglected, because during radiation domination, $R \ll 1$. Finally the effective mass $k^2 c_\text{s}^2$ is constant in time because $R \ll 1$ implies $c_\text{s}^2 = 1/3$. Hence we are in the simple case discussed before, and the solution is proportional to $\cos(kc_\text{s}\eta)$, corresponding to stationary oscillations, symmetric around $\Theta_0 = 0$.

③: wavelengths are smaller than the sound horizon during the intermediate stage between the time of equality and that of photon decoupling. In this region, the metric perturbations have decayed, but R cannot be neglected (baryons and photons contribute to the total energy density with the same order of magnitude). Hence the oscillator equation has a non-negligible friction term (increasing with time), and a time-varying effective mass (decreasing with time). The solution of the equation corresponds to damped oscillations. Physically, this damping is caused by the increasing inertia and decreasing pressure of the baryon-photon fluid when the energy density of non-relativistic baryons takes over.

④: modes with smaller wavelength than the diffusion scale in the photon-baryon fluid. We have defined this scale in section 12.9.2. At early times, in the tightly-coupled limit, the mean free path of particles in the fluid is negligible, and cosmologically interesting scales are all well above the diffusion length. At the approach of decoupling, the diffusion length suddenly increases, and encompasses most of sub-sound-horizon wavelengths. In this regime, the oscillator equaion (253) does not apply anymore, because we cannot describe baryons and photons in terms of a perfect fluid. Perturbations are then strongly damped, since diffusion tends to average out any small-scale fluctuation.

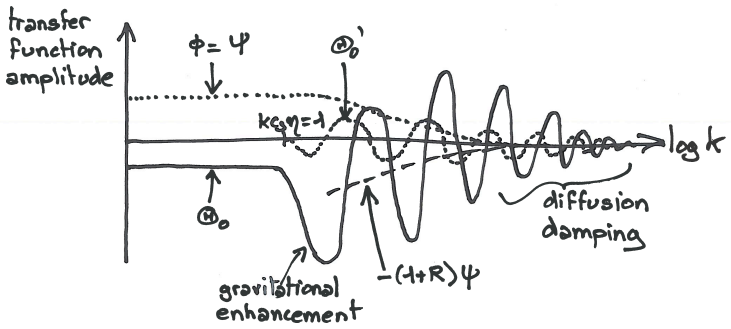


Figure 5: Transfer functions at the time of decoupling.

After describing these different regions, we are ready for understanding the qualitative behavior of the various relevant transfer functions, evaluated at the time of decoupling. Figure 142 shows the transfer functions $\Theta_0(\eta_{\text{dec}}, k)$ (solid line), $\psi(\eta_{\text{dec}}, k)$ (upper dotted line) and $\Theta_0'(\eta_{\text{dec}}, k)$ (middle dotted line) as a function of $\log k$. For simplicity, we neglect the role of the anisotropic stress generated by neutrinos (and to a lesser extent by photons near decoupling time). Hence we can assume $\phi = \psi$ at all times. For scales above the sound horizon ($kc_\text{s}\eta \ll 1$), the transfer functions are constant: indeed we know that, on those scales and in the Newtonian gauge, density and metric fluctuations are frozen. Adiabatic initial conditions impose $-2\psi = \delta_{\text{tot}} \simeq \delta_\text{b} \simeq \frac{3}{4}\delta_\gamma \simeq 3\Theta_0$. The opposite sign of Θ_0 and ψ reflects the fact that an overdensity corresponds to a temperature excess and a gravitational potential well (and vice-versa).

We remember that transfer functions are all normalized to $\mathcal{R}(\vec{k}) = 1$ (see section 12.9.1): this corresponds to a negative δ_γ (and Θ_0) and to a positive ψ, like in the figure.

Because of the decay of metric fluctuations inside the Hubble radius, the ψ curve smoothly decreases and tends towards zero in the small wavelength limit. The behavior of Θ_0 is more complicated. Modes which are just crossing the sound horizon near $\eta = \eta_{\rm dec}$ are experiencing the boost caused by gravitational source terms in Eq. (253): this explains the first bump in the solid line. For smaller wavelengths, we see oscillatory patterns corresponding to acoustic oscillations. Smaller wavelengths crossed the sound horizon earlier, and had more time for oscillating before decoupling. The maxima observed at the time of decoupling correspond to modes that could experience 0.5, 1, 1.5, 2,..., periods of oscillations before that time. The zero point of oscillations follows $-(1+R)\psi$, represented on the figure with a dashed line: this zero point reaches zero well inside the sound horizon. The amplitude of the oscillations is maximal for the first oscillatory pattern, i.e. for modes that crossed the sound horizon very recently. The second oscillatory pattern is reduced by the fact that those modes stayed for a longer time inside the sound horizon during the matter dominated regime, and experienced more damping due to baryons. The third and higher oscillatory patterns are reduced even more by diffusion damping just before photon decoupling. Temperature fluctuations on very small wavelengths are completely suppressed by photon diffusion.

Figure 142 also shows qualitatively the behavior of the time derivative $\Theta_0'(\eta_{\rm dec}, k)$, which exhibits oscillations that are out of phase with respect to those of $\Theta_0(\eta_{\rm dec}, k)$. This will be important in a few paragraphs, when discussing the Doppler effect.

Now that we understand qualitatively the behavior of the metric and photon transfer functions, we can go back to the decomposition of the CMB temperature spectrum C_l in three terms (Sachs-Wolfe, Doppler and integrated Sachs-Wolfe) discussed in section 12.9.2.

Sachs-Wolfe contribution.

We have seen that the Sachs-Wolfe contribution to C_l is approximately given by the power spectrum of the combination $(\Theta_0+\psi)$ at $\eta = \eta_{\rm dec}$, with a correspondence between k and l given by Eq. (247). We know that this power spectrum is given by the product of the primordial spectrum $\mathcal{P}_\mathcal{R}(k)$ (that we can choose to be scale-invariant in first approximation) by the square of the transfer function $(\Theta_0 + \psi)^2$. The qualitative behavior of the latter has no more secrets for us. We can pick up the solid and upper dotted lines in figure 142, add them up, and square the result.

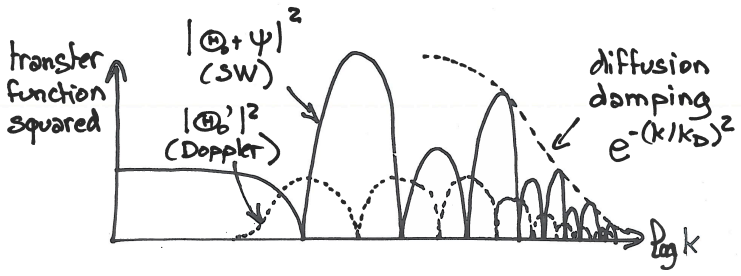


Figure 6: Squared transfer functions at the time of decoupling.

The result is shown in figure 143 as a function of $\log k$. For modes with $kc_s^2\eta \ll 1$, we see a flat plateau: there, our previous calculation of the Sachs-Wolfe effect would apply (this part of the curve is actually called the Sachs-Wolfe plateau). We then observe a series of peaks. Due to the shift of the zero-point of oscillations given by $-(1+R)\psi$ for Θ_0, and hence by $-R\psi$ for the Sachs-Wolfe term $(\Theta_0 + \psi)$, there is an asymmetry between the first few odd and even peaks, with odd peaks being enhanced. Moreover the overall amplitude of the peaks is suppressed in the large k limit by diffusion damping. A bit of algebra would show us that the envelope of the peaks is given in first approximation by the function $\exp[-(k/k_{\rm d})^2]$, where $k_{\rm d}$ is the diffusion wavenumber, related to the diffusion comoving scale of Eq. (209) by $k_{\rm d} r_{\rm d} = 1$.

Doppler contribution.

Next, we know that the C_l's receive a second contribution from the Doppler effect, related to the power spectrum of $\theta_{\rm b}$, which is equal to θ_γ until baryon and photons decouple from each other. It turns out that the photon velocity divergence θ_γ is itself related to the time derivative of the temperature fluctuation Θ_0. At a very qualitative level, we can infer the Doppler contribution from the shape of the transfer function $\Theta_0'(\eta_{\rm dec}, k)$. This contribution is null for scales above the sound horizon, since in this regime there are no oscillations and no significant dynamics in the fluid. On smaller scales, the Doppler contribution has oscillatory patterns, that are out of phase with respect to those of the Sachs-Wolfe term. The Doppler contribution is represented schematically as a dotted line in figure 143.

We can now sum up the Sachs-Wolfe and Doppler contributions. Also, we can transpose our results for power spectra as a function of k in terms of C_l's as a function of l. We have already seen in section 12.9.2 that there is a mapping betwwen the two, at least in the small-angle and instantaneous decoupling approximation, with a correspondence between values of k and l given by Eq. (247).

The result is shows in figure 144. Note that the vertical axis stands for $l^2 C_l$, for the following reason. If the primordial spectrum was scale invariant ($\mathcal{P}_\mathcal{R}(k) =$ constant) and the transfer functions were flat (as it is the case for large wavelengths/small l's), the quantity $l(l+1)C_l$ would also be flat and independent of l. This would follow from Eq. (240) if we had been more carefull in keeping all the factors. It is convenient to plot $l^2 C_l$ or $l(l+1)C_l$ instead of C_l, in order to display a roughly constant curve, just modulated by acoustic oscillations and diffusion damping.

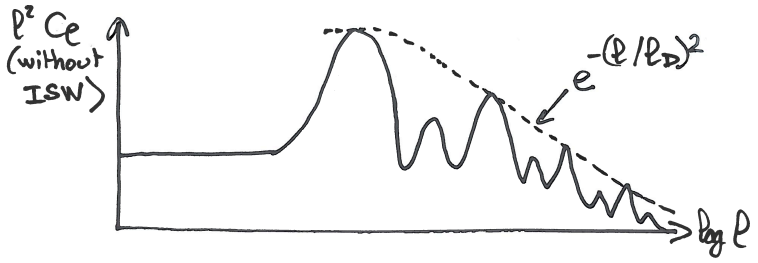


Figure 7: Contribution to the C_l's from the SW and Doppler terms.

Diffusion damping effect.

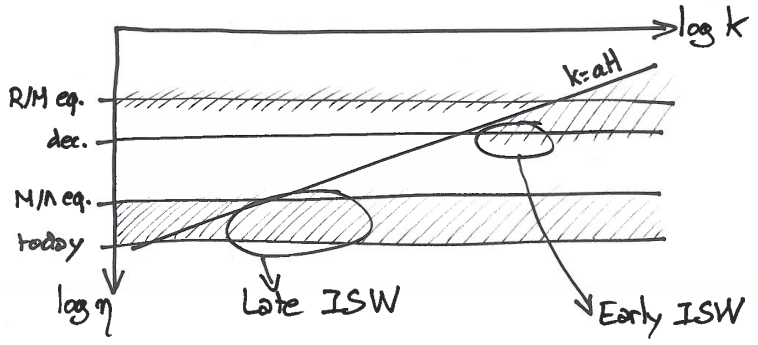


Figure 8: Regions in (k, η) space where metric fluctuations are expected to vary with time (giving rise potentially to an ISW effect after decoupling).

In summary, contribution to the integrated Sachs-Wolfe effect can only come from two regions:

- just after photon decoupling, on sub-sound-horizon scales, and

- during Λ domination, on all scales.

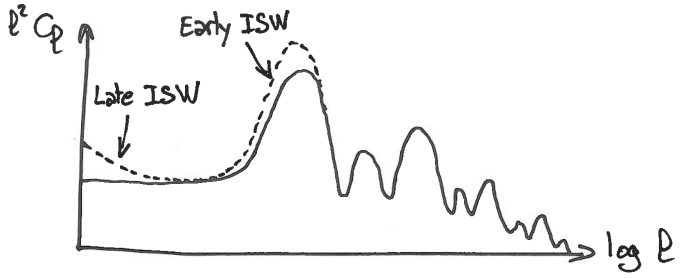


Figure 9: ISW contribution to the temperature spectrum.

Reionization effect.

The last effect that we omitted to describe is that of reionization. We have seen in section 12.9.2 that at small redshift ($z \simeq 10$), the reionization of the universe produces a small secondary bump in the visibility function, corresponding physically to a small probability for CMB photons to rescatter at late times. This rescattering will tend to smooth out any temperature anisotropy pattern. Hence, reionization lowers the overall amplitude of the C_l', but only a small amount (by approximately 15%). Note that the suppression of power is not uniform over the whole multipole range: smoothing effects cannot reach the largest observable scales (corresponding to the smallest values of l). Hence the effect of reionization is step-like shaped, and saturates for l of the order of 40 or so (as illustrated on figure 147).

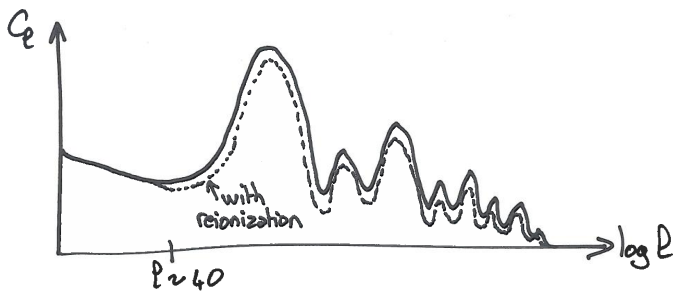


Figure 10: Impact of reionization on the temperature spectrum.

Parameter dependence of the temperature spectrum

$$\mathcal{P}_{\mathcal{R}}(k) = A_{\rm s}(k/k_*)^{n_{\rm s}-1} \ ,$$

$$H_0 \equiv 100\, h\, {\rm km.s^{-1}.Mpc^{-1}}$$

$$\rho_x^0 = \Omega_x \rho_{\rm crit}^0 = \Omega_x \frac{3H_0^2}{8\pi G} = \beta\, \omega_x$$

where $\beta = \frac{3(H_0/h)^2}{8\pi G}$ is a fixed number, with the dimension of an energy per volume.

The six free parameters of the minimal ΛCDM model can be chosen to be

$$\{A_{\rm s}, n_{\rm s}, \omega_{\rm b}, \omega_{\rm m}, \Omega_\Lambda, \tau_{\rm reio}\},$$

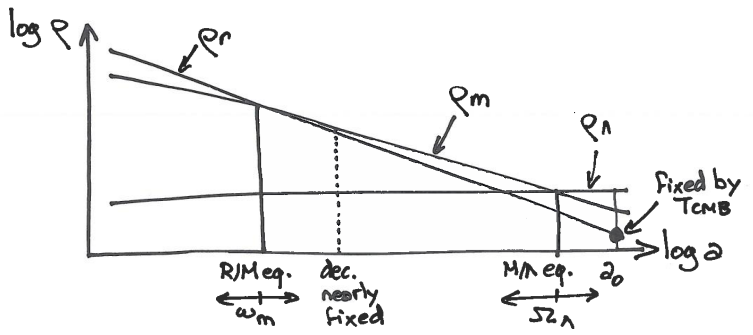


Figure 11: In the minimal ΛCDM model, evolution of the background density of radiation, matter and cosmological constant as a function of the scale factor.

$$\rho_r = \rho_{\rm m} \quad \Longrightarrow \quad \omega_r \left(\frac{a_0}{a_{\rm eq}}\right)^4 = \omega_{\rm m} \left(\frac{a_0}{a_{\rm eq}}\right)^3 \quad \Longrightarrow \quad \frac{a_{\rm eq}}{a_0} = \frac{\omega_r}{\omega_{\rm m}} \,.$$

$$\rho_{\rm m} = \rho_\Lambda \quad \Longrightarrow \quad \omega_{\rm m} \left(\frac{a_0}{a_\Lambda}\right)^3 = \Omega_\Lambda h^2 \quad \Longrightarrow \quad \frac{a_\Lambda}{a_0} = \left(\frac{1-\Omega_\Lambda}{\Omega_\Lambda}\right)^{1/3} \,,$$

A quick word on polarization

The Bolztmann equation presented in Eq. (217) was a bit over-simplified. We did as if the only degree of freedom describing photons with a blackbody spectrum was their temperature. In fact, photons are described by more degrees of freedom, called the Stokes parameters, involving also their polarization. In Eq. (217), the Thomson scattering rate was integrated over polarization parameters, but in reality some polarized correction terms are present.

In general, a vector field can be decomposed into a gradient and a curl component, like the electric and magnetic fields. Similarly, CMB polarization maps can be expanded in two scalar maps called, by analogy, E-type and B-type polarization maps.

One can show that primordial scalar perturbations can produce both temperature anisotropies and E-type polarization anisotropies. If the early universe only features Gaussian scalar perturbations on cosmological scales, all the information contained in CMB maps is encoded in the temperature power spectrum C_l^{TT}, the E-type polarization power spectrum C_l^{EE}, and the cross-correlation power spectrum C_l^{TE}. The later really contains additional independent information, because polarization patterns are only partially correlated with temperature patterns (in mathematical terms, $[C_l^{TE}]^2 \leq C_l^{TT} C_l^{EE}$).

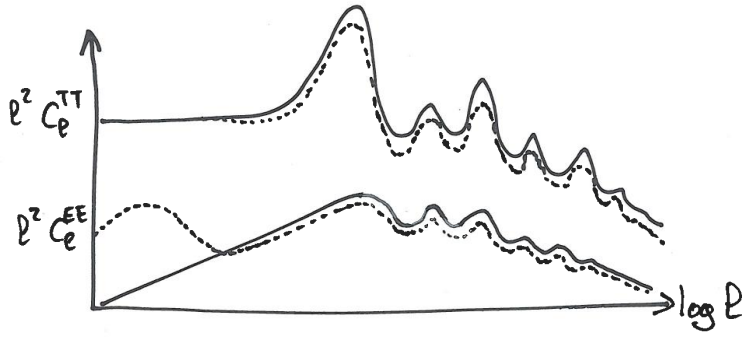


Figure 12: Typical shape of the temperature and E-type polarization power spectra in a ΛCDM model without (solid lines) or with (dotted lines) reionization of the universe at small redshift.

First, reionization has a very distinct effect on the polarization spectrum. The small fraction of photons rescattered around the time of reionization acquire additional polarization that will

show up in the form of a peak in the small-l branch of C_l^{EE} (see figure 149) and C_l^{TE}. We have seen in the previous section that with temperature data only, cosmic variance limits our ability to measure $\tau_{\rm reio}$, and *a fortiori* to constrain additional parameters describing the reionization history. The effect of reionization in the small-l branch of C_l^{EE} and C_l^{TE} is so big that despite of cosmic variance, $\tau_{\rm reio}$ can be well measured using polarization data.

(допишу потом)

A quick word on tensors

We have seen in section 12.9.1 that tensor modes are related to the traceless divergenceless components of the metric and stress-energy tensor, obeying to

$$\sum_i \delta T_{ii} = 0 \qquad \text{and} \qquad \forall j, \quad \sum_i \partial_i \delta T_{ij} = 0 \ .$$

Since these relations impose four constraints on the six degrees of freedom of the 3×3 symmetric matrix δT_{ij} (or δg_{ij}), there are two tensor degrees of freedom in $g_{\mu\nu}$ and $T_{\mu\nu}$.

The two degrees of freedom in δg_{ij} are the two degrees of polarization of gravitational waves, that can propagate even in the vacuum. In the presence of matter with a non-diagonal stress tensor δT_{ij}, gravitational waves can be seeded by the tensor components δT_{ij}.

If tensor perturbations are sufficiently large at recombination or later, they can generate CMB temperature and polarization anisotropies: one can show that they interact with CMB photons, and produce effects similar to the SW and ISW ones. As mentioned in the previous section, gravitational waves can even seed B-type polarization.

(допишу потом)

Cosmological Parameters for anisotropies by Dodelson (???)

8.7.1

Curvature

8.7.2

Degenerate Parameters

8.7.3

Distinct Imprints

Large-Scale Anisotropies by Dodelson

1.3.8 Details of anisotropies, other approaches in a nutshell (!!!!!!!!!)

Acoustic Oscillations by Dodelson (????)

$$\dot\Theta_l + \frac{k}{(-i)^{l+1}} \int_{-1}^{1} \frac{d\mu}{2} \mu \mathcal{P}_l(\mu) \Theta(\mu) = \dot\tau \Theta_l .$$

er terms (e.g., $-\dot\Phi$) have simple μ dependence (sca ments vanish for them. To do the integral, we on for Legendre polynomials, Eq. (C.3), to get

$$\dot\Theta_l - \frac{kl}{2l+1}\Theta_{l-1} + \frac{k(l+1)}{2l+1}\Theta_{l+1} = \dot\tau \Theta_l .$$

(?? слова про смысл перехода)

$$v_{\rm b} \simeq -3i\Theta_1 + \frac{R}{\dot\tau}\left[-3i\dot\Theta_1 - 3i\frac{\dot a}{a}\Theta_1 + ik\Psi\right].$$

$$\ddot\Theta_0 + \frac{\dot a}{a}\frac{R}{1+R}\dot\Theta_0 + k^2c_s^2\Theta_0 = -\frac{k^2}{3}\Psi - \frac{\dot a}{a}\frac{R}{1+R}\dot\Phi - \ddot\Phi \equiv F(k,\eta)$$

where on the right F is a forcing function and $c_s := \sqrt{\frac{1}{3(1+R)}}$ is the sound speed.

$$\left\{\frac{d^2}{d\eta^2} + \frac{\dot R}{1+R}\frac{d}{d\eta} + k^2c_s^2\right\}[\Theta_0+\Phi] = \frac{k^2}{3}\left[\frac{1}{1+R}\Phi - \Psi\right].$$

Tightly Coupled Limit of the Boltzmann Equations (?!!!!)

(!!! короче, додумаю скоро!!!)
In this limit, the two homogeneous solutions (? to what?) are

$$S_1(k,\eta) = \sin\left[kr_s(\eta)\right]; \qquad S_2(k,\eta) = \cos\left[kr_s(\eta)\right]$$

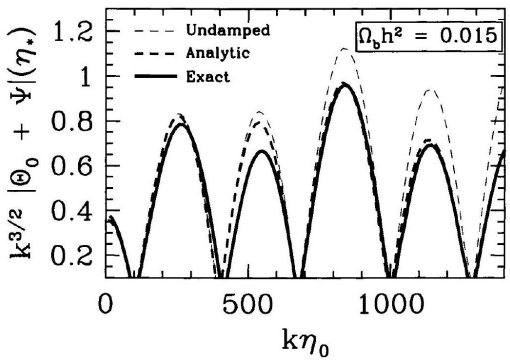
$$r_s(\eta) := \int_0^\eta d\eta' c_s\left(\eta'\right) \qquad \text{(sound horizon),}$$

c_s is the sound speed, the sound horizon is the comoving distance traveled by a sound wave by time η.

The tightly coupled solution for the photon temperature can be constructed from the homogeneous solutions of Eq. (8.21):

$$\Theta_0(\eta) + \Phi(\eta) = C_1S_1(\eta) + C_2S_2(\eta) + \frac{k^2}{3}\int_0^\eta d\eta'\left[\Phi\left(\eta'\right) - \Psi\left(\eta'\right)\right]\frac{S_1\left(\eta'\right)S_2(\eta) - S_1(\eta)S_2\left(\eta'\right)}{S_1\left(\eta'\right)\dot S_2\left(\eta'\right) - \dot S_1\left(\eta'\right)S_2\left(\eta'\right)}$$

$$\Theta_0(\eta) + \Phi(\eta) = \left[\Theta_0(0) + \Phi(0)\right]\cos\left(kr_s\right) + \frac{k}{\sqrt{3}}\int_0^\eta d\eta'\left[\Phi\left(\eta'\right) - \Psi\left(\eta'\right)\right]\sin\left[k\left(r_s(\eta) - r_s\left(\eta'\right)\right)\right]$$



Inhomogeneities to Anisotropies by 8.5.1 Dodelson (!!!)

(хорошая теория, но Лесгоргуес уже это тоже описал, так что потом дочитаю, скорее всего суть этого уже написана)

Free Streaming (8.5.2)

The C_l's

Effects of CMB by Dodelson

Sachs-Wolfe Effect (8.6.2)

(додумаю когда-то точно!)

ISW effect (??????????!!!)

(додумаю когда-то точно!)

Most of the CMB anisotropy is formed at very high redshift where there are no galaxies. However the ISW effect, which involves changes in the gravitational potential giving energy to (or taking energy from) the CMB photons, is active at late times. The temperature perturbation due to ISW is:

$$\Theta(\hat{\mathbf{n}})|_{\rm ISW} = 2\int \dot{\Phi}(r\hat{\mathbf{n}})dr.$$

This is zero in the matter-dominated era because Φ = constant. However the Poisson equation gives:

$$\Phi = -\frac{4\pi G a^2}{k^2}\bar{\rho}_m\delta,$$

and since $\bar{\rho}_m \propto a^{-3}$ and $\delta \propto D(a)$, the potential varies with time in proportion to $D(a)/a$. Thus in a Λ CDM universe, where D grows more slowly than a, the potentials move toward zero. In overdense regions $\dot{\Phi} > 0$, and in underdense regions $\dot{\Phi} < 0$. This leads to a small additional anisotropy in the CMB, but more importantly it leads to a positive correlation of the CMB with the galaxy distribution. The effect is unique to models with Λ or curvature and its detection rules out Einstein-de Sitter.

This galaxy-CMB cross-correlation was first observed by combining the WMAP CMB data with several galaxy surveys (2MASS, APM, SDSS, NVSS) and the Xray background (HEAO). Two combined analyses of the ISW effect with many galaxy samples have been undertaken by Ho et al. (2008) and Giannantonio et al. (2008) with a total detection significance of 4σ. In the matter+curvature

(wiki????)

The "late-time" ISW effect arises quite recently in cosmic history, as dark energy, or the cosmological constant, starts to govern the universe's expansion. Unfortunately, the nomenclature is a bit confusing. Often, "late-time ISW" implicitly refers to the late-time ISW effect to linear/first order in density perturbations. This linear part of the effect entirely vanishes in a flat universe with only matter, but dominates over the higherorder part of the effect in a universe with dark energy. The full nonlinear [disambiguation needed] (linear + higher-order) late-time ISW effect, especially in the case of individual voids and clusters, is sometimes known as the Rees-Sciama effect, since Martin Rees and Dennis Sciama elucidated the following physical picture.[2]

Accelerated expansion due to dark energy causes even strong large-scale potential wells (superclusters) and hills (voids) to decay over the time it takes a photon to travel through them. A photon gets a kick of energy going into a potential well (a supercluster), and it keeps some of that energy after it exits, after the well has been stretched out and shallowed. Similarly, a photon has to expend energy entering a supervoid, but will not get all of it back upon exiting the slightly reduced potential hill.

A signature of the late-time ISW is a non-zero cross-correlation function between the galaxy density (the number of galaxies per square degree) and the temperature of the CMB, [3] because superclusters gently heat photons, while supervoids gently cool them. This correlation has been detected at moderate to high significance. [4][5][6][7][8]

In May 2008, Granett, Neyrinck & Szapudi showed that the late-time ISW can be pinned to discrete supervoids and superclusters identified in the SDSS Luminous Red Galaxy catalog. [9] Their ISW detection traces the localised ISW effect produced by supervoids and superclusters

have on the CMB. However, the amplitude of this localised detection is controversial, as it is significantly larger than the expectations and depends on several assumptions of the analysis.

Small Scales

(8.6.2 dodelson)

1.3.9 Other properties of CMB in a nutshell (!!??)

(тут именно про всякий бэкграунд, как анализировать данные, которые мы видим???? ну и мб чуть про их получение.)

Polarization of CMB

For polarization one can define a similar decomposition:

$$\begin{pmatrix} Q(\hat{\mathbf{p}}) & U(\hat{\mathbf{p}}) \\ U(\hat{\mathbf{p}}) & -Q(\hat{\mathbf{p}}) \end{pmatrix} = \sum_{lm} \left[a^E_{lm} Y^E_{\mathrm{lm}\,\hat{a}\hat{b}}(\hat{\mathbf{p}}) + a^B_{\mathrm{lm}} Y^B_{lm} \hat{a}\hat{\mathbf{p}}) \right].$$

There are separate E-type and B-type power spectra,

$$\left\langle a^{E*}_{lm} a^E_{l'm'} \right\rangle = C^{EE}_l \delta_{ll'} \delta_{mm'}, \qquad \left\langle a^{B*}_{lm} a^B_{l'm'} \right\rangle = C^{BB}_l \delta_{ll'} \delta_{mm'}.$$

Because E has the same parity as temperature, it is also possible to have a temperature-polarization spectrum:

$$\left\langle a^*_{lm} a^E_{l'm'} \right\rangle = C^{\Theta E}_l \delta_{ll'} \delta_{mm'}.$$

The power spectra for the scalars can be determined using the same method as for the temperature, i.e. integrating over wavenumbers and angles to get:

$$C^{EE}_l = 4\pi \int \Delta^2_\zeta(k) \left| \frac{E_l}{\zeta}(k) \right|^2 d\ln k$$
$$C^{\Theta E}_l = 4\pi \int \Delta^2_\zeta(k) \mathrm{Re} \left[\frac{\Theta^*_l}{\zeta}(k) \frac{E_l}{\zeta}(k) \right] d\ln k.$$

(real part and complex conjugate are technically unnecessary for the scalars since Θ_l/ζ and E_l/ζ are real).

General important questions about CMB

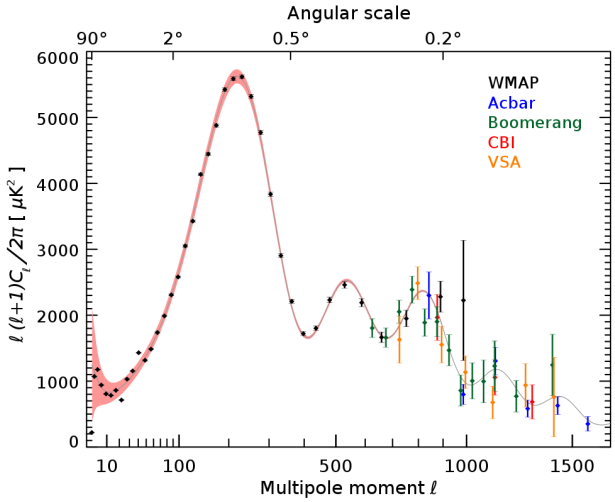
How to obtain the temperature of photons? (!?!?!??!?)

By some distribution (???)
(idk, very interesting question)

Effects of dark matter on CMB and matter spectrum

(see Lesgourgues, not important now)

Observations of CMB (???)



(!!! тут важное описание, того, почему именно такая СМБ, почему именно такие особенности, что по центру чуть ярче???)

(мб картинки добавлю разного анализа СМБ, они есть, просто я все равно их не понимаю, так что когда-то потом займусь ими)

1.4 Mathematics and physical background for applications

1.4.1 Physical background for fluctuations, constants, mathematics

(идея в том, что тут всякая математика и сама собой разумеющаяся физика, потом соберу их)

Constants for cosmology

(I'll add later)

On Gaussian perturbations

A real field $f(\mathbf{x})$ (which could for instance be a density perturbation at the time of recombination) in a region with finite volume V, this field can be Fourier expanded as $f(\mathbf{x}) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}$. Reality of f implies that $f_{-\mathbf{k}} = f^*_{\mathbf{k}}$. If $f_{\mathbf{k}} \equiv a_{\mathbf{k}} + ib_{\mathbf{k}}$, we find $a_{-\mathbf{k}} = a_{\mathbf{k}}$ and $b_{-\mathbf{k}} = -b_{\mathbf{k}}$.

(!?!?!?!?!? why do we want to analyze many regions of CMB sphere??!!)

One way to think about probability distributions is to imagine having a very large number N of such spatial regions, and that the coefficients $a_{\mathbf{k}}$ and $b_{\mathbf{k}}$ can take different values in different regions. In general, a probability distribution function $P\left(\{a_{\mathbf{k}}\}, \{b_{\mathbf{k}}\}\right)$ can be defined by

$$dN = NP\left(\{a'_{\mathbf{k}}\}, \{b'_{\mathbf{k}}\}\right) \prod_{\mathbf{k}} \left(da_{\mathbf{k}} db_{\mathbf{k}}\right),$$

where dN denotes the number of regions with $a'_{\mathbf{k}} < a_{\mathbf{k}} < a'_{\mathbf{k}} + da_{\mathbf{k}}$ and $b'_{\mathbf{k}} < b_{\mathbf{k}} < b'_{\mathbf{k}} + db_{\mathbf{k}}$ for all $\mathbf{k}$. We are interested in situations where the various modes are uncorrelated, meaning that

$$P\left(\{a_{\mathbf{k}}\}, \{b_{\mathbf{k}}\}\right) = \prod_{\mathbf{k}} p\left(a_{\mathbf{k}}, b_{\mathbf{k}}\right).$$

The simplest inflationary models predict homogeneous and isotropic Gaussian distributions,

$$p\left(a_{\mathbf{k}}, b_{\mathbf{k}}\right) = \frac{1}{\pi\sigma_k^2} \exp\left(-\frac{a_{\mathbf{k}}^2}{\sigma_k^2}\right) \exp\left(-\frac{b_{\mathbf{k}}^2}{\sigma_k^2}\right),$$

where σ_k only depends on $k \equiv |\mathbf{k}|$. Now we can compute correlation functions, such as

$$\langle f_{\mathbf{k}} f_{\mathbf{k}'} \rangle = \langle a_{\mathbf{k}} a_{\mathbf{k}'} \rangle - \langle b_{\mathbf{k}} b_{\mathbf{k}'} \rangle + i \langle a_{\mathbf{k}} b_{\mathbf{k}'} \rangle + i \langle b_{\mathbf{k}} a_{\mathbf{k}'} \rangle = \sigma_k^2 \delta_{\mathbf{k},-\mathbf{k}'},$$

(For example, $\langle a_{\mathbf{k}}^2 \rangle = \int\limits_{-\infty}^{\infty} da_{\mathbf{k}} db_{\mathbf{k}} \frac{1}{\pi\sigma_k^2} \exp\left(-\frac{a_{\mathbf{k}}^2}{\sigma_k^2}\right) \exp\left(-\frac{b_{\mathbf{k}}^2}{\sigma_k^2}\right) a_{\mathbf{k}}^2 = \frac{\sigma_k^2}{2}$).

Now, in the infinite volume limit, $V \to \infty$ (????):

$$\sum_{\mathbf{k}} \to \frac{V}{(2\pi)^3} \int d^3\mathbf{k},$$
$$f_{\mathbf{k}}^{\text{old}} \to V^{-1/2} f_{\mathbf{k}}^{\text{new}},$$
$$p_{\text{old}}\ (a_{\mathbf{k}}, b_{\mathbf{k}}) \to V p_{\text{new}}\ (a_{\mathbf{k}}, b_{\mathbf{k}}),$$

so that

$$f(\mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} f_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}},$$
$$p\left(f_{\mathbf{k}}\right) = \frac{1}{\pi V \sigma_k^2} \exp\left(-\frac{|f_{\mathbf{k}}|^2}{V\sigma_k^2}\right),$$
$$\langle f_{\mathbf{k}} f_{\mathbf{k}'} \rangle = V\sigma_k^2 \delta_{\mathbf{k},-\mathbf{k}'} = (2\pi)^3 \sigma_k^2 \delta\left(\mathbf{k} + \mathbf{k}'\right),$$

In the $V \to \infty$ limit the expression for $p\left(f_{\mathbf{k}}\right)$ is only formally defined, but the correlator $\langle f_{\mathbf{k}} f_{\mathbf{k}'} \rangle$ remains well-defined.

(?????? тут важный абзац, физически что вообще мы получаем, когда считаем эти корреляции??? что означает, что две области скореллированы??? что они когда-то были рядом?)

О квантовых полях для космологических возмущений

$$\pi(\mathbf{x}, t) \equiv \frac{\delta L[\phi]}{\delta\dot{\phi}(\mathbf{x}, t)} = \dot{\phi}(\mathbf{x}, t)$$

$$H = \int d^3\mathbf{x} \pi(\mathbf{x}, t)\dot{\phi}(\mathbf{x}, t) - L = \frac{1}{2}\int d^3\mathbf{x} \left[\pi^2 + (\nabla\phi)^2 + m^2\phi^2\right].$$

$$[\hat{\phi}(\mathbf{x}, t), \hat{\pi}(\mathbf{y}, t)] = i\delta(\mathbf{x} - \mathbf{y})$$
$$[\hat{\phi}(\mathbf{x}, t), \hat{\phi}(\mathbf{y}, t)] = 0 \qquad [\hat{\pi}(\mathbf{x}, t), \hat{\pi}(\mathbf{y}, t)] = 0.$$

$$\hat{\phi}(\mathbf{x}, t) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \frac{1}{\sqrt{2\omega_k}} \left[e^{-i\omega_k t} e^{i\mathbf{k}\cdot\mathbf{x}} a_{\mathbf{k}} + e^{i\omega_k t} e^{-i\mathbf{k}\cdot\mathbf{x}} a_{\mathbf{k}}^\dagger\right],$$

$$H = \int d^3\mathbf{k} \frac{\omega_k}{2} \left(a_{\mathbf{k}} a_{\mathbf{k}}^\dagger + a_{\mathbf{k}}^\dagger a_{\mathbf{k}}\right) = \int d^3\mathbf{k}\omega_k \left(a_{\mathbf{k}}^\dagger a_{\mathbf{k}} + \frac{1}{2}\delta^{(3)}(0)\right).$$

$$E_0 = \delta^{(3)}(0)\frac{1}{2}\int d^3\mathbf{k}\omega_k.$$

(что нужно будет - то и допишу)

On Wronskian method

(додумаю когда-то его, возможно, и без него можно все делать)

$$\hat{q}(t) = q(t)\hat{a} + q^*(t)\hat{a}^\dagger,$$
$$W\left[q, q^*\right] \times \left[\hat{a}, \hat{a}^\dagger\right] = 1,$$
$$W\left[q_1, q_2^*\right] \equiv -i\left(q_1 \partial_t q_2^* - \left(\partial_t q_1\right) q_2^*\right)$$

$$W\left[q, q^*\right] \equiv 1,$$
$$\left[\hat{a}, \hat{a}^\dagger\right] = 1.$$

On Framework of Zero Point Fluctuations

$$\left\langle |\hat{q}|^2 \right\rangle = |q(t)|^2 = \frac{\hbar}{2\omega}.$$

(может, что-то допишу еще)

On Framework of choosing a vacuum

(про то, какая тут идея вообще?)

1.4.2 Basic diffgeom and mathematics for cosmology

Основы диффгема для космологии

(в диффгеме очень много всего, там сперва отработаю темы, сюда потом вставлю определения и формулы. Тоже полезно, тоже много вопросов может про это быть)

Однородные пространства по Ландау

(тут крутой диффгем, потом додумаю его, пока просто не нужен был он)

$$\eta_{ad}\eta_{be}\eta_{cf}e^{def} = \eta e_{abc}, \quad e_{abf}e^{cdf} = \delta_a^c\delta_b^d - \delta_a^d\delta_b^c.$$

Уравнения Эйнштейна в синхронной системе выражаются, согласно (97.11)-(97.13), через трехмерные тензоры $\varkappa_{\alpha\beta}$ и $P_{\alpha\beta}$. Для первого имеем просто

$$\varkappa_{(a)(b)} = \dot{\eta}_{ab}, \quad \varkappa_{(a)}^{(b)} = \dot{\eta}_{ac}\eta^{cb}$$

(точка означает дифференцирование по t). Компоненты же $P_{(a)(b)}$ можно выразить через величины η_{ab} и структурные константы группы с помощью (98.14). После замены трехиндексовых $\lambda^a{}_{bc} = C^a{}_{bc}$ на двухиндексовые C^{ab} и ряда преобразований [1]) получим:

$$P_{(a)}^{(b)} = \frac{1}{2\eta}\left\{2C^{bd}C_{ad} + C^{db}C_{ad} + C^{bd}C_{da} - C^d{}_d\left(C^b{}_a + C_a{}^b\right) + \delta_a^b\left[\left(C_d^d\right)^2 - 2C^{df}C_{df}\right]\right\}.$$

Тождества Бианки для трехмерного тензора $P_{\alpha\beta}$ в однородном пространстве:

$$P_b^c C^b{}_{ca} + P_a^c C^b{}_{cb} = 0.$$

Реперные составляющие 4-тензора Риччи (????):

$$R_0^0 = -\frac{1}{2}\dot{\varkappa}_{(a)}^{(a)} - \frac{1}{4}\varkappa_{(a)}^{(b)}\varkappa_{(b)}^{(a)},$$
$$R_{(a)}^0 = -\frac{1}{2}\varkappa_{(b)}^{(c)}\left(C_{ca}^b - \delta_a^b C_{dc}^d\right),$$
$$R_{(a)}^{(b)} = -\frac{1}{2\sqrt{n}}\left(\sqrt{\eta}\varkappa_{(a)}^{(b)}\right) \bullet - P_{(a)}^{(b)}.$$

1.5 Main cosmological parameters and ideas

1.5.1 Hubble parameter and tension in a nutshell

(!! это додумать нужно!!! важный вопрос)

Measurements of Hubble constant

(ничего не знаю, потом подумаю и добавлю инфу. пока из лекций Калтеха)

HST Key Project. The basic idea in this measurement is to use Cepheid variable stars to measure distances to nearby galaxies ($\leq$ 20Mpc). Then use these galaxies to calibrate secondary luminosity standards that are used for more distant galaxies ($>$ 20Mpc). Second step is necessary for the Hubble flow ($v = HD$) to dominate over random velocities of galaxies (typically a few hundred km/s).

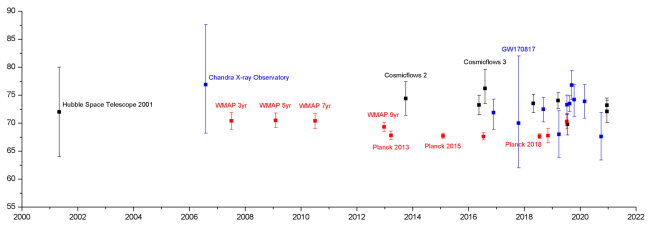
Cepheids are high-mass (typical $\sim 5M_\odot$), luminous (typical $\sim 10^3 L_\odot$) stars near the end of their lives that are unstable against radial pulsations with periods typically of a few days. The radial pulsations are linearly unstable due to the increase of opacity as the star is compressed and He^+ionizes to He^{2+} (traps heat which causes positive work during the expanding part of the cycle). They are stabilized at the nonlinear level which prevents the star from disrupting.

Cepheids have a range of luminosities, so the first step is to calibrate periodluminosity relation for Cepheids. The HST Key Project did this against the Large Magellanic Cloud at distance 50kpc. They find

$$M_V = -2.76 \log_{10} \frac{P}{10\ \mathrm{d}} - 4.22,$$
$$M_I = -2.96 \log_{10} \frac{P}{10\ \mathrm{d}} - 4.90.$$

Does it really exist?

Hubble tension and overview of explanations



1.5.2 Dark Matter in a nutshell

History of DM and Evidence

• In typical galaxies, we infer from the galactic luminosity distribution that the mass is strongly concentrated at the center. So away from the center, stellar velocities should fall off as $v \propto 1/\sqrt{r}$.

• We can measure the function $v(r)$ by measuring redshifts in galaxies tilted relative to us. The result is that $v(r)$ is flat at high distances, or can even increase. Historically, such observations were made in 1939 for the Andromeda galaxy and extended to larger radii in the 1970s by Rubin and Ford. Later, similar measurements were done at radii well beyond the visible disk by measuring 21 cm line emission.

• One should also account for interstellar gas in the calculation of $\rho(r)$, which extends beyond the stars. This can be measured from 21 cm line emission, but in any case is not nearly enough. In addition, there have been decades of searches for baryonic massive compact halo objects (MACHOs) which were just too dim to see, such as cool red and white dwarfs, brown dwarfs, black holes, and rogue planets. But none are present in nearly enough quantity to account for the discrepancy, supporting a nonbaryonic "dark" contribution.

• The curves can be fit by assuming a spherically symmetric DM halo. The halo extends far beyond the galactic disk, with significant density up to order 100kpc from the center. The

density functions $\rho(r)$ can be computed by numerical simulations; at the simplest level, they simply solve the collisionless Boltzmann equation. Some commonly used profiles are

$$\rho(r) = \rho_0 \times \begin{cases} 1/\tilde{r}\,(1+\tilde{r}^2) & \text{Navarro-Frenk-White (NFW)} \\ 1/\tilde{r}^{\gamma}\,(1+\tilde{r}^{3-\gamma}) & \text{generalized NFW} \\ 1/(1+\tilde{r})\,(1+\tilde{r}^2) & \text{Buckert} \\ \exp\left(-2\left(\tilde{r}^{\alpha}-1\right)/\alpha\right) & \text{Einasto} \end{cases}$$

where $\tilde{r} = r/r_s$ and r_s is the scale radius, where $r_s \simeq 20$kpc for our galaxy. Generally, these expressions can't be derived from any analytic model; they empirically fit the results of simulations. Note that a constant $v(r)$ at large radii corresponds to $\rho(r) \propto 1/r^2$.

• Such a wide DM distribution occurs naturally: objects can only gravitationally collapse if they can radiate away energy, while DM only weakly interacts. The collapsing visible matter then becomes a disc, because of the angular momentum barrier, while the DM halo remains spherical.

• We can also infer our local DM density by measuring the local vertical distribution of stars; however, this method is quite noisy and gives a value that could be consistent with zero.

• There are subtleties with modeling galaxies numerically. For example, simulated halo density profiles are often "cuspy" (with a sharp increase at small radii, like the NFW or Einasto profile), but observations suggest "cored" density profiles (flat at small radii, like the Buckert profile).

• However, most DM simulations completely neglected the effects of the baryons, since they are a subdominant fraction of the mass, and have much more complex interactions. But near the center of a galaxy, the baryonic density becomes very high, so they might affect the DM halo.

• Adiabatic contraction. As the baryons slowly collapse, they also pull in the DM (as can be computed using adiabatic invariants). This makes the DM profile more cuspy.

• "Bursty" outflows of baryonic matter, e.g. from supernova and active galactic nuclei, could carry DM outward, coring the profile.

• The DM halo could also be affected by DM self-interactions.

These simulations are difficult, and many parameters must be determined empirically. So far, there's no definitive resolution to the problem, because the uncertainties are large.

• In addition, ΛCDM simulations seem to indicate more small scale structure than is actually observed, specifically more dwarf galaxies bound to the Milky Way; this is called the "too big to fail" or "missing satellites" problem. On the other hand, others argue that this is just observation bias; the satellite galaxies might just be too dim to see, and more are being discovered regularly. Almost all problems with DM are galactic in nature (see this recent review), but galactic dynamics is extraordinarily complex and uncertain.

Note. MOND is an alternative hypothesis that explains galaxy rotation curves well. Here, the gravitational acceleration due to a mass takes the limiting value

$$\lim_{r\to\infty} g = \sqrt{\frac{GMa_0}{r^2}}$$

for a constant a_0. This explains flat rotation curves, because we expect

$$\frac{v^2}{r} = \sqrt{\frac{GMa_0}{r^2}}$$

at large distances, and the factors of r cancel. It also provides a direct explanation of the empirical Tully-Fisher relation $L \propto v^4$, assuming that $M \propto L$. There are several options for g at intermediate values of r which can be adjusted to fit observations at intermediate radii.

In general, MOND seems to do a better job of fitting rotation curves, requiring fewer parameters than galaxy simulations with DM, but completely fails to match evidence from galaxy clusters and the CMB; the only MOND models that work here also use DM. Moreover, galaxy simulations are steadily improving with time, reducing the benefit that MOND provides. This explains the comparatively little interest in MOND: it only gives a minor benefit to the modeling of only one of many aspects of the universe. (However, there are theories of superfluid DM which reduce to MOND on galactic scales. Also, to be fair, one can get the right CMB power spectrum with MOND as long as one puts in a precisely chosen, bumpy initial power spectrum; in this case DM plus inflation has the advantage of working with generic initial

conditions, i.e. being more predictive.) For a deeper look at the debate from a philosophical perspective, see this review.

Evidence for dark matter also comes from observing clusters of galaxies.

• Historically, this was the first evidence for dark matter, proposed in 1933 by Fritz Zwicky to explain observations of the Coma cluster.

• Regarding each galaxy as a point mass, we can show the virial theorem,

$$\ddot{I} = 2T + V, \quad I = \frac{1}{2}\sum_i m_i \mathbf{r}_i \cdot \mathbf{r}_i$$

where I is the virial, T is the total kinetic energy, and V is the total potential energy. In the steady state, $\ddot{I}$ averages to zero and we have

$$T = -\frac{V}{2}$$

which allows us to estimate the mass of a cluster of galaxies from the observed $\langle v^2 \rangle$.

• In practice, we can estimate $\langle v_\perp^2 \rangle$ from redshifts, and multiply by 3 assuming spherical symmetry. We also assume $\rho(r)$ is proportional to the luminosity $L(r)$, allowing us to estimate the average potential energy. Combining these gives a total mass several times higher than the stellar mass inferred by a "census" estimate using luminosity.

• One also needs to know the distance to the cluster. Zwicky did this using the overall Doppler shift and Hubble's law. Though he used an incorrect value for the Hubble constant, overestimating the amount of DM, he did find the right qualitative conclusion.

• The common ground between cluster observations and rotation curves is that they use "tracers", visible collisionless objects that respond to the DM density. Stars are generally excellent tracers because they are very sparse, and hence collisionless.

• DM is also responsible for presence of the hot intracluster gas, which emits X-rays; if DM were not present, most of the gas would be gone. Assuming the gas is in hydrostatic equilibrium and obeys the ideal gas law,

$$\frac{dP}{dr} = -\frac{GM(r)\rho(r)}{r^2}, \quad P = \frac{\rho k T}{m}$$

we find that

$$M(r) = \frac{kT(r)r}{Gm}\left(-\frac{d\log(\rho T)}{d\log r}\right)$$

so that $M(r)$ can be inferred from the functions $T(r)$ and $\rho(r)$, which are measured by X-ray emission. This was done in the 1990s and indicate the same result.

Note. A proof of the virial theorem. We simply differentiate to find

$$\dot{I} = \sum_i \mathbf{r}_i \cdot \mathbf{p}_i, \quad \ddot{I} = \sum_i \mathbf{v}_i \cdot \mathbf{p}_i + \mathbf{r}_i \cdot \mathbf{F}_i$$

The first term in $\ddot{I}$ is simply $2T$. To handle the second term, note that

$$\sum_i \mathbf{r}_i \cdot \mathbf{F}_i = -\sum_{i\neq j} \mathbf{r}_i \cdot \nabla_i V_{ji} = -\sum_{i<j} \mathbf{r}_i \cdot \nabla_i V_{ji} - \sum_{i>j} \mathbf{r}_i \cdot \nabla_i V_{ji}$$

Swapping the index names on the second term, we get

$$-\sum_{i<j} \mathbf{r}_i \cdot \nabla_i V_{ji} + \mathbf{r}_j \cdot \nabla_j V_{ij} = -\sum_{i<j} (\mathbf{r}_i - \mathbf{r}_j) \cdot \nabla_i V_{ji}$$

where we used the fact that $V_{ij} = V_{ji}$ is a function of $\mathbf{r}_i - \mathbf{r}_j$. Now, for any power law potential $V_{ij} \propto (\mathbf{r}_i - \mathbf{r}_j)^n$, the term simplifies using the power rule to $-nV$, giving the result

$$\ddot{I} = 2T - nV.$$

But as long as we work in the CM frame, we expect $\dot{I}$ to be roughly constant over time, giving

$$\langle T \rangle = \frac{n}{2} \langle V \rangle$$

Using the virial theorem, we can also estimate the speed of dark matter in the galactic halo,

$$v \sim \sqrt{\frac{GM_{\text{halo}}}{R_{\text{halo}}}} \sim \sqrt{G\frac{10^{12} M_\odot}{100\text{kpc}}} \sim 200 \text{ km/s}$$

That is, we have $v/c \sim 10^{-3}$.

Note. For generic initial conditions, we estimate the virialization time to be on the order of the gravitational collapse timescale $t \sim 1/\sqrt{G\rho}$, which is on the order of the time it takes the galaxy to rotate; the Milky Way has rotated about 100 times since its formation. There is the possibility of a nonvirialized subcomponent to dark matter. For example, dark matter that has recently been pulled off a satellite galaxy is part of a "stream" with low velocity dispersion.

We can also see evidence for dark matter through gravitational lensing.

• Galaxies and galaxy clusters deflect light. This can be used to measure their masses, giving results in agreement with the observations above.

• Lensing observations can also be used to constrain the nature of dark matter. Baryonic DM could come in the form of MACHOs (massive compact halo objects), i.e. stellar remnants and brown dwarfs. When a MACHO passes exactly between the Earth and a distant star, we see a bright ring around the star. If the object is slightly off center, we instead get a bright arc.

• In practice, these arcs and rings are small enough that we can only detect a momentary brightening of the star. Surveying the sky for these brightening events gives us an estimate of the number of MACHOs in the galaxy. The result is a small fraction of the DM density.

• Gravitational lensing was also used to analyze the Bullet cluster, which was formed by the collision of two large galaxy clusters. It indicated that the interstellar gas, as measured by X-ray spectroscopy, lagged behind most of the gravitational mass, indicating the presence of weakly-interacting DM. (The stars are effectively collisionless and kept going, mostly following the DM.) This was regarded by many as the "nail in the coffin" for MOND.

• However, note that there also are "anti-Bullet cluster" systems, such as the "train wreck cluster" Abell 520, where the DM is in the middle with the interstellar gas; it's not clear how these are to be interpreted, though the data are in some doubt. The Bullet cluster is also unusual, as it has exceptionally high velocity dispersion. Finally, bounds on DM self-interaction from the Bullet cluster are in tension with the self-interaction needed to explain galaxy anomalies.

Finally, cosmological observations give us crucial independent data.

• We can infer the mass density of stars by measuring luminosities, and assuming that a large population of stars has the same mass-to-luminosity ratio as a typical population near us. Additional mass comes from nonluminous stellar remnants, such as white dwarfs, neutron stars, and black holes, as well as brown dwarfs.

• Galaxies also contain significant amounts of interstellar gas, which accounts for 10% of the mass of the Milky Way. For rich clusters of galaxies, the mass of the intergalactic gas can be larger than the mass in the galaxies. This gas is extremely hot and is detected by X-rays using the Sunyaev-Zel'dovich effect.

• By combining all these measurements, a total baryonic density of $\Omega_b = 0.05$ is found, while redshift measurements in the concordance model indicate Ω_m is much larger; in the model the difference is accounted for by DM.

• Stringent independent constraints on Ω_b come from the baryon acoustic oscillation peaks in the CMB, and from the results of BBN, which both yield the same number.

• The DM density also plays an important role in structure formation in the early universe, as we've seen above. Measurements of the matter power spectrum yield yet another independent, consistent value for Ω_m.

• Various modified gravity theories which reduce to MOND in the nonrelativistic limit, such as TeVeS, have also been applied to cosmological scales. However, they currently do a much worse job of fitting the observations than the standard cosmological model. For instance, TeVeS does not produce stable stars.

Models of Dark Matter

• DM candidates can be divided into "cold" and "hot", depending on their typical velocities. Since the 1990s we have known that most, if not all DM must be cold, as hot DM could "freestream" to smooth out small mass fluctuations. Concretely, with cold DM one can first form small structures that merge into larger ones; with hot DM one would first form large structures that fragment into smaller ones. Observations allow only 1% of the DM to be hot.

• One could also consider warm DM, which only suppresses small structures; one could test for this by looking for a cutoff in the matter power spectrum. Other constraints on warm DM come from Lyman α-forest absorption, which provides a map of nearby ($z \sim 5$) hydrogen gas.

• If it's assumed that DM is in thermal equilibrium with the SM thermal bath, then lighter DM particles would automatically be hot. However, there are many nonthermal DM production mechanisms, such as the misalignment mechanism for axions.

• The interactions of DM with the SM are so far consistent with being purely gravitational. Any interactions present must be weak, as they would otherwise contradict, e.g. Bullet cluster observations, and so on. A rough limit for the scattering cross section is $\sigma/m \lesssim (100\text{MeV})^{-3}$. However, adding weak interactions may improve the accuracy of numerical galaxy simulations.

• Decay of DM must be slow. At the simplest level, it must decay slowly enough to still be present today; if the decay is to visible particles, one has stronger constraints from the non-observation of such particles, while if the decay is invisible, one also has stronger constraints due to bounds on how much the DM density can change over time. However, a small amount of decay or annihilation could be used to explain some standing anomalies involves excess detections of energetic SM particles. These should be taken with a grain of salt, because many such anomalies exist, and they frequently vanish.

• DM must have a very small charge, if any at all. The simplest reason is that we can't see it, but stronger constraints come from, e.g. the requirement that the DM be completely decoupled from the thermal bath at recombination. Depending on the mass range, the bound on charge is about $10^{-6}e$. Such a small charge could result from a normal-sized charge under a $U(1)_{A'}$ which weakly kinetically mixes with $U(1)_A$.

• To fit with CMB observations, DM must be present by $z \sim 1000$, eliminating many candidates that form by collapse at late times, such as ordinary black holes. (There aren't enough anyway; they are currently estimated to be about 0.1% of DM.) However, "primordial" black holes, assumed to form early in the universe, are a possibility.

• One generic bound on fermionic DM comes from the exclusion principle. We have

$$\frac{n}{p^3} \sim \frac{\rho_{\text{DM}}/m_{\text{DM}}}{(m_{\text{DM}}v)^3} \lesssim 2$$

which yields $m_{\text{DM}} \gtrsim 1\text{keV}$. This is the Tremaine-Gunn bound. However, bosonic DM can be substantially lighter. On the other hand, composite DM can range up to planetary masses, above which there are lensing bounds, leaving a range of many orders of magnitude in mass.

• The DM does not experience as much dissipation as SM particles, because our galaxy's DM halo seems to be roughly spherical. It's still possible that part of the DM could have collapsed into a "dark disk", and some claim its destabilizing effect could have killed the dinosaurs, though this may have been ruled out with Gaia data.

We now briefly consider some dark matter candidates.

• Many DM candidates fall into one of two classes: light cold bosonic DM (below keV scale) and heavy WIMP-like DM (above MeV scale) produced by freeze-out. Some of these candidates are covered in more detail in the notes on the Standard Model.

• The lightest stable supersymmetric particle, typically a neutralino (but also possibly, e.g. a "mixed sneutrino" or a gravitino), is the canonical example of a WIMP. Many experiments have tried to directly detect WIMPs recoiling on nuclei by the weak force, and have now ruled out most of the "natural" parameter space.

• Neutrinos could account for the DM density if their mass was around 10eV, and they came into thermal equilibrium. This made neutrinos a leading DM candidate in the 1980s, when neutrino oscillations established that they were massless. However, such neutrinos would be hot, and so could not make up the DM by themselves. Today, upper bounds on the neutrino masses from cosmological measurements show that neutrinos are cold DM, but only about 1% of it.

• Sterile neutrinos that don't come into thermal equilibrium could account for the DM density if their mass is on the keV scale; they can be produced via neutrino oscillations in the

DodelsonWidrow or Shi-Fuller mechanisms. Resonant scattering off such sterile neutrinos has been used as an explanation for a deficit of 3.5keV photons emitted from galaxy centers. But they are also strongly constrained, as their decay to normal neutrinos would produce X-rays that haven't been detected.

• Axions are typically extremely light, but can become cold DM because they have a nonthermal production mechanism, the misalignment mechanism. This produces axions in a BEC-like state, with very low velocity dispersion.

• One could also consider axions in a supersymmetric theory, in which case they are part of a chiral supermultiplet. There is another real scalar field, the saxion, and a fermion called the axino. If the axion is the phase of a complex scalar field, the saxion can be thought of as the magnitude. The saxion decays quickly, but the axino can be another WIMP-like DM candidate.

• Some call Primordial black holes (PBHs) the "least speculative" DM candidate, because they don't have nongravitational interactions. But they must be formed by the collapse of large density perturbations in the early universe, which don't appear in the simplest inflation models.

• These perturbations could appear if the inflaton potential had a plateau (the "ultra slow-roll" regime), or in multi-field inflation, where different fields are responsible for the perturbations on small scales and cosmological scales. They could also be formed more violently, such as by the collapse of cosmic strings, bubble collisions from a first-order phase transition, or even by collapse of "dark stars" made of dark sector particles.

• More generally, we could have "macroscopic DM", i.e. DM that comes in discrete chunks, which could have formed through a phase transition or dissipative process. Such objects can have masses ranging from the Planck mass (about a microgram) up to many solar masses. Compared to PBHs, macroscopic DM can be much less compact and doesn't emit Hawking radiation, but it can have nongravitational interactions with the SM.

• Here we overview constraints on ultraheavy DM, including PBHs and macroscopic DM.

• Above $m_{\rm DM} = 10^3 M_\odot$, the gravitational effects of individual DM objects would measurably warp structures by tidal effects.

• The lifetime of a black hole before it evaporates by Hawking radiation is

$$\tau \sim \frac{G^2 M^3}{\hbar c^4} \sim 10^{64} \left(\frac{M}{M_\odot}\right)^3 \text{ yr}$$

Below $m_{\rm DM} \sim 10^{12}$ kg $\sim 10^{-18} M_\odot$, PBHs would have already decayed. (But there is a loophole since extremal BHs are obstructed from decaying.) Black holes above $10^{-7} M_\odot$ aren't evaporating at all; they're actually gaining mass because the CMB is hotter.

• The formation of PBHs requires large density perturbations, which are strongly constrained by the CMB. However, these CMB constraints are only well-measured on large scales, so the small-scale perturbations which would form sub-solar mass PBHs are not constrained. They could be probed in the future by observations of primordial gravitational waves.

• PBHs are also strongly constrained up to $10^{-16} M_\odot$ because of X-rays that would be produced by Hawking radiation, and above $10^{-11} M_\odot$ by lensing constraints. A small window is left in between the two, where PBHs could be all of the DM. (Previously, it was thought much of this range was ruled out by "femtolensing" constraints, but it turns out these are invalidated by finite source size effects.) People also consider cases where PBHs could be a small subcomponent of the DM.

• Macroscopic DM is a bit more complicated, because it is characterized by both a mass and a SM interaction cross section. It is also constrained by lensing above $10^{-10} M_\odot$, and below 10^3 kg or so, there are bounds from terrestrial experiments. But there is a huge "asteroid" range where large interactions with the SM are allowed.

We now list some twists on the collisionless cold dark matter (CCDM) paradigm.

• Warm DM, as mentioned above, smooths out small scale structure.

• Self-interacting DM (SIDM) makes the later stages of structure formation significantly more complicated. Scattering can alleviate the cusp-core problem, and also strips the halos from small clumps of DM orbiting larger structures, reducing their number. It has also been invoked to explain the "train wreck cluster".

• In particle physics, SIDM can also stand for strongly interacting DM. The reason is that the cross sections needed to make self-interacting DM work are of order $\sigma/m \sim \Lambda_{\rm QCD}^{-3}$, which is typical for a particle interacting under the strong force.

• Self-annihilating DM, as mentioned above, can explain standing anomalies and helps alleviate the cusp-core problem, as does decaying DM.

• Fuzzy DM is extremely light bosonic DM with a de Broglie wavelength on the order of the galaxy size; this would provide an additional "poofiness" to DM halos, which could help alleviate the cusp-core problem. Using the uncertainty principle $\Delta x \Delta p \gtrsim 1$ with $\Delta p \sim m_{\rm DM} v$ and Δx the size of a dwarf galaxy halo gives the standard quoted estimate $m_{\rm DM} \gtrsim 10^{-22}$eV. However, over the past ten years, improved data and modeling have pushed up the constraints, which now lie around $m_{\rm DM} \gtrsim 10^{-21}$eV, depending on who you ask.

• Repulsive DM is light bosonic DM which forms a condensate with a repulsive self-interaction, also addressing the cusp-core problem. The idea of "superfluid DM" falls in this category.

• These options are distinguished by their detailed predictions. For example warm DM, fuzzy DM, and repulsive DM all pick out preferred length scales, while decay DM picks out a time scale, and self-interacting and self-annihilating DM pick out densities.

There are many DM production mechanisms beyond freeze-out, and we list some here.

• In the freeze-in scenario, we suppose that the DM is never in thermal equilibrium, but rather has zero abundance in the early universe; the abundance rises monotonically to the desired value throughout cosmological history. Such an initial condition can be achieved if the DM interaction with the SM is very weak, and it is not produced during reheating.

• The freeze-in scenario has different parametrics (the final abundance increases with interaction strength, as opposed to decreasing for freeze-out) and accommodates a wide range of masses. The relevant DM particles are called feebly interacting massive particles (FIMPs).

• Note that if the production rate Γ is set by a dimensionless coupling $g \ll 1$ rather than a suppression by a mass scale, then we have $\Gamma \sim g^2 T$. Since $H \sim T^2$, this means production gets more important later in cosmological evolution. This is nice because the final abundance isn't determined by the reheat temperature, which is free.

• One can modify the freeze-out scenario by considering other DM interactions.

• Coannihilation: there are two DM species, $\chi_i \chi_j \to \phi\phi$, where ϕ is some generic other particle, and χ_i and χ_j are in chemical equilibrium but have different masses. Now the lightest species is the DM, but freeze-out only fixes the cross section for $\chi_i \chi_j$ annihilation, which has absolutely no relation to the cross section for χ_i scattering in WIMP experiments.

• Elastic scattering: $\chi +$ SM $\leftrightarrow \chi +$ SM. This isn't an annihilation process, but it can transfer energy between the DM and SM baths, so its rate can play an important role in some of the more elaborate mechanisms mentioned below.

• Coscattering: $\chi\phi \to \psi\phi$, where the DM converts to another particle in scattering.

• Zombie/pandemic DM: the DM is all converted to another species in reactions like $\chi\phi \to \phi\phi$, which causes the amount of ϕ to increase exponentially.

• Self-annihilation: $n\chi \leftrightarrow \chi\chi$ with $n \geq 3$. These couplings could emerge in a theory with, e.g. χ^3 or χ^4 vertices. (This typically also implies DM self-interactions $\chi\chi \leftrightarrow \chi\chi$, which have astrophysical consequences.) This gives a qualitatively different annihilation rate.

• The above reaction is also called "cannibalization", because for an isolated DM bath, it slows the cooling as rest energy is cannibalized to kinetic energy.

In the standard WIMP freeze-out scenario, self-annihilation isn't present. The abundance is fixed by when the annihilation process decouples. In several newer ideas, the annihilation process is weak and decouples early, while the other processes continue to be effective.

• In strongly interacting massive particle (SIMP) DM, the SM and DM baths are kept in thermal equilibrium by an elastic scattering process, and the DM has a $3 \to 2$ self-annihilation process. Over time, energy is removed from the dark sector by elastic scattering, until the self-annihilation process decouples. The final DM abundance is sensitive to the DM mass and the self-annihilation rate, with typical target masses of $m_{\rm DM} \sim 100$MeV, motivating "light dark matter".

• In elastically decoupling relic (ELDER) DM, the elastic scattering process decouples before selfannihilation. Self-annihilation then keeps the DM bath at approximately a constant temperature for a long period, until it too decouples. It turns out this makes the final DM abundance sensitive to only the elastic scattering cross section, leading to different parametrics.

• Ultralight bosonic DM can be produced nonthermally through the misalignment mechanism, with its abundance thus fixed by initial conditions, or by inflationary fluctuations. Other nonthermal production mechanisms include the decay of heavy particles or topological defects.

• Asymmetric DM is the idea that the relic DM abundance arises from an asymmetry between DM and DM. (For example, in "cogenesis", this asymmetry arises simultaneously with

the baryon-antibaryon symmetry.) This is a tempting idea since the DM and baryon densities are equal up to an $O(1)$ factor. In the simplest version of the idea, $n_{\rm DM}$ and n_B could exactly cancel, with $m_{\rm DM} \approx 5 m_p$.

Example. More about SIMPs, ELDERs, and cannibals, focusing on a $3 \to 2$ annihilation process. (finish)

• The rates for DM annihilation, elastic scattering with the SM, and DM self-annihilation are

$$\Gamma_{\rm ann} \sim n_{\rm DM} \langle \sigma v \rangle_{\rm ann} , \quad \Gamma_{\rm kin} \sim n_{\rm SM} \langle \sigma v \rangle_{\rm kin} , \quad \Gamma_{3\to 2} \sim n_{\rm DM}^2 \left\langle \sigma_{3\to 2} v^2 \right\rangle$$

We can parametrize these rates in terms of couplings, where generically

$$\langle \sigma v \rangle_{\rm kin} \sim \langle \sigma v \rangle_{\rm ann} \equiv \frac{\epsilon^2}{m_{\rm DM}^2}, \quad \left\langle \sigma v^2 \right\rangle_{3\to 2} \equiv \frac{\alpha_{\rm eff}^3}{m_{\rm DM}^5}$$

where the $m_{\rm DM}$ dependence enters by dimensional analysis. We don't have to include powers of T because we have assumed the relevant rates are not dependent on the typical velocity v. This is just the usual assumption of s-wave annihilation for the two-body processes, and its analogue for the three-body annihilation.

• The story of SIMP DM holds as long as

$$\Gamma_{\rm kin} \gtrsim \Gamma_{3\to 2} \gtrsim \Gamma_{\rm ann}$$

until freeze-out. Assuming this is true, freeze-out occurs at

$$\Gamma_{3\to 2} \sim H_F \sim \frac{T_F^2}{M_{\rm pl}}$$

and by similar reasoning to the usual WIMP freeze-out calculation, this means

$$m_{\rm DM} \sim \alpha_{\rm eff} T_{\rm eq}^{2/3} M_{\rm pl}^{1/3}$$

For strong DM self-interactions, $\alpha_{\rm eff} \sim 1$, this points to masses $m_{\rm DM} \sim 100$MeV. The mass scale of the strong force, rather than the weak force, emerges, motivating the name.

• Next, we check the bounds above. After freeze-out, we have

$$\rho_{\rm DM} = m_{\rm DM} n_{\rm DM} \sim T^4 \frac{T_{\rm eq}}{T}$$

since the DM density and radiation density are on the same order at $T_{\rm eq.}$. Then at freeze-out,

$$n_{\rm SM} \sim T_F^3, \quad n_{\rm DM} \sim n_{\rm SM} \frac{T_{\rm eq}}{m_{\rm DM}}$$

which gives

$$\Gamma_{\rm ann} \sim \frac{\epsilon^2 T_F^3}{m_{\rm DM}^2} \frac{T_{\rm eq}}{m_{\rm DM}}, \quad \Gamma_{\rm kin} \sim \frac{\epsilon^2 T_F^3}{m_{\rm DM}^2}, \quad \Gamma_{3\to 2} \sim \frac{\alpha_{\rm eff}^3 T_F^6 T_{\rm eq}^2}{m_{\rm DM}^7}$$

• By using $T_F \sim m_{\rm DM}/20$ and plugging in our expression for $m_{\rm DM}$ above, our bounds become

$$\alpha_{\rm eff}^{1/2} \left(\frac{T_{\rm eq}}{M_{\rm pl}} \right)^{1/3} \lesssim \epsilon \lesssim \alpha_{\rm eff} \left(\frac{T_{\rm eq}}{M_{\rm pl}} \right)^{1/6}$$

For $\alpha_{\rm eff} \sim 1$ this gives a possible range of three orders of magnitude for ϵ.

• Incidentally, if we instead assumed an $n \to 2$ self-annihilation process dominates, then a similar calculation gives $m_{\rm DM} \sim \alpha_{\rm eff} \left(T_{\rm eq}^{n-1} M_{\rm pl} \right)^{1/n}$, so we can further lower the mass.

We now say a bit more about the relatively new paradigm of dark/hidden sectors.

• In the standard WIMP paradigm, we think of the DM as part of a sector that interacts strongly with the SM, such as the superpartners in SUSY. By contrast, a "dark sector" is a group of related particles, one of which is the DM, which interacts only weakly with the SM.

• In principle, a dark sector doesn't have to interact with the SM at all, except gravitationally. This is the worst-case scenario; on the other hand, there is reason for optimism, because many concrete production mechanisms require nongravitational interactions.

• For concreteness, let's suppose that pairs of DM particles couple to a mediator, which in turn couples to pairs of SM particles. In this scenario we can have freeze-out production, where the mediator takes the place of the weak boson for WIMPs.

• The relic abundance scales as $m_{\rm DM}^2/\epsilon^2 g_D^2$, where ϵ is the mediator's coupling to the SM, and g_D is its coupling to DM. Taking g_D to be $O(1)$ and adjusting ϵ, we can accommodate a range of DM masses, from the keV to TeV scale. In SUSY, which in this context is just a concrete way to construct the hidden sector, this was called the "WIMPless miracle". In principle g_D can be adjusted too, but lowering it just increases ϵ, so taking a strong dark sector coupling is a conservative assumption.

• For $m_{\rm DM} \lesssim$ MeV, DM annihilates after neutrinos decouple, which would be detected through $N_{\rm eff.}$. Thus, the window for "light" DM is MeV to GeV.

• Light DM is difficult to search for in direct detection experiments, because such particles deposit little energy in a collision. And since the couplings are low, it is difficult to see the production of such DM in the collider "energy frontier". Instead, for colliders one needs to go to the "intensity frontier", e.g. in beam dump experiments.

• Stepping back, we can ask which interactions between the hidden sector and the SM are "generic". One answer is that it corresponds to the lowest-dimension allowed operators:

• For a vector, there is the dark photon coupling $(\epsilon/2\cos\theta_W)\, B_{\mu\nu}F'^{\mu\nu}$ where F' is the gauge field strength of a dark photon, and B is the $U(1)_Y$ gauge field strength; after electroweak symmetry breaking it is $(\epsilon/2)F_{\mu\nu}F'^{\mu\nu} + (\epsilon/2)\tan\theta_W Z_{\mu\nu}F'^{\mu\nu}$, where $F_{\mu\nu}$ and $Z_{\mu\nu}$ are the photon and Z boson field strengths. The first term is the most important at low energies.

• To understand the phenomenology clearly, we need to refine the fields to remove the kinetic mixing. There are two possibilities. First, we can keep the interactions with other particles the same by setting $A'^{\mu} \to A'^{\mu} + \epsilon A$, in which case we get a mass mixing between the photon and dark photon, allowing them to oscillate into each other. Alternatively, we can remove the mass mixing too, in which case all charged particles pick up a small coupling to the dark photon.

• More general possibilities are also possible with more model building; for example, there are several motivations to gauge $U(1)_{B-L}$: it is non-anomalous in the SM provided sterile neutrinos are introduced, and it can be used to construct R-parity in SUSY models. However, the standard dark photon is distinguished by its minimality.

• There is nothing similar for an axial vector, because the SM has no conserved axial currents; coupling to a nonconserved current would yield issues with unitarity. Of course, a dark axial vector can be accommodated with further model building, which is used for the X17 boson to explain the Beryllium anomaly.

• For a pseudoscalar, there are again no options in the SM. However, if we allow dimension 5 operators, we have the axion couplings $(a/f_a)\, F_{\mu\nu}\tilde{F}^{\mu\nu}$, $(a/f_a)\, G_{\mu\nu}\tilde{G}^{\mu\nu}$, and $(\partial_\mu a/f_a)\, \bar{\psi}\gamma^\mu\gamma^5\psi$. Since the axion is so well-motivated, this is usually included in the standard list.

• For scalars, we have the dark Higgs couplings, $(\mu S + \lambda S^2)\, H^\dagger H$, where the first term isn't allowed for complex scalars. This leads to couplings to SM fermions proportional to their mass. If we allow dimension 5 operators, we can multiply the scalar with any marginal SM operator; in this context, the scalar is often called a "dilaton" in analogy to string theory, since in both cases its value shifts fundamental constants.

• For fermions we have $y_N L H N$, where N is called a sterile neutrino, or heavy neutral lepton. The dark photon, axion, dark Higgs, and sterile neutrino are the "portals". For concreteness, we'll focus on dark photons.

• The dark photon coupling kinetically mixes the SM photon and the dark photon. Upon diagonalizing the kinetic terms with a field redefinition, we can show that the physical degrees of freedom are the familiar photon, along with a massive vector which couples strongly to DM particles, and to SM particles proportional to $\epsilon e Q$.

• The way we would search for such a dark sector depends on the mass of the dark photon.

• If it's in the same "light" mass range as the DM particle, then we could look for its production in a collider experiment. In this case the identity of the DM particle isn't so important, except that it sets a target for the dark photon coupling.

• If it's light, then we can integrate it out. The result is that the DM particles effectively have very small "millicharges". The dark photon itself can also be searched for directly, e.g. through fifth force searches.

• The dark photon itself could be the DM, produced by a nonthermal production mechanism like the axion. This "ultralight" DM could be searched for through precision experiments, such as DM Radio.

• To get a specific target for ϵ, note that if there is a heavy particle that couples to both the photon and dark photon, then it can induce this coupling at one loop. Integrating it out gives $\epsilon \sim 1/16\pi^2 \sim 10^{-3}$, which gives a mass right in the middle of the light DM range.

• A $U(1)_{B-L}$ gauge boson would be searched for in much the same ways as an ordinary dark photon, but the details of the couplings would be different, e.g it would couple equally strongly to neutrons and protons. We could also break family symmetry, e.g. by gauging $U(1)_{L_\mu - L_\tau}$, which would produce no coupling to electrons. On the other hand, these exotic dark photons still pick up a mixing with the SM photon at one loop, albeit a smaller one.

• A similar story holds for a "dark Higgs". In the simplest models, it would couple to SM fermions proportionally to their mass, so it is best probed by the decays of heavy mesons.

• In the standard hidden sector freeze-out scenario, we assume the DM is heavier than the mediator. Forbidden DM is a variant on freeze-out where the DM can only annihilate into heavier particles. In this case, we have $\langle \sigma v \rangle \sim e^{-2\Delta m/T}/m_{\rm DM}^2$ at freeze-out, which can accommodate very light dark matter if Δm is chosen correctly. (To avoid the final particles themselves being a relic, they need to either be SM particles or a decaying hidden sector particle.)

• Another twist is if the DM is heavier than the mediator, but pairs of DM particles couple to pairs of mediator particles. Then DM can annihilate to the mediator, which later decays to SM particles. In this case, the relic density scales as $m_{\rm DM}^2/g_D^4$, where g_D is the mediator-DM coupling. This is called "secluded" annihilation because the extra step means there are much weaker constraints on $m_{\rm DM}$ and g.

• Things could of course be much more complicated. There could be composite objects in the dark sector, resulting from dissipation, phase transitions, or strong interactions. We could have dark atoms, dark molecules, dark mesons and baryons, dark nuclei, or even dark glueballs, leading to a rich phenomenology. As one of infinitely many possible examples, dark pions automatically have a quintic vertex due to the Wess-Zumino-Witten term, providing the self-annihilation process needed for several mechanisms above.

1.5.3 Cosmological constant (dark energy) in a nutshell

More about the discovery of dark energy

Type Ia supernova begin as white dwarfs, which slowly siphon matter from a companion star. When the mass passes the Chandrasekhar limit, a supernova occurs. Since the initial masses of all type Ia supernova are similar, they can be used as standard candles. Strictly speaking, they don't all have the same luminosity, but much of the variance can be accounted for by examining how the luminosity changes over time, in a so-called light-curve shape analysis. Supernova are detected by looking for momentary brightenings of distant galaxies. Their Doppler shift can be used to compute a redshift, while their luminosity can be used to compute a distance. In a uniformly expanding universe, these two are proportional.

Each data point is quite noisy. However, in the 1990s, supernova surveys investigated about 100 supernova at redshifts up to about $z = 1.0$, and observed a deviation from linearity at 5σ, indicating accelerating expansion due to our current period of dark energy dominance. Dark energy can be distinguished from a systematic effect, such as absorption of distant light by dust (i.e. a "tired light" hypothesis), because Λ CDM predicts that for $z \gtrsim 1.0$ we have matter domination, during which the expansion decelerates. However, measurements at such huge distances are extremely difficult, and ongoing. Another line of evidence for dark energy is that the accelerated expansion stunts the growth of galaxies late in the universe, which is indeed observed.

These results remain rather controversial to many. To theorists, dark energy appears difficult to account for in ordinary quantum field theory without tuning, leading to the cosmological constant problem, and it appears difficult to embed into string theory, according to the dS Swampland conjecture. Philosophers are convinced that dark energy is objectionable because of a great variety of vague, metaphysical reasons. Some cosmologists have objected to the original dark energy discovery papers, on the grounds that their analyses overestimated their statistical significance. Finally, late-time measurements of the Hubble constant (including supernova Ia

measurements, but also measurements of quasars) indicate a higher value than that extrapolated from CMB measurements, leading to the "Hubble tension".

Note. None of the discussion above gives direct information about the topology of the universe as a whole. For example, it's possible that the universe is toroidal (i.e. "wraps around") on scales far larger than our Hubble patch. This is scientifically uninteresting, because it produces no observable consequences while making the math more complicated; hence the default assumption is a trivial topology. However, if the topology is on a smaller scale, predictions can be made. In 2003, it was conjectured that the small C_l at $l \lesssim 3$ could be explained by a finite, "Poincare dodecahedral" space, whose volume would be too small for such perturbations to "fit in". This leads to the prediction of "matched circles" in the temperature variations of the CMB, where we see the same part of the surface of last scattering in two different directions; however, a search in 2004 failed to find this. (?? I'll find a ref for it later???)

Main properties of cosmo constant and Vacuum Energy

- In QFT, like the simple harmonic oscillator in ordinary quantum mechanics, there is a zero point energy to the ground state
- For bosons, $\hbar\omega/2 = E(q)/2$, so the most naive version of the cosmological constant problem is that $\rho \propto M^4$ where $M = M_{\rm Pl} = 1/\sqrt{8\pi G}$ if the theory applies out to the Planck scale
- The critical energy density $\rho_c = 3H_0^2/8\pi G \approx 8 \times 10^{-47} h^2 {\rm GeV}^4$ is more than 10^{120} off $M_{\rm Pl}^4 \approx 2 \times 10^{76} {\rm GeV}^4$.
- This is the cosmological constant problem in its basic form - a more sophisticated QFT version is even given renormalization we expect $\rho_{\rm vac} \sim m^4$ for each particle of mass m

Alternatives to Λ

(currently by Hu)
- Alternatives to Λ attempt to have an energy density that is much like vacuum energy in that it remains constant as the universe expands
- For example the potential energy of a field stuck on a hill stays the same even as a grows
- But a field ϕ on a hill can roll converting some of the potential energy V to kinetic energy which does decay with a and make $w > -1$
- Quadratic approximation around minimum of potential: $V = m^2\phi^2/2$ but $m \ll H \sim 10^{-33}$eV.
- Such a field would be a scalar (number at each spacetime position) we'll come back to this possibility in the early universe: inflation
- A light scalar could potentially have exotic long range interactions and coupling to ordinary matter
- Such fifth forces and other modifications to Einstein gravity could also be the explanation of cosmic acceleration

Calculation of cosmo constant in a nutshell

1.5.4 Sense of inflanton field

(все важные вопросы про него тут, отдельное обсуждение)

О настоящем смысле модели инфляций

Что такое инфлятонное поле?

(Проен сказал, что это как раз вакуумная энергия, то есть та самая космологическая постоянная)

Как так, что вакуум расталкивает пространство-время, если у нас инфлятонное поле?

(тоже очень важный вопрос, я как-то пока оставил его, но это вообще я не понимаю.)

1.6 Main experimental facts, their analysis and confirmation of cosmological models

1.6.1 Challenges of Λ-CDM

(по вики, там много фактов, потом обязательно улучшу деление на разделы!!!. важно понимать это для понимания, насколько на самом деле модель рабочая?)

Lack of detection

Violations of the cosmological principle

Violations of isotropy

Violations of homogeneity

El Gordo galaxy cluster collision

KBC void

The KBC void is an immense, comparatively empty region of space containing the Milky Way approximately 2 billion light-years)600 megaparsecs, Mpc) in diameter [55][56][16] Some authors have said the existence of the KBC void violates the assumption that the CMB reflects baryonic density fluctuations at $z = 1100$ or Einstein's theory of general relativity, either of which would violate the $\wedge$ CDM model [57] while other authors have claimed that supervoids as large as the KBC void are consistent with the Λ CDM model. [58]

Hubble tension

S8 tension

Axis of evil

(wiki)
The $\wedge$ CDM model assumes that the data of the cosmic microwave background and our interpretation of the CMB are correct. However, there exists an apparent correlation between the plane of the Solar System, [78] the rotation of galaxies, and certain aspects of the CMB. This may indicate that there is something wrong with the data or the interpretation of the cosmic microwave background used as evidence for the $\wedge$CDM model, or that the Copernican principle and cosmological principle are violated

Cosmological lithium problem

Shape of the universe

Violations of the strong equivalence principle

Cold dark matter discrepancies

Cuspy halo problem

Dwarf galaxy problem

Satellite disk problem

High-velocity galaxy problem

Galaxy morphology problem

Fast galaxy bar problem

Small scale crisis

High redshift galaxies

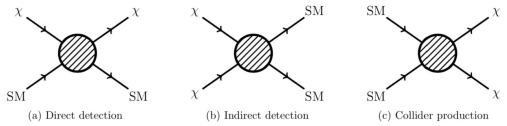
Missing baryon problem

Unfalsifiability

1.6.2 Dark Matter detection

Direct Dark Matter Detection

• The three main paradigms for DM detection are shown below.



(a) Direct detection (b) Indirect detection (c) Collider production

The amplitudes for all three are related by crossing symmetry.

• In a direct detection experiment, one watches for recoils of SM particles (e.g. atomic nuclei) off DM particles. Such experiments are placed deep underground to shield cosmic rays. They were first explored in the mid-1980s by Goodman and Witten. Many experiments have been done since then, placing stringent bounds on WIMP-nucleon cross sections.

• A single experiment, the DAMA collaboration, has seen a significant annual modulation in scattering events, which could be due to the seasonal variation of the Earth's velocity through the DM halo. However, the results are not taken seriously due to possible systematic effects. Since DAMA uses a different material than other experiments, it can be consistent with other experiments if the DM coupling is strongly spin-dependent.

• Astrophysicists typically use indirect detection methods, searching for the produces of annihilating DM. Such signatures include an excess of antimatter, mono-energetic gamma rays, and high-energy neutrinos. The sensitivities of these experiments tend to be complementary to direct direction experiments, e.g. they are more sensitive to higher masses.

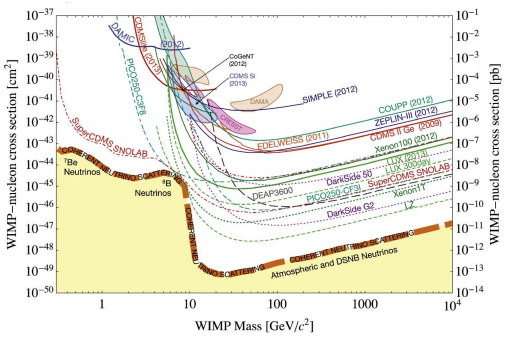
• TeV scale DM could also be produced at the LHC, with the classic signal being missing energy or momentum, such as in the production of a "mono-X" event, where X is a photon, jet, W, Z, h, t or so on. Here, one can take a "top-down" approach, computing results from a full theory such as the MSSM. Alternatively, one could use a simplified model, which gives more generic constraints but which may miss peculiarities that can occur in more complex models. This is done starting from a full model by just removing irrelevant particles, or setting couplings equal or to zero.

• For practicality, LHC results are often phrased in terms of these simplified models, which include the DM particle and the most important mediator. This leaves a reasonable number of parameters, primarily the DM mass, the mediator mass, and, e.g. the mediator's coupling to quarks, which fix the rates for processes such as mono-jets, mediator pair production, or qq scattering via a mediator.

• Various combinations are then possible for the spin and parity of the DM and mediator. For example, a typical figure caption is "axial mediator, Dirac DM, $g_q = 0.25, g_\chi = 1.0$ " with the masses of the mediator and DM on the axes.

Next, we consider direct detection in more detail.

• Standard WIMP direct detection experiments consist of looking for recoils of WIMPs off nuclei. An exclusion plot is shown below.



The basic intuition is that at higher masses, the sensitivity to the cross section falls as 1/m simply because there are fewer DM particles. At lower masses, the sensitivity falls off because the energy deposited by each DM particle is smaller, so not all impacts can be seen; the curve here depends on the details of the high-velocity tail.

• Since the WIMPs would rotate along with the galaxy, there is both a daily modulation of the event rate, due to the spin of the Earth, and a yearly modulation, due to the orbit of the Earth. This is an important way to confirm a WIMP signal, as there are many possible backgrounds, such as radioactivity and cosmic rays. One can also directly measure the WIMP "wind" if the detector is directionally sensitive.

• Suppose a WIMP with mass m_χ and speed v hits a stationary nucleus with mass M. If the nucleus recoils at an angle θ from the original velocity of the WIMP, then some basic classical mechanics gives a recoil energy of

$$E = \frac{2\mu^2 v^2 \cos^2\theta}{M}, \quad \mu = \frac{M m_\chi}{M + m_\chi}$$

The maximum recoil energy has the limits

$$E_{\max} \sim M v^2 \min\left(1, m_\chi^2/M^2\right)$$

For a standard weak-scale WIMP recoiling off a heavy nucleus, the two arguments above are of the same order. Taking $v \sim 10^{-3}$, we have $E_{\max} \sim 10 - 100\text{keV}$.

• Depending on the search, the impact is detected by the resulting ionization, light emitted, calorimetry (i.e. energy deposited into phonons/heat), or a combination of these options. These have varying advantages; for instance, calorimetry reliably gets all the deposited energy, but it is by far the slowest.

• Calculating the differential event rate dR/dE uses input from several fields of physics.

• First, compute the amplitude for WIMPs to scatter off quarks or gluons. This is highly model-dependent.

• Next, use results from nuclear physics to compute the scattering amplitude for protons and neutrons. Usually, we take these to be the same for simplicity; the final experimental bound will get reported in terms of that quantity.

• Infer the amplitude for scattering off the heavy nuclei by summing over the nucleons. At finite momentum transfer, this requires using nuclear form factors, such as the standard simple Helm form factor.

• Compute the scattering rate using the WIMP number density and velocity distribution, e.g. using the standard halo model

$$f(v) = \frac{4}{\sqrt{\pi}} \frac{v^2}{v_0^3} e^{-v^2/v_0^2}$$

This is not exactly true, especially in the high velocity tail; for example, one crude correction would be to set it to zero for velocities greater than the escape velocity of the Milky Way. There is a great variety of halo models, which mostly differ in the tail.

One can substantially change the results by upsetting some of the assumptions made above; this can be used to make the DAMA results consistent with other exclusions.

• An important distinction is between spin-independent and spin-dependent scattering. For spinindependent scattering, in the limit of low momentum transfers, the amplitude contributions of the nucleons add coherently, enhancing the cross section by A^2 where A is the atomic mass number. This works for a decent range of momentum transfers because of the low WIMP velocity. (However, for higher momentum transfers, we must account for a form factor suppression.) The standard WIMP exclusion plots assume spin-independent scattering.

• On the other hand, for spin-dependent scattering, the contributions of nucleons with opposite spins cancel. Since heavy nuclei generally have spin much less than A, this leads to a huge suppression. (DAMA's NaI target is much more sensitive to spin-dependent couplings than other targets, which is why its positive result is not flatly impossible.)

• In general, it's important to have good resolution at low energies, because dR/dE is always monotonically decreasing. It is especially important for experiments detecting light DM, which require light targets and very low energy thresholds.

Example. Suppose the WIMP χ is a Majorana fermion, e.g. the lightest neutralino in a SUSY model. There are two possible leading couplings to quarks: the spin-independent scalar coupling

$$\mathcal{L} \supset \bar{\chi}\chi\bar{q}q$$

which could arise from Higgs exchange, and the spin-dependent axial-vector coupling

$$\mathcal{L} \supset \bar{\chi}\gamma^\mu\gamma_5\chi\bar{q}\gamma_\mu\gamma_5 q$$

which could arise from Z exchange. If χ is a Dirac fermion, we could also have the spin-independent vector coupling

$$\mathcal{L} \supset \bar{\chi}\gamma_\mu\chi\bar{q}\gamma^\mu q.$$

One can extend this discussion to spin 0 or spin 1 WIMPs. Standard WIMP exclusion plots assume a spin-independent scalar coupling.

Indirect Dark Matter Detection

• The standard WIMP miracle calculation fixes the two-body annihilation rate during freeze-out, which can be used to guess the present day annihilation rate. Of course, there are many caveats. Many of the more elaborate DM production mechanisms discussed above completely break this relation: DM relying on $3 \to 2$ self-annihilation might have no $2 \to 2$ annihilation process at all, and coannihilating DM would not have anything to annihilate against today.

• Even if standard $2 \to 2$ annihilation is the dominant process, the annihilation cross section could be effectively different today than during freeze-out, since DM is now more nonrelativistic.

• If s-wave annihilation is forbidden, the leading contribution is usually p-wave annihilation, which has $\langle\sigma v\rangle \propto v^2$, which is much lower today.

• If the DM interacts by a light mediator, its annihilation rate can be higher today due to the Sommerfeld enhancement, a nonperturbative effect due to the exchange of many mediators.

• There could be resonances which enhance the annihilation rate, for appropriate DM velocity.

• We could also consider a one-body decay. It differs qualitatively from annihilation, since it scales only linearly with the density. However, we don't have a solid benchmark for how fast the decay should be.

• One could get rough estimates using an EFT approach with GUTs. For example, if a weak-scale WIMP decays by a dimension 5/6/7 operator from the GUT scale, we would have

$$\tau \sim \begin{cases} m_{\text{GUT}}^2/m_{\text{DM}}^3 \sim 0.1 \text{ s} & \text{dimension 5} \\ m_{\text{GUT}}^4/m_{\text{DM}}^5 \sim 10^{25} \text{ s} & \text{dimension 6} \\ m_{\text{GUT}}^6/m_{\text{DM}}^7 \sim 10^{51} \text{ s} & \text{dimension 7} \end{cases}$$

which are definitely ruled out, possibly observable, and completely unobservable.

• There are a few qualitatively distinct options for the decay/annihilation products. If colored particles or taus are produced, they hadronize and produce many pions, which in turn decay to produce a continuum spectrum of gamma rays. If electrons or muons are produced, we

end up with charged leptons and neutrinos, with only a few photons. Photons can be produced directly, giving a sharp spectral peak, but the branching ratio is expected to be small. Neutrinos can also be produced, but they're very hard to detect.

• Charged particles get strongly deflected by galactic magnetic fields, so experiments looking for such particles are effectively direction blind. Photons and neutrinos aren't deflected, but the detection rates are generically expected to be lower.

• Historically, many indirect detection experiments have reported anomalies which could be explained by DM, but then have been explained by astrophysical effects. This is especially difficult for charged particles, since they could be produced by essentially any nearby new astrophysical phenomenon, such as an unseen population of pulsars. For the uncharged particles, these sources can be localized and ruled out.

Now we focus on detection of uncharged particles.

• Consider DM annihilation in a volume element dV at a distance R from the observer. If the spectrum produced per annihilation is dN/dE, then the spectrum produced per time is

$$\frac{1}{2}\frac{\rho^2}{m_{\rm DM}^2}\langle\sigma v\rangle\frac{dN}{dE}dV$$

where ρ is the DM density and the $1/2$ avoids overcounting. Alternatively, if DM comes in particles and antiparticles symmetrically, then we lose the $1/2$, but we convert ρ^2 to $(\rho/2)^2$, making the result $1/2$ again as large.

• The spectrum per unit time incident on a detector per unit area dA is found by dividing by $4\pi R^2$ and integrating over dV. This quantity splits neatly into a "particle physics" part and "astrophysics" part,

$$\frac{dN_{\rm obs}}{dEdtdA} = \frac{1}{8\pi m_{\rm DM}^2}\langle\sigma v\rangle\frac{dN}{dE}\int\rho^2 dRd\Omega$$

where the integral is called the J-factor. For decaying DM, a similar argument gives

$$\frac{dN_{\rm obs}}{dEdtdA} = \frac{1}{4\pi m_{\rm DM}}\frac{1}{\tau}\frac{dN}{dE}\int\rho dRd\Omega$$

where the integral is called the D-factor. Note, however, that conventions differ on where to put the factors of $1/8\pi$ or $1/4\pi$, and whether to include the angular integral in the J or D-factor.

• Here we have assumed that the cross-section does not depend too strongly on the velocity; otherwise we have to put the $\langle\sigma v\rangle$ under the integral. Also, if we look at distant targets, one should also account for redshift; the line of sight integral dR should be recast as a dz integral.

• Computations of the J-factor depend on the halo model used, especially whether it is cuspy or cored. This makes a significant difference for annihilation because scales quadratically with the density, and hence dominates in the galactic center region. One can estimate local DM densities using oscillations of stars above and below the galactic plane, but this is difficult at the galactic center because one must subtract off a large baryonic matter background.

• For concreteness, plugging in typical WIMP numbers with a cuspy NFW profile gives

$$\frac{dN_{\rm obs}}{dEdt} \sim 10^{-9}\ {\rm cm}^{-2}\ {\rm s}^{-1}$$

for the galactic center. We need about a year with a 100 cm^2 detector to see one photon.

• There are several places that one could look, with various advantages and disadvantages.

• Dwarf galaxies are nearby and have low background, and thus often give the strongest constraints on annihilation.

• The galactic center has high signal, giving it a high potential for discovery. However, it has high background, and the J-factor is uncertain.

• The galactic halo is large and nearby, but has complex backgrounds.

• Other galaxies and galaxy clusters carry a huge amount of DM and also hold redshift information, but are further away. Since they contain a huge total amount of DM, they often give the strongest constraints on decay. However, the J-factor is uncertainty since it depends on DM substructure in the cluster.

• Seemingly empty space could produce a signal, depending on the DM substructure, but it's hard to know where to look.

• Extragalactic background radiation also carries substantial redshift information and averages over a huge region.

• For gamma rays, there are a variety of experiments providing exclusion bounds.

• From 100MeV to 1TeV: Fermi and DAMPE. These are in space, with a relatively small area but wide field of view.

• From 100GeV to 10TeV: HESS, VERITAS, MAGIC. These are ground-based telescopes, with a narrow field of view.

• From 1TeV to 100TeV: HAWC, and later CTA and LHAASO. These ground-based apparatuses have a wide field of view, since they search for the Cherenkov radiation produced when the gamma rays pass through a tank of water.

• For neutrinos, there's also a variety of experiments.

• SuperK probes neutrinos from a few MeV to 1TeV.

• ANTARES reaches 100GeV to 100TeV.

• IceCube reaches 100GeV to 10^9GeV.

• The strongest constraints on GeV-scale DM that annihilates to photon-rich channels come from Fermi looking at gamma rays from dwarf galaxies. The analysis is done by fitting "templates" (i.e. Poisson distributions due to background and candidate signal) using maximum likelihood. A combined analysis rules out thermal relic WIMP DM for masses up to ~ 100GeV, subject to standard assumptions, though the precise bound depends on the annihilation channel.

• Constraints on decaying DM can can probe DM masses up to 10^{10}GeV, or even Hawking radiation from primordial black holes. We can also cover masses down to the keV range using X-ray telescopes such as Chandra, XMM-Newton, INTEGRAL, HXMT, Astrosat, NuSTAR, and Suzaku, or the proposed XRISM, XPoSat, STROBE-X, Athena, and Lynx.

• According to Fermi data, there is a highly significant excess of GeV-scale photons from the galactic center, first noticed by Goodenough and Hooper in 2009. Particle explanations of this excess would have masses in the range $7 - 10$GeV and are generally called "hooperons", but the excess could also be explained by a large population of pulsars. Emission from pulsars would be more "clumpy", but whether the data reflects this is a matter of current controversy which depends on fine statistical details.

• There is also an outstanding 4σ excess of 3.5keV photons from galaxy clusters, found by XMMNewton in 2014, which has been confirmed by other experiments. The simplest DM explanation is a decaying 7keV sterile neutrino, but that hypothesis is in tension with other constraints. In addition, there are many possible non-DM sources in this energy range.

Next, we consider detection of charged particles.

• Charged cosmic rays diffuse through the galactic magnetic field, losing energy as they propagate and escaping at the boundaries; both of these effects soften the spectrum. Protons are usually in the "diffusion-dominated" regime, while electrons and positrons are in the "loss-dominated" regime, sampling only a local regime of about 1kpc. In practice, one uses specialized programs such as GALPROP or DRAGON, or pregenerated results such as PPPC4DMID.

• It is generally better to look for antimatter because DM decay should produce it equally as often, but the background should be much lower; examples include antiprotons, antideuterons, and positrons. Antimatter can be distinguished from matter by how it curves in a magnetic field. If we're using matter, we can still distinguish signal from background, but we need to use spectral features. In general, the backgrounds for charged particles are much harder to model than for photons, so while the bounds tend to be stronger, systematic uncertainties are high.

• The leading experiment for cosmic rays of GeV energies and up is AMS-02, on the International Space Station, and for the purposes of DM detection its sensitivity to antiprotons and positrons is most important.

• Some surprisingly strong limits for sub-GeV cosmic rays come from the Voyager spectrometer. These cosmic rays are usually deflected by the solar wind, but Voyager can see them since it is beyond the heliopause. This provides the best limits on 10MeV to 1GeV DM annihilating or decaying to electrons and positrons.

• Even just considering AMS-02, there are many outstanding anomalies.

• There is a very large, long-standing excess of TeV-scale positrons, which was also seen by AMS-02's predecessor Pamela. But it's so large that it's hard to explain with DM, so it's likely due to a nearby population of pulsars.

• There is an excess of antiprotons around 100GeV at 4.5σ, but the statistical significance is disputed because of possible correlations between energy bins.

• There are tentatively a few events with antihelium, which is supposed to be extremely rare from astrophysical sources. However, it's also tough to get an antihelium signal from DM.

We can also consider "very indirect" DM detection.

• Annihilating or decaying DM would have steadily injected energy into the universe throughout its history. This can modify the results of BBN, and also change the ionization and temperature history, which affects the CMB and 21 cm radiation. Such limits are clean because they don't depend on galactic astrophysics.

• Extra ionization during the cosmic dark ages ($z \sim 10 - 1000$) acts as a "screen" for CMB photons. To estimate the resulting bound, note that during radiation domination and after freeze-out, the fraction of DM that annihilates per Hubble time is

$$n\langle\sigma v\rangle H \propto T$$

and this quantity is order 1 at freeze-out itself.

• Thus, at recombination the fraction is small, $T_{\rm rec}/T_f \sim 10^{-9}$. On the other hand, the massenergy of one DM particle is enough to ionize $(1\text{GeV})/(13.6\text{eV}) \sim 10^8$ hydrogen atoms, assuming all the energy goes into ionization, so at matter-radiation equality a tiny fraction of the DM particles annihilating gives a large effect.

• Planck is sensitive to changes in the ionization of order $\sim 10^{-4}$, so ionization provides a strong bound, especially for lower DM masses. Planck rules out the thermal relic benchmark for masses below 10GeV, though the precise bound depends on the annihilation channel. Ionization also places constraints on decaying DM, which are especially strong from MeV to GeV.

• Annihilation can also heat the universe, and energy injection during recombination distorts the blackbody spectrum of the CMB. However, at matter-radiation equality the change in temperature would be of order $T_{\rm rec}/T_f \sim 10^{-9}$, which is too small for Planck to probe. More promisingly, the energy can also go into heating the baryons in the form of hydrogen gas, which are vastly outnumbered by the photons.

• We can measure the gas temperature over time by measuring the redshifted 21 cm line, with more emission corresponding to a higher temperature. Without a DM effect, we expect the gas to be colder than the CMB after recombination, since nonrelativistic matter cools down more quickly, then heat back up at the cosmic dark ages; DM would appear as a higher temperature at earlier times.

• This line is currently probed by EDGES, HERA, LOFAR, MWA, PAPER, SARAS, SCI-HI, and will be probed further by DARE, LEDA, PRIZM, and SKA. EDGES has seen a deep absorption trough at $z \sim 17$, which is extremely surprising, but the measurement is very difficult and has not yet been confirmed by other experiments.

• The Lyman-alpha forest is a set of emission lines due to the Lyman-alpha transition in neutral hydrogen. It can thus be used to probe the amount of neutral hydrogen. While the CMB is better for probing extra ionization near the beginning of the cosmic dark ages, the Lyman-alpha forest is better for probing extra ionization near the end, before the usual reionization starts. It is not as useful for earlier times because when most hydrogen is neutral, such photons are very efficiently absorbed.

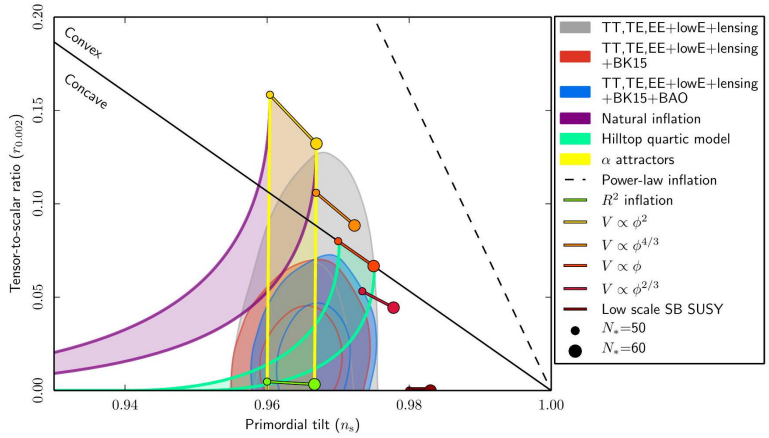
1.6.3 On experimental data

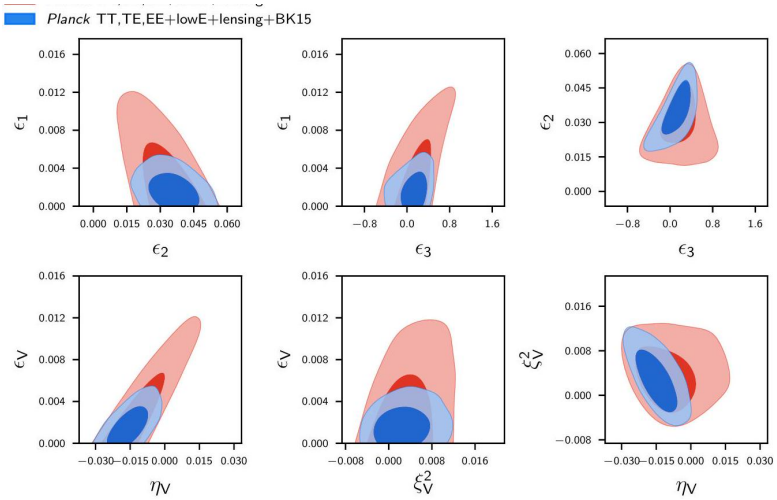
О методах измерений (!??!?!?!??!?!)

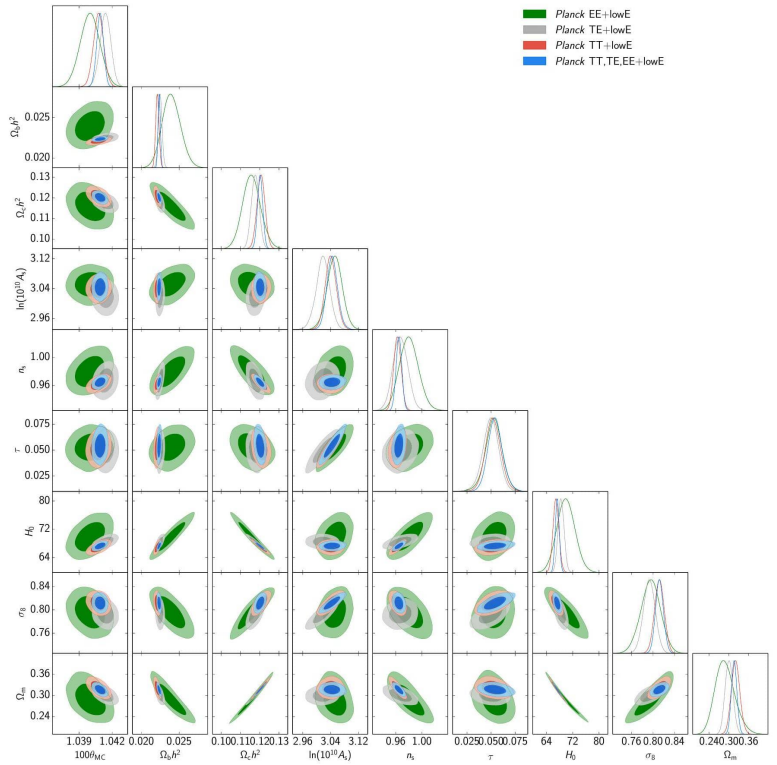
(пока хз.)

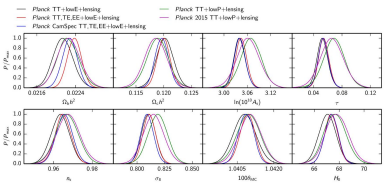
Результаты наблюдений “Планка” (???)

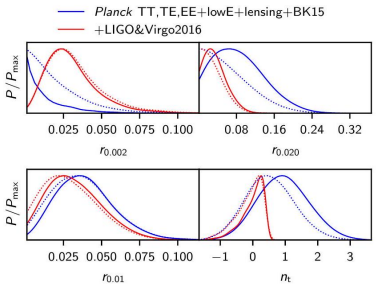
(??? еще раз посмотрю их статью, потому что я не понял, что там вообще меряется.)

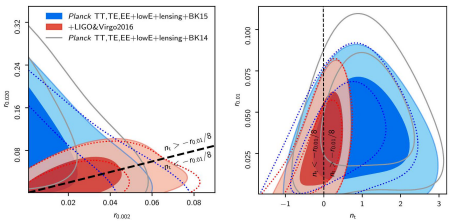

Convex
Concave
Tensor-to-scalar ratio ($r_{0.002}$)
Primordial tilt (n_s)
TT,TE,EE+lowE+lensing
TT,TE,EE+lowE+lensing +BK15
TT,TE,EE+lowE+lensing +BK15+BAO
Natural inflation
Hilltop quartic model
α attractors
Power-law inflation
R^2 inflation
$V \propto \phi^2$
$V \propto \phi^{4/3}$
$V \propto \phi$
$V \propto \phi^{2/3}$
Low scale SB SUSY
N_*=50
N_*=60

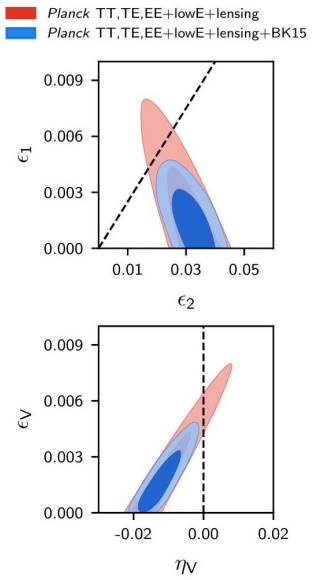

Planck TT,TE,EE+lowE+lensing+BK15
ϵ_1
ϵ_2
ϵ_3
ϵ_V
η_V
ξ^2_V


Planck EE+lowE
Planck TE+lowE
Planck TT+lowE
Planck TT,TE,EE+lowE
$100\theta_{MC}$
$\Omega_b h^2$
$\Omega_c h^2$
$\ln(10^{10}A_s)$
n_s
τ
H_0
σ_8
Ω_m









В итоге, какая модель насколько подтверждается?

(вроде сказано, что обычного скалярного поля и подтверждается, но это не точно.)

О других наблюдениях, позволяющих проверить теории космологии (????)

(другие указания, пока ничего не знаю, но наверняка может быть что-то)

Summary **of Observations**

Observations of the cosmic microwave background (CMB) and the large-scale structure (LSS) may be used to constrain the spectrum of primordial seed fluctuations. This makes CMB and LSS experiments probes of the early universe. To extract this information about the inflationary era the late-time evolution of fluctuations has to be accounted for. This is done with numerical codes such as CMBFAST and CAMB.

Current observations are in beautiful agreement with the basic inflationary predictions: The universe is flat with a spectrum of nearly scale-invariant, Gaussian and adiabatic density fluctuations. The fluctuations show non-zero correlations on scales that were bigger than the horizon at recombination. Furthermore, the peak structure of the CMB spectrum is evidence that the fluctuations we created with coherent phases.

Future tests of inflation will mainly come from measurements of CMB polarization. B-modes of CMB polarization are a unique signature of inflationary gravitational waves. The B-mode amplitude is a direct measure of the energy scale of inflation. In addition, measurements of non-Gaussianty potentially carry a wealth of information about the physics of inflation by constraining interactions of the inflaton field.

Finally, the following measurements would falsify single-field slow-roll inflation:
- Large non-Gaussianity, $f_{\rm NL} > 1$.
- Non-zero isocurvature perturbations, $\alpha \neq 0$.
- Large running of the scalar spectrum, $|\alpha_{\rm s}| > 0.001$.
- Violation of the tensor consistency relation, $r \neq -8n_{\rm t}$.

1.6.4 Statistical methods for experiments in cosmology (????)

(добавлю, если это совсем специфика - есть следующая часть на это)

Part II
Special Topics in a Nutshell

2 Other models, theories, properties and data (??)

2.1 Main physical background for cosmology

2.1.1 Gravity ideas and formalisms for cosmology

2.1.2 Gravitational lensing for cosmology in a nutshell (???)

(Додельсон аж книгу по нему написал, да и Вайнберг тоже вставляет его, думаю, не зря это. потом разберу)

Weak lensing by Lee Yoo (lectures in Zurich)

5.1 Gravitational Lensing by a Point Mass

In classical mechanics, the gravitational interaction due to a point mass M provides a perturbation along the transverse direction to a test particle moving with the relative speed $v_{\rm rel}$:

$$\Delta v_\perp = \frac{2GM}{bv_{\rm rel}} \tag{5.1}$$

where G is the Newton's constant and b is the transverse separation (or the impact parameter). The prediction for the light deflection angle $\hat{\alpha}$ in Einstein's general relativity is well-known to follow the same result in classical mechanics, but with additional factor two:

$$\hat{\alpha} = \frac{4GM}{bc^2} = 8.155 \times 10^{-3}\,\text{arcsec}\left(\frac{M}{M_\odot}\right)\left(\frac{b}{\text{AU}}\right)^{-1} \tag{5.2}$$

Given the deflection angle $\hat{\alpha}$, we can readily write down the lens equation in terms of the angular diameter distances

$$\mathcal{D}_s\hat{s} = \mathcal{D}_s\hat{n} - \mathcal{D}_{ls}\hat{\alpha}, \quad \hat{s} = \hat{n} - \theta_E^2/\hat{n} \tag{5.3}$$

where the Einstein radius is

$$\theta_E = \sqrt{\frac{4GM}{c^2}\frac{\mathcal{D}_{ls}}{\mathcal{D}_l\mathcal{D}_s}} = 2.853 \times 10^{-3}\,\text{arcsec}\left(\frac{M}{M_\odot}\right)\left(\frac{\mathcal{D}_l}{\text{kpc}}\right)^{-1/2}\left(1 - \frac{\mathcal{D}_l}{\mathcal{D}_s}\right)^{1/2} \tag{5.4}$$

For a point mass lens and a point source, two lensed image positions are readily obtained as

$$\hat{n}_1 = \frac{1}{2}\left(\hat{s} + \sqrt{\hat{s}^2 + 4\theta_E^2}\right), \quad \hat{n}_2 = \frac{1}{2}\left(\hat{s} - \sqrt{\hat{s}^2 + 4\theta_E^2}\right) < 0, \quad \hat{n}_1 + \hat{n}_2 = \hat{s} \tag{5.5}$$

and when the source and the lens are aligned, the lensed images form a ring with radius θ_E.

- microlensing, probe of MACHOs or exoplanets

5.2 Standard Weak Lensing Formalism

5.2.1 Lens Equation and Distortion Matrix

This light deflection due to a point mass can be generalized to derive the standard weak lensing formalism by considering the gravitational potential fluctuation $\psi = -GM/r$ of the general matter distribution ρ (but still a single lens plane), instead of a point mass (ψ indeed corresponds to the metric fluctuation α_χ). The lensing potential Φ is the line-of-sight integration of the metric fluctuation,

$$\Phi := \frac{1}{c^2}\frac{\mathcal{D}_{ls}}{\mathcal{D}_l\mathcal{D}_s}\int dz 2\psi \tag{5.6}$$

and using the Poisson equation, we can relate the lensing potential with the surface density Σ as

$$\nabla^2\psi = 4\pi G\bar{\rho}a^2\delta, \qquad \hat{\nabla}^2\Phi = 2\frac{\Sigma}{\Sigma_c}, \quad \Phi(\hat{n}) = \int d^2\hat{n}'\frac{\Sigma}{\pi\Sigma_c}\ln|\hat{n}-\hat{n}'| \tag{5.7}$$

where we ignored the boundary term when the Poisson equation is integrated and the critical surface density is defined as

$$\Sigma_c^{-1} := \frac{4\pi G}{c^2}\frac{\mathcal{D}_{ls}\mathcal{D}_l}{\mathcal{D}_s}, \quad \Sigma_c = 1.663\times 10^6 h M_\odot \mathrm{pc}^{-2}\left(\frac{\mathcal{D}_s}{\mathcal{D}_{ls}}\right)\left(\frac{\mathcal{D}_l}{h^{-1}\mathrm{Mpc}}\right)^{-1} \tag{5.8}$$

Though the lensing potential is formally divergent for a point mass, its angular derivative is well defined:

$$\hat{\alpha} = \left(\frac{\mathcal{D}_l\mathcal{D}_s}{\mathcal{D}_{ls}}\right)\nabla_\perp\Phi = \frac{\mathcal{D}_s}{\mathcal{D}_{ls}}\hat{\nabla}\Phi \tag{5.9}$$

such that the lens equation becomes

$$\hat{s} = \hat{\boldsymbol{n}} - \hat{\nabla}\Phi \tag{5.10}$$

where $\hat{\nabla}$ is the angular gradient. When the lensing material is distributed over the redshift, the lensing potential is then obtained by integrating the potential fluctuation over the line-of-sight distance as

$$\Phi = \int_0^{\bar{r}_s} d\bar{r}\left(\frac{\bar{r}_s-\bar{r}}{\bar{r}_s\bar{r}}\right)2\psi \equiv \int_0^{\bar{r}_s} d\bar{r}\frac{g\left(\bar{r},\bar{r}_s\right)}{\bar{r}^2}2\psi \tag{5.11}$$

where we switched to a comoving angular diameter distance $\bar{r}$ and we defined the weight function g for later convenience

$$g := \bar{r}^2\left(\frac{\bar{r}_s-\bar{r}}{\bar{r}_s\bar{r}}\right) \tag{5.12}$$

When the background source galaxies are also spread over some redshift with the distribution $n_g\left(r_s\right)$, the lensing potential can be readily generalized by replacing the weight function with

$$g := \bar{r}^2\int_{\bar{r}}^{\infty} d\bar{r}_s\left(\frac{\bar{r}_s-\bar{r}}{\bar{r}_s\bar{r}}\right)n_g\left(\bar{r}_s\right), \quad \Phi = \int_0^{\infty} d\bar{r}\frac{g(\bar{r})}{\bar{r}^2}2\psi, \quad 1 = \int_0^{\infty} d\bar{r}_s n_g\left(\bar{r}_s\right) \tag{5.13}$$

where the upper limit for the integration is indeed $\bar{r}(z=\infty)$ and the source distribution is normalized. [1]

Using the lens equation, the distortion matrix $\mathbb{D}$ (or sometimes called the amplification matrix) is defined as

$$\mathbb{D}_{ij} \equiv \frac{\partial s_i}{\partial n_j} = \mathbb{I}_{ij} - \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix}, \quad \Phi_{ij} := \hat{\nabla}_j \hat{\nabla}_i \Phi \tag{5.14}$$

where $\mathbb{I}$ is the two-dimensional identity matrix and we defined a short hand notation for the angular derivatives of the lensing potential. The distortion matrix is conventionally decomposed into the trace, the traceless symmetric and the anti-symmetric matrices:

$$\mathbb{D} := \mathbb{I} - \begin{pmatrix} \kappa & 0 \\ 0 & \kappa \end{pmatrix} - \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix} - \begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix}, \quad \det \mathbb{D} = (1-\kappa)^2 - \gamma^2 + \omega^2 \tag{5.15}$$

where the trace is the gravitational lensing convergence κ and the symmetric traceless part is the lensing shear $\gamma = \sqrt{\gamma_1^2 + \gamma_2^2}$:

$$\begin{aligned} \kappa &\equiv 1 - \frac{1}{2}\operatorname{Tr}\mathbb{D} = \frac{1}{2}\left(\Phi_{11} + \Phi_{22}\right), & \omega &\equiv \frac{\mathbb{D}_{21} - \mathbb{D}_{12}}{2} = 0 \\ \gamma_1 &\equiv \frac{\mathbb{D}_{22} - \mathbb{D}_{11}}{2} = \frac{1}{2}\left(\Phi_{11} - \Phi_{22}\right), & \gamma_2 &\equiv -\frac{\mathbb{D}_{12} + \mathbb{D}_{21}}{2} = \Phi_{12} = \Phi_{21} \end{aligned} \tag{5.16}$$

Since the distortion matrix in Eq. (5.14) is symmetric, the rotation ω vanishes in the standard formalism at all orders.

The standard lensing formalism is based on the lens equation and the lensing potential in Eq. (5.10). However, the source angular position $\hat{s} := (\theta + \delta\theta, \phi + \delta\phi)$ is gauge-dependent, and the lensing potential that is responsible for the angular distortion $(\delta\theta, \delta\phi)$ is also gauge-dependent. Indeed, we already know that 2ψ in Eq. (5.10) should be $(\alpha_\chi - \varphi_\chi)$ to match the leading terms for $\delta\theta$ and the Poisson equation in Eq. (5.21) is indeed an Einstein equation with ψ there replaced by $-\varphi_\chi$.[2] Furthermore, there exist no contributions from the vector and the tensor perturbations in the standard lensing formalism. Finally, while the derivations in this subsection assume no linearity, all formulas of the standard lensing formalism turn out to be valid only at the linear order in perturbations.

5.2.2 Convergence and Shear

While the distortion matrix is defined in terms of angles, it is often assumed in literature that the line-of-sight direction is along z-axis ($\hat{\boldsymbol{n}} // \hat{\boldsymbol{z}}$, i.e., $\theta = 0$), and two angles are aligned with $x - y$ plane. In such a setting, consider two small angular vectors at the source position subtended respectively by $d\theta$ and $d\phi$ at the observer position

$$\Delta s_i^{d\theta} = \mathbb{D}_{i1} d\theta, \quad \Delta s_i^{d\phi} = \mathbb{D}_{i2} d\phi \tag{5.18}$$

The solid angle at the source subtended by these two angular vectors is then related to the solid angle at the observer as

$$d\Omega_s = \left|\Delta s^{d\theta} \times \Delta s^{d\phi}\right| = \det \mathbb{D} d\theta d\phi = \det \mathbb{D} d\Omega_o \tag{5.19}$$

and hence the gravitational lensing magnification μ is then

[1] Sometimes it is normalized when integrated over redshift.

[2] Additional condition of a vanishing anisotropic pressure is needed to guarantee $\alpha_\chi = -\varphi_\chi$ and hence the consistency in the lensing equation.

$$\mu^{-1} \equiv \frac{d\Omega_s}{d\Omega_o} = \det \mathbb{D} \tag{5.20}$$

For this reason, the distortion matrix is often called the inverse magnification matrix. Using the Poisson equation in cosmology,

$$\nabla^2\psi = 4\pi G\bar{\rho}a^2\delta_m = \frac{3H_o^2}{2}\Omega_m\frac{\delta_m}{a} \tag{5.21}$$

the gravitational lensing convergence can be computed in terms of the matter density fluctuation δ_m in the comoving gauge as

$$\kappa = \int_0^{\bar{r}_s} d\bar{r} g\left(\bar{r},\bar{r}_s\right)\nabla_\perp^2\psi = \frac{3H_o^2}{2}\Omega_m\int_0^{\bar{r}_s} d\bar{r}\frac{g\left(\bar{r},\bar{r}_s\right)}{a}\delta_m \tag{5.22}$$

where we used

$$\nabla^2 = \nabla_\perp^2 + \frac{\partial^2}{\partial\bar{r}^2} + \frac{2}{\bar{r}}\frac{\partial}{\partial\bar{r}}, \quad \nabla_\perp^2 = \frac{1}{\bar{r}^2}\hat{\nabla}^2 \tag{5.23}$$

and ignored the boundary terms.
In general, we can compute the individual components of the distortion matrix

$$\gamma_1 = \int_0^{\bar{r}_s} d\bar{r} g\left(\bar{r},\bar{r}_s\right)\left(\nabla_1^2 - \nabla_2^2\right)\psi, \qquad \gamma_2 = 2\int_0^{\bar{r}_s} d\bar{r} g\left(\bar{r},\bar{r}_s\right)\nabla_1\nabla_2\psi \tag{5.24}$$

by using

$$\frac{1}{2}\Phi_{ij} \equiv \frac{1}{2}\hat{\nabla}_i\hat{\nabla}_j\Phi = \int_0^{\bar{r}_s} d\bar{r} g\left(\bar{r},\bar{r}_s\right)\nabla_i\nabla_j\psi \tag{5.25}$$

where the indices i, j represent the perpendicular components.

5.2.3 Angular Power Spectrum and Angular Correlation

Assuming that the survey area is small, we will utilize the angular Fourier transformation in Eq. (2.17) by again computing

$$\Phi(\mathbf{l}) = \int d^2\theta e^{-i\mathbf{l}\cdot\boldsymbol{\theta}}\int_0^{\bar{r}_s} d\bar{r}\frac{g\left(\bar{r},\bar{r}_s\right)}{\bar{r}^2}2\psi = \int_0^{\bar{r}_s} d\bar{r}\frac{g\left(\bar{r},\bar{r}_s\right)}{\bar{r}^4}\int\frac{dk_\parallel}{2\pi}2\psi\left(k_\perp = \frac{l}{\bar{r}}\right)e^{ik_\parallel\bar{r}} \tag{5.26}$$

such that the lensing observables are

$$\kappa(\mathbf{l}) = -\frac{l^2}{2}\Phi(l), \quad \gamma_1(\mathbf{l}) = -\frac{l_1^2 - l_2^2}{2}\Phi(l) = \cos 2\phi_l\kappa(\mathbf{l}), \quad \gamma_2(\mathbf{l}) = -l_1 l_2\Phi(l) = \sin 2\phi_l\kappa(\mathbf{l}) \tag{5.27}$$

where we used $\mathbf{l} = (l_1, l_2) = l\left(\cos\phi_l, \sin\phi_l\right)$. Therefore, the angular power spectra can be readily derived as

$$P_\kappa(l) = \frac{l^4}{4}P_\Phi(l), \quad P_{\gamma_1} = \cos^2 2\phi_l P_\kappa(l), \quad P_{\gamma_2} = \sin^2 2\phi_l P_\kappa(l) \tag{5.28}$$

where the angular power spectrum of the lensing potential is obtained by using the Limber approximation as

$$P_\Phi(l) = \int_0^{\bar{r}_s} d\bar{r} \frac{g^2(\bar{r}, \bar{r}_s)}{\bar{r}^6} 4P_\psi \left(k_\perp = \frac{l}{\bar{r}} \right) \tag{5.29}$$

From the relation of the lensing observables, we find it useful to construct E and B-modes as

$$E(\mathbf{l}) := \cos 2\phi_l \gamma_1(\mathbf{l}) + \sin 2\phi_l \gamma_2(\mathbf{l}), \quad B(\mathbf{l}) := -\sin 2\phi_l \gamma_1(\mathbf{l}) + \cos 2\phi_l \gamma_2(\mathbf{l}) \tag{5.30}$$

We can readily derive

$$E(\mathbf{l}) = \kappa(\mathbf{l}), \quad B(\mathbf{l}) = 0, \quad P_E(l) = P_\kappa(l), \quad P_B(l) = P_{EB}(l) = 0 \tag{5.31}$$

in the absence of any systematics and/or physics other than the gravitational lensing, such that it provides a consistency check of the measurements, where the convergence power spectrum is again related to the matter power spectrum as

$$P_\kappa(l) = \left(\frac{3H_0^2}{2} \Omega_m \right)^2 \int_0^{\bar{r}_s} d\bar{r} \frac{g^2(\bar{r}, \bar{r}_s)}{\bar{r}^2 a^2} P_m \left(k_\perp = \frac{l}{\bar{r}} \right) \tag{5.32}$$

Now we compute the angular correlation function by Fourier transforming the angular power spectrum. Out of two shear components, we construct three angular correlation functions as

$$\begin{aligned} w_{ij}(\theta) := \langle \gamma_i(0) \gamma_j(\theta) \rangle &= \int \frac{d^2 l}{(2\pi)^2} e^{i1\cdot\theta} \begin{pmatrix} \cos^2 2\phi_l & \cos 2\phi_l \sin 2\phi_l \\ \cos 2\phi_l \sin 2\phi_l & \sin^2 2\phi_l \end{pmatrix} P_\kappa(l) \\ &= \frac{1}{2} \int_0^\infty \frac{dl}{2\pi} P_\kappa(l) \begin{pmatrix} J_0(l\theta) + J_4(l\theta) & 0 \\ 0 & J_0(l\theta) - J_4(l\theta) \end{pmatrix} \end{aligned} \tag{5.33}$$

where J_n is the Bessel function and used its integral representation

$$J_0(x) = \int \frac{d\phi}{2\pi} e^{ix\cos\phi}, \quad J_4(x) = \int \frac{d\phi}{2\pi} e^{ix\cos\phi} \cos 4\phi \tag{5.35}$$

5.2.4 Worked Examples

For the simplest case, where the lens and the source are at two definite redshift slices, the lensing observables can be written in a polar coordinate as

$$2\kappa = \Phi_{rr} + \frac{\Phi_r}{r} + \frac{\Phi_{\theta\theta}}{r^2} = 2\frac{\Sigma}{\Sigma_c} \tag{5.36}$$

$$2\gamma_1 = \cos 2\theta \Phi_{rr} - \frac{2\sin 2\theta}{r} \Phi_{r\theta} - \frac{\cos 2\theta}{r} \Phi_r - \frac{\cos 2\theta}{r^2} \Phi_{\theta\theta} + \frac{2\sin 2\theta}{r^2} \Phi_\theta \tag{5.37}$$

$$2\gamma_2 = \sin 2\theta \Phi_{rr} + \frac{2\cos 2\theta}{r} \Phi_{r\theta} - \frac{\sin 2\theta}{r} \Phi_r - \frac{\sin 2\theta}{r^2} \Phi_{\theta\theta} - \frac{2\cos 2\theta}{r^2} \Phi_\theta \tag{5.38}$$

$$\gamma^2 := \gamma_1^2 + \gamma_2^2 = \frac{1}{4} \left(\Phi_{rr} - \frac{\Phi_r}{r} - \frac{\Phi_{\theta\theta}}{r^2} \right)^2 + \left(\frac{\Phi_{r\theta}}{r} - \frac{\Phi_\theta}{r^2} \right)^2 \tag{5.39}$$

For an axisymmetric lens, the lensing observables are further simplified, and the convergence and shear are

$$2\kappa = \hat{\nabla}^2\Phi = \Phi_{rr} + \frac{\Phi_r}{r} = 2\frac{\Sigma}{\Sigma_c}, \quad \gamma = \frac{1}{2}\left(\frac{\Phi_r}{r} - \Phi_{rr}\right) = \frac{\Phi_r}{r} - \frac{\Sigma}{\Sigma_c} = \frac{\bar{\Sigma}(<r) - \Sigma}{\Sigma_c} \tag{5.40}$$

where $\bar{\Sigma}(<r)$ is the average surface density enclosed in radius r and $\bar{\Sigma}(<r) = \Phi_r/r$ from the first relation. The magnification is determined by the surface density of the lensing material, and the gravitational shear is set by the excess surface density of the enclosed mass $\Delta\Sigma := \bar{\Sigma}(<r) - \Sigma(r)$.

For a point mass, the convergence and the shear are

$$\psi = -\frac{GM}{r}, \quad \nabla^2\psi = 4\pi GM\delta^D(x), \quad \kappa = \frac{\Sigma}{\Sigma_c} = \frac{M\delta^D(R)}{\Sigma_c} \tag{5.41}$$

$$\gamma = \frac{\bar{\Sigma}(<R) - \Sigma}{\Sigma_c} = \frac{\bar{\Sigma}}{\Sigma_c} = \frac{\theta_E^2}{\theta^2}, \quad \bar{\Sigma} := \frac{M}{\pi R^2} \tag{5.42}$$

The lensing magnification is then

$$\mu^{-1} = \det\mathbb{D} = 1 - \Phi_{rr} - \frac{\Phi_r}{r} + \frac{\Phi_{rr}\Phi_r}{r} = \frac{s}{r}\frac{\partial s}{\partial r} \tag{5.43}$$

where we used the lens equation

$$s = r - \Phi_r, \quad \partial_r s = 1 - \Phi_{rr} \tag{5.44}$$

For a point mass, there exist two lensed images. When two images are not spatially resolved, the magnification of the lensed images is the sum of two, and we derive the master equation for microlening

$$\mu = \left(\frac{s}{r}\frac{\partial s}{\partial r}\right)_1^{-1} + \left(\frac{s}{r}\frac{\partial s}{\partial r}\right)_2^{-1} = \frac{u^2+2}{u\sqrt{u^2+4}}, \quad u := \frac{s}{\theta_E} \tag{5.45}$$

5.2.5 Galaxy-Galaxy Lensing

Galaxy-galaxy lensing is used to refer to the two-point correlation of the galaxies at one point and the lensing signal measured by background galaxies at the other point. In short, it measures the galaxy-matter cross-correlation. Compared to the cosmic shear measurements, the advantage here is that we have well-defined lenses (lens galaxies) in the foreground, such that the shear measurements in galaxy-galaxy lensing are less susceptible to other systematics.

Assuming spherical symmetry, we can readily derive the lensing convergence and the shear as

$$\kappa(\theta) = \frac{\Sigma(\theta)}{\Sigma_c}, \quad \gamma(\theta) = \bar{\kappa}(\theta) - \kappa(\theta) \tag{5.46}$$

where the "comoving" critical surface density is [3]

$$\Sigma_c(z_1, z_2) = \frac{c^2}{4\pi G}\frac{r_s}{r_l r_{ls}}\frac{1}{1+z_l} = 1.663\times10^{18} h M_\odot \mathrm{Mpc}^{-2}\frac{r_s}{r_l}\left(\frac{r_{ls}}{h^{-1}\mathrm{Mpc}}\right)^{-1}\frac{1}{1+z_l} \tag{5.47}$$

Since we are measuring the excess matter around galaxies, the lensing observables are related to the projected galaxymatter correlation function:

$$w(R) = \int_{-\infty}^{\infty} dz \xi_{\rm gm}\left(r = \sqrt{R^2 + z^2}\right) = \int_0^{\infty} \frac{dk_\perp k_\perp}{2\pi} P_{\rm gm}\left(k_\perp\right) J_0\left(k_\perp R\right) \tag{5.48}$$

where the integration along the line-of-sight is performed. However, note that this is valid only on small angle, as the observed angular separation θ, not the physical separation R is kept fixed. Under the small-angle approximation, the convergence at a given separation can be derived as

$$\kappa(\theta) = \frac{\Sigma(\theta)}{\Sigma_c} = \int dz \frac{\bar{\rho}_m}{\Sigma_c}\left(1 + \xi_{gm}\right) = \frac{\bar{\rho}_m}{\Sigma_c} w(R) \tag{5.49}$$

and the remaining lensing observables are

$$\bar{\kappa}(\theta) = \frac{2\pi}{\pi\theta^2} \int_0^{\theta} d\theta\theta\kappa(\theta) = \frac{3H_0^2\Omega_m}{2} \frac{\left(\bar{r}_s - \bar{r}_l\right)\bar{r}_l}{a_l \bar{r}_s} \int \frac{dk_\perp k_\perp}{2\pi} P_{gm}\left(k_\perp; \bar{r}_l\right) \frac{2J_1\left(k_\perp \bar{r}\theta\right)}{k_\perp \bar{r}\theta} \tag{5.50}$$

$$\gamma_T(\theta) = \bar{\kappa}(\theta) - \kappa(\theta) = \frac{3H_0^2\Omega_m}{2} \frac{\left(\bar{r}_s - \bar{r}_l\right)\bar{r}_l}{a_l \bar{r}_s} \int \frac{dk_\perp k_\perp}{2\pi} P_{gm}\left(k_\perp; \bar{r}_l\right) J_2\left(k_\perp \bar{r}\theta\right) \tag{5.51}$$

$$\Delta\Sigma(R) = \Sigma_c \gamma_T = \bar{\rho}_m \int_{-\infty}^{\infty} dz \left[\frac{2}{R^2} \int_0^R dR' R' \xi_{\rm gm}\left(R', z\right) - \xi_{\rm gm}(R, z)\right] \tag{5.52}$$

5.3 Weak Lensing Observables

5.3.1 Ellipticity of Galaxies

The ellipticity ϵ of galaxies is measured in terms of its semi-major axis a and the semi-minor axis b or in terms of the axis ratio q as

$$\epsilon := \frac{a^2 - b^2}{a^2 + b^2} \equiv \frac{1 - q^2}{1 + q^2} \equiv \frac{\delta - \frac{1}{2}\delta^2}{1 - \delta + \frac{1}{2}\delta^2} \simeq \delta, \quad q := \frac{b}{a} := 1 - \delta \tag{5.53}$$

In an idealized case of round galaxies, the ellipticity ϵ, the axis ratio q, and the distortion δ are a measure of gravitational lensing effects of intervening matter, and they are equivalent in the weak lensing regime. In observations, the center of the galaxy and its ellipticity moment are measured by using some weight function $W\left[I_\nu(\hat{n})\right]$ of the observed intensity as

$$\hat{n}_o := \frac{\int d^2\hat{n}\hat{n}W[\hat{n}]}{\int d^2\hat{n}W[\hat{n}]}, \quad \mathcal{M}_{ij} := \frac{\int d^2\hat{n}\left(\hat{n} - \hat{n}_o\right)_i \left(\hat{n} - \hat{n}_o\right)_j W[\hat{n}]}{\int d^2\hat{n}W[\hat{n}]} \tag{5.54}$$

where the simplest weight function is just the observed intensity $W = I[\hat{n}]$. Given the ellipticity moment, we can define the ellipticity vector and the position angle as

$$\boldsymbol{\epsilon} := \left(\frac{\mathcal{M}_{xx} - \mathcal{M}_{yy}}{\mathcal{M}_{xx} + \mathcal{M}_{yy}}, \frac{2\mathcal{M}_{xy}}{\mathcal{M}_{xx} + \mathcal{M}_{yy}}\right) := \left(\epsilon_+, \epsilon_\times\right) = \epsilon(\cos 2\Theta, \sin 2\Theta), \quad \tan 2\Theta \equiv \frac{2\mathcal{M}_{xy}}{\mathcal{M}_{xx} - \mathcal{M}_{yy}} \tag{5.55}$$

[3] Multiply by $\left(1 + z_l\right)^2$ for physical critical surface density. Note that sometimes people use the physical angular diameter distances, while using comoving coordinates for other quantities, in which $\left(1 + z_l\right)^2$ appears in the equation, instead of $\left(1 + z_l\right)$.

Note that the ellipticity vector is headless, such that it is identical under 180 degree rotation, or spin 2. Since only the ellipticity vector matters, the ellipticity moments $\mathcal{M}$ are often defined without the denominator.

5.3.2 Lensing Polarization

The ellipticity moments of the source galaxies would be what we measure in the absence of gravitational lensing. However, the gravitational lensing changes the observed ellipticity moments. Now, for simplicity, we will ignore rotation ($\omega = 0$) and express the distortion matrix in our coordinate:

$$\mathbb{D} = \mathbb{I} - \begin{pmatrix} \kappa & 0 \\ 0 & \kappa \end{pmatrix} - \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix} \tag{5.56}$$

and the magnification matrix is then the inverse of the distortion matrix:

$$\mathbf{M}_{ij} := \mathbb{D}^{-1}_{ij} = \frac{1}{|\mathbb{D}|} \begin{pmatrix} 1-\kappa+\gamma_1 & \gamma_2 \\ \gamma_2 & 1-\kappa-\gamma_1 \end{pmatrix}, \quad \mu := |\mathbf{M}| = |\mathbb{D}|^{-1} = \frac{1}{(1-\kappa)^2-\gamma^2} \tag{5.57}$$

Further, assuming the surface brightness conservation due to gravitational lensing (i.e., no frequency change), the observed ellipticity moments are related to those in the source rest-frame as

$$\mathcal{M}^I_{ij} := \int d^2\hat{n}\hat{n}_i\hat{n}_j W[\hat{n}] = \int |\mathbf{M}| d^2\hat{s}\mathbf{M}_{ik}\hat{s}_k\mathbf{M}_{jl}\hat{s}_l W[\hat{s}] \simeq \mu\mathbf{M}_{ik}\mathbf{M}_{jl}\mathcal{M}^s_{kl} \tag{5.58}$$

where we assumed $\hat{n}_o = 0$ and the source size is small that the magnification matrix is constant over the area. Using the definition of the source ellipticity moments

$$\mathcal{M}^s_{11} = \frac{1+\epsilon^s_+}{2}\mathcal{M}, \quad \mathcal{M}^s_{22} = \frac{1-\epsilon^s_+}{2}\mathcal{M}, \quad \mathcal{M}^s_{12} = \frac{\epsilon^s_\times}{2}\mathcal{M}, \quad \mathcal{M} := \mathcal{M}^s_{11} + \mathcal{M}^s_{22} \tag{5.59}$$

the observed ellipticity can be derived in terms of the magnification matrix as

$$\epsilon^I_+ = \frac{\left(1+\epsilon^s_+\right)\mathbf{M}^2_{11} + 2\epsilon^s_\times\mathbf{M}_{12}\left(\mathbf{M}_{11}-\mathbf{M}_{22}\right) - 2\epsilon^s_+\mathbf{M}^2_{12} - \left(1-\epsilon^s_+\right)\mathbf{M}^2_{22}}{\left(1+\epsilon^s_+\right)\mathbf{M}^2_{11} + 2\epsilon^s_\times\mathbf{M}_{12}\left(\mathbf{M}_{11}+\mathbf{M}_{22}\right) + 2\epsilon^s_+\mathbf{M}^2_{12} + \left(1-\epsilon^s_+\right)\mathbf{M}^2_{22}} \tag{5.60}$$

$$\epsilon^I_\times = \frac{2\mathbf{M}_{12}\left[\epsilon^s_\times\mathbf{M}_{12} + \left(1-\epsilon^s_+\right)\mathbf{M}_{22}\right] + 2\mathbf{M}_{11}\left[\left(1+\epsilon^s_+\right)\mathbf{M}_{12} + \epsilon^s_\times\mathbf{M}_{22}\right]}{\left(1+\epsilon^s_+\right)\mathbf{M}^2_{11} + 2\epsilon^s_\times\mathbf{M}_{12}\left(\mathbf{M}_{11}+\mathbf{M}_{22}\right) + 2\mathbf{M}^2_{12} + \left(1-\epsilon^s_+\right)\mathbf{M}^2_{22}} \tag{5.61}$$

where the relation is exact. In terms of the lensing convergence and shear, [4] we derive

$$\epsilon^I_+ = \frac{\epsilon^s_+\left[(1-\kappa)^2+\gamma^2_1-\gamma^2_2\right] + 2\epsilon^s_\times\gamma_1\gamma_2 + 2\gamma_1(1-\kappa)}{2\epsilon^s_+\gamma_1(1-\kappa) + 2\epsilon^s_\times\gamma_2(1-\kappa) + (1-\kappa)^2 + \gamma^2_1 + \gamma^2_2} \tag{5.62}$$

$$\epsilon^I_\times = \frac{\epsilon^s_\times\left[(1-\kappa)^2-\gamma^2_1+\gamma^2_2\right] + 2\epsilon^s_+\gamma_1\gamma_2 + 2\gamma_2(1-\kappa)}{2\epsilon^s_+\gamma_1(1-\kappa) + 2\epsilon^s_\times\gamma_2(1-\kappa) + (1-\kappa)^2 + \gamma^2_1 + \gamma^2_2} \tag{5.63}$$

For circular sources, where $\mathcal{M}^s_{11} = \mathcal{M}^s_{22} \neq 0$, and $\mathcal{M}_{12} = 0$ (or $\epsilon^s_+ = \epsilon^s_\times = 0$), the observed ellipticity becomes

$$\epsilon^I_+ = \frac{2\gamma_1(1-\kappa)}{(1-\kappa)^2+\gamma^2_1+\gamma^2_2} \simeq 2\gamma_1, \quad \epsilon^I_\times = \frac{2\gamma_2(1-\kappa)}{(1-\kappa)^2+\gamma^2_1+\gamma^2_2} \simeq 2\gamma_2 \tag{5.64}$$

where we expanded to the linear order in the last step. Assuming there is no ellipticity correlation of the source galaxies

$$\langle \epsilon_+ \rangle = \langle \epsilon_\times \rangle = \langle \epsilon_+ \epsilon_\times \rangle = \frac{1}{2} \langle \epsilon^2 \rangle \langle \sin 4\phi \rangle = 0 \tag{5.65}$$

$$\langle \epsilon_+^2 \rangle = \langle \epsilon_\times^2 \rangle = \langle \epsilon^2 \rangle \langle \cos^2 2\phi \rangle = \frac{1}{2} \langle \epsilon^2 \rangle \tag{5.66}$$

the observed ellipticity correlation becomes

$$\xi_{\epsilon_+^I}(\theta) = \langle \epsilon_+^I(0) \epsilon_+^I(\theta) \rangle = \xi_\delta(\theta) \left(1 - \sigma_\epsilon^2 + \frac{1}{4}\sigma_\epsilon^4 \right) = \xi_{\epsilon_\times^I},$$

to be compared to the typical ellipticity

$$\sigma_\epsilon = \langle \epsilon^s \rangle^{1/2} \simeq 0.3$$

- modified gravity, no galaxy bias
(???????????)

2.1.3 On some stars and astrophysics for cosmology

(полезные знания, но пока об этом в астрофизике сперва)

Type Ia supernova

(пока только мини инфа из лекций Caltech-a, нужно будет - разберусь потом в теории с этим)

Results from SNe Ia. The SNe Ia were the first technique to directly measure the acceleration of the universe. For low- z observations this is typically quoted in terms of the deceleration parameter q_0, defined by:

$$q_0 = -\left. \frac{a\ddot{a}}{\dot{a}^2} \right|_{a=1} = \frac{1}{2}\Omega_m - \Omega_\Lambda.$$

So $q_0 > 0$ corresponds to a decelerating universe, $q_0 < 0$ to accelerating. From this one can show that

$$H_0 D_L = z + \frac{1 - q_0}{2} z^2 + O\left(z^3\right).$$

In 1998 two supernova teams found that the universe was accelerating, each with ~ 50 SNe:
- High- z Supernova Search: Riess et al. 1998 (AJ 116, 1009). $q_0 = -1.0 \pm 0.4$.
- Supernova Cosmology Project: Perlmutter et al. 1999 (ApJ 517, 565). $0.8\Omega_m - 0.6\Omega_\Lambda = -0.2 \pm 0.1$. (The SCP result is not quite a constraint on q_0 because of the higher-order terms z^3, etc.)

2.2 Other models of cosmology in a nutshell

2.2.1 Variable Constants models overview

(есть статьи про это, посмотрю потом, тут прямо очень все сложно становится)

2.2.2 Description of formation of Solar System

(потом поизучаю, не до этого пока)

[4] With rotation, the magnification matrix is not symmetric, $\mathbf{M}_{12} \neq \mathbf{M}_{21}$.

2.2.3 On other models of inflation

(статья есть, где целая куча моделей собрана, укажу просто тут их, мб графики какие-то приведу, как они отличаются и что эксперимент про них говорит.)

Summary of Inflation in String Theory

Recent work by many authors has led to the emergence of robust mechanisms for inflation in string theory (see Refs. [16, 136, 137, 138, 139, 140] for recent reviews). The primary motivations for these works are the sensitivity of inflationary effective actions to the ultraviolet completion of gravity, and the prospect of empirical tests using precision cosmological data.

In this lecture we illustrated the *UV sensitivity of inflation* with two examples:

• *The eta problem*

The smallness of the eta parameter (or the inflaton mass) is sensitive to dimension-six Planck-suppressed corrections,

$$\Delta V \sim V\frac{\phi^2}{M_{\rm pl}^2} \qquad \Rightarrow \qquad \Delta\eta_{\rm v} \sim 1\,.$$

Such terms arise when integrating out heavy degrees of freedom (above the cutoff) to arrive at the low energy effective theory. For the example of warp brane inflation we showed how this problem is made explicit in string theory calculations [117, 47, 48, 121].

• *Tensor modes in large-field models*

The inflaton field is required to move over a super-Planckian distance for inflation to generate an observable gravitational wave amplitude. Protecting the flatness of the inflationary potential over a super-Planckian range is challenging:

• *No shift symmetry*

In the absence of any special symmetries, the potential in large-field inflation becomes sensitive to an infinite series of Planck-suppressed operators

$$\mathcal{L}_{\rm eff}(\phi) = -\frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{1}{4}\lambda\phi^4 - \sum_{p=1}^{\infty}\left[\lambda_p\phi^4 + \nu_p(\partial\phi)^2\right]\left(\frac{\phi}{M_{\rm pl}}\right)^{2p} + \cdots\,.$$

In this case, the flatness of the potential over a super-Planckian range requires a fine-tuning of a large number of expansion parameters λ_p (compared to the eta problem which only requires tuning of one mass parameter).

• *Shift symmetry*

If the inflaton field respects a shift symmetry, $\phi \to \phi + const.$, then the action of chaotic inflation

$$\mathcal{L}_{\rm eff}(\phi) = -\frac{1}{2}(\partial\phi)^2 - \lambda_p\,\phi^p\,,$$

with small coefficient λ_p is 'technically natural'.

To construct an inflationary model with detectable gravitational waves, we are therefore interested in finding, in string theory, a configuration that has both a large kinematic range, $\Delta\phi > M_{\rm pl}$, and a potential protected by a shift symmetry that is approximately preserved by the full string theory. Such models have recently been constructed in Refs. [251, 44, 141].

2.2.4 Primordial phase transitions, quantum bubbles, primordial GWs (????)

(выпишу суть уже написанных у меня лекций. Про GW может быть отдельный раздел сделаю потом)

Characteristic frequency of relic GWs

(??? no idea now)

GW production by classical fields

(by the way, this will first be in my note on GWs)

General formalism

GW generation by a stochastic scalar field

GWs from preheating after inflation

Parametric resonance in single-field inflation

$$S=\int d^4x\sqrt{-g}\left[\frac{1}{16\pi G}R-\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi-V(\phi)-\frac{1}{2}g^{\mu\nu}\partial_\mu\chi\partial_\nu\chi-\frac{1}{2}g^2\chi^2\phi^2\right] \tag{22.43}$$

$$\phi(t,\mathbf{x})=\phi(t)+\delta\phi(t,\mathbf{x}) \tag{22.44}$$
$$\chi(t,\mathbf{x})=\delta\chi(t,\mathbf{x}) \tag{22.45}$$

$$\phi(t)=\Phi(t)\sin m_\phi t \tag{22.46}$$

$$\begin{aligned}\delta\ddot{\phi}_{\mathbf{k}}+3H\delta\dot{\phi}_{\mathbf{k}}+\left(\tfrac{k^2}{a^2}+m_\phi^2\right)\delta\phi_{\mathbf{k}}=0\\ \delta\ddot{\chi}_{\mathbf{k}}+3H\delta\dot{\chi}_{\mathbf{k}}+\left(\tfrac{k^2}{a^2}+g^2\phi^2(t)\right)\delta\chi_{\mathbf{k}}=0\end{aligned} \tag{22.48}$$

$$\begin{aligned}\omega_k^2(t)&=k^2+g^2a^2(t)\phi^2(t)\\&=k^2+g^2\Phi^2a^2(t)\sin m_\phi t\end{aligned} \tag{22.49}$$

$$\frac{d^2}{dz^2}\left(\delta\chi_k\right)+3\frac{H}{m_\phi}\frac{d}{dz}\left(\delta\chi_k\right)+\left[A(k)-2q\cos(2z)\right]\delta\chi_k \tag{22.50}$$

$z=m_\phi t$,

$$q=\frac{g^2\Phi^2}{4m_\phi^2} \tag{22.51}$$

$$A(k)=\frac{k^2}{m_\phi^2a^2}+2q \tag{22.52}$$

Tachyonic preheating in hybrid inflation 745

GWs from first-order phase transitions

19.1.8 22.4.1 Crossovers and phase transitions

19.2 Crossover/phase transition at the electroweak scale

19.3 Crossover/phase transition at the QCD scale

First-order phase transitions in cosmology

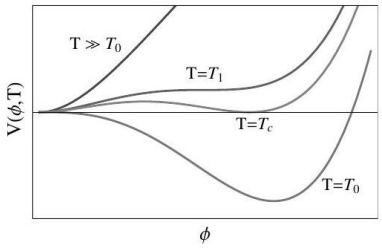
$$V(\phi, T) = \frac{1}{2}M^2(T)\phi^2 - \frac{1}{3}\delta(T)\phi^3 + \frac{1}{4}\lambda\phi^4 \tag{22.58}$$

$$M^2(T) = \gamma\left(T^2 - T_0^2\right) \tag{22.59}$$

$$\delta(T) = AT \tag{22.60}$$

$$T_1^2 = \frac{T_0^2}{1 - A^2/(4\lambda\gamma)} \tag{22.61}$$

$$T_c^2 = \frac{T_0^2}{1 - 2A^2/(9\lambda\gamma)} \tag{22.62}$$



$$\rho_* = \left[-V[\eta(T), T] + T\frac{d}{dT}V[\eta(T), T]\right]_{T=T_*} \tag{22.63}$$

$$\alpha = \frac{30\rho_*}{\pi^2 g_*\left(T_*\right)T_*^4} \tag{22.64}$$

$$\Gamma(t) = \Gamma_* e^{\beta(t-t_*)} \tag{22.65}$$

Thermal tunneling theory

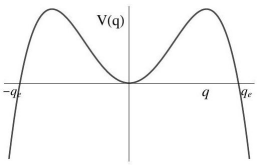
Tunneling and the bounce solution

$$V(q) = \frac{1}{2}\omega^2 q^2 + \frac{1}{2}gq^4 \tag{22.66}$$

$$\operatorname{tr} e^{-\beta H/\hbar} = \int\limits_{q(-\beta/2)=q(\beta/2)} [dq] e^{-S_E/\hbar} \tag{22.67}$$

$$S_E = \int\limits_{-\beta/2}^{\beta/2} dt \left[\frac{1}{2}\dot{q}^2 + V(q)\right] \tag{22.68}$$

$$E_0(g) = -\lim_{\beta\to\infty}\frac{\hbar}{\beta}\log \int\limits_{q(-\beta/2)=q(\beta/2)=0} [dq]e^{-S_E/\hbar} \tag{22.69}$$



$$\frac{1}{\hbar}\int\limits_{-\infty}^{\infty} dt \left[\frac{1}{2}\dot{q}^2 + \frac{1}{2}\omega^2 q^2 + \frac{1}{2}gq^4\right] \to \frac{1}{g\hbar}\int\limits_{-\infty}^{\infty} dt \left[\frac{1}{2}\dot{q}^2 + \frac{1}{2}\omega^2 q^2 + \frac{1}{2}q^4\right] \tag{22.70}$$

$$\psi(t) \propto e^{-i(\mathrm{Re}E_0)t - |\mathrm{Im}\,E_0|t} \tag{22.71}$$

$$q(t) = q_{\mathrm{cl}}(t) + \hbar^{1/2}\varphi(t) \tag{22.72}$$

$$\int\limits_{q(-\beta/2)=q(\beta/2)=0} [dq]e^{-S_E/\hbar} =$$

$$= e^{-S[q_{\mathrm{cl}}]/\hbar}\int [d\varphi]e^{-\frac{1}{2}\int\limits_{-\infty}^{\infty} dt_1 dt_2 \varphi(t_1)M(t_1,t_2)\varphi(t_2)} \tag{22.73}$$

$$\begin{aligned} M\left(t_1,t_2\right) &= \left.\frac{\delta^2 S}{\delta q\left(t_1\right)\delta q\left(t_2\right)}\right|_{q=q_{\mathrm{cl}}} \\ &= \left[-\frac{d^2}{dt_1^2} + V''\left[q_{\mathrm{cl}}\left(t_1\right)\right]\right]\delta\left(t_1 - t_2\right) \end{aligned} \tag{22.74}$$

$$\int\limits_{-\infty}^{\infty} dt' M\left(t,t'\right)\varphi_n\left(t'\right) = \lambda_n\varphi_n(t) \tag{22.75}$$

$$\langle f|g\rangle = \int\limits_{-\infty}^{\infty} dt f^*(t)g(t) \tag{22.76}$$

$$\langle\varphi_n|\varphi_m\rangle = \delta_{nm}u_n \tag{22.77}$$

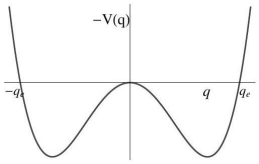
$$\varphi(t) = \sum_n \xi_n\varphi_n(t) \tag{22.78}$$

$$[d\varphi] = \mathcal{N}\prod_n \sqrt{\frac{u_n}{2\pi\hbar}}d\xi_n \tag{22.79}$$

$$\int_{-\infty}^{\infty} dt_1 dt_2 \varphi\left(t_1\right) M\left(t_1, t_2\right) \varphi\left(t_2\right)=\sum_n \lambda_n u_n \xi_n^2 \tag{22.80}$$

$$e^{-S\left[q_{\mathrm{c} 1}\right] / \hbar}\left(\operatorname{det}^{\prime} M\right)^{1 / 2} \tag{22.81}$$

$$\ddot{q}=V^{\prime} \tag{22.82}$$



$$\frac{1}{2} \dot{q}_{\mathrm{cl}}^2-V\left(q_{\mathrm{cl}}\right)=E \tag{22.83}$$

$$\begin{aligned} \beta & =2 \int_{q_1}^{q_2} d q \frac{1}{\sqrt{2(E+V)}} \\ & =2 \int_{q_1}^{q_2} d q \frac{1}{\sqrt{2 E+\omega^2 q^2+(1 / 2) g q^4}} \end{aligned} \tag{22.84}$$

$$\frac{1}{2} \dot{q}_{\mathrm{cl}}^2-V\left(q_{\mathrm{cl}}\right)=0 \tag{22.85}$$

$$S_{\mathrm{cl}}=\int_{-\infty}^{\infty} d t \dot{q}_{\mathrm{cl}}^2 \tag{22.86}$$

$$\varphi_0(t)=\dot{q}_{\mathrm{cl}}(t) \tag{22.87}$$

$$\left[-\frac{d^2}{d t^2}+V^{\prime \prime}\left[q_{\mathrm{cl}}(t)\right]\right] \dot{q}_{\mathrm{cl}}(t)=0 \tag{22.88}$$

$$\begin{aligned} 0 & =\frac{\partial}{\partial \gamma_i}\left(\frac{\delta S}{\delta q(t)}\right)_{q=q_{\mathrm{cl}}} \\ & =\int d t^{\prime}\left(\frac{\delta^2 S}{\delta q(t) \delta q\left(t^{\prime}\right)}\right)_{q=q_{\mathrm{cl}}} \frac{\partial q_{\mathrm{cl}}\left(t^{\prime} ; \gamma\right)}{\partial \gamma_i} \end{aligned} \tag{22.89}$$

$$\begin{aligned} q_{\mathrm{cl}}\left(t ; t_0\right)+\delta \xi_0 \varphi_0(t) & =q_{\mathrm{cl}}\left(t ; t_0\right)-\delta \xi_0 \frac{\partial}{\partial t_0} q_{\mathrm{cl}}\left(t ; t_0\right) \\ & =q_{\mathrm{cl}}\left(t ; t_0-\delta \xi_0\right) \end{aligned} \tag{22.90}$$

$$\varphi\left(t ; t_0\right) \equiv q(t)-q_{\mathrm{cl}}\left(t ; t_0\right) \tag{22.91}$$

$$1 = \int dt_0 \delta \left[f\left(t_0\right) \right] \left| \frac{\partial f}{\partial t_0} \right| \tag{22.92}$$

$$f\left(t_0\right) = \int dt \varphi\left(t; t_0\right) \dot{q}_{\mathrm{cl}}\left(t; t_0\right) \tag{22.93}$$

$$\begin{aligned} \frac{\partial f}{\partial t_0} = \int dt \Bigg[& -\dot{q}_{\mathrm{cl}}\left(t; t_0\right) \frac{\partial}{\partial t_0} q_{\mathrm{cl}}\left(t; t_0\right) \\ & + \varphi\left(t; t_0\right) \frac{\partial}{\partial t_0} \dot{q}_{\mathrm{cl}}\left(t; t_0\right) \Bigg] \end{aligned} \tag{22.94}$$

$$1 = \int dt_0 \delta\left(\xi_0\right) \left[1 - \frac{1}{u_0} \varphi\left(t; t_0\right) \ddot{q}_{\mathrm{cl}} \right] \tag{22.95}$$

$$1 \simeq \int dt_0 \delta\left(\xi_0\right) \tag{22.96}$$

$$\Gamma = A e^{-S_b} \tag{22.97}$$

$$A = \hbar \left(\frac{S_b}{2\pi\hbar} \right)^{1/2} \left| \frac{\det' \left[-\partial_t^2 + V''\left(\phi_b\right) \right]}{\det\left[-\partial_t^2 + V''(0) \right]} \right|^{-1/2} \tag{22.98}$$

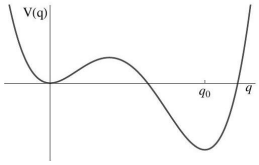
Tunneling in QFT at zero temperature

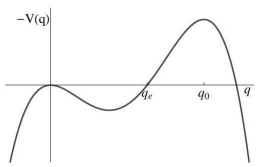
$$S_M = \int d^4x \left[-\frac{1}{2} \eta^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) \right] \tag{22.99}$$

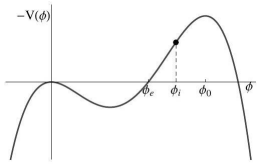
$$S_E = \int d^4 x_E \left[\frac{1}{2} \partial_\mu \phi \partial_\mu \phi + V(\phi) \right] \tag{22.100}$$

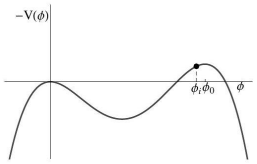
$$\lim_{\tau \to \pm\infty} \phi(\tau, \mathbf{x}) = 0 \tag{22.101}$$

$$\lim_{|\mathbf{x}| \to \infty} \phi(\tau, \mathbf{x}) = 0 \tag{22.102}$$









$$S_E = 2\pi^2 \int_0^\infty d\rho \rho^3 \left[\frac{1}{2} \left(\frac{d\phi}{d\rho} \right)^2 + V(\phi) \right] \tag{22.103}$$

$$\frac{d^2\phi}{d\rho^2} + \frac{3}{\rho} \frac{d\phi}{d\rho} = V'(\phi) \tag{22.104}$$

$$\lim_{\rho \to \infty} \phi(\rho) = 0 \tag{22.105}$$

$$\left. \frac{d\phi}{d\rho} \right|_{\rho=0} = 0 \tag{22.106}$$

$$\frac{\Gamma}{V} = A e^{-S_b/\hbar} \tag{22.107}$$

$$A = \left(\frac{S_b}{2\pi} \right)^2 \left| \frac{\det' \left[-\partial_\mu \partial_\mu + V'' \left(\phi_b \right) \right]}{\det \left[-\partial_\mu \partial_\mu + V''(0) \right]} \right|^{-1/2} \tag{22.111}$$

$$S_b = \frac{1}{4} \int d^4x \left(\partial_\mu \phi_b \right)^2 > 0 \tag{22.109}$$

Tunneling in QFT at finite temperature

$$\int \mathcal{D}\phi e^{-S_4} \tag{22.112}$$

$$S_4[\phi] = \int_0^{1/T} d\tau \int d^3x \left[\frac{1}{2} \partial_\mu \phi \partial_\mu \phi + V(\phi, T) \right] \tag{22.113}$$

$$S_4[\phi] \simeq \frac{1}{T} S_3[\phi] \tag{22.114}$$

$$S_3[\phi] = \int d^3x \left[\frac{1}{2} \left(\partial_i \phi \right)^2 + V(\phi, T) \right] \tag{22.115}$$

$$S_3 = 4\pi \int_0^\infty dr r^2 \left[\frac{1}{2} \left(\frac{d\phi}{dr} \right)^2 + V(\phi, T) \right] \tag{22.116}$$

$$\frac{d^2\phi}{dr^2} + \frac{2}{r} \frac{d\phi}{dr} = V'(\phi, T) \tag{22.117}$$

$$\lim_{r \to \infty} \phi(r) = 0 \tag{22.118}$$

$$\left. \frac{d\phi}{dr} \right|_{r=0} = 0 \tag{22.119}$$

$$\frac{\Gamma}{V} = Ae^{-S_{3,b}(T)/T} \tag{22.120}$$

$$A = T\left(\frac{S_3(T)}{2\pi T}\right)^{3/2}\left|\frac{\det'\left[-\nabla^2 + V''\left(\phi_b\right)\right]}{\det\left[-\nabla^2 + V''(0)\right]}\right|^{-1/2} \tag{22.121}$$

$$\frac{\Gamma}{V} \simeq cT^4 e^{-S_{3,b}(T)/T} \tag{22.122}$$

$$\int_0^{t_*} dt \frac{\Gamma}{VH^3(t)} = O(1) \tag{22.123}$$

$T \propto 1/a$
$dT/T = -Hdt$

$$\int_{T_*}^{\infty} \frac{dT}{T}\left(\frac{45}{4\pi^3 g_*(T)}\right)^2\left(\frac{M_{\rm Pl}}{T}\right)^4 e^{-S_3(T)/T} = O(1) \tag{22.124}$$

$$\frac{\det'\left[-\nabla^2 + V''\left(\phi_b\right)\right]}{\det\left[-\nabla^2 + V''(0)\right]}$$

$\left|\frac{\det'\left[-\nabla^2+V''(\phi_b)\right]}{\det[-\nabla^2+V''(0)]}\right|^{-1/2} \sim (\text{ mass })^3.$

$$\frac{S_{3,b}(T)}{T} \simeq \left(\frac{S_{3,b}(T)}{T}\right)_{t=t_*} + (t-t_*)\frac{d}{dt}\left(\frac{S_{3,b}(T)}{T}\right)_{t=t_*} \tag{22.125}$$

$$\beta = -\frac{d}{dt}\left(\frac{S_{3,b}(T)}{T}\right)_{t=t_*} \tag{22.126}$$

$$\Gamma(t) \simeq \Gamma_* e^{\beta(t-t_*)} \tag{22.127}$$

$dT/T = -Hdt$

$$\frac{\beta}{H_*} = T_*\frac{d}{dT}\left(\frac{S_{3,b}(T)}{T}\right)_{T=T_*} \tag{22.128}$$

$$\left(\frac{45}{4\pi^3 g_*\left(T_*\right)}\right)^2\left(\frac{M_{\rm Pl}}{T_*}\right)^4 e^{-S_3(T_*)/T_*} = O(1) \tag{22.129}$$

$$\frac{S_3\left(T_*\right)}{T_*} \simeq 4\log\frac{M_{\rm Pl}}{T_*} - 11.4 \tag{22.130}$$

$$4\pi\int_0^{\infty} drr^2 V(\phi,T) \simeq 4\pi\int_0^{R} drr^2 V(\phi,T) = -|\bar{V}(T)|\frac{4\pi R^3}{3} \tag{22.131}$$

$$4\pi\int_0^{\infty} drr^2\frac{1}{2}\left(\frac{d\phi}{dr}\right)^2 \simeq 4\pi R^2\int_0^{\infty} dr\frac{1}{2}\left(\frac{d\phi}{dr}\right)^2$$
$$\equiv 4\pi R^2 S_1(T) \tag{22.132}$$

$$S_3(R) = -|\bar{V}(T)|\frac{4\pi R^3}{3} + 4\pi R^2 S_1(T) \tag{22.133}$$

$$R(T) = \frac{2S_1(T)}{|\bar{V}(T)|} \tag{22.134}$$

Bubble dynamics and GW production

GWs from bubble collisions

(I'll make it shorter, just tell here the idea, maybe also some key formulas)

In a first-order phase transition, once these supercritical bubbles have nucleated, their subsequent dynamics depends on their environment. Conceptually, the simplest situation is that of vacuum bubbles, i.e. bubbles that are assumed to expand in the vacuum, neglecting their interaction with the cosmic fluid. In that case, the bubbles expand until their walls move at a speed close to the speed of light. The energy gained in the transition from the metastable state to the ground state is transferred to the kinetic energy of the bubble wall. As the bubble [24] Actually, when the bubble velocity v_b is not close to the speed of light, there are two different time- or lengthscales in the problem. One is the typical time-scale of the phase transition, β^{-1}, which can also identified with the duration of the phase transition. The other is the typical size of the bubbles near the end of the phase transition, when most bubble collisions take place, which is rather given by $R_* = v_b\beta^{-1}$. One should then ask which of these two scales is more appropriate for estimating the peak frequency through eq. (22.6). The answer is quite subtle, and depends on the detailed time evolution of the energy-momentum tensor of the source. It turns out that, in a first approximation, β^{-1} is appropriate for GWs from bubble collisions, while $v_b\beta^{-1}$ is appropriate for turbulence; see the Further Reading. A fit to the dependence on v_b for bubble collisions will be given in eqs. (22.144) and (22.145). expands, more and more regions of space convert to the ground state, and the wall becomes more and more energetic. At the same time, it also becomes thinner, and therefore the energy density stored in the wall increases very rapidly. Of course, as long as we have a single spherical bubble, this large energy cannot be converted to GWs. However, when two bubbles collide, spherical symmetry is broken and GW production can take place.

More realistically, in the cosmological setting in which we are eventually interested, the bubbles interact with the cosmic fluid. This interaction can in principle be quite complex, and results in friction over the bubble wall, which slows down and changes the bubble evolution, while at the same time the cosmic fluid is stirred by the expanding bubbles, with the generation of turbulent motions and acoustic waves. Indeed, three main mechanisms for GW production have been identified, namely GWs from bubble collisions, from turbulence and from colliding acoustic waves in the fluid. We examine them in turn.

19.8 GWs from turbulence in the cosmic fluid 771

19.9 GWs from sound waves in the cosmic fluid 772

2.2.5 Cosmic strings, (etc...)

19.9.2 22.5.1 Global and local strings 774

19.9.3 22.5.2 Effective description and Nambu-Goto action 778 19.9.4 22.5.3 String dynamicsCusps and kinks 782

19.9.5 22.5.4 Gravitational radiation from cosmic strings 787

19.10 Stochastic GW background from a cosmic string network 788

19.11GW bursts from cusps and kinks 792

19.11.1 22.6 Alternatives to inflation 794

19.11.2 22.7 Bounds on primordial GW backgrounds 799

19.11.3 22.7.1 The nucleosynthesis bound 799 19.11.4 22.7.2 Bounds on extra radiation from the CMB 806 19.11.5 22.7.3 Bounds from the CMB at large angles

2.2.6 On quantum cosmology, wavefunction of universe (????)

(all this is very speculative and I haven't met anyone who would prove me something from here.)

(чуть что есть отдельная про это запись)

Main Formulas of quantum cosmology (??????)

(хз)

Wheeler de Witt equation (??????)

(no one know what is it)

(from Quantum Theory of Gravity. I. The Canonical Theory Bryce S. DeWitt)

$$\pi = \frac{\delta L}{\delta \alpha_{,0}} = 0,$$
$$\pi^i = \frac{\delta L}{\delta \beta_{i,0}} = 0,$$
$$\pi^{ij} = \frac{\delta L}{\delta \gamma_{ij,0}} = -\gamma^{1/2}\left(K^{ij} - \gamma^{ij}K\right),$$
$$H = \int \left(\pi\alpha, 0 + \pi^i\beta_{i,0} + \pi^{ij}\gamma_{ij,0}\right) d^3x - L$$
$$= \int \left(\pi\alpha, 0 + \pi^i\beta_{i,0} + \alpha\mathcal{C} + \beta_i\chi^i\right) d^3x,$$

where

$$\mathcal{HC} \equiv \frac{1}{2}\gamma^{-1/2}\left(\gamma_{ik}\gamma_{jl} + \gamma_{il}\gamma_{jk} - \gamma_{ij}\gamma_{kl}\right)\pi^{ij}\pi^{kl} - \gamma^{1/2(3)}R$$
$$\equiv \gamma^{1/2}\left(K_{ij}K^{ij} - K^2 - {}^{(3)}R\right)$$
$$\chi^i \equiv -2\pi^{ij}{}_{.j} \equiv -2\pi^{ij}{}_{,j} - \gamma^{il}\left(2\gamma_{jl,k} - \gamma_{jk,l}\right)\pi^{jk}$$

Wave function of the universe by J. B. Hartle and S. W. Hawking

(they just wrote some formulas, no one know what do they mean)

$$\psi_0(x, 0) = N \int \delta x(\tau) \exp(-I[x(\tau)]),$$

$$\langle x,0|0,t'\rangle = \sum_n \psi_n(x)\bar{\psi}_n(0)\exp\left(iE_n t'\right)$$
$$= \int \delta x(t)\exp(iS[x(t)]),$$
$$\Psi\left[h_{ij}\right] = N\int_C \delta g(x)\exp\left(iS_E[g]\right)$$

S_E is the classical action for gravity including a cosmological constant Λ and the functional integral is over all four-geometries with a spacelike boundary on which the induced metric is h_{ij} and which to the past of that surface satisfy some appropriate condition to define the state. For this WF we have a diff eq - Wheeler-DeWitt equation which we shall derive from this point of view in Sec. II. With a simple choice of factor ordering it is

$$\left[-G_{ijkl}\frac{\delta^2}{\delta h_{ij}\delta h_{kl}} - {}^3R(h)h^{1/2} + 2\Lambda h^{1/2}\right]\Psi\left[h_{ij}\right] = 0,$$

where G_{ijkl} is the metric on superspace,

$$G_{ijkl} = \frac{1}{2}h^{-1/2}\left(h_{ik}h_{jl} + h_{il}h_{jk} - h_{ij}h_{kl}\right)$$

2.2.7 On bouncing universe model

(напишу потом мб.)

2.2.8 On Conformal cyclic cosmology

(идея Пенроуза, напишу потом. У Турока кст тоже статья про это есть.)

2.2.9 Other most typical models of cosmology

(потом мб пару разделов сделаю, пока на них только тут указания, потому что не были пока нужны мне они)

De Sitter space

(удалю скорее всего, просто еще раз прочитать нужно)

De Sitter space can be defined as a submanifold of a generalized Minkowski space of one higher dimension. Take Minkowski space $\mathbf{R}^{1,n}$ with the standard metric:

$$ds^2 = -dx_0^2 + \sum_{i=1}^{n} dx_i^2.$$

De Sitter space is the submanifold described by the hyperboloid of one sheet

$$-x_0^2 + \sum_{i=1}^{n} x_i^2 = \alpha^2,$$

where α is some nonzero constant with its dimension being that of length. The metric on de Sitter space is the metric induced from the ambient Minkowski metric. The induced metric is nondegenerate and has Lorentzian signature. (Note that if one replaces α^2 with $-\alpha^2$ in the above definition, one obtains a hyperboloid of two sheets. The induced metric in this case is positive-definite, and each sheet is a copy of hyperbolic n-space.

(допишу потом свойства и применения)

Гравитационная устойчивость изотропного мира по Ландау

Плоская анизотропная модель по Ландау

(все это потом буду додумывать.)
Пишем Уравнения Эйнштейна как

$$g_{00} = 1, \quad g_{0\alpha} = 0, \quad g_{\alpha\beta} = -\gamma_{\alpha\beta}(t).$$

Теперь мы можем воспользоваться уравнениями Эйнштейна в виде (97.11)-(97.13). Поскольку величины $\gamma_{\alpha\beta}$, а с ними и компоненты трехмерного тензора $\varkappa_{\alpha\beta} = \dot{\gamma}_{\alpha\beta}$ не зависят от координат x^α, то $R_{0\alpha} \equiv 0$. По той же причине $P_{\alpha\beta} \equiv 0$, и в результате уравнения гравитационного поля в пустоте сводятся к следующей системе:

$$\dot{\varkappa}_\alpha^\alpha + \frac{1}{2}\varkappa_\alpha^\beta \varkappa_\beta^\alpha = 0,$$
$$\frac{1}{\sqrt{\gamma}}\left(\sqrt{\gamma}\varkappa_\alpha^\beta\right)^\bullet = 0.$$

$$\sqrt{\gamma}\varkappa_\alpha^\beta = 2\lambda_\alpha^\beta, \qquad \lambda_\alpha^\beta = \text{ const}$$

$$\varkappa_\alpha^\alpha = \frac{\dot{\gamma}}{\gamma} = \frac{2}{\sqrt{\gamma}}\lambda_\alpha^\alpha,$$

$$\dot{\gamma}_{\alpha\beta} = \frac{2}{t}\lambda_\alpha^\gamma \gamma_{\gamma\beta}$$

Совокупность коэффициентов λ_α^γ можно рассматривать как матрицу некоторой линейной подстановки. Путем соответствующего линейного преобразования координат x^1, x^2, x^3 (или, что эквивалентно, величин $g_{1\beta}, g_{2\beta}, g_{3\beta}$) можно, вообще говоря, привести эту матрицу к диагональному виду. Обозначим ее главные значения посредством p_1, p_2, p_3 и будем считать, что все они вещественны и различны (о других случаях - см. ниже); единичные векторы в соответствующих главных направлениях пусть будут $\mathbf{n}^{(1)}, \mathbf{n}^{(2)}, \mathbf{n}^{(3)}$. Тогда решение уравнений (117.6) можно представить в виде

$$\gamma_{\alpha\beta} = t^{2p_1} n_\alpha^{(1)} n_\beta^{(1)} + t^{2p_2} n_\alpha^{(2)} n_\beta^{(2)} + t^{2p_3} n_\alpha^{(3)} n_\beta^{(3)}$$

(постоянные коэффициенты при степенях t можно обратить в единицу путем соответствующего выбора масштаба координат). Наконец, выбрав направления векторов $\mathbf{n}^{(1)}, \mathbf{n}^{(2)}, \mathbf{n}^{(3)}$ в качестве окончательного направления осей (назовем их x, y, z), приведем метрику к виду

$$ds^2 = dt^2 - t^{2p_1}dx^2 - t^{2p_2}dy^2 - t^{2p_3}dz^2$$

(E. Kasner, 1922). Здесь p_1, p_2, p_3 - любые три числа, удовлетворяющие двум соотношениям:

$$p_1 + p_2 + p_3 = 1, \quad p_1^2 + p_2^2 + p_3^2 = 1$$

The metric in $D > 3$ spacetime dimensions is

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \sum_{j=1}^{D-1} t^{2p_j} \left[\ \mathrm{d}x^j\right]^2,$$

and contains $D-1$ constants p_j, called the Kasner exponents. The metric describes a spacetime whose equal-time slices are spatially flat, however space is expanding or contracting at different

rates in different directions, depending on the values of the p_j. Test particles in this metric whose comoving coordinate differs by Δx^j are separated by a physical distance $t^{p_j}\Delta x^j$.

The Kasner metric is an exact solution to Einstein's equations in vacuum when the Kasner exponents satisfy the following Kasner conditions,

$$\sum_{j=1}^{D-1} p_j = 1 \qquad\qquad \sum_{j=1}^{D-1} p_j^2 = 1.$$

(???? потом разберу, это модель Казнера по Вики)

Колебательный режим приближения к особой точке по Ландау

(Абзац про достижения В. А. Белинский, Е. М. Лифшиц, И. М. Халатников, 1968, видимо, что-то глубокое, отдельно подумать про это нужно будет много.).

Особенность по времени в общем космологическом решении уравнений Эйнштейна по Ландау

(потом подумаю когда-то)

Факт существования особой точки в общем решении уравнений Эйнштейна был впервые доказан Пенроузом (R. Penrose, 1965) топологическими методами, которые, однако, не дают возможности установить конкретный аналитический характер особенности. Изложение этих методов и полученных с их помощью теорем см. Р. Пенроуз, Структура пространства-времени, «Мир», 1972.

2.2.10 Cosmological models that are considered wrong

(see https://www.astro.ucla.edu/~wright/errors.html, сейчас просто не до этого, там прямо много всякого)

Errors in Tired Light Cosmology

(see https://www.astro.ucla.edu/~wright/tiredlit.htm, there is a lot of info, now it is not important)

Errors in Irving Segal's Chronometric Cosmology

Errors in Eric Lerner's "The Big Bang Never Happened"

Errors in the Steady State and QSSC Models

Eddington did not predict the CMB

2.3 On general and more philosophical ideas about physics about cosmology

(тут то, про что если начнешь думать, то увязнешь в небытии и тупиковости разных идей. поэтому откидываю это подальше, выше - годные идеи, с которыми хоть что-то можно сделать!)

2.3.1 Cosmological paradoxes and philosophical problems

- Bentley's paradox: In a Newtonian universe, gravitation should pull all matter into a single point.

- Boltzmann brain: If the universe we observe resulted from a random thermodynamic fluctuation, it would be vastly more likely to be a simple one than the complex one we observe. The simplest case would be just a brain floating in vacuum, having the thoughts and sensations an ostensible observer has.

- Fermi paradox: If there are, as various arguments suggest, many other sentient species in the universe, then where are they? Should their presence not be obvious?

- Heat death paradox: If the universe were infinitely old, it would be in thermodynamic equilibrium, which contradicts what we observe.

- Olbers' paradox: Why is the night sky dark if there is an infinity of stars, covering every part of the celestial sphere?

2.3.2 What Could be in the Universe if Something Was Different?

(тут обзор разных фантастических идей, их полно.)

2.3.3 What is the fate of the universe?

(потом поизучаю, не до этого пока)

2.3.4 What Was in the Beginning of the Universe?

Как квантовая космология решает вопрос начала Вселенной? (????)

(думаю, это один из не многих вопросов, которые в 1 часть занести нужно.)

2.3.5 Nature of Time, Chronogeometry

Суть

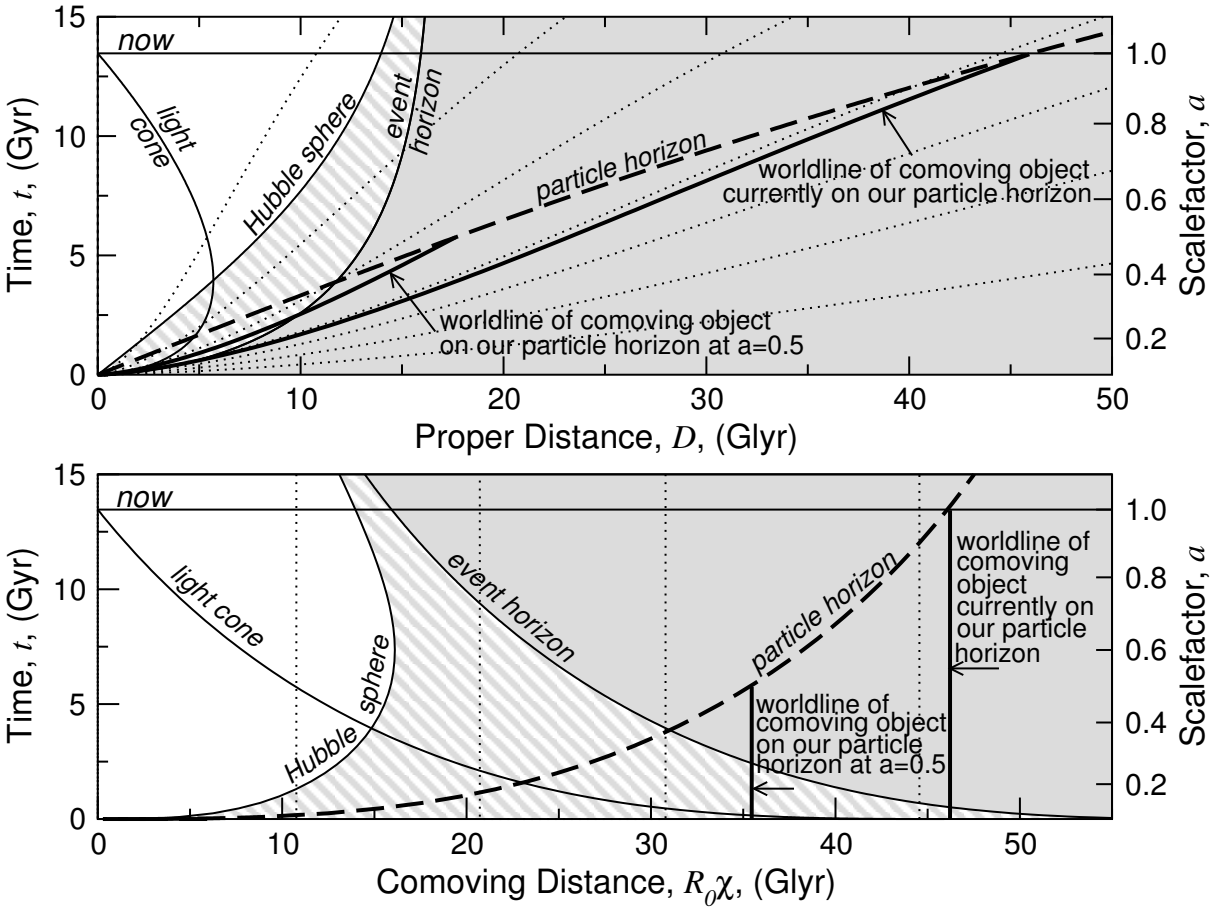
(пока в теорполе это пишу. Гуца и разных схожих авторов мб когда-то почитаю, вот куча записей частично под их влиянием и окажутся.)

2.3.6 Some Other Hypothetical Questions

(???)

О конце света

3 Some plots from different part of cosmology



Plots about relics of inflation

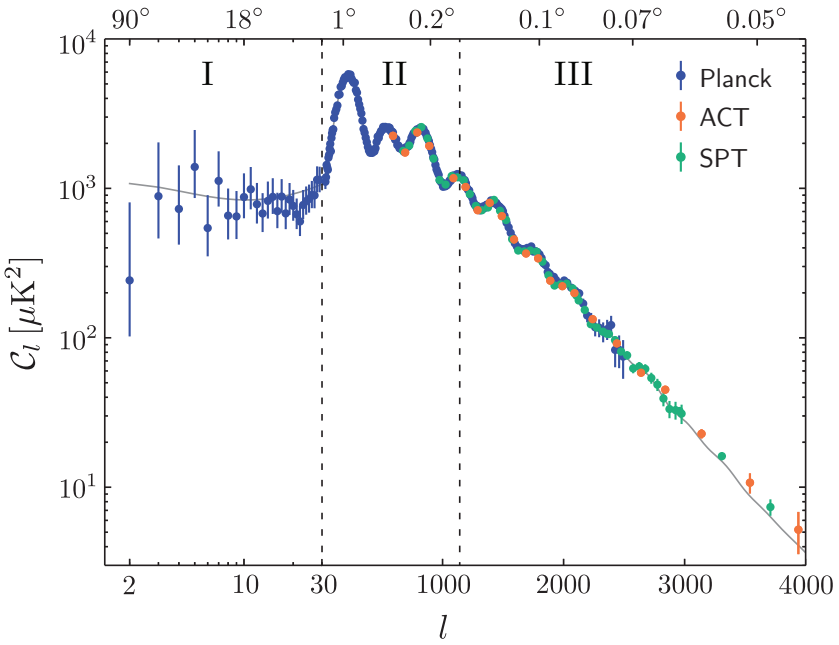


Figure 13: The angular variations of the CMB power spectrum are consequence of the dynamics of sound waves in the photon-baryon fluid. On large scales (region I), the fluctuations are frozen and we directly see the spectrum of the initial conditions. At intermediate scales (region II), we observe the oscillations of the fluid as captured at the moment of last-scattering. Finally, on small scales (region III), fluctuations are damped because their wavelengths are smaller than the mean free path of the photons.

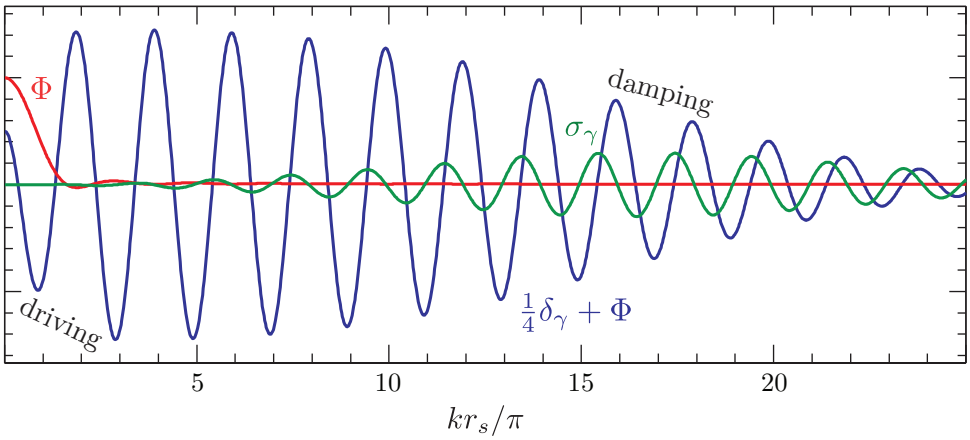


Figure 14: Acoustic oscillations with gravitational forcing and diffusion damping. For a mode that enters the sound horizon during radiation domination, the gravitational potential decays after horizon crossing and drives the acoustic amplitude higher. As the photon diffusion length increases and becomes comparable to the wavelength, viscosity σ_γ is generated and small-scale fluctuations are washed out.

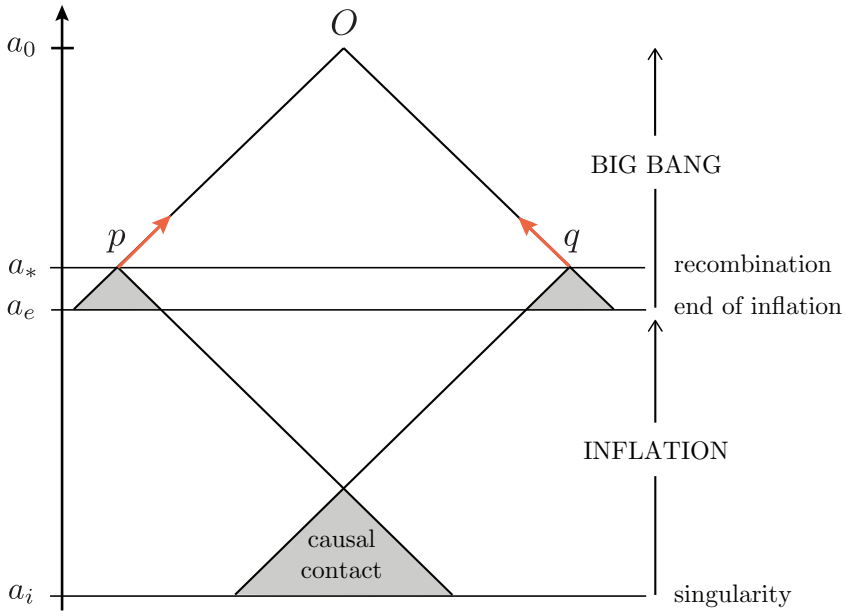


Figure 15: Illustration of the inflationary solution to the horizon problem in comoving coordinates (using conformal time on the vertical axis). The spacelike singularity of the standard Big Bang is replaced by the reheating surface, i.e. rather than marking the beginning of time it now simply corresponds to the transition from the end of inflation to the standard Big Bang evolution. All points in the CMB have overlapping past light cones and therefore originated from a causally connected region of space.

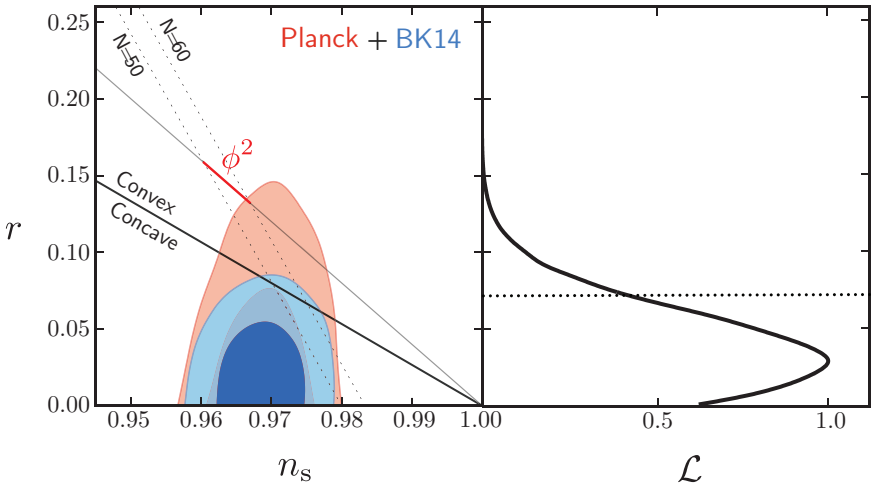
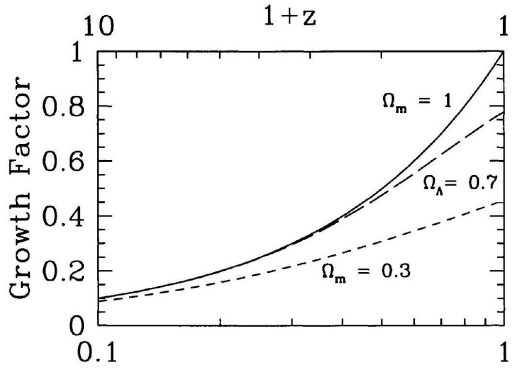
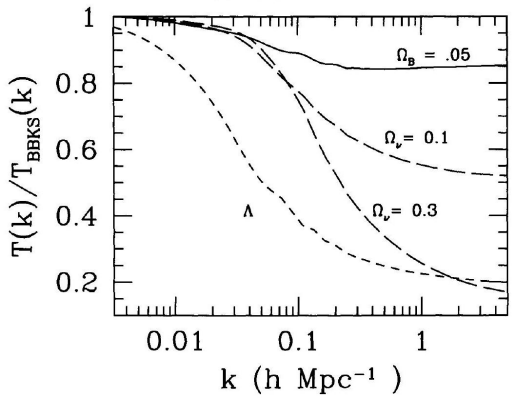
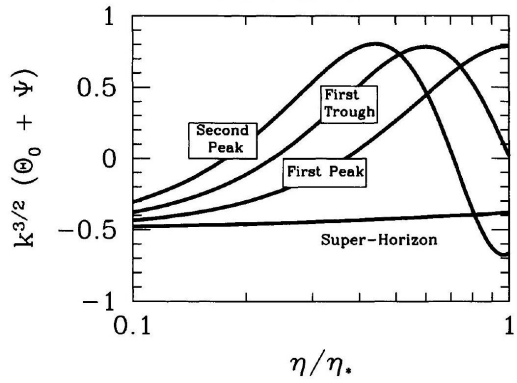


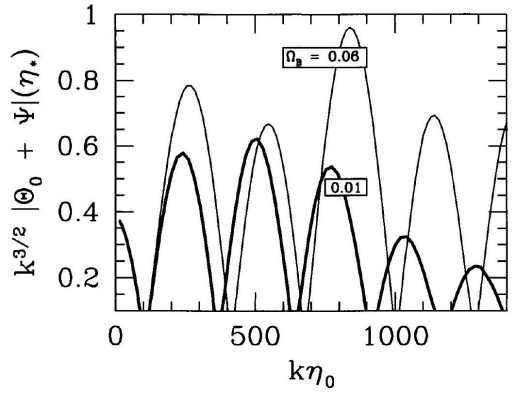
Figure 16: Current constraints on n_s and r from CMB measurements of Planck and BICEP [41].

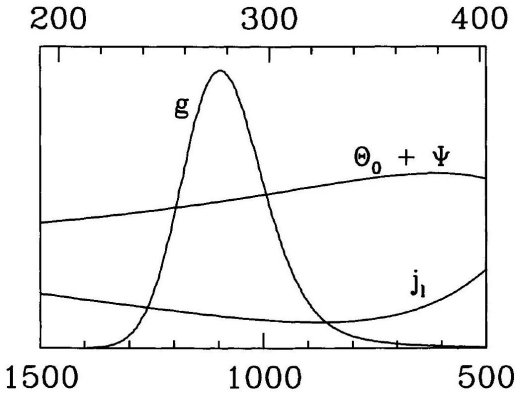
1

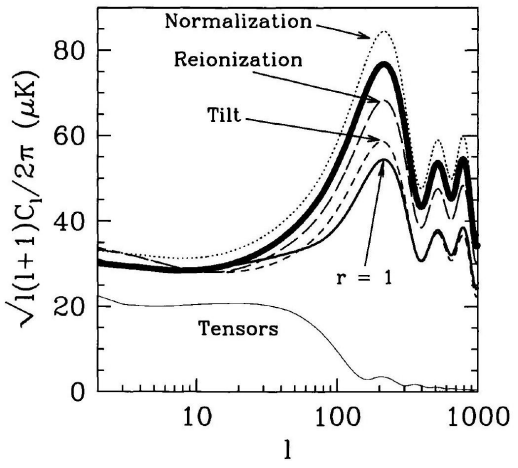

1+z
10
1
Growth Factor
Ω_m = 1
Ω_Λ= 0.7
Ω_m = 0.3
0.1
1


$T(k)/T_{BBKS}(k)$
Ω_B = .05
Ω_ν= 0.1
Ω_ν= 0.3
Λ
0.01
0.1
1
k (h Mpc^{-1})

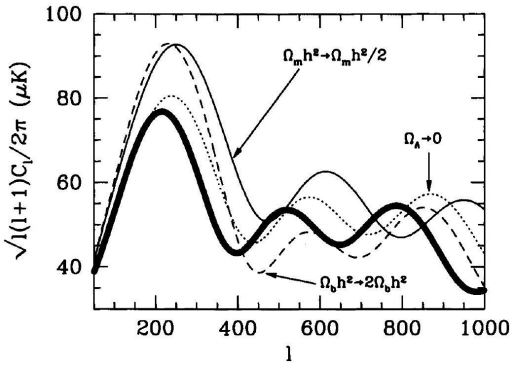

$k^{3/2}$ (Θ_0 + Ψ)
First Trough
Second Peak
First Peak
Super-Horizon
0.1
1
η/η_*

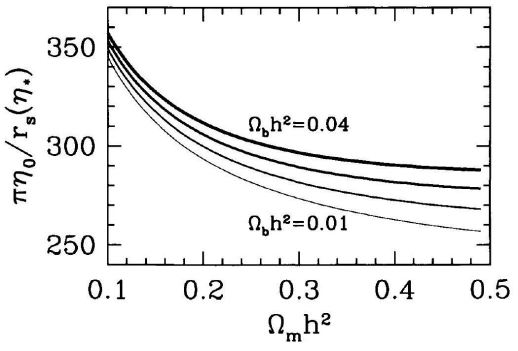

$k^{3/2}$ |Θ_0 + Ψ|(η_*)
Ω_B = 0.06
0.01
0
500
1000
kη_0

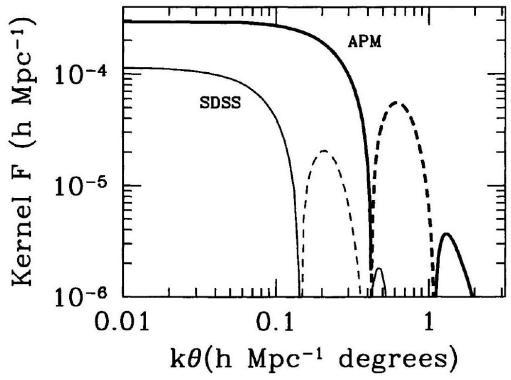

200
300
400
g
$\Theta_0 + \Psi$
j_l
1500
1000
500

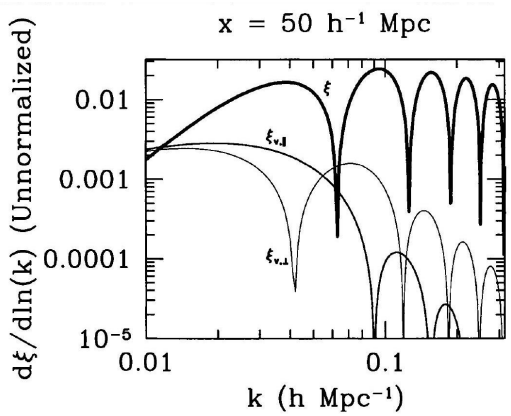

Normalization
Reionization
Tilt
r = 1
Tensors
$\sqrt{l(l+1)C_l/2\pi}$ (μK)
80
60
40
20
0
10
100
1000
l

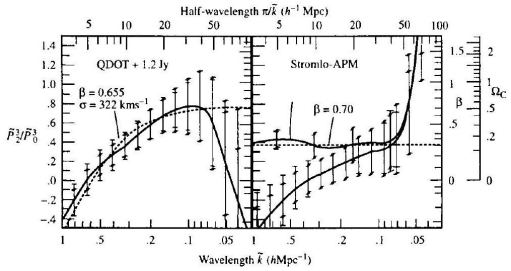
2

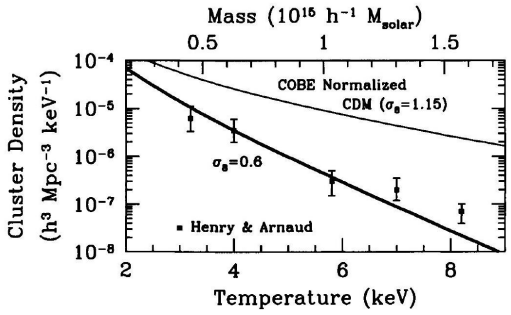

$\Omega_m h^2 \to \Omega_m h^2/2$
$\Omega_\Lambda \to 0$
$\Omega_b h^2 \to 2\Omega_b h^2$
$\sqrt{l(l+1)C_l/2\pi}$ (μK)
100
80
60
40
200
400
600
800
1000
l

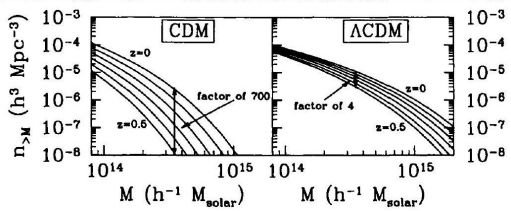

$\Omega_b h^2$=0.04
$\Omega_b h^2$=0.01
$\pi\eta_0/r_s(\eta_*)$
350
300
250
0.1
0.2
0.3
0.4
0.5
$\Omega_m h^2$


APM
SDSS
Kernel F (h Mpc⁻¹)
10⁻⁴
10⁻⁵
10⁻⁶
0.01
0.1
1
kθ(h Mpc⁻¹ degrees)

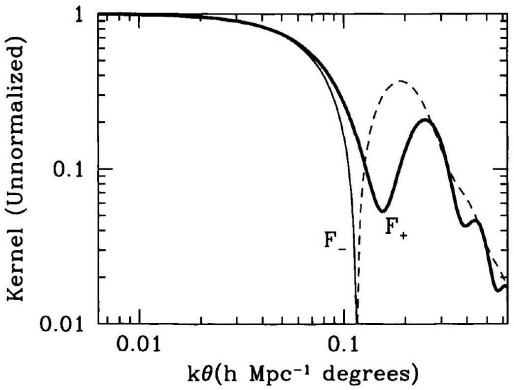

x = 50 h⁻¹ Mpc
ξ
ξv,∥
ξv,⊥
dξ/dln(k) (Unnormalized)
0.01
0.001
0.0001
10⁻⁵
0.01
0.1
k (h Mpc⁻¹)

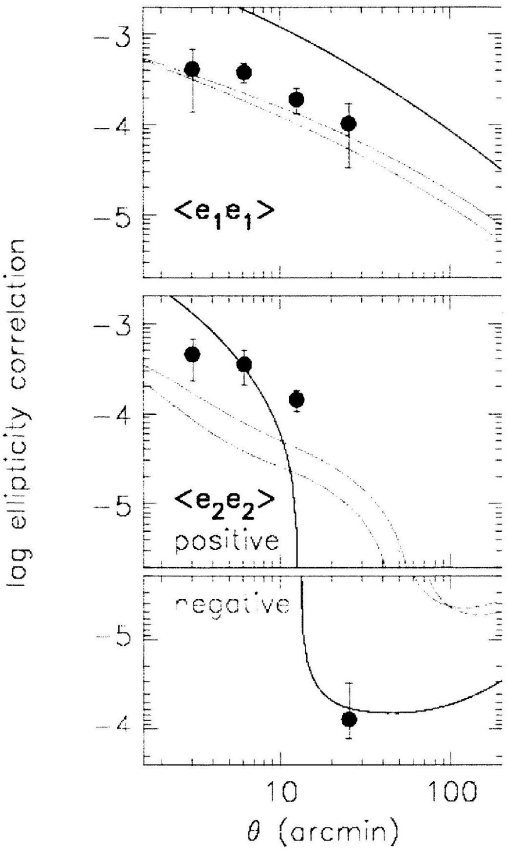

Half-wavelength π/k̄ (h⁻¹ Mpc)
QDOT + 1.2 Jy
β = 0.655
σ = 322 kms⁻¹
Stromlo-APM
β = 0.70
Ωc
β
Wavelength k̄ (hMpc⁻¹)

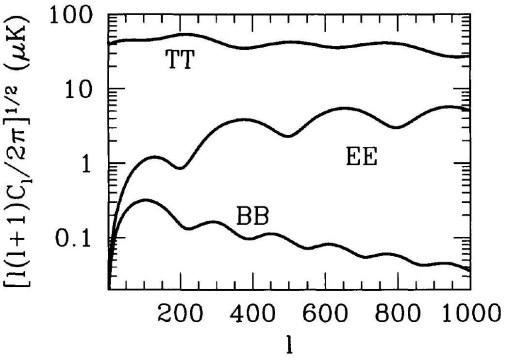

Mass (10¹⁵ h⁻¹ M_solar)
0.5
1
1.5
COBE Normalized
CDM (σ8=1.15)
σ8=0.6
Henry & Arnaud
Cluster Density
(h³ Mpc⁻³ keV⁻¹)
Temperature (keV)

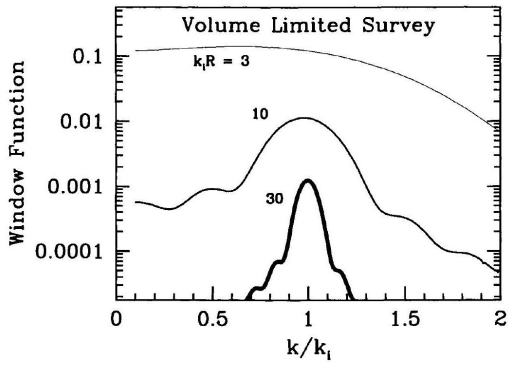

CDM
ΛCDM
z=0
z=0.6
factor of 700
factor of 4
n>M (h³ Mpc⁻³)
M (h⁻¹ M_solar)

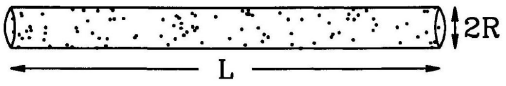
3

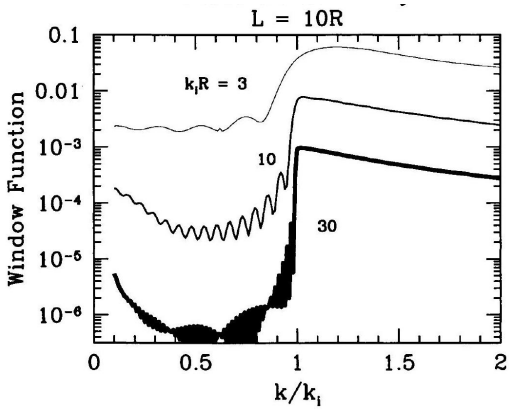

Kernel (Unnormalized)
1
0.1
0.01
0.01
0.1
F_
F+
kθ(h Mpc⁻¹ degrees)


log ellipticity correlation
−3
−4
−5
<e1e1>
<e2e2>
positive
negative
10
100
θ (arcmin)

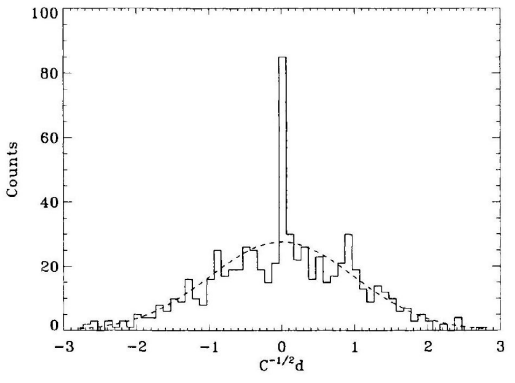

[l(l+1)C_l/2π]^1/2 (μK)
100
10
1
0.1
TT
EE
BB
200
400
600
800
1000
l

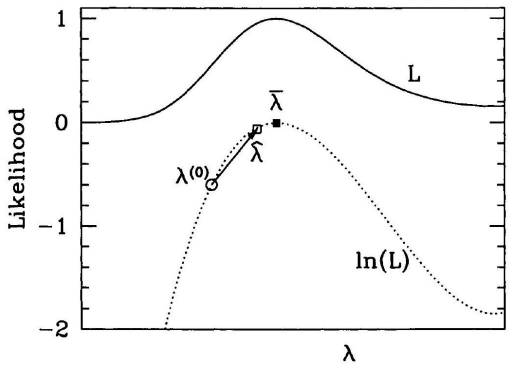

Volume Limited Survey
$k_iR = 3$
10
30
Window Function
0.1
0.01
0.001
0.0001
0
0.5
1
1.5
2
k/k_i

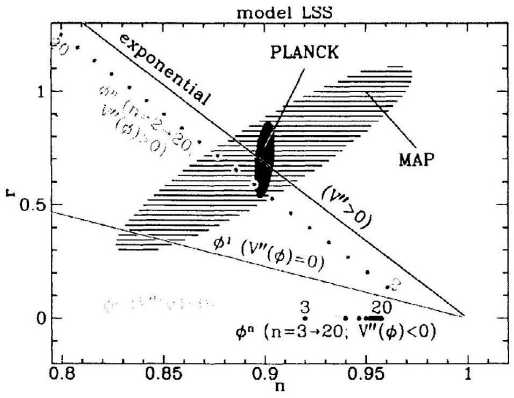

2R
L

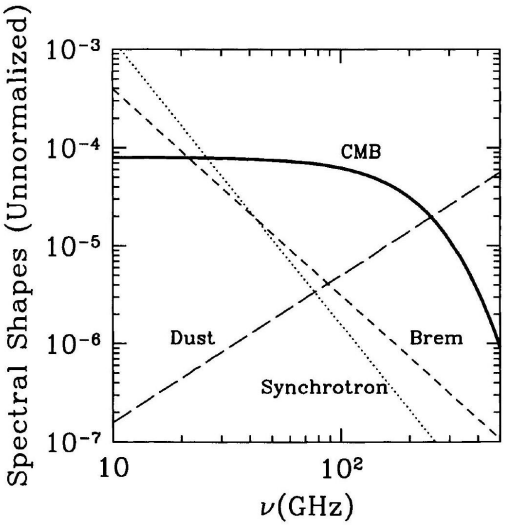

L = 10R
$k_iR = 3$
10
30
Window Function
0.1
0.01
10^{-3}
10^{-4}
10^{-5}
10^{-6}
0
0.5
1
1.5
2
k/k_i

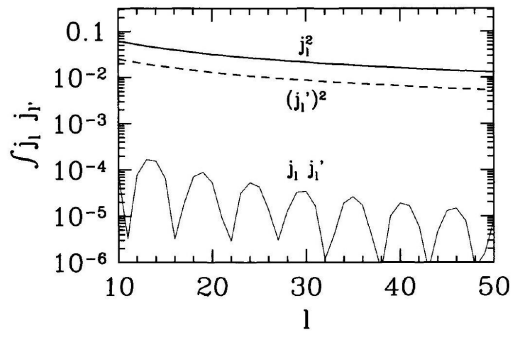
4


Counts
100
80
60
40
20
0
−3
−2
−1
0
1
2
3
$C^{-1/2}d$


Likelihood
1
0
−1
−2
L
$\bar{\lambda}$
$\hat{\lambda}$
$\lambda^{(0)}$
ln(L)
λ


model LSS
PLANCK
exponential
MAP
$(V''>0)$
ϕ^1 $(V''(\phi)=0)$
ϕ^n (n=3→20; $V''(\phi)<0$)
r
n
0.8
0.85
0.9
0.95
1
0.5
0
3
20


Spectral Shapes (Unnormalized)
CMB
Dust
Brem
Synchrotron
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10
10^2
ν(GHz)


$\int j_l\, j_{l'}$
j_l^2
$(j_l')^2$
$j_l\, j_l'$
0.1
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10
20
30
40
50
l

Part III
Fundamentals of Cosmology

4 General Fundamentals of Cosmology

4.1 General, Philosophical Approach

(see also a section about experiments, there will be a more detailed discussion of the facts which will be mentioned here)

4.1.1 Why In General We Can Describe a Universe (??)

(хз вообще, может ли такое быть)

Почему космология как наука в принципе возможна?

4.1.2 Flatness of the Universe

(???? I'll discuss some ideas)

Почему Вселенная плоская?

(она плоская, а почему так?)

4.1.3 Gravitation on the Scale of the Universe (???)

(??? what is known at all about it???)

4.1.4 Idea of a Cosmological Constant

(see below a big discussion of it)

Как догадаться до косм постоянной?

В уравнении она возможна

Физически (??)

Космологическая постоянная как врожденная кривизна пространства (????)

4.1.5 Gravitational Constant

(see also a note on gravity)

4.1.6 Time in Context of Cosmology (????)

(?? что вообще про него мы думаем?)

if Universe expands than there might be time dilation that will cause time to move slower. Why don't we account for it? (???????)

Imagine 2 particles and Univ accelerates. How would they know if Universe expands, or if they move? There is no way. So time might be changing also.

4.1.7 Idea of Dark Energy

Схожесть темной энергии и эфира (!?!??!)

(ну мб схожа, тут очень много про это подумать нужно.)

4.1.8 Idea of Dark Matter

(???)

4.1.9 Redshift

Redshift in a nutshell

(!!! Отдельный раздел про него сделаю! может, потом очень большим будет, пока просто subsubsection будет.)

Everything we know about the universe is inferred from the light we receive from distant objects. The light emitted by a distant galaxy can be viewed either quantum mechanically as freely-propagating photons, or classically as propagating electromagnetic waves. To interpret the observations correctly, we need to take into account that the wavelength of the light gets stretched (or, equivalently, the photons lose energy) by the expansion of the universe. We now quantify this effect.

Redshifting of photons.

In the quantum mechanical description, the wavelength of light is inversely proportional to the photon momentum, $\lambda = h/p$. Since according to (1.2.51) the momentum of a photon evolves as $a(t)^{-1}$, the wavelength scales as $a(t)$. Light emitted at time t_1 with wavelength λ_1 will be observed at t_0 with wavelength

$$\lambda_0 = \frac{a\left(t_0\right)}{a\left(t_1\right)}\lambda_1.$$

Since $a\left(t_0\right) > a\left(t_1\right)$, the wavelength of the light increases, $\lambda_0 > \lambda_1$.

Redshifting of classical waves.

We can derive the same result by treating light as classical electromagnetic waves. Consider a galaxy at a fixed comoving distance d. At a time τ_1, the galaxy emits a signal of short conformal duration $\Delta\tau$ (see fig. 1.5). According to (1.1.24), the light arrives at our telescopes at time $\tau_0 = \tau_1 + d$. The conformal duration of the signal measured by the detector is the same as at the source, but the physical time intervals are different at the points of emission and detection,

$$\Delta t_1 = a\left(\tau_1\right)\Delta\tau \quad \text{and} \quad \Delta t_0 = a\left(\tau_0\right)\Delta\tau$$

If Δt is the period of the light wave, the light is emitted with wavelength $\lambda_1 = \Delta t_1$ (in units where $c = 1$), but is observed with wavelength $\lambda_0 = \Delta t_0$, so that

$$\frac{\lambda_0}{\lambda_1} = \frac{a\left(\tau_0\right)}{a\left(\tau_1\right)}.$$

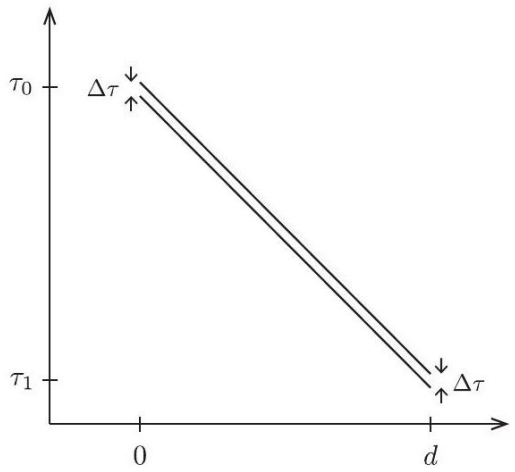


Figure 1.5: In conformal time, the period of a light wave ($\Delta\tau$) is equal at emission (τ_1) and at observation (τ_0). However, measured in physical time ($\Delta t = a(\tau)\Delta\tau$) the period is longer when it reaches us, $\Delta t_0 > \Delta t_1$. We say that the light has redshifted since its wavelength is now longer, $\lambda_0 > \lambda_1$.

It is conventional to define the redshift parameter as the fractional shift in wavelength of a photon emitted by a distant galaxy at time t_1 and observed on Earth today,

$$z \equiv \frac{\lambda_0 - \lambda_1}{\lambda_1}.$$

We then find

$$1 + z = \frac{a\left(t_0\right)}{a\left(t_1\right)}.$$

It is also common to define $a\left(t_0\right) \equiv 1$, so that

$$1 + z = \frac{1}{a\left(t_1\right)}.$$

Hubble's law.-For nearby sources, we may expand $a\left(t_1\right)$ in a power series,

$$a\left(t_1\right) = a\left(t_0\right)\left[1 + \left(t_1 - t_0\right) H_0 + \cdots\right],$$

where H_0 is the Hubble constant

$$H_0 \equiv \frac{\dot{a}\left(t_0\right)}{a\left(t_0\right)}.$$

Eq. (1.2.58) then gives $z = H_0\left(t_0 - t_1\right) + \cdots$. For close objects, $t_0 - t_1$ is simply the physical distance d (in units with $c = 1$). We therefore find that the redshift increases linearly with distance

$$z \simeq H_0 d.$$

The slope in a redshift-distance diagram (cf. fig. 1.8) therefore measures the current expansion rate of the universe, H_0. These measurements used to come with very large uncertainties. Since H_0 normalizes everything else (see below), it became conventional to define [11]

$$H_0 \equiv 100h\mathrm{kms}^{-1}\mathrm{Mpc}^{-1},$$

where the parameter h is used to keep track of how uncertainties in H_0 propagate into other cosmological parameters. Today, measurements of H_0 have become much more precise,[12]

$$h \approx 0.67 \pm 0.01.$$

4.1.10 Redshift by Landau

(лл2 пока)

вкратце о красном смещении

Theory

Основной характерной чертой всех рассмотренных решений является нестационарность метрики: радиус кривизны пространства является функцией времени.

Изменение же радиуса кривизны приводит к изменению всех вообще расстояний между телами в пространстве, как это видно уже из того обстоятельства, что элемент пространственного расстояния dl пропорционален a.

Так, при увеличении a в таком пространстве тела «разбегаются» друг от друга (в открытой модели увеличению a соответствуют $\eta > 0$, а в закрытой $0 < \eta < \pi$).

С точки зрения наблюдателя, находящегося на одном из них, дело будет выглядеть так, как если бы остальные тела двигались в радиальных направлениях, удаляясь от наблюдателя.

Скорость этого «разбегания» (в данный момент t) сама пропорциональна расстоянию между телами.

Это предсказание теории следует поставить в соответствие с фундаментальным астрономическим фактом - эффектом красного смещения линий В спектрах галактик.

Истолковав это смещение как доплеровское, мы приходим к заключению о «разбегании» галактик, т. е. о том, что в настоящее время Вселенная расширяется.

Свет в изотропном пространстве (??)

Рассмотрим распространение лучей света в изотропном пространстве.

Для этого проще всего воспользоваться тем, что вдоль мировой линии распространения светового сигнала интервал $ds = 0$. Точку,

из которой выходит луч света, выберем в качестве начала координат χ, θ, φ. Из соображений симметрии очевидно, что лучи будут распространяться «радиально», т. е. вдоль линии $\theta = \text{const}$, $\varphi = \text{const}$.

Полагая соответственно этому в (112.4) или (113.2)$d\theta = d\varphi = 0$ получим $ds^2 = a^2\,(d\eta^2 - d\chi^2)$. Приравнивая нулю, находим $d\eta = \pm d\chi$ или, интегрируя:

$$\chi = \pm\eta + \text{const}.$$

Знак плюс перед η соответствует лучу, распространяющемуся по направлению от начала координат, а знак минус - лучу, проходящему в начало координат.

В таком виде уравнение (114.1) применимо к распространению лучей как в открытой, так и в закрытой моделях.

С помощью формул предыдущих параграфов можно выразить отсюда проходимое лучом расстояние как функцию времени.

В открытой модели луч света, вышедший из некоторой точки, по мере своего распространения неограниченно удаляется от нее.

В закрытой же модели вышедший из исходной точки луч света в конце концов может дойти до «противоположного полюса» пространства (чему соответствует изменение χ от 0 до π); при дальнейшем распространении луч начнет приближаться к исходной точке.

Обходу луча «вокруг пространства» и возвращению в исходную точку соответствовало бы изменение χ от 0 до 2π. Из (114.1) мы видим, что при этом и η ДОЛЖНО было бы измениться на 2π, что, однако, невозможно (за исключением одного случая - выхода луча в момент, соответствующий $\eta = 0$).

Таким образом, луч не мог бы успеть возвратиться в исходную точку, обойдя «вокруг пространства».

Лучу, приходящему в точку наблюдения (начало координат), соответствует уравнение (114.1) со знаком минус перед η.

Если момент прихода луча в эту точку есть $t(\eta_0)$, то при $\eta = \eta_0$ должно быть $\chi = 0$, так что уравнение распространения таких лучей имеет вид

$$\chi = \eta_0 - \eta$$

Отсюда видно, что к наблюдателю, находящемуся в точке $\chi = 0$, могут дойти к моменту времени $t(\eta_0)$ лучи, вышедшие из точек, находящихся на «расстояниях», не превышающих $\chi = \eta_0$.

Этот результат, относящийся как к открытой, так и к закрытой моделям, весьма существен.

В каждый момент времени $t(\eta)$ в данной точке пространства физическому наблюдению доступно не все пространство, а лишь его часть, соответствующая $\chi \leqslant \eta$. С математической точки зрения «видимая область» пространства представляет собой сечение четырехмерного пространства-времени световым конусом.

Это сечение оказывается конечным как для открытой, так и для закрытой моделей (бесконечным же в открытой модели является сечение гиперповерхностью $t = \text{const}$, соответствующее пространству, рассматриваемому Во всех своих точках в один и тот же момент времени t), В этом смысле разница между открытой и закрытой моделями оказывается менее глубокой, чем это могло бы на первый взгляд показаться.

Чем более удалена от наблюдателя наблюдаемая им в данный момент времени область пространства, тем более ранним моментам времени она соответствует.

Представим себе сферическую поверхность, являющуюся геометрическим местом точек, из которых свет вышел В момент времени $t(\eta - \chi)$ и наблюдается в начале координат в момент $t(\eta)$. Площадь этой поверхности равна $4\pi a^2(\eta - \chi)\sin^2\chi$ (в закрытой модели) или $4\pi a^2(\eta - \chi)\,\mathrm{sh}^2\,\chi$ (в открытой модели).

По мере удаления от наблюдателя площадь «видимой сферы» сначала возрастает от нуля (при $\chi = 0$), затем достигает максимума, после чего снова уменьшается, обращаясь в нуль при $\chi = \eta($ где $a(\eta - \chi) = a(0) = 0)$. Это значит, что сечение световым конусом не только конечно, но и замкнуто.

Оно как бы замыкается в точке, «противоположной» наблюдателю; ее можно увидеть при наблюдении в любом направлении В пространстве.

В этой точке $\varepsilon \to \infty$, так что наблюдению доступна, в принципе, материя на всех ступенях ее эволюции.

Полное наблюдаемое количество материи равно в открытой модели интегралу

$$M_{\text{набл}} = 4\pi \int\limits_0^\eta \mu a^3\,\mathrm{sh}^2\,\chi d\chi$$

Подставив μa^3 из (113.7), Получим:

$$M_{\text{набл}} = \frac{3c^2 a_0}{2k}(\mathrm{sh}\,\eta\,\mathrm{ch}\,\eta - \eta)$$

Эта величина неограниченно возрастает при $\eta \to \infty$. В закрытой же модели возрастание $M_{\text{набл}}$ ограничено, разумеется, полной массой M; аналогичным образом получим в этом случае:

$$M_{\text{набл}} = \frac{M}{\pi}(\eta - \sin\eta\cos\eta)$$

По мере возрастания η от 0 до π эта величина возрастает от 0 до M В дальнейшем же увеличение $M_{\text{набл}}$ согласно полученной формуле фиКтивно и соответствует просто тому, что в «сжимающемся» мире удаленные тела наблюдались бы дважды (по свету, «обошедшему пространство» с двух сторон).

Рассмотрим теперь изменение частоты света при его распространении в изотропном пространстве.

Для этого замечаем предварительно следующее обстоятельство.

Пусть в некоторой точке пространства происходят два события, разделенные промежутком времени $dt = a(\eta)d\eta/$.

Если в моменты этих событий отправляются световые сигналы, наблюдаемые в другой точке пространства, то между моментами их наблюдения пройдет промежуток времени, соответствующий тому же изменению $d\eta$ величины η, что и в точке отправления.

Это следует непосредственно из уравнения (114.1), согласно которому изменение величины η за время распространения луча света из одной точки в другую зависит только от разности координат χ этих точек.

Но поскольку за время распространения сигнала радиус кривизны a изменится, то промежутки времени t между моментами отправления двух сигналов и моментами их наблюдения будут различными; отношение этих промежутков равно отношению соответствующих значений a. Отсюда следует, в частности, что и периоды световых колебаний, измеренные в мировом времени t, меняются вдоль луча пропорционально a.

Частота же света будет, очевидно, обратно пропорциональна a.

Таким образом, при распространении луча света вдоль него постоянно произведение $\omega a = \text{const}$. Предположим, что в момент времени $t(\eta)$ мы наблюдаем свет, излученный источником, находящимся на расстоянии, соответствующем определенному значению координаты χ. Моментом испускания этого света является, согласно (114.1), момент $t(\eta - \chi)$. Если ω_0 есть частота света в момент его испускания, то наблюдаемая нами частота ω равна, согласно (114.5),

$$\omega = \omega_0 \frac{a(\eta - \chi)}{a(\eta)}$$

При монотонно возрастающей функции $a(\eta)$ имеем $\omega < \omega_0$, т. е. происходит уменьшение частоты света.

Это значит, что при наблюдении спектра приходящего света все его линии должны оказаться смещенными в красную сторону по сравнению со спектрами тех же веществ В обычных условиях.

Это явление красного смещения представляет собой, по существу, эффект Доплера от взаимного «разбегания» галактик.

Величина красного смешения, измеряемая отношением ω/ω_0 сдвинутой частоты к несдвинутой, зависит (при данном моменте наблюдения) от расстояния, на котором находится наблюдаемый источник света (в соотношение (114.6) входит координата χ источника света).

При не слишком больших расстояниях можно разложить $a(\eta - \chi)$ в ряд по степеням χ, ограничившись первыми двумя членами:

$$\frac{\omega}{\omega_0} = 1 - \chi\frac{a'(\eta)}{a(\eta)}$$

(Штрих означает дифференцирование по η).

Далее замечаем, что произведение $\chi a(\eta)$ является здесь не чем иным, как расстоянием l до наблюдаемого источника.

Действительно, «радиальный» элемент длины равен $dl = ad\chi$; при интегрировании этого соотношения возникает Вопрос о том, каким способом физического наблюдения определяется расстояние, - в зависимости от этого надо брать значения a в разных точках пути интегрирования в разные моменты времени (интегрирование при $\eta = \text{const}$ соответствовало бы одновременному рассмотрению всех точек пути, что физически неосуществимо).

Но при «малых» расстояниях можно пренебречь изменением a вдоль пути интегрирования И написать просто $l = a\chi$ со значением a, взятым в момент наблюдения.

В результате находим для относительной величины изменения частоты (ее обозначают обычно буквой z) следующую формулу:

$$z = \frac{\omega_0 - \omega}{\omega_0} = \frac{H}{c} l$$

где введено обозначение

$$H = c\frac{a'(\eta)}{a^2(\eta)} = \frac{1}{a}\frac{da}{dt}$$

для так называемой постоянной Хаббла.

Эта величина не зависит при заданном моменте наблюдения от l.

Таким образом, относительный сдвиг спектральных линий пропорционален расстоянию до наблюдаемого источника света.

Рассматривая красное смещение как результат эффекта Доплера, можно определить скорости v галактик, с которыми они удаляются от наблюдателя.

Написав $z = v/c$ и сравнив с (114.7), находим

$$v = Hl$$

(эту формулу можно получить и непосредственно, вычисляя производную $v = d(a\chi)/dt$). Астрономические данные подтверждают закон (114.7), но определение значения постоянной Хаббла затрудняется неопределенностью в установлении масштаба космических расстояний, пригодного для удаленных галактик.

Принятое в настоящее время значение H составляет:

$$H \approx 0,8 \cdot 10^{-10} \text{год}^{-1} = 0,25 \cdot 10^{-17} \text{с}^{-1}$$

$$\frac{1}{H} \approx 4 \cdot 10^{17} \text{с} = 13 \cdot 10^{9} \text{ год.}$$

Это значение H соответствует увеличению «скорости разбегания» на 75 км/с на каждый мегапарсек расстояния.

Подставляя в уравнение (113.4)$\varepsilon = \mu c^2$ и $H = ca'/a^2$, получим для открытой модели следующее соотношение:

$$\frac{c^2}{a^2} = H^2 - \frac{8\pi k}{3}\mu$$

Комбинируя это уравнение с равенством

$$H = \frac{c \operatorname{sh}\eta}{a_0(\operatorname{ch}\eta - 1)^2} = \frac{c}{a}\operatorname{cth}\frac{\eta}{2}$$

получим:

$$\operatorname{ch}\frac{\eta}{2} = H\sqrt{\frac{3}{8\pi k\mu}}$$

Для закрытой модели аналогичным образом получим:

$$\frac{c^2}{a^2} = \frac{8\pi k}{3}\mu - H^2$$
$$\cos\frac{\eta}{2} = H\sqrt{\frac{3}{8\pi k\mu}}$$

Сравнивая (114.11) и (114.13), мы видим, что кривизна пространства отрицательна или положительна, смотря по тому, отрицательна Или положительна разность $8\pi k\mu/3 - H^2$. Эта разность обращается В нуль при $\mu = \mu_k$, где

$$\mu_k = \frac{3H^2}{8\pi k}$$

Со значением (114.10) получим $\mu_k \approx 1.10^{-29}$ г/см ^{3}. При современном состоянии астрономических сведений оценка средней плотности материи в пространстве может быть произведена лишь с весьма небольшой точностью.

Оценки

Для оценки, основанной на подсчете числа галактик и на их средней массе, принимают в настоящее время значение около 3.10$^-$31 г/см ^{3}. Это значение в 30 раз меньше μ_k и, таким образом, свидетельствовало бы в пользу открытой модели.

Однако, не говоря уже о недостаточной достоверности самого этого значения, следует Иметь В виду, что В нем не учитывается возможное существование межгалактического темного газа, учет которого мог бы существенно повысить среднюю плотность материи.

Отметим некоторое неравенство, которое оказывается Возможным ПОлучить при заданном значении величины Н.

Для открытой модели имеем $H = c\,\mathrm{sh}\,\eta/a_0(\mathrm{ch}\,\eta - 1)^2$, и отсюда

$$t = \frac{a_0}{c}(\mathrm{sh}\,\eta - \eta) = \frac{\mathrm{sh}\,\eta(\mathrm{sh}\,\eta - \eta)}{H(\mathrm{ch}\,\eta - 1)^2}$$

Поскольку $0 < \eta < \infty$, то должно быть

$$\frac{2}{3H} < t < \frac{1}{H}$$

Аналогично для закрытой модели получим:

$$t = \frac{\sin\eta(\eta - \sin\eta)}{H(1 - \cos\eta)^2}$$

Возрастанию $a(\eta)$ соответствует интервал $0 < \eta < \pi$; поэтому получаем:

$$0 < t < \frac{2}{3H}$$

Далее определим интенсивность I света, доходящего до наблюдателя от источника, находящегося на расстоянии, соответствующем определенному значению координаты χ. Плотность потока световой энергии в точке наблюдения обратно пропорциональна поверхности сферы, проведенной через рассматриваемую точку с центром в точке

нахождения источника; в пространстве отрицательной кривизны площадь поверхности сферы равна $4\pi a^2\mathrm{sh}^2\chi$. Свет, испущенный источником в течение времени $dt = a(\eta - \chi)d\eta/c$, будет приходить в точку наблюдения в течение времени $a(\eta)dt/a(\eta - \chi) = a(\eta)d\eta/$ с.

Поскольку Интенсивность определяется как поток световой энергии в единицу времени, то, следовательно, в I появится множитель $a(\eta-\chi)/a(\eta)$. Наконец, энергия волнового пакета пропорциональна частоте (см. (53.9)); Поскольку частота меняется при распространении света по закону (114.5), то это приведет к появлению в I еще одного множителя $a(\eta-\chi)/a(\eta)$. В результате окончательно получаем интенсивность В виде

$$I = \text{const} \cdot \frac{a^2(\eta-\chi)}{a^4(\eta)\,\mathrm{sh}^2\,\chi}$$

Для закрытой модели получилось бы аналогично:

$$I = \text{const} \cdot \frac{a^2(\eta-\chi)}{a^4(\eta)\sin^2\chi}$$

Этими формулами устанавливается зависимость видимой яркости наблюдаемого объекта от его расстояния (при заданной абсолютной Яркости).

При малых χ можно положить $a(\eta-\chi)\approx a(\eta)$, и тогда $I\sim 1/a^2(\eta)\chi^2 = 1/l^2$, т. е. обычный закон уменьшения интенсивности обратно пропорционально квадрату расстояния.

Наконец, рассмотрим вопрос о так называемых собственных движениях тел.

Говоря о плотности и движении материи, мы везде подразумевали усредненную плотность и усредненное движение; в частности, В той системе отсчета, которой мы все время пользуемся, скорость усредненного движения равна нулю.

Истинные же скорости тел обнаруживают некоторый разброс вокруг своего среднего значения.

С течением времени скорости собственного движения тел меняются.

Для определения закона этого изменения рассмотрим свободно движущееся тело и выберем начало координат в какой-либо точке его траектории.

Тогда траекторией будет являться радиальная линия $\theta=\text{const}$, $\varphi=\text{const}$.

Уравнение Гамильтона-Якоби (87.6) после подстановки значений g^{ik} примет виД

$$\left(\frac{\partial S}{\partial\chi}\right)^2 - \left(\frac{\partial S}{\partial\eta}\right)^2 + m^2c^2a^2(\eta) = 0$$

Поскольку в коэффициенты этого уравнения χ не входит (т. е. координата χ циклична), то имеет место закон сохранения $dS/d\chi=\text{const}$.

Импульс же p движущегося тела равен, по общему определению, $p=\partial S/\partial l=\partial S/a\partial\chi$. Таким образом, при движении тела остается постоянным произведение $pa=\text{const}$.

Вводя скорость v собственного движения тела согласно

$$p = mv/\sqrt{1-v^2/c^2}$$

получим:

$$\frac{va}{\sqrt{1-\frac{v^2}{c^2}}} = \text{const}.$$

Этим соотношением определяется закон изменения скоростей со временем.

По мере возрастания a скорости v монотонно падают.

ЛЛ2-114.1

Найти первые два члена разложения видимой яркости галактики как функции ее красного смещения; абсолютная яркость галактики меняется со временем по экспоненциальному закону $I_{\text{абс}} = \text{const}\cdot e^{\alpha t}$ (H. Robertson, 1955)

Р е ш е н и е.

Зависимость видимой яркости туманности, наблюдаемой в «момент» η, от расстояния χ дается (для закрытой модели) формулой

$$I = \text{const} \cdot e^{\alpha[t(\eta-\chi)-t(\eta)]} \frac{a^2(\eta-\chi)}{a^4(\eta)\sin^2\chi}$$

Красное смещение, определенное согласно (114.7):

$$z = \frac{\omega_0 - \omega}{\omega} = \frac{a(\eta) - a(\eta-\chi)}{a(\eta-\chi)}$$

Разлагая I и z по степеням χ(с функциями $a(\eta)$ и $t(\eta)$ из (112.9) – (112.10)) и исключая затем χ из получающихся выражений, находим в результате:

$$I = \text{const} \cdot \frac{1}{z^2}\left[1 - \left(1 - \frac{q}{2} + \frac{\alpha}{H}\right) z\right]$$

где введено обозначение

$$q = \frac{2}{1+\cos\eta} = \frac{\mu}{\mu_k} > 1$$

Для открытой модели получается такая же формула с

$$q = \frac{2}{1+\operatorname{ch}\eta} = \frac{\mu}{\mu_k} < 1$$

ЛЛ2-114.2

Найти первые члены разложения числа галактик, находящихся внутри «сферы» заданного радиуса, как функции от красного смещения на границе сферы (в пространственное распределение галактик предполагается однородным).

Р е ш е н и е.

Число N галактик, находящихся на «расстоянии» $\leqslant \chi$, есть (в закрытой модели)

$$N = \text{const} \cdot \int_0^\chi \sin^2\chi d\chi \approx \text{const} \cdot \chi^3$$

Подставляя сюда первые два члена разложения функции $\chi(z)$, получим:

$$N = \text{const} \cdot z^3\left[1 - \frac{3}{4}(2+q)z\right]$$

В таком виде эта формула справедлива и для открытой модели.

Обзор применений красного смещения (??)

(?? пока не знаю.)

5 Main Methods of Gravity and Mathematics for Cosmology

(самая суть и примеры, подробно - в дополнениях)

5.1 Geometry

5.1.1 Geodesics

Geodesics

How do particles evolve in the FRW spacetime? In the absence of additional non-gravitational forces, freely-falling particles in a curved spacetime move along geodesics. I will briefly remind you of some basic facts about geodesic motion in general relativity [8] and then apply it to the FRW spacetime (1.1.16).

Geodesic Equation

Consider a particle of mass m. In a curved spacetime it traces out a path $X^\mu(s)$. The fourvelocity of the particle is defined by

$$U^\mu \equiv \frac{dX^\mu}{ds}.$$

A geodesic is a curve which extremises the proper time $\Delta s/c$ between two points in spacetime. In the box below, I show that this extremal path satisfies the geodesic equation [9]

$$\frac{dU^\mu}{ds} + \Gamma^\mu_{\alpha\beta} U^\alpha U^\beta = 0,$$

where $\Gamma^\mu_{\alpha\beta}$ are the Christoffel symbols $\Gamma^\mu_{\alpha\beta} \equiv \frac{1}{2} g^{\mu\lambda} \left(\partial_\alpha g_{\beta\lambda} + \partial_\beta g_{\alpha\lambda} - \partial_\lambda g_{\alpha\beta}\right)$.

Derivation of the geodesic equation

Consider the motion of a massive particle between to points in spacetime A and B (see fig. 1.4). The relativistic action of the particle is

$$S = -m \int_A^B \mathrm{d}s.$$

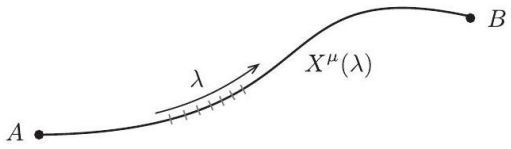


Figure 1.4: Parameterisation of an arbitrary path in spacetime, $X^\mu(\lambda)$.

We label each point on the curve by a parameter λ that increases monotonically from an initial value $\lambda(A) \equiv 0$ to a final value $\lambda(B) \equiv 1$. The action is a functional of the path $X^\mu(\lambda)$,

$$S\left[X^\mu(\lambda)\right] = -m \int_0^1 \left(g_{\mu\nu}(X) \dot{X}^\mu \dot{X}^\nu\right)^{1/2} \mathrm{d}\lambda \equiv \int_0^1 L\left[X^\mu, \dot{X}^\mu\right] \mathrm{d}\lambda,$$

where $\dot{X}^\mu \equiv dX^\mu/d\lambda$. The motion of the particle corresponds to the extremum of this action. The integrand in (1.2.29) is the Lagrangian L and it satisfies the Euler-Lagrange equation

$$\frac{d}{d\lambda}\left(\frac{\partial L}{\partial \dot{X}^\mu}\right) - \frac{\partial L}{\partial X^\mu} = 0.$$

The derivatives in (1.2.30) are

$$\frac{\partial L}{\partial \dot{X}^\mu} = -\frac{1}{L} g_{\mu\nu}\dot{X}^\nu \quad , \quad \frac{\partial L}{\partial X^\mu} = -\frac{1}{2L}\partial_\mu g_{\nu\rho}\dot{X}^\nu \dot{X}^\rho.$$

Before continuing, it is convenient to switch from the general parameterisation λ to the parameterisation using proper time s. (We could not have used s from the beginning since the value of s at B is different for different curves. The range of integration would then have been different for different curves.) Notice that

$$\left(\frac{ds}{d\lambda}\right)^2 = g_{\mu\nu}\dot{X}^\mu \dot{X}^\nu = L^2,$$

and hence $ds/d\lambda = L$. In the above equations, we can therefore replace $d/d\lambda$ with Ld/ds. The Euler-Lagrange equation then becomes

$$\frac{d}{ds}\left(g_{\mu\nu}\frac{dX^\nu}{ds}\right) - \frac{1}{2}\partial_\mu g_{\nu\rho}\frac{dX^\nu}{ds}\frac{dX^\rho}{ds} = 0.$$

Expanding the first term, we get

$$g_{\mu\nu}\frac{d^2X^\nu}{ds^2} + \partial_\rho g_{\mu\nu}\frac{dX^\rho}{ds}\frac{dX^\nu}{ds} - \frac{1}{2}\partial_\mu g_{\nu\rho}\frac{dX^\nu}{ds}\frac{dX^\rho}{ds} = 0.$$

In the second term, we can replace $\partial_\rho g_{\mu\nu}$ with $\frac{1}{2}\left(\partial_\rho g_{\mu\nu} + \partial_\nu g_{\mu\rho}\right)$ because it is contracted with an object that is symmetric in ν and ρ. Contracting (1.2.34) with the inverse metric and relabelling indices, we find

$$\frac{d^2X^\mu}{ds^2} + \Gamma^\mu_{\alpha\beta}\frac{dX^\alpha}{ds}\frac{dX^\beta}{ds} = 0.$$

Substituting (1.2.25) gives the desired result (1.2.26).

Covariant derivative

The derivative term in (1.2.26) can be manipulated by using the chain rule

$$\frac{d}{ds}U^\mu\left(X^\alpha(s)\right) = \frac{dX^\alpha}{ds}\frac{\partial U^\mu}{\partial X^\alpha} = U^\alpha\frac{\partial U^\mu}{\partial X^\alpha},$$

so that we get

$$U^\alpha\left(\frac{\partial U^\mu}{\partial X^\alpha} + \Gamma^\mu_{\alpha\beta}U^\beta\right) = 0.$$

The term in brackets is the covariant derivative of U^μ, i.e. $\nabla_\alpha U^\mu \equiv \partial_\alpha U^\mu + \Gamma^\mu_{\alpha\beta}U^\beta$. This allows us to write the geodesic equation in the following slick way: $U^\alpha\nabla_\alpha U^\mu = 0$. In the GR course you will derive this form of the geodesic equation directly by thinking about parallel transport.

Geodesic equation in 4-momenta

Using the definition of the four-momentum of the particle,

$$P^\mu = mU^\mu,$$

we may also write $U^\alpha\left(\frac{\partial U^\mu}{\partial X^\alpha} + \Gamma^\mu_{\alpha\beta}U^\beta\right) = 0$ as

$$P^\alpha \frac{\partial P^\mu}{\partial X^\alpha} = -\Gamma^\mu_{\alpha\beta} P^\alpha P^\beta.$$

For massless particles, the action (1.2.29) vanishes identically and our derivation of the geodesic equation breaks down. We don't have time to go through the more subtle derivation of the geodesic equation for massless particles. Luckily, we don't have to because the result is exactly the same as (1.2.39). [10] We only need to interpret P^μ as the four-momentum of a massless particle.

Accepting that the geodesic equation (1.2.39) applies to both massive and massless particles, we will move on. I will now show you how to apply the geodesic equation to particles in the FRW universe.

Geodesic Motion in FRW

To evaluate the r.h.s. of $P^\alpha \frac{\partial P^\mu}{\partial X^\alpha} = -\Gamma^\mu_{\alpha\beta} P^\alpha P^\beta$ we need to compute the Christoffel symbols for the FRW metric $\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\gamma_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j$.

All Christoffel symbols with two time indices vanish, i.e. $\Gamma^\mu_{00} = \Gamma^0_{0\beta} = 0$. The only non-zero components are

$$\Gamma^0_{ij} = a\dot{a}\gamma_{ij}, \quad \Gamma^i_{0j} = \frac{\dot{a}}{a}\delta^i_j, \quad \Gamma^i_{jk} = \frac{1}{2}\gamma^{il}\left(\partial_j\gamma_{kl} + \partial_k\gamma_{jl} - \partial_l\gamma_{jk}\right),$$

or are related to these by symmetry (note that $\Gamma^\mu_{\alpha\beta} = \Gamma^\mu_{\beta\alpha}$).

Derivation of Γ^0_{ij} (????)

-The Christoffel symbol with upper index equal to zero is

$$\Gamma^0_{\alpha\beta} = \frac{1}{2}g^{0\lambda}\left(\partial_\alpha g_{\beta\lambda} + \partial_\beta g_{\alpha\lambda} - \partial_\lambda g_{\alpha\beta}\right).$$

The factor $g^{0\lambda}$ vanishes unless $\lambda = 0$ in which case it is equal to 1. Therefore,

$$\Gamma^0_{\alpha\beta} = \frac{1}{2}\left(\partial_\alpha g_{\beta 0} + \partial_\beta g_{\alpha 0} - \partial_0 g_{\alpha\beta}\right).$$

The first two terms reduce to derivatives of g_{00} (since $g_{i0} = 0$). The FRW metric has constant g_{00}, so these terms vanish and we are left with

$$\Gamma^0_{\alpha\beta} = -\frac{1}{2}\partial_0 g_{\alpha\beta}.$$

The derivative is non-zero only if α and β are spatial indices, $g_{ij} = -a^2\gamma_{ij}$ (don't miss the sign!). In that case, we find

$$\Gamma^0_{ij} = a\dot{a}\gamma_{ij}$$

Analysis of geodesic equation

The homogeneity of the FRW background implies $\partial_i P^\mu = 0$, so that the geodesic equation (1.2.39) reduces to

$$\begin{aligned} P^0\frac{dP^\mu}{dt} &= -\Gamma^\mu_{\alpha\beta}P^\alpha P^\beta \\ &= -\left(2\Gamma^\mu_{0j}P^0 + \Gamma^\mu_{ij}P^i\right)P^j, \end{aligned}$$

[10] One way to think about massless particles is as the zero-mass limit of massive particles. A more rigorous derivation of null geodesics from an action principle can be found in Paul Townsend's Part III Black Holes lectures [arXiv:gr-qc/9707012]. where I have used (1.2.41) in the second line.

- The first thing to notice from (1.2.46) is that massive particles at rest in the comoving frame, $P^j = 0$, will stay at rest because the r.h.s. then vanishes,

$$P^j = 0 \quad \Rightarrow \quad \frac{dP^i}{dt} = 0.$$

- Next, we consider the $\mu = 0$ component of (1.2.46), but don't require the particles to be at rest. The first term on the r.h.s. vanishes because $\Gamma^0_{0j} = 0$. Using (1.2.41), we then find

$$E\frac{dE}{dt} = -\Gamma^0_{ij}P^iP^j = -\frac{\dot{a}}{a}p^2,$$

where we have written $P^0 \equiv E$ and defined the amplitude of the physical three-momentum as

$$p^2 \equiv -g_{ij}P^iP^j = a^2\gamma_{ij}P^iP^j.$$

Momenta of massive and massless particles (?)

Notice the appearance of the scale factor in (1.2.49) from the contraction with the spatial part of the FRW metric, $g_{ij} = -a^2\gamma_{ij}$. The components of the four-momentum satisfy the constraint $g_{\mu\nu}P^\mu P^\nu = m^2$, or $E^2 - p^2 = m^2$, where the r.h.s. vanishes for massless particles. It follows that $E\,\mathrm{d}E = p\,\mathrm{d}p$, so that (1.2.48) can be written as

$$\frac{\dot{p}}{p} = -\frac{\dot{a}}{a} \quad \Rightarrow \quad p \propto \frac{1}{a}.$$

We see that the physical three-momentum of any particle (both massive and massless) decays with the expansion of the universe.

For massless particles, eq. (1.2.50) implies

$$p = E \propto \frac{1}{a} \quad \text{(massless particles)}\ ,$$

i.e. the energy of massless particles decays with the expansion.

- For massive particles, eq. (1.2.50) implies

$$p = \frac{mv}{\sqrt{1-v^2}} \propto \frac{1}{a} \quad \text{(massive particles)}\ ,$$

where $v^i = dx^i/dt$ is the comoving peculiar velocity of the particles (i.e. the velocity relative to the comoving frame) and $v^2 \equiv a^2\gamma_{ij}v^iv^j$ is the magnitude of the physical peculiar velocity, cf. eq. (1.1.17). To get the first equality in (1.2.52), I have used

$$P^i = mU^i = m\frac{dX^i}{ds} = m\frac{dt}{ds}v^i = \frac{mv^i}{\sqrt{1-a^2\gamma_{ij}v^iv^j}} = \frac{mv^i}{\sqrt{1-v^2}}.$$

Eq. (1.2.52) shows that freely-falling particles left on their own will converge onto the Hubble flow.

5.1.2 Distances

For distant objects, we have to be more careful about what we mean by "distance":

Metric distance.

We first define a distance that isn't really observable, but that will be useful in defining observable distances. Consider the FRW metric in the form (1.1.21),

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\left[\mathrm{d}\chi^2 + S_k^2(\chi)\mathrm{d}\Omega^2\right],$$

where [13]

$$S_k(\chi) \equiv \begin{cases} R_0 \sinh\left(\chi/R_0\right) & k=-1 \\ \chi & k=0 \\ R_0 \sin\left(\chi/R_0\right) & k=+1 \end{cases}$$

The distance multiplying the solid angle element $\mathrm{d}\Omega^2$ is the metric distance,

$$d_m = S_k(\chi).$$

[11] A parsec (pc) is 3.26 light-years. Blame astronomers for the funny units in (6.3.29).

[12] Planck 2013 Results - Cosmological Parameters [arXiv:1303.5076].

[13] Notice that the definition of $S_k(\chi)$ contains a length scale R_0 after we chose to make the scale factor dimensionless, $a\left(t_0\right) \equiv 1$. This is achieved by using the rescaling symmetry $a \to \lambda a, \chi \to \chi/\lambda$, and $S_k^2 \to S_k^2/\lambda$. In a flat universe ($k=0$), the metric distance is simply equal to the comoving distance χ. The comoving distance between us and a galaxy at redshift z can be written as

$$\chi(z) = \int_{t_1}^{t_0} \frac{\mathrm{d}t}{a(t)} = \int_0^z \frac{\mathrm{d}z}{H(z)},$$

where the redshift evolution of the Hubble parameter, $H(z)$, depends on the matter content of the universe (see §1.3). We emphasize that the comoving distance and the metric distance are not observables.

Luminosity distance.

Type IA supernovae are called 'standard candles' because they are believed to be objects of known absolute luminosity L (= energy emitted per second). The observed flux F (= energy per second per receiving area) from a supernova explosion can then be used to infer its (luminosity) distance. Consider a source at a fixed comoving distance χ. In a static Euclidean space, the relation between absolute luminosity and observed flux is

$$F = \frac{L}{4\pi\chi^2}.$$

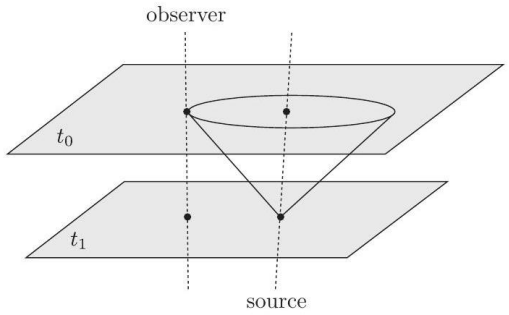


Figure 1.6: Geometry associated with the definition of luminosity distance.
In an FRW spacetime, this result is modified for three reasons:

1. At the time t_0 that the light reaches the Earth, the proper area of a sphere drawn around the supernova and passing through the Earth is $4\pi d_m^2$. The fraction of the light received in a telescope of aperture A is therefore $A/4\pi d_m^2$.

2. The rate of arrival of photons is lower than the rate at which they are emitted by the redshift factor $1/(1+z)$.

3. The energy E_0 of the photons when they are received is less than the energy E_1 with which they were emitted by the same redshift factor $1/(1+z)$.

Hence, the correct formula for the observed flux of a source with luminosity L at coordinate distance χ and redshift z is

$$F = \frac{L}{4\pi d_m^2 (1+z)^2} \equiv \frac{L}{4\pi d_L^2},$$

where we have defined the luminosity distance, d_L, so that the relation between luminosity, flux and luminosity distance is the same as in (1.2.69). Hence, we find

$$d_L = d_m(1+z).$$

Angular diameter distance.

-Sometimes we can make use of 'standard rulers', i.e. objects of known physical size D. (This is the case, for example, for the fluctuations in the CMB.) Let us assume again that the object is at a comoving distance χ and the photons which we observe today were emitted at time t_1. A naive astronomer could decide to measure the distance d_A to the object by measuring its angular size $\delta\theta$ and using the Euclidean formula for its distance, [14]

$$d_A = \frac{D}{\delta\theta}.$$

This quantity is called the angular diameter distance. The FRW metric (1.1.23) implies

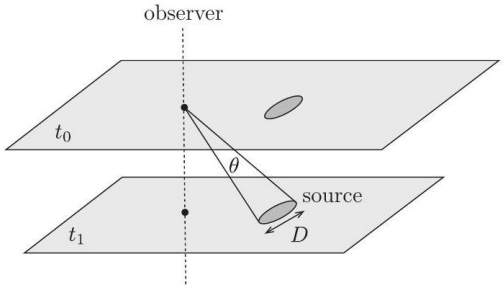


Figure 1.7: Geometry associated with the definition of angular diameter distance.

the following relation between the physical (transverse) size of the object and its angular size on the sky

$$D = a\,(t_1)\,S_k(\chi)\delta\theta = \frac{d_m}{1+z}\delta\theta.$$

Hence, we get

$$d_A = \frac{d_m}{1+z}.$$

The angular diameter distance measures the distance between us and the object when the light was emitted. We see that angular diameter and luminosity distances aren't independent, but related by

$$d_A = \frac{d_L}{(1+z)^2}.$$

Fig. 1.8 shows the redshift dependence of the three distance measures d_m, d_L, and d_A. Notice that all three distances are larger in a universe with dark energy (in the form of a cosmological constant Λ) than in one without. This fact was employed in the discovery of dark energy (see fig. 1.9 in §1.3.3).

5.2 Differential Geometry for Cosmology

(подробно все это - в записи по диффгему)

5.2.1 Killing Vector (??)

(??? discuss it for cosmology)

6 Some Basic Models

6.1 Friedman Universe

6.1.1 Cosmic parameters

The universe is filled with a mixture of different matter components. It is useful to classify the different sources by their contribution to the pressure:

- Matter

We will use the term "matter" to refer to all forms of matter for which the pressure is much smaller than the energy density, $|P| \ll \rho$. As we will show in Chapter 3, this is the case for a gas of non-relativistic particles (where the energy density is dominated by the mass). Setting $P = 0$ in (1.3.98) gives

$$\rho \propto a^{-3}.$$

This dilution of the energy density simply reflects the expansion of the volume $V \propto a^3$.

- Dark matter. Most of the matter in the universe is in the form of invisible dark matter. This is usually thought to be a new heavy particle species, but what it really is, we don't know.

- Baryons. Cosmologists refer to ordinary matter (nuclei and electrons) as baryons. [17]

[17] Of course, this is technically incorrect (electrons are leptons), but nuclei are so much heavier than electrons that most of the mass is in the baryons. If this terminology upsets you, you should ask your astronomer friends what they mean by "metals". 21 1. Geometry and Dynamics

- Radiation

We will use the term "radiation" to denote anything for which the pressure is about a third of the energy density, $P = \frac{1}{3}\rho$. This is the case for a gas of relativistic particles, for which the energy density is dominated by the kinetic energy (i.e. the momentum is much bigger than the mass). In this case, eq. (1.3.98) implies

$$\rho \propto a^{-4}.$$

The dilution now includes the redshifting of the energy, $E \propto a^{-1}$.

- Photons. The early universe was dominated by photons. Being massless, they are always relativistic. Today, we detect those photons in the form of the cosmic microwave background.

- Neutrinos. For most of the history of the universe, neutrinos behaved like radiation. Only recently have their small masses become relevant and they started to behave like matter.

- Gravitons. The early universe may have produced a background of gravitons (i.e. gravitational waves, see §6.5.2). Experimental efforts are underway to detect them.

- Dark energy

We have recently learned that matter and radiation aren't enough to describe the evolution of the universe. Instead, the universe today seems to be dominated by a mysterious negative pressure component, $P = -\rho$. This is unlike anything we have ever encountered in the lab. In particular, from eq. (1.3.98), we find that the energy density is constant,

$$\rho \propto a^{0}.$$

Since the energy density doesn't dilute, energy has to be created as the universe expands. [18]

- Vacuum energy. In quantum field theory, this effect is actually predicted! The ground state energy of the vacuum corresponds to the following stress-energy tensor

$$T^{\rm vac}_{\mu\nu} = \rho_{\rm vac} g_{\mu\nu}.$$

Comparison with eq. (1.3.93), show that this indeed implies $P_{\rm vac} = -\rho_{\rm vac}$. Unfortunately, the predicted size of $\rho_{\rm vac}$ is completely off,

$$\frac{\rho_{\rm vac}}{\rho_{\rm obs}} \sim 10^{120}.$$

- Something else? The failure of quantum field theory to explain the size of the observed dark energy has lead theorists to consider more exotic possibilities (such as time-varying dark energy and modifications of general relativity). In my opinion, none of these ideas works very well.

[18] In a gravitational system this doesn't have to violate the conservation of energy. It is the conservation equation (1.3.98) that counts. Cosmological constant. - The left-hand side of the Einstein equation (1.3.76) isn't uniquely defined. We can add the term $-\Lambda g_{\mu\nu}$, for some constant Λ, without changing the conservation of the stress tensor, $\nabla^\mu T_{\mu\nu} = 0$ (recall, or check, that $\nabla^\mu g_{\mu\nu} = 0$). In other words, we could have written the Einstein equation as

$$G_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}.$$

Einstein, in fact, did add such a term and called it the cosmological constant. However, it has become modern practice to move this term to the r.h.s. and treat it as a contribution to the stress-energy tensor of the form

$$T^{(\Lambda)}_{\mu\nu} = \frac{\Lambda}{8\pi G} g_{\mu\nu} \equiv \rho_\Lambda g_{\mu\nu}.$$

This is of the same form as the stress-energy tensor from vacuum energy, eq. (1.3.102).

Summary

Most cosmological fluids can be parameterised in terms of a constant equation of state: $w = P/\rho$. This includes cold dark matter ($w = 0$), radiation ($w = 1/3$) and vacuum energy ($w = -1$). In that case, the solutions to (1.3.98) scale as

$$\rho \propto a^{-3(1+w)},$$

and hence

$$\rho \propto \begin{cases} a^{-3} & \text{matter} \\ a^{-4} & \text{radiation} \\ a^{0} & \text{vacuum} \end{cases}.$$

6.1.2 Spacetime Curvature

We want to relate these matter sources to the evolution of the scale factor in the FRW metric (1.1.14). To do this we have to compute the Einstein tensor on the l.h.s. of the Einstein equation (1.3.76),

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}.$$

We will need the Ricci tensor

$$R_{\mu\nu} \equiv \partial_\lambda \Gamma^\lambda_{\mu\nu} - \partial_\nu \Gamma^\lambda_{\mu\lambda} + \Gamma^\lambda_{\lambda\rho}\Gamma^\rho_{\mu\nu} - \Gamma^\rho_{\mu\lambda}\Gamma^\lambda_{\nu\rho}$$

and the Ricci scalar

$$R = R^\mu{}_\mu = g^{\mu\nu} R_{\mu\nu}.$$

Again, there is a lot of beautiful geometry behind these definitions. We will simply keep pluggingand-playing: given the Christoffel symbols (1.2.41) nothing stops us from computing (1.3.109).

We don't need to calculate $R_{i0} = R_{0i}$, because it is a 3-vector, and therefore must vanish due to the isotropy of the Robertson-Walker metric. (Try it, if you don't believe it!) The non-vanishing components of the Ricci tensor are

$$R_{00} = -3\frac{\ddot{a}}{a},$$
$$R_{ij} = -\left[\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 + 2\frac{k}{a^2}\right] g_{ij}.$$

Notice that we had to find $R_{ij} \propto g_{ij}$ to be consistent with homogeneity and isotropy.

Derivation of R_{00}.

Setting $\mu = \nu = 0$ in (1.3.109), we have

$$R_{00} = \partial_\lambda \Gamma^\lambda_{00} - \partial_0 \Gamma^\lambda_{0\lambda} + \Gamma^\lambda_{\lambda\rho}\Gamma^\rho_{00} - \Gamma^\rho_{0\lambda}\Gamma^\lambda_{0\rho},$$

Since Christoffels with two time-components vanish, this reduces to

$$R_{00} = -\partial_0 \Gamma^i_{0i} - \Gamma^i_{0j}\Gamma^j_{0i}.$$

where in the second line we have used that Christoffels with two time-components vanish. Using $\Gamma^i_{0j} = (\dot{a}/a)\delta^i_j$, we find

$$R_{00} = -\frac{d}{dt}\left(3\frac{\dot{a}}{a}\right) - 3\left(\frac{\dot{a}}{a}\right)^2 = -3\frac{\ddot{a}}{a}.$$

Derivation of R_{ij}

Evaluating (1.3.112) is a bit tedious. A useful trick is to compute $R_{ij}(\boldsymbol{x} = 0) \propto \delta_{ij} \propto g_{ij}(\boldsymbol{x} = 0)$ using (1.1.9) and then transform the resulting relation between 3-tensors to general $\boldsymbol{x}$.

We first read off the spatial metric from (1.1.9),

$$\gamma_{ij} = \delta_{ij} + \frac{k x_i x_j}{1 - k\,(x_k x^k)}.$$

The key point is to think ahead and anticipate that we will set $\boldsymbol{x} = 0$ at the end. This allows us to drop many terms. You may be tempted to use $\gamma_{ij}(\boldsymbol{x} = 0) = \delta_{ij}$ straight away. However, the Christoffel symbols contain a derivative of the metric and the Riemann tensor has another derivative, so there will be terms in the final answer with two derivatives acting on the metric. These terms get a contribution from the second term in (1.3.116). However, we can ignore the denominator in the second term of γ_{ij} and use

$$\gamma_{ij} = \delta_{ij} + k x_i x_j.$$

The difference in the final answer vanishes at $\boldsymbol{x} = 0$ [do you see why?]. The derivative of (1.3.117) is

$$\partial_l \gamma_{ij} = k\left(\delta_{li} x_j + \delta_{lj} x_i\right).$$

With this, we can evaluate

$$\Gamma^i_{jk} = \frac{1}{2}\gamma^{il}\left(\partial_j \gamma_{kl} + \partial_k \gamma_{jl} - \partial_l \gamma_{jk}\right).$$

The inverse metric is $\gamma^{ij} = \delta^{ij} - k x^i x^j$, but the second term won't contribute when we set $\boldsymbol{x} = 0$ in the end [do you see why?], so we are free to use $\gamma^{ij} = \delta^{ij}$. Using (1.3.118) in (1.3.119), we then get

$$\Gamma^i_{jk} = kx^i\delta_{jk}.$$

This vanishes at $\boldsymbol{x} = 0$, but its derivative does not

$$\Gamma^i_{jk}(\boldsymbol{x}=0) = 0 \quad , \quad \partial_l\Gamma^i_{jk}(\boldsymbol{x}=0) = k\delta^i_l\delta_{jk}.$$

We are finally ready to evaluate the Ricci tensor R_{ij} at $\boldsymbol{x} = 0$

$$R_{ij}(\boldsymbol{x}=0) \equiv \underbrace{\partial_\lambda\Gamma^\lambda_{ij} - \partial_j\Gamma^\lambda_{i\lambda}}_{(A)} + \underbrace{\Gamma^\lambda_{\lambda\rho}\Gamma^\rho_{ij} - \Gamma^\rho_{i\lambda}\Gamma^\lambda_{j\rho}}_{(B)}.$$

Let us first look at the two terms labelled (B). Dropping terms that are zero at $\boldsymbol{x} = 0$, I find

$$\begin{aligned}(B) &= \Gamma^l_{l0}\Gamma^0_{ij} - \Gamma^0_{il}\Gamma^l_{j0} - \Gamma^l_{i0}\Gamma^0_{jl} \\ &= 3\frac{\dot{a}}{a}a\dot{a}\delta_{ij} - a\dot{a}\delta_{ij}\frac{\dot{a}}{a}\delta^l_j - \frac{\dot{a}}{a}\delta^l_j a\dot{a}\delta_{jl} \\ &= \dot{a}^2\delta_{ij}.\end{aligned}$$

The two terms labelled (A) in (1.3.122) can be evaluated by using (1.3.121),

$$\begin{aligned}(A) &= \partial_0\Gamma^0_{ij} + \partial_l\Gamma^l_{ij} - \partial_j\Gamma^l_{il} \\ &= \partial_0(a\dot{a})\delta_{ij} + k\delta^l_l\delta_{ij} - k\delta^l_j\delta_{il} \\ &= \left(a\ddot{a} + \dot{a}^2 + 2k\right)\delta_{ij}.\end{aligned}$$

Hence, I get

$$\begin{aligned}R_{ij}(\boldsymbol{x}=0) &= (A) + (B) \\ &= \left[a\ddot{a} + 2\dot{a}^2 + 2k\right]\delta_{ij} \\ &= -\left[\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 + 2\frac{k}{a^2}\right]g_{ij}(\boldsymbol{x}=0).\end{aligned}$$

As a relation between tensors this holds for general $\boldsymbol{x}$, so we get the promised result (1.3.112). To be absolutely clear, I will never ask you to reproduce a nasty computation like this.

The Ricci scalar is

$$R = -6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2}\right]$$

Exercise.-Verify eq. (1.3.126).

The non-zero components of the Einstein tensor $G^\mu{}_\nu \equiv g^{\mu\lambda}G_{\lambda\nu}$ are

$$\begin{aligned}G^0_0 &= 3\left[\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2}\right], \\ G^i_j &= \left[2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2}\right]\delta^i_j.\end{aligned}$$

Exercise.-Verify eqs. (1.3.127) and (1.3.128).

6.1.3 Friedmann Equations

Equations

We combine eqs. (1.3.127) and (1.3.128) with stress-tensor (1.3.92), to get the Friedmann equations,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2},$$
$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P).$$

Here, ρ and P should be understood as the sum of all contributions to the energy density and pressure in the universe. We write ρ_r for the contribution from radiation (with ρ_γ for photons and ρ_ν for neutrinos), ρ_m for the contribution by matter (with ρ_c for cold dark matter and ρ_b for baryons) and ρ_Λ for the vacuum energy contribution. The first Friedmann equation is often written in terms of the Hubble parameter, $H \equiv \dot{a}/a$,

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}.$$

Let us use subscripts ' 0 ' to denote quantities evaluated today, at $t = t_0$. A flat universe ($k = 0$) corresponds to the following critical density today

$$\rho_{\text{crit },0} = \frac{3H_0^2}{8\pi G} = 1.9 \times 10^{-29} h^2 \text{ grams cm}^{-3}$$
$$= 2.8 \times 10^{11} h^2 M_\odot \text{Mpc}^{-3}$$
$$= 1.1 \times 10^{-5} h^2 \text{ protons cm}^{-3}.$$

We use the critical density to define dimensionless density parameters

$$\Omega_{I,0} \equiv \frac{\rho_{I,0}}{\rho_{\text{crit },0}}.$$

The Friedmann equation (1.3.131) can then be written as

$$H^2(a) = H_0^2 \left[\Omega_{r,0}\left(\frac{a_0}{a}\right)^4 + \Omega_{m,0}\left(\frac{a_0}{a}\right)^3 + \Omega_{k,0}\left(\frac{a_0}{a}\right)^2 + \Omega_{\Lambda,0}\right],$$

where we have defined a "curvature" density parameter, $\Omega_{k,0} \equiv -k/\left(a_0 H_0\right)^2$. It should be noted that in the literature, the subscript ' 0 ' is normally dropped, so that e.g. Ω_m usually denotes the matter density today in terms of the critical density today. From now on we will follow this convention and drop the ' 0 ' subscripts on the density parameters. We will also use the conventional normalization for the scale factor, $a_0 \equiv 1$. Eq. (1.3.134) then becomes

$$\frac{H^2}{H_0^2} = \Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda.$$

Λ CDM

Observations (see figs. 1.9 and 1.10) show that the universe is filled with radiation (' r '), matter ($'m$ ') and dark energy (' Λ '):

$$|\Omega_k| \leq 0.01, \quad \Omega_r = 9.4 \times 10^{-5}, \quad \Omega_m = 0.32, \quad \Omega_\Lambda = 0.68.$$

The equation of state of dark energy seems to be that of a cosmological constant, $w_\Lambda \approx -1$. The matter splits into 5% ordinary matter (baryons, ' b ') and 27% (cold) dark matter (CDM, ' c '):

$$\Omega_b = 0.05, \quad \Omega_c = 0.27.$$

We see that even today curvature makes up less than 1% of the cosmic energy budget. At earlier times, the effects of curvature are then completely negligible (recall that matter and radiation scale as a^{-3} and a^{-4}, respectively, while the curvature contribution only increases as a^{-2}). For the rest of these lectures, I will therefore set $\Omega_k \equiv 0$. In Chapter 2, we will show that inflation indeed predicts that the effects of curvature should be minuscule in the early universe (see also Problem Set 2).

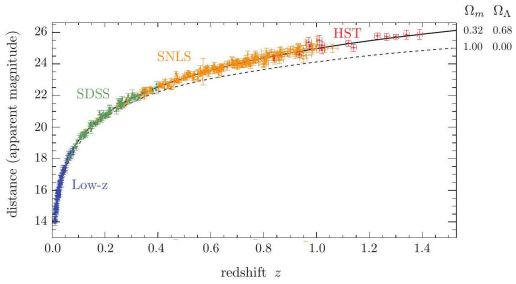


Figure 1.9: Type IA supernovae and the discovery dark energy. If we assume a flat universe, then the supernovae clearly appear fainter (or more distant) than predicted in a matter-only universe ($\Omega_m = 1.0$). (SDSS = Sloan Digital Sky Survey; SNLS = SuperNova Legacy Survey; HST = Hubble Space Telescope.)

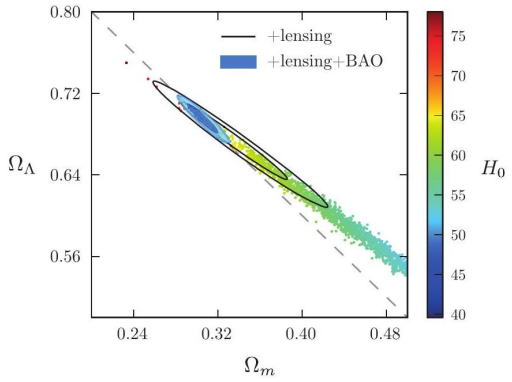


Figure 1.10: A combination CMB and LSS observations indicate that the spatial geometry of the universe is flat. The energy density of the universe is dominated by a cosmological constant. Notice that the CMB data alone cannot exclude a matter-only universe with large spatial curvature. The evidence for dark energy requires additional input.

Dynamics

The dynamics of the universe is determined by the Einstein equation

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}.$$

This relates the Einstein tensor $G_{\mu\nu}$ (a measure of the "spacetime curvature" of the FRW universe) to the stress-energy tensor $T_{\mu\nu}$ (a measure of the "matter content" of the universe). We

[14] This formula assumes $\delta\theta \ll 1$ (in radians) which is true for all cosmological objects. 17 1. Geometry and Dynamics

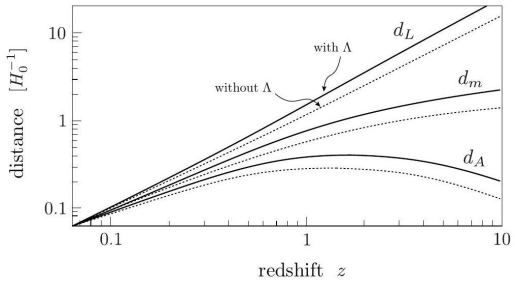


Figure 1.8: Distance measures in a flat universe, with matter only (dotted lines) and with 70% dark energy (solid lines). In a dark energy dominated universe, distances out to a fixed redshift are larger than in a matter-dominated universe.

will first discuss possible forms of cosmological stress-energy tensors $T_{\mu\nu}$(§1.3.1), then compute the Einstein tensor $G_{\mu\nu}$ for the FRW background (§1.3.2), and finally put them together to solve for the evolution of the scale factor $a(t)$ as a function of the matter content (§1.3.3).

Matter Sources

We first show that the requirements of isotropy and homogeneity force the coarse-grained stressenergy tensor to be that of a perfect fluid,

$$T_{\mu\nu} = (\rho + P)U_\mu U_\nu - P g_{\mu\nu},$$

where ρ and P are the energy density and the pressure of the fluid and U^μ is its four-velocity (relative to the observer).

Number Density

In fact, before we get to the stress-energy tensor, we study a simpler object: the number current four-vector N^μ. The $\mu = 0$ component, N^0, measures the number density of particles, where for us a "particle" may be an entire galaxy. The $\mu = i$ component, N^i, is the flux of the particles in the direction x^i. Isotropy requires that the mean value of any 3-vector, such as N^i, must vanish, and homogeneity requires that the mean value of any 3 -scalar [15], such as N^0, is a function only of time. Hence, the current of galaxies, as measured by a comoving observer, has the following components

$$N^0 = n(t), \quad N^i = 0,$$

where $n(t)$ is the number of galaxies per proper volume as measured by a comoving observer. A general observer (i.e. an observer in motion relative to the mean rest frame of the particles), would measure the following number current four-vector

$$N^\mu = nU^\mu$$

where $U^\mu \equiv dX^\mu/ds$ is the relative four-velocity between the particles and the observer. Of course, we recover the previous result (1.3.78) for a comoving observer, $U^\mu = (1, 0, 0, 0)$. For

[15] A 3-scalar is a quantity that is invariant under purely spatial coordinate transformations. $U^\mu = \gamma\,(1, v^i)$, eq. (1.3.79) gives the correctly boosted results. For instance, you may recall that the boosted number density is γn. (The number density increases because one of the dimensions of the volume is Lorentz contracted.)

The number of particles has to be conserved. In Minkowski space, this implies that the evolution of the number density satisfies the continuity equation

$$\dot{N}^0 = -\partial_i N^i,$$

or, in relativistic notation,

$$\partial_\mu N^\mu = 0.$$

Eq. (1.3.81) is generalised to curved spacetimes by replacing the partial derivative ∂_μ with a covariant derivative ∇_μ, [16]

$$\nabla_\mu N^\mu = 0.$$

Eq. (1.3.82) reduces to (1.3.81) in the local intertial frame.

Covariant derivative. - The covariant derivative is an important object in differential geometry and it is of fundamental importance in general relativity. The geometrical meaning of ∇_μ will be discussed in detail in the GR course. In this course, we will have to be satisfied with treating it as an operator that acts in a specific way on scalars, vectors and tensors:

- There is no difference between the covariant derivative and the partial derivative if it acts on a scalar

$$\nabla_\mu f = \partial_\mu f.$$

- Acting on a contravariant vector, V^ν, the covariant derivative is a partial derivative plus a correction that is linear in the vector:

$$\nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma^\nu_{\mu\lambda} V^\lambda.$$

Look carefully at the index structure of the second term. A similar definition applies to the covariant derivative of covariant vectors, ω_ν,

$$\nabla_\mu \omega_\nu = \partial_\mu \omega_\nu - \Gamma^\lambda_{\mu\nu} \omega_\lambda.$$

Notice the change of the sign of the second term and the placement of the dummy index.

- For tensors with many indices, you just repeat (1.3.84) and (1.3.85) for each index. For each upper index you introduce a term with a single $+\Gamma$, and for each lower index a term with a single $-\Gamma$:

$$\begin{aligned} &+ \Gamma^{\mu_1}{}_{\sigma\lambda} T^{\lambda\mu_2\cdots\mu_k}{}_{\nu_1\nu_2\cdots\nu_l} + \Gamma^{\mu_2}{}_{\sigma\lambda} T^{\mu_1\lambda\cdots\mu_k}{}_{\nu}\nu_2\cdots\nu_l + \cdots \\ &- \Gamma^{\lambda}{}_{\sigma\nu_1} T^{\mu_1\mu_2\cdots\mu_k}{}_{\lambda\nu_2\cdots\nu_l} - \Gamma^{\lambda}{}_{\sigma\nu_2} T^{\mu_1\mu_2\cdots\mu_k}{}_{\nu_1\lambda\cdots\nu_l} - \cdots . \end{aligned}$$

This is the general expression for the covariant derivative. Luckily, we will only be dealing with relatively simple tensors, so this monsterous expression will usually reduce to something managable.

[16] If this is the first time you have seen a covariant derivative, this will be a bit intimidating. Find me to talk about your fears.

$$\nabla_\sigma T^{\mu_1\mu_2\cdots\mu_k}{}_{\nu_1\nu_2\cdots\nu_l} = \partial_\sigma T^{\mu_1\mu_2\cdots\mu_k}{}_{\nu_1\nu_2\cdots\nu_l}$$

19 1. Geometry and Dynamics
Explicitly, eq. (1.3.82) can be written

$$\nabla_\mu N^\mu = \partial_\mu N^\mu + \Gamma^\mu_{\mu\lambda} N^\lambda = 0.$$

Using (1.3.78), this becomes

$$\frac{dn}{dt} + \Gamma^i_{i0} n = 0$$

and substituting (1.2.41), we find

$$\frac{\dot{n}}{n} = -3\frac{\dot{a}}{a} \quad \Rightarrow \quad n(t) \propto a^{-3}.$$

As expected, the number density decreases in proportion to the increase of the proper volume.

Energy-Momentum Tensor

We will now use a similar logic to determine what form of the stress-energy tensor $T_{\mu\nu}$ is consistent with the requirements of homogeneity and isotropy. First, we decompose $T_{\mu\nu}$ into a 3 -scalar, T_{00}, 3-vectors, T_{i0} and T_{0j}, and a 3 -tensor, T_{ij}. As before, isotropy requires the mean values of 3 -vectors to vanish, i.e. $T_{i0} = T_{0j} = 0$. Moreover, isotropy around a point $\boldsymbol{x} = 0$ requires the mean value of any 3 -tensor, such as T_{ij}, at that point to be proportional to δ_{ij} and hence to g_{ij}, which equals $-a^2\delta_{ij}$ at $\boldsymbol{x} = 0$,

$$T_{ij}(\boldsymbol{x} = 0) \propto \delta_{ij} \propto g_{ij}(\boldsymbol{x} = 0).$$

Homogeneity requires the proportionality coefficient to be only a function of time. Since this is a proportionality between two 3-tensors, T_{ij} and g_{ij}, it must remain unaffected by an arbitrary transformation of the spatial coordinates, including those transformations that preserve the form of g_{ij} while taking the origin into any other point. Hence, homogeneity and isotropy require the components of the stress-energy tensor everywhere to take the form

$$T_{00} = \rho(t), \quad \pi_i \equiv T_{i0} = 0, \quad T_{ij} = -P(t) g_{ij}(t, \boldsymbol{x}).$$

It looks even nicer with mixed upper and lower indices

$$T^\mu_\nu = g^{\mu\lambda} T_{\lambda\nu} = \begin{pmatrix} ccccc\rho 0 & & 00 \\ & 0 - P & 00 \\ & 00 & -P0 \\ & 00 & 0 - P \end{pmatrix}.$$

This is the stress-energy tensor of a perfect fluid as seen by a comoving observer. More generally, the stress-energy tensor can be written in the following, explicitly covariant, form

$$T^\mu_\nu = (\rho + P) U^\mu U_\nu - P\delta^\mu_\nu,$$

where $U^\mu \equiv dX^\mu/ds$ is the relative four-velocity between the fluid and the observer, while ρ and P are the energy density and pressure in the rest-frame of the fluid. Of course, we recover the previous result (1.3.92) for a comoving observer, $U^\mu = (1, 0, 0, 0)$.

How do the density and pressure evolve with time? In Minkowski space, energy and momentum are conserved. The energy density therefore satisfies the continuity equation $\dot{\rho} = -\partial_i \pi^i$, i.e. the rate of change of the density equals the divergence of the energy flux. Similarly, the

evolution of the momentum density satisfies the Euler equation, $\dot{\pi}_i = \partial_i P$. These conservation laws can be combined into a four-component conservation equation for the stress-energy tensor

$$\partial_\mu T^\mu_\nu = 0.$$

In general relativity, this is promoted to the covariant conservation equation

$$\nabla_\mu T^\mu_\nu = \partial_\mu T^\mu_\nu + \Gamma^\mu_{\mu\lambda} T^\lambda_\nu - \Gamma^\lambda_{\mu\nu} T^\mu_\lambda = 0.$$

Eq. (1.3.95) reduces to (1.3.94) in the local intertial frame. This corresponds to four separate equations (one for each ν). The evolution of the energy density is determined by the $\nu = 0$ equation

$$\partial_\mu T^\mu_0 + \Gamma^\mu_{\mu\lambda} T^\lambda_0 - \Gamma^\lambda_{\mu 0} T^\mu_\lambda = 0.$$

Since T^i_0 vanishes by isotropy, this reduces to

$$\frac{d\rho}{dt} + \Gamma^\mu_{\mu 0}\rho - \Gamma^\lambda_{\mu 0} T^\mu_\lambda = 0.$$

From eq. (1.2.41) we see that $\Gamma^\lambda_{\mu 0}$ vanishes unless λ and μ are spatial indices equal to each other, in which case it is $\dot{a}/a$. The continuity equation (1.3.97) therefore reads

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + P) = 0.$$

Exercise.-Show that (1.3.98) can be written as, $\mathrm{d}U = -P\ \mathrm{d}V$, where $U = \rho V$ and $V \propto a^3$.

6.1.4 Single-Component Universe

The different scalings of radiation (a^{-4}), matter (a^{-3}) and vacuum energy (a^0) imply that for most of its history the universe was dominated by a single component (first radiation, then matter, then vacuum energy; see fig. 1.11). Parameterising this component by its equation of state w_I captures all cases of interest. For a flat, single-component universe, the Friedmann equation (1.3.135) reduces to

$$\frac{\dot{a}}{a} = H_0 \sqrt{\Omega_I} a^{-\frac{3}{2}(1+w_I)}.$$

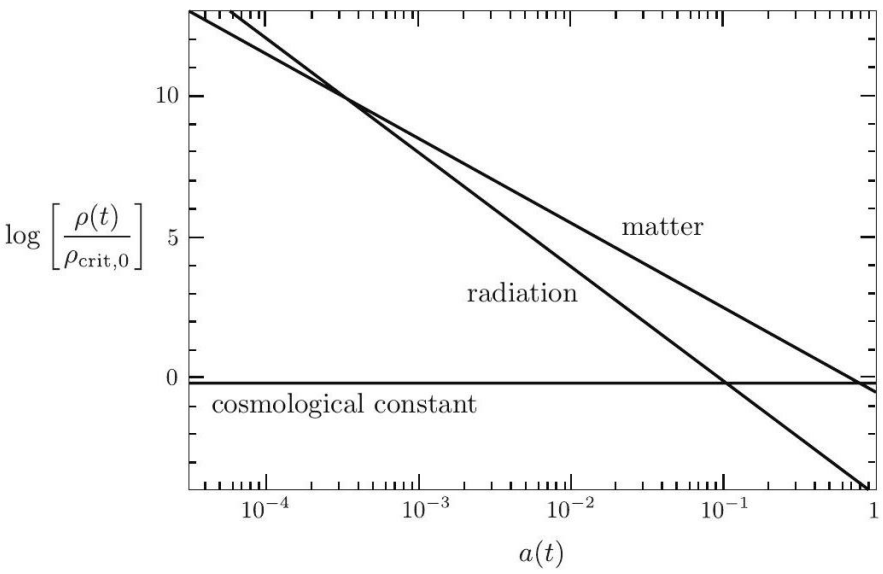


Figure 1.11: Evolution of the energy densities in the universe.

Integrating this equation, we obtain the time dependence of the scale factor

$$a(t) \propto \begin{cases} ccccc t^{2/3(1+w_I)} w_I \neq -1 & t^{2/3}\text{MD} \\ & t^{1/2}\text{RD} \\ e^{Ht} w_I = -1 & \Lambda\text{D} \end{cases}$$

or, in conformal time,

$$a(\tau) \propto \begin{cases} cccc \tau^{2/(1+3w_I)} w_I \neq -1 & \tau^2\text{MD} \\ & \tau\text{RD} \\ (-\tau)^{-1} w_I = -1 & \Lambda\text{D} \end{cases}$$

Exercise.-Derive eq. (1.3.138) from eq. (1.3.137).

Table 1.1 summarises the solutions for a flat universe during radiation domination (RD), matter domination (MD) and dark energy domination (ΛD).

	w	$\rho(a)$	$a(t)$	$a(\tau)$
RD	$\frac{1}{3}$	a^{-4}	$t^{1/2}$	τ
MD	0	a^{-3}	$t^{2/3}$	τ^2
ΛD	-1	a^0	e^{Ht}	$-\tau^{-1}$

Table 1.1: FRW solutions for a flat single-component universe.

6.1.5 Two-Component Universe

Matter and radiation were equally important at $a_{\rm eq} \equiv \Omega_r/\Omega_m \approx 3 \times 10^{-4}$, which was shortly before the cosmic microwave background was released (in §3.3.3, we will show that this happened at $a_{\rm rec} \approx 9 \times 10^{-4}$). It will be useful to have an exact solution describing the transition era. Let us therefore consider a flat universe filled with a mixture of matter and radiation. To solve for the evolution of the scale factor, it proves convenient to move to conformal time. The Friedmann equations (1.3.129) and (1.3.130) then are

$$\begin{aligned} (a')^2 &= \frac{8\pi G}{3}\rho a^4, \\ a'' &= \frac{4\pi G}{3}(\rho - 3P)a^3, \end{aligned}$$

where primes denote derivatives with respect to conformal time and

$$\rho \equiv \rho_m + \rho_r = \frac{\rho_{\rm eq}}{2}\left[\left(\frac{a_{\rm eq}}{a}\right)^3 + \left(\frac{a_{\rm eq}}{a}\right)^4\right].$$

Exercise.-Derive eqs. (1.3.139) and (1.3.140). You will first need to convince yourself that $\dot{a} = a'/a$ and $\ddot{a} = a''/a^2 - (a')^2/a^3$.

Notice that radiation doesn't contribute as a source term in eq. (1.3.140), $\rho_r - 3P_r = 0$. Moreover, since $\rho_m a^3 = \text{const.} = \frac{1}{2}\rho_{\rm eq} a_{\rm eq}^3$, we can write eq. (1.3.140) as

$$a'' = \frac{2\pi G}{3}\rho_{\rm eq} a_{\rm eq}^3.$$

This equation has the following solution

$$a(\tau) = \frac{\pi G}{3}\rho_{\rm eq} a_{\rm eq}^3 \tau^2 + C\tau + D.$$

Imposing $a(\tau = 0) \equiv 0$, fixes one integration constant, $D = 0$. We find the second integration constant by substituting (1.3.143) and (1.3.141) into (1.3.139),

$$C = \left(\frac{4\pi G}{3}\rho_{\rm eq}a_{\rm eq}^4\right)^{1/2}.$$

Eq. (1.3.143) can then be written as

$$a(\tau) = a_{\rm eq}\left[\left(\frac{\tau}{\tau_\star}\right)^2 + 2\left(\frac{\tau}{\tau_\star}\right)\right],$$

where

$$\tau_\star \equiv \left(\frac{\pi G}{3}\rho_{\rm eq}a_{\rm eq}^2\right)^{-1/2} = \frac{\tau_{\rm eq}}{\sqrt{2}-1}.$$

For $\tau \ll \tau_{\rm eq}$, we recover the radiation-dominated limit, $a \propto \tau$, while for $\tau \gg \tau_{\rm eq}$, we agree with the matter-dominated limit, $a \propto \tau^2$.

6.1.6 Expanding Confusion: common misconceptions of cosmological horizons and the superluminal expansion of the universe by Lineweaver

Abstract

We use standard general relativity to illustrate and clarify several common misconceptions about the expansion of the universe. To show the abundance of these misconceptions we cite numerous misleading, or easily misinterpreted, statements in the literature. In the context of the new standard ΛCDM cosmology we point out confusions regarding the particle horizon, the event horizon, the "observable universe" and the Hubble sphere (distance at which recession velocity = c). We show that we can observe galaxies that have, and always have had, recession velocities greater than the speed of light. We explain why this does not violate special relativity and we link these concepts to observational tests. Attempts to restrict recession velocities to less than the speed of light require a special relativistic interpretation of cosmological redshifts. We analyze apparent magnitudes of supernovae and observationally rule out the special relativistic Doppler interpretation of cosmological redshifts at a confidence level of 23σ.

Keywords: Cosmology: observations, Cosmology: theory
PACS 04.20.Cv, 98.80.Es, 98.80.Jk

Introduction

The general relativistic (GR) interpretation of the redshifts of distant galaxies, as the expansion of the universe, is widely accepted. However this interpretation leads to several concepts that are widely misunderstood. Since the expansion of the universe is the basis of the big bang model, these misunderstandings are fundamental. Popular science books written by astrophysicists, astrophysics textbooks and to some extent professional astronomical literature addressing the expansion of the Universe, contain misleading, or easily misinterpreted, statements concerning recession velocities, horizons and the "observable universe". Probably the most common misconceptions surround the expansion of the Universe at distances beyond which Hubble's law ($v_{\rm rec} = HD$: recession velocity = Hubble's constant × distance) predicts

recession velocities faster than the speed of light [Appendix B: 1–8], despite efforts to clarify the issue (Murdoch 1977, Harrison 1981, Silverman 1986, Stuckey 1992, Ellis & Rothman 1993, Harrison 1993, Kiang 1997, Davis & Lineweaver 2000, Kiang 2001, Gudmundsson and Björnsson 2002). Misconceptions include misleading comments about the observability of objects receding faster than light [App. B: 9–13]. Related, but more subtle confusions can be found surrounding cosmological event horizons [App. B: 14–15]. The concept of the expansion of the universe is so fundamental to our understanding of cosmology and the misconceptions so abundant that it is important to clarify these issues and make the connection with observational tests as explicit as possible. In Section 6.1.6 we review and illustrate the standard general relativistic description of the expanding universe using spacetime diagrams and we provide a mathematical summary in Appendix 6.1.6. On the basis of this description, in Section 6.1.6 we point out and clarify common misconceptions about superluminal recession velocities and horizons. Examples of misconceptions, or easily misinterpreted statements, occurring in the literature are given in Appendix A.2.6. Finally, in Section 6.1.6 we provide explicit observational tests demonstrating that attempts to apply special relativistic concepts to the Universe are in conflict with observations.

Standard general relativistic description of expansion

The results in this paper are based on the standard general relativistic description of an expanding homogeneous, isotropic universe (Appendix 6.1.6). Here we briefly summarize the main features of the GR description, about which misconceptions often arise. On the spacetime diagrams in Fig. 17 we demonstrate these features for the observationally favoured ΛCDM concordance model: $(\Omega_{\rm M}, \Omega_\Lambda) = (0.3, 0.7)$ with $H_0 = 70\, kms^{-1}Mpc^{-1}$ (Bennett et al. 2003, to one significant figure). The three spacetime diagrams in Fig. 17 plot, from top to bottom, time versus proper distance D, time versus comoving distance $R_0\chi$, and conformal time τ versus comoving distance. They show the relationship between comoving objects, our past light cone, the Hubble sphere and cosmological horizons.

Two types of horizon are shown in Fig. 17. The particle horizon is the distance light can have travelled from $t = 0$ to a given time t (Eq. 27), whereas the event horizon is the distance light can travel from a given time t to $t = \infty$ (Eq. 28). Using Hubble's law ($v_{\rm rec} = HD$), the Hubble sphere is defined to be the distance beyond which the recession velocity exceeds the speed of light, $D_{\rm HS} = c/H$. As we will see, the Hubble sphere is not an horizon. Redshift does not go to infinity for objects on our Hubble sphere (in general) and for many cosmological models we can see beyond it.

In the ΛCDM concordance model all objects with redshift greater than $z \sim 1.46$ are receding faster than the speed of light. This does not contradict SR because the motion is not in any observer's inertial frame. No observer ever overtakes a light beam and all observers measure light locally to be travelling at c. Hubble's law is derived directly from the Robertson-Walker metric (Eq. 15), and is valid for all distances in any homogeneous, expanding universe.

The teardrop shape of our past light cone in the top panel of Fig. 17 shows why we can observe objects that are receding superluminally. Light that superluminally receding objects emit propagates towards us with a local peculiar velocity of c, but since the recession velocity at that distance is greater than c, the total velocity of the light is away from us (Eq. 20). However, since the radius of the Hubble sphere increases with time, some photons that were initially in a superluminally receding region later find themselves in a subluminally receding region. They can therefore approach us and eventually reach us. The objects that emitted the photons however, have moved to larger distances and so are still receding superluminally. Thus we can observe objects that are receding faster than the speed of light (see Section 6.1.6 for more detail).

Our past light cone approaches the cosmological event horizon as $t_0 \to \infty$ (Eqs. 22 and 28).

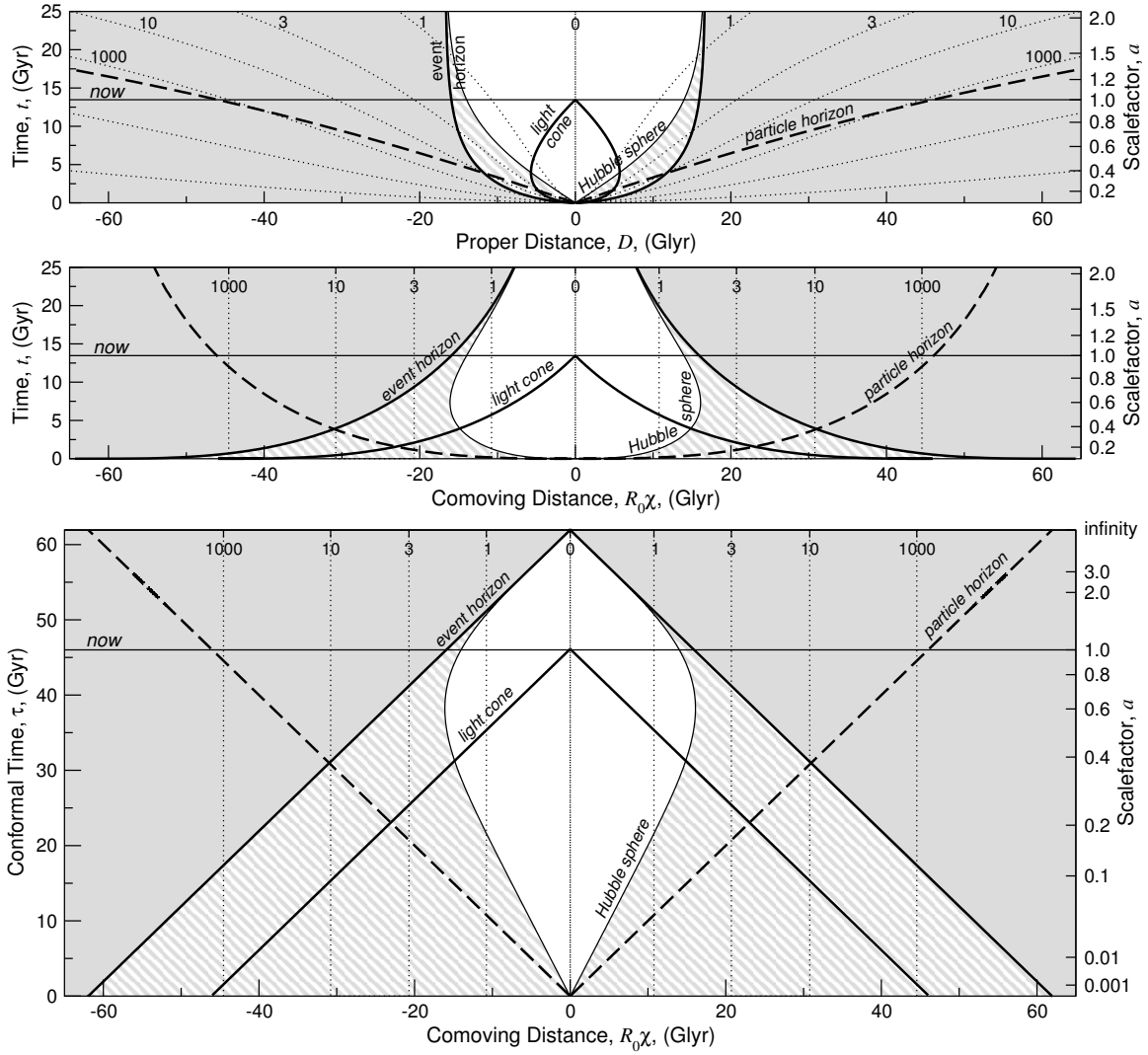


Figure 17: Spacetime diagrams showing the main features of the general relativistic description of the expansion of the universe for the $(\Omega_{\rm M}, \Omega_\Lambda) = (0.3, 0.7)$ model with $H_0 = 70\,km\ s^{-1}Mpc^{-1}$. Dotted lines show the worldlines of comoving objects. We are the central vertical worldline. The current redshifts of the comoving galaxies shown appear labeled on each comoving worldline. The normalized scalefactor, $a = R/R_0$, is drawn as an alternate vertical axis. All events that we currently observe are on our past light cone (with apex at $t =$ now). All comoving objects beyond the Hubble sphere (thin solid line) are receding faster than the speed of light. Top panel (proper distance): The speed of photons relative to us (the slope of the light cone) is not constant, but is rather $v_{\rm rec} - c$. Photons we receive that were emitted by objects beyond the Hubble sphere were initially receding from us (outward sloping lightcone at $t \lesssim 5$ Gyr). Only when they passed from the region of superluminal recession $v_{\rm rec} > c$ (gray crosshatching) to the region of subluminal recession (no shading) can the photons approach us. More detail about early times and the horizons is visible in comoving coordinates (middle panel) and conformal coordinates (lower panel). Our past light cone in comoving coordinates appears to approach the horizontal ($t = 0$) axis asymptotically. However it is clear in the lower panel that the past light cone at $t = 0$ only reaches a finite distance: about 46 Glyr, the current distance to the particle horizon. Currently observable light that has been travelling towards us since the beginning of the universe, was emitted from comoving positions that are now 46 Glyr from us. The distance to the particle horizon as a function of time is represented by the dashed line. Our event horizon is our past light cone at the end of time, $t = \infty$ in this case. It asymptotically approaches $\chi = 0$ as $t \to \infty$. The vertical axis of the lower panel shows conformal time. An infinite proper time is transformed into a finite conformal time so this diagram is complete on the vertical axis. The aspect ratio of $\sim 3/1$ in the top two panels represents the ratio between the radius of the observable universe and the age of the universe, 46 Glyr/13.5 Gyr.

Most observationally viable cosmological models have event horizons and in the ΛCDM model of Fig. 17, galaxies with redshift $z \sim 1.8$ are currently crossing our event horizon. These are the most distant objects from which we will ever be able to receive information about the present day. The particle horizon marks the size of our observable universe. It is the distance to the most distant object we can see at any particular time. The particle horizon can be larger than the event horizon because, although we cannot see *events* that occur beyond our event horizon, we can still see many galaxies that are beyond our current event horizon by light they emitted long ago.

In the GR description of the expansion of the Universe redshifts do not relate to velocities according to any SR expectations. We do not observe objects on the Hubble sphere (that recede at the speed of light) to have an infinite redshift (solve Eq. 24 for z using $\chi = c/\dot{R}$). Instead photons we receive that have infinite redshift were emitted by objects on our particle horizon. In addition, all galaxies become increasingly redshifted as we watch them approach the cosmological *event* horizon ($z \to \infty$ as $t \to \infty$). As the end of the universe approaches, all objects that are not gravitationally bound to us will be redshifted out of detectability.

Since this paper deals frequently with recession velocities and the expansion of the Universe the observational status of these concepts is important and is discussed in Sections 6.1.6 and 6.1.6.

Misconceptions

Misconception #1: Recession velocities cannot exceed the speed of light

A common misconception is that the expansion of the Universe cannot be faster than the speed of light. Since Hubble's law predicts superluminal recession at large distances ($D > c/H$) it is sometimes stated that Hubble's law needs special relativistic corrections when the recession velocity approaches the speed of light [App. B: 6–7]. However, it is well-accepted that general relativity, not special relativity, is necessary to describe cosmological observations. Supernovae surveys calculating cosmological parameters, galaxy-redshift surveys and cosmic microwave background anisotropy tests, all use general relativity to explain their observations. When observables are calculated using special relativity, contradictions with observations quickly arise (Section 6.1.6). Moreover, we know there is no contradiction with special relativity when faster than light motion occurs *outside the observer's inertial frame*. General relativity was specifically derived to be able to predict motion when global inertial frames were not available.

In special relativity, redshifts arise directly from velocities. It was this idea that led Hubble in 1929 to convert the redshifts of the "nebulae" he observed into velocities, and predict the expansion of the universe with the linear velocity-distance law that now bears his name. The general relativistic interpretation of the expansion interprets cosmological redshifts as an indication of velocity since the proper distance between comoving objects increases. However, the velocity is due to the rate of expansion of space, not movement through space, and therefore cannot be calculated with the special relativistic Doppler shift formula. Hubble & Humason's calculation of velocity therefore should not be given special relativistic corrections at high redshift, contrary to their suggestion [App. B: 16].

The general relativistic and special relativistic relations between velocity and cosmological redshift are (e.g. Davis & Lineweaver, 2001):

$$\text{GR} \qquad v_{\rm rec}(t,z) = \frac{c}{R_0}\,\dot{R}(t)\int_0^z \frac{dz'}{H(z')}, \tag{1}$$

$$\text{SR} \qquad v_{\rm pec}(z) = c\,\frac{(1+z)^2-1}{(1+z)^2+1}. \tag{2}$$

These velocities are measured with respect to the comoving observer who observes the receding object to have redshift, z. The GR description is written explicitly as a function of time because when we observe an object with redshift, z, we must specify the epoch at which we wish to calculate its recession velocity. For example, setting $t = t_o$ yields the recession velocity today of the object that emitted the observed photons at $t_{\rm em}$. Setting $t = t_{\rm em}$ yields the recession velocity at the time the photons were emitted (see Eqs. 17 & 24). The changing recession velocity of a comoving object is reflected in the changing slope of its worldline in the top panel of Fig. 17. There is no such time dependence in the SR relation.

Despite the fact that special relativity incorrectly describes cosmological redshifts it has been used for decades to convert cosmological redshifts into velocity because the special relativistic Doppler shift formula (Eq. 2), shares the same low redshift approximation, $v = cz$, as Hubble's Law (Fig. 18). It has only been in the last decade that routine observations have been deep enough that the distinction makes a difference. Figure 18 shows a snapshot at t_0 of the GR velocity-redshift relation for various models as well as the SR velocity-redshift relation and their common low redshift approximation, $v = cz$. Recession velocities exceed the speed of light in all viable cosmological models for objects with redshifts greater than $z \sim 1.5$. At higher redshifts special relativistic "corrections" can be more incorrect than the simple linear approximation (Fig. 21).

Some of the most common misleading applications of relativity arise from the misconception that nothing can recede faster than the speed of light. These include texts asking students to calculate the velocity of a high redshift receding galaxy using the special relativistic Doppler shift equation [App. B: 17–21], as well as the comments that galaxies recede from us at speeds "approaching the speed of light" [App. B: 4–5, 8], or quasars are receding at a certain percentage of the speed of light[1] [App. B: 3, 18–21].

Although velocities of distant galaxies are in principle observable, the set of synchronized comoving observers required to measure proper distance (Weinberg, 1972, p. 415; Rindler, 1977, p. 218) is not practical. Instead, more direct observables such as the redshifts of standard candles can be used to observationally rule out the special relativistic interpretation of cosmological redshifts (Section 6.1.6).

Misconception #2: Inflation causes superluminal expansion of the universe but the normal expansion of the universe does not

Inflation is sometimes described as "superluminal expansion" [App. B: 22–23]. This is misleading because *any* expansion described by Hubble's law has superluminal expansion for sufficiently distant objects. Even during inflation, objects within the Hubble sphere ($D < c/H$) recede at less than the speed of light, while objects beyond the Hubble sphere ($D > c/H$) recede faster than the speed of light. This is identical to the situation during non-inflationary expansion, except the Hubble constant during inflation was much larger than subsequent values. Thus the distance to the Hubble sphere was much smaller. During inflation the proper distance to the Hubble sphere stays constant and is coincident with the event horizon – this is also identical to the asymptotic behaviour of an eternally expanding universe with a cosmological constant $\Omega_\Lambda > 0$ (Fig. 17, top panel).

The oft-mentioned concept of structures "leaving the horizon" during the inflationary period refers to structures once smaller than the Hubble sphere becoming larger than the Hubble sphere. If the exponentially expanding regime, $R = R_0 e^{Ht}$, were extended to the end of time,

[1] Redshifts are usually converted into velocities using $v = cz$, which is a good approximation for $z \lesssim 0.3$ (see Fig. 18) but inappropriate for today's high redshift measurements. When a "correction" is made for high redshifts, the formula used is almost invariably the inappropriate special relativistic Doppler shift equation (Eq. 2).

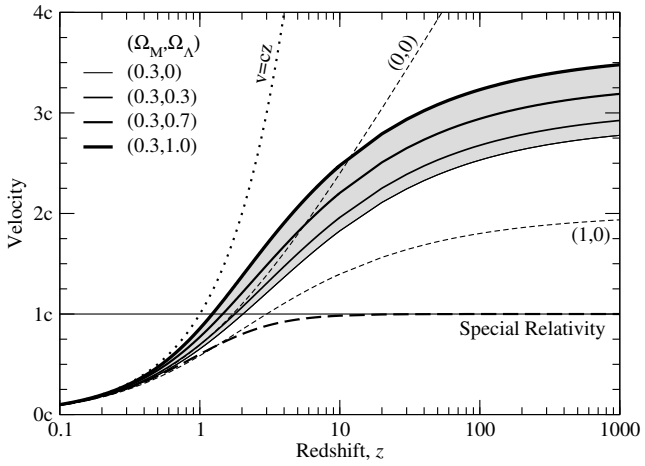


Figure 18: Velocity as a function of redshift under various assumptions. The linear approximation, $v = cz$, is the low redshift approximation of both the GR and SR results. The SR result is calculated using Eq. 2 while the GR result uses Eq. 1 and quotes the recession velocity at the present day, i.e., uses $\dot{R}(t) = \dot{R}_0$. The solid dark lines and gray shading show a range of Fridemann-Robertson-Walker (FRW) models as labeled in the legend. These include the observationally favoured cosmological model $(\Omega_{\rm M}, \Omega_\Lambda) = (0.3, 0.7)$. The recession velocity of all galaxies with $z \gtrsim 1.5$ exceeds the speed of light in all viable cosmological models. Observations now routinely probe regions that are receding faster than the speed of light.

the Hubble sphere would be the event horizon. However, in the context of inflation the Hubble sphere is not a true event horizon because structures that have crossed the horizon can "reenter the horizon" after inflation stops. The horizon they "reenter" is the revised event horizon determined by how far light can travel in a FRW universe without inflation.

It would be more appropriate to describe inflation as superluminal expansion if all distances down to the Planck length, $l_{\rm pl} \sim 10^{-35}m$, were receding faster than the speed of light. Solving $D_{\rm H} = c/H = l_{\rm pl}$ gives $H = 10^{43}s^{-1}$ (inverse Planck time) which is equivalent to $H = 10^{62}kms^{-1}Mpc^{-1}$. If Hubble's constant during inflation exceeded this value it would justify describing inflation as "superluminal expansion".

Misconception #3: Galaxies with recession velocities exceeding the speed of light exist but we cannot see them

Amongst those who acknowledge that recession velocities can exceed the speed of light, the claim is sometimes made that objects with recession velocities faster than the speed of light are not observable [App. B: 9–13]. We have seen that the speed of photons propagating towards us (the slope of our past light cone in the upper panel of Fig. 17) is not constant, but is rather $v_{\rm rec} - c$. Therefore light that is beyond the Hubble sphere has a total velocity away from us. How is it then that we can ever see this light? Although the photons are in the superluminal region and therefore recede from us (in proper distance), the Hubble sphere also recedes. In decelerating universes H decreases as $\dot{a}$ decreases (causing the Hubble sphere to recede). In accelerating universes H also tends to decrease since $\dot{a}$ increases more slowly than a. As long as the Hubble sphere recedes faster than the photons immediately outside it, $\dot{D}_{\rm H} > v_{\rm rec} - c$, the photons end up in a subluminal region and approach us[2]. Thus photons near the Hubble

[2]The behaviour of the Hubble sphere is model dependent. The Hubble sphere recedes as long as the deceleration parameter $q = -\ddot{R}R/\dot{R}^2 > -1$. In some closed eternally accelerating universes (specifically $\Omega_{mM} + \Omega_\Lambda > 1$ and $\Omega_\Lambda > 0$) the deceleration parameter can be less than minus one in which case we see faster-than-exponential

sphere that are receding slowly are overtaken by the more rapidly receding Hubble sphere[3].

Our teardrop shaped past light cone in the top panel of Fig. 17 shows that any photons we now observe that were emitted in the first $\sim$ five billion years were emitted in regions that were receding superluminally, $v_{\rm rec} > c$. Thus their total velocity was away from us. Only when the Hubble sphere expands past these photons do they move into the region of subluminal recession and approach us. The most distant objects that we can see now were outside the Hubble sphere when their comoving coordinates intersected our past light cone. Thus, they were receding superluminally when they emitted the photons we see now. Since their worldlines have always been beyond the Hubble sphere these objects were, are, and always have been, receding from us faster than the speed of light.

Evaluating Eq. 1 for the observationally favoured $(\Omega_{\rm M}, \Omega_\Lambda) = (0.3, 0.7)$ universe shows that all galaxies beyond a redshift of $z = 1.46$ are receding faster than the speed of light (Fig. 18). Hundreds of galaxies with $z > 1.46$ have been observed. The highest spectroscopic redshift observed in the Hubble deep field is $z = 6.68$ (Chen et al., 1999) and the Sloan digital sky survey has identified four galaxies at $z > 6$ (Fan et al., 2003). All of these galaxies have always been receding superluminally. The particle horizon, not the Hubble sphere, marks the size of our observable universe because we cannot have received light from, or sent light to, anything beyond the particle horizon[4]. Our effective particle horizon is the cosmic microwave background (CMB), at redshift $z \sim 1100$, because we cannot see beyond the surface of last scattering. Although the last scattering surface is not at any fixed comoving coordinate, the current recession velocity of the points from which the CMB was emitted is $3.2c$ (Fig. 18). At the time of emission their speed was $58.1c$, assuming $(\Omega_{\rm M}, \Omega_\Lambda) = (0.3, 0.7)$. Thus we routinely observe objects that are receding faster than the speed of light and the Hubble sphere is not a horizon[5].

Ambiguity: The depiction of particle horizons on spacetime diagrams

Here we identify an inconvenient feature of the most common depiction of the particle horizon on spacetime diagrams and provide a useful alternative (Fig. 19). The particle horizon at any particular time is a sphere around us whose radius equals the distance to the most distant object we can see. The particle horizon has traditionally been depicted as the *worldline* or comoving coordinate of the most distant particle that we have ever been able to see (Rindler, 1956; Ellis & Rothman, 1993). The only information this gives is contained in a single point: the current distance of the particle horizon, and this indicates the current radius of the observable universe. The rest of the worldline can be misleading as it does not represent a boundary between events we can see and events we cannot see, nor does it represent the distance to the particle horizon at different times. An alternative way to represent the particle horizon is to plot the distance to the particle horizon as a function of time (Kiang, 1991). The particle horizon at any particular time defines a unique distance which appears as a single point on a

expansion and some subluminally expanding regions can be beyond the event horizon (light that was initially in subluminal regions can end up in superluminal regions and never reach us). Exponential expansion, such as that found in inflation, has $q = -1$. Therefore the Hubble sphere is at a constant proper distance and coincident with the event horizon. This is also the late time asymptotic behaviour of eternally expanding FRW models with $\Omega_\Lambda > 0$ (see Fig. 17, upper panel).

[3]The myth that superluminally receding galaxies are beyond our view, may have propagated through some historical preconceptions. Firstly, objects on our event horizon *do* have infinite redshift, tempting us to apply our SR knowledge that infinite redshift corresponds to a velocity of c. Secondly, the once popular steady state theory predicts exponential expansion, for which the Hubble sphere and event horizon *are* coincident.

[4]The current distance to our particle horizon and its velocity is difficult to say due to the unknown duration of inflation. The particle horizon depicted in Fig. 17 assumes no inflation.

[5]Except in the special cases when the expansion is exponential, $R = R_0 e^{Ht}$, such as the de Sitter universe $(\Omega_{mM} = 0, \Omega_\Lambda > 0)$, during inflation or in the asymptotic limit of eternally expanding FRW universes.

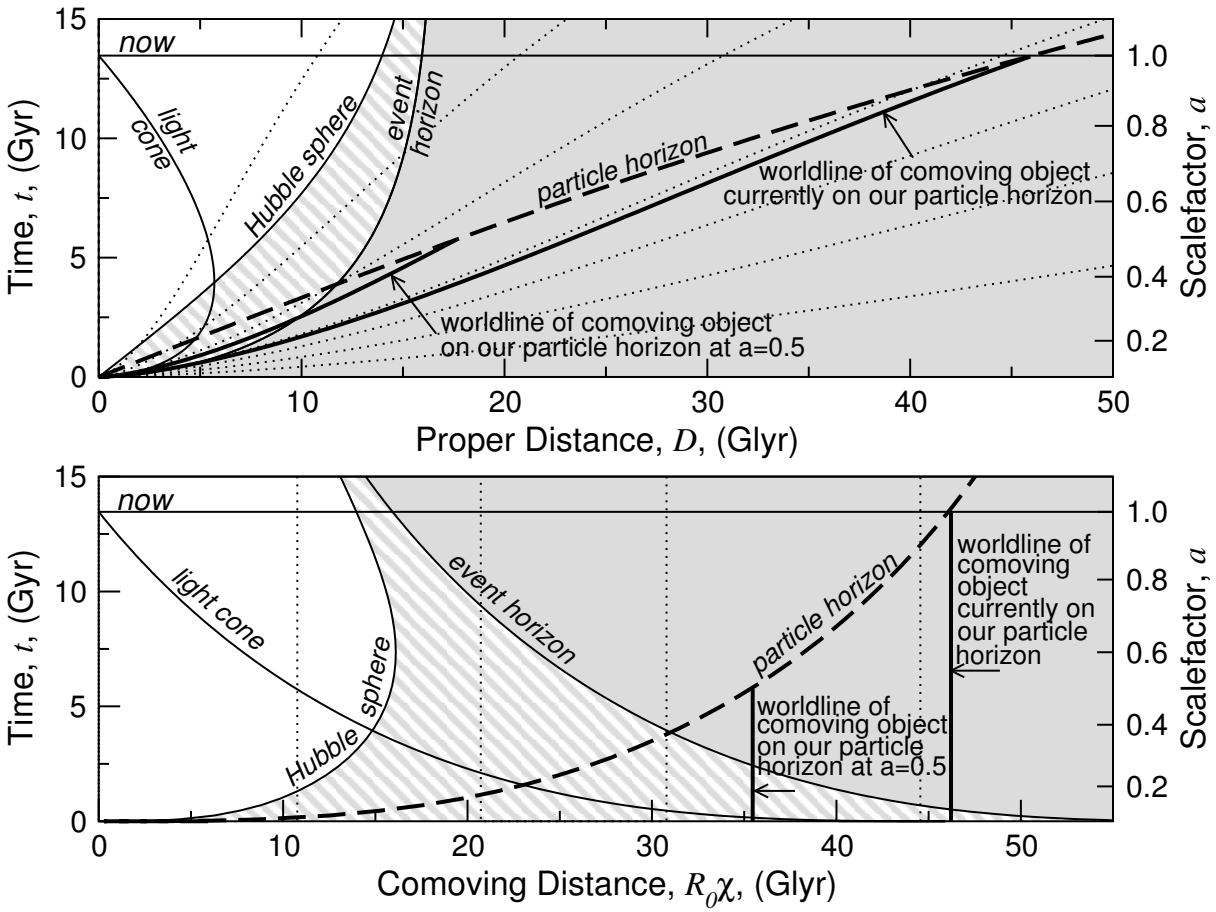


Figure 19: The traditional depiction of the particle horizon on spacetime diagrams is the worldline of the object currently on our particle horizon (thick solid line). All the information in this depiction is contained in a single point, the current distance to the particle horizon. An alternative way to plot the particle horizon is to plot the distance to the particle horizon as a function of time (thick dashed line and Fig. 17). This alleviates the need to draw a new worldline when we need to determine the particle horizon at another time (for example the worldline of the object on our particle horizon when the scalefactor $a = 0.5$).

spacetime diagram. Connecting the points gives the distance to the particle horizon vs time. It is this time dependent series of particle horizons that we plot in Fig. 17. (Rindler (1956) calls this the boundary of our creation light cone – a future light cone starting at the big bang.) Drawn this way, one can read from the spacetime diagram the distance to the particle horizon at any time. There is no need to draw another worldline.

Specifically, what we plot as the particle horizon is $\chi_{\rm ph}(t)$ from Eq. 27 rather than the traditional $\chi_{\rm ph}(t_0)$. To calculate the distance to the particle horizon at an arbitrary time t it is not sufficient to multiply $\chi_{\rm ph}(t_0)$ by $R(t)$ since the comoving distance to the particle horizon also changes with time.

The particle horizon is sometimes distinguished from the event horizon by describing the particle horizon as a "barrier in space" and the event horizon as a "barrier in spacetime". This is not a useful distinction because both the particle horizon and event horizon are surfaces in spacetime – they both form a sphere around us whose radius varies with time. When viewed in comoving coordinates the particle horizon and event horizon are mirror images of each other (symmetry about $z = 10$ in the middle and lower panels of Fig. 17). The traditional depiction of the particle horizon would appear as a straight vertical line in comoving coordinates, i.e., the

comoving coordinate of the present day particle horizon (Fig. 19, lower panel).

The proper distance to the particle horizon is *not* $D_{\rm PH} = ct_0$. Rather, it is the proper distance to the most distant object we can observe, and is therefore related to how much the universe has expanded, i.e. how far away the emitting object has become, since the beginning of time. In general this is $\sim 3ct_0$. The relationship between the particle horizon and light travel time arises because the comoving coordinate of the most distant object we can see *is* determined by the *comoving* distance light has travelled during the lifetime of the universe (Eq. 27).

Observational evidence for the general relativistic interpretaion of cosmological redshifts

Duration-redshift relation for Type Ia Supernovae

Many misconceptions arise from the idea that recession velocities are limited by SR to less than the speed of light so in Section 6.1.6 we present an analysis of supernovae observations yielding evidence against the SR interpretation of cosmological redshifts. But first we would like to present an observational test that *can not* distinguish between special relativistic and general relativistic expansion of the Universe.

General relativistic cosmology predicts that events occurring on a receding emitter will appear time dilated by a factor,

$$\gamma_{\rm GR}(z) = 1 + z. \tag{3}$$

A process that takes Δt_0 as measured by the emitter appears to take $\Delta t = \gamma_{\rm GR}\Delta t_0$ as measured by the observer when the light emitted by that process reaches them. Wilson (1939) suggested measuring this cosmological time dilation to test whether the expansion of the Universe was the cause of cosmological redshifts. Type Ia supernovae (SNe Ia) lightcurves provide convenient standard clocks with which to test cosmological time dilation. Recent evidence from supernovae includes Leibundgut et al. (1996) who gave evidence for GR time dilation using a single high-z supernova and Riess et al. (1997) who showed $1+z$ time dilation for a single SN Ia at the 96.4% confidence level using the time variation of spectral features. Goldhaber et al. (1997) show five data points of lightcurve width consistent with $1 + z$ broadening and extend this analysis in Goldhaber et al. (2001) to rule out any theory that predicts zero time dilation (for example "tired light" scenarios (see Wright, 2001)), at a confidence level of 18σ. All of these tests show that $\gamma = (1 + z)$ time dilation is preferred over models that predict *no* time dilation.

We want to know whether the same observational test can show that GR time dilation is preferred over SR time dilation as the explanation for cosmological redshifts. When we talk about SR expansion of the universe we are assuming that we have an inertial frame that extends to infinity (impossible in the GR picture) and that the expansion involves objects moving through this inertial frame. The time dilation factor in SR is,

$$\gamma_{\rm SR}(z) = (1 - v_{\rm pec}^2/c^2)^{-1/2}, \tag{4}$$

$$= \frac{1}{2}(1 + z + \frac{1}{1+z}) \approx 1 + z^2/2. \tag{5}$$

This time dilation factor relates the proper time in the moving emitter's inertial frame (Δt_0) to the proper time in the observer's inertial frame (Δt_1). To measure this time dilation the observer has to set up a set of synchronized clocks (each at rest in the observer's inertial frame) and take readings of the emitter's proper time as the emitter moves past each synchronized clock. The readings show that the emitter's clock is time dilated such that $\Delta t_1 = \gamma_{\rm SR}\Delta t_0$.

We do not have this set of synchronized clocks at our disposal when we measure time dilation of supernovae and therefore Eq. 5 is not the time dilation we observe. In an earlier version of this paper we mistakenly attempted to use this equation to show SR disagreed with

observational results. This could be classed as an example of an "expanding confusion". For the observed time dilation of supernovae we have to take into account an extra time dilation factor that occurs because the distance to the emitter (and thus the distance light has to propagate to reach us) is increasing. In the time Δt_1 the emitter moves a distance $v\Delta t_1$ away from us. The total proper time we observe (Δt) is Δt_1 plus an extra factor describing how long light takes to traverse this extra distance ($v\Delta t_1/c$),

$$\Delta t = \Delta t_1(1 + v/c). \tag{6}$$

The relationship between proper time at the emitter and proper time at the observer is thus,

$$\begin{aligned} \Delta t &= \Delta t_0\gamma(1 + v/c), & (7)\\ &= \Delta t_0\sqrt{\frac{1 + v/c}{1 - v/c}}, & (8)\\ &= \Delta t_0(1 + z). & (9) \end{aligned}$$

This is identical to the GR time dilation equation. Therefore using time dilation to distinguish between GR and SR expansion is impossible.

Leibundgut et al. (1996), Riess et al. (1997) and Goldhaber et al. (1997, 2001) do provide excellent evidence that expansion is a good explanation for cosmological redshifts. What they can not show is that GR is a better description of the expansion than SR. Nevertheless, other observational tests provide strong evidence against the SR interpretation of cosmological redshifts, and we demonstrate one such test in the next section.

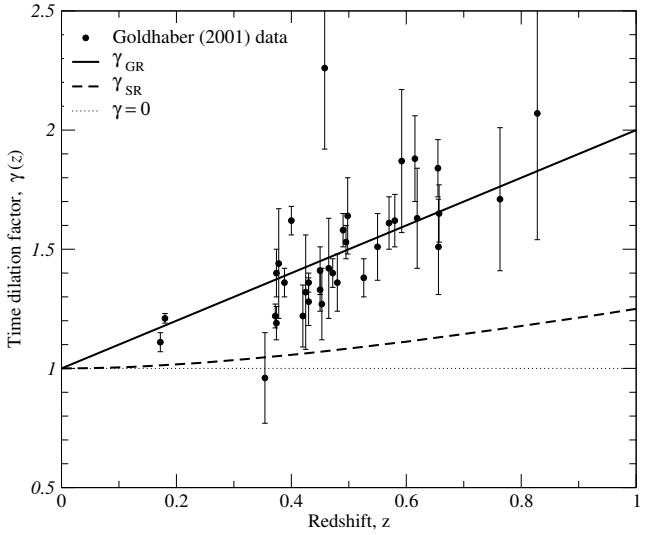


Figure 20: Supernovae time dilation factor vs redshift. The solid line is the time dilation factor predicted by both general relativity and special relativity. The thick dashed line is the special relativistic time dilation factor that a set of synchronized clocks spread throughout our inertial frame would observe, without taking into account the changing distance light has to travel to reach us. Once the change in the emitter's distance is taken into account SR predicts the same time dilation effect as GR, $\gamma = (1 + z)$. The thin dotted line represents any theory that predicts no time dilation (e.g. tired light). The 35 data points are from Goldhaber et al. (2001). They rule out no time dilation at a confidence level of 18σ.

Magnitude-redshift relationship for SNe Ia

Another observational confirmation of the GR interpretation that *is* able to rule out the SR interpretation is the curve in the magnitude-redshift relation. SNe Ia are being used as standard candles to fit the magnitude-redshift relation out to redshifts close to one (Riess et al.,

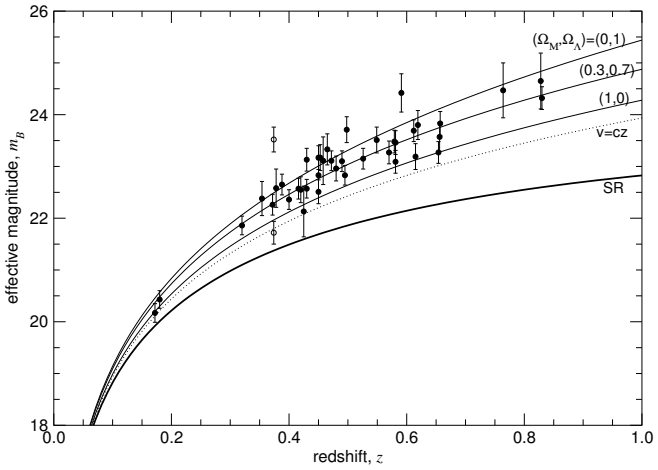


Figure 21: Magnitude-redshift relation for several models with data taken from Perlmutter et al. 1999 [Fig. 2(a)]. The SR prediction has been added (as described in text), as has the prediction assuming a linear $v = cz$ relationship. The interpretation of the cosmological redshift as an SR Doppler effect is ruled out at more than 23σ compared with the ΛCDM concordance model. The linear $v = cz$ model is a better approximation than SR, but is still ruled out at 12σ.

1998; Perlmutter et al., 1999). Recent measurements are accurate enough to put restrictions on the cosmological parameters $(\Omega_{\rm M}, \Omega_\Lambda)$. We perform a simple analysis of the supernovae magnitude-redshift data to show that it also strongly excludes the SR interpretation of cosmological redshifts (Fig. 21).

Figure 21 shows the theoretical curves for several GR models accompanied by the observed SNe Ia data from Perlmutter et al. (1999) [their Fig. 2(a)]. The conversion between luminosity distance, D_L (Eq. 13), and effective magnitude in the B-band given in Perlmutter et al. (1999), is $m_{\rm B}(z) = 5 \log H_0 D_L + M_{\rm B}$ where $M_{\rm B}$ is the absolute magnitude in the B-band at the maximum of the light curve. They marginalize over $M_{\rm B}$ in their statistical analyses. We have taken $M_{\rm B} = -3.45$ which closely approximates their plotted curves.

We superpose the curve deduced by interpreting Hubble's law special relativistically. One of the strongest arguments against using SR to interpret cosmological redshifts is the difficulty in interpreting observational features such as magnitude. We calculate $D(z)$ special relativistically by assuming the velocity in $v = HD$ is related to redshift via Eq. 2, so,

$$D(z) = \frac{c}{H} \frac{(1+z)^2 - 1}{(1+z)^2 + 1}. \tag{10}$$

Since all the redshifting happens at emission in the SR scenario, v should be calculated at the time of emission. However, since SR does not provide a technique for incorporating acceleration into our calculations for the expansion of the Universe, the best we can do is assume that the recession velocity, and thus Hubble's constant, are approximately the same at the time of emission as they are now[6]. We then convert $D(z)$ to $D_L(z)$ using Eq. 13, so $D_L(z) = D(z)(1+z)$. This version of luminosity distance has been used to calculate $m(z)$ for the SR case in Fig. 21.

SR fails this observational test dramatically being 23σ from the general relativistic ΛCDM model $(\Omega_{\rm M}, \Omega_\Lambda) = (0.3, 0.7)$. We also include the result of assuming $v = cz$. Equating this to

[6]There are several complications that this analysis does not address. (1) SR could be manipulated to give an evolving Hubble's constant and (2) SR could be manipulated to give a non-trivial relationship between luminosity distance, D_L, and proper distance, D. However, it is not clear how one would justify these ad hoc corrections.

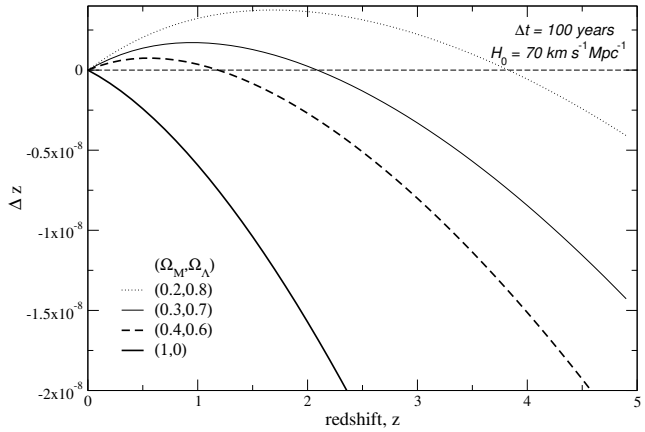


Figure 22: The change in the redshift of a comoving object as predicted by FRW cosmology for various cosmological models. The horizontal axis represents the initial redshifts. The timescale taken for the change is 100 years. The changes predicted are too small for current instrumentation to detect.

Hubble's law gives, $D_L(z) = cz(1+z)/H$. For this observational test the linear prediction is closer to the GR prediction (and to the data) than SR is. Nevertheless the linear result lies 12σ from the ΛCDM concordance result.

Future tests

Current instrumentation is not accurate enough to perform some other observational tests of GR. For example Sandage (1962) showed that the evolution in redshift of distant galaxies due to the acceleration or deceleration of the universe is a direct way to measure the cosmological parameters. The change in redshift over a time interval t_0 is given by,

$$\frac{dz}{dt_0} = H_0(1+z) - H_{\rm em}, \tag{11}$$

where $H_{\rm em} = \dot{R}_{\rm em}/R_{\rm em}$ is Hubble's constant at the time of emission. Unfortunately the magnitude of the redshift variation is small over human timescales. Ebert & Trümper (1975), Lake (1981), Loeb (1998) and references therein each reconfirmed that the technology of the day did not yet provide precise enough redshifts to make such an observation viable. Figure 22 shows the expected change in redshift due to cosmological acceleration or deceleration is only $\Delta z \sim 10^{-8}$ over 100 years. Current Keck/HIRES spectra with iodine cell reference wavelengths can measure quasar absorption line redshifts to an accuracy of $\Delta z \sim 10^{-5}$ (Outram et al., 1999). Thus, this observational test must wait for future technology.

Discussion

Recession velocities of individual galaxies are of limited use in *observational* cosmology because they are not directly observable. For this reason some of the physics community considers recession velocities meaningless and would like to see the issue swept under the rug [App. B: 24–25]. They argue that we should refrain from interpreting observations in terms of velocity or distance, and stick to the observable, redshift. This avoids any complications with superluminal recession and avoids any confusion between the variety of observationally-motivated definitions of distance commonly used in cosmology (e.g. Eqs. 13 and 14).

However, redshift is not the only observable that indicates distance and velocity. The host of low redshift distance measures and the multitude of available evidence for the Big Bang model

all suggest that higher redshift galaxies are more distant from us and receding faster than lower redshift galaxies. Moreover, we cannot currently sweep distance and velocity under the rug if we want to explain the cosmological redshift itself. Expansion has no meaning without well-defined concepts of velocity and distance. If recession velocity were meaningless we could not refer to an "expanding universe" and would have to restrict ourselves to some operational description such as "fainter objects have larger redshifts". However, within general relativity the relationship between cosmological redshift and recession velocity is straightforward. Observations of SNe Ia apparent magnitudes provide independent evidence that the cosmological redshifts are due to the general relativistic expansion of the universe. Understanding distance and velocity is therefore fundamental to the understanding of our Universe.

When distances are large enough that light has taken a substantial fraction of the age of the Universe to reach us there are more observationally convenient distance measures than proper distance, such as luminosity distance and angular-size distance. The most convenient distance measure depends on the method of observation. Nevertheless, these distance measures can be converted between each other, and so collectively define a unique concept[7]. In this paper we have taken proper distance to be the fundamental *radial* distance measure. Proper distance is the spatial geodesic measured along a hypersurface of constant cosmic time (as defined in the Robertson-Walker metric). It is the distance measured along a line of sight by a series of infinitesimal comoving rulers at a particular time, t. Both luminosity and angular-size distances are calculated from observables involving distance perpendicular to the line of sight and so contain the angular coefficient $S_k(\chi)$. They parametrize radial distances but are not geodesic distances along the three dimensional spatial manifold[8]. They are therefore not relevant for the calculation of recession velocity[9]. Nevertheless, if they were used, our results would be similar. Only angular-size distance can avoid superluminal velocities because $D_\theta = 0$ for both $z = 0$ and $z \to \infty$ (Murdoch, 1977). Even then the rate of change of angular-size distance does not approach c for $z \to \infty$.

Throughout this paper we have used proper time, t, as the temporal measure. This is the time that appears in the RW metric and the Friedmann equations. This is a convenient time measure because it is the proper time of comoving observers. Moreover, the homogeneity of the universe is dependent on this choice of time coordinate — if any other time coordinate were chosen (that is not a trivial multiple of t) the density of the universe would be distance dependent. Time can be defined differently, for example to make the SR Doppler shift formula (Eq. 2) correctly calculate recession velocities from observed redshifts (Page, 1993). However, to do this we would have to sacrifice the homogeneity of the universe and the synchronous proper time of comoving objects (Davis & Lineweaver, 2003).

[7]

$$\text{Proper Distance} \quad D = R\chi \tag{12}$$

$$\text{Luminosity Distance} \quad D_L = RS_k(\chi)(1+z) \tag{13}$$

$$\text{Angular-size Distance} \quad D_\theta = RS_k(\chi)(1+z)^{-1} \tag{14}$$

[8]Note also that the standard definition of angular-size distance is purported to be the physical size of an object, divided by the angle it subtends on the sky. The physical size used in this equation is not actually a length along a spatial geodesic, but rather along a line of constant χ (Liske, 2000). The correction is negligible for the small angles usually measured in astronomy.

[9]Murdoch, H. S. 1977, "[McVittie] regards as equally valid other definitions of distance such as luminosity distance and distance by apparent size. But while these are extremely useful concepts, they are really only definitions of observational convenience which extrapolate results such as the inverse square law beyond their range of validity in an expanding universe."

Conclusion

We have clarified some common misconceptions surrounding the expansion of the universe, and shown with numerous references how misleading statements manifest themselves in the literature. Superluminal recession is a feature of all expanding cosmological models that are homogeneous and isotropic and therefore obey Hubble's law. This does not contradict special relativity because the superluminal motion does not occur in any observer's inertial frame. All observers measure light locally to be travelling at c and nothing ever overtakes a photon. Inflation is often called "superluminal recession" but even during inflation objects with $D < c/H$ recede subluminally while objects with $D > c/H$ recede superluminally. Precisely the same relationship holds for non-inflationary expansion. We showed that the Hubble sphere is not a horizon — we routinely observe galaxies that have, and always have had, superluminal recession velocities. All galaxies at redshifts greater than $z \sim 1.46$ today are receding superluminally in the ΛCDM concordance model. We have also provided a more informative way of depicting the particle horizon on a spacetime diagram than the traditional worldline method. An abundance of observational evidence supports the general relativistic big bang model of the universe. The duration of supernovae light curves shows that models predicting no expansion are in conflict with observation. Using magnitude-redshift data from supernovae we were able to rule out the SR interpretation of cosmological redshifts at the $\sim 23\sigma$ level. Together these observations provide strong evidence that the general relativistic interpretation of the cosmological redshifts is preferred over special relativistic and tired light interpretations. The general relativistic description of the expansion of the universe agrees with observations, and does not need any modifications for $v_{\rm rec} > c$.

App: Standard general relativistic definitions of expansion and horizons

The metric for an homogeneous, isotropic universe is the Robertson-Walker (RW) metric,

$$ds^2 = -c^2dt^2 + R(t)^2[d\chi^2 + S_k^2(\chi)d\psi^2], \tag{15}$$

where c is the speed of light, dt is the time separation, $d\chi$ is the comoving coordinate separation and $d\psi^2 = d\theta^2 + \sin^2\theta d\phi^2$, where θ and ϕ are the polar and azimuthal angles in spherical coordinates. The scalefactor, R, has dimensions of distance. The function $S_k(\chi) = \sin\chi$, χ or $\sinh\chi$ for closed ($k = +1$), flat ($k = 0$) or open ($k = -1$) universes respectively (Peacock, 1999, p. 69). The proper distance D, at time t, in an expanding universe, between an observer at the origin and a distant galaxy is defined to be along a surface of constant time ($dt = 0$). We are interested in the radial distance so $d\psi = 0$. The RW metric then reduces to $ds = Rd\chi$ which, upon integration yields,

$$\text{Proper distance,} \qquad D(t) = R(t)\chi. \tag{16}$$

Differentiating this yields the theoretical form of Hubble's law (Harrison, 1993),

$$\text{Recession velocity,} \quad v_{\rm rec}(t,z) = \dot{R}(t)\chi(z), \tag{17}$$

$$= H(t)D(t), \tag{18}$$

where $v_{\rm rec} = \dot{D}$ (for $\dot{\chi} = 0$) and $\chi(z)$ is the fixed comoving coordinate associated with a galaxy observed today at redshift z. Note that the redshift of an object at this fixed comoving coordinate changes with time[10] (Eq. 11). A distant galaxy will have a particular recession velocity when it emits the photon at $t_{\rm em}$ and a different recession velocity when we observe the photon at t_0. Eq. 18 evaluated at t_0 gives the recession velocities plotted in Fig. 18.

[10]In addition, objects that have a peculiar velocity also move through comoving coordinates. Therefore more generally Eq. 17 above should be written with χ explicitly time dependent, $v_{\rm rec}(t,z) = \dot{R}(t)\chi(z,t)$.

The recession velocity of a comoving galaxy is a time dependent quantity because the expansion rate of the universe $\dot{R}(t)$ changes with time. The current recession velocity of a galaxy is given by $v_{\rm rec} = \dot{R}_0\chi(z)$. On the spacetime diagram of Fig. 17 this is the velocity taken at points along the line of constant time marked "now". The recession velocity of an emitter at the time it emitted the light we observe is the velocity at points along our the past light cone[11]. However, we can also compute the recession velocity a comoving object has at *any time* during the history of the universe, having initially calculated its comoving coordinate from its present day redshift.

Allowing χ to vary when differentiating Eq. 16 with respect to time gives two distinct velocity terms (Landsberg & Evans, 1977; Silverman, 1986; Peacock, 1999; Davis et al., 2003),

$$\dot{D} = \dot{R}\chi + R\dot{\chi}, \tag{19}$$

$$v_{\rm tot} = v_{\rm rec} + v_{\rm pec}. \tag{20}$$

This explains the changing slope of our past light cone in the upper panel of Fig. 17. The peculiar velocity of light is always c (Eq. 21) so the total velocity of light whose peculiar velocity is towards us is $v_{\rm tot} = v_{\rm rec} - c$ which is always positive (away from us) when $v_{\rm rec} > c$. Nevertheless we can eventually receive photons that initially were receding from us because the Hubble sphere expands and overtakes the receding photons so the photons find themselves in a region with $v_{\rm rec} < c$ (Section 6.1.6).

Photons travel along null geodesics, $ds = 0$. To obtain the comoving distance, χ, between an observer at the origin and a galaxy observed to have a redshift $z(t)$, set $ds = 0$ (to measure along the path of a photon) and $d\psi = 0$ (to measure radial distances) in the RW metric yielding,

$$c\,dt = R(t)d\chi. \tag{21}$$

This expression confirms our previous statement that the *peculiar* velocity of a photon, $R\dot{\chi}$, is c. Since the velocity of light through comoving coordinates is not constant ($\dot{\chi} = c/R$), to calculate comoving distance we cannot simply multiply the speed of light through comoving space by time. We have to integrate over this changing comoving speed of light for the duration of propagation. Thus, the comoving coordinate of a comoving object that emitted the light we now see at time t is attained by integrating Eq. 21,

$$\text{Past Light Cone, } \chi_{\rm lc}(t_{\rm em}) = c\int_{t_{\rm em}}^{t_0}\frac{dt'}{R(t')}. \tag{22}$$

We can parametrize time using redshift and thus recast Eq. 22 in terms of observables. The cosmological redshift of an object is given by the ratio of the scalefactor at the time of observation, $R(t_0) = R_0$, to the scalefactor at the time of emission, $R(t)$,

$$\text{Redshift, } \quad 1 + z = \frac{R_0}{R(t)}. \tag{23}$$

Differentiating Eq. 23 with respect to t gives $dt/R(t) = -dz/R_0H(z)$ where redshift is used instead of time to parametrize Hubble's constant. $H(z)$ is Hubble's constant at the time an object with redshift, z, emitted the light we now see. Thus, for the limits of the integral in Eq. 22, the time of emission becomes $z = 0$ while the time of observation becomes the observed redshift, z. The comoving coordinate of an object in terms of observables is therefore,

$$\chi(z) = \frac{c}{R_0}\int_o^z\frac{dz'}{H(z')}. \tag{24}$$

[11] The recession velocity at the time of emission is $v_{\rm rec}(t_{\rm em}) = R(t_{\rm em})\chi(z)$ where $R(t_{\rm em}) = R(t)$ as defined in Eq. 23.

Thus, there is a direct one to one relationship between observed redshift and comoving coordinate. Notice that in contrast to special relativity, the redshift does not indicate the velocity, it indicates the distance[12]. That is, the redshift tells us not the velocity of the emitter, but where the emitter sits (at rest locally) in the coordinates of the universe. The recession velocity is obtained by inserting Eq. 24 into Eq. 17 yielding Eq 1.

The Friedmann Equation gives the time dependence of Hubble's constant,

$$H(z) = H_0\,(1+z)\left[1+\Omega_{mM}z+\Omega_\Lambda\left(\frac{1}{(1+z)^2}-1\right)\right]^{1/2}. \tag{25}$$

Expressing Hubble's constant this way is useful because it is in terms of observables. However, it restricts our calculations to objects with redshift $z<\infty$. That is, objects we can currently see. There is no reason to assume the universe ceases beyond our current particle horizon. Expressing Hubble's constant as $H(t)=\dot{R}(t)/R(t)$ allows us to extend the analysis to a time of observation, $t\to\infty$, which is beyond what we can currently observe. Friedmann's equation is then (using the scalefactor normalized to one at the present day $a(t)=R(t)/R_0$),

$$\dot{R}(t) = R_0H_0\left[1+\Omega_{mM}\left(\frac{1}{a}-1\right)+\Omega_\Lambda(a^2-1)\right]^{1/2}, \tag{26}$$

which we use with the identity $dt/R(t)=dR/(\dot{R}R)$ to evaluate Eqs. 22, 27 and 28.

Altering the limits on the integral in Eq. 22 gives the horizons we have plotted on the spacetime diagrams. The time dependent particle horizon we plot in Fig. 17 uses $D_{\rm ph}=R(t)\chi_{\rm ph}(t)$ with,

$$\text{Particle Horizon, } \chi_{\rm ph}(t) = c\int_0^t \frac{dt'}{R(t')}. \tag{27}$$

The traditional depiction of the particle horizon as a worldline uses $D_{\rm ph}=R(t)\chi_{\rm ph}(t_0)$. The comoving distance to the event horizon is given by,

$$\text{Event Horizon, } \chi_{\rm eh}(t) = c\int_t^{t_{\rm end}} \frac{dt'}{R(t')}, \tag{28}$$

where $t_{\rm end}=\infty$ in eternally expanding models or the time of the big crunch in recollapsing models.

A conformal time interval, $d\tau$, is defined as a proper time interval dt divided by the scalefactor,

$$\text{Conformal time, } d\tau = dt/R(t). \tag{29}$$

6.2 Inflation

Обсудим подробно инфляцию.

(пока такой небольшой раздел, у баунмана он не раздут, так что и я пока что не раздуваю его.)

The FRW cosmology described in the previous chapter is incomplete. It doesn't explain why the universe is homogeneous and isotropic on large scales. In fact, the standard cosmology predicts that the early universe was made of many causally disconnected regions of space. The fact that these apparently disjoint patches of space have very nearly the same densities and

[12]Distance is proportional to recession velocity at any particular time, but a particular redshift measured at different times will correspond to different recession velocities.

temperatures is called the horizon problem. In this chapter, I will explain how inflation-an early period of accelerated expansion - drives the primordial universe towards homogeneity and isotropy, even if it starts in a more generic initial state.

Throughout this chapter, we will trade Newton's constant for the (reduced) Planck mass,

$$M_{\rm pl} \equiv \sqrt{\frac{\hbar c}{8\pi G}} = 2.4 \times 10^{18}{\rm GeV},$$

so that the Friedmann equation (1.3.131) is written as $H^2 = \rho/\left(3M_{\rm pl}^2\right)$.

6.2.1 The Horizon Problem (??)

Обзор (?????)

Проблема горизонта возникает, когда мы видим однородность и изотропность распределения космического фонового излучения (теплового излучения, оставшегося после Большого взрыва) на всем небе. Это означает, что эти различные области Вселенной, которые сейчас находятся далеко друг от друга, в прошлом были ближе друг к другу и имели возможность взаимодействовать и достичь термодинамического равновесия. Однако, если эти области находились за горизонтом друг друга на момент Большого взрыва, то как они могли прийти в равновесие, если между ними не было достаточного времени для взаимодействия?

Light and Horizons

The size of a causal patch of space it determined by how far light can travel in a certain amount of time. As we mentioned in §1.1.3, in an expanding spacetime the propagation of light (photons) is best studied using conformal time. Since the spacetime is isotropic, we can always define the coordinate system so that the light travels purely in the radial direction (i.e. $\theta = \phi = $ const.). The evolution is then determined by a two-dimensional line element [1]

$$ds^2 = a^2(\tau)\left[d\tau^2 - d\chi^2\right].$$

Since photons travel along null geodesics, $ds^2 = 0$, their path is defined by

$$\Delta\chi(\tau) = \pm\Delta\tau$$

where the plus sign corresponds to outgoing photons and the minus sign to incoming photons. This shows the main benefit of working with conformal time: light rays correspond to straight lines at 45° angles in the $\chi - \tau$ coordinates. If instead we had used physical time t, then the light cones for curved spacetimes would be curved. With these preliminaries, we now define two different types of cosmological horizons. One which limits the distances at which past events can be observed and one which limits the distances at which it will ever be possible to observe future events.

[1] For the radial coordinate χ we have used the parameterisation of (1.1.23), so that (2.1.1) is conformal to two-dimensional Minkowski space and the curvature k of the three-dimensional spatial slices is absorbed into the definition of the coordinate χ. Had we used the regular polar coordinate r, the two-dimensional line element would have retained a dependence on k. For flat slices, χ and r are of course the same.

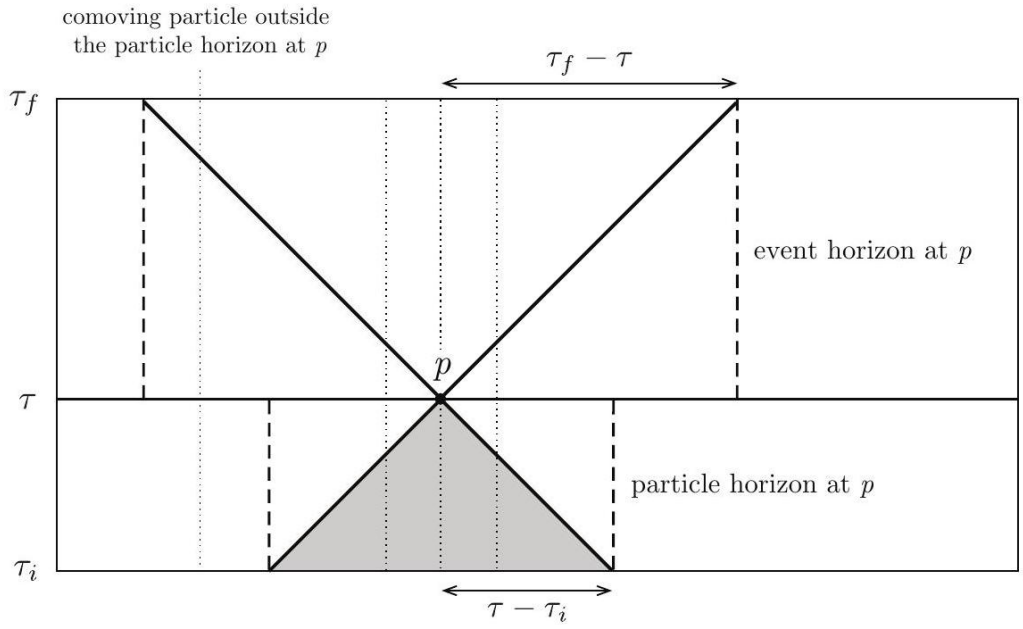


Figure 2.1: Spacetime diagram illustrating the concept of horizons. Dotted lines show the worldlines of comoving objects. The event horizon is the maximal distance to which we can send signal. The particle horizon is the maximal distance from which we can receive signals.

Particle horizon

Eq. (2.1.2) tells us that the maximal comoving distance that light can travel between two times τ_1 and $\tau_2 > \tau_1$ is simply $\Delta\tau = \tau_2 - \tau_1$ (recall that $c \equiv 1$). Hence, if the Big Bang 'started' with the singularity at $t_i \equiv 0$, [2] then the greatest comoving distance from which an observer at time t will be able to receive signals travelling at the speed of light is given by

$$\chi_{\text{ph}}(\tau) = \tau - \tau_i = \int_{t_i}^{t} \frac{\mathrm{d}t}{a(t)}.$$

This is called the (comoving) particle horizon. The size of the particle horizon at time τ may be visualised by the intersection of the past light cone of an observer p with the spacelike surface $\tau = \tau_i$ (see fig. 2.1). Causal influences have to come from within this region. Only comoving particles whose worldlines intersect the past light cone of p can send a signal to an observer at p. The boundary of the region containing such worldlines is the particle horizon at p. Notice that every observer has his of her own particle horizon.

Event horizon

Just as there are past events that we cannot see now, there may be future events that we will never be able to see (and distant regions that we will never be able to influence). In comoving coordinates, the greatest distance from which an observer at time t_f will receive signals emitted at any time later than t is given by

$$\chi_{\text{eh}}(\tau) = \tau_f - \tau = \int_{t}^{t_f} \frac{\mathrm{d}t}{a(t)}.$$

This is called the (comoving) event horizon. It is similar to the event horizon of black holes. Here, τ_f denotes the 'final moment of (conformal) time'. Notice that this may be finite even

if physical time is infinite, $t_f = +\infty$. Whether this is the case or not depends on the form of $a(t)$. In particular, τ_f is finite for our universe, if dark energy is really a cosmological constant.

[2] Notice that the Big Bang singularity is a moment in time, but not a point space. Indeed, in figs. 2.1 and 2.2 we describe the singularity by an extended (possibly infinite) spacelike hypersurface.

Growing Hubble Sphere

It is the particle horizon that is relevant for the horizon problem of the standard Big Bang cosmology. Eq. (2.1.3) can be written in the following illuminating way

$$\chi_{\rm ph}(\tau) = \int_{t_i}^{t} \frac{{\rm d}t}{a} = \int_{a_i}^{a} \frac{{\rm d}a}{a\dot{a}} = \int_{\ln a_i}^{\ln a} (aH)^{-1}\ {\rm d}\ln a,$$

where $a_i \equiv 0$ corresponds to the Big Bang singularity. The causal structure of the spacetime can hence be related to the evolution of the comoving Hubble radius $(aH)^{-1}$. For a universe dominated by a fluid with constant equation of state $w \equiv P/\rho$, we get

$$(aH)^{-1} = H_0^{-1} a^{\frac{1}{2}(1+3w)}.$$

Note the dependence of the exponent on the combination $(1+3w)$. All familiar matter sources satisfy the strong energy condition (SEC), $1+3w > 0$, so it used to be a standard assumption that the comoving Hubble radius increases as the universe expands. In this case, the integral in (2.1.5) is dominated by the upper limit and receives vanishing contributions from early times. We see this explicitly in the example of a perfect fluid. Using (2.1.6) in (2.1.5), we find

$$\chi_{\rm ph}(a) = \frac{2H_0^{-1}}{(1+3w)} \left[a^{\frac{1}{2}(1+3w)} - a_i^{\frac{1}{2}(1+3w)} \right] \equiv \tau - \tau_i.$$

The fact that the comoving horizon receives its largest contribution from late times can be made manifest by defining

$$\tau_i \equiv \frac{2H_0^{-1}}{(1+3w)} a_i^{\frac{1}{2}(1+3w)} \overset{a_i \to 0, w > -\frac{1}{3}}{\longrightarrow} 0.$$

The comoving horizon is finite,

$$\chi_{\rm ph}(t) = \frac{2H_0^{-1}}{(1+3w)} a(t)^{\frac{1}{2}(1+3w)} = \frac{2}{(1+3w)} (aH)^{-1}.$$

We see that in the standard cosmology $\chi_{\rm ph} \sim (aH)^{-1}$. This has lead to the confusing practice of referring to both the particle horizon and the Hubble radius as the "horizon" (see §2.2.2).

Why is the CMB so uniform?

About 380000 years after the Big Bang, the universe had cooled enough to allow the formation of hydrogen atoms and the decoupling of photons from the primordial plasma (see §3.3.3). We observe this event in the form of the cosmic microwave background (CMB), an afterglow of the hot Big Bang. Remarkably, this radiation is almost perfectly isotropic, with anisotropies in the CMB temperature being smaller than one part in ten thousand.

A moment's thought will convince you that the finiteness of the conformal time elapsed between $t_i = 0$ and the time of the formation of the CMB, t_{rec}, implies a serious problem:

it means that most spots in the CMB have non-overlapping past light cones and hence never were in causal contact. This is illustrated by the spacetime diagram in fig. 2.2. Consider two opposite directions on the sky. The CMB photons that we receive from these directions were emitted at the points labelled p and q in fig. 2.2. We see that the photons were emitted sufficiently close to the Big Bang singularity that the past light cones of p and q don't overlap. This implies that no point lies inside the particle horizons of both p and q. So the puzzle is: how do the photons coming from p and q "know" that they should be at almost exactly the same temperature? The same question applies to any two points in the CMB that are separated by more than 1 degree in the sky. The homogeneity of the CMB spans scales that are much larger than the particle horizon at the time when the CMB was formed. In fact, in the standard cosmology the CMB is made of about 10^4 disconnected patches of space. If there wasn't enough time for these regions to communicate, why do they look so similar? This is the horizon problem.

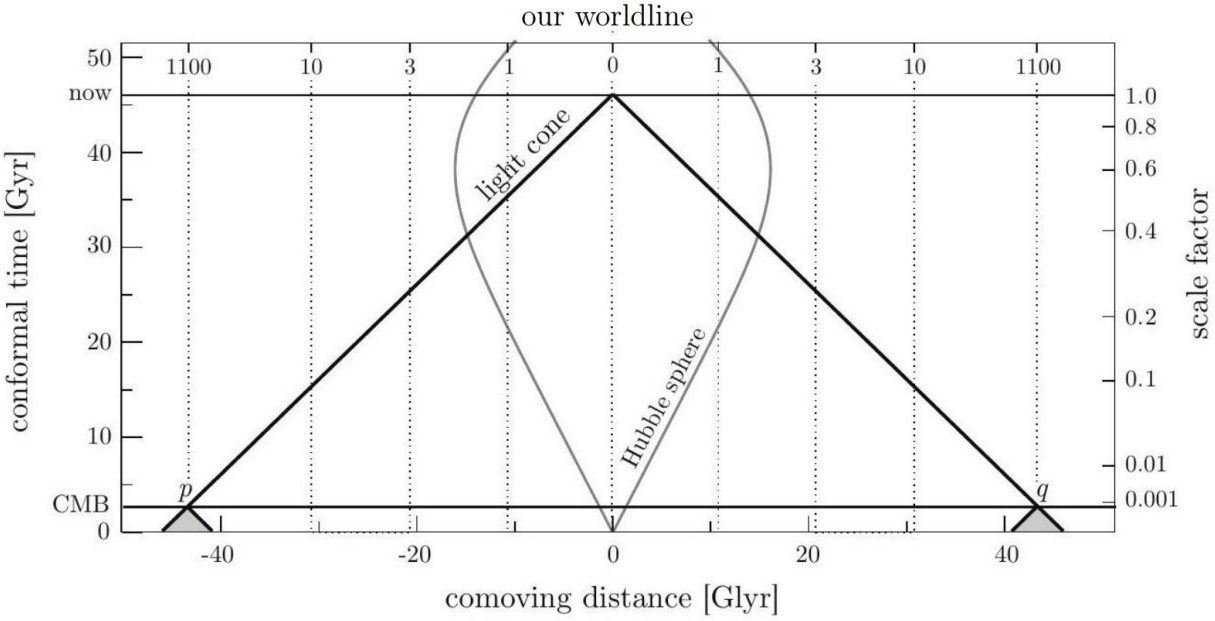


Figure 2.2: The horizon problem in the conventional Big Bang model. All events that we currently observe are on our past light cone. The intersection of our past light cone with the spacelike slice labelled CMB corresponds to two opposite points in the observed CMB. Their past light cones don't overlap before they hit the singularity, $a = 0$, so the points appear never to have been in causal contact. The same applies to any two points in the CMB that are separated by more than 1 degree on the sky.

Физически, как это, выйти за горизонт событий? (????????)

(не знаю, физически плохо это понимаю еще.)

Почему нельзя ответить, что просто системы так эволюционировали, что одинаково и излучали, потому что по сути системы в разных точек галактик аналогичны?

(хз????)

6.2.2 A Shrinking Hubble Sphere

Our description of the horizon problem has highlighted the fundamental role played by the growing Hubble sphere of the standard Big Bang cosmology. A simple solution to the horizon problem therefore suggests itself: let us conjecture a phase of decreasing Hubble radius in the early universe,

$$\frac{d}{dt}(aH)^{-1} < 0.$$

If this lasts long enough, the horizon problem can be avoided. Physically, the shrinking Hubble sphere requires a SEC-violating fluid, $1 + 3w < 0$.

Solution of the Horizon Problem

For a shrinking Hubble sphere, the integral in (2.1.5) is dominated by the lower limit. The Big Bang singularity is now pushed to negative conformal time,

$$\tau_i = \frac{2H_0^{-1}}{(1+3w)} a_i^{\frac{1}{2}(1+3w)} \overset{a_i \to 0, w < -\frac{1}{3}}{\longrightarrow} -\infty.$$

This implies that there was "much more conformal time between the singularity and decoupling than we had thought"! Fig. 2.3 shows the new spacetime diagram. The past light cones of

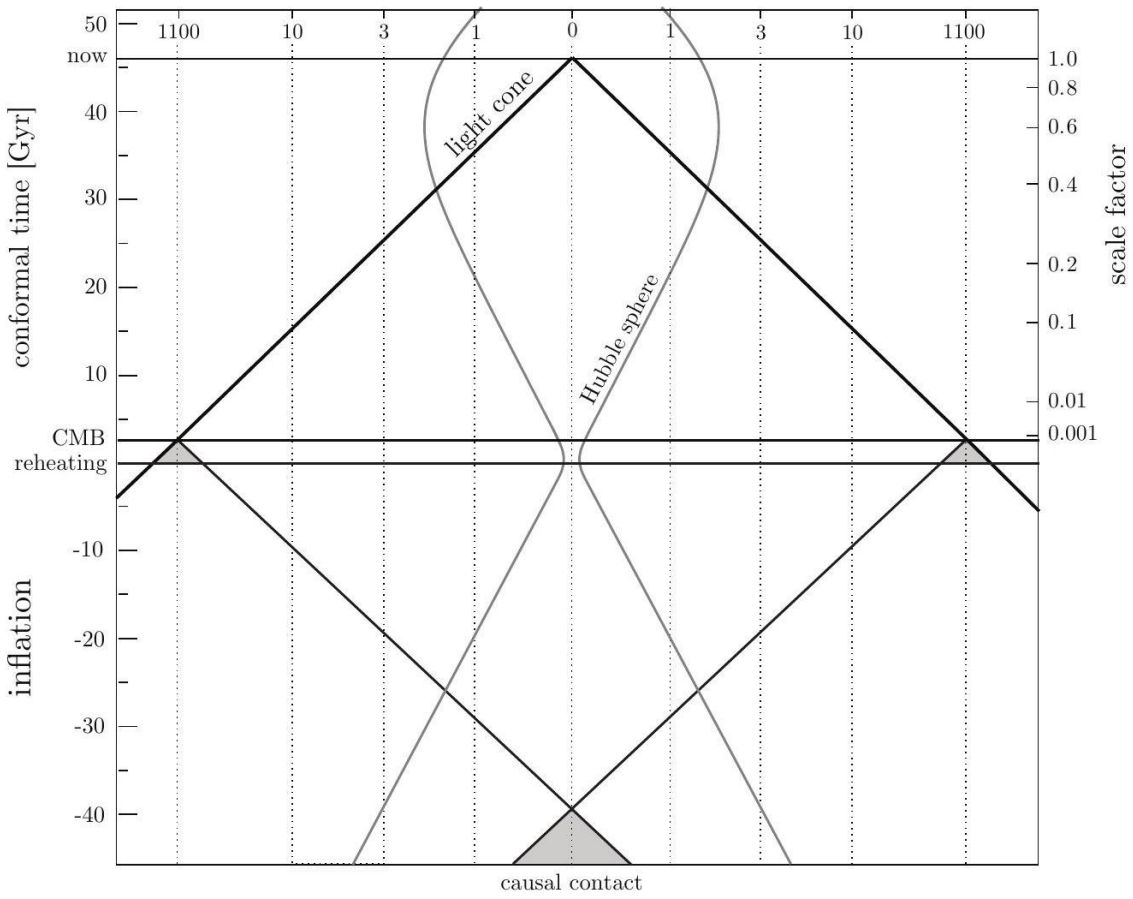


Figure 2.3: Inflationary solution to the horizon problem. The comoving Hubble sphere shrinks during inflation and expands during the conventional Big Bang evolution (at least until dark energy takes over at $a \approx 0.5$). Conformal time during inflation is negative. The spacelike singularity of the standard Big Bang is replaced by the reheating surface, i.e. rather than

marking the beginning of time it now corresponds simply to the transition from inflation to the standard Big Bang evolution. All points in the CMB have overlapping past light cones and therefore originated from a causally connected region of space.

widely separated points in the CMB now had enough time to intersect before the time τ_i. The uniformity of the CMB is not a mystery anymore. In inflationary cosmology, $\tau = 0$ isn't the initial singularity, but instead becomes only a transition point between inflation and the standard Big Bang evolution. There is time both before and after $\tau = 0$.

Hubble Radius vs. Particle Horizon

A quick word of warning about bad (but unfortunately standard) language in the inflationary literature: Both the particle horizon $\chi_{\rm ph}$ and the Hubble radius $(aH)^{-1}$ are often referred to simply as the "horizon". In the standard FRW evolution (with ordinary matter) the two are roughly the same - cf. eq. (2.1.9) - so giving them the same name isn't an issue. However, the whole point of inflation is to make the particle horizon much larger than the Hubble radius.

The Hubble radius $(aH)^{-1}$ is the (comoving) distance over which particles can travel in the course of one expansion time. [3] It is therefore another way of measuring whether particles are causally connected with each other: comparing the comoving separation λ of two particles with $(aH)^{-1}$ determines whether the particles can communicate with each other at a given moment (i.e. within the next Hubble time). This makes it clear that $\chi_{\rm ph}$ and $(aH)^{-1}$ are conceptually very different:

[3] The expansion time, $t_H \equiv H^{-1} = dt/d\ln a$, is roughly the time in which the scale factor doubles. - if $\lambda > \chi_{\rm ph}$, then the particles could never have communicated.

- if $\lambda > (aH)^{-1}$, then the particles cannot talk to each other now.

Inflation is a mechanism to achieve $\chi_{\rm ph} \gg (aH)^{-1}$. This means that particles can't communicate now (or when the CMB was created), but were in causal contact early on. In particular, the shrinking Hubble sphere means that particles which were initially in causal contact with another - i.e. separated by a distance $\lambda < (a_I H_I)^{-1}$ — can no longer communicate after a sufficiently long period of inflation: $\lambda > (aH)^{-1}$; see fig. 2.4. However, at any moment before horizon exit (careful: I really mean exit of the Hubble radius!) the particles could still talk to each other and establish similar conditions. Everything within the Hubble sphere at the beginning of inflation, $(a_I H_I)^{-1}$, was causally connected.

Since the Hubble radius is easier to calculate than the particle horizon it is common to use the Hubble radius as a means of judging the horizon problem. If the entire observable universe was within the comoving Hubble radius at the beginning of inflation-i.e. $(a_I H_I)^{-1}$ was larger than the comoving radius of the observable universe $(a_0 H_0)^{-1}$ - then there is no horizon problem. Notice that this is more conservative than using the particle horizon since $\chi_{\rm ph}(t)$ is always bigger than $(aH)^{-1}(t)$. Moreover, using $(a_I H_I)^{-1}$ as a measure of the horizon problem means that we don't have to assume anything about earlier times $t < t_I$.

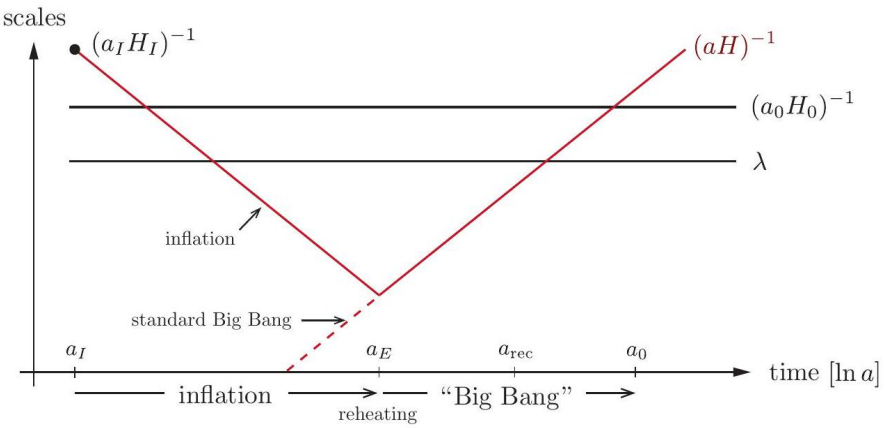


Figure 2.4: Scales of cosmological interest were larger than the Hubble radius until $a \approx 10^{-5}$ (where today is at $a(t_0) \equiv 1$). However, at very early times, before inflation operated, all scales of interest were smaller than the Hubble radius and therefore susceptible to microphysical processing. Similarly, at very late times, the scales of cosmological interest are back within the Hubble radius. Notice the symmetry of the inflationary solution. Scales just entering the horizon today, 60 e-folds after the end of inflation, left the horizon 60 e-folds before the end of inflation.

Duration of inflation.-How much inflation do we need to solve the horizon problem? At the very least, we require that the observable universe today fits in the comoving Hubble radius at the beginning of inflation,

$$(a_0 H_0)^{-1} < (a_I H_I)^{-1} .$$

Let us assume that the universe was radiation dominated since the end of inflation and ignore the relatively recent matter- and dark energy-dominated epochs. Remembering that $H \propto a^{-2}$ during radiation domination, we have

$$\frac{a_0 H_0}{a_E H_E} \sim \frac{a_0}{a_E}\left(\frac{a_E}{a_0}\right)^2 = \frac{a_E}{a_0} \sim \frac{T_0}{T_E} \sim 10^{-28},$$

where in the numerical estimate we used $T_E \sim 10^{15}$GeV and $T_0 = 10^{-3}$eV(~ 2.7 K). Eq. (2.2.12) can then be written as

$$(a_I H_I)^{-1} > (a_0 H_0)^{-1} \sim 10^{28} (a_E H_E)^{-1} .$$

For inflation to solve the horizon problem, $(aH)^{-1}$ should therefore shrink by a factor of 10^{28}. The most common way to arrange this it to have $H \sim$ const. during inflation (see below). This implies $H_I \approx H_E$, so eq. (2.2.14) becomes

$$\frac{a_E}{a_I} > 10^{28} \Rightarrow \ln\left(\frac{a_E}{a_I}\right) > 64.$$

This is the famous statement that the solution of the horizon problem requires about 60e-folds of inflation.

Conditions for Inflation

I like the shrinking Hubble sphere as the fundamental definition of inflation since it relates most directly to the horizon problem and is also key for the inflationary mechanism of generating fluctuations (see Chapter 6). However, before we move on to discuss what physics can lead to a shrinking Hubble sphere, let me show you that this definition of inflation is equivalent to other popular ways of describing inflation.

- Accelerated expansion.-From the relation

$$\frac{d}{dt}(aH)^{-1} = \frac{d}{dt}(\dot{a})^{-1} = -\frac{\ddot{a}}{(\dot{a})^2},$$

we see that a shrinking comoving Hubble radius implies accelerated expansion

$$\ddot{a} > 0.$$

This explains why inflation is often defined as a period of acceleration.

- Slowly-varying Hubble parameter.-Alternatively, we may write

$$\frac{d}{dt}(aH)^{-1} = -\frac{\dot{a}H + a\dot{H}}{(aH)^2} = -\frac{1}{a}(1-\varepsilon), \qquad \text{where } \varepsilon \equiv -\frac{\dot{H}}{H^2}.$$

The shrinking Hubble sphere therefore also corresponds to

$$\varepsilon = -\frac{\dot{H}}{H^2} < 1.$$

- Quasi-de Sitter expansion.-For perfect inflation, $\varepsilon = 0$, the spacetime becomes de Sitter space

$$\mathrm{d}s^2 = \mathrm{d}t^2 - e^{2Ht}\,\mathrm{d}\boldsymbol{x}^2,$$

where $H = \partial_t \ln a = \text{const}$. Inflation has to end, so it shouldn't correspond to perfect de Sitter space. However, for small, but finite $\varepsilon \neq 0$, the line element (2.2.20) is still a good approximation to the inflationary background. This is why we will often refer to inflation as a quasi-de Sitter period. - Negative pressure.-What forms of stress-energy source accelerated expansion? Let us consider a perfect fluid with pressure P and density ρ. The Friedmann equation, $H^2 = \rho/\left(3M_{\rm pl}^2\right)$, and the continuity equation, $\dot{\rho} = -3H(\rho + P)$, together imply

$$\dot{H} + H^2 = -\frac{1}{6M_{\rm pl}^2}(\rho + 3P) = -\frac{H^2}{2}\left(1 + \frac{3P}{\rho}\right).$$

We rearrange this to find that

$$\varepsilon = -\frac{\dot{H}}{H^2} = \frac{3}{2}\left(1 + \frac{P}{\rho}\right) < 1 \quad \Leftrightarrow \quad w \equiv \frac{P}{\rho} < -\frac{1}{3},$$

i.e. inflation requires negative pressure or a violation of the strong energy condition. How this can arise in a physical theory will be explained in the next section. We will see that there is nothing sacred about the strong energy condition and that it can easily be violated.

- Constant density.-Combining the continuitiy equation, $\dot{\rho} = -3H(\rho+P)$, with eq. (2.2.21), we find

$$\left|\frac{d\ln\rho}{d\ln a}\right| = 2\varepsilon < 1.$$

For small ε, the energy density is therefore nearly constant. Conventional matter sources all dilute with expansion, so we need to look for something more unusual.

6.2.3 The Physics of Inflation

We have shown that a given FRW spacetime with time-dependent Hubble parameter $H(t)$ corresponds to cosmic acceleration if and only if

$$\varepsilon \equiv -\frac{\dot{H}}{H^2} = -\frac{d\ln H}{dN} < 1.$$

Here, we have defined $\mathrm{d}N \equiv \mathrm{d}\ln a = H\,\mathrm{d}t$, which measures the number of e-folds N of inflationary expansion. Eq. (2.3.24) implies that the fractional change of the Hubble parameter per e-fold is small. Moreover, in order to solve the horizon problem, we want inflation to last for a sufficiently long time (usually at least $N \sim 40$ to $60e$-folds). To achieve this requires ε to remain small for a sufficiently large number of Hubble times. This condition is measured by a second parameter

$$\eta \equiv \frac{d\ln\varepsilon}{dN} = \frac{\dot{\varepsilon}}{H\varepsilon}.$$

For $|\eta| < 1$, the fractional change of ε per Hubble time is small and inflation persists. In this section, we discuss what microscopic physics can lead to the conditions $\varepsilon < 1$ and $|\eta| < 1$.

Scalar Field Dynamics

As a simple toy model for inflation we consider a scalar field, the inflaton $\phi(t, \boldsymbol{x})$. As indicated by the notation, the value of the field can depend on time t and the position in space $\boldsymbol{x}$. Associated with each field value is a potential energy density $V(\phi)$ (see fig. 2.5). If the field is dynamical (i.e. changes with time) then it also carries kinetic energy density. If the stress-energy associated with the scalar field dominates the universe, it sources the evolution of the FRW background. We want to determine under which conditions this can lead to accelerated expansion.

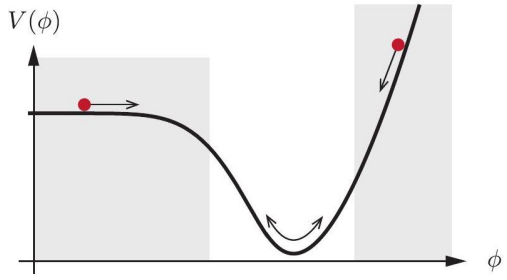


Figure 2.5: Example of a slow-roll potential. Inflation occurs in the shaded parts of the potential.

The stress-energy tensor of the scalar field is [4]

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left(\frac{1}{2}g^{\alpha\beta}\partial_\alpha\phi\partial_\beta\phi - V(\phi)\right).$$

Consistency with the symmetries of the FRW spacetime requires that the background value of the inflaton only depends on time, $\phi = \phi(t)$. From the time-time component $T^0_0 = \rho_\phi$, we infer that

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi).$$

We see that the total energy density, ρ_ϕ, is simply the sum of the kinetic energy density, $\frac{1}{2}\dot{\phi}^2$, and the potential energy density, $V(\phi)$. From the space-space component $T^i_j = -P_\phi\delta^i_j$, we find that the pressure is the difference of kinetic and potential energy densities,

$$P_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi).$$

We see that a field configuration leads to inflation, $P_\phi < -\frac{1}{3}\rho_\phi$, if the potential energy dominates over the kinetic energy.

Next, we look in more detail at the evolution of the inflaton $\phi(t)$ and the FRW scale factor $a(t)$. Substituting ρ_ϕ from (2.3.27) into the Friedmann equation, $H^2 = \rho_\phi / \left(3M_{\rm pl}^2\right)$, we get

$$H^2 = \frac{1}{3M_{\rm pl}^2}\left[\frac{1}{2}\dot{\phi}^2 + V\right].$$

Taking a time derivative, we find

$$2H\dot{H} = \frac{1}{3M_{\rm pl}^2}\left[\dot{\phi}\ddot{\phi} + V'\dot{\phi}\right],$$

where $V' \equiv dV/d\phi$. Substituting ρ_ϕ and P_ϕ into the second Friedmann equation (2.2.21), $\dot{H} = -\left(\rho_\phi + P_\phi\right) / \left(2M_{\rm pl}^2\right)$, we get

$$\dot{H} = -\frac{1}{2}\frac{\dot{\phi}^2}{M_{\rm pl}^2}.$$

Notice that $\dot{H}$ is sourced by the kinetic energy density. Combining (2.3.30) with (2.3.29) leads to the Klein-Gordon equation

$$\ddot{\phi} + 3H\dot{\phi} + V' = 0.$$

[4] You can derive this stress-energy tensor either from Noether's theorem or from the action of a scalar field. You will see those derivations in the QFT course: David Tong, Part III Quantum Field Theory. This is the evolution equation for the scalar field. Notice that the potential acts like a force, V', while the expansion of the universe adds friction, $H\dot{\phi}$.

Slow-Roll Inflation

Substituting eq. (2.3.30) into the definition of ε, eq. (2.3.24), we find

$$\varepsilon = \frac{\frac{1}{2}\dot{\phi}^2}{M_{\rm pl}^2 H^2}.$$

Inflation ($\varepsilon < 1$) therefore occurs if the kinetic energy, $\frac{1}{2}\dot{\phi}^2$, only makes a small contribution to the total energy, $\rho_\phi = 3M_{\rm pl}^2 H^2$. This situation is called slow-roll inflation.

In order for this condition to persist, the acceleration of the scalar field has to be small. To assess this, it is useful to define the dimensionless acceleration per Hubble time

$$\delta \equiv -\frac{\ddot{\phi}}{H\dot{\phi}}.$$

Taking the time-derivative of (2.3.31),

$$\dot{\varepsilon} = \frac{\dot{\phi}\ddot{\phi}}{M_{\rm pl}^2 H^2} - \frac{\dot{\phi}^2 \dot{H}}{M_{\rm pl}^2 H^3}$$

and comparing to (2.3.25), we find

$$\eta = \frac{\dot{\varepsilon}}{H\varepsilon} = 2\frac{\ddot{\phi}}{H\dot{\phi}} - 2\frac{\dot{H}}{H^2} = 2(\varepsilon - \delta).$$

Hence, $\{\varepsilon, |\delta|\} \ll 1$ implies $\{\varepsilon, |\eta|\} \ll 1$.

Slow-roll approximation.- So far, no approximations have been made. We simply noted that in a regime where $\{\varepsilon, |\delta|\} \ll 1$, inflation occurs and persists. We now use these conditions to simplify the equations of motion. This is called the slow-roll approximation. The condition $\varepsilon \ll 1$ implies $\frac{1}{2}\dot{\phi}^2 \ll V$ and hence leads to the following simplification of the Friedmann equation (F),

$$H^2 \approx \frac{V}{3M_{\rm pl}^2}.$$

In the slow-roll approximation, the Hubble expansion is determined completely by the potential energy. The condition $|\delta| \ll 1$ simplifies the Klein-Gordon equation (KG) to

$$3H\dot{\phi} \approx -V'.$$

($\mathrm{KG_{SR}}$)

This provides a simple relationship between the gradient of the potential and the speed of the inflaton. Substituting ($\mathrm{F_{SR}}$) and ($\mathrm{KG_{SR}}$) into (2.3.31) gives

$$\varepsilon = \frac{\frac{1}{2}\dot{\phi}^2}{M_{\rm pl}^2 H^2} \approx \frac{M_{\rm pl}^2}{2}\left(\frac{V'}{V}\right)^2.$$

Furthermore, taking the time-derivative of ($\mathrm{KG_{SR}}$),

$$3\dot{H}\dot{\phi} + 3H\ddot{\phi} = -V''\dot{\phi},$$

leads to

$$\delta + \varepsilon = -\frac{\ddot{\phi}}{H\dot{\phi}} - \frac{\dot{H}}{H^2} \approx M_{\rm pl}^2\frac{V''}{V}.$$

Hence, a convenient way to assess whether a given potential $V(\phi)$ can lead to slow-roll inflation is to compute the potential slow-roll parameters [5]

$$\epsilon_{\rm v} \equiv \frac{M_{\rm pl}^2}{2}\left(\frac{V'}{V}\right)^2, \quad |\eta_{\rm v}| \equiv M_{\rm pl}^2\frac{|V''|}{V}.$$

Successful slow-roll inflation occurs when these parameters are small, $\{\epsilon_{\rm v}, |\eta_{\rm v}|\} \ll 1$.

Amount of inflation. - The total number of ' e-folds' of accelerated expansion are

$$N_{\rm tot} \equiv \int_{a_I}^{a_E} \mathrm{d}\ln a = \int_{t_I}^{t_E} H(t)\mathrm{d}t,$$

where t_I and t_E are defined as the times when $\varepsilon(t_I) = \varepsilon(t_E) \equiv 1$. In the slow-roll regime, we can use

$$H\ \mathrm{d}t = \frac{H}{\dot{\phi}}\mathrm{d}\phi = \frac{1}{\sqrt{2\varepsilon}}\frac{|\mathrm{d}\phi|}{M_{\rm pl}} \approx \frac{1}{\sqrt{2\epsilon_{\rm v}}}\frac{|\mathrm{d}\phi|}{M_{\rm pl}}$$

to write (2.3.39) as an integral in the field space of the inflaton [6]

$$N_{\rm tot} = \int_{\phi_I}^{\phi_E} \frac{1}{\sqrt{2\epsilon_{\rm v}}}\frac{|\mathrm{d}\phi|}{M_{\rm pl}}$$

where ϕ_I and ϕ_E are defined as the boundaries of the interval where $\epsilon_{\rm v} < 1$. The largest scales observed in the CMB are produced about 60e-folds before the end of inflation

$$N_{\rm cmb} = \int_{\phi_{\rm cmb}}^{\phi_E} \frac{1}{\sqrt{2\epsilon_{\rm v}}} \frac{|{\rm d}\phi|}{M_{\rm pl}} \approx 60.$$

A successful solution to the horizon problem requires $N_{\rm tot} > N_{\rm cmb}$.

Case study: $m^2\phi^2$ inflation.-As an example, let us give the slow-roll analysis of arguably the simplest model of inflation: single-field inflation driven by a mass term

$$V(\phi) = \frac{1}{2} m^2 \phi^2.$$

The slow-roll parameters are

$$\epsilon_{\rm v}(\phi) = \eta_{\rm v}(\phi) = 2\left(\frac{M_{\rm pl}}{\phi}\right)^2.$$

To satisfy the slow-roll conditions $\epsilon_{\rm v}, |\eta_{\rm v}| < 1$, we therefore need to consider super-Planckian values for the inflaton

$$\phi > \sqrt{2} M_{\rm pl} \equiv \phi_E.$$

[5] In contrast, the parameters ε and η are often called the Hubble slow-roll parameters. During slow-roll, the parameters are related as follows: $\epsilon_{\rm v} \approx \varepsilon$ and $\eta_{\rm v} \approx 2\varepsilon - \frac{1}{2}\eta$.

[6] The absolute value around the integration measure indicates that we pick the overall sign of the integral in such a way as to make $N_{\rm tot} > 0$. The relation between the inflaton field value and the number of e-folds before the end of inflation is

$$N(\phi) = \frac{\phi^2}{4M_{\rm pl}^2} - \frac{1}{2}.$$

Fluctuations observed in the CMB are created at

$$\phi_{\rm cmb} = 2\sqrt{N_{\rm cmb}} M_{\rm pl} \sim 15 M_{\rm pl}.$$

Reheating

During inflation most of the energy density in the universe is in the form of the inflaton potential $V(\phi)$. Inflation ends when the potential steepens and the inflaton field picks up kinetic energy. The energy in the inflaton sector then has to be transferred to the particles of the Standard Model. This process is called reheating and starts the Hot Big Bang. We will only have time for a very brief and mostly qualitative description of the absolute basics of the reheating phenomenon. [7] This is non-examinable.

Scalar field oscillations.-After inflation, the inflaton field ϕ begins to oscillate at the bottom of the potential $V(\phi)$, see fig. 2.5. Assume that the potential can be approximated as $V(\phi) = \frac{1}{2}m^2\phi^2$ near the minimum of $V(\phi)$, where the amplitude of ϕ is small. The inflaton is still homogeneous, $\phi(t)$, so its equation of motion is

$$\ddot{\phi} + 3H\dot{\phi} = -m^2\phi.$$

The expansion time scale soon becomes much longer than the oscillation period, $H^{-1} \gg m^{-1}$. We can then neglect the friction term, and the field undergoes oscillations with frequency m. We can write the energy continuity equation as

$$\dot{\rho}_\phi + 3H\rho_\phi = -3HP_\phi = -\frac{3}{2}H\left(m^2\phi^2 - \dot{\phi}^2\right).$$

The r.h.s. averages to zero over one oscillation period. The oscillating field therefore behaves like pressureless matter, with $\rho_\phi \propto a^{-3}$. The fall in the energy density is reflected in a decrease of the oscillation amplitude.

Inflaton decay. - To avoid that the universe ends up empty, the inflaton has to couple to Standard Model fields. The energy stored in the inflaton field will then be transferred into ordinary particles. If the decay is slow (which is the case if the inflaton can only decay into fermions) the inflaton energy density follows the equation

$$\dot{\rho}_\phi + 3H\rho_\phi = -\Gamma_\phi\rho_\phi,$$

where Γ_ϕ parameterizes the inflaton decay rate. If the inflaton can decay into bosons, the decay may be very rapid, involving a mechanism called parametric resonance (sourced by Bose condensation effects). This kind of rapid decay is called preheating, since the bosons thus created are far from thermal equilibrium.

Thermalisation.- The particles produced by the decay of the inflaton will interact, create other particles through particle reactions, and the resulting particle soup will eventually reach thermal

[7] For more details see Baumann, The Physics of Inflation, DAMTP Lecture Notes. equilibrium with some temperature $T_{\rm rh}$. This reheating temperature is determined by the energy density $\rho_{\rm rh}$ at the end of the reheating epoch. Necessarily, $\rho_{\rm rh} < \rho_{\phi,E}$ (where $\rho_{\phi,E}$ is the inflaton energy density at the end of inflation). If reheating takes a long time, we may have $\rho_{\rm rh} \ll \rho_{\phi,E}$. The evolution of the gas of particles into a thermal state can be quite involved. Usually it is just assumed that it happens eventually, since the particles are able to interact. However, it is possible that some particles (such as gravitinos) never reach thermal equilibrium, since their interactions are so weak. In any case, as long as the momenta of the particles are much higher than their masses, the energy density of the universe behaves like radiation regardless of the momentum space distribution. After thermalisation of at least the baryons, photons and neutrinos is complete, the standard Hot Big Bang era begins.

6.3 Thermal History

In this chapter, we will describe the first three minutes [1] in the history of the universe, starting from the hot and dense state following inflation. At early times, the thermodynamical properties of the universe were determined by local equilibrium. However, it are the departures from thermal equilibrium that make life interesting. As we will see, non-equilibrium dynamics allows massive particles to acquire cosmological abundances and therefore explains why there is something rather than nothing. Deviations from equilibrium are also crucial for understanding the origin of the cosmic microwave background and the formation of the light chemical elements.

We will start, in §3.1, with a schematic description of the basic principles that shape the thermal history of the universe. This provides an overview of the story that will be fleshed out in much more detail in the rest of the chapter: in §3.2, will present equilibrium thermodynamics in an expanding universe, while in 3.3, we will introduce the Boltzmann equation and apply it to several examples of non-equilibrium physics. We will use units in which Boltzmann's constant is set equal to unity, $k_{\rm B} \equiv 1$, so that temperature has units of energy.

6.3.1 The Hot Big Bang

The key to understanding the thermal history of the universe is the comparison between the rate of interactions Γ and the rate of expansion H. When $\Gamma \gg H$, then the time scale of particle interactions is much smaller than the characteristic expansion time scale:

$$t_c \equiv \frac{1}{\Gamma} \ll t_H \equiv \frac{1}{H}.$$

Local thermal equilibrium is then reached before the effect of the expansion becomes relevant. As the universe cools, the rate of interactions may decrease faster than the expansion rate. At $t_c \sim t_H$, the particles decouple from the thermal bath. Different particle species may have different interaction rates and so may decouple at different times.

Local Thermal Equilibrium

Let us first show that the condition (3.1.1) is satisfied for Standard Model processes at temperatures above a few hundred GeV. We write the rate of particle interactions as [2]

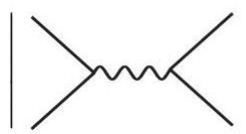
$$\Gamma \equiv n\sigma v$$

where n is the number density of particles, σ is their interaction cross section, and v is the average velocity of the particles. For $T \gtrsim 100\text{GeV}$, all known particles are ultra-relativistic,

[1] A wonderful popular account of this part of cosmology is Weinberg's book The First Three Minutes.

[2] For a process of the form $1+2 \leftrightarrow 3+4$, we would write the interaction rate of species 1 as $\Gamma_1 = n_2\sigma v$, where n_2 is the density of the target species 2 and v is the average relative velocity of 1 and 2. The interaction rate of species 2 would be $\Gamma_2 = n_1\sigma v$. We have used the expectation that at high energies $n_1 \sim n_2 \equiv n.$ and hence $v \sim 1$. Since particle masses can be ignored in this limit, the only dimensionful scale is the temperature T. Dimensional analysis then gives $n \sim T^3$. Interactions are mediated by gauge bosons, which are massless above the scale of electroweak symmetry breaking. The cross sections for the strong and electroweak interactions then have a similar dependence, which also can be estimated using dimensional analysis [3]

$$\sigma \sim \left| \quad \right|^2 \sim \frac{\alpha^2}{T^2},$$



where $\alpha \equiv g_A^2/4\pi$ is the generalized structure constant associated with the gauge boson A. We find that

$$\Gamma = n\sigma v \sim T^3 \times \frac{\alpha^2}{T^2} = \alpha^2 T.$$

We wish to compare this to the Hubble rate $H \sim \sqrt{\rho}/M_{\text{pl}}$. The same dimensional argument as before gives $\rho \sim T^4$ and hence

$$H \sim \frac{T^2}{M_{\text{pl}}^2}.$$

The ratio of (3.1.4) and (3.1.5) is

$$\frac{\Gamma}{H} \sim \frac{\alpha^2 M_{\text{pl}}}{T} \sim \frac{10^{16}\text{GeV}}{T}$$

where we have used $\alpha \sim 0.01$ in the numerical estimate. Below $T \sim 10^{16}$GeV, but above 100 GeV, the condition (3.1.1) is therefore satisfied.

When particles exchange energy and momentum efficiently they reach a state of maximum entropy. It is a standard result of statistical mechanics that the number of particles per unit volume in phase space - the distribution function - then takes the form [4]

$$f(E) = \frac{1}{e^{E/T} \pm 1},$$

where the + sign is for fermions and the - sign for bosons. When the temperature drops below the mass of the particles, $T \ll m$, they become non-relativistic and their distribution function receives an exponential suppression, $f \to e^{-m/T}$. This means that relativistic particles ('radiation') dominate the density and pressure of the primordial plasma. The total energy density is therefore well approximated by summing over all relativistic particles, $\rho_r \propto \sum_i \int \mathrm{d}^3 p f_i(p) E_i(p)$. The result can be written as (see below)

$$\rho_r = \frac{\pi^2}{30} g_\star(T) T^4,$$

where $g_\star(T)$ is the number of relativistic degrees of freedom. Fig. 3.1 shows the evolution of $g_\star(T)$ assuming the particle content of the Standard Model. At early times, all particles are relativistic and $g_\star = 106.75$. The value of $g_\star$ decreases whenever the temperature of the universe drops below the mass of a particle species and it becomes non-relativistic. Today, only photons and (maybe) neutrinos are still relativistic and $g_\star = 3.38$.

[3] Shown in eq. (3.1.3) is the Feynman diagram associated with a $2 \to 2$ scattering process mediated by the exchange of a gauge boson. Each vertex contributes a factor of the gauge coupling $g_A \propto \sqrt{\alpha}$. The dependence of the cross section on α follows from squaring the dependence on α derived from the Feynman diagram, i.e. $\sigma \propto (\sqrt{\alpha} \times \sqrt{\alpha})^2 = \alpha^2$. For more details see the Part III Standard Model course.

[4] The precise formula will include the chemical potential - see below.

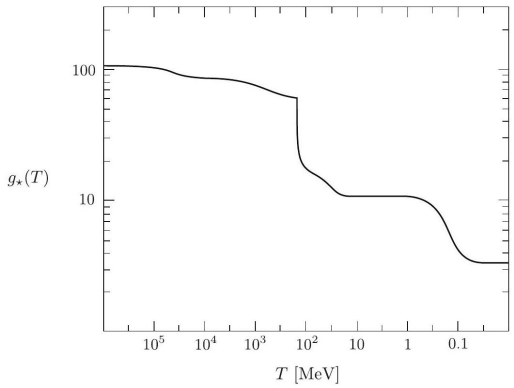


Figure 3.1: Evolution of the number of relativistic degrees of freedom assuming the Standard Model.

Decoupling and Freeze-Out

If equilibrium had persisted until today, the universe would be mostly photons. Any massive particle species would be exponentially suppressed. [5] To understand the world around us, it is therefore crucial to understand the deviations from equilibrium that led to the freeze-out of massive particles (see fig. 3.2).

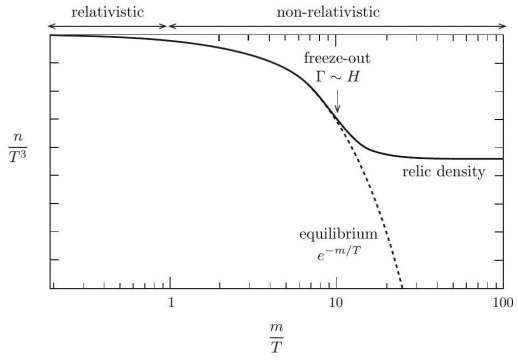


Figure 3.2: A schematic illustration of particle freeze-out. At high temperatures, $T \gg m$, the particle abundance tracks its equilibrium value. At low temperatures, $T \ll m$, the particles freeze out and maintain a density that is much larger than the Boltzmann-suppressed equilibrium abundance.

Below the scale of electroweak symmetry breaking, $T \lesssim 100\text{GeV}$, the gauge bosons of the weak interactions, $W^{\pm}$and Z, receive masses $M_W \sim M_Z$. The cross section associated with

[5] This isn't quite correct for baryons. Since baryon number is a symmetry of the Standard Model, the number density of baryons can remain significant even in equilibrium. processes mediated by the weak force becomes

$$\sigma \sim \left| \quad \times \quad \right|^2 \sim G_F^2 T^2 \ ,$$

where we have introduced Fermi's constant, [6] $G_F \sim \alpha/M_W^2 \sim 1.17 \times 10^{-5}\text{GeV}^{-2}$. Notice that the strength of the weak interactions now decreases as the temperature of the universe drops. We find that

$$\frac{\Gamma}{H} \sim \frac{\alpha^2 M_{\rm pl} T^3}{M_W^4} \sim \left(\frac{T}{1\text{MeV}}\right)^3 ,$$

which drops below unity at $T_{\rm dec} \sim 1\text{MeV}$. Particles that interact with the primordial plasma only through the weak interaction therefore decouple around 1MeV. This decoupling of weak scale interactions has important consequences for the thermal history of the universe.

A Brief History of the Universe

Table 3.1 lists the key events in the thermal history of the universe:

- Baryogenesis. Relativistic quantum field theory requires the existence of anti-particles (see Part III Quantum Field Theory). This poses a slight puzzle. Particles and antiparticles annihilate through processes such as $e^+ + e^- \to \gamma + \gamma$. If initially the universe was filled with equal amounts of matter and anti-matter then we expect these annihilations to lead to a universe dominated by radiation. However, we do observe an overabundance of matter (mostly baryons) over anti-matter in the universe today. Models of baryogenesis try to derive the observed baryon-to-photon ratio

$$\eta \equiv \frac{n_b}{n_\gamma} \sim 10^{-9},$$

from some dynamical mechanism, i.e. without assuming a primordial matter-antimatter asymmetry as an initial condition. Although many ideas for baryogenesis exist, none is singled out by experimental tests. We will not have much to say about baryogenesis in this course.

- Electroweak phase transition. At 100GeV particles receive their masses through the Higgs mechanism. Above we have seen how this leads to a drastic change in the strength of the weak interaction.

- QCD phase transition. While quarks are asymptotically free (i.e. weakly interacting) at high energies, below 150MeV, the strong interactions between the quarks and the gluons become important. Quarks and gluons then form bound three-quark systems, called baryons, and quark-antiquark pairs, called mesons. These baryons and mesons are the relevant degrees of freedom below the scale of the QCD phase transition.

- Dark matter freeze-out. Since dark matter is very weakly interacting with ordinary matter we expect it to decouple relatively early on. In §3.3.2, we will study the example of WIMPs - weakly interacting massive particles that freeze out around 1MeV. We will

[6] The $1/M_W^2$ comes from the low-momentum limit of the propagator of a massive gauge field.

Event	time t	redshift z	temperature T
Inflation	10^{-34} s (?)	–	–
Baryogenesis	?	?	?
EW phase transition	20 ps	10^{15}	100 GeV
QCD phase transition	20 μs	10^{12}	150 MeV
Dark matter freeze-out	?	?	?
Neutrino decoupling	1 s	6×10^9	1 MeV
Electron-positron annihilation	6 s	2×10^9	500 keV
Big Bang nucleosynthesis	3 min	4×10^8	100 keV
Matter-radiation equality	60 kyr	3400	0.75 eV
Recombination	260–380 kyr	1100–1400	0.26–0.33 eV
Photon decoupling	380 kyr	1000–1200	0.23–0.28 eV
Reionization	100–400 Myr	11–30	2.6–7.0 meV
Dark energy-matter equality	9 Gyr	0.4	0.33 meV
Present	13.8 Gyr	0	0.24 meV

Table 3.1: Key events in the thermal history of the universe.

show that choosing natural values for the mass of the dark matter particles and their interaction cross section with ordinary matter reproduces the observed relic dark matter density surprisingly well.

- Neutrino decoupling. Neutrinos only interact with the rest of the primordial plasma through the weak interaction. The estimate in (3.1.10) therefore applies and neutrinos decouple at 0.8MeV.

- Electron-positron annihilation. Electrons and positrons annihilate shortly after neutrino decoupling. The energies of the electrons and positrons gets transferred to the photons, but not the neutrinos. In §3.2.4, we will explain that this is the reason why the photon temperature today is greater than the neutrino temperature.

- Big Bang nucleosynthesis. Around 3 minutes after the Big Bang, the light elements were formed. In §3.3.4, we will study this process of Big Bang nucleosynthesis (BBN).

- Recombination. Neutral hydrogen forms through the reaction $e^- + p^+ \to \mathrm{H} + \gamma$ when the temperature has become low enough that the reverse reaction is energetically disfavoured. We will study recombination in §3.3.3. - Photon decoupling. Before recombination the strongest coupling between the photons and the rest of the plasma is through Thomson scattering, $e^- + \gamma \to e^- + \gamma$. The sharp drop in the free electron density after recombination means that this process becomes inefficient and the photons decouple. They have since streamed freely through the universe and are today observed as the cosmic microwave background (CMB).

In the rest of this chapter we will explore in detail where this knowledge about the thermal history of the universe comes from.

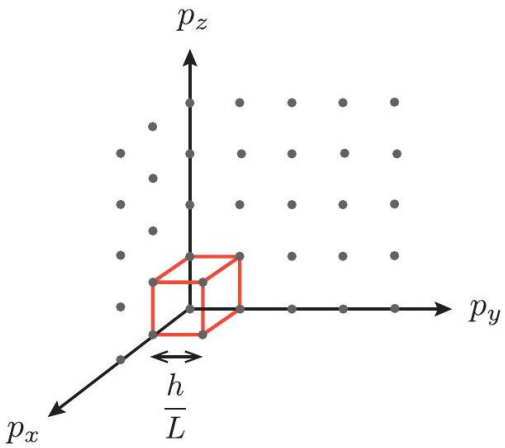
6.3.2 Equilibrium

Equilibrium Thermodynamics

We have good observational evidence (from the perfect blackbody spectrum of the CMB) that the early universe was in local thermal equilibrium. [7] Moreover, we have seen above that the Standard Model predicts thermal equilibrium above 100GeV. To describe this state and the subsequent evolution of the universe, we need to recall some basic facts of equilibrium thermodynamics, suitably generalized to apply to an expanding universe.

Microscopic to Macroscopic

Statistical mechanics is the art of turning microscopic laws into an understanding of the macroscopic world. I will briefly review this approach for a gas of weakly interacting particles. It is convenient to describe the system in phase space, where the gas is described by the positions and momenta of all particles. In quantum mechanics, the momentum eigenstates of a particle in a volume $V = L^3$ have a discrete spectrum:



The density of states in momentum space $\{\boldsymbol{p}\}$ then is $L^3/h^3 = V/h^3$, and the state density in phase space $\{\boldsymbol{x}, \boldsymbol{p}\}$ is

$$\frac{1}{h^3}.$$

If the particle has g internal degrees of freedom (e.g. spin), then the density of states becomes

$$\frac{g}{h^3} = \frac{g}{(2\pi)^3},$$

[7] Strictly speaking, the universe can never truly be in equilibrium since the FRW spacetime doesn't posses a time-like Killing vector. But this is physics not mathematics: if the expansion is slow enough, particles have enough time to settle close to local equilibrium. (And since the universe is homogeneous, the local values of thermodynamics quantities are also global values.)

where in the second equality we have used natural units with $\hbar = h/(2\pi) \equiv 1$. To obtain the number density of a gas of particles we need to know how the particles are distributed amongst the momentum eigenstates. This information is contained in the (phase space) distribution function $f(\boldsymbol{x}, \boldsymbol{p}, t)$. Because of homogeneity, the distribution function should, in fact, be independent of the position $\boldsymbol{x}$. Moreover, isotropy requires that the momentum dependence is only in terms of the magnitude of the momentum $p \equiv |\boldsymbol{p}|$. We will typically leave the time dependence implicitit will manifest itself in terms of the temperature dependence of the distribution functions. The particle density in phase space is then the density of states times the distribution function

$$\frac{g}{(2\pi)^3} \times f(p).$$

The number density of particles (in real space) is found by integrating (3.2.14) over momentum,

$$n = \frac{g}{(2\pi)^3} \int \mathrm{d}^3 p f(p).$$

To obtain the energy density of the gas of particles, we have to weight each momentum eigenstate by its energy. To a good approximation, the particles in the early universe were weakly interacting. This allows us to ignore the interaction energies between the particles and write the energy of a particle of mass m and momentum p simply as

$$E(p) = \sqrt{m^2 + p^2}.$$

Integrating the product of (3.2.16) and (3.2.14) over momentum then gives the energy density

$$\rho = \frac{g}{(2\pi)^3} \int \mathrm{d}^3 p f(p) E(p).$$

Similarly, we define the pressure as

$$P = \frac{g}{(2\pi)^3} \int \mathrm{d}^3 p f(p) \frac{p^2}{3E}$$

Pressure. *-Let me remind you where the $p^2/3E$ factor in (3.2.18) comes from. Consider a small area element of size $\mathrm{d}A$, with unit normal vector $\hat{\boldsymbol{n}}$ (see fig. 3.3). All particles with velocity $|\boldsymbol{v}|$, striking this area element in the time interval between t and $t + \mathrm{d}t$, were located at $t = 0$ in a spherical shell of radius $R = |\boldsymbol{v}|t$ and width $|\boldsymbol{v}|\mathrm{d}t$. A solid angle $\mathrm{d}\Omega^2$ of this shell defines the volume $\mathrm{d}V = R^2 |\boldsymbol{v}| \mathrm{d}t \, \mathrm{d}\Omega^2$ (see the grey shaded region in fig. 3.3). Multiplying the phase space density (3.2.14) by $\mathrm{d}V$ gives the number of particles in the volume (per unit volume in momentum space) with energy $E(|\boldsymbol{v}|)$,

$$\mathrm{d}N = \frac{g}{(2\pi)^3} f(E) \times R^2 |\boldsymbol{v}| \mathrm{d}t \, \mathrm{d}\Omega.$$

Not all particles in $\mathrm{d}V$ reach the target, only those with velocities directed to the area element. Taking into account the isotropy of the velocity distribution, we find that the total number of particles striking the area element $\mathrm{d}A\hat{\boldsymbol{n}}$ with velocity $\boldsymbol{v} = |\boldsymbol{v}|\hat{\boldsymbol{v}}$ is

$$\mathrm{d}N_A = \frac{|\hat{\boldsymbol{v}}\cdot\hat{\boldsymbol{n}}|\mathrm{d}A}{4\pi R^2}\times \mathrm{d}N = \frac{g}{(2\pi)^3}f(E)\times\frac{|\boldsymbol{v}\cdot\hat{\boldsymbol{n}}|}{4\pi}\mathrm{d}A\ \mathrm{d}t\ \mathrm{d}\Omega,$$

where $\boldsymbol{v}\cdot\hat{\boldsymbol{n}}<0$. If these particles are reflected elastically, each transfer momentum $2|\boldsymbol{p}\cdot\hat{\boldsymbol{n}}|$ to the target. Therefore, the contribution of particles with velocity $|\boldsymbol{v}|$ to the pressure is

$$\mathrm{d}P(|\boldsymbol{v}|) = \int \frac{2|\boldsymbol{p}\cdot\hat{\boldsymbol{n}}|}{\mathrm{d}A\ \mathrm{d}t}\ \mathrm{d}N_A = \frac{g}{(2\pi)^3}f(E)\times\frac{p^2}{2\pi E}\int\cos^2\theta\sin\theta\mathrm{d}\theta\mathrm{d}\phi = \frac{g}{(2\pi)^3}\times f(E)\frac{p^2}{3E},$$

where we have used $|\boldsymbol{v}| = |\boldsymbol{p}|/E$ and integrated over the hemisphere defined by $\hat{\boldsymbol{v}}\cdot\hat{\boldsymbol{n}}\equiv-\cos\theta<0$ (i.e. integrating only over particles moving towards $\mathrm{d}A$ - see fig. 3.3). Integrating over energy E (or momentum p), we obtain (3.2.18).

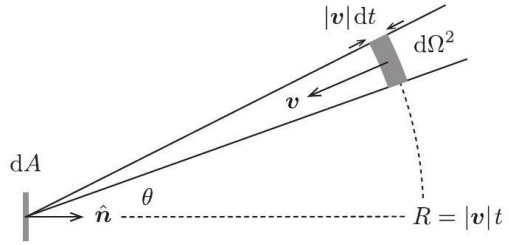


Figure 3.3: Pressure in a weakly interacting gas of particles.

Local Thermal Equilibrium

A system of particles is said to be in kinetic equilibrium if the particles exchange energy and momentum efficiently. This leads to a state of maximum entropy in which the distribution functions are given by the Fermi-Dirac and Bose-Einstein distributions [8]

$$f(p) = \frac{1}{e^{(E(p)-\mu)/T}\pm 1},$$

where the + sign is for fermions and the - sign for bosons. At low temperatures, $T<E-\mu$, both distribution functions reduce to the Maxwell-Boltzmann distribution

$$f(p)\approx e^{-(E(p)-\mu)/T}.$$

The equilibrium distribution functions have two parameters: the temperature T and the chemical potential μ. The chemical potential may be temperature-dependent. As the universe expands, T and $\mu(T)$ change in such a way that the continuity equations for the energy density ρ and the particle number density n are satisfied. Each particle species i (with possibly distinct m_i, μ_i, T_i) has its own distribution function f_i and hence its own n_i, ρ_i, and P_i.

Chemical potential.* - In thermodynamics, the chemical potential characterizes the response of a system to a change in particle number. Specifically, it is defined as the derivative of the entropy with respect to the number of particles, at fixed energy and fixed volume,

$$\mu = -T\left(\frac{\partial S}{\partial N}\right)_{U,V}.$$

[8] We use units where Boltzmann's constant is $k_\mathrm{B}\equiv 1$.

6.3.3 Isotropic Space by Landau

вкратце

Theory

Требования, поставленные в $93 в качестве условий для определения действия гравитационного поля, будут по-прежнему удовлетворены, если к скаляру G добавить постоянный член, т. е. если положить

$$S_g = -\frac{c^3}{16\pi k}\int (G + 2\Lambda)\sqrt{-g}d\Omega$$

где Λ - новая постоянная (с размерностью см $^{-2}$).

Такое изменение приведет к появлению в уравнениях Эйнштейна дополнительного члена Λg_{ik}:

$$R_{ik} - \frac{1}{2}Rg_{ik} = \frac{8\pi k}{c^4}T_{ik} + \Lambda g_{ik}$$

Если приписать «космологической постоянной» Λ очень малое значение, то наличие этого члена не будет сказываться существенным образом на гравитационных полях в не слишком больших областях пространства-времени, но приведет к появлению новых типов «космологических решений», которые могли бы описывать мир в целом.

В настоящее время, однако, нет никаких настоятельных и убедительных оснований - как наблюдательных, так и теоретических для такого видоизменения основных уравнений теории.

(?? да??)

Подчеркнем, что речь шла бы об изменении, имеющем глубокий физический смысл: введение в плотность лагранжевой функции постоянного члена, вообще не зависящего от состояния поля, означало бы приписывание пространству-времени принципиально неустранимой кривизны, не связанной ни с материей, ни с гравитационными волнами.

Все дальнейшее изложение в этой главе основано поэтому на уравнениях Эйнштейна в их виде без космологической постоянной.

Как известно, звезды распределены по пространству весьма неравномерным образом - они сконцентрированы в отдельных звездных системах (галактиках).

Но при исследовании Вселенной «в больших масштабах» следует отвлекаться от «местныХ» неоднородностей, вызванных скоПлением вещества в звезды и звездные системы.

Так, под плотностью масс должна подразумеваться плотность, усредненная по областям пространства, размеры которых велики по сравнению с расстояниями между галактиками.

Рассматриваемые ниже решения уравнений Эйнштейна - так называемая изотропная космологическая модель Фридмана - основаны на предположении об однородности и изотропии распределения вещества по пространству.

Существующие астрономические данные не противоречат такому предположению,

и в настоящее время есть все основания считать,

что изотропная модель дает в общих чертах адекватное описание не только современного состояния Вселенной,

но и значительной доли ее эволюции в прошлом.

Мы увидим ниже, что основным свойством этой модели является ее нестационарность.

Нет сомнения в том, что это свойство («расширяющаяся Вселенная») дает правильное объяснение фундаментального для космологической проблемы явления красного смещения ($ 114).

В то же время ясно, что предположение об однородности и изотропии Вселенной уже по самому своему существу неизбежно может иметь лишь приближенный характер, поскольку эти свойства заведомо нарушаются при переходе к меньшим масштабам.

К вопросу о возможной роли неоднородности Вселенной в различных аспектах космологической проблемы мы вернемся в \$115 – \$119.

Однородность и изотропия пространства означают, что можно выбрать такое мировое время, чтобы в каждый его момент метрика пространства была одинаковой во всех точках и по всем направлениям.

Займемся прежде всего изучением метрики изотропного пространства как таковой, не интересуясь пока его возможной зависимостью от времени.

Скалярная кривизна

Как мы уже делали выше, обозначим трехмерный метрический тензор как $\gamma_{\alpha\beta}$, т. е. напишем элемент пространственного расстояния в виде

$$dl^2 = \gamma_{\alpha\beta} dx^\alpha dx^\beta$$

Кривизна пространства полностью определяется его трехмерным тензором кривизны, который мы обозначаем как $P_{\alpha\beta\gamma\delta}$, в отличие от четырехмерного тензора R_{iklm}. В случае полной изотропии тензор $P_{\alpha\beta\gamma\delta}$ должен, очевидно, выражаться только через метрический тензор $\gamma_{\alpha\beta}$, а потому в силу своих свойств симметрии должен иметь вид

$$P_{\alpha\beta\gamma\delta} = \lambda \left(\gamma_{\alpha\gamma}\gamma_{\beta\delta} - \gamma_{\alpha\delta}\gamma_{\beta\gamma}\right)$$

где λ - постоянная.

Тензор Риччи $P_{\alpha\beta} = P^\gamma_{\alpha\gamma\beta}$ равен соответственно

$$P_{\alpha\beta} = 2\lambda\gamma_{\alpha\beta}$$

а скалярная кривизна

$$P = 6\lambda$$

Таким образом, свойства кривизны изотропного пространства определяются лишь одной постоянной.

Три типа метрики

Соответственно этому возможны всего три существенных различных случая пространственной метрики

1) так называемое пространство постоянной положительной кривизны (при $\lambda > 0$),

2) пространство постоянной отрицательной кривизны (при $\lambda < 0$) и

3) пространство с кривизной, равной нулю ($\lambda = 0$).

Из них последнее представляет собой плоское, т.

е.

евклидово пространство.

При изучении метрики удобно исходить из геометрической аналогии,

рассматривая геометрию изотропного трехмерного пространства как геометрию на заведомо изотропной гиперповерхности (в некотором фиктивном четырехмерном пространстве).

Такой поверхностью является гиперсфера; соответствующее ей трехмерное пространство и является пространством положительной постоянной кривизны.

Уравнение гиперсферы с радиусом a в четырехмерном пространстве x_1, x_2, x_3, x_4 имеет вид

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = a^2$$

а элемент длины на ней выражается как

$$dl^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2$$

Рассматривая координаты x_1, x_2, x_3 как три пространственные координаты и исключая из dl^2 фиктивную координату x_4 с помощью первого уравнения,

находим элемент пространственного расстояния в виде

$$dl^2 = dx_1^2 + dx_2^2 + dx_3^2 + \frac{(x_1 dx_1 + x_2 dx_2 + x_3 dx_3)^2}{a^2 - x_1^2 - x_2^2 - x_3^2}$$

Из этого выражения легко вычислить постоянную λ В (111.2).

Поскольку нам заранее известно, что тензор $P_{\alpha\beta}$ имеет вид (111.3)

во всем пространстве, то достаточно вычислить его только для точек, находящихся вблизи начала координат, где $\gamma_{\alpha\beta}$ равны

$$\gamma_{\alpha\beta} = \delta_{\alpha\beta} + \frac{x_\alpha x_\beta}{a^2}$$

Так как первые производные от $\gamma_{\alpha\beta}$, а значит, и величины $\lambda^\alpha_{\beta\gamma}$ (ср.

задачу 1 §88) - трехмерные символы Кристоффеля, соответствующие метрике $\gamma_{\alpha\beta}$, - в начале координат обращаются в нуль, то вычисление по общей формуле (92.7) оказывается очень простым и дает в результате

$$\lambda = \frac{1}{a^2}$$

Величину a можно назвать «радиусом кривизны» пространства.

Введем вместо координат x_1, x_2, x_3 соответствующие им «сферические» координаты r, θ, φ. Тогда элемент длины примет вид

$$dl^2 = \frac{dr^2}{1 - \frac{r^2}{a^2}} + r^2 \left(\sin^2\theta d\varphi^2 + d\theta^2\right)$$

Начало координат может быть выбрано в любой точке пространства.

Длина окружности в этих координатах равна $2\pi r$, а поверхность сферы $4\pi r^2$. Длина же «радиуса» окружности (или сферы) равна

$$\int_0^r \frac{dr}{\sqrt{1 - \frac{r^2}{a^2}}} = a \arcsin\frac{r}{a}$$

т.

е.

больше r. Таким образом, отношение длины окружности к радиусу в таком пространстве меньше, чем 2π.

Другую удобную форму dl^2 имеет в «четырехмерных сферических координатах»,

получающихся, если ввести вместо координаты r «угол» χ согласно $r = a\sin\chi$(χменяется в предел

Тогда

$$dl^2 = a^2\left[d\chi^2 + \sin^2\chi\left(\sin^2\theta d\varphi^2 + d\theta^2\right)\right]$$

Координата χ измеряет расстояние от начала координат, равное $a\chi$. Поверхность сферы в этих координатах равна $4\pi a^2 \sin^2\chi$. Мы видим, что по мере удаления от начала координат величина поверхности сферы увеличивается, пока не достигнет на расстоянии $\pi a/2$ максимального значения, равного $4\pi a^2$. Вслед за этим она начинает уменьшаться, пока не превратится в точку на «противоПоложном полюсе» простран-

ства на расстоянии πa - наибольшем расстоянии, которое вообще может существовать в таком пространстве (все это видно, конечно, И из (111.7), если заметить, что координата r не может принимать значений больших, чем a).

Объем пространства с положительной кривизной равен

$$V = \int_0^{2\pi}\int_0^{\pi}\int_0^{\pi} a^3 \sin^2\chi \sin\theta d\chi d\theta d\varphi$$

откуда

$$V = 2\pi^2 a^3$$

Таким образом, пространство положительной кривизны оказывается «замкнутым само в себе» - конечным по объему,

но, разумеется, не имеющим границ.

Интересно отметить, что в замкнутом пространстве полный электрический заряд должен быть равен нулю.

Действительно, всякая замкнутая поверхность в конечном пространстве с обеих своих сторон охватывает конечные же области пространства.

Поэтому поток электрического поля через эти поверхности равен, с одной стороны, полному заряду, находящемуся внутри поверхности, а с другой, равен находящемуся вне ее заряду, взятому с обратным знаком.

Сумма же зарядов с обеих сторон поверхности равна, следовательно,

нулю.

Аналогичным образом, из выражения (96.16) 4-импульса в виде интеграла по поверхности следует обращение в нуль полного 4-импульса P^i во всем пространстве.

Пространство постоянной отрицательной кривизны

Перейдем теперь к рассмотрению геометрии пространства,

обладающего постоянной отрицательной кривизной.

$_3$(111.6) мы видим, что постоянная λ становится отрицательной, если a мнимо.

Поэтому все формулы для пространства отрицательной кривизны можно непосредственно получить из предыдущих, заменив в них a на ia.

Другими словами, геометрия пространства отрицательной кривизны получается математически как геометрия на четырехмерной псевдосфере с мнимым радиусом.

Таким образом, постоянная λ равна теперь

$$\lambda = -\frac{1}{a^2}$$

а элемент длины в пространстве отрицательной кривизны в координатах r, θ, φ имеет вид

$$dl^2 = \frac{dr^2}{1+\frac{r^2}{a^2}} + r^2\left(\sin^2\theta d\varphi^2 + d\theta^2\right)$$

где координата r может пробегать все значения от 0 до ∞.

Отношение длины окружности к радиусу теперь больше, чем 2π. Выражение

Для dl^2, соответствующее (111.8), получится,

если ввести координату χ согласно $r = a\,\mathrm{sh}\,\chi$(χ меняется здесь от 0 до ∞).

Тогда

$$dl^2 = a^2\left\{d\chi^2 + \mathrm{sh}^2\chi\left(\sin^2\theta d\varphi^2 + d\theta^2\right)\right\}$$

Поверхность сферы равна теперь $4\pi a^2\mathrm{sh}^2\chi$ при удалении от начала координат (увеличении χ) возрастает неограниченно.

Объем пространства отрицательной кривизны, очевидно, бесконечен.

получение конформино-евклидовых координат

Преобразовать элемент длины

$$dl^2 = dr^2\left(1 - \frac{r^2}{a^2}\right) + r^2\left(\sin^2\theta d\varphi^2 + d\theta^2\right)$$

к виду, в котором он был бы пропорционален своему евклидову выражению (конформно-евклидовы координаты).

Ре ш е н и е.

Подстановка

$$r = \frac{r_1}{1 + \frac{r_1^2}{4a^2}}$$

приводит к результату:

$$dl^2 = \left(1 + \frac{r_1^2}{4a^2}\right)^{-2}\left(dr_1^2 + r_1^2 d\theta^2 + r_1^2\sin^2\theta d\varphi^2\right)$$

и кстати, это важно, это отвечает на вопрос, можно ли было бы заменить как-то координаты, чтобы убрать особенности.

какой-то такой там важный вопрос, Володя говорил.

6.3.4 Closed Isotropic Model by Landau

вкратце

Theory

Переходя к исследованию пространственно-временной метрики изотропной модели, мы должны прежде всего условиться о выборе системы отсчета.

Наиболее удобна «сопутствующая» система отсчета, Движущаяся в каждой точке пространства вместе с находящимся В ней веществом.

Другими словами, системой отсчета является сама заполняющая пространство материя; скорость вещества в этой системе По определению равна везде нулю.

Очевидно, что такой выбор системы отсчета для изотропной модели естествен: при другом выборе направленность скоростей материи создавала бы кажущуюся неэквивалентность различных направлений в пространстве.

Временная координата должна быть выбрана указанным в начале предыдущего параграфа образом, т.е. так, чтобы в каждый данный момент времени метрика во всем пространстве была одинаковой.

Ввиду полной эквивалентности всех направлений, компоненты $g_{0\alpha}$ метрического тензора в выбранной нами системе отсчета равны нулю.

Действительно, три компоненты $g_{0\alpha}$ можно рассматривать как Компоненты трехмерного вектора, который, будучи отличен от нуля, создавал бы неравноценность различных направлений.

Таким образом, ds^2 должно иметь вид $ds^2 = g_{00}\left(dx^0\right)^2 - dl^2$. Компонента g_{00} является здесь функцией только от x^0. Поэтому можно всегда выбрать временную координату так, чтобы g_{00} обратилось в 1.

Обозначая ее через ct, имеем

$$ds^2 = c^2dt^2 - dl^2$$

Переменная t является синхронным собственным временем в каждой точке пространства.

Начнем с рассмотрения пространства положительной кривизны ниже мы будем для краткости говорить о соответствующем решении уравнений Эйнштейна как о закрытой модели.

Для dl воспользуемся выражением (111.8), в котором радиус кривизны a является, вообще говоря, функцией времени.

Таким образом, ds^2 пишем в виде

$$ds^2 = c^2dt^2 - a^2(t)\left\{d\chi^2 + \sin^2\chi\left(d\theta^2 + \sin^2\theta d\varphi^2\right)\right\}$$

Функция $a(t)$ определяется уравнениями Эйнштейна.

Для решения этих уравнений удобно воспользоваться вместо времени величиной η, определяемой соотношением

$$cdt = ad\eta$$

Тогда ds^2 напишется в виде

$$ds^2 = a^2(\eta)\left\{d\eta^2 - d\chi^2 - \sin^2\chi\left\{d\theta^2 + \sin^2\theta d\varphi^2\right)\right\}$$

Для составления уравнений поля надо начать с вычисления компонент тензора R_{ik} (координатами x^0, x^1, x^2, x^3 являются $\eta, \chi, \theta, \varphi$). С помощью значений компонент метрического тензора

$$g_{00} = a^2, \quad g_{11} = -a^2, \quad g_{22} = -a^2\sin^2\chi, \quad g_{33} = -a^2\sin^2\chi\sin^2\theta$$

вычисляем величины Γ^i_{kl} :

$$\Gamma^0_{00} = \frac{a'}{a}, \quad \Gamma^0_{\alpha\beta} = -\frac{a'}{a^3}g_{\alpha\beta}, \quad \Gamma^\alpha_{0\beta} = \frac{a'}{a}\delta^\alpha_\beta, \quad \Gamma^0_{\alpha 0} = \Gamma^\alpha_{00} = 0$$

гДе Штрих означает дифференцирование по η (компоненты $\Gamma^\alpha_{\beta\gamma}$ нет надобности вычислять в явном виде).

С помощью этих значений по общей формуле (92.7) получим:

$$R^0_0 = \frac{3}{a^4}\left(a'^2 - aa''\right)$$

Из тех же соображений симметрии, которые были применены выше К $g_{0\alpha}$, заранее очевидно, что компоненты $R_{0\alpha} = 0$. Для вычисления же компонент R^β_α замечаем, что если выделить в них члены, содержащие только $g_{\alpha\beta}$ (т.

е.

только $\Gamma^\alpha_{\beta\gamma}$), то эти члены должны составить компоненты трехмерного тензора $-P^\beta_\alpha$, значения которых заранее известны из (111.3) и (111.6)

$$R^\beta_\alpha = -P^\beta_\alpha + \ldots = -\frac{2}{a^2}\delta^\beta_\alpha + \ldots$$

гДе многоточие подразумевает члены, содержащие наряду с $g_{\alpha\beta}$ также и g_{00}. В результате вычисления последних получим:

$$R^\beta_\alpha = -\frac{1}{a^4}\left(2a^2 + a^2 + aa''\right)\delta^\beta_\alpha$$

Н 3аТеМ

$$R = R^0_0 + R^\alpha_\alpha = -\frac{6}{a^3}\left(a + a''\right)$$

Поскольку в выбранной нами системе отсчета материя неподвижна, то $u^\alpha = 0, u^0 = 1/a$ и из (94.9) имеем $T^0_0 = \varepsilon$, где ε - плотность энергиИ материи.

Подставляя полученные выражения в уравнение

$$R^0_0 - \frac{1}{2}R = \frac{8\pi k}{c^4}T^0_0$$

получИм:

$$\frac{8\pi k}{c^4}\varepsilon = \frac{3}{a^4}\left(a^2 + a^{22}\right)$$

Сюда входят две неизвестные функции ε и a; поэтому необходимо получить еще одно уравнение.

В качестве него удобно выбрать (вместо пространственных компонент уравнений Эйнштейна) уравнение $T^i_{0:i} = 0$ - одно из четырех уравнений (94.7), содержашихся, как мы знаем, в уравнениях поля.

Это уравнение можно вывести и непосредственно с помощью термодинамических соотношений следующим образом.

Пользуясь в уравнениях поля выражением (94.9) ДлЯ тензора энергиИ-импульса, мы тем самым пренебрегаем всеми процессами диссипации энергии,

приводящими к возрастанию энтропии.

Такое пренебрежение, разумеется, здесь вполне законно, поскольку дополнительные члены, которые надо было бы прибавить к T^i_k в связи с диссипаЦией энергии,

ничтожно малы по сравнению с плотностью энергии ε,

включающей в себя энергию покоя материальных тел.

Таким образом, при выводе уравнений поля мы можем считать полную энтропию постоянной.

Воспользуемся теперь известным термодинамическим соотношением $d\mathscr{E} = TdS - pdV$, где $\mathscr{E}, S, V$ - энергия, энтропия и объем системы, а p, T - давление и температура.

При постоянной энтропии имеем просто $d\mathscr{E} = -pdV$. Вводя плотность энергии $\varepsilon = \mathscr{E}/V$, без труда находим:

$$d\varepsilon = -(\varepsilon + p)\frac{dV}{V}$$

Объем пространства V пропорционален, согласно (111.9), кубу радиуса кривизны a. Поэтому $dV/V = 3da/a = 3d\ln a$ и мы можем написать:

$$-\frac{d\varepsilon}{\varepsilon + p} = 3d\ln a$$

или, интегрируя,

$$3\ln a = -\int \frac{d\varepsilon}{p + \varepsilon} + \text{const}$$

(нижний предел в интеграле постоянен).

Если связь между ε и p (уравнение состояния материи) известна, то уравнение (112.6) определяет ε как функцию от a. Тогда из (112.5) мы можем определить η в виде

$$\eta = \pm\int \frac{da}{a\sqrt{\frac{8\pi k}{3c^4}\varepsilon a^2 - 1}}$$

Уравнения (112.6) – (112.7) решают в общем виде задачу об определении метрики в изотропной закрытой модели.

Если материя распределена в пространстве в виде отдельных макроскопических тел, то при определении создаваемого ею гравитационного поля мы можем рассматривать эти тела как материальные частицы,

обладающие определенными массами, не интересуясь вовсе их внутренним строением.

Считая скорости тел сравнительно малыми (малыми по сравнению с c), можно положить просто $\varepsilon = \mu c^2$, где μ - сумма масс тел, отнесенная к единице объема.

По той же причине давление «газа», состоящего из этих тел, крайне мало по сравнению с ε и им можно пренебречь (давления же внутри тел,

согласно сказанному, не имеют отношения к рассматриваемому вопросу).

Что касается имеющегося в пространстве излучения, то его количество относительно мало и его энергией и давлением тоже можно пренебречь.

Таким образом, для описания в терминах рассматриваемой модели современного состояния Вселенной следует пользоваться уравнением состояния «пылевидной» материи

$$\varepsilon = \mu c^2, \quad p = 0$$

Интегрирование в (112.6) дает тогда $\mu a^3 = \text{const}$.

Это равенство можно было бы написать и сразу,

так как оно выражает собой просто постоянство суммы M масс тел во всем пространстве, как и должно было быть в рассматриваемом случае пылевидной материи.

Поскольку объем пространства в замкнутой модели равен $V = 2\pi^2 a^3$, То $\text{const} = M/2\pi^2$. Таким образом,

$$\mu a^3 = \text{const} \ = \frac{M}{2\pi^2}$$

Подставив (112.8) в уравнение (112.7) и произведя интегрирование, получим:

$$a = a_0(1 - \cos\eta)$$

где постоянная

$$a_0 = \frac{2kM}{3\pi c^2}$$

Наконец, для связи между t и η находим из (112.3):

$$t = \frac{a_0}{c}(\eta - \sin\eta)$$

Уравнения (112.9)−(112.10) определяют в параметрическом виде зависимость $a(t)$. Функция $a(t)$ возрастает от нуля при $t = 0(\eta = 0)$ до максимального значения $a = 2a_0$, достигаемого при $t = \pi a_0/c(\eta = \pi)$, и затем снова убывает до нуля при $t = 2\pi a_0/c(\eta = 2\pi)$. При $\eta \ll 1$ имеем приближенно $a = a_0\eta^2/2, t = a_0\eta^3/6c$, так что

$$a \approx \left(\frac{9a_0c^2}{2}\right)^{1/3} t^{2/3}$$

При этом плотность вещества

$$\mu = \frac{1}{6\pi k t^2} = \frac{8 \cdot 10^5}{t^2}$$

(численное значение коэффициента дано для плотности в г $\cdot$ см $^{-3}$ при t в секундах).

Обратим внимание на то, что в этом пределе зависимость $\mu(t)$ имеет универсальный характер в том смысле, что не зависит от параметра a_0.

При $a \to 0$ плотность μ обращается в бесконечность.

Но при $\mu \to \infty$ давление тоже становится большим, и потому для исследования метрики в этой области надо рассмотреть противоположный случай наибольшего возможного (при данной плотности энергии ε) давления, т. е. описывать материю уравнением состояния

$$p = \frac{\varepsilon}{3}$$

(см. примечание на с. 119).

Из формулы (112.6) получим тогда

$$\varepsilon a^4 = \text{const} \ \equiv \frac{3c^4 a_1^2}{8\pi k}$$

(a_1- новая постоянная), после чего уравнения (112.3) и приводят к зависимости

$$a = a_1 \sin\eta, \quad t = \frac{a_1}{c}(1 - \cos\eta)$$

Поскольку это решение имеет смысл рассматривать только при очень больших значениях ε (т. е. малых a), то положим $\eta \ll 1$. Тогда $a \approx a_1\eta$, $t \approx a_1\eta^2/2c$, так что

$$a = \sqrt{2a_1 ct}$$

При этом

$$\frac{\varepsilon}{c^2} = \frac{3}{32\pi k t^2} = \frac{4.5 \cdot 10^5}{t^2}$$

(эта зависимость снова не содержит никаких параметров).

Таким образом, и здесь $a \to 0$ при $t \to 0$, так что значение $t = 0$ действительно является особой точкой пространственно-временной метрики изотропной модели (и то же самое относится к закрытой модели и ко второй точке, в которой $a = 0$).

Мы видим также из (112.14), что при изменении знака t величина $a(t)$ сделалась бы мнимой, а ее квадрат - отрицательным.

Все четыре компоненты g_{ik} в (112.2) стали бы при этом положительными, так же как и определитель g. Но такая метрика физически бессмысленна.

Это значит, что не имеет физического смысла аналитически продолжать метрику за особую точку.

6.3.5 Open Isotropic Model by Landau

вкратце

В открытой изотропной модели метрика имеет вид:

$$ds^2 = c^2dt^2 - a^2(t)\left\{d\chi^2 + \mathrm{sh}^2\,\chi\left(d\theta^2 + \sin^2\theta d\varphi^2\right)\right\}$$

актуальность и применения открытой изотропной модели

хз, насколько нужна она?

Theory

Решение, соответствующее изотропному пространству отрицательной кривизны (открытая модель), получается вполне аналогично предыдущему.

Рассмотрим метрику вида

$$ds^2 = c^2dt^2 - a^2(t)\left\{d\chi^2 + \mathrm{sh}^2\,\chi\left(d\theta^2 + \sin^2\theta d\varphi^2\right)\right\}$$

Вводим снова вместо t переменную η согласно $cdt = ad\eta$; тогда получаем:

$$ds^2 = a^2(\eta)\left\{d\eta^2 - d\chi^2 - \mathrm{sh}^2\,\chi\left(d\theta^2 + \sin^2\theta d\varphi^2\right)\right\}$$

Это выражение может быть формально получено из (112.4) заменой η, χ, a соответственно на $i\eta, i\chi$, ia.

Поэтому и уравнения поля можно получить просто путем этой же замены из (112.5) – (112.6). Уравнение (112.6) сохраняет при этом свой прежний вид:

$$3\ln a = | - \int \frac{d\varepsilon}{\varepsilon + p} + \mathrm{const}$$

а вместо (112.5) имеем:

$$\frac{8\pi k}{c^4}\varepsilon = \frac{3}{a^4}\left(a'^2 - a^2\right)$$

Соответственно этому находим вместо (112.7)

$$\eta = \pm \int \frac{da}{a\sqrt{\frac{8\pi k}{3c^4}\varepsilon a^2 + 1}}$$

Для пылевидной материи получаем отсюда:

$$a = a_0(\operatorname{ch}\eta - 1), \quad t = \frac{a_0}{c}(\operatorname{sh}\eta - \eta)$$

$$\mu a^3 = \frac{3c^2}{4\pi k}a_0$$

Формулы (113.6) определяют в параметрическом виде зависимость $a(t)$. В отличие от замкнутой модели, здесь радиус кривизны меняется монотонно, возрастая от нуля при $t = 0(\eta = 0)$ до бесконечности при $t \to \infty$ ($\eta \to \infty$). Плотность же материи, соответственно, монотонно убывает от бесконечного значения при $t = 0$ (при $\eta \ll 1$ закон этого убывания дается той же приближенной формулой (112.12), что и в закрытой модели).

Для больших плотностей решение (113.6)−(113.7) неприменимо, и надо снова обратиться к случаю $p = \varepsilon/3$. При этом снова получается

$$\varepsilon a^4 = \text{const} \equiv \frac{3c^4 a_1^2}{8\pi k}$$

а для зависимости $a(t)$ находим:

$$a = a_1 \operatorname{sh}\eta, \quad t = \frac{a_1}{c}(\operatorname{ch}\eta - 1)$$

или при $\eta \ll 1$:

$$a = \sqrt{2a_1 ct}$$

(и прежняя формула (112.15) для $\varepsilon(t)$). Таким образом, и в открытой модели метрика имеет особую точку (но в отличие от закрытой модели - лишь одну).

Наконец, предельным случаем рассмотренных решений, соответствующим бесконечному радиусу кривизны пространства, является

модель с плоским (евклидовым) пространством.

Интервал ds^2 в этой модели можно написать в виде

$$ds^2 = c^2dt^2 - b^2(t)\left(dx^2 + dy^2 + dz^2\right)$$

(в качестве пространственных координат выбраны «декартовы» координаты x, y, z). Зависящий от времени множитель в элементе пространственного расстояния не меняет, очевидно, евклидовости пространственной метрики, так как при заданном t этот множитель постоянен и простым преобразованием координат может быть приведен к единице.

Вычисления, аналогичные произведенным в предыдущем параграфе, приводят к следующим уравнениям:

$$\frac{8\pi k}{c^2}\varepsilon = \frac{3}{b^2}\left(\frac{db}{dt}\right)^2, \quad 3\ln b = -\int \frac{d\varepsilon}{p+\varepsilon} + \text{const}.$$

Для случая малых давлений находим:

$$\mu b^3 = \text{const}, \quad b = \text{const}\, t^{2/3}$$

При малых t опять надо рассматривать случай $p = \varepsilon/3$, при котором получаем:

$$\varepsilon b^4 = \text{const}, \quad b = \text{const}\,\sqrt{t}.$$

Таким образом, и в этом случае метрика имеет особую точку ($t = 0$).

Отметим, что все найденные изотропные решения существуют лишь при отличной от нуля плотности материи; для пустого пространства уравнения Эйнштейна не имеют такого рода решений 1).

Упомянем также, что в математическом отношении они являются частным случаем более общего класса решений, содержащего три физически различные произвольные функции пространственных координат (см. задачу).

Вид квазиоднородным образом расширяющейся метрики вблизи особой точки

Найти общий вид вблизи особой точки для метрики, в которой расширение пространства происходит «квазиоднородным» образом, т.е. так, что все компоненты $\gamma_{\alpha\beta} = -g_{\alpha\beta}$ (в синхронной системе отсчета) стремятся к нулю по одинаковому закону.

Пространство заполнено материей с уравнением состояния $p = \varepsilon/3$

Ищем решение вблизи особой точки ($t = 0$) в виде

$$\gamma_{\alpha\beta} = t a_{\alpha\beta} + t^2 b_{\alpha\beta} + \dots$$

где $a_{\alpha\beta}, b_{\alpha\beta}$ - функции координат (пространственных); ниже полагаем $c = 1$. Обратный тензор

$$\gamma_{\alpha\beta} = \frac{1}{t} a^{\alpha\beta} - b^{\alpha\beta}$$

где тензор $a^{\alpha\beta}$ обратен $a_{\alpha\beta}$, а $b^{\alpha\beta} = a^{\alpha\gamma} a^{\beta\delta} b_{\gamma\delta}$; ниже все операции поднимания индексов и ковариантного дифференцирования производятся при помощи не зависящей от времени метрики $a_{\alpha\beta}$.

Вычисляя левые стороны уравнений (97.11) и (97.12) с необходимой точностью до $1/t$, Получим:

$$-\frac{3}{4t^2} + \frac{1}{2t} b = \frac{8\pi k}{3} \varepsilon \left(-4u_0^2 + 1\right), \quad \frac{1}{2}\left(b; \alpha - b^{\beta}_{\alpha;\beta}\right) = -\frac{32\pi k}{3} \varepsilon u_\alpha u_0$$

(где $b = b^\alpha_\alpha$).

Учитывая также тождество

$$1 = u_i u^i \approx u_0^2 - \frac{1}{t} u_\alpha u_\beta a^{\alpha\beta}$$

найдем:

$$8\pi k \varepsilon = \frac{3}{4t^2} - \frac{b}{2t}, \quad u_\alpha = -\frac{t^2}{2}\left(b_{;\alpha} - b^{\beta}_{\alpha;\beta}\right)$$

Трехмерные символы Кристоффеля, а с ними и тензор $P_{\alpha\beta}$ в первом по $1/t$ приближении не зависят от времени; при этом $P_{\alpha\beta}$ совпадают с выражениями, получающимися при вычислении с метрикой просто $a_{\alpha\beta}$. Учитывая это, найдем, что в уравнении (97.13) члены порядка t^{-2} взаимно сокращаются, а члены $\sim 1/t$ дают

$$P^\beta_\alpha + \frac{3}{4} b^\beta_\alpha + \frac{5}{12} \delta^\beta_\alpha b = 0$$

откуда

$$b^\beta_\alpha = -\frac{4}{3} P^\beta_\alpha + \frac{5}{18} \delta^\beta_\alpha P$$

(где $P = a^{\beta\gamma}P_{\beta\gamma}$).

Ввиду тождества

$$P^{\beta}_{\alpha;\beta} - \frac{1}{2}P_{;\alpha} = 0$$

(см.

(92.10)) имеет место соотношение

$$b^{\beta}_{\alpha;\beta} = \frac{7}{9}b_{;\alpha}$$

и потому u_α можно переписать в виде

$$u_\alpha = -\frac{t^2}{9}b_{;\alpha}$$

Таким образом, все шесть функций $a_{\alpha\beta}$ остаются произвольными, а по ним определяются коэффициенты $b_{\alpha\beta}$ следующего члена разложения (1).

Выбор времени в метрике (1) полностью определен условием $t = 0$ в особой точке; пространственные же координаты допускают еще произвольные преобразования,

не затрагивающие времена (ими можно воспользоваться, например, для приведения тензора $a_{\alpha\beta}$ к диагональному виду).

Поэтому полученное решение содержит всего три «физически различные» произвольные функции.

Отметим, что в этом решении пространственная метрика неоднородна и анизотропна, а распределение плотности материи стремится при $t \to 0$ к однородному.

Трехмерная скорость v имеет (в приближении(4)) равный нулю ротор, а ее величина стремится к нулю по закону

$$v^2 = v_\alpha v_\beta \gamma^{\alpha\beta} \sim t^3$$

6.4 Not Popular, but Basic Models

6.4.1 Gravitational Stability of Isotropic World

(???!!! it is a good idea to discuss gravitational stability!!!)

вкратце

??? а в чем вопрос, собственно??

Theory

Рассмотрим вопрос о поведении малых возмущений в изотропной модели, то есть о ее гравитационной устойчивости.

При этом мы ограничимся рассмотрением возмущений в сравнительно небольших областях пространства - областях, линейные размеры которых малы по сравнению с радиусом a.

В каждой такой области пространственная метрика может быть принята в первом приближении евклидовой, т.е. метрика $dl^2 = a^2\left[d\chi^2 + \sin^2\chi\left(\sin^2\theta d\varphi^2 + d\theta^2\right)\right]$ или $dl^2 = a^2\left\{d\chi^2 + \mathrm{sh}^2\,\chi\left(\sin^2\theta d\varphi^2 + d\theta^2\right)\right\}$ заменится метрикой

$$dl^2 = a^2(\eta)\left(dx^2 + dy^2 + dz^2\right)$$

где x, y, z - декартовы координаты, измеренные в единицах радиуса a.

В качестве временной координаты будем по-прежнему пользоваться переменной η.

возмущение метрики - описание его (???)

Без ограничения общности будем описывать возмущенное поле по-прежнему в синхронной системе отсчета, т.е. наложим на изменения δg_{ik} метрического тензора условия $\delta g_{00} = \delta g_{0\alpha} = 0$.

(???? мб она все-таки нужна?)

Варьируя при этих условиях тождество $g_{ik}u^i u^k = 1$ и имея в виду, что невозмущенные значения компонент 4-скорости материи $u^0 = 1/a, u^\alpha = 0$, получим $g_{00}u^0\delta u^0 = 0$, откуда $\delta u^0 = 0$.

Возмущения же δu^α, вообще говоря, отличны от нуля, так что система отсчета - уже не сопутствующая.

Возмущения пространственного метрического тензора обозначим посредством $h_{\alpha\beta} = \delta\gamma_{\alpha\beta} = -\delta g_{\alpha\beta}$.

Тогда $\delta\gamma^{\alpha\beta} = -h^{\alpha\beta}$, причем поднимание индексов у $h_{\alpha\beta}$ осуществляется с помощью невозмущенной метрики $\gamma_{\alpha\beta}$.

В линейном приближении малые возмущения гравитационного поля удовлетворяют уравнениям

$$\delta R_i^k - \frac{1}{2}\delta_i^k \delta R = \frac{8\pi k}{c^4}\delta T_i^k$$

В сихронной системе отсчета вариации компонент тензора энергии-импульса (94.9) равны:

$$\delta T_\alpha^\beta = -\delta_\alpha^\beta \delta p, \quad \delta T_0^\alpha = a(p+\varepsilon)\delta u^\alpha, \quad \delta T_0^0 = \delta\varepsilon$$

Ввиду малости $\delta\varepsilon$ и δp можно написать $\delta p = \frac{dp}{d\varepsilon}\delta\varepsilon$, и мы получаем соотношения

$$\delta T_\alpha^\beta = -\delta_\alpha^\beta \frac{dp}{d\varepsilon}\delta T_0^0$$

Формулы для δR_i^k можно получить варьированием выражений (97.10).

Поскольку невозмущенный метрический тензор $\gamma_{\alpha\beta} = a^2\delta_{\alpha\beta}$, то невозмущенные значения

$$\varkappa_{\alpha\beta} = \frac{2\dot{a}}{a}\gamma_{\alpha\beta} = \frac{2a'}{a^2}\gamma_{\alpha\beta}, \quad \varkappa_\alpha^\beta = \frac{2a'}{a}\delta_\alpha^\beta$$

где точка означает дифференцирование по ct, а штрих - по η. Возмущения же величин $\varkappa_{\alpha\beta}$ и $\S_\alpha^\beta = \varkappa_{\alpha\gamma\gamma}\gamma^{\gamma\beta}$

$$\delta\varkappa_{\alpha\beta} = \dot{h}_{\alpha\beta} = \frac{1}{a}h'_{\alpha\beta}, \quad \delta\varkappa_\alpha^\beta = -h^{\beta\gamma}{}_{\alpha\gamma} + \gamma^{\beta\gamma}\dot{h}_{\alpha\gamma} = \dot{h}_\alpha^\beta = \frac{1}{a}h_\alpha^{\beta\prime}$$

где $h_\alpha^\beta = \gamma^{\beta\gamma}h_{\alpha\gamma}$. Невозмущенные значения трехмерного тензора P_α^β для евклидовой метрики (115.1) равны нулю.

Вариации же δP_α^β вычисляются по формулам (108.3) – (108.4); очевидно, что δP_α^β выражается через $\delta\gamma_{\alpha\beta}$ так же, как 4-тензор δR_{ik} выражается через δg_{ik}, причем все тензорные операции производятся в трехмерном пространстве с метрикой (115.1); ввиду евклидовости этой метрики все ковариантные дифференцирования сводятся к простым дифференцированиям по координатам x^α (контравариантные же дифференцирования - еще и к делению на a^2).

Имея все это в виду (и переходя везде от производных по t к производным по η), получим после простого вычисления:

$$\delta R_\alpha^\beta = -\frac{1}{2a^2}\left(h_{\alpha,\gamma}^{\gamma,\beta} + h_{\gamma,\alpha}^{\beta,\gamma} - h_{\gamma\alpha}^{,\beta,\gamma} - h_{,\alpha}^{,\beta}\right) - \frac{1}{2a^2}h_\alpha^{\beta\prime\prime} - \frac{a'}{a^3}h_\alpha^{\beta\prime} - \frac{a'}{2a^3}h'\delta_\alpha^\beta$$

$$\delta R_0^0 = -\frac{1}{2a^2}h'' - \frac{a'}{2a^3}h',$$

$$\delta R_0^\alpha = \frac{1}{2a^2}\left(h^{,\alpha} - h_\beta^{\alpha,\beta}\right)'$$

($h \equiv h^\alpha_\alpha$). Здесь как нижние, так и верхние индексы после запятой означают простые дифференцирования по координатам x^α (мы продолжаем Писать индексы вверху и внизу лишь для сохранения единообразия обозначений).

Окончательные уравнения для возмущения h^β_α мы получим, подставив в (115.4) компоненты δT^k_i, выраженные через δR^k_i согласно (115.2). В качестве этих уравнений удобно выбрать уравнения, получающиеся из (115.4) при $\alpha \neq \beta$ при упрощении по индексам α, β. Они гласят:

$$\left(h^{\gamma,\beta}_{\alpha,\gamma} + h^{\beta,\gamma}_{\gamma,\alpha} - h^{\beta}_{,\alpha} - h^{\beta,\gamma}_{\alpha,\gamma}\right) + h^{\beta''}_\alpha + 2\frac{a'}{a}h^{\beta'}_\alpha = 0, \quad \alpha \neq \beta$$

$$\frac{1}{2}\left(h^{\delta,\gamma}_{\gamma,\delta} - h^{\gamma}_{,\gamma}\right)\left(1 + 3\frac{dp}{d\varepsilon}\right) + h'' + h'\frac{a'}{a}\left(2 + 3\frac{dp}{d\varepsilon}\right) = 0$$

Возмущения плотности и скорости материи могут быть определены по известным h^β_α с помощью формул (115.2)-(115.3).

Так, для относительного изменения плотности имеем:

$$\frac{\delta\varepsilon}{\varepsilon} = \frac{c^4}{8\pi k\varepsilon}\left(\delta R^0_0 - \frac{1}{2}\delta R\right) = \frac{c^4}{16\pi k\varepsilon a^2}\left(h^{\beta,\alpha}_{\alpha,\beta} - h^{,\alpha}_{,\alpha} + \frac{2a'}{a}h'\right)$$

Среди решений уравнений (115.6) есть такие, которые могут быть исключены простым преобразованием системы отсчета (не нарушающим ее синхронности) и поэтому не представляют собой реального физического изменения метрики.

Вид таких решений может быть заранее установлен с помощью полученных в задаче 3 $ 97 формул (1) и (2).

Подставив в них невозмущенные значения $\gamma_{\alpha\beta} = a^2\delta_{\alpha\beta}$, получим следующие выражения для фиктивных возмущений метрики:

$$h^\beta_\alpha = f^{,\beta}_{0,\alpha}\int\frac{d\eta}{a} + \frac{a'}{a^2}f_0\delta^\beta_\alpha + \left(f^{,\beta}_\alpha + f^\beta_{,\alpha}\right)$$

где f_0, f_α - произвольные (малые) функции координат x, y, z. Поскольку метрика в рассматриваемых нами небольших областях пространства предполагается евклидовой, то произвольное возмущение в каждой такой области может быть разложено по плоским волнам.

Понимая под x, y, z декартовы координаты, измеренные в единицах a, мы можем написать пространственный периодический множитель плоских волн в виде $e^{i\mathbf{nr}}$, где $\mathbf{n}$ - безразмерный вектор, представляющий собой волновой вектор, измеренный в единицах 1/a (волновой вектор $\mathrm{k} = \mathbf{n}/a$).

Если мы имеем возмущение в участке пространства с размерами $\sim l$, то в его разложение войдут в основном волны с длинами $\lambda = 2\pi a/n \sim l$. Ограничиваясь возмущениями в областях с размерами $l \ll a$, мы тем самым предполагаем число n достаточно большим ($n \gg 2\pi$).

Классификация гравитационных возмущений.

Гравитационные возмущения можно разделить на три типа по возможных типах плоских волн, в виде которых может быть представлен симметричный тензор $h_{\alpha\beta}$.

1.

С помощью скалярной функций

$$Q = e^{i\mathbf{n}r}$$

можно составить вектор $\mathrm{P} = \mathrm{n}Q$ и тензоры

$$Q^\beta_\alpha = \frac{1}{3}\delta^\beta_\alpha Q, \quad P^\beta_\alpha = \left(\frac{1}{3}\delta^\beta_\alpha - \frac{n_\alpha n^\beta}{n^2}\right)Q$$

Таким плоским волнам отвечают возмущения, в которых наряду с гравитационным полем испытывают изменении также скорость и плотность материи, т. е. мы имеем дело с возмущениями, сопровождающимися возникновением сгущений или разрежений материи.

Возмущение h_α^β выражается при этом через тензоры Q_α^β и P_α^β, возмущение скорости - через вектор P, а возмущение плотности - через скаляр Q.

2.

С помощью поперечной векторной волны

$$\mathbf{S} = \mathbf{s}e^{i\mathbf{nr}}, \quad \mathbf{sn} = 0$$

можно составить тензор $\left(n^\beta S_\alpha + n_\alpha S^\beta\right)$; соответствующего же скаляра не существует, поскольку nS = 0. Этим волнам отвечают возмущения, В которых наряду с гравитационным полем испытывает изменение скорость, но не плотность материи; их можно назвать вращательными возмущениями.

3.

Поперечная тензорная волна

$$G_\alpha^\beta = g_\alpha^\beta e^{i\mathrm{nr}}, \quad g_\alpha^\beta n_\beta = 0$$

С ее помощью нельзя составить ни вектора, ни скаляра.

Этим волнам отвечают возмущения гравитационного поля, при которых материя остается неподвижной и однородно распределенной в пространстве.

Другими словами, это - гравитационные волны в изотропном мире.

Наиболее интересны возмущения первого типа.

Полагаем

$$h_\alpha^\beta = \lambda(\eta)P_\alpha^\beta + \mu(\eta)Q_\alpha^\beta, \quad h = \mu Q$$

Из (115.7) получим для относительного изменения плотности

$$\frac{\delta\varepsilon}{\varepsilon} = \frac{c^4}{24\pi k\varepsilon a^2}\left[n^2(\lambda+\mu) + \frac{3a'}{a}\mu'\right]Q$$

Уравнения, определяющие функции λ и μ, получаются подстановкой (115.13) в (115.6)

$$\lambda'' + 2\frac{a'}{a}\lambda' - \frac{n^2}{3}(\lambda+\mu) = 0$$
$$\mu'' + \mu'\frac{a'}{a}\left(2 + 3\frac{dp}{d\varepsilon}\right) + \frac{n^2}{3}(\lambda+\mu)\left(1 + 3\frac{dp}{d\varepsilon}\right) = 0$$

Эта уравнения имеют прежде всего следующие два частных интеграла, соответствующие тем фиктивным изменениям метрики, которые могут быть исключены преобразованием системы отсчета:

$$\lambda = -\mu = \mathrm{const}$$
$$\lambda = -n^2\int\frac{d\eta}{a}, \quad \mu = n^2\int\frac{d\eta}{a} - \frac{3a'}{a^2}$$

(первый из них получается из (115.8) выбором $f_0 = 0, f_\alpha = P_\alpha$, Второй - выбором $f_0 = Q, f_\alpha = 0$). На ранних стадиях расширения мира, когда материя описывается уравнением состояниЯ $p = \varepsilon/3$, имеем $a \approx a_1\eta, \eta \ll 1$ (как в открытой, так и в закрытой моделях).

Уравнения (115.15) принимают вид

$$\lambda'' + \frac{2}{\eta}\lambda' - \frac{n^2}{3}(\lambda+\mu) = 0, \quad \mu'' + \frac{3}{\eta}\mu' + \frac{2n^2}{3}(\lambda+\mu) = 0$$

Исследование этих уравнений удобно производить раздельно длЯ двух предельных случаев в зависимости от взаимного соотношения между двумя большими величинами n и $1/\eta$.

Предположим сначала, что число n не слишком велико (или η достаточно мало), так что $n\eta \ll 1$. С той точностью, с которой справедливы уравнения (115.18), находим из них в данном случае:

$$\lambda = \frac{3C_1}{\eta} + C_2\left(1 + \frac{n^2}{9}\eta^2\right), \quad \mu = -\frac{2n^2}{3}C_1\eta + C_2\left(1 - \frac{n^2}{6}\eta^2\right)$$

где C_1, C_2 - постоянные; отсюда исключены решения вида (115.16) К (115.17) (в данном случае это - решение, в котором $\lambda = -\mu = \text{const}$ И В котором $\lambda + \mu \sim 1/\eta^2$).

Вычислив также $\delta\varepsilon/\varepsilon$ согласно (115.14) И (112.15), получим следующие выражения для возмущений метрики и плотности:

$$h_\alpha^\beta = \frac{3C_1}{\eta}P_\alpha^\beta + C_2\left(Q_\alpha^\beta + P_\alpha^\beta\right)$$

$$\frac{\delta\varepsilon}{\varepsilon} = \frac{n^2}{9}\left(C_1\eta + C_2\eta^2\right)Q \qquad \text{при} \qquad p = \frac{\varepsilon}{3}, \eta \ll \frac{1}{n}$$

Постоянные C_1, C_2 должны удовлетворять определенным условиям, выражающим малость возмущения в момент η_α его возникновения: должно быть $h_\alpha^\beta \ll 1$(откуда $\lambda \ll 1, \mu \ll 1$) и $\delta\varepsilon/\varepsilon \ll 1$. В применении к (115.19) эти условия приводят к неравенствам $C_1 \ll \eta_0, C_2 \ll 1$.

В выражениях (115.19) имеются члены, возрастающие в расширяющемся мире как различные степени радиуса $a = a_1\eta$. Однако это возрастание не приводит к тому, чтобы возмущение могло стать большим: если применить формулы (115.19) по порядку величины при $\eta \sim 1/n$, то мы увидим, что (в силу полученных выше неравенств для C_1, C_2) возмущения остаются малыми даже на верхнем пределе действия этих формул.

Пусть теперь число n настолько велико, что $n\eta \gg 1$. Решая уравнения (115.18) при этом условии, найдем, что главные члены в λ и μ равны:

$$\lambda = -\frac{\mu}{2} = \text{const}\,\frac{1}{\eta^2}e^{in\eta/\sqrt{3}}$$

Отсюда находим для возмущений метрики и плотности:

$$h_\alpha^\beta = \frac{C}{n^2\eta^2}\left(P_\alpha^\beta - 2Q_\alpha^\beta\right)e^{in\eta/\sqrt{3}}, \quad \frac{\delta\varepsilon}{\varepsilon} = -\frac{C}{9}Qe^{in\eta/\sqrt{3}}$$

$$\text{при } p = \frac{\varepsilon}{3}, \quad \frac{1}{n} \ll \eta \ll 1$$

где C - комплексная постоянная, удовлетворяющая условию $|C| \ll 1$. Наличие периодического множителя в этих выражениях вполне естественно.

При больших n мы имеем дело с возмущением, пространственная периодичность которого определяется большим волновым вектором $k = n/$ а.

Такие возмущения должны распространяться как звуковые волны со скоростью

$$u = \sqrt{\frac{dp}{d\left(\varepsilon/c^2\right)}} = \frac{c}{\sqrt{3}}$$

Соответственно временная часть фазы определяется, как полагается В геометрической акустике, большим интегралом $\int kudt = n\eta/\sqrt{3}$.

Амплитуда относительного изменения плотности остается, как мы видим, постоянной, амплитуды же возмущений метрики при расширений мира убывают как a^{-2}.

Далее, рассмотрим более поздние стадии расширения, когда материя разрежена уже настолько, что можно пренебречь ее давлением ($p = 0$). При этом мы ограничимся здесь лишь случаем малых η соответствующих тем стадиям расширения, когда радиус a еще очень мал по сравнению с его современным значением, но все же материя уже достаточно разрежена.

При $p = 0$ и $\eta \ll 1$ имеем $a \approx a_0\eta^2/2$ и уравнения (115.15) принимают вид

$$\lambda'' + \frac{4}{\eta}\lambda' - \frac{n^2}{3}(\lambda + \mu) = 0, \quad \mu'' + \frac{4}{\eta}\mu' + \frac{n^2}{3}(\lambda + \mu) = 0$$

Решение этих уравнений:

$$\lambda + \mu = 2C_1, \quad \lambda - \mu = 2n^2\left(\frac{C_1\eta^2}{15} + \frac{2C_2}{\eta^3}\right)$$

Вычислив также $\delta\varepsilon/\varepsilon$ (с помощью (115.14) и (112.12)), находим:

$$h_\alpha^\beta = C_1\left(P_\alpha^\beta + Q_\alpha^\beta\right) + \frac{2n^2C_2}{\eta^3}\left(P_\alpha^\beta - Q_\alpha^\beta\right) \text{ при } \eta \ll \frac{1}{n}$$

$$h_\alpha^\beta = \frac{C_1}{15}n^2\eta^2\left(P_\alpha^\beta - Q_\alpha^\beta\right) + \frac{2n^2C_2}{\eta^3}\left(P_\alpha^\beta - Q_\alpha^\beta\right) \text{ при } \frac{1}{n} \ll \eta \ll 1$$

$$\frac{\delta\varepsilon}{\varepsilon} = \left(\frac{C_1n^2\eta^2}{30} + \frac{C_2n^2}{\eta^3}\right)Q$$

Мы видим, что $\delta\varepsilon/\varepsilon$ содержит член, возрастающий пропорционально a.

Однако если $n\eta \ll 1$, то $\delta\varepsilon/\varepsilon$ не становится все же большим даже при $\eta \sim 1/n$ в силу условия $C_1 \ll 1$. Если же $\eta n \gg 1$, то при $\eta \sim 1$ относительное изменение плотности становится порядка C_1n^2, между тем как малость начального возмущения требует лишь, чтобы было $C_1n^2\eta_0^2 \ll 1$. Таким образом, хотя возрастание возмущений происходит и медленно, но общее увеличение может быть значительным и в результате возмущение может стать сравнительно большим.

Аналогичным образом могут быть рассмотрены возмущения второго и третьего из перечисленных выше типов.

Однако законы затухания этих возмущений могут быть найдены и без детальных вычислений исходя из следующих простых соображений.

Если в небольшом участке вещества (с линейными размерами l) имеется вращательное возмущение со скоростью δv, то момент импульса этого участка $\sim (\varepsilon/c^2)\, l^3 \cdot l \cdot v$. При расширении мира l растет пропорционально a, а ε убывает как a^{-3} (в случае $p = 0$) или как a^{-4} (при $p = \varepsilon/3$).

В силу сохранения момента имеем поэтому

$$\delta v = \text{const при } p = \frac{\varepsilon}{3}, \quad \delta v \propto \frac{1}{a} \text{при } p = 0$$

Наконец, плотность энергии гравитационных волн должна убывать при расширении мира как a^{-4}. С другой стороны, эта плотность, выражается через возмущение метрики как $\sim k^2\left(h_\alpha^\beta\right)^2$, где $k = n/a$ - волновой вектор возмущения.

Отсюда следует, что амплитуда возмущения типа гравитационной волны убывает со временем как 1/a.

6.4.2 Homogeneous Spaces

вкратце

(хз вообще, что это, я и близко не на этом уровне)

Theory

Предположение об однородности и изотропиИ пространства определяет его метрику полностьЮ (оставляя свободным лишь знак кривизны).

Значительно больше свободы оставляет предположение об одной только однородности пространства, без какой-либо еще дополнительной симметрии.

Рассмотрим вопрос о том, каковы могут быть метрические свойства однородного пространства.

Речь будет идти о метрике пространства, рассматриваемой в заданный момент времени t.

При этом предполагается, что пространственновременная система отсчета выбрана синхронной, так что t есть единое для всего пространства синхронизованное время.

Однородность означает одинаковость метрических свойств во всех точках пространства.

Точное определение этого понятия связано с рассмотрением совокупности преобразований координат, которые совмещают пространство само с собой, т. е. оставляют его метрику неизменной: если до преобразования элемент длины

$$dl^2 = \gamma_{\alpha\beta}\left(x^1, x^2, x^3\right) dx^\alpha dx^\beta$$

то после преобразования тот же элемент

$$dl^2 = \gamma_{\alpha\beta}\left(x'^1, x'^2, x'^3\right) dx'^\alpha dx'^\beta$$

с той же функциональной зависимостью $\gamma_{\alpha\beta}$ от новых координат.

Пространство однородно, если оно допускает совокупность преобразований (или, как говорят, группу движений), позволяющих совместить любую заданную его точку с любой другой точкой.

В силу трехмерности пространства очевидно, что для этого различные преобразования группы должны определяться значениями трех независимых параметров.

Так, в евклидовом пространстве однородность выражается инвариантностью метрики по отношению к параллельным переносам (трансляциям) декартовой системы координат.

Каждая трансляция определяется тремя параметрами - компонентами вектора смещения начала координат.

Все эти преобразования оставляют инвариантными три независимых дифференциала (dx, dy, dz), из которых и строится элемент длины.

В общем случае неевклидова однородного пространства преобразования его группы движений тоже оставляют инвариантными три независимые линейные дифференциальные формы, не сводящиеся, однако, к полным дифференциалам каких-либо координатных функций.

Напишем эти формы в виде

$$e_\alpha^{(a)} dx^\alpha$$

где латинский индекс (a) нумерует три независимых реперных вектора (функции координат).

С помощью форм (116.1) инвариантная по отношению к данной группе движений пространственная метрика строится как

$$dl^2 = \eta_{ab}\left(e_\alpha^{(a)} dx^\alpha\right)\left(e_\beta^{(b)} dx^\beta\right)$$

где метрический тензор

$$\gamma_{\alpha\beta} = \eta_{ab} e_\alpha^{(a)} e_\beta^{(b)}$$

где симметричные по индексам a, b коэффициенты η_{ab} - функции времени.

Таким образом, мы приходим к «триадному»| представлению пространственной метрики с помощью тройки реперных векторов; к этому представлению применимы все полученные в & 98 формулы.

При этом выбор реперных векторов диктуется свойствами симметрии пространства и, вообще говоря, эти векторы не ортогональны (так что матрица η_{ab} не диагональна).

Как и в \$98, наряду с тройкой векторов $e_\alpha^{(a)}$ введем тройку взаимных с ними векторов $e_{(a)}^\alpha$, для которых

$$e_{(a)}^\alpha e_\alpha^{(b)} = \delta_a^b, \quad e_{(a)}^\alpha e_\beta^{(a)} = \delta_\beta^\alpha$$

В трехмерном случае связь между теми и другими векторами может быть представлена в явном виде как

$$\mathbf{e}_{(1)} = \frac{1}{v}\left[\mathbf{e}^{(2)}\mathbf{e}^{(3)}\right], \quad \mathbf{e}_{(2)} = \frac{1}{v}\left[\mathbf{e}^{(3)}\mathbf{e}^{(1)}\right], \quad \mathbf{e}_{(3)} = \frac{1}{v}\left[\mathbf{e}^{(1)}\mathbf{e}^{(2)}\right]$$

где

$$v = \left|e_\alpha^{(a)}\right| = \left(\mathbf{e}^{(1)}\left[\mathbf{e}^{(2)}\mathbf{e}^{(3)}\right]\right)$$

а $\mathrm{e}_{(a)}$ и $\mathrm{e}^{(a)}$ надо понимать как декартовы векторы с компонентами соответственно $e_{(a)}^\alpha$ и $e_\alpha^{(a)}$ Определитель метрического тензора (116.3)

$$\gamma = \eta v^2$$

где η - определитель матрицы η_{ab}.

Инвариантность дифференциальных форм (116.1) означает, что

$$e_\alpha^{(a)}(x)dx^\alpha = e_\alpha^{(a)}(x')\,dx'^\alpha$$

причем $e_\alpha^{(a)}$ в обеих сторонах равенства - одни и те же функции соответственно от старых и новых координат.

Умножив это равенство на $e_{(\alpha)}^\beta(x')$, заменив $dx'^\beta = \left(\partial x'^\beta/\partial x^\alpha\right)dx^\alpha$ и сравнив коэффициенты при одинаковых дифференциалах dx^α, получим

$$\frac{\partial x'^\beta}{\partial x^\alpha} = e_{(a)}^\beta(x')\,e_\alpha^{(a)}(x).$$

Эти равенства представляют собой систему дифференциальных уравнений, определяющих функции $x'^\beta(x)$ по заданным реперным векторам. Для того чтобы быть интегрируемыми, уравнения (116.8) должны тождественно удовлетворять условиям

$$\frac{\partial^2 x'^\beta}{\partial x^\alpha \partial x^\gamma} = \frac{\partial^2 x'^\beta}{\partial x^\gamma \partial x^\alpha}$$

Вычислив производные, получим:

$$\left[\frac{\partial e_{(a)}^\beta(x')}{\partial x'^\delta} - \frac{\partial e_{(b)}^\beta(x')}{\partial x'^\delta}e_{(a)}^\delta(x')\right]e_\gamma^{(b)}(x)e_\alpha^{(a)}(x) = e_{(a)}^\beta(x')\left[\frac{\partial e_\gamma^{(a)}(x)}{\partial x^\alpha} - \frac{\partial e_\alpha^{(a)}(x)}{\partial x^\gamma}\right]$$

Умножив обе стороны равенства на $e_{(d)}^\alpha(x)e_{(c)}^\gamma(x)e_\beta^{(f)}(x')$ и перенеся Дифференцирования с одних множителей на другие с учетом (116.4), Получим в левой стороне:

$$e_\beta^{(f)}(x')\left[\frac{\partial e_{(d)}^\beta(x')}{\partial x'^\delta}e_{(c)}^\delta(x') - \frac{\partial e_{(c)}^\beta(x')}{\partial x'^\delta}e_{(d)}^\delta(x')\right] = e_{(c)}^\beta(x')\,e_{(d)}^\delta(x')\left[\frac{\partial e_\beta^{(f)}(x')}{\partial x'^\delta} - \frac{\partial e_\delta^{(f)}(x')}{\partial x'^\beta}\right],$$

а в правой - такое же выражение как функцию от x. Поскольку x и x' произвольны, то эти выражения должны сводиться к постоянным:

$$\left(\frac{\partial e_\alpha^{(c)}}{\partial x^\beta} - \frac{\partial e_\beta^{(c)}}{\partial x^\alpha}\right)e_{(a)}^\alpha e_{(b)}^\beta = C_{ab}^c$$

Постоянные C^c_{ab} называются структурными константами группы Умножив на $e^\gamma_{(c)}$ можно переписать (116.9) в виде

$$e^\alpha_{(a)}\frac{\partial e^\gamma_{(b)}}{\partial x^\alpha} - e^\beta_{(b)}\frac{\partial e^\gamma_{(a)}}{\partial x^\beta} = C^c_{ab}e^\gamma_{(c)}$$

Это и есть искомые условия однородности пространства.

(!!!!!?????)

Выражение в левой стороне равенства (116.9) совпадает с определением величин λ^c_{ab}, которые, таким образом, оказываются постоянными.

По своему определению структурные константы антисимметричны по нижним индексам:

$$C^c_{ab} = -C^c_{ba}$$

Еще одно условие для них можно получить, заметив, что равенство (116.10) эквивалентно правилу коммутации

$$[X_a, X_b] \equiv X_aX_b - X_bX_a = C^c_{ab}X_c$$

для линейных дифференциальных операторов

$$X_a = e^\alpha_{(a)}\frac{\partial}{\partial x^\alpha}$$

Тогда упомянутое соотношение возникнет из тождества

$$[[X_a, X_b], X_c] + [[X_b, X_c], X_a] + [[X_c, X_a], X_b] = 0$$

(так называемое тождество Якоби) и имеет вид

$$C^f_{ab}C^d_{cf} + C^f_{bc}C^d_{af} + C^f_{ca}C^d_{bf} = 0$$

Определенное преимущество перед трехиндексовыми константами $C^c{}_{ab}$ представляют двухиндексовые величины, получающиеся путем дуального преобразования

$$C^c_{ab} = e_{abd}C^{dc}$$

где $e_{abc} = e^{abc}$ - единичный антисимметричный символ (причем $e_{123} = +1$). Правила коммутации (116.12) с помощью таких констант запишутся в виде

$$e^{abc}X_bX_c = C^{ad}X_d$$

Свойство (116.11) уже учтено в определении (116.15), а свойство (116.14) примет вид

$$e_{bcd}C^{cd}C^{ba} = 0$$

Укажем также, что определение (116.9) для величин Cab можно представить в векторном виде:

$$C^{ab} = -\frac{1}{v}\mathbf{e}^{(a)}\,\mathrm{rot}\,\mathbf{e}^{(b)}$$

где снова векторные операции производятся так, как если бы координаты x^α были декартовыми.

Выбор трех реперных векторов в дифференциальных формах (116.1) (а с ними и операторов X_a), разумеется, не однозначен.

Они могут быть подвергнуты любому линейному преобразованию с постоянными коэффициентами

$$\mathbf{e}_{(a)} = A^b_a\mathbf{e}_{(b)}$$

По отношению к таким преобразованиям величины η_{ab} и C^{ab} ведут себя как тензоры.

Условия (116.17) - единственные, которым должны удовлетворять структурные константы C^{ab}. Но среди допускаемых этими условиями наборов констант есть эквивалентные - в том смысле, что различие связано лишь с преобразованиями (116.19).

Вопрос о классификации однородных пространств сводится к определению всех неэквивалентных наборов структурных констант.

Это можно сделать, воспользовавшись «тензорными» свойствами величин C^{ab}, следующим простым способом (C. G. Behr, 1962).

Несимметричный «тензор» C^{ab} можно разложить на симметричную и антисимметричную части.

Первую обозначим посредством n^{ab}, а вторую выразим через дуальный ей «вектор» a_c :

$$C^{ab} = n^{ab} + e^{abc}a_c$$

Подстановка этого выражения в (116.17) приводит к условию

$$n^{ab}a_b = 0$$

Преобразованиями (116.19) симметричный «тензор» n^{ab} может быть приведен к диагональному виду; пусть n_1, n_2, n_3 - его главные значения.

Равенство $n^{ab}a_b = 0$ показывает, что «вектор» a_b (если он существует) лежит в одном из главных направлений «тензора» n^{ab}, - в том, которое отвечает нулевому главному значению.

Не уменьшая общности, можно поэтому положить $a_b = (a, 0, 0)$. Тогда $n^{ab}a_b = 0$ сводится к $an_1 = 0$, т. е. одна из величин a или n_1 должна быть нулем.

Правила же коммутации (116.16) примут виД

$$[X_1, X_2] = -aX_2 + n_3X_3, \quad [X_2, X_3] = n_1X_1, \quad [X_3, X_1] = n_2X_2 + aX_3$$

После этого остается еще свобода в изменении знака операторов X_a и в произвольных их масштабных преобразованиях (умножению на постоянные).

Это дозволяет одновременно изменить знак всех n_1, n_2, n_3, а также сделать величину a положительной (если она отлична от нуля).

Можно также обратить все структурные константы в ± 1, если по крайней мере одна из величин a, n_2, n_3 равна нулю.

Если же все эти три величины отличны от нуля, то масштабные преобразования оставляют инвариантным отношение a^2/n_2n_3 ").

Таким образом, мы приходим к следующему перечислению возможных типов однородных пространств; в первом столбце таблицы римской Цифрой указан номер, которым принято обозначать типы по классификации Бианки (L. Bianchi, 1918)

(тут таблица, которую мб потом перепишу)

ТиП I - евклидово пространство; все компоненты пространственного тензора кривизны (см.

ниже формулу (116.24)) обращаются в нуль.

Помимо тривиального случая галилеевой метрики, сюда относится зависящая от времени метрика, которая будет рассмотрена в следующем параграфе.

ТиП IX содержит в себе как частный случай пространство постоянной положительной кривизны.

Оно получается, если в элементе длИны (116.2) положить $\eta_{ab} = \delta_{ab}/4\lambda$, где λ - положительная постоЯнная.

Действительно, вычисление по (116.24)с$C^{11} = C^{22} = C^{33} = 1$ (структурные константы типа IX) дает $P_{(a)(b)} = 1/2\delta_{ab}$ и затем

$$P_{\alpha\beta} = P_{(a)(b)}e^{(a)}_{\alpha}e^{(b)}_{\beta} = 2\lambda\gamma_{\alpha\beta}$$

что как раз и соответствует указанному пространству (ср.
(111.3)).

Аналогичным образом пространство постоянной отрицательной кривизны содержится как частный случай в типе V.

Действительно, положив $\eta_{ab} = \delta_{ab}/\lambda$ и вычислив $P_{(a)(b)}$ по (116.24)с$C^{23} = -C^{32} = 1$ ПОЛучИМ

$$P_{(a)(b)} = -2\delta_{ab}, \quad P_{\alpha\beta} = -2\lambda\gamma_{\alpha\beta}$$

что и отвечает постоянной отрицательной кривизне.

Наконец, покажем, каким образом уравнения Эйнштейна для мира с однородным пространством сводятся к системе обыКновенных Дифференциальных уравнений, содержащих только функЦиИ времени.

ДлЯ этого пространственные компоненты 4-векторов и 4-тензоров надо разложить по тройке реперных векторов данного пространства:

$$R_{(a)(b)} = R_{\alpha\beta}e^{\alpha}_{(a)}e^{\beta}_{(b)}, \quad R_{0(a)} = R_{0\alpha}e^{\alpha}_{(a)}, \quad u^{(a)} = u^{\alpha}e^{(a)}_{\alpha}$$

причем все эти величины являются уже фунКциями только от t; фунКЦиЯми времени являются также и скалярные величины - плотность энергиИ ε и давление p материИ.

Уравнения Эйнштейна в синхронной системе выражаются, согласно (97.11) – (97.13), через трехмерные тензоры $_{\alpha\beta}$ и $P_{\alpha\beta}$. Для первого имеем просто

$$\mathscr{K}_{(a)(b)} = \dot{\eta}_{ab}, \quad x^{(b)}_{(a)} = \dot{\eta}_{ac}\eta^{cb}$$

(точка означает дифференцирование по t).

Компоненты же $P_{(a)(b)}$ можНО Выразить через величины η_{ab} и структурные константы групПы с помощью (98.14).

После замены трехиндексовых $\lambda_{bc} = C^{a}_{bc}$ на двухиндексовые C^{ab} и ряда преобразований 1) получим:

$$P^{(b)}_{(a)} = \frac{1}{2\eta}\left\{2C^{bd}C_{ad} + C^{db}C_{ad} + C^{bd}C_{da} - C^{d}_{d}\left(C^{b}_{a} + C^{b}_{a}\right) + \right.$$
$$\left. + \delta^{b}_{a}\left[\left(C^{d}_{d}\right)^{2} - 2C^{df}C_{df}\right]\right\}$$

Здесь, в соответствии с обШим правилом,

$$C^{b}_{a} = \eta_{ac}C^{cb}, \quad C_{ab} = \eta_{ac}\eta_{bd}C^{cd}$$

Отметим также, что тождества Бианки для трехмерного тензора $P_{\alpha\beta}$ В однородном пространстве принимают виД

$$P^{c}_{b}C^{b}_{ca} + P^{c}_{a}C^{b}_{cb} = 0$$

Окончательные выражения для реперных составляющих 4-тензора Риччи [2])

Подчеркнем, что для составления уравнений Эйнштейна нет, таким образом, необходимости в использовании явных выражений для репер- ных векторов как функций координат.

6.4.3 Flat Anisotropic Model

вкратце

Theory

Адекватность изотропной модели для описания поздних этапов эволюции Вселенной сама по себе не дает оснований ожидать, что она столь же пригодна и для описания ранних стадий эволюции, вблизи особой точки по времени. Этот вопрос будет детально обсужден

в пар 119, а в этом и следующем параграфах будут предварительно рассмотрены решения уравнений Эйнштейна, тоже обладающие особой точкой по времени, но принципиально отличных (от фридмановской особенности) типов.

Будем искать решение, в котором все компоненты метрического тензора являются, при надлежащем выборе системы отсчета, функциями лишь одной переменной - времени $x^0 = t^1$).

Такой вопрос рассматривался уже в $ 109, где, однако, был рассмотрен только случай, когда определитель $|g_{\alpha\beta}| = 0$. Теперь уже будем считать этот определитель отличным от нуля.

Как было показано в $ 109, в таком случае можно, без ограничения общности, положить все $g_{0\alpha} = 0$. Преобразованием переменной t, согласно $\sqrt{g_{00}}dt \to dt$ можно затем обратить g_{00} В единицу, так что мы получим синхронную систему отсчета, в которой

$$g_{00} = 1, \quad g_{0\alpha} = 0, \quad g_{\alpha\beta} = -\gamma_{\alpha\beta}(t)$$

Теперь мы можем воспользоваться уравнениями Эйнштейна в виде (97.11)−(97.13). Поскольку величины $\gamma_{\alpha\beta}$, а с ними и компоненты трехмерного тензора $\mathcal{X}_{\alpha\beta} = \dot{\gamma}_{\alpha\beta}$ не зависят от координат x^α, то $R_{0\alpha} \equiv 0$.

По той же причине $P_{\alpha\beta} \equiv 0$, и в результате уравнения гравитационного поля в пустоте сводятся к следующей системе:

$$l\dot{x}^\alpha_\alpha + \frac{1}{2}x^\beta_\alpha x^\alpha_\beta = 0$$
$$\frac{1}{\sqrt{\gamma}}\left(\sqrt{\gamma}x^\beta_\alpha\right)^\bullet = 0$$

Из (117.3) следует, что

$$\sqrt{\gamma}x^\beta_\alpha = 2\lambda^\beta_\alpha$$

где λ^β_α - постоянные величины.

Упрощая по индексам α и β, Получим:

$$x^\alpha_\alpha = \frac{\dot{\gamma}}{\gamma} = \frac{2}{\sqrt{\gamma}}\lambda^\alpha_\alpha$$

откуда видно, что $\gamma = \text{const}\ t^2$. Без ограничения общности можно положить const $= 1$ (это достигается просто изменением масштаба координат x^α), тогда $\lambda^\alpha_\alpha = 1$. Подстановка (117.4) в уравнение (117.2) дает теперь соотношение

$$\lambda^\beta_\alpha\lambda^\alpha_\beta = 1$$

связывающее между собой постоянные λ^β_α.

Далее, опустив в (117.4) индекс β, перепишем эта равенства в виде системы обыкновенных дифференциальных уравнений для $\gamma_{\alpha\beta}$:

$$\dot{\gamma}_{\alpha\beta} = \frac{2}{t}\lambda^\gamma_\alpha\gamma_{\gamma\beta}$$

Совокупность коэффициентов λ^γ_α можно рассматривать как матрицу некоторой линейной подстановки.

Путем соответствующего линейного преобразования координат x^1, x^2, x^3 (или, что эквивалентно, величин $g_{1\beta}, g_{2\beta}, g_{3\beta}$) можно, вообще говоря, привести эту матрицу к диагональному виду.

Обозначим ее главные значения посредством p_1, p_2, p_3 И будем считать, что все они вещественны и различны (о других случаях - см. ниже); единичные векторы в соответствующих

главных направлениях пусть будут $\mathbf{n}^{(1)}, \mathbf{n}^{(2)}, \mathbf{n}^{(3)}$. Тогда решение уравнений (117.6) можно представить в виде

$$\gamma_{\alpha\beta} = t^{2p_1} n_\alpha^{(1)} n_\beta^{(1)} + t^{2p_2} n_\alpha^{(2)} n_\beta^{(2)} + t^{2p_3} n_\alpha^{(3)} n_\beta^{(3)}$$

(постоянные коэффициенты при степенях t можно обратить в единицу путем соответствующего выбора масштаба координат).

Наконец, выбрав направления векторов $\mathbf{n}^{(1)}, \mathbf{n}^{(2)}, \mathbf{n}^{(3)}$ в качестве окончательного направления осей (назовем их x, y, z), приведем метрику к виду

$$ds^2 = dt^2 - t^{2p_1} dx^2 - t^{2p_2} dy^2 - t^{2p_3} dz^2$$

(E. Kasner, 1922).

Здесь p_1, p_2, p_3 - любые три числа, удовлетворяющие двум соотношениям:

$$p_1 + p_2 + p_3 = 1, \quad p_1^2 + p_2^2 + p_3^2 = 1$$

(первое следует из $-g = t^2$, а второе получается затем из (117.5)). Три числа p_1, p_2, p_3 не могут, очевидно, иметь одинаковые значения.

Равенство двух из них имеет место в тройках значений (0,0,1) и $(-1/3, 2/3, 2/3)$. Во всех других случаях числа p_1, p_2, p_3 различны причем одно из них отрицательно, а два других положительны.

Если расположить их в порядке $p_1 < p_2 < p_3$, то их значения будут лежать в интервалах

$$-\frac{1}{3} \leqslant p_1 \leqslant 0, \quad 0 \leqslant p_2 \leqslant \frac{2}{3}, \quad \frac{2}{3} \leqslant p_3 \leqslant 1$$

Таким образом, метрика (117.8) соответствует плоскому однородному, но анизотропному пространству, все объемы в котором растут (с увеличением времени) пропорционально t, причем линейные расстояния вдоль двух осей (y, z) увеличиваются, а вдоль одной (x) убывают.

Момент $t = 0$ является особой точкой решения; метрика имеет в ней особенность, не устранимую никаким преобразованием системы отсчета, причем инварианты тензора четырехмерной кривизны обращаются в бесконечность.

Исключением является лишь случай $p_1 = p_2 = 0, p_3 = 1$; при этих значениях мы имеем дело просто с плоскИм пространством-временем: преобразованием $t \operatorname{sh} z = \zeta, t \operatorname{ch} z = \tau$ метрика (117.8) приводится к галилеевой 1).

Метрика (117.8) является точным решением уравнений Эйнштейна ДлЯ пустого пространства.

Но вблизи особой точки, при малых t, она остается приближенным (с точностьЮ до членов главного порядка по $1/t$) решением уравнений и при наличии равномерно распределенной В пространстве материи.

Скорость и ход изменения плотности материи определяются при этом просто уравнениями ее движения в заданном гравитационном поле, а обратное влияние материи на поле оказывается пренебрежимым.

Плотность материи стремится к бесконечности при $t \to 0$ - в соответствии с физическим характером особенности (см. задачу 3).

ЛЛ2-117.1

Найти решение уравнений (117.6), соответствующее случаю, когда матрица λ_α^β имеет одно вещественное (p_3) и два комплексных $(p_{1,2} = p' \pm ip'')$ главных значения.

Р е ш е н и е.

В этом случае переменная x^0, от которой зависят все величины, должна иметь пространственный характер; обозначим ее как

$x^0 = x$. Соответственно в (117.1) должно быть теперь $g_{00} = -1$. Уравнения же (117.2)–(117.3) не меняются.

Векторы $\mathbf{n}^{(1)}, \mathbf{n}^{(2)}$ В (117.7) становятся комплексными: $\mathbf{n}^{(1,2)} = (\mathbf{n}' \pm \pm i\mathbf{n}'')/\sqrt{2}$, где $\mathbf{n}', \mathbf{n}''$ - единичные векторы.

Выбирая оси x^1, x^2, x^3 в направлениях $\mathbf{n}', \mathbf{n}'', \mathbf{n}^{(3)}$, получим решение в виде

$$-g_{11} = g_{22} = x^{2p'} \cos\left(2p'' \ln \frac{x}{a}\right), \quad g_{12} = -x^{2p'} \sin\left(2p'' \ln \frac{x}{a}\right)$$
$$g_{33} = -x^{2p_3}, \quad -g = -g_{00}\,|g_{\alpha\beta}| = x^2$$

где a - постоянная (которую уже нельзя устранить выбором масштаба вдоль оси x, не изменив других коэффициентов в написанных выражениях).

Числа p_1, p_2, p_3 по-прежнему удовлетворяют соотношениям (117.9), причем вещественное число p_3 либо меньше $-1/3$, либо больше единицы.

ЛЛ2-117.2

То же в случае совпадающих двух главных значений ($p_2 = p_3$).

Р е ш е н и е.

Как известно, из общей теории линейных дифференциальных уравнений, в этом случае система (117.6) может быть приведена к следующему каноническому виду:

$$\dot{g}_{11} = \frac{2p_1}{x} g_{11}, \quad \dot{g}_{2\alpha} = \frac{2p_2}{x} g_{2\alpha}, \quad \dot{g}_{3\alpha} = \frac{2p_2}{x} g_{3\alpha} = a_\alpha + \frac{\lambda}{x} g_{2\alpha}, \quad \alpha = 2, 3$$

где λ - постоянная.

При $\lambda = 0$ мы возвращаемся к (117.8).

При $\lambda \neq 0$ можно положить $\lambda = 1$; тогда

$$g_{11} = -x^{2p_1}, \quad g_{2\alpha} = a_\alpha x^{2p_2}, \quad g_{3\alpha} = a_\alpha x^{2p_2} \ln x + b_\alpha x^{2p_2}$$

Из условия $g_{32} = g_{23}$ находим, что $a_2 = 0, a_3 = b_2$. Надлежащим выбором масштаба вдоль осей x^2, x^3 окончательно приводим метрику к следующему виду:

$$ds^2 = -dx^2 - x^{2p_1} \left(dx^1\right)^2 \pm 2x^{2p_2} dx^2 dx^3 \pm x^{2p_2} \ln \frac{x}{a} \left(dx^3\right)^2$$

Числа p_1, p_2 могут иметь значения 1,0 или $-1/3, 2/3$.

ЛЛ2-117.3

Вблизи особой точки $t = 0$ найти закон изменения со временем плотности материи, равномерно распределенной в пространстве с метрикой (117.8).

Р е ш е н и е.

Пренебрегая обратным влиянием материи на поле, исходим из гидродинамических уравнений движения

$$\frac{1}{\sqrt{-q}} \frac{\partial}{\partial x^i} \left(\sqrt{-g}\sigma u^i\right) = 0$$

$$(p+\varepsilon)u^k\left(\frac{\partial u_i}{\partial x^k}-\frac{1}{2}u^l\frac{\partial g_{kl}}{dx^i}\right)=-\frac{\partial p}{\partial x^i}-u_iu^k\frac{\partial p}{\partial x^k}$$

содержащихся в уравнениях $T^k_{i;k}=0$ (см. «Гидродинамика», \$ 134).

Здесь σ - плотность энтропии; вблизи особенности надо пользоваться ультрарелятивистским уравнением состояния $p=\varepsilon/3$, и тогда $\sigma\propto e^{3/4}$.

Обозначим временные множители в (117.8) посредством $a=t^{p_1}, b=t^{p_2}$, $c=t^{p_3}$. Поскольку все величины зависят только от времени, а $\sqrt{-g}=abc$, уравнения (1) дают

$$\frac{d}{dt}\left(abcu_0\varepsilon^{3/4}\right)=0,\quad 4\varepsilon\frac{du_\alpha}{dt}+u_\alpha\frac{d\varepsilon}{dt}=0$$

Отсюда

$$abcu_0\varepsilon^{3/4}=\text{const}$$
$$u_\alpha\varepsilon^{1/4}=\text{const}.$$

Согласно (3) все ковариантные составляющие u_α - одинакового порядка величины.

Из контравариантных же компонент наиболее велика (при $t\to 0$) $u^3\approx u_3/c^2$. Сохранив в тождестве $u_iu^i=1$ лишь наибольшие члены, получим поэтому $u_0^2\approx u_3u^3=(u_3)^2/c^2$ и затем из (2) и (3):

$$\varepsilon\propto\frac{1}{a^2b^2},\quad u_\alpha\propto\sqrt{ab}$$

или

$$\varepsilon\propto t^{-2(p_1+p_2)}=t^{-2(1-p_3)},\quad u_\alpha\propto t^{(1-p_3)/2}$$

Как и следовало, ε стремится при $t\to 0$ к бесконечности для всех значений p_3, за исключением лишь $p_3=1$, - в соответствии с тем, что особенность в метрике с показателями (0,0,1) фиктивна.

Справедливость использованного приближения проверяется оценкой компонент T^k_i, опущенных в правых частях уравнений (117.2)-(117.3).

Главные члены в них:

$$llT^0_0\sim\varepsilon u_0^2\propto t^{-(1+p_3)},T^1_1\sim\varepsilon\propto t^{-2(1-p_3)}$$
$$T^2_2\sim\varepsilon u_2u^2\propto t^{-(1+2p_2-p_3)},T^3_3\sim\varepsilon u_3u^3\propto t^{-(1+p_3)}$$

Все они действительно растут при $t\to 0$ медленнее, чем левые стороны уравнений, возрастающие как t^{-2}.

6.4.4 Oscillations During Approach to Singularity

вкратце

Есть модель Кнезера, в которой какие-то есть колебания.

Theory

На примере модели мира с однородным пространством типа IX изучим особенность метрики по времени, имеющую колебательный характер.

Мы увидим в следующем параграфе, что такой характер имеет весьма общее значение.

Нас будет интересовать поведение модели вблизи особой точки (которую выберем в качестве начала отсчета времени, $t=0$).

Как и в рассмотренном в \$ 117 решении Казнера, наличие материи не отражается на качественных свойствах этого поведения, и для упрощения исследования будем предполагать пространство пустым.

Положим в (116.3)????? матрицу величин $\eta_{ab}(t)$ диагональной, обозначив ее диагональные элементы через a^2, b^2, c^2; три реперных вектора $\mathbf{e}^{(1)}, \mathbf{e}^{(2)}, \mathbf{e}^{(3)}$ обозначим теперь посредством $1, \mathbf{m}, \mathbf{n}$.

Тогда пространственная метрика запишется в виде

$$\gamma_{\alpha\beta} = a^2 l_\alpha l_\beta + b^2 m_\alpha m_\beta + c^2 n_\alpha n_\beta$$

Для пространства типа IX?????? структурные константы

$$C^{11} = C^{22} = C^{33} = 1$$

при этом $C^1_{23} = C^2_{31} = C^3_{12} = 1$

Из (116.26) видно, что для таких констант и при диагональной матрице η_{ab} компоненты $R^0_{(a)}$ тензора Риччи в синхронной системе отсчета обращаются в нуль тождественно.

Согласно (116.24) обращаются в нуль также и недиагональные компоненты $P_{(a)(b)}$.

Остальные компоненты уравнений Эйнштейна дают для функций $a(t), b(t), c(t)$ систему уравнений:

$$\begin{aligned}
\frac{(\dot{a}bc)^{\cdot}}{abc} &= \frac{1}{2a^2b^2c^2}\left[\left(b^2-c^2\right)^2 - a^4\right] \\
\frac{(abc)^2}{abc} &= \frac{1}{2a^2b^2c^2}\left[\left(a^2-c^2\right)^2 - b^4\right] \\
\frac{(ab\dot{c})^{-}}{abc} &= \frac{1}{2a^2b^2c^2}\left[\left(a^2-b^2\right)^2 - c^4\right] \\
&\frac{\ddot{a}}{a} + \frac{\ddot{b}}{b} + \frac{\ddot{c}}{c} = 0
\end{aligned}$$

((118.3) - уравнения $R^{(1)}_{(1)} = R^{(2)}_{(2)} = R^{(3)}_{(3)} = 0$; (118.4) - уравнение $R^0_0 = 0$)

Производные по времени в системе (118.3) – (118.4) принимают более простой вид, если ввести вместо функций a, b, c их логарифмы α, β, γ :

$$a = e^\alpha, \quad b = e^\beta, \quad c = e^\gamma$$

а вместо t - переменную τ согласно

$$dt = abcd\tau$$

Тогда:

$$\begin{aligned}
2\alpha_{,\tau,\tau} &= \left(b^2-c^2\right)^2 - a^4 \\
2\beta_{,\tau,\tau} &= \left(a^2-c^2\right)^2 - b^4 \\
2\gamma_{,\tau,\tau} &= \left(a^2-b^2\right)^2 - c^4 \\
\frac{1}{2}(\alpha+\beta+\gamma)_{,\tau,\tau} &= \alpha_{,\tau}\beta_{,\tau} + \alpha_{,\tau}\gamma_{,\tau} + \beta_{,\tau}\gamma_{,\tau}
\end{aligned}$$

где индекс τ означает дифференцирование по τ. Сложив почленно уравнения (118.7) и заменив в левой стороне сумму вторых производных согласно (118.8), получим:

$$\alpha_{,\tau}\beta_{,\tau} + \alpha_{,\tau}\gamma_{,\tau} + \beta_{,\tau}\gamma_{,\tau} = \frac{1}{4}\left(a^4 + b^4 + c^4 - 2a^2b^2 - 2a^2c^2 - 2b^2c^2\right)$$

Это соотношение содержит только первые производные и представляет собой первый интеграл уравнений (118.7).

Уравнения (118.3) – (118.4) не могут быть решены точно в аналитическом виде, но вблизи особой точки допускают детальное качественное исследование.

Прежде всего замечаем, что в отсутствие правых частей в уравнениях (118.3) (или, что то же, в уравнениях (118.7)) система имела бы точное решение, в котором

$$a \sim t^{p_l}, \quad b \sim t^{p_m}, \quad c \sim t^{p_n}, \quad (118.10)$$

где p_l, p_m, p_n - числа, связанные соотношениями

$$p_l + p_m + p_n = p_l^2 + p_m^2 + p_n^2 = 1$$

(аналог решения Казнера (117.8) для однородного плоского пространства).

Мы обозначили здесь показатели степеней как p_l, p_m, p_n, не предопределяя их последовательности в порядке возрастания; обозначение же p_1, p_2, p_3 из §117 сохраним за тройками чисел, расположенных в порядке $p_1 < p_2 < p_3$ и соответственно пробегающими значения в интервалах (117.10).

Эти числа могут быть представлены в параметрическом виде как

$$p_1(u) = \frac{-u}{1+u+u^2}, \quad p_2(u) = \frac{1+u}{1+u+u^2}, \quad p_3(u) = \frac{u(1+u)}{1+u+u^2}$$

Все различные значения p_1, p_2, p_3 (с соблюдением условного порядка) Получаются, если параметр u пробегает значения в области $u \geqslant 1$. Значения же $u < 1$ приводятся к той же области согласно

$$p_1\left(\frac{1}{u}\right) = p_1(u), \quad p_2\left(\frac{1}{u}\right) = p_3(u), \quad p_3\left(\frac{1}{u}\right) = p_2(u)$$

На рис. 25 изображены графики p_1, p_2, p_3 в зависимости от $1/u$.

Предположим, что в некотором интервале времени правые части В уравнениях (118.7) действительно малы, так что ими можно пренебречь и имеет место казнеровский режим (118.10).

Такая ситуация не может продолжаться (при $t \to 0$) неограниченно, так как среди указанных членов всегда имеются возрастающие.

Так, если отрицательный показатель степени относится к функции $a(t)$ $(p_l = p_1 < 0)$, то возмущение казнеровского режима возникает от членов a^4; остальные же члены при уменьшении t будут убывать.

Сохранив в правых частях (118.7) лишь эти члены, получим систему уравнений

$$\alpha_{,\tau,\tau} = -\frac{1}{2}e^{4\alpha}, \quad \beta_{,\tau,\tau} = \gamma_{,\tau,\tau} = \frac{1}{2}e^{4\alpha}$$

Решение этих уравнений должно описывать эволюцию метрики из «начального» состояния, в котором оно описывается формулами (118.10) с определенным набором показателей (причем $p_l < 0$); пусть $p_l = p_1, p_m = p_2, p_n = p_3$, так что

$$a = t^{p_1}, \quad b = t^{p_2}, \quad c = t^{p_3}$$

(коэффициенты пропорциональности в этих выражениях можно положить равными 1 без ограничения общности получаемого ниже результата).

При этом $abc = t, \tau = \ln t +$ const, поэтому начальные условия длЯ уравнений (118.14) формулируются в виде

$$\alpha_{,\tau} = p_1, \quad \beta_{,\tau} = p_2, \quad \gamma_{,\tau} = p_3$$

Первое из уравнений (118.14) имеет вид уравнения одномерного движения частицы в поле экспоненциальной потенциальной стенки, причем α играет роль координаты.

В этой аналогии начальному казнеровскому режиму соответствует свободное движение с постоянной скоростью $\alpha_{,\tau} = p_1$. Отразившись от стенки, частица будет снова двигаться свободно со скоростью обратного знака: $\alpha_{,\tau} = -p_1$. Заметив также, что в силу всех трех уравнений (118.14)

$$\alpha_{,\tau} + \beta_{,\tau} = \text{const}, \quad \alpha_{,\tau} + \gamma_{,\tau} = \text{const}$$

найдем, что $\beta_{,\tau}$ и $\gamma_{,\tau}$ приобретут значения $\beta_{,\tau} = p_2 + 2p_1, \gamma_{,\tau} = p_3 + 2p_1$ Определив отсюда α, β, γ и затем t согласно (118.6), Получим:

$$e^{\alpha} \sim e^{-p_1\tau}, \quad e^{\beta} \sim e^{(p_2+2p_1)\tau}, \quad e^{\gamma} \sim e^{(p_3+2p_1)\tau}, \quad t \sim e^{(1+2p_1)\tau}$$

Т. е. $a \sim t^{p'_l}, b \sim t^{p'_m}, c \sim t'^{p}_{n}$, где

$$p'_l = \frac{|p_1|}{1-2|p_1|}, \quad p'_m = -\frac{2|p_1| - p_2}{1-2|p_1|}, \quad p'_n = \frac{p_3 - 2|p_1|}{1-2|p_1|}$$

Таким образом, воздействие возмущения приводит к смене одной «Казнеровской эпохи» другой, причем отрицательная степень t перебрасывается с направления l на направление m : если было $p_l < 0$, то теперь $p'_m < 0$. В процессе смены функция $a(t)$ проходит через максимум, а $b(t)$ - через минимум: убывавшая прежде величина $b(t)$ начинает возрастать, возраставшая $a(t)$ - падать, а функция $c(t)$ продолжает убывать.

Само возмущение (члены a^4 в уравнениях (118.7)), прежде возраставшее, начинает убывать и затухает.

Дальнейшая эволюция метрики приведет аналогичным образом к возрастанию возмущения, выражающегося членами b^4 в уравнениях (118.7), следующей смене казнеровских показателей, и т.д.

Правило смены показателей (118.15) удобно представить с помощью параметризации (118.12): если

$$p_l = p_1(u), \quad p_m = p_2(u), \quad p_n = p_3(u)$$

то

$$p'_l = p_2(u-1), \quad p'_m = p_1(u-1), \quad p'_n = p_3(u-1)$$

Остается положительным больший из двух положительных показателей.

В этом процессе смен казнеровских эпох лежит ключ к пониманию характера эволюции метрики при приближении к особой точке.

Последовательные смены (118.16) с перебросом отрицательного показателя степени (p_1) между направлениями 1 и m продолжаются до тех пор, пока не исчерпается целая часть начального значения и И не станет $u < 1$. Значение $u < 1$ преобразуется В $u > 1$ согласно (118.13); в этот момент отрицателен показатель p_l Или p_m, а p_n становится меньшим из двух положительных чисел ($p_n = p_2$). СледуюЩая серия смен будет уже перебрасывать отрицательный показатель между направлениями **n** и 1 или между **n** и **m**. При произвольном (иррациональном) начальном значении u процесс смен продолжается неограниченно.

При точном решении уравнений показатели p_1, p_2, p_3 теряют, конечно, свой буквальный смысл.

Отметим, что вносимая этим обстоятельством некоторая «размытость» в определении этих чисел (а с ними и параметра u), хотя она и мала, лишает смысла рассмотрение как-либо выделенных (например, рациональных) значений и.

Именно поэтому реальным смыслом обладают лишь те закономерности, которые свойственны общему случаю произвольных иррациональных значений и.

Таким образом, процесс эволюции модели в направлении к особой точке складывается из последовательных серий колебаний, в течение Каждой из которых расстояния вдоль

двух пространственных осей осциллируют, а вдоль третьей - монотонно убывают; объем убывает по закону, близкому К $\sim t$. При переходе от одной серии к следущая серия смен будет уже перебрасывать отрицательный показатель (иррациональном) начальном значении и процесс смен продолжается неограниченно.

При точном решении уравнений показатели p_1, p_2, p_3 теряют, конечно, свой буквальный смысл.

Отметим, что вносимая этим обстоятельством некоторая «размытость» в определении этих чисел (а с ними и параметра u), хотя она и мала, лишает смысла рассмотрение как-либо выделенных (например, рациональных) значений и.

Именно поэтому реальным смыслом обладают лишь те закономерности, которые свойственны общему случаю произвольных иррациональных значений и.

Таким образом, процесс эволюцИи модели в направлении к особой точке складывается из последовательных серий колебаний, в течение Каждой из которых расстояния вдоль двух пространственных осей осциллируют, а вдоль третьей - монотонно убывают; объем убывает по закону, близкому К $\sim$ t.

При переходе от одной серии к следуюЩей направление, вдоль которого происходит монотонное убывание расстояний, переходит с одной оси на другую.

Порядок этих переходов приобретает асимптотически характер случайного процесса.

Такой же характер приобретает и порядок чередования длин последовательных серий колебаний (т. е. чисел сменяющихся в каждой серии «казнеровских эпох»).

Последовательные серии колебаний сгущаются по мере приближения к особой точке.

Между любым конечным моментом мирового времени t и моментом $t = 0$ заключено бесконечное множество колебаний.

Естественной переменной для описания временного хода этой эволюции оказывается не само время t, а его логарифм, $\ln t$, по которому весь процесс приближения к особой точке растянут до - ∞.

В изложенном решении мы с самого начала несколько упростили задачу, предположив матрицу $\eta_{ab}(t)$ в (116.3) диагональной.

Включение в метрику недиагональных компонент η_{ab} не меняет описанного колебательного характера эволюции метрики и закона (118.16) смен показателей p_l, p_m, p_n чередующихся казнеровских эпох.

Оно приводит, однако, к появлению дополнительного свойства: смена показателей сопровождается также и изменением направлений осей, к которым эти показатели относятся.

6.4.5 Singularity in Einstein's Equation for Cosmology (???)

вкратце особенности

Theory

Уже было отмечено, что адекватность модели Фридмана для описания современного состояния Вселенной сама по себе еще не дает оснований ожидать, что она столь же пригодна и для описания ранних стадий эволюции мира.

В этой связи возникает прежде всего вопрос о том, в какой степени существование особой точки по времени вообще является обязательным свойством космологических моделей; и не связано ли оно со специфическими упрощающими предположениями (в первую очередь с симметрией), лежащими в их основе.

Подчеркнем, что, говоря об особой точке, мы имеем в виду физическую особенность — обращение в бесконечность плотности материи и инвариантов тензора четырехмерной кривизны.

Независимость от специфических предположений означала бы, что наличие особенности присуще не только частным, но и общему решению уравнений Эйнштейна.

Критерием общности является число содержащихся в решений «физически произвольных» функций. В общем решении число таких функций должно быть достаточным для произвольного задания начальных условий в какой-либо выбран- ный момент времени (4 для пустого пространства, 8 для пространства, заполненного материей, — см. пар 95).

Нахождение общего решения в точном виде для всего пространства в течение всего времени, разумеется, невозможно.

Но для решения поставленного вопроса в этом нет необходимости: достаточно исследовать вид решения вблизи особенности. Особенность, которую имеет решение Фридмана, характерна тем, что обращение в нуль пространственных расстояний происходит по одинаковому закону во всех направлениях.

Такой тип особенности не является достаточно общим: он свойствен классу решений, содержащему лишь три произвольные функции координат (см. задачу к пар 113).

Отметим также, что эти решения существуют только для пространства, заполненного материей. Особенность же колебательного типа, рассмотренная в предыдущем параграфе, имеет общий характер — существует решение уравнений Эйнштейна с такой особенностью, содержащее всю требуемую совокупность произвольных функций.

Мы обрисуем здесь кратко способ построения такого решения, не вникая в детали вычислений 1). Как и в однородной модели (пар 118), режим приближения к особой точке в общем решении складывается из чередующихся серий сменяющих друг друга «казнеровских эпох».

В течение каждой такой эпохи главные (по $1/t$) члены в пространственном метрическом тензоре (в синхронной системе отсчета) имеют вид (118.1) с функциями времени a, b, c из (118.10), но векторы l, m, n являются теперь произвольными (а не вполне определенными, как в однородной модели) функциями пространственных координат.

Такими же функциями (а не просто числами) являются теперь показатели pl, pm, pn, по-прежнему связанные друг с другом соотношениями (118.11).

Построенная таким образом метрика удовлетворяет уравнениям $R_0^0 = 0$ и $R_\alpha^\beta = 0$ для поля в пустоте (В иХ Главных членах) в течение некоторого конечного интервала времени.

Уравнения же $R_\alpha^0 = 0$ приводят к трем соотношениям (не содержащим времени), которые должны быть наложены на содержащиеся В $\gamma_{\alpha\beta}$ произвольные функции пространственных координат.

Эти соотношения связывают между собой 10 различных фунКций: по три степеней времени (какая-либо из трех функций p_l, p_m, p_n, связанных Двумя условиями (118.11)).

При определении числа физически проИзвольных функций надо учесть также, что синхронная система отсчета допускает еще произвольные преобразования трех пространственных Координат, не затрагивающие времени.

поэтому метрика содержит всего $10-3-3 = 4$ произвольные функции - именно столько, сколько должно быть в обшем решении для поля в пустоте.

Смена одной казнеровской эпохи на другую происходит (как и в однородной модели) благодаря наличию в трех из шести уравнений $r_\alpha^\beta = 0$ членов, которые при уменьшении t растут быстрее других, играя, таким образом, роль возмущения, разрушающего казнеровский режим. эти уравнения в общем случае имеют вид, отличающийся от уравнений (118. 14) лишь зависящим от пространственных координат множителем $(\mathrm{lrot}\, 1/\mathbf{l}[\mathbf{mn}])^2$ в их правых сторонах (подразумевается, что из трех показателей p_l, p_m, p_n отрицателен p_l) 1). поскольку, однако, уравнения (118. 14) составляют систему обыкновенных дифференциальных уравнений по отношению ко времени, это отличие никак не сказывается на их решении и на следуюшем из этого решения законе смены казнеровских показателей (118. 16), а тем самым и на всех дальнейших следствиях, изложенных в \$118[2]). степень общности

решения не уменьшается при введении материи: материя «вписывается» в метрику со всеми вносимыми ею 4 новыми координатными функциями, необходимыми для задания начальных распределений ее плотности и трех компонент скорости. тензор энергии-импульса материи T_i^k привносит в уравнения поля члены, оказывающиеся более высокого порядка по $1/t$, чем главные члены (в точности аналогично тому, как это было показано в задаче 3 к \$117 для плоской однородной модели).

Таким образом, существование особой точки по времени является весьма общим свойством решений уравнений Эйнштейна, причем режим приближения к особой точке имеет в общем случае колебательный характер [3]). Подчеркнем, что этот характер не связан с наЛичием материи (а потому и с ее уравнением состояния) и свойствен уже самому по себе пустому пространству-времени. Зависящая же

От присутствия материи особенность монотонного изотропного тиПа, свойственная решению Фридмана, имеет лишь частное значение.

Говоря об особенностях в космологическом аспекте, мы имеем В виду особую точку, достигаемую всем пространством, а не лишь его ограниченной частью, как при гравитационном коллапсе конечного тела. Но общность колебательного решения дает основание полагать,

что такой же характер имеет и особенность, достигаемая конечным те- лом в его коллапсе под горизонтом событий в сопутствующей системе отсчета. Мы везде говорили о направлении приближения к особой точке как о направлении уменьшения времени; но ввиду симметрии уравнений Эйнштейна по отношению к изменению знака времени, с тем же правом могла бы идти речь о приближении к особенности в направлении уве- личения времени.

В действительности, однако, ввиду физической неэк- вивалентности будущего и прошедшего между этими двумя случаями имеется существенное отличие в отношении самой постановки вопроса. Особенность в будущем может иметь физический смысл, лишь если она допустима при произвольных начальных условиях, задаваемых в какой-либо предшествующий момент времени.

Ясно, что нет никаких оснований для того, чтобы распределение материи и поля, достигаемое в какой-либо момент в процессе эволюции Вселенной, соответствовало бы специфическим условиям, требуемым для осуществления того или иного частного решения уравнений Эйнштейна. На вопрос же о типе особенности в прошлом исследование, основанное на одних лишь уравнениях гравитации, вряд ли вообще может дать однозначный ответ.

Естественно думать, что отбор решения, отвечающего реальному миру, связан с какими-то глубокими физиче- скими требованиями, установление которых на основании одной лишь существующей теории гравитации невозможно и которые смогут быть выяснены лишь в результате дальнейшего синтеза физических теорий. В этом смысле в принципе может оказаться, что этому отбору соответ- ствует какой-либо частный (например, изотропный) тип особенности. Наконец, необходимо сделать еще следующее замечание.

Область применимости уравнений Эйнштейна самих по себе никак не ограни- чена со стороны малых расстояний или больших плотностей материи в том смысле, что уравнения не приводят в этом пределе ни к каким внутренним противоречиям (в отличие, например, от классических уравнений электродинамики).

В этом смысле исследование особенно- стей пространственно-временной метрики на базе уравнений Эйнштей- на вполне корректно.

Нет сомнений, однако, в том, что в действитель- ности в указанном пределе должны стать существенными квантовые явления, о которых при современном состоянии теорий мы еще ничего не можем сказать.

Лишь в будущем синтезе теории тяготения и кван- товой теории сможет выясниться, чт´о именно из результатов класси- ческой теории сохранит реальный смысл.

В то же время нет сомнений и в том, что самый факт возникновения особенностей в решениях урав- нений Эйнштейна (как в космологическом аспекте, так и для коллапса конечных тел) имеет глубокий физический смысл.

Не надо забывать и о том, что достижение в процессе гравитационного коллапса уже тех огромных плотностей, при которых еще нет оснований для сомнений в законности классической теории тяготения, достаточны для того, чтобы говорить о физически «особом» явлении.

6.5 Relics from the Hot Big Bang

6.5.1 Introduction

There are many reasons to believe that our current understanding of fundamental physics is incomplete. For example, the nature of dark matter and dark energy are still unknown, the stability of the Higgs mass remains unsolved, the origin of neutrino masses is unexplained, the strong CP problem is still there, the physics of inflation remains elusive, and the origin of the matter-antimatter asymmetry is still to be discovered. Attempts to address these shortcomings of the Standard Model (hereafter SM) often involve new degrees of freedom. These new particles can escape detection in terrestrial experiments either because they are too *heavy* to be produced at the energies available or because they are too *weakly coupled* to be created in sufficiently large numbers. However, in astrophysical systems and in the early universe the small cross sections can be compensated for by probing environments with large densities and by following the evolution over long time scales. For example, the high densities at the cores of stars allow a significant production of new weakly coupled species. Integrated over the long lifetime of stars this can lead to large deviations from the standard stellar evolution. Observations of the lifecycles of stars have therefore put interesting constraints on the couplings of new light species to the SM. Similarly, in the hot environment of the early universe, even extremely weakly coupled particles can be produced prolifically and their gravitational influence becomes detectable in the cosmic microwave background (CMB). Moreover, during inflation even very massive particles can be created by the rapid expansion of the spacetime. When these particles decay they produce distinctive signatures in cosmological observables. Observing these effects could teach us a great deal about the physics driving the inflationary expansion.

New physics can affect cosmological observables in two distinct ways: *i*) it may affect the *initial conditions* of the hot big bang, or *ii*) it may affect the *evolution* of these initial conditions through changes to the thermal history of the universe. These opportunities for probing new physics with cosmological observations are illustrated in the following schematic:

$$\begin{array}{ccc} & \text{new physics?} & \\ & \downarrow & \\ \underset{\text{initial conditions}}{\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \ldots \rangle} & \xrightarrow{\qquad \text{evolution} \qquad} & \underset{\text{cosmological correlators}}{\langle O_1 O_2 \ldots \rangle} \\ \uparrow & & \\ \text{new physics?} & & \end{array}$$

where ζ denotes the primordial curvature perturbations and O stands for late-time observables, which may be the temperature variations δT of the CMB or the density fluctuations $\delta\rho_g$ in the distribution of galaxies. In these lectures, we will give examples from both of these avenues of tests for physics beyond the Standard Model (BSM).

The initial fluctuations were drawn from a probability distribution $P[\zeta]$ and hence are

characterized by the following correlation functions

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \ldots \rangle = \int \mathcal{D}\zeta \, P[\zeta] \, \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \cdots \, .$$

By tracing the observed correlations $\langle O_1 O_2 \ldots \rangle$ back in time, cosmologists try to measure the initial correlations $\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \ldots \rangle$ and use them to extract information about the initial probability distribution $P[\zeta]$ and about the physics that gave rise to it. If the early universe went through a period of inflation then these initial correlations were produced dynamically *before* the hot big bang. New physics during inflation, such as the existence of new degrees of freedom X, can leave imprints in the spectrum of primordial perturbations, i.e. $P[\zeta] \to P[\zeta, X]$.

New physics may also affect how the initial correlations evolve into the late-time observables. For example, the evolution equations for fluctuations in the primordial plasma can be affected by the presence of new degrees of freedom X:

$$G_{\mu\nu} = 8\pi G \left(T^{\rm SM}_{\mu\nu} + T^{X}_{\mu\nu} \right) \, , \qquad \nabla^\mu \left(T^{\rm SM}_{\mu\nu} + T^{X}_{\mu\nu} \right) = 0 \, .$$

As we will see, much of the evolution in the early universe is very well understood and probed by very precise observations. The possibilities for new physics are therefore highly constrained.

The most conservative way to parameterize physics beyond the Standard Model is in terms of an *effective field theory* (EFT). The basic input of an EFT are the field content and the symmetries that are relevant at a given energy scale. The effective Lagrangian is then the sum of all operators consistent with the symmetries,

$$\mathcal{L}_{\rm eff} \, \subset \, \sum g \, \mathcal{O}_X \mathcal{O}_{\rm SM} \, ,$$

where $\mathcal{O}_{\rm SM}$ denotes operators made from the SM degrees of freedom and $\mathcal{O}_X$ stands for operators constructed from any additional fields. The couplings g parameterize the strengths of the interactions between the fields X and the SM. Deviations from the SM predictions then scale with the size of the couplings g. In these lectures, I will show how cosmological observations put constraints on these couplings.

Outline

The goal of these lectures is to show that cosmology is becoming an increasingly sensitive probe of BSM physics. The presentation is divided into two parts: In Part I, we study the production of new light particles in the hot big bang and describe their effects on the anisotropies of the cosmic microwave background. In Part II, we investigate the possibility of very massive particles being created during inflation and determine their imprints in higher-order cosmological correlations.

Notation and conventions

Throughout these lectures, we will use natural units, $c = \hbar = 1$, with reduced Planck mass $M_{\rm pl}^2 = 1/8\pi G$. Our metric signature is $(-+++)$. Greek letters will stand for spacetime indices, $\mu, \nu, \ldots = 0, 1, 2, 3$, and Latin letters for spatial indices, $i, j, \ldots = 1, 2, 3$. Three-dimensional vectors will be written in boldface, $\mathbf{k}$, and unit vectors will be hatted, $\hat{\mathbf{k}}$. Overdots and primes will denote derivatives with respect to conformal time τ and physical time t, respectively. The dimensionless power spectrum of a Fourier mode $\zeta_{\mathbf{k}}$ will be defined as

$$\mathcal{P}_\zeta(k) \equiv \frac{k^3}{2\pi^2} \langle \zeta_{\mathbf{k}} \zeta_{-\mathbf{k}} \rangle' \, ,$$

where the prime on the expectation value indicates that the overall momentum-conserving delta function has been dropped.

Acknowledgements

I am grateful to the organizers Igor Klebanov and Mirjam Cvetic for the invitation to lecture at TASI 2017. Thanks to my students and collaborators for many helpful discussions on the material presented in these lectures. I am especially grateful to Daniel Green, Joel Meyers, Benjamin Wallisch and Matias Zaldarriaga for collaboration on topics covered in Part I, and to Nima Arkani-Hamed, Valentin Assassi, Garrett Goon, Daniel Green, Hayden Lee, Liam McAllister and Guilherme Pimentel for collaboration on work described in Part II. Finally, thanks to Benjamin Wallisch and Swagat Saurav Mishra for comments on a preliminary version of the manuscript.

6.5.2 Theory

Consider a $2 \to N$ scattering process producing a new particle species X. Assume that X is very weakly coupled to the Standard Model degrees of freedom. The production of particles X then carries energy and momentum away from the interaction region. Such missing energy, of course, famously led to the discovery of neutrinos. However, particles that are more weakly coupled than neutrinos are produced very rarely in colliders and their missing energy signatures are hard to detect. Fortunately, the creation of new species may be enhanced in astrophysical systems and in the early universe. To see this, consider the change in the number densities of the particles involved in the above interaction. Schematically, this is given by

$$\frac{\Delta n}{n} \sim n\sigma \times \Delta t \,,$$

i.e. the fractional change in the number density is equal to the interaction rate, $\Gamma \sim n\sigma$, times the interaction time, Δt. We see that small cross sections σ can be compensated for by high densities and long time scales. This explains why observations of the lifecycles of stars have put interesting constraints on the existence of extra species. Taking the typical number density inside of stars to be $n \sim (1\,\mathrm{keV})^3$ and integrating over the typical lifetime of a star, $\Delta t \sim 10^{16}\,\mathrm{sec} \approx 1.5 \times 10^{31}\,\mathrm{eV}^{-1}$, we find significant changes in the stellar evolution if

$$\sigma > (n\Delta t)^{-1} \sim (10^{10}\,\mathrm{GeV})^{-2} \,.$$

Similar constraints can be derived from the observed energy loss in supernova explosions. In that case, the relevant time scales are much shorter, $\Delta t \sim 10\,\mathrm{sec}$, but the densities are much higher, $n \sim (30\,\mathrm{MeV})^3$.

These order of magnitude estimates also give us a sense for the power of cosmological constraints. In the early universe, the interaction time scales are short, $\Delta t < 1\,\mathrm{sec}$, but the densities can be very high, $n \sim T^3 \gg (1\,\mathrm{MeV})^3$. For temperatures above $10^4\,\mathrm{GeV}$, we expect cosmological constraints to be stronger than those from astrophysics. In this part of the lectures, we will show that primordial cosmology is indeed a highly sensitive probe of new light particles produced in the hot Big Bang.

We will begin, in Section 6.5.3, with a quick review of FRW cosmology. In Section 7.5.2, we will introduce the cosmic microwave background as a tool for precision cosmology. In Section 6.5.5, we will discuss the physics of the acoustic oscillations observed in the CMB anisotropy spectrum. Finally, in Section 6.5.6, we will show that extra relativistic particles leave a unique signature in the CMB spectrum.

6.5.3 Big Bang Cosmology

We will begin with a lightning review of some elementary concepts in cosmology. I will assume that you have seen most, if not all, of this material before, so I will cite many results without detailed derivations. Further details can be found in my *Cosmology* course [1] or in any of the standard textbooks (e.g. [2, 3]).

Geometry and Dynamics

The Friedmann-Robertson-Walker (FRW) metric of a homogenous and isotropic spacetime is

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\gamma_{ij}\,\mathrm{d}x^i\mathrm{d}x^j\,, \tag{30}$$

where γ_{ij} denotes the metric of a maximally symmetric 3-space and $a(t)$ is the scale factor. Throughout these lectures, we will restrict to the special case of flat spatial slices, i.e. $\gamma_{ij} = \delta_{ij}$, and define $\mathrm{d}\mathbf{x}^2 \equiv \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j$. We will also introduce conformal time, $\mathrm{d}\tau = \mathrm{d}t/a(t)$, so that the metric becomes

$$\mathrm{d}s^2 = a^2(\tau)\left(-\mathrm{d}\tau^2 + \mathrm{d}\mathbf{x}^2\right). \tag{31}$$

We will first discuss the kinematics of particles in an FRW spacetime for an arbitrary scale factor $a(\tau)$. After that, we will show how the Einstein equations determine $a(\tau)$ in terms of the matter content of the universe.

Kinematics

Particles in the FRW spacetime evolve according to the geodesic equation

$$P^\nu\nabla_\nu P^\mu = P^\nu\left(\frac{\partial P^\mu}{\partial x^\nu} + \Gamma^\mu_{\nu\rho}P^\rho\right) = 0\,, \tag{32}$$

where $P^\mu \equiv dx^\mu/d\lambda$ is the four-momentum of the particle. In an expanding spacetime, it is convenient to write the components of the four-momentum as

$$P^\mu = a^{-1}[E, \mathbf{p}]\,. \tag{33}$$

For massless particles, such as photons, we have the constraint $g_{\mu\nu}P^\mu P^\nu = -E^2 + |\mathbf{p}|^2 = 0$, so we can write $\mathbf{p} = E\,\hat{\mathbf{p}}$, where $\hat{\mathbf{p}}$ is a unit vector in the direction of propagation.

Exercise.—Show that the non-zero connection coefficients associated with the metric (328) are

$$\Gamma^0_{00} = \mathcal{H}\,, \quad \Gamma^0_{ij} = \mathcal{H}\delta_{ij}\,, \quad \Gamma^i_{j0} = \mathcal{H}\delta^i_j\,, \tag{34}$$

where $\mathcal{H} \equiv \dot{a}/a$ is the conformal Hubble parameter.

The $\mu = 0$ component of the geodesic equation (32) becomes

$$P^0\frac{dP^0}{d\tau} = -\Gamma^0_{\alpha\beta}P^\alpha P^\alpha\,. \tag{35}$$

Using (33) and (34), we get

$$(a^{-1}E)\frac{d}{d\tau}(a^{-1}E) = -\mathcal{H}a^{-2}E^2 - \mathcal{H}a^{-2}|\mathbf{p}|^2\,, \tag{36}$$

which simplifies to

$$\frac{1}{E}\frac{dE}{d\tau} = -\frac{1}{a}\frac{da}{d\tau}\,. \tag{37}$$

This describes the *redshifting* of the photon energy in an expanding spacetime, $E \propto a^{-1}$.

Dynamics

The evolution of the scale factor is determined by the Friedmann equations

$$3\mathcal{H}^2 = 8\pi G a^2 \bar{\rho}\,, \tag{38}$$

$$2\dot{\mathcal{H}} + \mathcal{H}^2 = -8\pi G a^2\, \bar{P}\,, \tag{39}$$

where $\bar{\rho}$ and $\bar{P}$ are the background density and pressure, respectively.

Exercise.—By substituting (34) into

$$R_{\mu\nu} \equiv \partial_\lambda \Gamma^\lambda_{\mu\nu} - \partial_\nu \Gamma^\lambda_{\mu\lambda} + \Gamma^\lambda_{\lambda\rho}\Gamma^\rho_{\mu\nu} - \Gamma^\rho_{\mu\lambda}\Gamma^\lambda_{\nu\rho} \tag{40}$$

show that

$$R_{00} = -3\dot{\mathcal{H}}\,, \quad R_{ij} = (\dot{\mathcal{H}} + 2\mathcal{H}^2)\delta_{ij} \quad \Rightarrow \quad a^2 R \equiv a^2 g^{\mu\nu} R_{\mu\nu} = -6(\dot{\mathcal{H}} + \mathcal{H}^2)\,. \tag{41}$$

Hence, show that the non-zero components of the Einstein tensor, $G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}$, are

$$G_{00} = 3\mathcal{H}^2\,, \quad G_{ij} = -(2\dot{\mathcal{H}} + \mathcal{H}^2)\delta_{ij}\,. \tag{42}$$

Use this to confirm that the 00-Einstein equation, $G_{00} = 8\pi G T_{00}$, implies (A.122) and the ij-Einstein equation, $G_{ij} = 8\pi G T_{ij}$, leads to (39).

Combining (A.122) and (39), we can write an evolution equation for the density

$$\dot{\bar{\rho}} = -3\mathcal{H}(\bar{\rho} + \bar{P})\,. \tag{43}$$

For pressureless matter ($\bar{P}_m \approx 0$) this implies $\bar{\rho}_m \propto a^{-3}$, while for radiation ($\bar{P}_r = \frac{1}{3}\bar{\rho}_r$) we have $\bar{\rho}_r \propto a^{-4}$.

Exercise.—Derive the continuity equation (43) from the conservation of the stress tensor, $\nabla^\mu T_{\mu\nu} = 0$. By integrating the Friedmann equation (A.122) for matter and radiation show that

$$a(\tau) = \begin{cases} \tau^2 & \text{matter} \\ \tau & \text{radiation} \end{cases}\,. \tag{44}$$

Cosmic inventory

The universe is filled with several different species of particles:

$$\underbrace{\text{photons } (\gamma) \quad \text{neutrinos } (\nu)}_{\text{radiation } (r)} \qquad \overbrace{\underbrace{\text{electrons } (e) \quad \text{protons } (p)}_{\text{baryons } (b)} \quad \text{cold dark matter } (c)}^{\text{matter } (m)}\,.$$

The number density, energy density and pressure of each species a can be written as

$$n_a = g_a \int \frac{\mathrm{d}^3 p}{(2\pi)^3}\, f_a(\mathbf{x}, \mathbf{p})\,, \tag{45}$$

$$\rho_a = g_a \int \frac{\mathrm{d}^3 p}{(2\pi)^3}\, f_a(\mathbf{x}, \mathbf{p})\, E(p)\,, \tag{46}$$

$$P_a = g_a \int \frac{\mathrm{d}^3 p}{(2\pi)^3}\, f_a(\mathbf{x}, \mathbf{p})\, \frac{p^2}{3E(p)}\,, \tag{47}$$

where $f_a(\mathbf{x}, \mathbf{p})$ is the (phase space) distribution function of the species a, and g_a is the number of internal degrees of freedom. In the unperturbed universe, the distribution functions should not depend on the position and the direction of the momentum, i.e. $f_a(\mathbf{x}, \mathbf{p}) \to \bar{f}_a(E(p))$.

At early times, particle interactions were efficient enough to keep the different species in local equilibrium. They then shared a common temperature T and the distribution functions take the following maximum entropy form

$$\bar{f}_a(E) = \frac{1}{e^{(E_a - \mu_a)/T} \pm 1}\,, \tag{48}$$

with + for fermions and - for bosons. The chemical potential μ_a vanishes for photons and is (likely) small for all other species. We will henceforth set it to zero. When the temperature drops below the mass of a particle species, $T \ll m_a$, they become non-relativistic and their distribution function receives an exponential (Boltzmann) suppression, $f_a \to e^{-m_a/T}$. This means that relativistic particles ('radiation') dominate the density and pressure of the primordial plasma. By performing the integrals (46) and (47) in the limit $E \to p$, one finds

$$\bar{\rho}_a = \frac{\pi^2}{30}\, g_a T^4 \begin{cases} 1 & \text{bosons} \\ \frac{7}{8} & \text{fermions} \end{cases} \qquad \text{and} \quad \bar{P}_a = \frac{1}{3}\bar{\rho}_a\,. \tag{49}$$

The total radiation density is

$$\bar{\rho}_r = \frac{\pi^2}{30}\, g_* T^4\,, \quad \text{where} \quad g_* \equiv \sum_{a=b} g_a + \frac{7}{8}\sum_{a=f} g_a\,. \tag{50}$$

If equilibrium had persisted until today, all species with masses greater than $10^{-3}\,\text{eV}$ would be exponentially suppressed. This would not be a very interesting world. Fortunately, many massive particle species (e.g. dark matter) are weakly interacting and decoupled from the primordial plasma at early times.

Thermal History

The key to understanding the thermal history of the universe is understanding the competition between the interaction rate of particles, Γ, and the expansion rate of the universe, H. Particles maintain equilibrium as long as $\Gamma \gg H$ and freeze out when $\Gamma \lesssim H$ (see Fig. 23).

Neutrinos are the most weakly interacting particles of the Standard Model and therefore decoupled first (around 0.8 MeV or 1 sec after the Big Bang). Shortly after neutrino decoupling, electrons and positrons annihilated. The energies of the electrons and positrons got transferred to the photons, but not the neutrinos. The temperature of the photons today is therefore slightly larger than that of the neutrinos (see insert below). At around the same time, neutron-proton interactions became inefficient, leading to a relic abundance of neutrons. These neutrons were essential for the formation of the light elements during Big Bang nucleosynthesis (BBN), which occurred around 3 minutes after the Big Bang.

Cosmic neutrino background. After the neutrinos decouple, their momenta redshift, $p_\nu \propto a^{-1}$, and their distribution functions f_ν maintain their shape. The combination of these two facts requires that the neutrino temperature evolves as $T_\nu \propto a^{-1}$. We would like to compare this to the evolution of the photon temperature T_γ, since this is what has been measured by observations of the CMB. We will use the fact that the *comoving entropy* is conserved in equilibrium [1]:

$$s a^3 = \frac{\rho + P}{T} a^3 = \frac{2\pi^2}{45} g_* (aT)^3 = const\,. \tag{51}$$

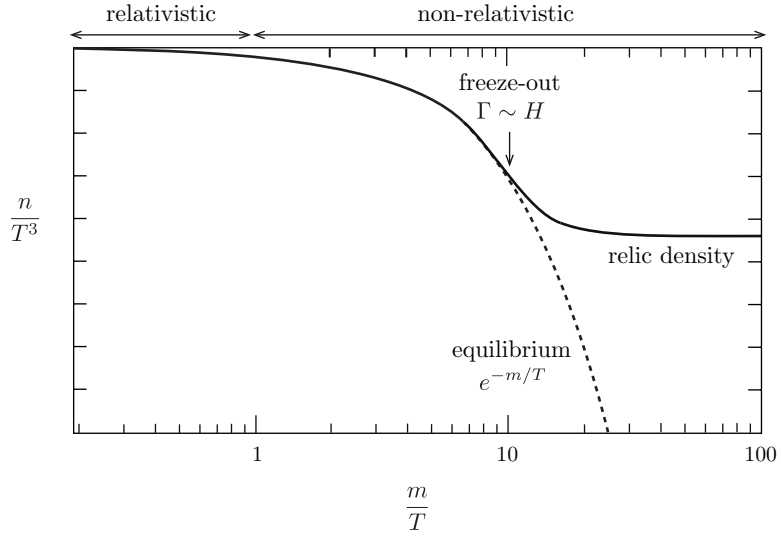


Figure 23: A schematic illustration of particle freeze-out. At high temperatures, the particle abundance tracks its equilibrium value. At low temperatures, the particles freeze out and maintain a relic density that is much larger than the Boltzmann-suppressed equilibrium abundance.

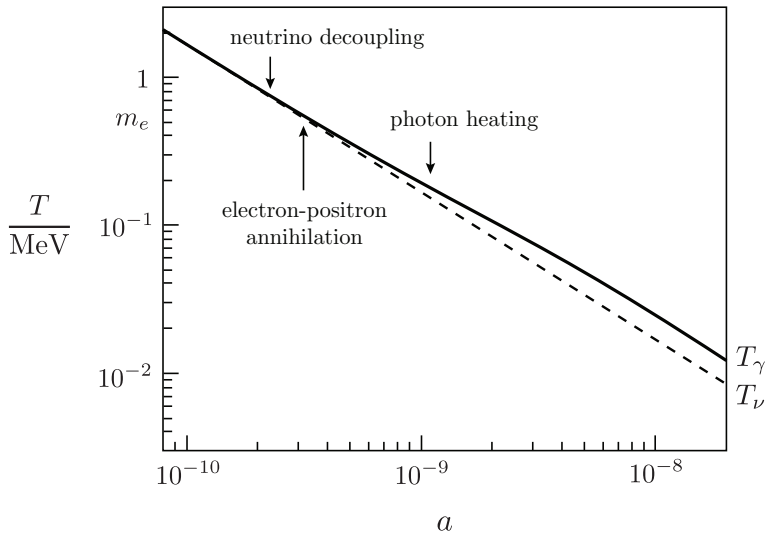


Figure 24: Thermal history through electron-positron annihilation. Neutrinos are decoupled and their temperature redshifts simply as $T_\nu \propto a^{-1}$. The energy density of the electron-positron pairs is transferred to the photon gas whose temperature therefore redshifts more slowly, $T_\gamma \propto g_*^{-1/3} a^{-1}$.

Since entropy is separately conserved for the thermal bath and the decoupled species, we only need to consider the change in the effective number of relativistic degrees of freedom in equilibrium with the photons. Before e^+e^- annihilation, i.e. at $T_+ > m_e$, we have

$$g_*(T_+) = 2 + \frac{7}{8} \times (2 \times 2) = \frac{11}{2}\,, \tag{52}$$

where we have counted photons, electrons and positrons. After e^+e^- annihilation, i.e. at $T_- < m_e$, only the two polarization degrees of freedom of the photons contribute,

$$g_*(T_-) = 2\,. \tag{53}$$

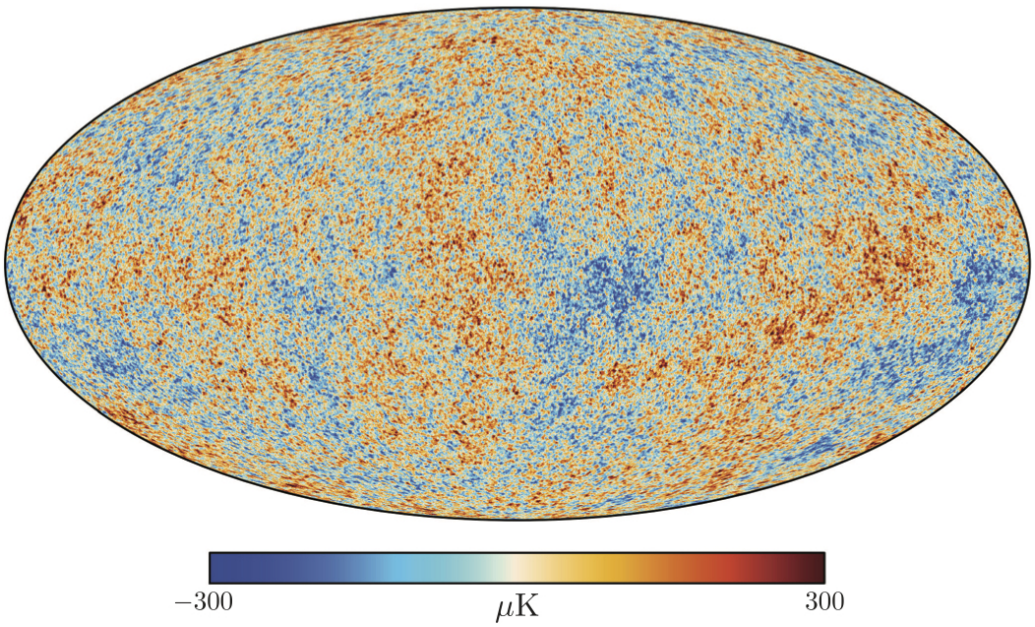


Figure 25: Planck measurement of the temperature variations in the CMB sky.

Comparing $g_*(aT_\gamma)^3 = const.$ to $aT_\nu = const.$, we get

$$T_\nu = \left(\frac{g_*(T_-)}{g_*(T_+)}\right)^{1/3} T_\gamma = \left(\frac{4}{11}\right)^{1/3} T_\gamma \,. \tag{54}$$

Given the measured temperature of the CMB today, $T_{\gamma,0} = 2.7\,\mathrm{K}$, this tells us that the present temperature of the cosmic neutrino background (CνB) is $T_{\nu,0} = 1.9\,\mathrm{K}$.

While the neutrinos are still relativistic (i.e. for most of the history of the universe), each species carries the following energy density

$$\frac{\rho_\nu}{\rho_\gamma} = \frac{\frac{7}{8} \times 2 \times T_\nu^4}{2 \times T_\gamma^4} = \frac{7}{8}\left(\frac{4}{11}\right)^{4/3} \equiv a_\nu \approx 0.227 \,. \tag{55}$$

The $N_\nu = 3$ neutrino species of the Standard Model therefore contribute a significant amount to the total radiation density in the early universe:

$$\frac{\sum \rho_\nu}{\rho_r} = \frac{N_\nu \rho_\nu}{N_\nu \rho_\nu + \rho_\gamma} = \frac{N_\nu a_\nu}{N_\nu a_\nu + 1} \approx 41\% \,. \tag{56}$$

Although the neutrinos are decoupled, their gravitational effects are significant and have recently been observed in the CMB [4, 5].

Below about 1 eV, or 380,000 years after the Big Bang, the temperature became low enough for neutral hydrogen atoms to form through the reaction $e^- + p^+ \to \mathrm{H} + \gamma$. This is the moment of *recombination*. At this point the density of free electrons dropped dramatically. Before recombination the strongest coupling between the photons and the rest of the plasma was through Thomson scattering, $e^- + \gamma \to e^- + \gamma$. The sharp drop in the free electron density after recombination means that this process became inefficient and the photons decoupled. After *decoupling* the photons streamed freely through the universe and are observed today as the CMB.

Structure Formation

The CMB is an almost perfect blackbody with an average temperature of 2.7 K. Small variations in the CMB temperature across the sky (see Fig. 25), reflect spatial variations in the density of the primordial plasma, $\delta\rho_a \equiv \rho_a - \bar{\rho}_a$ (see Section 7.5.2), and related perturbations of the spacetime geometry, $\delta g_{\mu\nu} \equiv g_{\mu\nu} - \bar{g}_{\mu\nu}$. During the radiation-dominated phase of the early universe, the growth of matter perturbations was inhibited by the large pressure provided by the radiation. Perturbations in the coupled photon and baryon fluids were oscillating with constant amplitude (see Section 6.5.5). Shortly before recombination, however, the universe became matter dominated and the radiation pressure disappeared, so that density fluctuations could start to grow under the influence of gravity. This growth of the matter perturbations eventually led to the large-scale structure (LSS) of the universe.

Initial Conditions

At sufficiently early times, all scales of interest to current observations were outside the Hubble radius, $k < \mathcal{H}$. On super-Hubble scales, the evolution of perturbations becomes very simple, especially for adiabatic initial conditions.

Adiabatic perturbations

Adiabatic perturbations have the property that the local state of matter (determined, for example, by the energy density ρ and the pressure P) at some spacetime point $(\tau, \mathbf{x})$ of the perturbed universe is the same as in the *background* universe at some slightly different time $\tau + \delta\tau(\mathbf{x})$. (Notice that the time shift varies with position $\mathbf{x}$.) If the universe is filled with multiple fluids, adiabatic perturbations correspond to perturbations induced by a *common, local shift in time* of all background quantities; e.g. adiabatic density perturbations are defined as

$$\delta\rho_a(\tau, \mathbf{x}) \equiv \bar{\rho}_a(\tau + \delta\tau(\mathbf{x})) - \bar{\rho}_a(\tau) \approx \dot{\bar{\rho}}_a \delta\tau(\mathbf{x}) \,, \tag{57}$$

where $\delta\tau$ is the same for all species a. This implies that all matter perturbations can be characterized by a single degree of freedom. It also means that we can perform a local time reparameterization to set all matter perturbations to zero, e.g. $\delta\rho_a \equiv 0$. In that gauge, the information about fluctuations is carried by the following perturbation of the metric

$$g_{ij}(\tau, \mathbf{x}) = a^2(\tau)\, e^{2\zeta(\tau,\mathbf{x})}\, \delta_{ij} \,, \tag{58}$$

where ζ is called the *curvature perturbation*. An attractive property of the curvature perturbation is that it is conserved on super-Hubble scales.

Statistics

The initial conditions for the hot Big Bang are believed to have been created by quantum fluctuations during a period of inflationary expansion [6]. As we will see in Section 6.6.2, this mechanism predicts the statistics of the initial conditions, i.e. it predicts the correlations between the CMB fluctuations in different directions in the sky, rather than the specific value of the temperature fluctuation in a specific direction. For Gaussian initial conditions, these correlations are completely specified by the two-point correlation function

$$\langle \zeta(\mathbf{x})\zeta(\mathbf{x}')\rangle \equiv \xi_\zeta(\mathbf{x}, \mathbf{x}') = \xi_\zeta(|\mathbf{x}' - \mathbf{x}|) \,, \tag{59}$$

where the last equality holds as a consequence of statistical homogeneity and isotropy. The Fourier transform of ζ then satisfies

$$\langle \zeta(\mathbf{k})\zeta^*(\mathbf{k}')\rangle = \frac{2\pi^2}{k^3}\mathcal{P}_\zeta(k)\, \delta_D(\mathbf{k} - \mathbf{k}') \,, \tag{60}$$

where $\mathcal{P}_\zeta(k)$ is the (dimensionless) *power spectrum*.

Exercise.—Show that

$$\xi_\zeta(\mathbf{x}, \mathbf{x}') = \int \frac{\mathrm{d}k}{k} \mathcal{P}_\zeta(k) \,\mathrm{sinc}(k|\mathbf{x} - \mathbf{x}'|)\,. \tag{61}$$

In Section 6.6.2, we will explicitly compute the form of $\mathcal{P}_\zeta(k)$ predicted by inflation. However, before we do that, we will show, in Sections 7.5.2 and 6.5.5, how generic scale-invariant initial conditions, $\mathcal{P}_\zeta(k) \approx const.$, evolve into the anisotropies of the cosmic microwave background. We will first do this in the context of the Standard Model of particle physics, before asking, in Section 6.5.6, what kind of deviations can arise in theories beyond the SM.

6.5.4 Afterglow of the Big Bang

Observations of the temperature fluctuations in the cosmic microwave background have played a pivotal role in establishing the standard cosmological model. We now have a detailed understanding of the geometry and composition of the universe, and there is growing evidence that the primordial fluctuations originated from quantum fluctuations during a period of inflation. In this section and the next, we will give a simplified analytical treatment of the physics of the CMB anisotropies. Our goal is to present just enough details to be able to explain how the CMB can be used as a probe of BSM physics.

CMB Anisotropies

The first thing one sees when looking at the microwave scy is the motion of the Solar System with respect to the rest frame of the CMB (cf. Fig. 26).

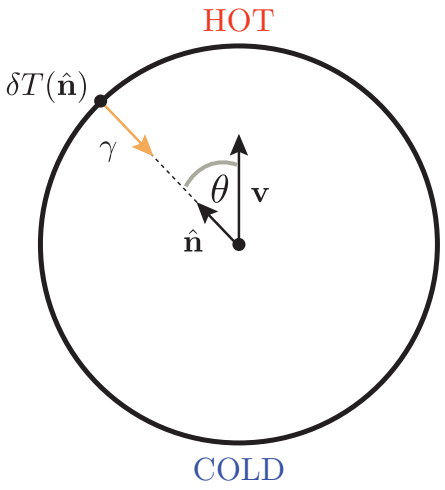


Figure 26: The motion of the Solar System relative to the CMB rest frame produces a dipolar pattern in the observed CMB temperature.

Consider a photon entering our detectors from a direction $\hat{\mathbf{n}}$. In the rest frame of the CMB, it has momentum $\mathbf{p} = -p\hat{\mathbf{n}}$. Due to the Doppler effect, the observed momentum is

$$p_0(\hat{\mathbf{n}}) = \frac{p}{\gamma(1 - \hat{\mathbf{n}} \cdot \mathbf{v})} \approx p\,(1 + \hat{\mathbf{n}} \cdot \mathbf{v})\,, \tag{62}$$

where $\mathbf{v}$ is our velocity relative to the CMB rest frame, p is the momentum of the photon in the CMB rest frame and $\gamma = (1 - v^2)^{-1/2}$ is the Lorentz factor. We have also shown an

approximation at leading order in $|\mathbf{v}| \ll 1$. As expected, the momentum is higher if we move towards the photon ($\hat{\mathbf{n}} \cdot \mathbf{v} = v$) and smaller if we move away from it ($\hat{\mathbf{n}} \cdot \mathbf{v} = -v$). Since the CMB has a blackbody spectrum, we can relate the change in the observed momentum of photons to a change in the observed temperature:

$$\frac{\delta T(\hat{\mathbf{n}})}{T} \equiv \frac{T_0(\hat{\mathbf{n}}) - T}{T} = \frac{p_0(\hat{\mathbf{n}}) - p}{p} = \hat{\mathbf{n}} \cdot \mathbf{v} = v \cos\theta \,. \tag{63}$$

Fitting this dipolar anisotropy to the data, we find that the speed of the Solar System relative to the CMB is

$$v = 368\,\mathrm{km/s}. \tag{64}$$

After subtracting the dipole, we are left with the *primordial anisotropy*.

Perturbed photon geodesics

Let us trace the life of a photon after decoupling. Its (physical) momentum will redshift due to the expansion of the universe. In addition, the momentum will change in response to the inhomogeneities of the universe. We will study these effects by solving the geodesic equation (32) in the perturbed spacetime.

We will treat perturbations in the metric in (conformal) Newtonian gauge

$$\mathrm{d}s^2 = a^2(\tau)\big[-(1+2\Phi)\mathrm{d}\tau^2 + (1-2\Psi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\big]\,, \tag{65}$$

where the perturbation $\Phi \approx \Psi$ plays the role of the gravitational potential.

Exercise.—Show that the connection coefficients associated with the metric (65) are

$$\begin{aligned}
\Gamma^0_{00} &= \mathcal{H} + \dot{\Phi}\,,\\
\Gamma^0_{i0} &= \partial_i\Phi\,,\\
\Gamma^i_{00} &= \delta^{ij}\partial_j\Phi\,,\\
\Gamma^0_{ij} &= \mathcal{H}\delta_{ij} - \big[\dot{\Psi} + 2\mathcal{H}(\Phi+\Psi)\big]\delta_{ij}\,,\\
\Gamma^i_{j0} &= \big[\mathcal{H} - \dot{\Psi}\big]\delta^i_j\,,\\
\Gamma^i_{jk} &= -2\delta^i_{(j}\partial_{k)}\Psi + \delta_{jk}\delta^{il}\partial_l\Psi\,.
\end{aligned} \tag{66}$$

As we will show in the following insert, the geodesic equation then leads to the following evolution equation for the photon momentum:

$$\boxed{\frac{d}{d\tau}\ln(ap) = -\frac{d\Phi}{d\tau} + \frac{\partial(\Phi+\Psi)}{\partial\tau}}\,. \tag{67}$$

In the absence of the source terms on the right-hand side, this implies $p \propto a^{-1}$, which is the expected redshifting due to the expansion of the universe, cf. eq. (37). The inhomogeneous source terms describe how photons lose (gain) energy as they move out of (into) potential wells.

,

Derivation.—We will derive eq. (67) from the geodesic equation for photons,

$$\frac{dP^0}{d\lambda} = -\Gamma^0_{\alpha\beta} P^\alpha P^\beta \,, \tag{68}$$

where $P^\mu = dx^\mu/d\lambda$ is the four-momentum of the photons. We need expressions for the components of the four-momentum in the presence of metric perturbations.

- We first consider the P^0 component. Since photons are massless we have

$$\begin{aligned} P^2 = g_{\mu\nu} P^\mu P^\nu &= 0 \\ &= -a^2(1+2\Phi)(P^0)^2 + p^2 \,, \end{aligned} \tag{69}$$

  where we have substituted the metric (65) and defined $p^2 \equiv g_{ij} P^i P^j$. Solving (69) for P^0, we find

$$P^0 = \frac{p}{a}(1-\Phi) \,. \tag{70}$$

- We then write the spatial component of the four-momentum as

$$P^i \equiv \alpha\, \hat{p}^i \,. \tag{71}$$

  To determine the constant of proportionality α, we use

$$\begin{aligned} p^2 = g_{ij} P^i P^j &= a^2(1-2\Psi)\delta_{ij}\hat{p}^i\hat{p}^j\alpha^2 \\ &= a^2(1-2\Psi)\alpha^2 \,, \end{aligned} \tag{72}$$

  where the last equality holds because the direction vector is a unit vector. Solving (72) for α, we get $\alpha = p(1+\Psi)/a$, or

$$P^i = \frac{p\hat{p}^i}{a}(1+\Psi) \,. \tag{73}$$

Substituting these results into the geodesic equation (68), we get

$$\frac{p}{a}(1-\Phi)\frac{d}{d\tau}\left[\frac{p}{a}(1-\Phi)\right] = -\Gamma^0_{\alpha\beta} P^\alpha P^\beta \,, \tag{74}$$

where we have used the standard trick of rewriting the derivative with respect to λ as a derivative with respect to time multiplied by $d\tau/d\lambda = P^0$. We expand out the time derivative to get

$$\frac{dp}{d\tau}(1-\Phi) = \mathcal{H} p(1-\Phi) + p\frac{d\Phi}{d\tau} - \Gamma^0_{\alpha\beta} P^\alpha P^\beta \frac{a^2}{p}(1+\Phi) \,. \tag{75}$$

Multiplying both sides by $(1+\Phi)/p$ and dropping all quadratic terms in Φ, we find

$$\frac{1}{p}\frac{dp}{d\tau} = \mathcal{H} + \frac{d\Phi}{d\tau} - \Gamma^0_{\alpha\beta} P^\alpha P^\beta \frac{a^2}{p^2}(1+2\Phi) \,. \tag{76}$$

To evaluate the last term on the right-hand side, we use the perturbed Christoffel symbols (6.5.4). After a bit of algebra, we get

$$-\Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p^2}(1+2\Phi) = -2\mathcal{H} + \frac{\partial\Psi}{\partial\tau} - \frac{\partial\Phi}{\partial\tau} - 2\hat{p}^i\frac{\partial\Phi}{\partial x^i} \,. \tag{77}$$

Equation (76) then becomes

$$\frac{1}{p}\frac{dp}{dt} \tag{78}$$

$$= -\mathcal{H} + \frac{d\Phi}{d\tau} - 2\left(\frac{\partial\Phi}{\partial\tau} + \hat{p}^i\frac{\partial\Phi}{\partial x^i}\right) + \frac{\partial(\Phi+\Psi)}{\partial\tau} \,. \tag{79}$$

At leading order, the term in brackets is equal to the total time derivative of Φ:

$$\frac{d\Phi}{d\tau} = \frac{\partial\Phi}{\partial\tau} + \frac{dx^i}{d\tau}\frac{\partial\Phi}{\partial x^i} = \frac{\partial\Phi}{\partial\tau} + \hat{p}^i\frac{\partial\Phi}{\partial x^i}\,, \tag{80}$$

where we used

$$\frac{dx^i}{d\tau} = \frac{dx^i}{d\lambda}\frac{d\lambda}{d\tau} = \frac{P^i}{P^0} = \hat{p}^i(1+\Psi+\Phi) = \hat{p}^i + \mathcal{O}(1)\,. \tag{81}$$

Substituting (80) into (79), we get

$$\frac{1}{p}\frac{dp}{d\tau} = -\frac{1}{a}\frac{da}{d\tau} - \frac{d\Phi}{d\tau} + \frac{\partial(\Phi+\Psi)}{\partial\tau}\,, \tag{82}$$

which confirms the result (67).

Line-of-sight solution

By integrating the geodesic equation (67) along a line-of-sight, we can relate the observed CMB temperature anisotropies to the fluctuations at recombination. To simplify matters, we will therefore work with the idealised approximation of *instantaneous recombination*. The CMB photons were then emitted at a fixed time τ_*. This moment is often called *last scattering*. Integrating (67) from the time of emission τ_* to the time of observation τ_0, we then get

$$\ln(ap)_0 = \ln(ap)_* + (\Phi_* - \Phi_0) + \int_{\tau_*}^{\tau_0} \mathrm{d}\tau\,\frac{\partial}{\partial\tau}(\Phi+\Psi)\,. \tag{83}$$

To relate this to the temperature anisotropy, we note that

$$ap \propto a\bar{T}\left(1 + \frac{\delta T}{\bar{T}}\right), \tag{84}$$

where $\bar{T}(\tau)$ is the mean temperature. Taylor-expanding the logarithms in (83) to first order in $\delta T/\bar{T}$, and keeping in mind that $a_0\bar{T}_0 = a_*\bar{T}_*$, we find

$$\left.\frac{\delta T}{\bar{T}}\right|_0 = \left.\frac{\delta T}{\bar{T}}\right|_* + (\Phi_* - \Phi_0) + \int_{\tau_*}^{\tau_0} \mathrm{d}\tau\,\frac{\partial}{\partial\tau}(\Phi+\Psi)\,. \tag{85}$$

The term Φ_0 only affects the monopole perturbation, so it is unobservable and therefore usually dropped from the equation. The fractional temperature perturbation at last-scattering can be expressed in terms of the density contrast of photons, $\delta_\gamma \equiv \delta\rho_\gamma/\rho_\gamma$, as

$$\left.\frac{\delta T}{\bar{T}}\right|_* = \frac{1}{4}(\delta_\gamma)_*\,, \tag{86}$$

where the factor of $\frac{1}{4}$ arises because $\rho_\gamma \propto T^4$. Equation (85) then reads

$$\left.\frac{\delta T}{\bar{T}}\right|_0 = \left(\frac{1}{4}\delta_\gamma + \Phi\right)_* + \int_{\tau_*}^{\tau_0} \mathrm{d}\tau\,\frac{\partial}{\partial\tau}(\Phi+\Psi)\,. \tag{87}$$

Each term on the right-hand side has a simple physical interpretation:

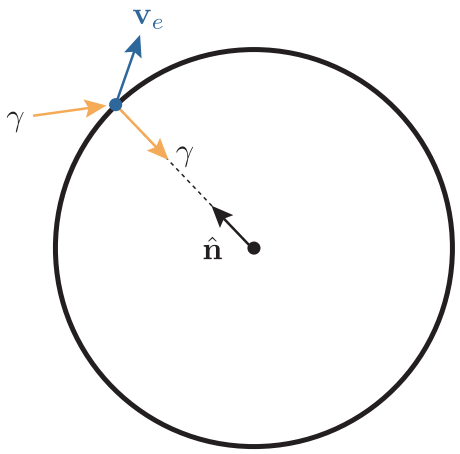


Figure 27: The motion of electrons at the surface of last-scattering produces an additional temperature anisotropy.

- The term $\frac{1}{4}\delta_\gamma$ can be thought of as the intrinsic temperature variation over the background last-scattering surface.
- The term Φ arises from the gravitational redshift that the photons experience when climbing out of a potential well at last-scattering. The combination $\frac{1}{4}\delta_\gamma + \Phi$ is often called the *Sachs-Wolfe* (SW) term.
- Finally, the *integrated Sachs-Wolfe* (ISW) term describes the effect of gravitational redshifting from evolution of the potentials along the line-of-sight. During matter domination, $\dot{\Phi} \approx \dot{\Psi} = 0$ and this term vanishes.

So far, we have ignored the motion of the electrons at the surface of last-scattering. Including this effect leads to an extra Doppler shift in the received energy of photons when referenced to an observer comoving with the electrons at last-scattering (see Fig. 27),

$$\left.\frac{\delta T}{\bar{T}}\right|_0 \subset (\hat{\mathbf{n}} \cdot \mathbf{v}_e)_* \,. \tag{88}$$

Putting everything together, we obtain the following important result

$$\boxed{\frac{\delta T}{\bar{T}}(\hat{\mathbf{n}}) = \left(\frac{1}{4}\delta_\gamma + \Phi + \hat{\mathbf{n}} \cdot \mathbf{v}_e\right)_* + \int_{\tau_*}^{\tau_0} \mathrm{d}\tau \, (\dot{\Phi} + \dot{\Psi})} \tag{89}$$

where we have dropped the subscript '0' on the observed $\delta T/\bar{T}$ to avoid clutter. Figure 28 illustrates the contributions that each of the terms in (89) makes to the power spectrum of the CMB temperature anisotropies (see §6.5.4). We see that the ISW contribution is subdominant and that the shape of the power spectrum is mostly determined by the Sachs-Wolfe and Doppler contributions.

Large scales.—For adiabatic initial conditions, the superhorizon initial condition is $\delta_\gamma \approx \frac{4}{3}\delta_m \approx -\frac{8}{3}\Psi$. The Sachs-Wolfe term then becomes

$$\frac{1}{4}\delta_\gamma + \Phi = -\frac{2\Psi}{3} + \Phi \approx \frac{1}{3}\Phi \,. \tag{90}$$

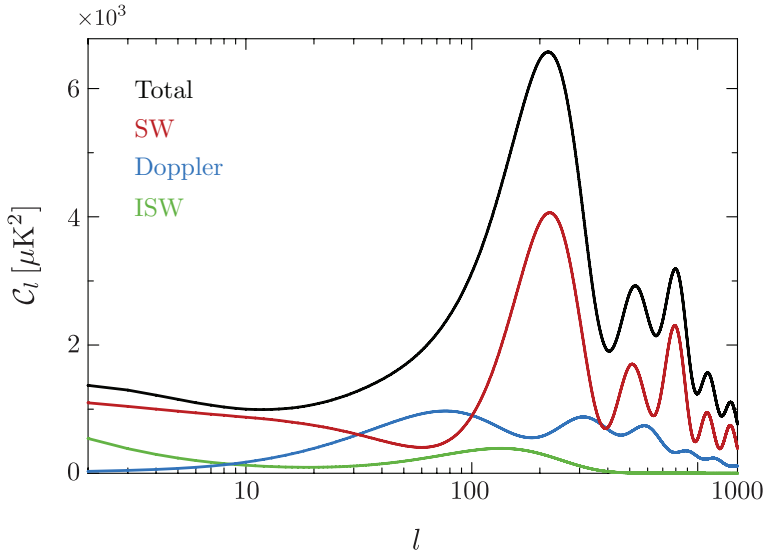


Figure 28: Contributions of the various terms in (89) to the (rescaled) power spectrum of CMB anisotropies, $\mathcal{C}_l \equiv l(l+1)C_l$.

This shows that, on large scales, an overdense region ($\Psi \approx \Phi < 0$) appears as a cold spot in the sky. While the temperature at the bottom of the potential well is hotter than the average ($-\frac{2}{3}\Psi$), photons lose more energy (Φ) as they climb out of the potential well, resulting in a cold spot ($\frac{1}{3}\Phi < 0$).

CMB Power Spectrum

A map of the cosmic microwave background radiation describes the variation of the CMB temperature as a function of direction, $\delta T(\hat{\mathbf{n}})$. We will be interested in the statistical correlations between temperature fluctuations in two different directions $\hat{\mathbf{n}}$ and $\hat{\mathbf{n}}'$ (see Fig. 29), averaged over the entire sky.

If the initial conditions are statistically isotropic, then we expect these correlations only to depend on the relative orientation of $\hat{\mathbf{n}}$ and $\hat{\mathbf{n}}'$. In that case, we can write the two-point correlation function as

$$\left\langle \delta T(\hat{\mathbf{n}})\, \delta T(\hat{\mathbf{n}}') \right\rangle = \sum_l \frac{2l+1}{4\pi} C_l P_l(\cos\theta)\,, \tag{91}$$

where $\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}' \equiv \cos\theta$ and P_l are Legendre polynomials. The expansion coefficients C_l are the *angular power spectrum*. If the fluctuations are Gaussian, then the power spectrum contains the entire information of the CMB map.

The right panel of Fig. 29 illustrates the temperature variations created by a single plane wave inhomogeneity. The CMB anisotropies observed on the sky are a *superposition* of many such plane waves with amplitudes that are weighted by the spectrum of primordial curvature perturbations $\mathcal{P}_\zeta(k)$. In Section 6.6.2, we will show that the initial conditions of the primordial perturbations are expected to be featureless, $\mathcal{P}_\zeta(k) \approx const.$ The observed features in the CMB anisotropy spectrum arise from the subhorizon evolution of perturbations in the photon density and the metric. This evolution takes the form of cosmic sound waves (see Section 6.5.5). These waves are *captured* at recombination and *projected* onto the sky. The observed oscillations in the CMB power spectrum are therefore a snapshot of primordial sound waves caught at different phases in their evolution at the time when photons last scattered off electrons. The beautiful physics of the CMB fluctuations is described in detail in my *Advanced Cosmology* course [7].

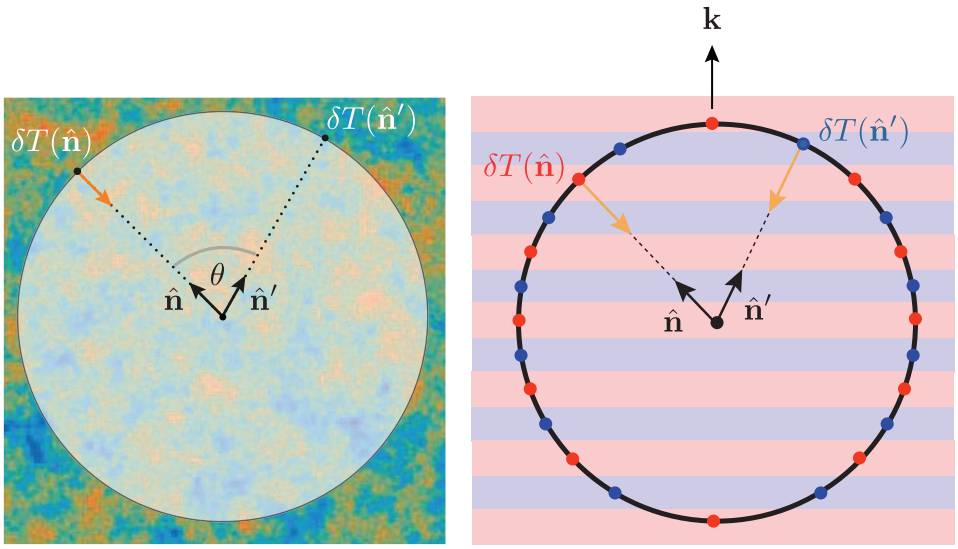


Figure 29: *Left*: Illustration of the two-point correlation function of the temperature anisotropy $\delta T(\hat{\mathbf{n}})$. *Right*: Illustration of the temperature anisotropy created by a single plane wave inhomogeneity are recombination.

Substituting the line-of sight solution (89) into the definition of the angular power spectrum (91), we find (see insert below)

$$\boxed{C_l = \frac{4\pi}{(2l+1)^2}\int \mathrm{d}\ln k\; T_l^2(k)\,\mathcal{P}_\zeta(k)}\,. \tag{92}$$

The *transfer function* $T_l(k)$ captures both the *evolution* of the initial fluctuations until recombination and the *projection* onto the surface of last-scattering. Ignoring the subdominant ISW contribution, we can write the transfer function as follows:

$$\boxed{T_l(k) = T_{\rm SW}(k)\,j_l(kr_*) + T_{\rm D}(k)\,j_l'(kr_*)}\,, \quad \text{where} \quad \begin{aligned} T_{\rm SW}(k) &\equiv \frac{(\frac{1}{4}\delta_\gamma + \Phi)_*}{\zeta(\mathbf{k})}\,, \\ T_{\rm D}(k) &\equiv -\frac{(v_e)_*}{\zeta(\mathbf{k})}\,. \end{aligned} \tag{93}$$

The subscript $*$ denotes quantities evaluated at recombination, with r_* being the distance to last-scattering. The Bessel function $j_l(kr_*)$ and its derivative $j_l'(kr_*)$ act almost like delta functions and map the Fourier modes k to the harmonic moments $l \sim kr_*$. Given that $\mathcal{P}_\zeta(k)$ is expected to be nearly constant, the angular power spectrum C_l therefore measures the square of the transfer function $T_l(k)$ evaluated at $k = l/r_*$:

$$C_l \;\sim\; \frac{4\pi}{(2l+1)^2}\Big[T_{\rm SW}^2(k) + T_{\rm D}^2(k)\Big]\Big|_{k\sim l/r_*}\,, \tag{94}$$

where we have dropped the cross term $T_{\rm SW}(k)\,T_{\rm D}(k)$ because it is negligible. In the next section, we will discuss the evolution effects that determine the transfer function and hence the CMB power spectrum.

Projection.—To understand the origin of the Bessel functions in (93), let us consider the projection of the Sachs-Wolfe term onto the surface of last-scattering. Assuming instantaneous recombination,

we can write

$$\delta T(\hat{\mathbf{n}}) = \int \mathrm{d}r\, \delta T(\mathbf{x}, \mathbf{n})\, \delta_D(r - r_*) \tag{95}$$

$$= \int \frac{\mathrm{d}^3 k}{(2\pi)^3}\, e^{i(kr_*)\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}}\, \delta T(\mathbf{k})\,, \tag{96}$$

where we substituted the Fourier expansion of the temperature field in the second line. The exponential in (96) can be written in a Rayleigh plane wave expansion,

$$e^{i(kr_*)\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}} = \sum_l (-i)^l (2l+1) j_l(kr_*) P_l(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})\,. \tag{97}$$

The two-point function of temperature anisotropies then becomes

$$\langle \delta T(\hat{\mathbf{n}})\, \delta T(\hat{\mathbf{n}}')\rangle = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \int \frac{\mathrm{d}^3 k'}{(2\pi)^3} \sum_l \sum_{l'} (-i)^{l+l'} (2l+1)(2l'+1)\, j_l(kr_*) j_{l'}(k'r_*) \times P_l(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}) P_{l'}(\hat{\mathbf{k}}'\cdot\hat{\mathbf{n}}')\, \langle \delta T(\mathbf{k})\, \delta T(\mathbf{k}')\rangle\,. \tag{98}$$

The power spectrum of the temperature field can be written in terms of the power spectrum of the primordial curvature perturbations, cf. (60),

$$\langle \delta T(\mathbf{k})\, \delta T(\mathbf{k}')\rangle = T_{\rm SW}(k) T_{\rm SW}(k') \langle \zeta(\mathbf{k})\, \zeta(\mathbf{k}')\rangle$$

$$= T^2_{\rm SW}(k)\, \frac{2\pi^2}{k^3}\, \mathcal{P}_\zeta(k)\, \delta_D(\mathbf{k} - \mathbf{k}')\,. \tag{99}$$

The delta function allows us to trivially perform one of the momentum integrals in (98). To evaluate the angular part of the second momentum integral, we use the following identity

$$\int \mathrm{d}^2\hat{\mathbf{k}}\, P_l(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}) P_{l'}(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}') = \frac{4\pi}{2l+1} P_l(\hat{\mathbf{n}}\cdot\hat{\mathbf{n}}')\, \delta_{ll'}\,. \tag{100}$$

The two-point function then takes the form (91) with the angular power spectrum given by

$$C_l^{\rm SW} = \frac{4\pi}{(2l+1)^2} \int \mathrm{d}\ln k\; \underset{\text{projection}}{\underset{\uparrow}{j_l^2(kr_*)}}\; \underset{\text{evolution}}{\underset{\uparrow}{T^2_{\rm SW}(k)}}\; \underset{\text{initial conditions}}{\underset{\uparrow}{\mathcal{P}_\zeta(k)}}\,. \tag{101}$$

A similar derivation gives the form of the Doppler contribution.

6.5.5 Cosmic Sound Waves

In the early universe, photons and electrons were strongly interacting, while the electrons were strongly coupled to protons. The combined system is often called the *photon-baryon fluid*. In this section, we will study the evolution of sound waves in this medium. These waves will evolve in an inhomogeneous spacetime whose perturbations are sourced by all forms of matter in the universe (see Fig. 30).

Our treatment in this section will be rather telegraphic and is just meant to give a flavor for the beautiful physics underlying the CMB. Further details can be found in the notes of my *Advanced Cosmology* course [7], or in the following textbooks [2, 8] and review articles [9, 921, 11, 12].

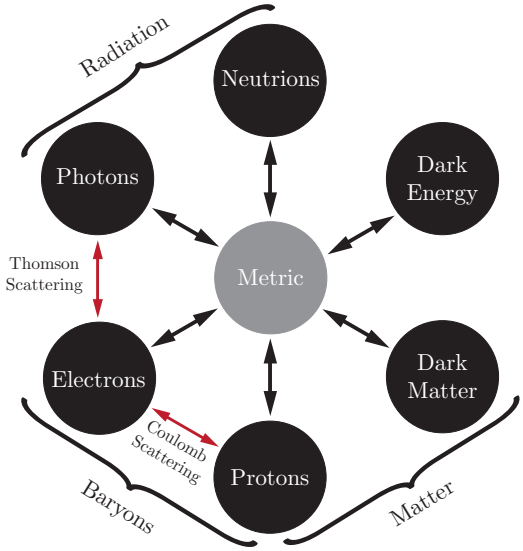


Figure 30: Interactions between the different forms of matter in the universe.

Photon-Baryon Fluid

Combining the continuity and Euler equations for the photon-baryon fluid leads to an evolution equation for the photon density perturbations [7]:

$$\boxed{\ddot{\delta}_\gamma + \frac{\mathcal{H}R}{1+R}\,\dot{\delta}_\gamma - c_s^2\nabla^2\delta_\gamma \;=\; \frac{4}{3}\nabla^2\Phi + 4\ddot{\Psi} + \frac{4\mathcal{H}R}{1+R}\,\dot{\Psi}}$$

$$\begin{array}{cccc} \uparrow & \uparrow & \uparrow & \uparrow \\ \text{friction} & \text{pressure} & \text{gravity} & \text{time dilation} \end{array} \tag{102}$$

where $R \equiv \frac{3}{4}\bar{\rho}_b/\bar{\rho}_\gamma$ is the ratio of the momentum densities of baryons and photons, and the sound speed of the photon-baryon fluid is defined as

$$c_s^2 \equiv \frac{1}{3(1+R)}\,. \tag{103}$$

Equation (B.4) is the master equation describing the entire CMB phenomenology. The most important terms in the equation are the photon pressure term on the left-hand side and the gravitational forcing term on the right-hand side. In addition, we have a friction term proportional to the baryon density R on the left-hand side and two terms related to time dilation effects on the right-hand side. The metric potentials Φ and Ψ are determined by the Einstein equations (which include important contributions from dark matter).

In practice, the equations describing the many coupled fluctuations in the primordial plasma have to be solved numerically. To gain some intuition, however, it is nevertheless useful to obtain approximate analytic results. In the following, we will solve equation (B.4) by making several (more or less justified) approximations. Our goal is to understand the main features of the CMB power spectrum shown in Fig. 31.

Acoustic Oscillations

During radiation domination, the baryon density is subdominant, $R \ll 1$, so for the moment we will set $R = 0$. For simplicity, we will also ignore the time dilation terms in (B.4); we will include their effects in §6.5.6. Equation (B.4) can then be written as

$$\ddot{\Theta} - c_s^2\nabla^2\Theta \;=\; 0\,, \tag{104}$$

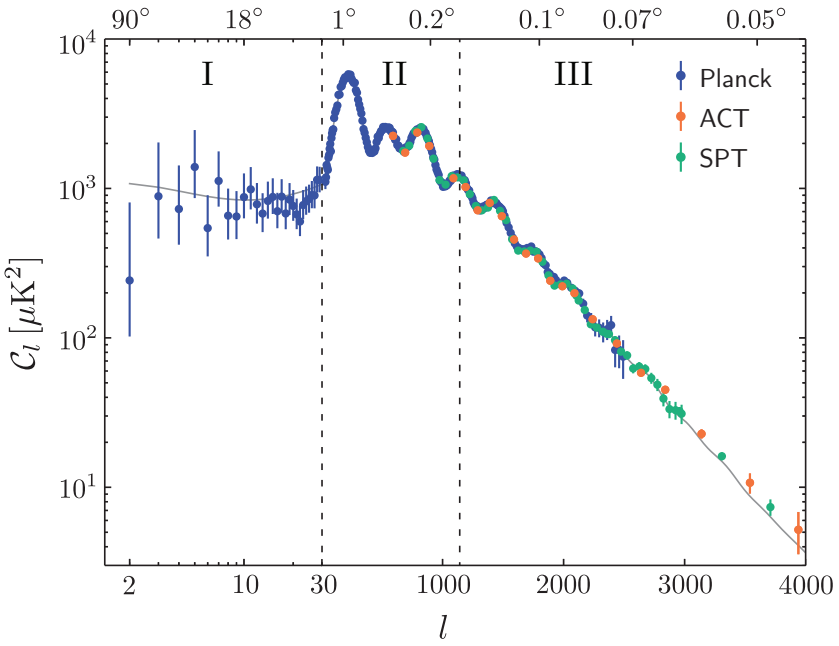


Figure 31: The angular variations of the CMB power spectrum are consequence of the dynamics of sound waves in the photon-baryon fluid. On large scales (region I), the fluctuations are frozen and we directly see the spectrum of the initial conditions. At intermediate scales (region II), we observe the oscillations of the fluid as captured at the moment of last-scattering. Finally, on small scales (region III), fluctuations are damped because their wavelengths are smaller than the mean free path of the photons.

where $c_s^2 \approx \frac{1}{3}$ and $\Theta \equiv \frac{1}{4}\delta_\gamma + \Phi$ is precisely the combination of the Sachs-Wolfe term appearing in (89). Solving (104) for a single Fourier mode, we get

$$\Theta(\mathbf{k},\tau) \;=\; A_{\mathbf{k}}\cos(c_s k\tau) + B_{\mathbf{k}}\sin(c_s k\tau)\,, \tag{105}$$

where $A_{\mathbf{k}}$ and $B_{\mathbf{k}}$ are parameters that are fixed by the initial conditions. For adiabatic initial conditions, all perturbations in the limit $\tau \to 0$ are analytic functions of k^2, which is only the case for the cosine part of the solution above. We therefore set $B_{\mathbf{k}} = 0$. Moreover, the matching to the superhorizon initial conditions implies $A_{\mathbf{k}} = 3\zeta_{\mathbf{k}}$ [7] and we get

$$\Theta(\mathbf{k},\tau) \;=\; 3\zeta_{\mathbf{k}}\cos(c_s k\tau)\,. \tag{106}$$

Evaluating this solution at recombination, $\tau = \tau_*$, gives the Sachs-Wolfe transfer function, $T_{\rm SW}(k) \equiv \Theta(\mathbf{k},\tau_*)/\zeta_{\mathbf{k}} = 3\cos(c_s k\tau_*)$. As we have seen in §6.5.4, the CMB power spectrum is roughly given by the square of the transfer function. The oscillatory k-dependence of the transfer function then maps to the observed oscillations in the CMB power spectrum in harmonic space, cf. eq. (94).

Sound horizon

Modes caught at extrema of their oscillations will have enhanced fluctuations

$$k_n = n\pi/s_*\,, \tag{107}$$

where $s_* \equiv c_s\tau_* \approx \frac{1}{\sqrt{3}}\tau_*$ is the sound horizon at recombination. We see that the peaks occur at multiples of the fundamental scale $k_* \equiv \pi/s_* \approx \sqrt{3}\pi/\tau_*$. This scale becomes a characteristic

angular scale by simple projection

$$\theta_* = \frac{\lambda_*}{D_A}\,, \tag{108}$$

$$, \tag{109}$$

$$l_* = k_* D_A \approx \frac{\tau_*}{\tau_0}\,, \tag{110}$$

where D_A is the angular diameter distance (which in a flat universe is $D_A = \tau_0 - \tau_* \approx \tau_0$). Assuming a purely matter-dominated universe after recombination, we have $\tau \propto a^{1/2}$ and therefore find

$$\theta_* \approx \left(\frac{1}{1100}\right)^{1/2} \approx 2^\circ\,, \tag{111}$$

$$l_* \approx 200\,. \tag{112}$$

The presence of dark energy and spatial curvature would slightly change this result. Measurements on the CMB spectrum have now determined θ_* to better than 0.05%, which puts strong constraints on the geometry and composition of the universe. Keeping the physical densities of dark matter, baryons, photons and neutrinos fixed, i.e. $\Omega_i h^2 = const.$, the scale θ_* is a measure of the curvature parameter Ω_k through its effect on the angular diameter distance to last-scattering. The observed value of θ_* is found to be consistent with a flat universe. Allowing the matter density and the Hubble parameter to vary, one finds $\theta_* \approx f(\Omega_m h^3)$, i.e. there is a specific degeneracy between variations in Ω_m and h. This degeneracy is broken by measurements of the peak morphology of the CMB spectrum and by external data sets (BAO, supernovae, etc.).

Baryon loading

Let me briefly comment on the effects of baryons on the CMB spectrum. The baryon-to-photon ratio increases with time, $R \propto a$, reaching an order one value at recombination. This decreases the sound speed, cf. (103). Instead of (104), we now have

$$\frac{d^2}{d\tau^2}(\Theta + R\Phi) - \frac{1}{3}\nabla^2(\Theta + R\Phi) = 0\,, \tag{113}$$

where we have ignored the time variation of R relative to the much faster evolution of the acoustic oscillations. We see that the finite baryon density, $R \neq 0$, changes the equilibrium point of the oscillations from $\Theta = 0$ to $\Theta = -R\Phi$. Since the CMB spectrum depends on the square of the solution, the shift of the equilibrium of the oscillating solution leads to odd and even peaks in the CMB having unequal heights. The relative heights of the CMB spectrum therefore provide a measure of the baryon density Ω_b.

Radiation driving

One important effect is not included in our highly simplified treatment. During the radiation era the gravitational potential Φ decays inside the horizon. Counterintuitively, the decaying potential actually enhances temperature fluctuations through a subtle resonance effect (see Fig. 32). Since the potential Φ decays after sound horizon crossing, it drives the first compression of the photon-baryon fluid without a counterbalancing effect on the subsequent rarefaction stage. The higher peaks in the CMB spectrum correspond to fluctuations that began their oscillations in the radiation-dominated era and therefore have enhanced amplitudes. This effect is sensitive to the ratio of radiation to matter, Ω_r/Ω_m. Since the radiations density Ω_r is fixed by the observed CMB temperature, measuring the peak heights relative to the Sachs-Wolf plateau determines the matter density of the universe Ω_m.

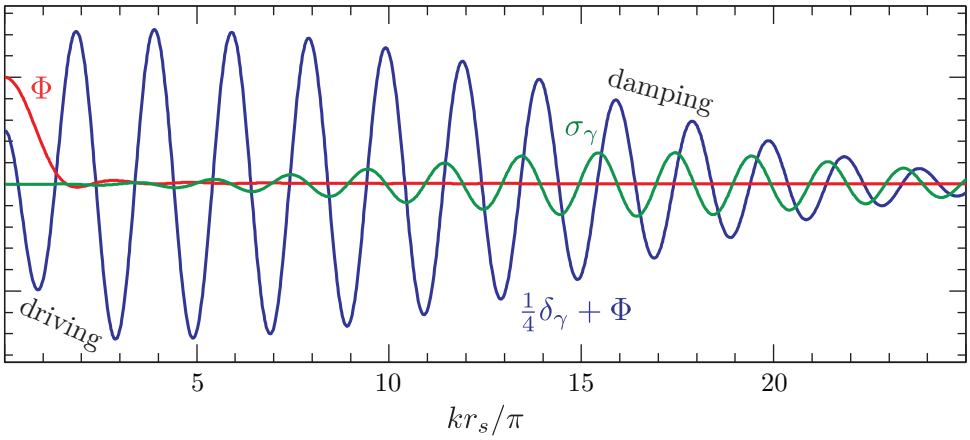


Figure 32: Acoustic oscillations with gravitational forcing and diffusion damping. For a mode that enters the sound horizon during radiation domination, the gravitational potential decays after horizon crossing and drives the acoustic amplitude higher. As the photon diffusion length increases and becomes comparable to the wavelength, viscosity σ_γ is generated and small-scale fluctuations are washed out.

Diffusion damping

So far, we have treated photons and baryons as a single perfect fluid, i.e. we took the mean free path of the photons to be zero. In reality, the coupling between electrons and photons is imperfect and the photons have a finite mean free path:

$$\lambda_C = \frac{1}{n_e \sigma_T a}\,, \tag{114}$$

where n_e is the electron density and σ_T is the Thomson cross section. Accounting for this leads to the damping of small-scale fluctuations: see Fig. 31. As the photons random walk through the baryons, hot and cold regions are mixed. By this process, fluctuations will be erased below the diffusion length:

$$\lambda_D = \sqrt{N}\,\lambda_C = \sqrt{\tau/\lambda_C}\,\lambda_C = \sqrt{\tau \lambda_C}\,, \tag{115}$$

which is the geometric mean between the horizon scale and the mean free path. As we will show in the following insert, the transfer function for the photon density fluctuations will receive an exponential suppression for modes with $k > k_D \equiv 2\pi/\lambda_D$. As shown in the insert below, the photon transfer function receives the following correction:

$$T(k) \to \mathcal{D}(k) T(k)\,, \tag{116}$$

where $\mathcal{D}(k) = e^{-k^2/k_D^2}$.

Imperfect fluid.—Diffusion causes heat conduction and generates viscosity in the fluid. Incorporating these effects into the dynamics leads to a modified oscillator equation [7],

$$\ddot{\Theta} + \mu\, c_s^2 k^2\, \dot{\Theta} + c_s^2 k^2\, \Theta = 0\,, \tag{117}$$

where we have ignored the gravitational source terms and defined

, (118)

$$\mu \equiv \left[\frac{16}{15} + \frac{R^2}{1+R}\right] \lambda_C\,. \tag{119}$$

Using the WKB ansatz

$$\Theta \propto \exp\left(i \int \omega \, \mathrm{d}\tau\right), \tag{120}$$

we get

$$-\omega^2 + \mu \, c_s^2 k^2 \, i\omega + c_s^2 k^2 = 0\,, \tag{121}$$

which we can write as

$$\omega = \pm c_s k \Big[1 + i\omega\mu\Big]^{1/2} \approx \pm c_s k \Big[1 \pm \frac{i}{2}\mu \, c_s k\Big]\,. \tag{122}$$

Substituting this back into (120), we get

$$\Theta \propto= e^{\pm i k r_s} \exp\left[-\frac{1}{2}(k/k_D)^2\right], \tag{123}$$

where we have defined the diffusion wavenumber as

$$k_D^{-2} = \int \mathrm{d}\eta \, \mu \, c_s^2 = \int \mathrm{d}\tau \, \frac{1}{3(1+R)} \left[\frac{16}{15} + \frac{R^2}{1+R}\right] \lambda_C\,. \tag{124}$$

In the limit $R \to 0$, this becomes

$$k_D^{-2} \approx \frac{16}{45} \int \mathrm{d}\tau \, \lambda_C \sim \tau \lambda_C\,, \tag{125}$$

which agrees with our previous estimate (115).

Within the Standard Model, the physics of the CMB anisotropies is understood extremely well. In the next section, we will explore what can be learned about physics beyond the Standard Model by looking for subtle deviations in the CMB spectrum.

6.5.6 Light Relics

Future cosmological observations will measure the radiation density of the early universe at the percent level. In this section, I will show how these observations will probe the physics of neutrinos, as well as the possible existence of extra light particles that are more weakly coupled to the SM than neutrinos. Examples of light relics that can be constrained in this way are the QCD axion [13, 14, 15], axion-like particles (ALPs) [16], dark photons [17] and light sterile neutrinos [8]. These particles may be so weakly coupled that they are hard to detect in terrestrial experiments, but the large number densities in the early universe make their gravitational imprints observable.

Dark Radiation

Let us assume that some physics beyond the Standard Model adds an extra radiation density ρ_X to the early universe. It is conventional to measure this density relative to the density of a SM neutrino species:

$$\Delta N_{\text{eff}} \equiv \frac{\rho_X}{\rho_\nu} = \frac{1}{a_\nu} \frac{\rho_X}{\rho_\gamma}\,, \tag{126}$$

and define $N_{\text{eff}} = N_\nu + \Delta N_{\text{eff}}$ as the *effective number of neutrinos*, although ρ_X may have nothing to do with neutrinos. Current measurements of the CMB anisotropies and the light

element abundances find

$$N_{\rm eff}^{\rm CMB} = 3.04 \pm 0.18\,, \tag{127}$$

$$N_{\rm eff}^{\rm BBN} = 2.85 \pm 0.28\,, \tag{128}$$

which is consistent with the SM prediction,[13] $N_{\rm eff} = 3.046$. Future CMB observations have the potential to improve these constraints by an order of magnitude [1].

A natural source for $\Delta N_{\rm eff} \neq 0$ are extra relativistic particles. Let us therefore consider a light species X as the only additional particle in some BSM theory. Assuming that this species was in thermal equilibrium with the SM at some point in the history of the universe, we can compute its contribution to $N_{\rm eff}$ in the same way as we derived the relic density of neutrinos in §6.5.3. For concreteness, let us assume that the particles of the species X decouple before neutrino decoupling, $T_{{\rm dec},X} > 10$ MeV. Particle-antiparticle annihilations until neutrino decoupling will lead to a difference between the temperature associated with the species X and that of neutrinos:

$$\begin{aligned} T_X = \left(\frac{g_*(T_{{\rm dec},\nu})}{g_*(T_{{\rm dec},X})}\right)^{1/3} T_\nu &= \left(\frac{10.75}{106.75}\right)^{1/3} \left(\frac{106.75}{g_*(T_{{\rm dec},X})}\right)^{1/3} T_\nu \\ &= 0.465 \left(\frac{106.75}{g_*(T_{{\rm dec},X})}\right)^{1/3} T_\nu\,. \end{aligned} \tag{129}$$

After neutrino decoupling, T_X and T_ν evolve in the same way, with both receiving the same suppression relative to T_γ from e^+e^- annihilation. As long as both X and ν are relativistic, they therefore maintain a constant energy ratio

$$\Delta N_{\rm eff} \equiv \frac{\rho_X}{\rho_\nu} = \frac{g_{*,X}}{g_{*,\nu}} \left(\frac{T_X}{T_\nu}\right)^4 = 0.027\, g_{*,X} \left(\frac{106.75}{g_*(T_{{\rm dec},X})}\right)^{4/3}\,, \tag{130}$$

where $g_{*,\nu} = \frac{7}{4}$ and $g_{*,X} = \{1, \frac{7}{4}, 2, \ldots\}$ are the internal degrees of freedom for spin $\{0, \frac{1}{2}, 1, \ldots\}$ particles. Figure 33 shows the contribution of a single decoupled species as a function of its decoupling temperature. We see that the contributions asymptote to fixed values for decoupling temperatures above the mass of the top quark (the heaviest particle of the SM). Using $g_*(T_{{\rm dec},X}) \leq 106.75$ in expression (130), we find that the extra species X contributes the following minimal amount[14]

$$\boxed{\Delta N_{\rm eff} > 0.027\, g_{*,X}} \;=\; \begin{cases} 0.054 & \text{gauge boson} \\ 0.047 & \text{Weyl fermion} \\ 0.027 & \text{Goldstone boson} \end{cases} \tag{131}$$

As we will see, this is an interesting target for future CMB experiments.

[13] The predicted value of $N_{\rm eff} = 3.046$ accounts for plasma corrections of quantum electrodynamics, flavour oscillations and, in particular, the fact that neutrinos have not fully decoupled when electrons and positrons annihilated.

[14] In deriving this bound, we assumed an extension of the SM in which there is no significant entropy production after decoupling and that the species X is the only addition to the SM particle content. Additional particles may both increase $\Delta N_{\rm eff}$ is there are light enough, or decrease it if they are relativistic at the decoupling of X, but become non-relativistic before neutrino decoupling. While entropy production typically dilutes the effects of extra relativistic species, it can also lead to additional effects that can be looked for in cosmological observations. For a more detailed discussion of these issues, see [20].

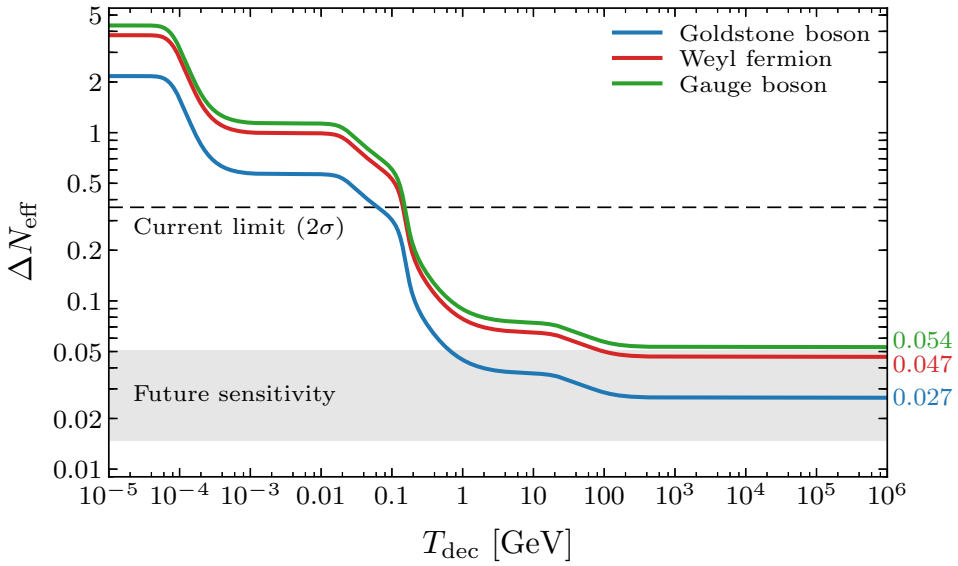


Figure 33: Contributions of a single thermally-decoupled Goldstone boson, Weyl fermion or massless gauge boson to the effective number of neutrinos, $\Delta N_{\rm eff}$, as a function of its decoupling temperature $T_{\rm dec}$.

Imprints in the CMB

Adding extra relativistic species will change the shape of the CMB spectrum. Precise measurements of the spectrum therefore make the CMB an accurate tool for probing this type of BSM physics.

Diffusion damping

The main effect of adding radiation density to the early universe is to increase the damping of the CMB spectrum [913] (see Fig. 34). Increasing $N_{\rm eff}$ increases H_*, the expansion rate at recombination. This would change both the damping scale θ_D and the peak location θ_*. Using

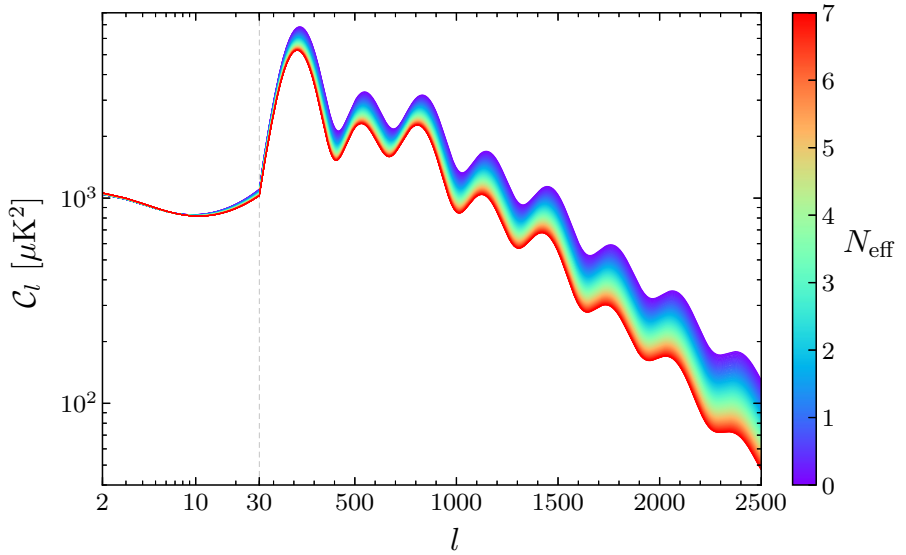


Figure 34: Variation of the CMB spectrum $\mathcal{C}_l \equiv l(l+1)C_l$ as a function of $N_{\rm eff}$ for fixed θ_*.

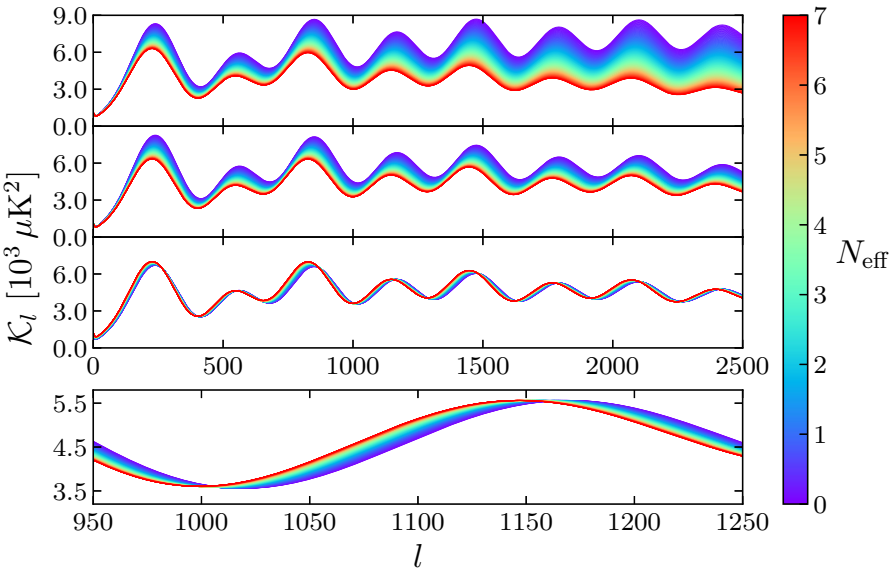


Figure 35: Variation of the undamped power spectra, $\mathcal{K}_l \equiv \mathcal{D}_l^{-1}\mathcal{C}_l$, as a function of $N_{\rm eff}$. The physical baryon density ω_b, the matter-to-radiation ratio ρ_m/ρ_r and the angular size of the sound horizon θ_* are held fixed in all panels. The dominant effect in the first panel is the variation of the damping scale θ_D. In the second panel, we fixed θ_D by adjusting the Helium fraction Y_P. The dominant variation is now the amplitude perturbation δA. In the third panel, the spectra are normalized at the fourth peak. The remaining variation is the phase shift φ (see the zoom-in in the fourth panel).

the estimates presented in §6.5.5, the ratio of θ_D and θ_* scales as

$$\frac{\theta_D}{\theta_*} = \frac{1}{r_{s,*}\,k_D} \propto \frac{1}{H_*^{-1}H_*^{1/2}} = H_*^{1/2}\,. \tag{132}$$

Since θ_* is measured very accurately, we need to keep it fixed. This can be done, for example, by simultaneously increasing the Hubble constant H_0. Increasing $N_{\rm eff}$ (and hence H_*) at fixed θ_* then implies larger θ_D, i.e. the damping kicks in at larger scales reducing the power in the damping tail (see Fig. 34). By accurately measuring the small-scale CMB anisotropies, observations therefore put a constraint on the number of relativistic species at recombination, cf. eq. (127).

The main limiting factor in these measurements is a degeneracy with the primordial Helium fraction $Y_P \equiv n_{\rm He}/n_b$. At fixed $\omega_b \equiv \Omega_b h^2$, increasing Y_P decreases the number density of free electrons. This increases the diffusion length, cf. eq. (114), and hence reduces the power in the damping tail. The parameters Y_P and $N_{\rm eff}$ are therefore anti-correlated. As we will discuss next, this degeneracy is broken by a more subtle effect of free-streaming relativistic species on the CMB spectrum.

Phase shift

Recently, CMB experiments have started to become sensitive to perturbations in the gas of relativistic particles. As we will see, perturbations in the density of free-streaming relativistic particles (e.g. neutrinos) leave a small imprint in the temporal phase of the acoustic oscillations and hence a coherent shift in the peak locations of the CMB spectrum (see Fig. 35). To get some intuition for the physical origin of this effect, let us return to the master equation (B.4), but now solve it slightly more accurately. Since the effect occurs during the radiation-dominated era, we

can still assume perfect radiation domination with $R = 0$. Equation (B.4) then becomes [22]

$$\ddot{d}_\gamma - c_s^2 \nabla^2 d_\gamma \;=\; \nabla^2 \Phi_+ \tag{133}$$

where we have defined $d_\gamma \equiv \frac{3}{4}\delta_\gamma - 3\Phi$ and $\Phi_\pm \equiv \Phi \pm \Psi$. The solution for a single Fourier mode can be written as

$$d_\gamma(y) = A\cos(y) - c_s^{-2} \int_0^y \mathrm{d}y'\, \Phi_+(y')\, \sin(y - y')\,, \tag{134}$$

where $y \equiv c_s k\tau$ and we have dropped the argument $\mathbf{k}$ on d_γ, A, and Φ_+ to avoid clutter. Using $\sin(y - y') = \sin(y)\cos(y') - \cos(y)\sin(y')$, we can write this as

$$d_\gamma(y) = \left[A + c_s^{-2}\alpha(y)\right]\cos(y) - c_s^{-2}\beta(y)\sin(y)\,, \tag{135}$$

where

$$\alpha(y) \equiv \int_0^y \mathrm{d}y'\, \Phi_+(y')\, \sin(y')\,, \tag{136}$$

$$\beta(y) \equiv \int_0^y \mathrm{d}y'\, \Phi_+(y')\, \cos(y')\,. \tag{137}$$

To obtain the CMB spectrum, we need to evaluate the solution at recombination. For the high-l modes of the CMB, it is a good approximation to use $y \to \infty$ for the limit of integration. The solution (135) can then be written as

$$d_\gamma(y) = \tilde{A}\cos(y + \varphi)\,, \tag{138}$$

where $\tilde{A} \equiv A + c_s^{-2}\alpha$ and

$$\sin\varphi \equiv \frac{\beta}{\sqrt{\beta^2 + (\alpha + c_s^2 A)^2}}\,. \tag{139}$$

We see that $\beta \neq 0$ corresponds to a constant phase shift of the acoustic oscillations.

To diagnose when such a phase shift can occur, it is useful to write the parameter β as follows

$$\begin{aligned}\beta \;&=\; \int_0^\infty \mathrm{d}y'\, \Phi_+(y')\cos(y') \\ &=\; \frac{1}{2}\int_{-\infty}^{+\infty} \mathrm{d}y' \left[\Phi_+^{(\mathrm{s})}(y')\right] e^{iy'}\,,\end{aligned} \tag{140}$$

where, in the second line, we have analytically continued the integrand and defined the symmetric part of the potential, $\Phi_+^{(\mathrm{s})}(y) \equiv \Phi_+(y) + \Phi_+(-y)$. For adiabatic modes, we expect $\Phi_+^{(\mathrm{s})}(y)$ to be an analytic function. Closing the contour in the upper-half plane, we find $\beta = 0$ if the contour at infinity vanishes. This is the case if $\Phi_+^{(\mathrm{s})}$ is sourced by fluctuations that travel at $c < c_s$. Neutrinos or other free-streaming particles, on the other hand, travel at the speed of light. This induces a mode in $\Phi_+^{(\mathrm{s})}$ of the form $e^{-ick\eta} = e^{-i(c/c_s)y}$, with $c = 1 > c_s$, and therefore leads to a finite phase shift.

Neutrino free-streaming.—We will briefly sketch how free-streaming relativistic particles, such as neutrinos, produce a phase shift in the CMB anisotropy spectrum. Details can be found in [22, 5].

The evolution of Φ_+ is related to that of Φ_- via the following Einstein equation:

$$\Phi_+'' + \frac{4}{y}\Phi_+' + \Phi_+ \;=\; \Phi_-'' + \frac{2}{y}\Phi_-' + 3\Phi_- \;\equiv\; \mathcal{S}[\Phi_-]\,, \tag{141}$$

where Φ_- is sourced by the anisotropic stress π_ν created by the neutrinos:

$$\Phi_-(y) = -\frac{2k^2}{y^2} f_\nu \pi_\nu(y)\,. \tag{142}$$

Here, we have introduced the fractional neutrino density $f_\nu \equiv \sum \rho_\nu/\rho_r \approx 0.41$. The evolution of π_ν follows from the Boltzmann equation for the neutrino distribution function. The solution can be written as [22]

$$\frac{k^2}{2}\pi_\nu(y) \approx -\zeta\, j_2(c_s^{-1}y) + c_s^{-1}\int_0^y \mathrm{d}y'\,\Phi_+(y')\left[\frac{2}{5}j_1(c_s^{-1}(y-y')) - \frac{3}{5}j_3(c_s^{-1}(y-y'))\right]. \tag{143}$$

We see that the solution involves an integral over Φ_+. Moreover, the solution depends on $c_s^{-1}y = k\tau$, i.e. it contains modes travelling at the speed of light. Following [22], the system of equations can be solved order by order in the fractional neutrino density:

- At zeroth order in f_ν, the potential Φ_- vanishes and the homogeneous solution is a function only of y, i.e. it doesn't contain modes travelling faster than the sound speed of the photon-baryon fluid. No phase shift is generated.
- At first order in f_ν, the potential Φ_- is non-zero. Note that the right-hand side of (142) is proportional to f_ν, so only the zeroth-order solution for π_ν (and Φ_+) is needed to determine the first-order solution for Φ_-. Computing the induced first-order correction to Φ_+, one finds [22, 5]

  $$\beta \approx 0.60\,\zeta\, f_\nu \quad \text{and} \quad \varphi \approx 0.19\,\pi\, f_\nu\,. \tag{144}$$

  As expected, a finite phase shift is generated. This phase shift has recently been detected in the CMB data [4, 5]. It has also been measured in the clustering of galaxies via its imprint in the spectrum of baryon acoustic oscillations [23].

CMB Stage 4

The sensitivity of ground-based CMB experiments can be characterized by the number of detectors that are mounted onto the telescope. The current generation of experiments has about 10^3 detectors, but there are plans to increase the number of detectors by up to two orders of magnitude [1]. This would lead to a significant improvement in the sensitivity of CMB experiments (see Fig. 36). What is particularly exciting about this is that it will lead to an order of magnitude improvement in constraints on $N_{\rm eff}$, allowing us to probe particles that decoupled before the QCD phase transition (cf. Fig. 33). If these so-called CMB Stage 4 experiments can reach the threshold $\Delta N_{\rm eff} = 0.027$, they would be sensitive to any light relics that have ever been in thermal equilibrium with the Standard Model (modulo the constraints described in footnote 14). As we will show in the next section, even the absence of a detection would be informative since it would put strong constraints on the couplings of extra light species to the SM.

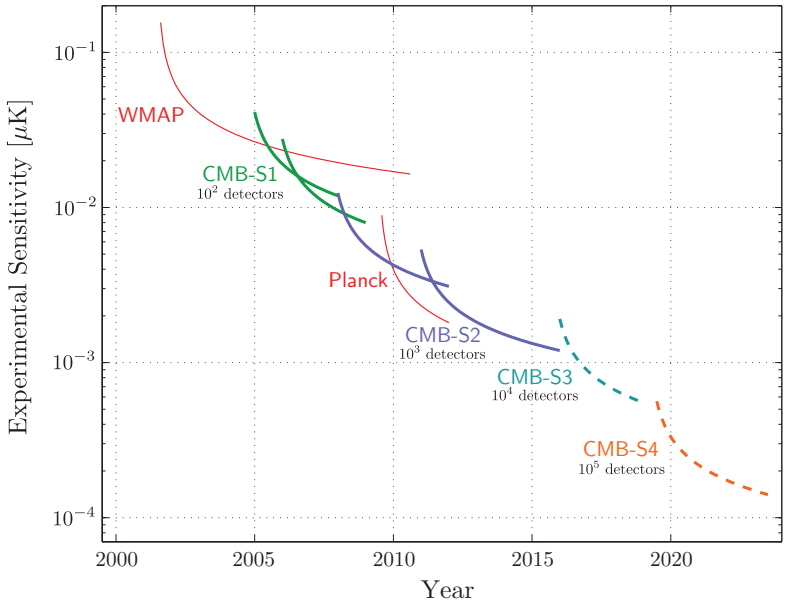


Figure 36: Evolution of the sensitivity of past and future CMB experiments (figure adapted from [1]).

EFT of Light Species

Extra light species arise in many models of physics beyond the Standard Model. Rather than studying these models one by one, it is more efficient to pass directly to an effective field theory for the couplings of the new light fields X to the SM,

$$\mathcal{L} \subset \sum g\, \mathcal{O}_X \mathcal{O}_{\rm SM}\,. \tag{145}$$

The strength of the couplings g determines the decoupling temperature, $T_{\rm dec}(g)$, and hence the thermal abundance of the species X, cf. Fig. 33.

Following [24], we consider EFTs that are minimal and technically natural. Minimality here means that the additional particle content is as small as possible, usually consisting of only one additional elementary particle. Naturalness requires the small masses of the extra particles to be protected by symmetries, which also restricts the allowed interactions in (145). Since the available symmetries depend on the spin of the new particles, it is convenient to organize the EFT according to spin.

Spin-0

A particularly well-motivated example of light particles are *Goldstone bosons* created by the spontaneous breaking of additional global symmetries. Goldstone bosons are either massless (if the broken symmetry was exact) or naturally light (if it was approximate). Examples of light pseudo-Nambu-Goldstone bosons (pNGBs) are *axions* [13, 14, 15], *familons* [25, 26, 27], and *majorons* [28, 29], associated with spontaneously broken Peccei-Quinn, family and lepton-number symmetry, respectively. Axion, familon and majoron models are characterized by different couplings in (145).

Axions arise naturally in many areas of high-energy physics, the QCD axion being a famous example. They are a compelling example of a new particle that is experimentally elusive because of its weak coupling rather than due to kinematic constraints. What typically distinguishes axions from other pNGBs are their unique couplings to the SM gauge fields. Below the scale

of electroweak symmetry breaking, the shift-symmetric couplings of the axion to the SM gauge fields are

$$\mathcal{L} = -\frac{1}{4}\left(\frac{a}{\Lambda_\gamma} F_{\mu\nu}\tilde{F}^{\mu\nu} + \frac{a}{\Lambda_g} G^a_{\mu\nu}\tilde{G}^{\mu\nu,a}\right), \tag{146}$$

where $X_{\mu\nu} \equiv \{F_{\mu\nu}, G^a_{\mu\nu}\}$ are the field strengths of photons and gluons, and $\tilde{X}^{\mu\nu} \equiv \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}X_{\rho\sigma}$ are their duals. Axion models will typically include couplings to all SM gauge fields, but only the coupling to gluons is strictly necessary to solve the strong CP problem.

Spin-$\frac{1}{2}$

Light *fermions* are a natural possibility because Weyl and Dirac mass terms are protected by chiral and axial symmetries, respectively. A hidden Weyl fermion χ then couples to the SM through an anapole moment, $\chi^\dagger\bar{\sigma}^\mu\chi\partial^\nu B_{\mu\nu}$, or a four-fermion interaction, $\chi^\dagger\bar{\sigma}^\mu\chi\psi\gamma_\mu\psi$, while a Dirac fermion Ψ can couple through a dipole interaction, $\bar{\Psi}\sigma^{\mu\nu}\Psi B_{\mu\nu}$.

Spin-1

Massless spin-1 particles are technically natural because they have fewer degrees of freedom than their massive counterparts. The dominant coupling of *dark photons* A'_μ to the SM is through the dipole interaction $H\bar{\psi}\sigma^{\mu\nu}\psi F'_{\mu\nu}$.

Spin-$\frac{3}{2}$

The *gravitino* is a universal prediction of supergravity. Its mass is set by the SUSY breaking scale, $m_{3/2} = F/M_{\rm pl}$, and can be very small in low-scale SUSY breaking scenarios. The coupling of the longitudinal component of the gravitino to the SM is equivalent to the Goldstino coupling $\chi^\dagger\sigma_\mu\partial_\nu\chi\, T^{\mu\nu}$. The strength of the coupling is set by the SUSY-breaking scale F rather than $M_{\rm pl}$.

Spin-2

The *graviton* interacts only through Planck-suppressed gravitational interactions and hence has never been in thermal equilibrium with the SM. Its thermal abundance is therefore negligible.

Cosmic axions

Let us illustrate the power of future CMB observations through the example of axions [30]. To simplify the narrative, I will assume that a future CMB-S4 mission will be sensitive enough to exclude the minimal abundance of thermal axions, $\Delta N_{\rm eff} > 0.027$. In practice, this will probably require additional data from large-scale structure surveys [31].

At high energies, the rate of axion production is through the gauge field interactions (146) and can be expressed as [32]

$$\Gamma(\Lambda_n, T) = \sum_n \gamma_n(T)\,\frac{T^3}{\Lambda_n^2}\,. \tag{147}$$

The prefactors $\gamma_n(T)$ have their origin in the running of the couplings and are only weakly dependent on temperature. We will ignore this temperature dependence in the following. We see that the production rate, $\Gamma \propto T^3$, decreases faster than the expansion rate during the radiation era, $H \propto T^2$. To avoid producing a thermal axion abundance requires that the

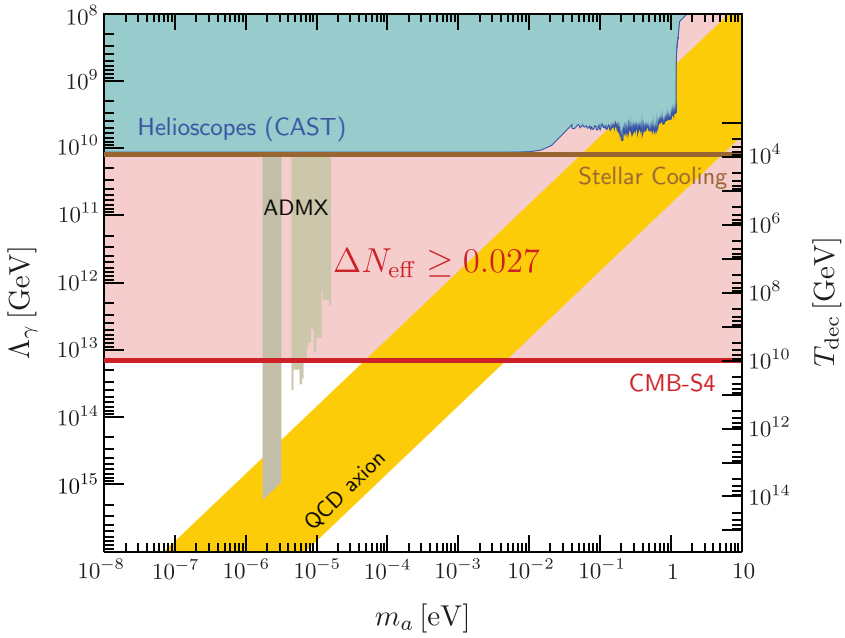


Figure 37: Comparison between current constraints on the axion-photon coupling and the sensitivity of a future CMB-S4 mission (figure adapted from [33]). The yellow band indicates a range of representative models for the QCD axion. The future CMB bound is a function of the reheating temperature T_R. We note that ADMX assumes that the axion is all of the dark matter, while all other constraints do not have this restriction.

interaction rate was never larger than the expansion rate. Denoting the reheating temperature of the universe by T_R, this implies

$$\Gamma(\Lambda_n, T_R) < H(T_R) = \frac{\pi}{\sqrt{90}} \sqrt{g_{*,R}} \frac{T_R^2}{M_{\rm pl}}\,, \tag{148}$$

where $g_{*,R} \equiv g_*(T_R)$. For a given reheating temperature, this is a constraint on the couplings Λ_n in (147). Treating the different axion couplings separately, we can write

$$\Lambda_n \;>\; \left(\frac{\pi^2}{90} g_{*,R}\right)^{-1/4} \sqrt{\gamma_{n,R} T_R M_{\rm pl}}\,, \tag{149}$$

where $\gamma_{n,R} \equiv \gamma_n(T_R)$.

The operator that has been most actively investigated experimentally is the coupling to photons. Photons are easily produced in large numbers in both the laboratory and in many astrophysical settings which makes this coupling a particularly fruitful target for axion searches. The couplings in the high-energy theory prior to electroweak symmetry breaking are related to the photon coupling Λ_γ through the Weinberg mixing angle. To be conservative, I will present the weakest constraint which arises when the axion only couples to the $U(1)_Y$ gauge field. A specific axion model is likely to also couple to the $SU(2)_L$ sector and the constraint on Λ_γ would then be stronger. Using $\gamma_{\gamma,R} \approx \gamma_\gamma(10^{10}\,\text{GeV}) = 0.029$ and $g_{*,R} = 106.75 + 1$, we find

$$\Lambda_\gamma \;>\; 1.4 \times 10^{13}\,\text{GeV}\,\sqrt{T_{R,10}}\,, \tag{150}$$

where $T_{R,10} \equiv T_R/10^{10}\,\text{GeV}$. For a reheating temperature of about $10^{10}\,\text{GeV}$, the bound in (150) is three orders of magnitude stronger than the best current constraints (cf. Fig. 37). Even for a reheating temperature as low as $10^4\,\text{GeV}$ the bound from the CMB would still marginally improve over existing constraints.

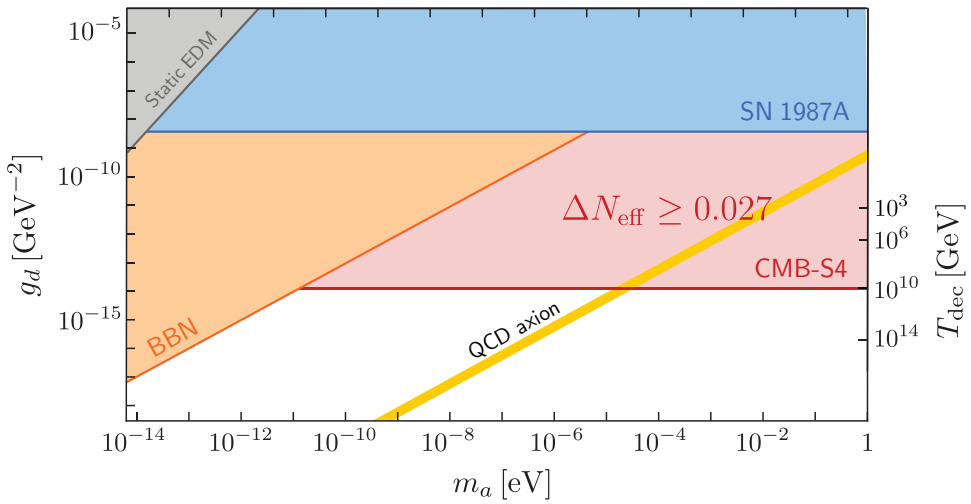


Figure 38: Comparison between current constraints on the axion-gluon coupling and the sensitivity of a future CMB-S4 mission (figure adapted from [34, 35]). We note that the static EDM [34] and BBN constraints [35] assume that the axion is all of the dark matter, while SN 1987A [36] and the future CMB constraint do not have this restriction.

The coupling to gluons is especially interesting for the QCD axion since it has to be present in order to solve the strong CP problem. The axion production rate associated with the gluon interaction in (146) is $\Gamma_g \simeq 0.41 T^3/\Lambda_g^2$ [32]. The bound (149) then implies

$$\Lambda_g > 5.4 \times 10^{13}\,\mathrm{GeV}\,\sqrt{T_{R,10}}\,. \tag{151}$$

Laboratory constraints on the axion-gluon coupling are usually phrased in terms of the induced electric dipole moment (EDM) of nucleons: $d_n = g_d a_0$, where a_0 is the value of the local axion field. For the QCD axion, the coupling g_d is given by [37, 34]

$$g_d \approx \frac{2\pi}{\alpha_s} \times \frac{3.8 \times 10^{-3}\,\mathrm{GeV}^{-1}}{\Lambda_g}\,. \tag{152}$$

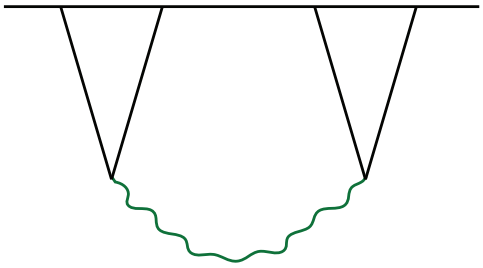
Constraints on g_d (and hence Λ_g) are shown in Fig. 38. We see that future CMB-S4 observations can improve over existing constraints on Λ_g by up to six orders of magnitude if $T_R = \mathcal{O}(10^{10}\,\mathrm{GeV})$. Even if the reheating temperature is as low as $10^4\,\mathrm{GeV}$, the future CMB constraints will be tighter by three orders of magnitude.

Deriving similar constraints for the other axion couplings and for the couplings of fields with spin is left as an exercise for the reader.

6.6 Relics from Inflation

Inflation predicts that the quantum fluctuations of any massless fields get amplified by the rapid expansion of the spacetime. Two massless fields that are guaranteed to exist in all inflationary models are the curvature perturbation ζ and the tensor fluctuations γ_{ij}. In Part I of these lectures, we assumed a nearly scale-invariant spectrum of curvature perturbations as a source for the density fluctuations in the late universe. In the following, I will show that these initial conditions indeed naturally arise from inflation. I will also demonstrate that the same effect produces tensor fluctuations. The search for this stochastic background of primordial gravitational waves is a very active area of observational cosmology.

Inflation also excites massive particles as long as their masses aren't far above the inflationary Hubble scale. Since the Hubble scale during inflation may be as high as 10^{14} GeV, this gives us the opportunity to probe the particle spectrum at energies far beyond the reach of conventional particle colliders. Once produced, these massive particles quickly decay into the massless modes ζ and γ_{ij}, creating higher-order correlations in the inflatonary fluctuations:



As we will see, this effect leads to a characteristic non-locality in cosmological correlators.

We will start with a brief review of slow-roll inflation. In Section 6.6.1, we will discuss the dynamics of the inflationary background, while in Section 6.6.2 we will explicitly compute the spectrum of quantum fluctuations. In Section 6.6.3, we will extend this treatment to include interactions and show that they lead to a characteristic non-Gaussianity in cosmological correlators. In Section 6.6.4, we will discuss the imprints of extra fields, showing how the masses and spins are encoded in the momentum dependence of higher-order correlation functions.

6.6.1 Inflationary Cosmology

A key fact about the universe is that on large scales it is described by the FRW metric (328). But why? A naive extrapolation of the radiation-dominated Big Bang cosmology suggests that the early universe consisted of many causally disconnected regions of space. The fact that these apparently disjoint patches of space are observed to have nearly the same densities and temperatures is called the *horizon problem*. In this section, I will explain how inflation—an early period of accelerated expansion—drives the primordial universe towards homogeneity and isotropy, even if it started in a more generic initial state.

Horizon Problem

The *particle horizon* is the maximal distance that a signal can travel between the time corresponding to the initial singularity, $t_i \equiv 0$, and a later time t. In physical coordinates, this distance is given by

$$D(t) \;=\; a(t)\int_0^t \frac{\mathrm{d}t}{a(t)} = a(t)\int_0^a \frac{\mathrm{d}\ln a}{a'}\,. \tag{153}$$

If the early universe was filled by ordinary matter, then $a'' < 0$. In that case, the integral in (153) is dominated by late times and converges to a finite value:

$$\text{e.g.}\quad a(t) \propto \begin{cases} t^{2/3} \\ t^{1/2} \end{cases} \quad\to\quad D(t) \;=\; \begin{cases} 3t & \text{matter} \\ 2t & \text{radiation} \end{cases} \tag{154}$$

This leads to a puzzle: because the age of the universe (t_0) is much larger than the time of recombination (t_*), the CMB naively consists of many causally disconnected patches (see Fig. 39). The following questions immediately arise: Why is the CMB so homogeneous? And, more importantly, why are the observed CMB fluctuations correlated on large scales and not just random noise?

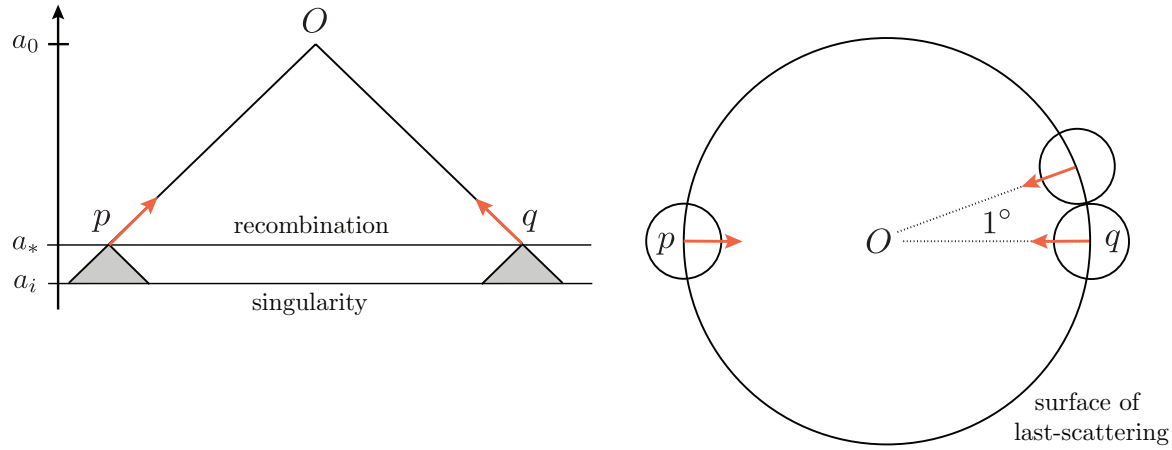


Figure 39: Illustration of the horizon problem in the conventional Big Bang model. All events that we currently observe are on our past light cone. The intersection of our past light cone with the spacelike slice labelled "recombination" corresponds to the "surface of last-scattering". Any two points on the surface of last-scattering that are separated by more than 1 degree, appear never to have been in causal contact. This means that their past light cones do not overlap before the singularity. This is illustrated for opposite points on the sky labelled p and q.

The horizon problem is solved if the early universe experienced a sustained period of accelerated expansion (= inflation), $a'' > 0$. In that case, the integral in (153) is dominated by early times and the particle horizon diverges in the past. Signals were therefore able to travel a much larger distance than suggested by the naive extrapolation of the standard FRW expansion (see Fig. 40).

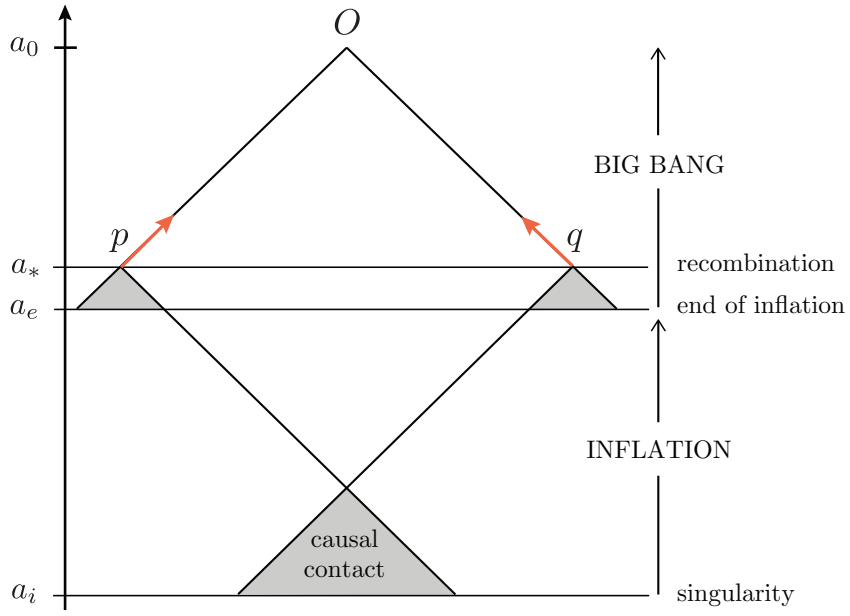


Figure 40: Illustration of the inflationary solution to the horizon problem in comoving coordinates (using conformal time on the vertical axis). The spacelike singularity of the standard Big Bang is replaced by the reheating surface, i.e. rather than marking the beginning of time it now simply corresponds to the transition from the end of inflation to the standard Big Bang evolution. All points in the CMB have overlapping past light cones and therefore originated from a causally connected region of space.

Exercise.—A special case of accelerated expansion is the quasi-de Sitter limit, which is characterized by a nearly constant expansion rate, $H = a'/a \approx const.$, so that $a(t) = e^{H(t-t_0)}$, where t_0 is some fiducial time at which $a(t_0) \equiv 1$. Show that

$$a(\tau) = -\frac{1}{H\tau}\,, \tag{155}$$

for $\tau < 0$. Notice that the initial singularity has been pushed to $\tau = -\infty$ (cf. Fig. 40).

Exercise.—Show that $a'' > 0$ is equivalent to a slow variation of the Hubble parameter

$$\varepsilon \equiv -\frac{H'}{H^2} < 1\,. \tag{156}$$

Notice that $\varepsilon \approx 0$ corresponds to a quasi-de Sitter spacetime with a nearly constant expansion rate, $H \approx const.$ Using the Friedmann equations, show that

$$\varepsilon = \frac{3}{2}\left(1 + \frac{P}{\rho}\right) < 1 \quad \Leftrightarrow \quad w \equiv \frac{P}{\rho} < -\frac{1}{3}\,. \tag{157}$$

The last condition corresponds to a violation of the strong energy condition.

Slow-Roll Inflation

As a simple toy model for inflation, let us consider the dynamics of a scalar field, the *inflaton* $\phi(t, \mathbf{x})$. As indicated by the notation, the value of the field can depend on time t and the position in space $\mathbf{x}$. Associated with each field value is a potential energy density $V(\phi)$ (see Fig. 63). If the field is dynamical (i.e. changes with time) then it also carries a kinetic energy density. If the energy density associated with the scalar field dominates the universe, it sources the evolution of the FRW background. We want to determine under which conditions this can lead to accelerated expansion.

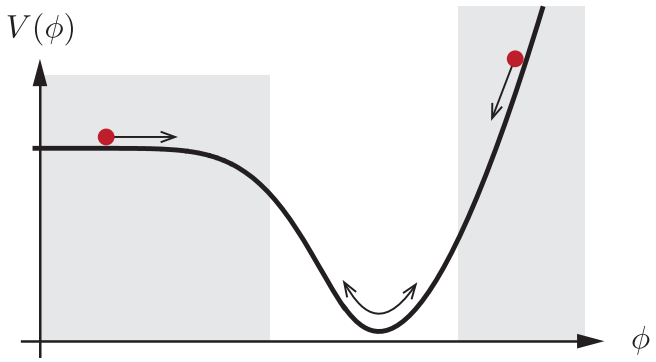


Figure 41: Example of a slow-roll potential. Inflation occurs in the shaded parts of the potential.

The stress-energy tensor of the scalar field is

$$T_{\mu\nu} \;=\; \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left(\frac{1}{2}g^{\alpha\beta}\partial_\alpha\phi\partial_\beta\phi - V(\phi)\right). \tag{158}$$

Consistency with the symmetries of the FRW spacetime requires that the background value of the inflaton only depends on time, $\phi = \phi(t)$. From the time-time component $T^0{}_0 = -\rho_\phi$, we

infer that the energy density of the field is

$$\rho_\phi \;=\; \frac{1}{2}(\phi')^2 + V)\phi)\,. \tag{159}$$

We see that this is simply the sum of the kinetic energy density, $\frac{1}{2}(\phi')^2$, and the potential energy density, $V(\phi)$. From the space-space component $T^i{}_j = P_\phi\,\delta^i_j$, we find that the pressure is the *difference* of kinetic and potential energy densities,

$$P_\phi \;=\; \frac{1}{2}(\phi')^2 - V)\phi)\,. \tag{160}$$

A field configuration therefore leads to inflation, $P_\phi < -\frac{1}{3}\rho_\phi$, if the potential energy dominates over the kinetic energy. i.e. if the field *rolls slowly*.

Exercise.—Using the Einstein equations (A.122) and (39), show that

$$\left.\begin{aligned} 3M_{\rm pl}^2H^2 &= \tfrac{1}{2}(\phi')^2 + V \\ M_{\rm pl}^2H' &= -\tfrac{1}{2}(\phi')^2 \end{aligned}\right\} \quad\Rightarrow\quad \varepsilon = \frac{\frac{1}{2}(\phi')^2}{M_{\rm pl}^2H^2} < 1\,, \tag{161}$$

where $M_{\rm pl} = (8\pi G)^{-1/2}$ is the reduced Planck mass.

Equation (161) corresponds to the *first slow-roll condition*:

$$(\phi')^2 < V\,. \tag{162}$$

This condition alone, however, does not guarantee successful inflation. We also need to assure that inflation does not just occur for an instant, but lasts long enough to solve the horizon problem.

Combining the two Friedmann equations in (161) leads to the Klein-Gordon equation for the evolution of the scalar field

$$\phi'' + 3H\phi' = -V_{,\phi}\,, \tag{163}$$

where $V_{,\phi}$ denotes the derivative of the potential with respect to the field value. To maintain the slow-roll condition (162) for a sufficient period of time, we require that the acceleration of the field is small. This is quantified by the *second slow-roll condition*:

$$\phi'' < 3H\phi'\,. \tag{164}$$

The evolution of the scalar field is then friction dominated, with the velocity of the field determined by the slope of the potential.

The two slow-roll conditions (162) and (164) can be expressed as conditions on the shape of the inflaton potential:

$$\varepsilon_V \equiv \frac{M_{\rm pl}^2}{2}\left(\frac{V_{,\phi}}{V}\right)^2 < 1\,, \qquad \eta_V \equiv M_{\rm pl}^2\left|\frac{V_{,\phi\phi}}{V}\right| < 1\,. \tag{165}$$

We will refer to the parameters in (165) as the *(potential) slow-roll parameters*.

Exercise.—Applying the slow-roll conditions (162) and (164) to the Friedmann equation (161) and the Klein-Gordon equation (163), we get

$$3M_{\rm pl}^2H^2 \approx V\,, \tag{166}$$

$$3H\phi' \approx -V_{,\phi}\,. \tag{167}$$

Use this to show that $\varepsilon \approx \varepsilon_V$ during slow-roll inflation.

Inflation ends when the first slow-roll condition is violated, $\varepsilon_V(\phi_e) \equiv 1$. The amount of inflation is measured in terms of 'e-folds' of expansion, $\mathrm{d}N = \mathrm{d}\ln a$. The total number of e-folds between a point ϕ on the potential and the end of inflation at ϕ_e is

$$N(\phi) \equiv \int_a^{a_e} \mathrm{d}\ln a = \int_t^{t_e} H\,\mathrm{d}t = \int_\phi^{\phi_e} \frac{H}{\phi'}\,\mathrm{d}\phi$$
$$= \int_\phi^{\phi_e} \frac{1}{\sqrt{2\varepsilon}} \frac{|\mathrm{d}\phi|}{M_{\rm pl}} \approx \int_\phi^{\phi_e} \frac{1}{\sqrt{2\varepsilon_V}} \frac{|\mathrm{d}\phi|}{M_{\rm pl}}\,. \tag{168}$$

To solve the horizon problem requires at least between 40 and 60 e-folds (with the precise value depending on the reheating temperature).

Case study: $m^2\phi^2$ inflation.—As an example, let us give the slow-roll analysis of arguably the simplest model of inflation: single-field inflation driven by a mass term

$$V(\phi) = \frac{1}{2}m^2\phi^2\,. \tag{169}$$

The slow-roll parameters are

$$\varepsilon_V(\phi) = \eta_V(\phi) = 2\left(\frac{M_{\rm pl}}{\phi}\right)^2\,. \tag{170}$$

To satisfy the slow-roll conditions $\{\varepsilon_V, |\eta_V|\} < 1$, we therefore need to consider super-Planckian values for the inflaton

$$\phi > \sqrt{2}M_{\rm pl} \equiv \phi_e\,. \tag{171}$$

The relation between the inflaton field value and the number of e-folds before the end of inflation is

$$N(\phi) = \frac{\phi^2}{4M_{\rm pl}^2} - \frac{1}{2}\,. \tag{172}$$

Solving the horizon problem then requires that the initial value of the field, ϕ_i, satisfies

$$\phi_i > \phi_{60} \equiv 2\sqrt{60}\,M_{\rm pl} \sim 15\,M_{\rm pl}\,. \tag{173}$$

We note that the total field excursion is super-Planckian, $\Delta\phi = \phi_i - \phi_e \gg M_{\rm pl}$.

Effective Field Theory

In the absence of a complete microscopic theory of inflation, we describe inflation in the context of an effective field theory. We are then obliged to include in the inflationary action all operators consistent with the assumed symmetries of the inflaton,

$$\mathcal{L}_{\rm eff}(\phi) \;=\; -\frac{1}{2}(\partial_\mu\phi)^2 - V(\phi) \;+\; \sum_n c_n V(\phi)\frac{\phi^{2n}}{\Lambda^{2n}} + \sum_n d_n \frac{(\partial\phi)^{2n}}{\Lambda^{4n}} + \cdots\,. \tag{174}$$

One of the remarkable features of inflation is that it is sensitive even to Planck-suppressed operators.

Eta problem

Successful inflation requires that the inflaton mass m_ϕ is parametrically smaller than the Hubble scale H:

$$\eta_V = \frac{m_\phi^2}{3H^2} \ll 1\,. \tag{175}$$

It is difficult to protect this hierarchy against high-energy corrections. We know that some new degrees of freedom must appear at $\Lambda \lesssim M_{\rm pl}$ to give a UV-completion of gravity. In string theory, this scale is often found to be significantly below the Planck scale, $\Lambda \lesssim M_s \ll M_{\rm pl}$. If ϕ has order-one couplings to any massive fields ψ (with $m_\psi \sim \Lambda$), then integrating out the fields ψ yields the effective action (1350) for ϕ with order-one couplings c_n and d_n. The above argument makes us worry that integrating out the massive fields ψ yields corrections to the potential of the form

$$\Delta V = c_1\, V(\phi)\frac{\phi^2}{\Lambda^2}\,, \tag{176}$$

where $c_1 \sim \mathcal{O}(1)$. If this term arises, then the eta parameter receives the following correction

$$\Delta\eta_V = \frac{M_{\rm pl}^2}{V})\Delta V)'' \approx 2c_1\left(\frac{M_{\rm pl}}{\Lambda}\right)^2 > 1\,, \tag{177}$$

where the final inequality follows from $\Lambda \lesssim M_{\rm pl}$. Notice that this problem is independent of the energy scale of inflation. All inflationary models have to address the eta problem.

Large-field inflation

The Planck-scale sensitivity of inflation is dramatically enhanced in models with observable gravitational waves, $r \gtrsim 0.01$. In this case, the inflaton field moves over a super-Planckian range during the last 60 e-folds of inflation, $\Delta\phi > M_{\rm pl}$ (see §6.6.2), and an infinite number of operators contribute equally to the effective action (1350). This observation makes an effective field theorist nervous and a string theorist curious [38].

6.6.2 Quantum Initial Conditions

One of the most remarkable features of inflation is that it provides a natural mechanism for producing the initial conditions for the hot big bang. To see this, recall that the evolution of the inflaton field $\phi(t)$ governs the energy density of the early universe $\rho(t)$ and, hence, controls the end of inflation (see Fig. 42). Essentially, the field ϕ plays the role of a "clock" reading off the amount of inflationary expansion still to occur. By the uncertainty principle, arbitrarily precise timing is not possible in quantum mechanics. Instead, quantum-mechanical clocks necessarily have some variance, so the inflaton will have spatially varying fluctuations $\delta\phi(t,\mathbf{x})$. There will therefore be local differences in the time when inflation ends, $\delta t(\mathbf{x})$, so that different regions of space inflate by different amounts. These differences in the local expansion histories lead to differences in the local densities after inflation, $\delta\rho(t,\mathbf{x})$, and to curvature perturbations in comoving gauge, $\zeta(\mathbf{x})$. It is worth remarking that the theory was not engineered to produce these fluctuations, but that their origin is instead a natural consequence of treating inflation quantum mechanically.

Quantum Fluctuations

Free Scalar in de Sitter

Before attacking the real problem of interest, namely the quantization of coupled inflaton-metric fluctuations during inflation, we will consider the simpler case of a free scalar field in

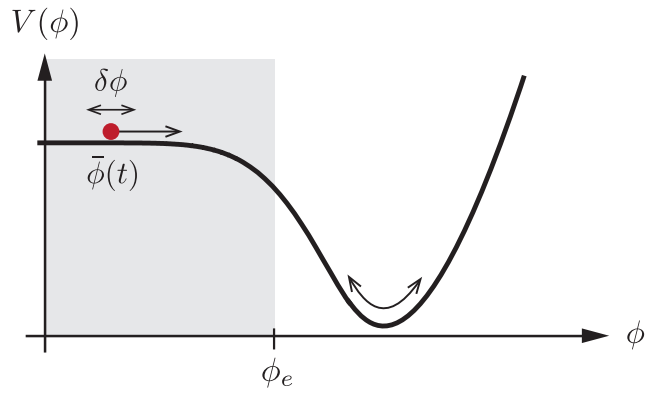


Figure 42: Quantum fluctuations $\delta\phi(t, \mathbf{x})$ around the classical background evolution $\bar{\phi}(t)$. Regions acquiring negative fluctuations $\delta\phi$ remain potential-dominated longer than regions with positive $\delta\phi$. Different parts of the universe therefore undergo slightly different evolutions. After inflation, this induces density fluctuations $\delta\rho(t, \mathbf{x})$.

de Sitter space. We will assume that the scalar field carries an insignificant amount of the total energy density and, hence, doesn't backreact on the de Sitter geometry. Such a field is sometimes called a *spectator field*.

The action of a massless, free scalar field in de Sitter space is

$$\begin{aligned} S &= \frac{1}{2}\int \mathrm{d}^4x\sqrt{-g}\, g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi \\ &= \frac{1}{2}\int \mathrm{d}\tau\, \mathrm{d}^3x\, a^2\left[\dot{\varphi}^2 - (\partial_i\varphi)^2\right], \end{aligned} \tag{178}$$

where $a(\tau) = -1/(H\tau)$. It is useful to define the canonically normalized field $v \equiv a\varphi$, so that

$$S = \frac{1}{2}\int \mathrm{d}\tau\, \mathrm{d}^3x\left[\dot{v}^2 - (\partial_i v)^2 + \frac{\ddot{a}}{a}v^2\right]. \tag{179}$$

This action implies the following equation of motion for the Fourier modes of the field

$$\boxed{\ddot{v}_{\mathbf{k}} + \left(k^2 - \frac{\ddot{a}}{a}\right)v_{\mathbf{k}} = 0}\,. \tag{180}$$

Sometimes this is called the *Mukhanov-Sasaki (MS) equation.*

We see that the expansion of the universe is captured by the time dependence of the effective mass of the canonically-normalized field, $m^2_{\text{eff}}(\tau) \equiv \ddot{a}/a = 2/\tau^2$. At early times, $-\tau \gg k^{-1}$, this mass is much smaller than the momentum k of the relevant Fourier modes. The dynamics then reduces to that of a simple harmonic oscillator in flat space,

$$\ddot{v}_{\mathbf{k}} + k^2 v_{\mathbf{k}} \approx 0\,. \tag{181}$$

In this limit, the quantum fluctuations of the field v therefore follow from the standard quantization of a simple harmonic oscillator.

Canonical quantization

From the action (179), we read off the the momentum conjugate to v:

$$\pi \equiv \frac{\partial\mathcal{L}}{\partial\dot{v}} = \dot{v}\,. \tag{182}$$

We promote the fields $v(\tau, \mathbf{x})$ and $\pi(\tau, \mathbf{x})$ to quantum operators $\hat{v}(\tau, \mathbf{x})$ and $\hat{\pi}(\tau, \mathbf{x})$. These operators satisfy the equal time commutation relation

$$\boxed{[\hat{v}(\tau, \mathbf{x}), \hat{\pi}(\tau, \mathbf{x}')] = i\delta_D(\mathbf{x} - \mathbf{x}')} \tag{183}$$

in units where $\hbar \equiv 1$. The delta function is required by *locality*: modes at different points in space are independent and the corresponding operators therefore commute. In Fourier space, this becomes

$$\begin{aligned}
[\hat{v}_{\mathbf{k}}(\tau), \hat{\pi}_{\mathbf{k}'}(\tau)] &= \int \mathrm{d}^3x \int \mathrm{d}^3x' \underbrace{[\hat{v}(\tau, \mathbf{x}), \hat{\pi}(\tau, \mathbf{x}')]}_{i\delta_D(\mathbf{x} - \mathbf{x}')} e^{-i\mathbf{k}\cdot\mathbf{x}} e^{-i\mathbf{k}'\cdot\mathbf{x}'} \\
&= i \int \mathrm{d}^3x \, e^{-i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{x}} \\
&= i\,(2\pi)^3 \delta_D(\mathbf{k} + \mathbf{k}')\,,
\end{aligned} \tag{184}$$

where the delta function implies that modes with different wavelengths commute.

Note that we are using the Heisenberg picture where operators vary in time while states are time independent. The operator solution $\hat{v}_{\mathbf{k}}(\tau)$ is determined by two initial conditions $\hat{v}_{\mathbf{k}}(\tau_i)$ and $\hat{\pi}_{\mathbf{k}}(\tau_i) = \partial_\tau \hat{v}_{\mathbf{k}}(\tau_i)$. Since the evolution equation is linear, the solution is linear in these operators. It is convenient to trade $\hat{v}_{\mathbf{k}}(\tau_i)$ and $\hat{\pi}_{\mathbf{k}}(\tau_i)$ for a single time-independent non-Hermitian operator $\hat{a}_{\mathbf{k}}$, in terms of which the solution can be written as

$$\boxed{\hat{v}_{\mathbf{k}}(\tau) = v_k(\tau)\,\hat{a}_{\mathbf{k}} + v_k^*(\tau)\,a^\dagger_{-\mathbf{k}}} \tag{185}$$

where the (complex) mode function $v_k(\tau)$ satisfies the classical equation of motion (180). Of course, $v_k^*(\tau)$ is the complex conjugate of $v_k(\tau)$ and $\hat{a}^\dagger_{\mathbf{k}}$ is the Hermitian conjugate of $\hat{a}_{\mathbf{k}}$. As indicated by dropping the vector notation $\mathbf{k}$ on the subscript, the mode functions, $v_k(\tau)$ and $v_k^*(\tau)$, are the same for all Fourier modes with $k \equiv |\mathbf{k}|$.[15]

We choose the normalization of the mode functions, so that $v_k \dot{v}_k^* - \dot{v}_k v_k^* \equiv i$. Substituting (185) into (184), we then get

$$\boxed{[\hat{a}_{\mathbf{k}}, \hat{a}^\dagger_{\mathbf{k}'}] = (2\pi)^3 \delta_D(\mathbf{k} + \mathbf{k}')} \tag{186}$$

which is the standard commutation relation for the *raising* and *lowering operators* of a harmonic oscillator. The quantum states in the Hilbert space are constructed by defining the vacuum state $|0\rangle$ via

$$\hat{a}_{\mathbf{k}}|0\rangle = 0\,, \tag{187}$$

and by producing excited states through repeated application of the creation operators $a^\dagger_{\mathbf{k}}$.

Choice of vacuum

The most general solution of equation (180) is

$$v_k(\tau) = c_1 \left(1 - \frac{i}{k\tau}\right) e^{-ik\tau} + c_2 \left(1 + \frac{i}{k\tau}\right) e^{ik\tau}\,, \tag{188}$$

[15] Since the frequency $\omega_k(\tau) \equiv k^2 - \ddot{a}/a$ in (180) depends only on $k \equiv |\mathbf{k}|$, the evolution does not depend on direction. The constant operators $\hat{a}_{\mathbf{k}}$ and $\hat{a}^\dagger_{\mathbf{k}}$, on the other hand, define initial conditions which may depend on direction.

where the constraint on the overall normalisation of the mode functions, $v_k\dot{v}_k^* - \dot{v}_k v_k^* \equiv i$, implies

$$|c_1|^2 - |c_2|^2 = \frac{1}{2k}\,. \tag{189}$$

At this point, we still have a one-parameter family of solutions for the mode function $v_k(\tau)$. A change in $v_k(\tau)$ could be accompanied by a change in $\hat{a}_{\mathbf{k}}$ that keeps the solution $\hat{v}_{\mathbf{k}}(\tau)$ unchanged. Each such solution corresponds to a different vacuum state, cf. eq. (187). However, a special choice of $v_k(\tau)$ is selected if we require the vacuum state $|0\rangle$ to be the ground state of the Hamiltonian.

To see this, consider the Hamiltonian operator

$$\hat{H} = \int \mathrm{d}^3x \left[\frac{1}{2}\hat{\pi}^2 + \frac{1}{2}(\nabla\hat{v})^2 + \frac{1}{2}\frac{\ddot{a}}{a}\,\hat{v}^2\right]. \tag{190}$$

Substituting the mode expansion (185) into this, we find

$$\hat{H} = \frac{1}{4}\int \mathrm{d}^3k \left[a_{\mathbf{k}}a_{-\mathbf{k}}F_k^* + a_{\mathbf{k}}^\dagger a_{-\mathbf{k}}^\dagger F_k + \Big(2a_{\mathbf{k}}^\dagger a_{\mathbf{k}} + \delta_D(0)\Big)\,E_k\right], \tag{191}$$

where

$$E_k(\tau) = |\dot{v}_k|^2 + \omega_k^2|v_k|^2\,, \tag{192}$$

$$F_k(\tau) = \dot{v}_k^2 + \omega_k^2 v_k^2\,. \tag{193}$$

The vacuum expectation value of the Hamiltonian is

$$\langle 0|\hat{H}|0\rangle = \frac{1}{4}\delta_D(0)\int \mathrm{d}^3k\, E_k(\tau)\,, \tag{194}$$

where the divergent factor $\delta_D(0)$ is an artefact of integrating over an infinite volume. The energy density of the vacuum is

$$\varepsilon_k \equiv \frac{1}{4}\int \mathrm{d}^3k\, E_k(\tau)\,. \tag{195}$$

At early times, we have

$$\lim_{\tau\to-\infty} v_k(\tau) = c_1\, e^{-ik\tau} + c_2\, e^{ik\tau}\,, \tag{196}$$

$$\lim_{\tau\to-\infty} E_k(\tau) = 4\left(|c_1|^2 + |c_2|^2\right)k^2\,. \tag{197}$$

Given the constraint (604), the function in (197) is minimized for

$$|c_1| = \frac{1}{\sqrt{2k}}\,, \quad c_2 = 0\,. \tag{198}$$

Up to an irrelevant phase, this uniquely determines the *Bunch-Davies* mode function

$$\boxed{v_k(\tau) = \frac{1}{\sqrt{2k}}\left(1 - \frac{i}{k\tau}\right)e^{-ik\tau}}\,. \tag{199}$$

Note that (198) implies

$$\lim_{\tau\to-\infty} E_k(\tau) = 2k\,, \tag{200}$$

$$\lim_{\tau\to-\infty} F_k(\tau) = 0\,, \tag{201}$$

and hence

$$\lim_{\tau\to-\infty} \hat{H} = \int \mathrm{d}^3k \left[a_{\mathbf{k}}^\dagger a_{\mathbf{k}} + \frac{1}{2}\delta_D(0)\right]\hbar\omega_k\,, \tag{202}$$

where we have reinstated Planck's constant $\hbar$. We see that the vacuum state $|0\rangle$ is the state of minimum energy $\frac{1}{2}\hbar\omega_k$. If any function other than (199) had been chosen as the mode function, then the state annihilated by $\hat{a}_{\mathbf{k}}$ would *not* be the ground state of the oscillator.

Zero-point fluctuations

Finally, we can predict the quantum statistics of the operator

$$\hat{v}(\tau,\mathbf{x}) = \int \frac{\mathrm{d}^3k}{(2\pi)^3}\left[v_k(\tau)\,\hat{a}_{\mathbf{k}} + v_k^*(\tau)\,a^\dagger_{-\mathbf{k}}\right]e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{203}$$

The expectation value of $\hat{v}$ vanishes, i.e. $\langle\hat{v}\rangle \equiv \langle 0|\hat{v}|0\rangle = 0$. However, the variance of inflaton fluctuations receives non-zero quantum fluctuations:

$$\begin{aligned}
\langle|\hat{v}|^2\rangle &\equiv \langle 0|\hat{v}(\tau,\mathbf{0})\hat{v}(\tau,\mathbf{0})|0\rangle \\
&= \int \frac{\mathrm{d}^3k}{(2\pi)^3}\int \frac{\mathrm{d}^3k'}{(2\pi)^3}\,\langle 0|\big)\overbrace{v_k^*(\tau)\hat{a}^\dagger_{\mathbf{k}} + v_k(\tau)\hat{a}_{\mathbf{k}}\big)\big(v_{k'}(\tau)\hat{a}_{\mathbf{k}'} + v^*_{k'}(\tau)\hat{a}^\dagger_{\mathbf{k}'}\big)}|0\rangle \\
&= \int \frac{\mathrm{d}^3k}{(2\pi)^3}\int \frac{\mathrm{d}^3k'}{(2\pi)^3}\,v_k(\tau)v^*_{k'}(\tau)\,\langle 0|[\hat{a}_{\mathbf{k}},\hat{a}^\dagger_{\mathbf{k}'}]|0\rangle \\
&= \int \frac{\mathrm{d}^3k}{(2\pi)^3}\,|v_k(\tau)|^2 \\
&= \int \mathrm{d}\ln k\;\frac{k^3}{2\pi^2}|v_k(\tau)|^2.
\end{aligned} \tag{204}$$

We define the (dimensionless) *power spectrum* as

$$\boxed{\mathcal{P}_v(k,\tau) \equiv \frac{k^3}{2\pi^2}|v_k(\tau)|^2}. \tag{205}$$

We see that the square of the classical solution determines the variance of quantum fluctuations. Substituting the Bunch-Davies mode function (199) into (205), we find

$$\mathcal{P}_\varphi(k,\tau) = \frac{\mathcal{P}_v(k,\tau)}{a^2(\tau)} = \left(\frac{H}{2\pi}\right)^2\left[1+(k\tau)^2\right] \xrightarrow{k\tau\to 0} \left(\frac{H}{2\pi}\right)^2. \tag{206}$$

Note that in the superhorizon limit, $k\tau \to 0$, the dimensionless power spectrum $\mathcal{P}_\varphi$ approaches the same constant for all momenta. This is the characteristic of a *scale-invariant* spectrum.

Massive fields

The above discussion is easily generalized to massive spectator fields. The modified equation of motion is

$$\ddot{v}_k + \left(k^2 + m^2a^2 - \frac{\ddot{a}}{a}\right)v_k = 0\,, \tag{207}$$

which, in de Sitter space, becomes

$$\ddot{v}_k + \left(k^2 - \frac{\nu^2-1/4}{\tau^2}\right)v_k = 0\,, \quad \text{where} \quad \nu^2 \equiv \frac{9}{4} - \frac{m^2}{H^2}\,. \tag{208}$$

The most general solution of (208) is

$$v_k(\tau) = \sqrt{-\tau}\,\left[c_1 H_\nu^{(1)}(|k\tau|) + c_2 H_\nu^{(2)}(|k\tau|)\right], \tag{209}$$

where $H_\nu^{(1)}$ and $H_\nu^{(2)}$ are Hankel functions of the first and second kind. Imposing the Bunch-Davies initial condition, we find

$$c_1 = \frac{\sqrt{\pi}}{2}e^{i(2\nu+1)\frac{\pi}{4}}\,, \quad c_2 = 0\,. \tag{210}$$

For the moment, let us assume $m < \frac{3}{2}H$, so that ν is real. We then find

$$\mathcal{P}_\varphi(k,\tau) = \frac{\mathcal{P}_v(k,\tau)}{a^2(\tau)} \xrightarrow{k\tau\to 0} \left(\frac{H}{2\pi}\right)^2 (k\tau)^{3/2-\nu} . \tag{211}$$

We observe that the superhorizon limit of the spectrum is scale-dependent and evolves in time. The scale-dependence takes a power law form with the following *spectral index*

$$n_\varphi \equiv \frac{d\ln \mathcal{P}_\varphi}{d\ln k} = \frac{3}{2} - \nu \xrightarrow{m\ll H} \frac{1}{3}\frac{m^2}{H^2} . \tag{212}$$

In the limit $m \to 0$, we recover the scale-invariant spectrum of a massless field in de Sitter.

For $m > \frac{3}{2}H$, the degree of the Hankel function becomes imaginary, $\nu \equiv i\mu$, where

$$\mu \equiv \sqrt{m^2/H^2 - 9/4} \xrightarrow{m\gg H} m/H . \tag{213}$$

What used to be an irrelevant phase factor in (210), now becomes an exponential suppression of the amplitude of the mode function

$$|v_k(\tau)| = \frac{\sqrt{\pi}}{2} e^{-\pi\mu/2}\sqrt{-\tau}\,|H^{(1)}_{i\mu}(|k\tau|)| . \tag{214}$$

The power spectrum of very massive fields in de Sitter is therefore highly suppressed, $\mathcal{P}_\varphi \propto e^{-\pi m/H}$, for $m \gg \frac{3}{2}H$.

Fluctuations during Inflation

We now move to studying the fluctuations in the inflaton field during inflation. These fluctuations cannot be treated independently from fluctuations in the metric, since the two are coupled by the Einstein equations. This leads to some technical complications, but conceptually the quantization of the coupled inflaton-metric fluctuations is the same as before.

Metric fluctuations

We will treat the fluctuations of the metric in the so-called *ADM formalism* [168]. We start by writing the perturbed line element as

$$\mathrm{d}s^2 = -N^2\mathrm{d}t^2 + h_{ij}(N^i\mathrm{d}t + \mathrm{d}x^i)(N^j\mathrm{d}t + \mathrm{d}x^j) , \tag{215}$$

where $N \equiv N(t,\mathbf{x})$ is the lapse function, $N^i \equiv N^i(t,\mathbf{x})$ is the shift vector, and $h_{ij} \equiv h_{ij}(t,\mathbf{x})$ is the induced metric on three-dimensional hypersurfaces of constant time t. The geometry of the spatial slices is characterized by the intrinsic curvature, $R^{(3)}_{ij}$, i.e. the Ricci tensor of the induced metric, and by the extrinsic curvature

$$K_{ij} \equiv \frac{1}{2N}\left(h'_{ij} - \nabla_i N_j - \nabla_j N_i\right) \equiv \frac{1}{N}E_{ij} . \tag{216}$$

The four-dimensional Ricci scalar, R, can be written in terms of the three-dimensional Ricci scalar, $R^{(3)}$, and the extrinsic curvature tensor as

$$R = R^{(3)} + N^{-2}\left(E^{ij}E_{ij} - E^2\right) , \tag{217}$$

where indices are raised with h^{ij}, and $E \equiv h^{ij}E_{ij}$.

The inflaton action can then be written as

$$S = \frac{1}{2}\int \mathrm{d}^4x\sqrt{-g}\left[R - g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - 2V\right]$$
$$= \frac{1}{2}\int \sqrt{h}N\left[R^{(3)} - 2V + N^{-2}\left(E^{ij}E_{ij} - E^2\right) + N^{-2}\left(\phi' - N^i\partial_i\phi\right)^2 - h^{ij}\partial_i\phi\partial_j\phi\right], \quad (218)$$

where for the moment we have set $M_{\rm pl} \equiv 1$. Note that N and N_i do not have time derivatives acting on them and are therefore non-dynamical fields that will be fixed by constraint equations. Indeed, varying the action with respect to N and N^i, we find

$$R^{(3)} - 2V - h^{ij}\partial_i\phi\partial_j\phi - N^2[E_{ij}E^{ij} - E^2 - (\phi' - N^i\partial_i\phi)^2] = 0\,, \quad (219)$$
$$\nabla_i[N^{-1}(E^i_j - E\delta^i_j)] = 0\,. \quad (220)$$

Plugging the solutions for N and N^i back into the action leaves ϕ and h_{ij} as the only dynamical variables. We will perform this procedure in perturbation theory. For the time being, we will focus on scalar perturbations. To fix time and space reparameterizations, we will have to choose a gauge for ϕ and h_{ij}. We will present the results for two different gauges: spatially flat gauge and comoving gauge.

Spatially flat gauge

As the name suggests, in spatially flat gauge the induced metric is taken to be unperturbed:

$$h_{ij} = a^2\delta_{ij}\,. \quad (221)$$

We then consider the perturbations of the inflaton, the lapse and the shift:

$$\phi \equiv \bar{\phi}(t) + \varphi(t,\mathbf{x})\,, \quad N \equiv 1 + \alpha(t,\mathbf{x})\,, \quad N_i \equiv \partial_i\beta(t,\mathbf{x})\,. \quad (222)$$

Substituting (221) and (222) into the constraint equations (219) and (220), we obtain [40] [16]

$$\alpha = \frac{\bar{\phi}'}{2H}\varphi\,, \quad \partial^2\beta = \frac{(\bar{\phi}')^2}{2H^2}\frac{d}{dt}\left(-\frac{H}{\bar{\phi}'}\varphi\right). \quad (223)$$

Plugging this solution into the action (218), expanding to second order and performing a few integrations by parts, we find

$$S_2 = \frac{1}{2}\int \mathrm{d}t\,\mathrm{d}^3x\; a^3\left[(\varphi')^2 - \frac{1}{a^2}(\partial\varphi)^2 - \left[V'' - 2(3\varepsilon - \varepsilon^2 + \varepsilon\eta)H^2\right]\varphi^2\right]. \quad (224)$$

Switching to conformal time, the equation of motion for the Fourier components of the canonically normalized field $v \equiv a\varphi$ is

$$\ddot{v}_k + \left(k^2 + a^2\left[V'' - 6\varepsilon H^2\right] - \frac{\ddot{a}}{a}\right)v_k = 0\,, \quad (225)$$

where we have dropped terms in the effective mass that are higher order in slow-roll parameters. Note that the term $a^2\left[V'' - 6\varepsilon H^2\right] \sim \mathcal{O}(\varepsilon,\eta)(aH)^2$ is always smaller than $\ddot{a}/a = 2(aH)^2$. However, it is the only source of time evolution on superhorizon scales and even a small evolution can accumulate over time. In order not to have to follow the evolving field on superhorizon

[16] To find the quadratic action, we only have to solve the constraints to linear order. The second order terms in N and N_i will multiply the zeroth order constraints which vanish when the background equations of motion are imposed.

scales, it is useful to evaluate the solution at horizon crossing, $k = aH$, and then map it to a field that is known not to evolve outside the horizon. This constant mode is the *comoving curvature perturbation*, which in spatially flat gauge is defined as

$$\zeta = -\frac{H}{\bar{\phi}'}\varphi\,. \tag{226}$$

The power spectrum of ζ is

$$\mathcal{P}_\zeta(k) = \left(\frac{H}{\bar{\phi}'}\right)^2 \mathcal{P}_\varphi(k,\tau)\Big|_{k=aH}\,. \tag{227}$$

Dropping the slow-roll suppressed terms in (225), the equation of motion reduces to that of a massless field in de Sitter space. The power spectrum $\mathcal{P}_\varphi(k,\eta)$ is therefore given by (206) and we get

$$\boxed{\mathcal{P}_\zeta(k) = \left(\frac{H}{\bar{\phi}'}\right)^2 \left(\frac{H}{2\pi}\right)^2\Big|_{k=aH}}\,. \tag{228}$$

Another advantage of using ζ rather than φ to characterize the initial conditions is that ζ remains well-defined after inflation.

Comoving gauge

In comoving gauge the inflaton field is taken to be unperturbed:

$$\phi = \bar{\phi}(t)\,. \tag{229}$$

All perturbations are then carried by the metric:

$$h_{ij} = a^2(1+2\zeta(t,\mathbf{x}))\delta_{ij}\,, \quad N \equiv 1+\alpha(t,\mathbf{x})\,, \quad N_i \equiv \partial_i\beta(t,\mathbf{x})\,, \tag{230}$$

where ζ is the comoving curvature perturbation. Substituting this ansatz into the constraint equations (219) and (220), we obtain [40]

$$\alpha = \frac{\zeta'}{H}\,, \quad \partial^2\beta = -\frac{\partial^2\zeta}{H} + a^2\frac{(\bar{\phi}')^2}{2H^2}\zeta'\,. \tag{231}$$

Plugging this solution into the action (218), expanding to second order, performing integrations by parts and using the background equations of motion, we get a remarkably simple result

$$S_2 = \int \mathrm{d}t\,\mathrm{d}^3x\,a^3\varepsilon\left((\zeta')^2 - \frac{1}{a^2}(\partial\zeta)^2\right). \tag{232}$$

Note that there is no mass term, so ζ is conserved outside the horizon. Consider the equation of motion

$$\zeta_k'' + (3+\eta)H\zeta_k' + \frac{k^2}{a^2}\zeta_k = 0\,. \tag{233}$$

On super-Hubble scales, $k \ll aH$, this becomes $\zeta_k'' + (3+\eta)H\zeta_k' \approx 0$, which clearly has a constant mode as a solution.

The equation of motion of the canonically normalized field, $v \equiv a\sqrt{2\varepsilon}\,\zeta \equiv z\,\zeta$, is

$$\ddot{v}_k + \left(k^2 - \frac{\ddot{z}}{z}\right)v_k = 0\,, \tag{234}$$

where the effective mass can be written as

$$\frac{\ddot{z}}{z} = \frac{\nu^2 - 1/4}{\tau^2}\,, \quad \text{with} \quad \nu \approx \frac{3}{2} + \varepsilon + \frac{\eta}{2}\,. \tag{235}$$

Recall that we have seen eq. (234) before when we considered a massive scalar field in de Sitter space, cf. eq. (208). The Bunch-Davies mode function therefore is

$$|v_k(\tau)| = \frac{\sqrt{\pi}}{2}\sqrt{-\tau}\,|H_\nu^{(1)}(-k\tau)| \xrightarrow{k\tau\to 0} \frac{2^\nu \Gamma(\nu)}{2\sqrt{\pi}}\frac{1}{\sqrt{k}}\,(-k\tau)^{-\nu+1/2}\,, \tag{236}$$

and the superhorizon limit of the power spectrum of ζ is

$$\begin{aligned}\mathcal{P}_\zeta(k) \equiv \lim_{k\tau\to 0}\frac{k^3}{2\pi^2}|\zeta_k(\tau)|^2 &= \frac{1}{z^2(\tau)}\lim_{k\tau\to 0}\frac{k^3}{2\pi^2}|v_k(\tau)|^2 \\ &= \frac{1}{2a^2\varepsilon}\frac{k^2}{4\pi^2}\,(-k\tau)^{-2\nu+1} \;=\; \frac{1}{16\pi^2}\frac{H^2(\tau)}{\varepsilon(\tau)}\,(-k\tau)^{-2\nu+3}\,.\end{aligned} \tag{237}$$

Note that the time dependence of $H(\tau)$ and $\varepsilon(\tau)$ precisely cancels the time dependence of the final factor in (237), so that the power spectrum is time independent. Let k_* be a reference scale that exits the horizon at time $\tau_* = -1/k_*$. Equation (237) can then be written as

$$\boxed{\mathcal{P}_\zeta(k) = \frac{1}{8\pi^2}\frac{1}{\varepsilon_*}\frac{H_*^2}{M_{\rm pl}^2}\,(k/k_*)^{-2\varepsilon_*-\eta_*}} \tag{238}$$

where $H_* \equiv H(\tau_*)$ and $\varepsilon_* \equiv \varepsilon(\tau_*)$.

Curvature Perturbations

For ease of reference, we briefly summarize the results of the previous section. The power spectrum of the curvature perturbation ζ, cf. eqs. (228) and (238), takes a power law form

$$\mathcal{P}_\zeta(k) = A_{\rm s}\left(\frac{k}{k_*}\right)^{n_{\rm s}-1}\,, \tag{239}$$

where the amplitude and the spectral index are

$$A_{\rm s} \equiv \frac{1}{8\pi^2}\frac{1}{\varepsilon_*}\frac{H_*^2}{M_{\rm pl}^2}\,, \tag{240}$$

$$n_{\rm s} \equiv 1 - 2\varepsilon_* - \eta_*\,. \tag{241}$$

The observational constraint on the scalar spectral index is $n_{\rm s} = 0.9603 \pm 0.0073$. The observed percent-level deviation from the scale-invariant value, $n_{\rm s} = 1$, are the first direct measurement of time dependence in the inflationary dynamics.

Exercise.—Show that for slow-roll inflation, the results (240) and (460) can be written as

$$A_{\rm s} = \frac{1}{24\pi^2}\frac{1}{\varepsilon_V}\frac{V}{M_{\rm pl}^4}\,, \tag{242}$$

$$n_{\rm s} = 1 - 6\varepsilon_V + 2\eta_V\,, \tag{243}$$

where ε_V and η_V are the potential slow-roll parameters defined in (165). This expresses the amplitude of curvature perturbations and the spectral index in terms of the shape of the inflaton potential.

Gravitational Waves

Arguably the cleanest prediction of inflation is a spectrum of primordial gravitational waves. These are tensor perturbations to the spatial metric,

$$\mathrm{d}s^2 = a^2(\tau)\left[-\mathrm{d}\tau^2 + (\delta_{ij} + 2\gamma_{ij})\mathrm{d}x^i\mathrm{d}x^j\right], \tag{244}$$

where γ_{ij} is transverse and traceless. Substituting (244) into the Einstein-Hilbert action and expanding to second order gives

$$S = \frac{M_{\rm pl}^2}{2}\int \mathrm{d}^4x\sqrt{-g}\,R \qquad \Rightarrow \qquad S_2 = \frac{M_{\rm pl}^2}{8}\int \mathrm{d}\tau\,\mathrm{d}^3x\,a^2\left[\dot\gamma_{ij}^2 - (\partial\gamma_{ij})^2\right]. \tag{245}$$

It is convenient to use rotational symmetry to align the z-axis of the coordinate system with the momentum of the mode, i.e. $\mathbf{k} \equiv (0,0,k)$, and write

$$\frac{M_{\rm pl}}{2}\,a\gamma_{ij} \equiv \frac{1}{\sqrt{2}}\begin{pmatrix} v_+ & v_\times & 0 \\ v_\times & -v_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{246}$$

The action (245) then becomes

$$S_2 = \frac{1}{2}\sum_{\lambda=+,\times}\int \mathrm{d}\tau\,\mathrm{d}^3x\left[\dot v_\lambda^2 - (\partial v_\lambda)^2 + \frac{\ddot a}{a}v_\lambda^2\right], \tag{247}$$

which is just two copies of the action of a massless scalar field, one for each polarization mode of the gravitational wave, $v_{+,\times}$. The equation of motion for each polarization is

$$\ddot v_k + \left(k^2 - \frac{\ddot a}{a}\right)v_k = 0\,, \tag{248}$$

where the effective mass can be written as

$$\frac{\ddot a}{a} = \frac{\nu^2 - 1/4}{\tau^2}\,, \quad \text{with} \quad \nu \approx \frac{3}{2} + \varepsilon\,. \tag{249}$$

The Bunch-Davies mode function is then given by (236). The superhorizon limit of the power spectrum of the tensor flucutations then is

$$\begin{aligned}\mathcal{P}_\gamma(k) = 2\times\lim_{k\tau\to 0}\frac{k^3}{2\pi^2}|\gamma_k(\tau)|^2 &= 2\times\left(\frac{2}{aM_{\rm pl}}\right)^2\lim_{k\tau\to 0}\frac{k^3}{2\pi^2}|v_k(\tau)|^2 \\ &= \frac{2}{\pi^2}\frac{H^2(\tau)}{M_{\rm pl}^2}\,(-k\tau)^{-2\nu+3}\,,\end{aligned} \tag{250}$$

where the factor of 2 accounts for the sum over the two polarization modes. Introducing the reference scale k_*, this can be written as

$$\boxed{\mathcal{P}_\gamma(k) = \frac{2}{\pi^2}\frac{H_*^2}{M_{\rm pl}^2}(k/k_*)^{-2\varepsilon_*}}\,. \tag{251}$$

This result is arguably the most robust and model-independent prediction of inflation. We see that the form of the tensor power spectrum is again a power law,

$$\mathcal{P}_\zeta(k) = A_{\rm t}\left(\frac{k}{k_*}\right)^{n_{\rm t}}\,, \tag{252}$$

where the amplitude and the spectral index are

$$A_{\rm t} \equiv \frac{2}{\pi^2}\frac{H_*^2}{M_{\rm pl}^2}\,, \tag{253}$$

$$n_{\rm t} \equiv -2\varepsilon_*\,. \tag{254}$$

Notice that the tensor amplitude is a direct measure of the expansion rate H during inflation. This is in contrast to the scalar amplitude which depends on both H and ε. The tensor tilt is a direct measure of ε, whereas the scalar tilt depends both on ε and η. Observationally, a small value for $n_{\rm t}$ is hard to distinguish from zero. The tensor amplitude is often normalized with respect to the measured scalar amplitude, $A_{\rm s} = (2.196 \pm 0.060) \times 10^{-9}$ (at $k_* = 0.05$ Mpc^{-1}):

$$\boxed{r \equiv \frac{A_{\rm t}}{A_{\rm s}} = 16\varepsilon_*} \tag{255}$$

where r is the *tensor-to-scalar ratio*. Inflationary models make predictions for $(n_{\rm s}, r)$. The latest observational constraints on these parameters are shown in Fig. 80.

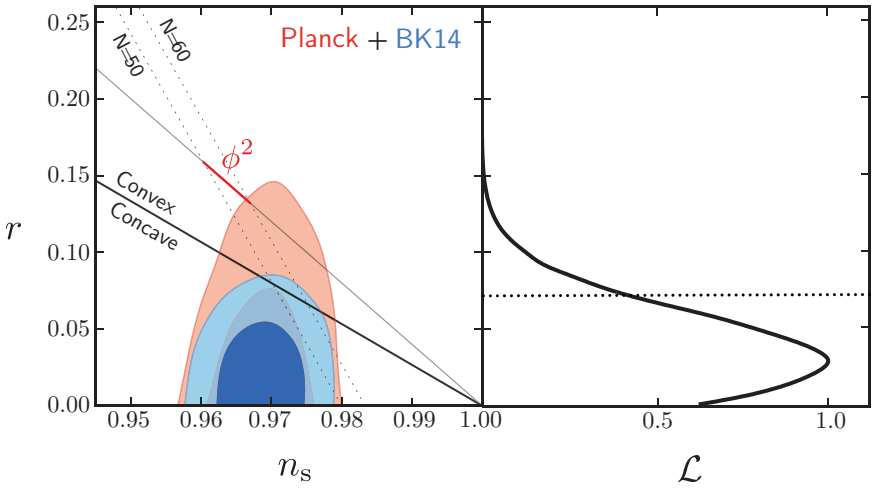


Figure 43: Current constraints on $n_{\rm s}$ and r from CMB measurements of Planck and BICEP [41].

Case study: $m^2\phi^2$ inflation.—In Section 6.6.1, we showed that the slow-roll parameters of $m^2\phi^2$ inflation are

$$\varepsilon_V(\phi) = \eta_V(\phi) = 2\left(\frac{M_{\rm pl}}{\phi}\right)^2\,, \tag{256}$$

and the number of e-folds before the end of inflation is

$$N(\phi) = \frac{\phi^2}{4M_{\rm pl}^2} - \frac{1}{2} \approx \frac{\phi^2}{4M_{\rm pl}^2}\,. \tag{257}$$

At the time when the CMB fluctuations crossed the horizon at $\phi = \phi_*$, we have

$$\varepsilon_{V,*} = \eta_{V,*} \approx \frac{1}{2N_*}\,. \tag{258}$$

The spectral tilt and the tensor-to-scalar ratio therefore are

$$n_{\rm s} \equiv 1 - 6\varepsilon_{V,*} + 2\eta_{V,*} = 1 - \frac{2}{N_*} \approx 0.97\,, \tag{259}$$

$$r \equiv 16\varepsilon_{V,*} = \frac{8}{N_*} \approx 0.13\,, \tag{260}$$

where the final equalities are for $N_* \approx 60$. Comparison with Fig. 80 shows that these predictions are in conflict with the latest CMB data.

The Lyth bound.—During $m^2\phi^2$ inflation, the inflaton field value varies by $\Delta\phi \approx 15\, M_{\rm pl}$ from the time the mode observed in the CMB exited the horizon until the end of inflation. This super-Planckian field variation is, in fact, a general feature of all inflationary models with observable gravitational waves.
Combining (161) and (255), the tensor-to-scalar ratio can be related to the evolution of the inflaton:

$$r = \frac{8}{M_{\rm pl}^2}\left(\frac{d\phi}{dN}\right)^2 , \tag{261}$$

where $\mathrm{d}N = H\mathrm{d}t$. The total field excursion between the time when CMB fluctuations exited the horizon at N_* and the end of inflation at N_e can therefore be written as the integral

$$\frac{\Delta\phi}{M_{\rm pl}} = \int_{N_e}^{N_*} \mathrm{d}N \sqrt{\frac{r}{8}} . \tag{262}$$

During slow-roll evolution, $r(N)$ doesn't evolve much and one may obtain the following approximate relation, called the *Lyth bound* :

$$\frac{\Delta\phi}{M_{\rm pl}} = \mathcal{O}(1) \times \left(\frac{r}{0.01}\right)^{1/2} , \tag{263}$$

where $r \equiv r(N_*)$ is the tensor-to-scalar ratio on CMB scales. Large values of the tensor-to-scalar ratio, $r > 0.01$, therefore correlate with super-Planckian field excursions, $\Delta\phi > M_{\rm pl}$.

A major goal of current efforts in observational cosmology is to detect the tensor component of the primordial fluctuations. Its amplitude depends on the energy scale of inflation and is therefore not predicted (i.e. it varies between models). While this makes the search for primordial tensor modes difficult, it is also what makes it exciting. Detecting tensors would reveal the energy scale at which inflation occurred, providing an important clue about the physics driving the inflationary expansion.

Most searches for tensors focus on the imprint that tensor modes leave in the polarization of the CMB. Polarization is generated through the scattering of the anisotropic radiation field off the free electrons just before decoupling. The presence of a gravitational wave background creates an anisotropic stretching of the spacetime which induces a special type of polarization pattern, the so-called B-mode pattern (a pattern whose "curl" doesn't vanish). Such a pattern cannot be created by scalar (density) fluctuations and is therefore a unique signature of primordial tensors (gravitational waves). A large number of ground-based, balloon and satellite experiments are currently searching for the B-mode signal predicted by inflation. A B-mode detection would be a milestone towards a complete understanding of the origin of all structure in the universe.

6.6.3 Primordial Interactions

In the previous section, we have computed the two-point function of primordial curvature perturbations, or its Fourier equivalent, the power spectrum. If the initial conditions are drawn from a Gaussian distribution function, then the power spectrum contains all of the information

about the primordial perturbations. In general, however, higher-order correlations can encode a significant amount of new information. In particular, these correlations are sensitive to nonlinear interactions while the power spectrum only probes the free theory.

In this section, I will introduce the basic formalism for computing the non-Gaussianity produced by inflation. I will apply this to single-field slow-roll inflation and show that gravitational nonlinearities produce a robust (but small) non-Gaussian signature. In Section 6.6.4, I will show that extra massive fields can get excited during inflation and that their decays lead to distinctive signatures in cosmological correlation functions.

For further reading I highly recommend the classic papers by Maldacena [40] and Weinberg [89], as well as the wonderfully clear reviews by Chen [43], Wang [44], Lim [45] and Komatsu [46].

Non-Gaussianity

The leading non-Gaussian signature is the three-point correlation function, or its Fourier equivalent, the bispectrum

$$B_\zeta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \equiv \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle \,. \tag{264}$$

For perturbations around an FRW background, the momentum dependence of the bispectrum simplifies considerably: Because of homogeneity, or *translation invariance*, the bispectrum is proportional to a delta function of the sum of the momenta, $B_\zeta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \propto \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)$, i.e. the sum of the momentum three-vectors must form a closed triangle. Because of isotropy, or *rotational invariance*, the bispectrum only depends on the magnitudes of the momentum vectors, but not on their orientations,

$$B_\zeta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)\, B_\zeta(k_1, k_2, k_3) \,. \tag{265}$$

It is convenient to define the dimensionless bispectrum as

$$\mathcal{B}_\zeta(k_1, k_2, k_3) \equiv \frac{(k_1 k_2 k_3)^2}{(2\pi^2)^2} B_\zeta(k_1, k_2, k_3) \,. \tag{266}$$

The *amplitude* of the non-Gaussianity is then defined as the size of the bispectrum in the equilateral momentum configuration:

$$f_{\rm NL}(k) \equiv \frac{5}{18} \frac{\mathcal{B}_\zeta(k, k, k)}{\mathcal{P}_\zeta^2(k)} \,, \tag{267}$$

where we have indicated that $f_{\rm NL}$ can in general depend on the overall momentum. On the other hand, if the fluctuations are *scale-invariant*, then $f_{\rm NL}$ is a constant and we can write

$$\mathcal{B}_\zeta(k_1, k_2, k_3) \equiv \frac{18}{5} f_{\rm NL} \times \mathcal{S}(x_2, x_3) \times \mathcal{P}_\zeta^2 \,, \tag{268}$$

where $x_2 \equiv k_2/k_1$ and $x_3 \equiv k_3/k_1$. The shape function $\mathcal{S}(x_2, x_3)$ is normalized so that $\mathcal{S}(1, 1) \equiv 1$.

As we will discuss in detail below, the *shape* of the non-Gaussianity contains a lot of information about the microphysics of inflation. A divergence of the signal for squeezed triangles is a signature for extra degrees of freedom during inflation (see Fig. 83 and Sec. 6.6.4), while a peak in the signal for equilateral triangles arises from higher-derivative inflaton self-interactions (see Fig. 45).

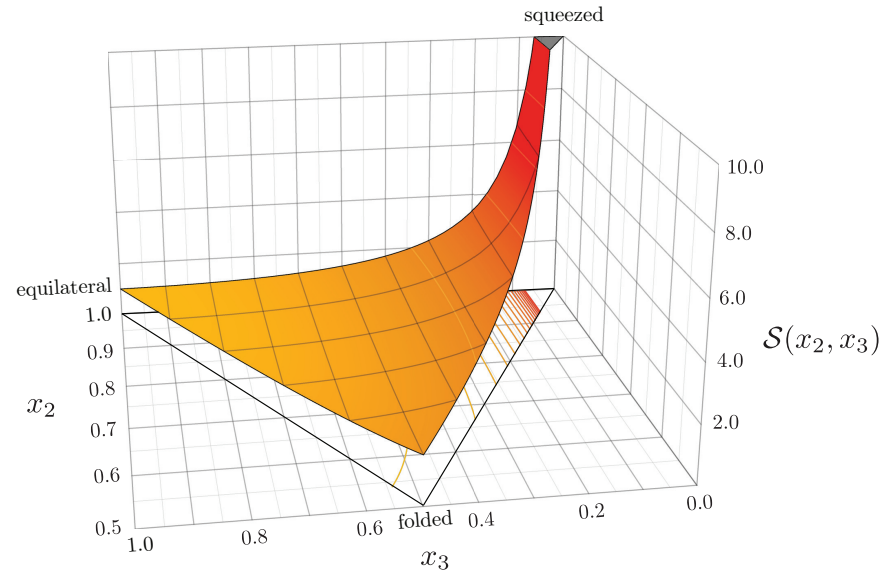


Figure 44: Bispectrum of local non-Gaussianity. The signal is peaked for squeezed triangles.

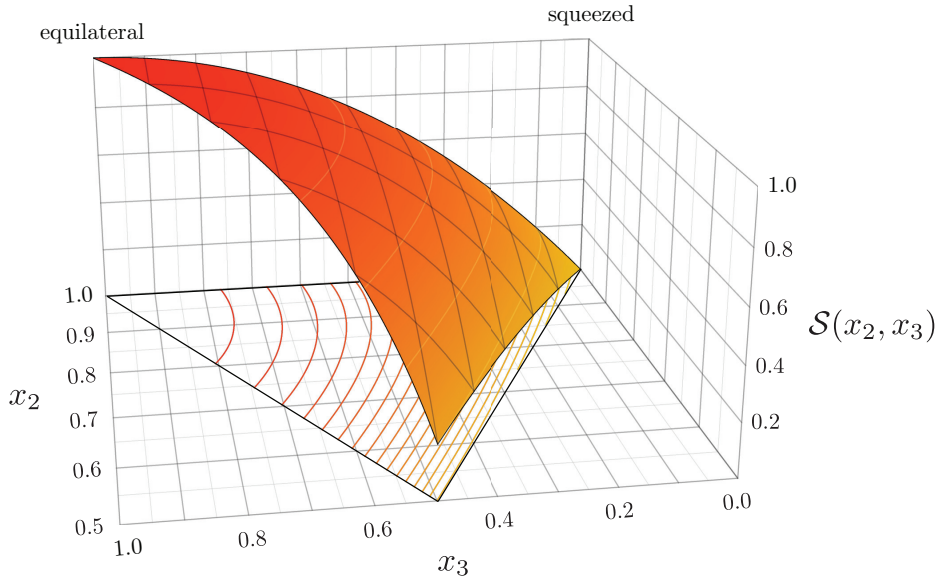


Figure 45: Bispectrum for the interaction $\dot{\zeta}^3$. The signal is peaked for equilateral triangles.

In-In Formalism

The problem of computing correlation functions in cosmology differs in important ways from the corresponding analysis of quantum field theory applied to particle physics.

In particle physics, the main observable is the S-matrix, i.e. the transition probability for a state $|in\rangle$ in the far past to become some state $|out\rangle$ in the far future,

$$\langle out|\, S|in\rangle = \langle out(+\infty)|in(-\infty)\rangle. \tag{269}$$

The scattering particles are taken to be non-interacting at very early and very late times, when they are far from the interaction region, and the asymptotic states can be taken to be vacuum states of the free Hamiltonian H_0.

In cosmology, on the other hand, the task is to determine the expectation values of products of operators *at a fixed time*. Boundary conditions are only imposed at very early times when their wavelengths are much smaller than the horizon and the interaction picture fields should have the same form as in Minkowski space. As we have seen in the previous section, this

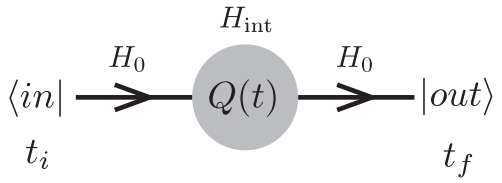


Figure 46: Particle physicists compute in-out transition amplitudes.

leads to the definition of the Bunch–Davies vacuum. In this section, we will describe the *in-in* formalism[17] to compute cosmological correlation functions as expectation values in two $|in\rangle$ states.

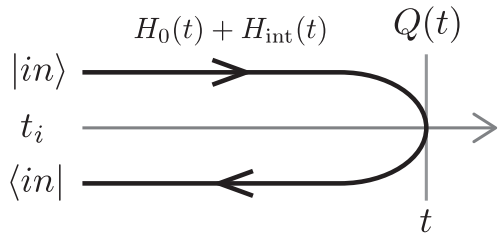


Figure 47: Cosmologists compute in-in expectation values.

Our goal is to compute n-point functions of the curvature perturbation ζ or the gravitational wave polarization modes $\gamma^\times$ and γ^+. We will collectively denote these fluctuations by the field $\psi = \{\zeta, \gamma^\times, \gamma^+\}$ and consider expectation values of operators such as $\hat{Q} = \hat{\psi}_{\mathbf{k}_1}\hat{\psi}_{\mathbf{k}_2}\cdots\hat{\psi}_{\mathbf{k}_n}$,

$$\langle \hat{Q}(\tau)\rangle = \langle in|\, \hat{Q}(\tau)\, |in\rangle\,, \tag{270}$$

where $|in\rangle$ is the vacuum of the interacting theory at some moment τ_i in the far past, and $\tau > \tau_i$ is some later time, such as horizon crossing or the end of inflation. To compute the matrix element in (270) we evolve $Q(\tau)$ back to τ_i using the perturbed Hamiltonian δH. Computing this time evolution is complicated by the interactions inside of $\delta H = H_0 + H_{\rm int}$, which lead to nonlinear equations of motion. We therefore introduce the *interaction picture* in which the leading time dependence of the fields is determined by the quadratic Hamiltonian H_0 (or linear equations of motion). Corrections arising from the interactions are then treated as a power series in $H_{\rm int}$.

This leads to the following important result:

$$\boxed{\langle \hat{Q}(\tau)\rangle \;=\; \langle 0|\, \bar{T} e^{\,i\int_{-\infty(1-i\epsilon)}^{\tau} \hat{H}^I_{\rm int}(\tau')\,{\rm d}\tau'}\, \hat{Q}^I(\tau)\, T e^{\,-i\int_{-\infty(1+i\epsilon)}^{\tau} \hat{H}^I_{\rm int}(\tau'')\,{\rm d}\tau''}\, |0\rangle} \tag{271}$$

where T ($\bar{T}$) is the (anti-)time-ordering symbol. Note that both $\hat{Q}^I$ and $\hat{H}^I_{\rm int}$ are evaluated using interaction picture operators. The standard $i\epsilon$ prescription has been used to effectively turn off the interaction in the far past and project the interacting $|in\rangle$ state onto the free vacuum $|0\rangle$. By expanding the exponentials, we can compute the correlation function perturbatively in $\hat{H}_{\rm int}$. For example, at leading order, we find

$$\boxed{\langle \hat{Q}(\tau)\rangle \;=\; -i\int_{-\infty}^{\tau} {\rm d}\tau'\; \langle 0|\, [\hat{Q}^I(\tau), \hat{H}^I_{\rm int}(\tau')]\, |0\rangle}\,. \tag{272}$$

[17] This is also referred to as the Schwinger–Keldysh formalism [86]. The use of the *in-in* formalism in cosmology was pioneered in [87, 88, 40, 89] (see also [169, 170]) and is reviewed in [43, 44].

We can use Feynman diagrams to organize the power series, drawing interaction vertices for every power of $\hat{H}_{\rm int}$.

In the following insert, I will derive the *in-in* master formula (271). Readers who are more interested in applications of the result, may skip this part.

Derivation.—The time evolution of the operators in the Heisenberg picture is determined by

$$\frac{d\hat{\psi}}{d\tau} = i[\hat{H}, \hat{\psi}]\,, \quad \frac{d\hat{p}_\psi}{d\tau} = i[\hat{H}, \hat{p}_\psi]\,, \tag{273}$$

where $\hat{H} = \hat{H}_0 + \hat{H}_{\rm int}$ is the perturbed Hamiltonian. This time evolution is complicated by the interactions inside $\hat{H}$, which lead to nonlinear equations of motion. We therefore introduce the *interaction picture* in which the leading time dependence of the fields is determined by the quadratic Hamiltonian $\hat{H}_0$ (or, equivalently, by the linear equations of motion):

$$\frac{d\hat{\psi}^I}{d\tau} = i[\hat{H}_0, \hat{\psi}^I] \quad \frac{d\hat{p}^I_\psi}{d\tau} = i[\hat{H}_0, \hat{p}^I_\psi]\,. \tag{274}$$

The solution to these equations can be written as

$$\hat{\psi}^I_{\mathbf{k}}(\tau) = \psi^I_k(\tau)\,\hat{a}_{\mathbf{k}} + h.c.\,, \tag{275}$$

where $\psi^I_k(\tau)$ is the solution to the free-field equation and the operators $\hat{a}_{\mathbf{k}}$ define the free-field vacuum $|0\rangle$. Corrections to the evolution of the operators can then be treated perturbatively in $\hat{H}_{\rm int}$. Relatively straightforward algebraic manipulations of (273) and (274) allow us to express an operator in the Heisenberg picture in terms of operators in the interaction picture [89],

$$\hat{Q}(\tau) = \hat{F}^{-1}(\tau, \tau_i)\,\hat{Q}^I(\tau)\,\hat{F}(\tau, \tau_i)\,, \tag{276}$$

where

$$\hat{F}(\tau, \tau_i) \equiv T e^{-i\int_{\tau_i}^{\tau} \hat{H}^I_{\rm int}(\tau'')\,{\rm d}\tau''}\,. \tag{277}$$

We can think of $\hat{F}(\tau, \tau_i)$ as an operator evolving quantum states in the interaction picture,

$$|\Omega(\tau)\rangle = \hat{F}(\tau, \tau_i)|\Omega(\tau_i)\rangle\,, \tag{278}$$

where $|\Omega(\tau_i)\rangle \equiv |\Omega\rangle$. We would like to relate the vacuum of the interacting theory, $|\Omega\rangle$, to the vacuum of the free theory, $|0\rangle$. Inserting a complete set of energy eigenstates $\{|\Omega\rangle, |n\rangle\}$ of the full theory, where $|n\rangle$ are the excited states, we have

$$|0\rangle = |\Omega\rangle\langle\Omega|0\rangle + \sum_n |n\rangle\langle n|0\rangle\,, \tag{279}$$

and correspondingly

$$e^{-i\hat{H}(\tau-\tau_i)}|0\rangle = e^{-i\hat{H}(\tau-\tau_i)}|\Omega\rangle\langle\Omega|0\rangle + \sum_n e^{-iE_n(\tau-\tau_i)}|n\rangle\langle n|0\rangle\,. \tag{280}$$

Adding a small imaginary part to the initial time, $\tau_i \to -\infty(1-i\epsilon) \equiv -\infty^-$, will project out the excited states, $e^{-iE_n(\tau-\tau_i)} \to e^{-\infty\times\epsilon E_n}(\cdots) \to 0$. We are then left with

$$\hat{F}(\tau, -\infty^-)|\Omega\rangle = \frac{\hat{F}(\tau, -\infty^-)|0\rangle}{\langle\Omega|0\rangle}\,. \tag{281}$$

The $i\epsilon$ prescription has effectively turned off the interactions in the far past and projected the interacting vacuum $|\Omega\rangle$ onto the free vacuum $|0\rangle$. Setting $|\langle\Omega|0\rangle| \to 1$, we arrive at the *in-in* master formula

$$\langle\hat{Q}(\tau)\rangle \;=\; \langle 0|\,\bar{T} e^{i\int_{-\infty^+}^{\tau} \hat{H}^I_{\rm int}(\tau')\,{\rm d}\tau'}\;\hat{Q}^I(\tau)\;T e^{-i\int_{-\infty^-}^{\tau} \hat{H}^I_{\rm int}(\tau'')\,{\rm d}\tau''}\,|0\rangle\,, \tag{282}$$

where $\infty^\pm \equiv \infty(1\pm i\epsilon)$. The integration contour goes from $-\infty(1-i\epsilon)$ to τ (where the correlation function is evaluated) and back to $-\infty(1+i\epsilon)$.

Gravitational Floor

Let us now apply this formalism to the calculation of the bispectrum of curvature perturbations in single-field slow-roll inflation [40]. The nonlinearities of the gravitational evolution produce a minimal amount of non-Gaussianity which we will call the "gravitational floor". Unfortunately, the amplitude of the signal is too small to be detectable in the foreseeable future.

To compute the bispectrum we need to expand the inflationary action (218) to third order in perturbations. It is convenient to perform the computation in comoving gauge, cf. eqs. (229) and (230). In §6.6.2, we solved the lapse N and the shift N_i to first order in ζ. This is sufficient also for the cubic action since second-order and third-order perturbations in N and N_i will multiply the first-order and zeroth-order constraint equation, respectively. Substituting (231) into (218) and expanding to third order, we find [40]

$$S_3 = \frac{1}{2}\int \mathrm{d}t\,\mathrm{d}^3x\,a^3\Big(2\varepsilon^2\,\zeta(\zeta')^2 - 4a^{-2}\varepsilon^2\,\zeta'(\partial\zeta)(\partial\zeta) + 2a^{-2}\varepsilon^2\,\zeta(\partial\zeta)^2 \\ + a^{-2}\varepsilon\eta'\,\zeta^2\zeta' + a^{-2}\varepsilon\,\partial\zeta\partial\beta\partial^2\beta + \frac{1}{2}a^{-2}\varepsilon\,\partial^2\zeta(\partial\beta)^2 + 2f(\zeta)\frac{\delta L_2}{\delta\zeta}\Big)\,, \tag{283}$$

where we have defined

$$\frac{\delta L_2}{\delta\zeta} \equiv (a^3\varepsilon\,\zeta')' - a\varepsilon\partial^2\zeta\,, \tag{284}$$

$$f(\zeta) \equiv \frac{\eta}{4}\zeta^2 + \frac{\zeta\zeta'}{H} + \frac{-(\partial\zeta)^2 + \partial^{-2}(\partial_i\partial_j(\partial_i\zeta\partial_j\zeta))}{4a^2H^2} + \frac{\partial\zeta\partial\beta - \partial^{-2}(\partial_i\partial_j(\partial_i\zeta\partial_j\beta))}{2a^2H^2}\,. \tag{285}$$

Maldacena showed that the term proportional to $f(\zeta)$ in (283) can be removed by a field redefinition,

$$\zeta \to \tilde\zeta + f(\tilde\zeta)\,. \tag{286}$$

This field redefinition has the following effect on the correlation function:

$$\langle\zeta(\mathbf{x}_1)\zeta(\mathbf{x}_2)\zeta(\mathbf{x}_3)\rangle = \langle\tilde\zeta(\mathbf{x}_1)\tilde\zeta(\mathbf{x}_2)\tilde\zeta(\mathbf{x}_3)\rangle + \frac{\eta}{2}\Big(\langle\tilde\zeta(\mathbf{x}_1)\tilde\zeta(\mathbf{x}_2)\rangle\langle\tilde\zeta(\mathbf{x}_1)\tilde\zeta(\mathbf{x}_3)\rangle + \text{cyclic}\Big) + \cdots\,. \tag{287}$$

The term proportional to $f(\zeta)$ in (283) therefore leads to a contribution to $f_{\rm NL}$ of order $\eta \ll 1$. The effect of the remaining interactions is computed by expanding the *in-in* master formula (271) to first order in $H_{\rm int} = -L_3 + \mathcal{O}(\zeta^4)$,

$$\langle\zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle = -i\int_{-\infty}^{0}\mathrm{d}\tau\;\langle 0|\big[\hat\zeta_{\mathbf{k}_1}\hat\zeta_{\mathbf{k}_2}\hat\zeta_{\mathbf{k}_3}(0), \hat H_{\rm int}(\tau)\big]|0\rangle\,, \tag{288}$$

where we have switched to conformal time and taken the superhorizon limit $k\tau \to 0$.

Back-of-the-envelope estimate.—Before embarking on a lengthy calculation of the bispectrum, it is often advisable to perform an order-of-magnitude estimate of the expected size of the signal, i.e. to estimate (288) without explicitly performing the integral. For example, the first term in (283), can be written as

$$\int \mathrm{d}\tau\, H_{\rm int}(\tau) \subset -\int \mathrm{d}\tau\mathrm{d}^3x\,a^2\varepsilon^2\zeta\dot\zeta^2. \tag{289}$$

We only need to keep track of factors of H and ε. Any time- and momentum-dependence will work itself out and only contributes to the shape function. Using $a \propto H^{-1}$ and $\zeta \propto \dot\zeta \propto \mathcal{P}_\zeta^{1/2} \sim H/\sqrt{\varepsilon}$,

we estimate that the contribution from the three-point vertex is $\sim H\sqrt{\varepsilon}$. Combining this with estimates for the size of the three external legs, $\zeta^3 \sim H^3\varepsilon^{-3/2}$, we find

$$\langle\zeta^3\rangle = -i\int \mathrm{d}\tau\,\langle[\hat{\zeta}^3, \hat{H}_{\rm int}(\tau)]\rangle \propto \frac{H^4}{\varepsilon} \propto \mathcal{O}(\varepsilon)\,\mathcal{P}_\zeta^2 \sim f_{\rm NL}\,\mathcal{P}_\zeta^2. \tag{290}$$

Similar results are obtained for the other interactions in (283), $f_{\rm NL} \sim \mathcal{O}(\varepsilon)$. We also include the contribution from the field redefinition in (287), $f_{\rm NL} \sim \mathcal{O}(\eta)$. We conclude that the non-Gaussianity in slow-roll inflation is slow-roll suppressed,

$$f_{\rm NL} \sim \mathcal{O}(\varepsilon,\eta) \ll 1\,. \tag{291}$$

This small amount of non-Gaussianity will be unobservable in the CMB.

To get the full momentum-dependence of the bispectrum, we actually have to do some real work and compute the integral in (288) using the free-field mode functions for ζ. The final result for the dimensionless bispectrum is [40]

$$\begin{aligned}\frac{\mathcal{B}_\zeta(k_1,k_2,k_3)}{\mathcal{P}_\zeta^2} &= \frac{\varepsilon}{2}\left[-\left(\frac{k_1^2}{k_2k_3} + 2\,\text{perms.}\right) + \left(\frac{k_1}{k_2} + 5\,\text{perms.}\right) + \frac{8}{k_t}\left(\frac{k_1k_2}{k_3} + 2\,\text{perms.}\right)\right] \\ &\quad + \frac{\eta}{2}\left(\frac{k_1^2}{k_2k_3} + 2\,\text{perms.}\right),\end{aligned} \tag{292}$$

where $k_t \equiv k_1 + k_2 + k_3$. It is interesting to take the *squeezed limit*, $k_3 \ll k_1 \approx k_2$, of the result:

$$\lim_{k_3\to 0}\frac{\mathcal{B}_\zeta(k_1,k_2,k_3)}{\mathcal{P}_\zeta^2} = (2\varepsilon+\eta)\,\frac{k_1}{k_3}\,. \tag{293}$$

Note that the coefficient in the squeezed limit equals the deviation from scale invariance of the power spectrum, $1 - n_{\rm s} = 2\varepsilon + \eta$. In fact, this is a general result that applies to all models of single-field inflation, not necessarily just to slow-roll models. The *single-field consistency relation* states that [40, 94]

$$\boxed{\lim_{k_3\to 0} B_\zeta(k_1,k_2,k_3) = (1-n_{\rm s})\,P_\zeta(k_1)P_\zeta(k_3)} \tag{294}$$

i.e. the squeezed limit of the three-point function is suppressed by $(1 - n_{\rm s})$ and vanishes for perfectly scale-invariant perturbations.

Proof.—The squeezed triangle correlates one long-wavelength mode, $k_{\rm L} = k_3$, to two short-wavelength modes, $k_{\rm S} = k_1 \approx k_2$,

$$\langle\zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle \;\to\; \langle(\zeta_{\mathbf{k}_{\rm S}})^2\zeta_{\mathbf{k}_{\rm L}}\rangle\,. \tag{295}$$

Modes with longer wavelengths freeze earlier. The long mode $\zeta_{\mathbf{k}_{\rm L}}$ will therefore already be frozen and act as a classical background wave when the two short modes $\zeta_{\mathbf{k}_{\rm S}}$ exit the horizon. This provides an intuitive way to study the correlations between long and short modes.

Why should $(\zeta_{\mathbf{k}_{\rm S}})^2$ be correlated with $\zeta_{\mathbf{k}_{\rm L}}$? The theorem says that it isn't correlated if $\zeta_{\mathbf{k}}$ is precisely scale-invariant, but that the short scale power does get modified by the long-wavelength mode if $n_{\rm s} \neq 1$. Let's see why. We decompose the evaluation of (295) into two steps:

i) we calculate the power spectrum of short fluctuations $\langle\zeta_{\rm S}^2\rangle_{\zeta_{\rm L}}$ in the presence of a long mode $\zeta_{\rm L}$;

ii) we then calculate the correlation $\langle(\zeta_{\rm S})^2\zeta_{\rm L}\rangle$, i.e. average the short-scale power spectrum over realizations of the long modes.

The calculation of $\langle\zeta_{\rm S}^2\rangle_{\zeta_{\rm L}}$ is simplest in real space: When the background mode is homogeneous, $\zeta_{\rm L}(x)\equiv\bar\zeta_{\rm L}$, it can be reabsorbed simply by a rescaling of the spatial coordinates, $\tilde x^i = e^{\bar\zeta_{\rm L}}x^i$ (recall that $\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)e^{2\zeta(\mathbf{x},t)}\mathrm{d}\mathbf{x}^2$). After this rescaling, $\bar\zeta_{\rm L}$ no longer appears in the action, so that the two-point function in the new coordinates is the same as in the absence of $\bar\zeta_{\rm L}$. In other words, in the limit of constant $\bar\zeta_{\rm L}$, we have

$$\langle\zeta_{\rm S}(\mathbf{x}_1)\zeta_{\rm S}(\mathbf{x}_2)\rangle_{\bar\zeta_{\rm L}} = \langle\zeta_{\rm S}(\tilde{\mathbf{x}}_1)\zeta_{\rm S}(\tilde{\mathbf{x}}_2)\rangle\,. \tag{296}$$

When $\zeta_{\rm L}$ is slowly varying, we can evaluate it at the middle point $\mathbf{x}_+\equiv\frac12(\mathbf{x}_1+\mathbf{x}_2)$ to get

$$\tilde{\mathbf{x}}_- \simeq \mathbf{x}_- + \zeta_{\rm L}(\mathbf{x}_+)\cdot\mathbf{x}_- + \cdots\,, \tag{297}$$

where we defined $\mathbf{x}_-\equiv\mathbf{x}_1-\mathbf{x}_2$. The two-point function at linear order in $\zeta_{\rm L}$ therefore is

$$\langle\zeta_{\rm S}(\mathbf{x}_2)\zeta_{\rm S}(\mathbf{x}_3)\rangle_{\zeta_{\rm L}(x)} \simeq \xi_{\rm S}(|\mathbf{x}_-|) + \zeta_{\rm L}(\mathbf{x}_+)[\mathbf{x}_-\cdot\nabla\xi_{\rm S}(|\mathbf{x}_-|)] \tag{298}$$

where

$$\xi_{\rm S}(|\mathbf{x}_-|)\equiv\int\frac{\mathrm{d}^3k_{\rm S}}{(2\pi)^3}\,P_\zeta(k_{\rm S})\,e^{i\mathbf{k}_{\rm S}\cdot\mathbf{x}_-}\,. \tag{299}$$

The three-point function then is

$$\begin{aligned}\langle\zeta_{\rm S}(\mathbf{x}_1)\zeta_{\rm S}(\mathbf{x}_2)\zeta_{\rm L}(\mathbf{x}_3)\rangle &\simeq \langle\zeta_{\rm L}(\mathbf{x}_3)\zeta_{\rm L}(\mathbf{x}_+)\rangle[\mathbf{x}_-\cdot\nabla\xi_{\rm S}(|\mathbf{x}_-|)]\,,\\ &= \int\frac{\mathrm{d}^3k_{\rm L}}{(2\pi)^3}\int\frac{\mathrm{d}^3k_{\rm S}}{(2\pi)^3}\,e^{i\mathbf{k}_{\rm L}\cdot(\mathbf{x}_3-\mathbf{x}_+)}P_\zeta(k_{\rm L})P_\zeta(k_{\rm S})\left[\mathbf{k}_{\rm S}\cdot\frac{\partial}{\partial\mathbf{k}_{\rm S}}\right]e^{i\mathbf{k}_{\rm S}\cdot\mathbf{x}_-}\,.\end{aligned} \tag{300}$$

Integrating by parts, inserting $1=\int\mathrm{d}^3k_3\,\delta_D(\mathbf{k}_3+\mathbf{k}_{\rm L})$ and using

$$\frac{\partial}{\partial\mathbf{k}_{\rm S}}\cdot[\mathbf{k}_{\rm S}P_\zeta(k_{\rm S})] = P_\zeta)k_{\rm S})\,\frac{d\ln(k_{\rm S}^3P_\zeta(k_{\rm S}))}{d\ln k_{\rm S}}\,, \tag{301}$$

we get

$$\begin{aligned}\langle\zeta_{\rm S}(\mathbf{x}_1)\zeta_{\rm S}(\mathbf{x}_2)\zeta_{\rm L}(\mathbf{x}_3)\rangle = -\int\frac{\mathrm{d}^3k_3}{(2\pi)^3}\int\frac{\mathrm{d}^3k_{\rm L}}{(2\pi)^3}\int\frac{\mathrm{d}^3k_{\rm S}}{(2\pi)^3}\;e^{-i\mathbf{k}_1\cdot\mathbf{x}_1-i\mathbf{k}_{\rm L}\cdot\mathbf{x}_++i\mathbf{k}_{\rm S}\cdot\mathbf{x}_-}\\ \times\left[(2\pi)^3\delta_D(\mathbf{k}_1+\mathbf{k}_{\rm L})P_\zeta(k_1)P_\zeta(k_{\rm S})\,\frac{d\ln(k_{\rm S}^3P_\zeta(k_{\rm S}))}{d\ln k_{\rm S}}\right].\end{aligned} \tag{302}$$

Letting $\mathbf{k}_{\rm L}=\mathbf{k}_2+\mathbf{k}_3$ and $\mathbf{k}_{\rm S}=\frac12(\mathbf{k}_2-\mathbf{k}_3)$, we get $-i\mathbf{k}_{\rm L}\cdot\mathbf{x}_+ + i\mathbf{k}_{\rm S}\cdot\mathbf{x}_- = -i\mathbf{k}_2\cdot\mathbf{x}_2 - i\mathbf{k}_3\cdot\mathbf{x}_3$. Changing variables in the integration and Fourier transforming, we get

$$\begin{aligned}\lim_{k_3\to0}\langle\zeta_{\mathbf{k}_1}\zeta_{\mathbf{k}_2}\zeta_{\mathbf{k}_3}\rangle &= -(2\pi)^3\delta_D(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\,P_\zeta(k_1)P_\zeta(k_3)\,\frac{d\ln k_3^3P_\zeta(k_3)}{d\ln k_3}\,,\\ &= (2\pi)^3\delta_D(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\,(1-n_{\rm s})\,P_\zeta(k_1)P_\zeta(k_3)\,.\end{aligned} \tag{303}$$

This completes the proof.

The fact that the squeezed limit of the bispectrum for single-field inflation vanished implies that it can be used as a clean diagnostic for extra fields during inflation. In the next section, we will expand on this view of the squeezed limit as a particle detector.

6.6.4 Heavy Relics

As we have seen in the previous sections, computing the quantum correlations in single-field slow-roll inflation is a well-defined problem leading to clean predictions for the cosmological correlation functions. In this final section, we will look for deviations from those predictions associated with the existence of extra particles during inflation.

We will begin, in §6.6.4, with a review of the allowed spectrum of particles in de Sitter space, highlighting the qualitative differences to the corresponding results in flat space. In §6.6.4 and §6.6.4, we derive the basic observational imprints that these massive fields create when they are excited during inflation. For further details we refer the reader to the vast literature on the subject, e.g. [53, 54, 55, 56, 57, 38, 58, 59, 60, 97, 62, 63].

Massive Fields in de Sitter

Particles in Minkowski space are classified as unitary irreducible representations of the Poincaré group [64, 65]. The eigenvalues of the Casimir operators of the Poincaré group are related as follows to the mass m and the spin s of the particles:

$$\mathcal{C}_1 \equiv P_\mu P^\mu = m^2\,, \tag{304}$$
$$\mathcal{C}_2 \equiv W_\mu W^\mu = -s(s+1)m^2\,, \tag{305}$$

where P_μ is the four-momentum and W_μ is the Pauli-Lubanski pseudovector. We distinguish between massive and massless particles. Massive particles carry $2s+1$ degrees of freedom (transverse and longitudinal polarizations), while massless particles have only two transverse polarizations.

Similarly, particles in de Sitter space are classified as unitary irreducible representations of the de Sitter group $SO(1,4)$. The Casimir operators of the de Sitter group have the following eigenvalues [66, 67]:

$$\mathcal{C}_1 \equiv \frac{1}{2}M_{AB}M^{AB} = m^2 - 2(s-1)(s+1)H^2\,, \tag{306}$$
$$\mathcal{C}_2 \equiv W_A W^A = -s(s+1)\left[m^2 - \left(s^2+s-\frac{1}{2}\right)H^2\right], \tag{307}$$

where M_{AB} are the generators of $SO(1,4)$, with $A,B = \{0,\dots,4\}$. Notice that this time the Casimir (306) is not just proportional to the mass alone, but also has a term that depends on spin. For $s \geq 1$, the representations of de Sitter space fall into three distinct categories [68, 69, 70]:

principal series	complementary series	discrete series
$\frac{m^2}{H^2} \geq \left(s-\frac{1}{2}\right)^2$	$s(s-1) < \frac{m^2}{H^2} < \left(s-\frac{1}{2}\right)^2$	$\frac{m^2}{H^2} = s(s-1) - t(t+1)\,,$

for $t = 0, 1, 2, ..., s-1$, which is called the depth of the field. Masses that are not associated with one of the above categories correspond to non-unitary representations and are therefore not allowed in the spectrum (see Fig. 48). At the specific mass values corresponding to the discrete series, the system gains an additional gauge invariance and some of the lowest helicity modes become pure gauge modes; this phenomenon is called partial masslessness [71]. We see that unitarity demands the existence of a lower bound

$$m^2 \geq s(s-1)H^2\,, \tag{308}$$

on the masses of fields except those that belong to the discrete series. For $s = 2$, this is known as the Higuchi bound [72].

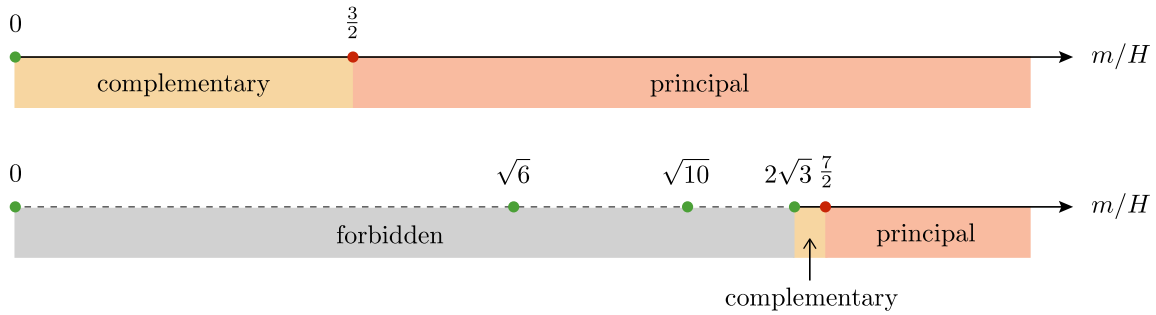


Figure 48: Spectrum of spin-0 (top) and spin-4 fields (bottom) in de Sitter space. The green points correspond to masses in the discrete series.

The Lagrangian for massive fields with arbitrary spin in flat space was constructed by Singh and Hagen in [73, 74], and generalized to (A)dS spaces in [75]. For massive fields with spin greater than 2, the action is rather complex and requires introducing auxiliary fields of lower spins. An alternative, which we will follow, is to use a group theoretical approach to find the equations of motion directly [66]. In four spacetime dimensions, a massive bosonic spin-s field is described by a totally symmetric rank-s tensor, $\sigma_{\mu_1\cdots\mu_s}$, subject to the constraints

$$\nabla^{\mu_1}\sigma_{\mu_1\cdots\mu_s} = 0\,, \quad \sigma^{\mu}{}_{\mu\mu_3\cdots\mu_s} = 0\,. \tag{309}$$

The conditions in (309) project out the components of the tensor which transform as fields with lower spins. The Casimir eigenvalue equation of the de Sitter group then gives the on-shell wave equation satisfied by these fields:

$$\big(\Box - m_s^2\big)\sigma_{\mu_1\cdots\mu_s} = 0\,, \tag{310}$$

where $\Box \equiv \nabla_\mu\nabla^\mu$ is the Laplace-Beltrami operator on dS_4 and $m_s^2 \equiv m^2 - (s^2-2s-2)H^2$. The shift in the mass arises from the mismatch between the Casimir and Laplace-Beltrami operators in de Sitter space and is necessary to describe the correct representations for massless fields. Explicit solutions to the equation of motion (310) can be found in Appendix A of [59]. The spatial components contain all the physical degrees of freedom, while the other components are constrained variables or pure gauge modes.[18] The late-time behaviour of the solution is

$$\lim_{\tau\to 0}\sigma_{i_1\cdots i_s}(\tau,\mathbf{x}) = \sum_{\pm}\tau^{\Delta_\pm - s}\sigma^{\pm}_{i_1\cdots i_s}(\mathbf{x})\,, \tag{311}$$

where $\Delta_\pm$ is the conformal dimension of the field

$$\Delta_\pm = \frac{3}{2} \pm i\mu\,, \quad \mu \equiv \sqrt{\frac{m^2}{H^2} - \left(s-\frac{1}{2}\right)^2}\,. \tag{312}$$

Note that, for $s=0$, the value $m=0$ corresponds to a conformally-coupled scalar field. In the limit $m \gg sH$, the parameter μ becomes the mass of the particle in Hubble units, $\mu \to m/H$. Particles belonging to the principal series correspond to $\mu \geq 0$, which covers the largest mass range. For real μ, the asymptotic scaling is given by a complex-conjugate pair, resulting in a wavefunction that oscillates logarithmically in conformal time. The complementary series has imaginary μ and corresponds to the interval $-i\mu = (0, 1/2)$. In that case, only the growing mode survives in the late-time limit.

[18] This can be seen by solving the equation of motion (310) explicitly, where the normalizations of non-physical components become singular [59].

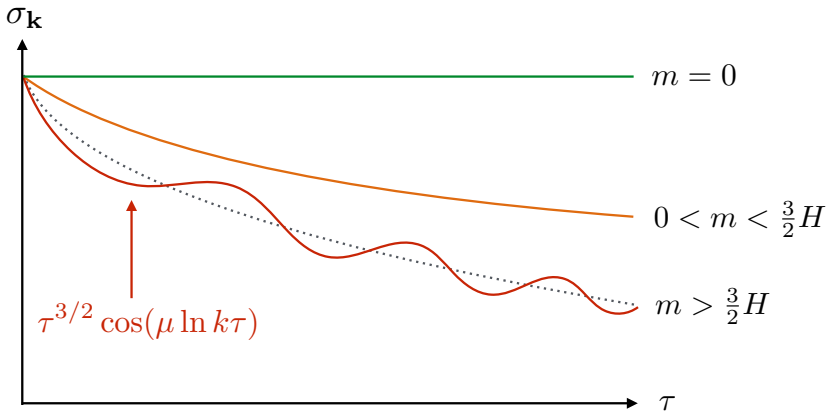
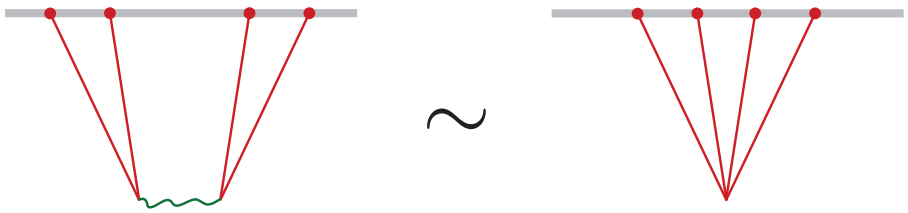


Figure 49: Superhorizon evolution of a massive scalar field. The decay of the field amplitude is determined by the mass of the field.

Local and Non-Local

In the remainder of this section, we will study the imprints of massive particles on inflationary correlations functions. Massive particles during inflation can have two types of effects:

- Particles with masses $M \gg H$ can be integrated out, leading to an EFT of inflaton interactions:



These *local* interactions create characteristic imprints in the non-Gaussianity of the inflaton fluctuations. In particular, the soft limits of the resulting correlation functions must satisfy the single-field consistency relation discussed in the previous section.

- Particles with masses $M \lesssim \text{few} \times H$ cannot be integrated out completely, but are produced nonperturbatively by the expanding spacetime. When these particles decay, they produce characteristic *non-local* signatures in cosmological correlators. The single-field consistency relation will be violated.

Local effects

To illustrate the local effects arising from massive particles, we consider the following two-field example:

$$\mathcal{L} = -\frac{1}{2}(\partial_\mu\phi)^2 - V(\phi) - \frac{1}{2}(\partial_\mu\sigma)^2 - \frac{1}{2}M^2\sigma^2 + \frac{\sigma(\partial_\mu\phi)^2}{\tilde{\Lambda}}\,, \tag{313}$$

where $M \gg H$. Integrating out the massive field σ gives

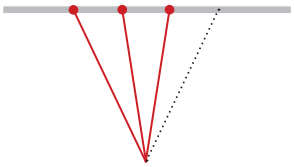
$$\begin{aligned}\mathcal{L}_{\text{eff}} &= -\frac{1}{2}(\partial_\mu\phi)^2 - V(\phi) + \frac{1}{8\tilde{\Lambda}^2}(\partial_\mu\phi)^2\,\frac{1}{\Box + M^2}\,(\partial_\mu\phi)^2 + \cdots \\ &= -\frac{1}{2}(\partial_\mu\phi)^2 - V(\phi) + \frac{1}{8\tilde{\Lambda}^2 M^2}\left((\partial_\mu\phi)^4 + (\partial_\mu\phi)^2\,\frac{\Box}{M^2}\,(\partial_\mu\phi)^2 + \cdots\right),\end{aligned} \tag{314}$$

where the second line is written as an expansion in $(H/M)^2$, with H being the Hubble scale during inflation. Truncating the expansion of EFT operators at the lowest order, we obtain the Lagrangian [80]

$$\mathcal{L}_{\rm eff} = \mathcal{L}_0(\phi) + \frac{(\partial_\mu \phi)^4}{8\Lambda^4}\,, \tag{315}$$

where $\mathcal{L}_0$ is the canonical slow-roll Lagrangian and $\Lambda^2 \equiv \tilde{\Lambda} M$. Following [80], we will compute the bispectrum induced by the interaction in (315).

This time it is useful to work in spatially flat gauge. At leading order in the slow-roll parameters, the action for the inflaton perturbation $\varphi \equiv \phi - \bar{\phi}(t)$ can then be calculated from (315) without including the contributions from $\sqrt{-g}R$. The intuitive reason for this simplification is that in the limit of a flat potential the perturbation φ does not induce a perturbation in the spacetime curvature. Evaluating one of the legs of the interaction $(\partial_\mu\phi)^4$ on the background solution, we get



$$\Leftrightarrow \qquad \mathcal{L}_{\rm eff,3} = -\frac{\phi'}{4\Lambda^4}\,\varphi'(\partial_\mu\varphi)^2\,. \tag{316}$$

Using this interaction, we can compute the bispectrum $\langle\varphi\varphi\varphi\rangle$. We will relate the result to $\langle\zeta\zeta\zeta\rangle$ at horizon crossing to avoid the error that would be induced by the superhorizon evolution of φ. To perform this matching, we, in principle, need to know the relation between φ and ζ up to second order. However, the quadratic term is slow-roll suppressed, so it suffices to use the linear relationship

$$\zeta = -\frac{H}{\phi'}\,\varphi\,. \tag{317}$$

Feeding $\mathcal{H}_{\rm int} = -\mathcal{L}_{\rm eff,3}$ into the *in-in* master formula (282), we find

$$\rangle' = \left(-\frac{H}{\bar{\phi}'}\right)^3 \langle\varphi_{\mathbf{k}_1}\varphi_{\mathbf{k}_2}\varphi_{\mathbf{k}_3}\rangle' \frac{\mathcal{B}_\zeta(k_1,k_2,k_3)}{\mathcal{P}_\zeta^2} \tag{318}$$

$$\begin{aligned} &= -\frac{4i}{k_1k_2k_3}\frac{(\phi')^2}{\Lambda^4}\int_{-\infty}^{0} \mathrm{d}\tau\left(-k_1^2k_2^2k_3^3\,\tau^2 - (\mathbf{k}_1\cdot\mathbf{k}_2)k_3^2(1-ik_1\tau)(1-ik_2\tau)\right)e^{ik_t\tau} \\ &\quad + \text{perms.} + c.c.\,, \qquad \begin{array}{cc} \uparrow & \uparrow \\ (\varphi')^3 & (\varphi')(\partial_i\varphi)^2 \end{array} \end{aligned} \tag{319}$$

where $k_t \equiv k_1 + k_2 + k_3$. Evaluating the integral, we get [80]

$$\frac{\mathcal{B}_\zeta(k_1,k_2,k_3)}{\mathcal{P}_\zeta^2} = \frac{8}{k_1k_2k_3}\frac{(\phi')^2}{\Lambda^4}\frac{1}{k_t^2}\left(\sum_i k_i^5 + \sum_{i\neq j}(2k_i^4k_j - 3k_i^3k_j^2) + \sum_{i\neq j\neq l}(k_i^3k_jk_l - 4k_i^2k_j^2k_l)\right). \tag{320}$$

The following features of this result are worth noting:

- The signal peaks in the equilateral configuration (cf. Fig. 45).
- The squeezed limit, $\lim_{k_{\rm L}\to 0}\langle\zeta_{\mathbf{k}_{\rm S}}\zeta_{\mathbf{k}_{\rm S}}\zeta_{\mathbf{k}_{\rm L}}\rangle$, is an analytic function of $k_{\rm L}/k_{\rm S}$. This is a consequence of the interaction being local.
- The amplitude is bounded,

$$f_{\rm NL} = \frac{35}{108}\frac{(\phi')^2}{\Lambda^4} < 1\,, \tag{321}$$

  where the final inequality follows from $\phi' < \Lambda^2$, as required by consistency of the truncation used in (315).

Nonlocal effects

For $M \lesssim$ few $\times H$, there are effects of massive particles that cannot be captured by local inflaton interactions, i.e. the field cannot be integrated out completely. The non-locality of the effective inflaton interactions will be reflected in a characteristic non-analyticity in its correlation functions.

As a concrete example, we consider a massive scalar field σ. Its two-point function in de Sitter space is

$$\langle \sigma_{\mathbf{k}}(\tau)\sigma_{\mathbf{k}'}(\tau')\rangle' = \frac{\pi}{4}H^2(\tau\tau')^{3/2}e^{-\pi\mu}H_{i\mu}(-k\tau)H^*_{i\mu}(-k\tau')\,, \tag{322}$$

where $H_{i\mu} \equiv H^{(1)}_{i\mu}$ is the Hankel function of the first kind and $\mu \equiv \sqrt{m^2/H^2 - 9/4}$. For now, let us focus on particles belonging to the principal series, so that μ is real. The local part of the two-point function has support only at coincident points in position space, while the non-local part describes correlations over long distances. In Fourier space, the local and non-local parts of the two-point function are analytic and non-analytic in the momentum k, respectively. In the late-time limit, we can split (322) into its local and non-local parts

$$\lim_{\tau,\tau'\to 0}\langle \sigma_{\mathbf{k}}(\eta)\sigma_{\mathbf{k}'}(\tau')\rangle'_{\text{local}} = \frac{H^2(\tau\tau')^{3/2}}{4\pi}\Gamma(-i\mu)\Gamma(i\mu)\left[e^{\pi\mu}\left(\frac{\tau}{\tau'}\right)^{i\mu} + e^{-\pi\mu}\left(\frac{\tau}{\tau'}\right)^{-i\mu}\right], \tag{323}$$

$$\lim_{\tau,\tau'\to 0}\langle \sigma_{\mathbf{k}}(\tau)\sigma_{\mathbf{k}'}(\tau')\rangle'_{\text{non-local}} = \frac{H^2(\tau\tau')^{3/2}}{4\pi}\left[\Gamma(-i\mu)^2\left(\frac{k^2\tau\tau'}{4}\right)^{i\mu} + \Gamma(i\mu)^2\left(\frac{k^2\tau\tau'}{4}\right)^{-i\mu}\right]. \tag{324}$$

Note that $\Gamma(\pm i\mu) \to e^{-\pi\mu/2}$ for large μ, resulting in an overall suppression of $e^{-\pi\mu}$ of the non-local contribution (324). In the next section, we will see how the non-analyticity of the non-local part of the two-point function of σ is reflected in the inflationary correlators.

Cosmological Collider Physics

The correlations in the massive field σ become observable when they get converted to inflaton fluctuations through interactions such as the mixing term in (313):

$$\mathcal{L} \supset \frac{\sigma(\partial_\mu\phi)^2}{\Lambda}\,. \tag{325}$$

Evaluating one of the legs of the inflaton field on the background solution, $\bar{\phi}$, leads to a bispectrum for the inflaton fluctuations φ (and hence of the curvature perturbations ζ); cf. Fig. 50. The derivation of the bispectrum of curvature perturbations can be found in [58, 59]. Here, we summarize the main features of the result:

- The non-local effect associated with the massive particle exchange shows up as a non-analyticity in the squeezed limit, $\lim_{k_{\rm L}\to 0}\langle\zeta_{\mathbf{k}_{\rm S}}\zeta_{-\mathbf{k}_{\rm S}}\zeta_{\mathbf{k}_{\rm L}}\rangle$. Between the horizon crossing times of the long mode $\zeta_{\rm L}$ and the short modes $\zeta_{\rm S}$ the amplitude of the massive field oscillates with a frequency set by the mass of the field. This leads to distinct oscillations in the bispectrum of curvature perturbations:

$$\lim_{k_{\rm L}\to 0}\langle\zeta_{\mathbf{k}_{\rm S}}\zeta_{-\mathbf{k}_{\rm S}}\zeta_{\mathbf{k}_{\rm L}}\rangle \propto \left(\frac{k_{\rm L}}{k_{\rm S}}\right)^{3/2}\cos\left[\frac{M}{H}\ln\left(\frac{k_{\rm L}}{k_{\rm S}}\right) + \delta\right], \tag{326}$$

  where δ is a computable phase [58]. We see that the mass M of the field is encoded in the frequency of the oscillations (see Fig. 51).

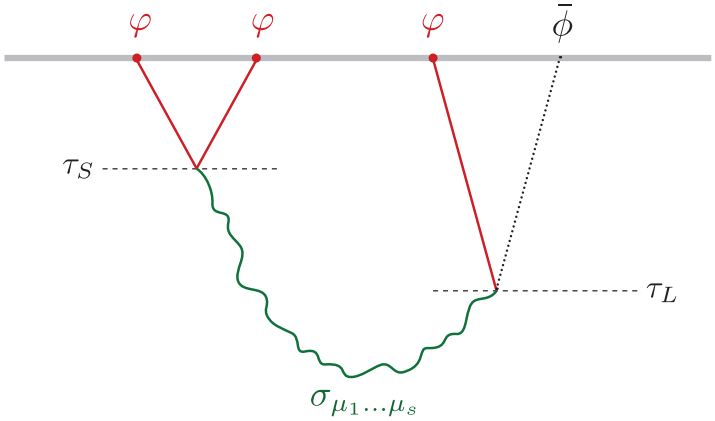


Figure 50: The spontaneous creation and decay of a massive particle σ produces a four-point function for the inflaton field ϕ. Evaluating one of the legs on the background solution, $\bar{\phi}$, leads to a three-point function for the inflaton fluctuations φ.

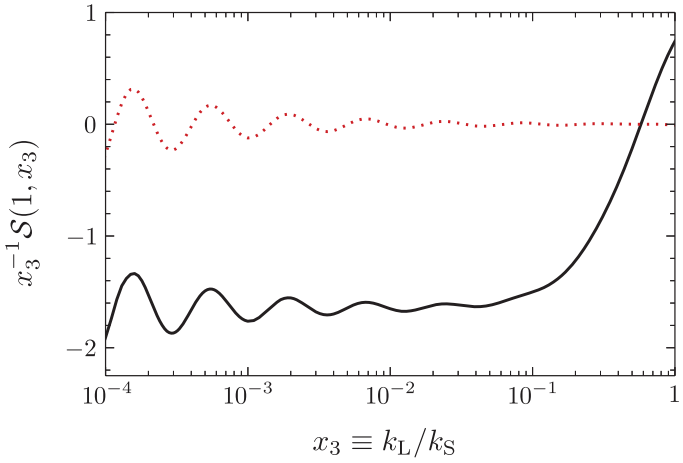


Figure 51: Shape function for the bispectrum arising from the exchange of a spin-2 particle with $\mu = 5$, evaluated in the isosceles-triangle configuration, $x_2 \equiv k_2/k_1 = 1$ (figure adapted from [59]). The dotted line is the non-local part of the signal.

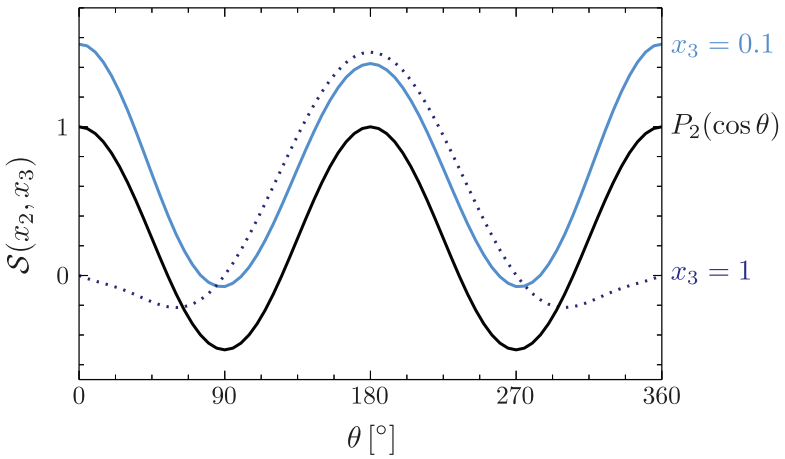


Figure 52: Shape function for the bispectrum arising from the exchange of a spin-2 particle with $\mu = 5$, as a function of the base angle $\theta = \cos^{-1}(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_3)$ for fixed ratios of $x_3 \equiv k_3/k_1$ (figure adapted from [59]).

- Particles with spin s lead to a unique angular dependence (see Fig. 52):

$$\lim_{k_{\rm L}\to 0} \langle \zeta_{\mathbf{k}_{\rm S}} \zeta_{-\mathbf{k}_{\rm S}} \zeta_{\mathbf{k}_{\rm L}} \rangle \;\propto\; P_s(\cos\theta)\,, \tag{327}$$

where $\cos\theta \equiv \mathbf{k}_{\rm S}\cdot\mathbf{k}_{\rm L}$ and P_s is a Legendre polynomial of degree s.

There are several effects that determine the amplitude of the signal:

- If we make the conservative assumption that the interaction in (325) is only of gravitational strength [58], then the scale determining the mixing between σ and ϕ is $\Lambda = M_{\rm pl}$. In that case, the amplitude of the bispectrum is at most of order the gravitational floor, $f_{\rm NL} \lesssim \mathcal{O}(\varepsilon)$. However, in principle, the mixing can be much larger. In particular, the mixing interaction in (325) remains perturbative as long as $\Lambda \geq (\phi')^{1/2}$, allowing for non-Gaussianity of order $f_{\rm NL} \lesssim \mathcal{O}(1)$ [53, 54, 77, 59]. Finally, in the EFT for the inflationary fluctuations [78] the cutoff Λ can even be below $(\phi')^{1/2}$ and the non-Gaussianity can be much larger, $f_{\rm NL} \lesssim \mathcal{O}(\mathcal{P}_\zeta^{-1/2})$.

- In addition, the spontaneous production of massive particles in the principal series is exponentially suppressed. This Boltzmann suppression is inherited by the amplitude of the bispectrum, $f_{\rm NL} \propto e^{-\pi M/H}$. The creation of particles in the complementary series, on the other hand, does not receive the same suppression and the signal can be large [53, 54]. Instead of oscillations, the squeezed limit in that case has a monotonic scaling $(k_{\rm L}/k_{\rm S})^{\Delta}$, where the scaling dimension Δ carries the information about the mass of the particle.

- For particles in the principal series, the squeezed limit is suppressed by a factor of $(k_{\rm L}/k_{\rm S})^{3/2}$. This is to be compared to the squeezed limit for equilateral non-Gaussianity which scales as $(k_{\rm L}/k_{\rm S})^2$. The signal from particles in the complementary series can be larger in the squeezed limit, $(k_{\rm L}/k_{\rm S})^{\Delta}$, where $0 < \Delta < 3/2$. Finally, partially massless particles do not decay on superhorizon scales and may therefore induce signals that are not suppressed in the squeezed limit [63].

Figure 53 is a schematic illustration of current and future constraints on (scale-invariant) primordial non-Gaussianities. We see that the perturbatively interesting regime spans about seven orders of magnitude in $f_{\rm NL}$. Of this regime, three orders of magnitude have been ruled out by current CMB observations, leaving a window of opportunity of about four orders of magnitude to be explored. Accessing these low levels of non-Gaussianity will be challenging. Even optimistic projections for future CMB observations won't reduce the constraints by more than an order of magnitude. Digging deeper will require new cosmological probes, such as observations of the large-scale structure of the universe [79] and the tomography of the 21cm transition of neutral hydrogen gas [80].

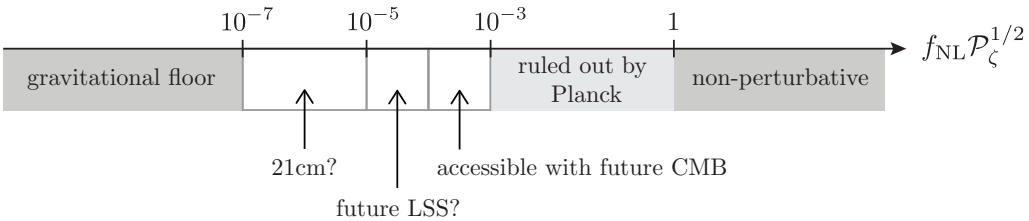


Figure 53: Schematic illustration of current and future constraints on (scale-invariant) primordial non-Gaussianity. The "gravitational floor" denotes the minimal level of non-Gaussianity created by purely gravitational interactions during inflation [40] (see §6.6.3).

7 TASI Lectures on Inflation Daniel Baumann

In a series of five lectures I review inflationary cosmology. I begin with a description of the initial conditions problems of the Friedmann-Robertson-Walker (FRW) cosmology and then

explain how inflation, an early period of accelerated expansion, solves these problems. Next, I describe how inflation transforms microscopic quantum fluctuations into macroscopic seeds for cosmological structure formation. I present in full detail the famous calculation for the primordial spectra of scalar and tensor fluctuations. I then define the inverse problem of extracting information on the inflationary era from observations of cosmic microwave background fluctuations. The current observational evidence for inflation and opportunities for future tests of inflation are discussed. Finally, I review the challenge of relating inflation to fundamental physics by giving an account of inflation in string theory.

7.1 Classical Dynamics of Inflation

The aim of this lecture is a first-principles introduction to the classical dynamics of inflationary cosmology. After a brief review of basic FRW cosmology we show that the conventional Big Bang theory leads to an initial conditions problem: the universe as we know it can only arise for very special and finely-tuned initial conditions. We then explain how inflation (an early period of accelerated expansion) solves this initial conditions problem and allows our universe to arise from generic initial conditions. We describe the necessary conditions for inflation and explain how inflation modifies the causal structure of spacetime to solve the Big Bang puzzles. Finally, we end this lecture with a discussion of the physical origin of the inflationary expansion.

Lecture 2: Quantum Fluctuations during Inflation

In this lecture we review the famous calculation of the primordial fluctuation spectra generated by quantum fluctuations during inflation. We present the calculation in full detail and try to avoid 'cheating' and approximations. After a brief review of fundamental aspects of cosmological perturbation theory, we first give a qualitative summary of the basic mechanism by which inflation converts microscopic quantum fluctuations into macroscopic seeds for cosmological structure formation. As a pedagogical introduction to quantum field theory in curved spacetime we then review the quantization of the simple harmonic oscillator. We emphasize that a unique vacuum state is chosen by demanding that the vacuum is the minimum energy state. We then proceed by giving the corresponding calculation for inflation. We calculate the power spectra of both scalar and tensor fluctuations.

Lecture 3: Contact with Observations

In this lecture we describe the inverse problem of extracting information on the inflationary perturbation spectra from observations of the cosmic microwave background and the large-scale structure. We define the precise relations between the gauge-invariant scalar and tensor power spectra computed in the previous lecture and the observed CMB anisotropies and galaxy power spectra. We give the transfer functions that relate the primordial fluctuations to the late-time observables. We then use these results to discuss the current observational evidence for inflation. Finally, we indicate opportunities for future tests of inflation.

Lecture 4: Primordial Non-Gaussianity

In this lecture we summarize key theoretical results in the study of primordial non-Gaussianity. Most results are stated without proof, but their significance for constraining the fundamental physical origin of inflation is explained. After introducing the bispectrum as a basic diagnostic of non-Gaussian statistics, we show that its momentum dependence is a powerful probe of the inflationary action. Large non-Gaussianity can only arise if inflaton interactions are significant during inflation. In single-field slow-roll inflation non-Gaussianity is therefore predicted to be unobservably small, while it can be significant in models with multiple fields, higher-derivative

interactions or non-standard initial states. Finally, we end the lecture with a discussion of the observational prospects for detecting or constraining primordial non-Gaussianity.

Lecture 5: Inflation in String Theory
We end this lecture series with a discussion of a slightly more advanced topic: inflation in string theory. We provide a pedagogical overview of the subject based on a recent review article with Liam McAllister. The central theme of the lecture is the sensitivity of inflation to Planck-scale physics, which we argue provides both the primary motivation and the central theoretical challenge for realizing inflation in string theory. We illustrate these issues through two case studies of inflationary scenarios in string theory: warped D-brane inflation and axion monodromy inflation. Finally, we indicate opportunities for future progress both theoretically and observationally.

7.2 Introduction

"I'm astounded by people who want to 'know' the Universe
when it's hard enough to find your way around Chinatown"
Woody Allen

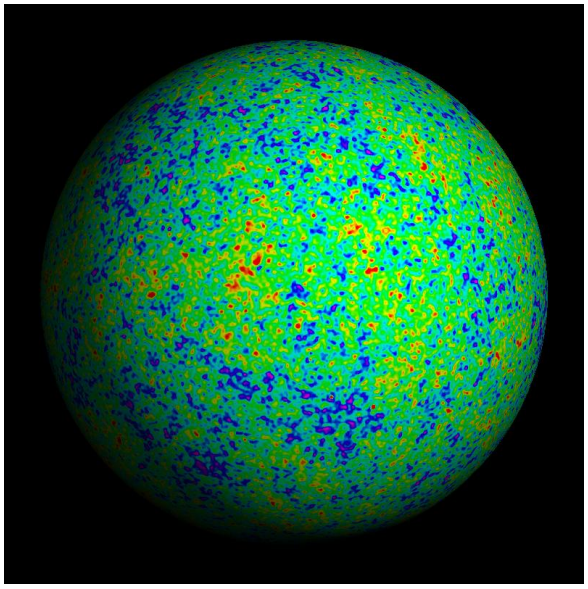


Figure 54: Fluctuations in the Cosmic Microwave Background (CMB). What produced them?

7.2.1 The Microscopic Origin of Structure

TASI 2009: *The Physics of the Large and the Small*

The fluctuations in the temperature of the cosmic microwave background (CMB) (see Fig. 54) tell an amazing story. Measured now almost routinely by experiments like the Wilkinson Microwave Anisotropy Probe (WMAP), the temperature variations of the microwave sky bear testimony of minute fluctuations in the density of the primordial universe. These fluctuations grew via gravitational instability into the large-scale structures (LSS) that we observe in the universe today. The success in relating observations of the thermal afterglow of the Big

Bang to the formation of structures billions of years later motivates us to ask an even bolder question: what is the fundamental microphysical origin of the CMB fluctuations? An answer to this question would provide us with nothing less than a fundamental understanding of the physical origin of all structure in the universe.

In these lectures, I will describe the currently leading working hypothesis that a period of cosmic *inflation* was integral part of this picture for the formation and evolution of structure. Inflation [119, 170, 15], a period of exponential expansion in the very early universe, is believed to have taken place some 10^{-34} seconds after the Big Bang singularity. Remarkably, inflation is thought to be responsible both for the large-scale homogeneity of the universe and for the small fluctuations that were the seeds for the formation of structures like our own galaxy.

The central focus of this lecture series will be to explain in full detail the physical mechanism by which inflation transformed microscopic quantum fluctuations into macroscopic fluctuations in the energy density of the universe. In this sense inflation provides the most dramatic example for the theme of *TASI 2009*: the connection between the 'physics of the large and the small'. We will calculate explicitly the statistical properties and the scale dependence of the spectrum of fluctuations produced by inflation. This result provides the input for all studies of cosmological structure formation and is one of the great triumphs of modern theoretical cosmology.

Structure and Evolution of the Universe

There is undeniable evidence for the expansion of the universe: the light from distant galaxies is systematically shifted towards the red end of the spectrum [4], the observed abundances of the light elements (H, He, and Li) matches the predictions of Big Bang Nucleosynthesis (BBN) [5], and the only convincing explanation for the CMB is a relic radiation from a hot early universe [6].

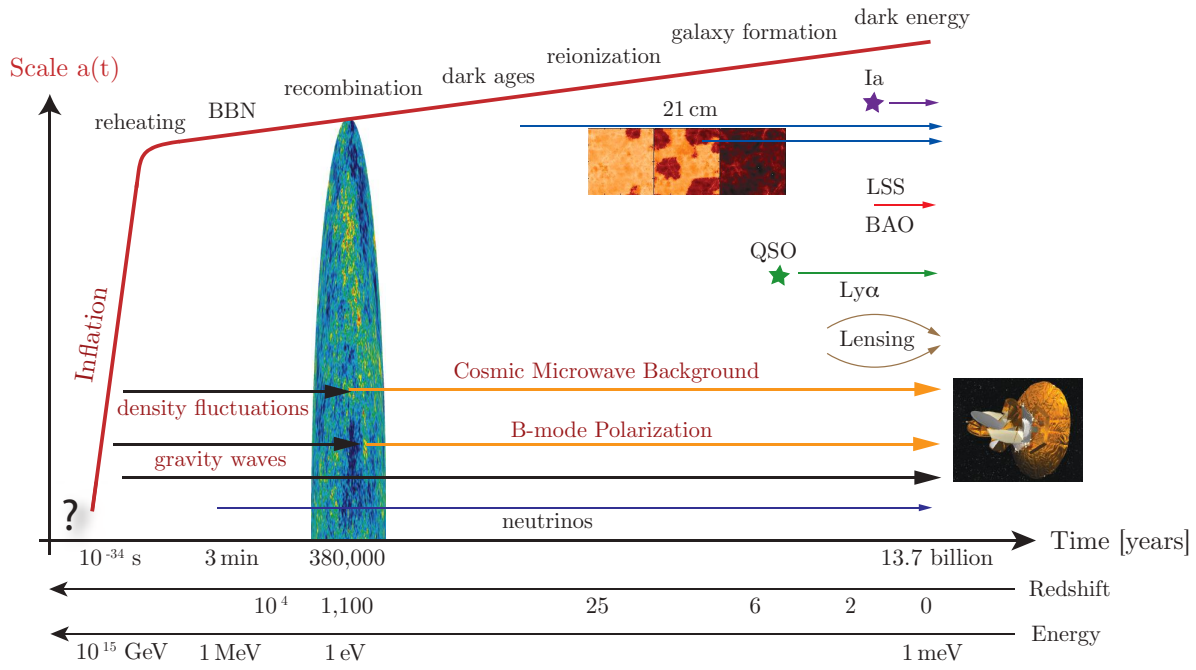


Figure 55: History of the universe. In this schematic we present key events in the history of the universe and their associated time and energy scales. We also illustrate several cosmological probes that provide us with information about the structure and evolution of the universe. *Acronyms*: BBN (Big Bang Nucleosynthesis), LSS (Large-Scale Structure), BAO (Baryon Acoustic Oscillations), QSO (Quasi-Stellar Objects = Quasars), Lyα (Lyman-alpha), CMB (Cosmic Microwave Background), Ia (Type Ia supernovae), 21cm (hydrogen 21cm-transition).

Two principles characterize thermodynamics and particle physics in an expanding universe: *i)* interactions between particles freeze out when the interaction rate drops below the expansion rate, and *ii)* broken symmetries in the laws of physics may be restored at high energies. Table 1 shows the thermal history of the universe and various phase transitions related to symmetry breaking events. In the following we will give a quick qualitative summary of these milestones in the evolution of our universe. We will emphasize which aspects of this cosmological story are based on established physics and which require more speculative ideas.

Table 1: Major Events in the History of the Universe.

	Time	Energy	
Planck Epoch?	$< 10^{-43}$ s	10^{18} GeV	
String Scale?	$\gtrsim 10^{-43}$ s	$\lesssim 10^{18}$ GeV	
Grand Unification?	$\sim 10^{-36}$ s	10^{15} GeV	
Inflation?	$\gtrsim 10^{-34}$ s	$\lesssim 10^{15}$ GeV	
SUSY Breaking?	$< 10^{-10}$ s	> 1 TeV	
Baryogenesis?	$< 10^{-10}$ s	> 1 TeV	
Electroweak Unification	10^{-10} s	1 TeV	
Quark-Hadron Transition	10^{-4} s	10^2 MeV	
Nucleon Freeze-Out	0.01 s	10 MeV	
Neutrino Decoupling	1 s	1 MeV	
BBN	3 min	0.1 MeV	
			Redshift
Matter-Radiation Equality	10^4 yrs	1 eV	10^4
Recombination	10^5 yrs	0.1 eV	1,100
Dark Ages	$10^5 - 10^8$ yrs		> 25
Reionization	10^8 yrs		$25 - 6$
Galaxy Formation	$\sim 6 \times 10^8$ yrs		~ 10
Dark Energy	$\sim 10^9$ yrs		~ 2
Solar System	8×10^9 yrs		0.5
Albert Einstein born	14×10^9 yrs	1 meV	0

From 10^{-10} seconds to today the history of the universe is based on well understood and experimentally tested laws of particle physics, nuclear and atomic physics and gravity. We are therefore justified to have some confidence about the events shaping the universe during that time.

Let us enter the universe at 100 GeV, the time of the electroweak phase transition (10^{-10} s). Above 100 GeV the electroweak symmetry is restored and the Z and $W^{\pm}$ bosons are massless. Interactions are strong enough to keep quarks and leptons in thermal equilibrium. Below 100 GeV the symmetry between the electromagnetic and the weak forces is broken, Z and $W^{\pm}$ bosons acquire mass and the cross-section of weak interactions decreases as the temperature of the universe drops. As a result, at 1 MeV, neutrinos decouple from the rest of the matter. Shortly after, at 1 second, the temperature drops below the electron rest mass and electrons and positrons annihilate efficiently. Only an initial matter-antimatter asymmetry of one part in a billion survives. The resulting photon-baryon fluid is in equilibrium. Around 0.1 MeV the strong interaction becomes important and protons and neutrons combine into the light elements (H, He, Li) during Big Bang nucleosynthesis (~ 200 s). The successful prediction of the H, He and Li abundances is one of the most striking consequences of the Big Bang theory.

The matter and radiation densities are equal around 1 eV (10^{11} s). Charged matter particles and photons are strongly coupled in the plasma and fluctuations in the density propagate as cosmic 'sound waves'. Around 0.1 eV (380,000 yrs) protons and electrons combine into neutral hydrogen atoms. Photons decouple and form the free-streaming cosmic microwave background. 13.7 billion years later these photons give us the earliest snapshot of the universe. Anisotropies in the CMB temperature provide evidence for fluctuations in the primordial matter density.

These small density perturbations, $\rho(\mathbf{x}, t) = \bar{\rho}(t)[1 + \delta(\mathbf{x}, t)]$, grow via gravitational instability to form the large-scale structures observed in the late universe. A competition between the background pressure and the universal attraction of gravity determines the details of the growth of structure. During radiation domination the growth is slow, $\delta \sim \ln a$ (where $a(t)$ is the scale factor describing the expansion of space). Clustering becomes more efficient after matter dominates the background density (and the pressure drops to zero), $\delta \sim a$. Small scales become non-linear first, $\delta \gtrsim 1$, and form gravitationally bound objects that decouple from the overall expansion. This leads to a picture of hierarchical structure formation with small-scale structures (like stars and galaxies) forming first and then merging into larger structures (clusters and superclusters of galaxies). Around redshift $z \sim 25$ ($1 + z = a^{-1}$), high energy photons from the first stars begin to ionize the hydrogen in the inter-galactic medium. This process of 'reionization' is completed at $z \approx 6$. Meanwhile, the most massive stars run out of nuclear fuel and explode as 'supernovae'. In these explosions the heavy elements (C, O, ...) necessary for the formation of life are created, leading to the slogan "we are all stardust". At $z \approx 1$, a negative pressure 'dark energy' comes to dominate the universe. The background spacetime is accelerating and the growth of structure ceases, $\delta \sim$ const.

The First 10^{-10} Seconds

The history of the universe from 10^{-10} seconds (1 TeV) to today is based on observational facts and tested physical theories like the Standard Model of particle physics, general relativity and fluid dynamics, *e.g.* the fundamental laws of high energy physics are well-established up to the energies reached by current particle accelerators ($\sim$ 1 TeV). Before 10^{-10} seconds, the energy of the universe exceeds 1 TeV and we lose the comfort of direct experimental guidance. The physics of that era is therefore as speculative as it is fascinating.

To explain the fluctuations seen in the CMB temperature requires an input of primordial seed fluctuations. In these lectures we will explain the conjecture that these primordial fluctuations were generated in the very early universe ($\sim 10^{-34}$ seconds) during a period of inflation. We will explain how microscopic quantum fluctuations in the energy density get stretched by the inflationary expansion to macroscopic scales, larger than the physical horizon at that time. After a perturbation exits the horizon no causal physics can affect it and it remains frozen with constant amplitude until it re-enters the horizon at a later time during the conventional (non-accelerating) Big Bang expansion. The fluctuations associated with cosmological structures re-enter the horizon when the universe is about 100,000 years olds, a short time before the decoupling of the CMB photons. Inside the horizon causal physics can affect the perturbation amplitudes and in fact leads to the acoustic peak structure of the CMB and the collapse of high-density fluctuations into galaxies and clusters of galaxies. Since we understand (and can calculate) the evolution of perturbations after they re-enter the horizon we can use the late time observations of the CMB and the LSS to infer the primordial input spectrum. Assuming this spectrum was produced by inflation, this gives us an observational probe of the physical conditions when the universe was 10^{-34} seconds old. This fascinating opportunity to use cosmology to probe physics at the highest energies will be the subject of these lectures.

7.2.2 Outline of the Lectures

In Lecture 1 we introduce the *classical* background dynamics of inflation. We explain how inflation solves the horizon and flatness problems. We discuss the slow-roll conditions and reheating and speculate on the physical origin of the inflationary expansion. In Lecture 2 we describe how *quantum* fluctuations during inflation become the seeds for the formation of large-scale structures. We present in full detail the derivation of the inflationary power spectra of scalar and tensor perturbations, $\mathcal{R}$ and h_{ij}. In Lecture 3 we relate the results of Lecture 2 to observations of the cosmic microwave background and the distribution of galaxies, *i.e.* we explain how to measure $P_{\mathcal{R}}(k)$ and $P_h(k)$ in the sky! We describe current observational constraints and emphasize future tests of inflation. In Lecture 4 we present key results in the study of non-Gaussianity of the primordial fluctuations. We explain how non-Gaussian correlations can provide important information on the inflationary action. We reserve Lecture 5 for the study of an advanced topic that is at the frontier of current research: inflation in string theory. We describe the main challenges of the subject and summarize recent advances.

To make each lecture self-contained, the necessary background material is presented in a short *review* section preceeding the core of each lecture. Every lecture ends with a *summary* of the most important results. An important part of every lecture are problems and exercises that appear throughout the text and (for longer problems) as a separate *problem set* appended to the end of the lecture. The exercises were carefully chosen to complement the material of the lecture or to fill in certain details of the computations.

A number of appendices collect standard results from cosmological perturbation theory and details of the inflationary perturbation calculation. It is hoped that the appendices provide a useful reference for the reader.

Notation

We have tried hard to keep the notation of these lectures coherent and consistent:
Throughout we will use the God-given natural units

$$c = \hbar \equiv 1\,.$$

We use the reduced Planck mass

$$M_{\rm pl} = (8\pi G)^{-1/2}\,,$$

and often set it equal to one. Our metric signature is $(-+++)$. Greek indices will take the values $\mu, \nu = 0, 1, 2, 3$ and latin indices stand for $i, j = 1, 2, 3$. Our Fourier convention is

$$\mathcal{R}_{\mathbf{k}} = \int \mathrm{d}^3\mathbf{x}\, \mathcal{R}(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}\,,$$

so that the power spectrum is

$$\langle \mathcal{R}_{\mathbf{k}} \mathcal{R}_{\mathbf{k}'} \rangle = (2\pi)^3 \delta(\mathbf{k}+\mathbf{k}') P_{\mathcal{R}}(k)\,, \qquad \Delta^2_{\mathcal{R}}(k) \equiv \frac{k^3}{2\pi^2} P_{\mathcal{R}}(k)\,.$$

For conformal time we use the letter τ (and caution the reader not confuse it with the astrophysical parameter for optical depth). We reserve the letter η for the second slow-roll parameter. Derivatives with respect to physical time are denoted by overdots, while derivatives with respect to conformal time are indicated by primes. Partial derivatives are denoted by commas, covariant derivatives by semi-colons.

Acknowledgements

I am most grateful to Scott Dodelson and Csaba Csaki for the invitation to give these lectures at the Theoretical Advanced Study Institute in Elementary Particle Physics (*TASI*).

In past few years I have learned many things from my teachers Liam McAllister, Paul Steinhardt, and Matias Zaldarriaga and my collaborators Igor Klebanov, Anatoly Dymarsky, Shamit Kachru, Hiranya Peiris, Alberto Nicolis and Asantha Cooray. Thanks to all of them for generously sharing their insights with me. The input from the members of the *CMBPol* Inflation Working Group [7] was very much appreciated. Their expert opinions on many of the topics described in these lectures were most valuable to me. I learned many things in discussions with Richard, Easther, Eva Silverstein, Eiichiro Komatsu, Mark Jackson, Licia Verde, David Wands, Paolo Creminelli, Leonardo Senatore, Andrei Linde, and Sarah Shandera.

The Aspen Center for Physics, Trident Cafe, Boston and Dado Tea, Cambridge are acknowledged for their hospitality while these lecture notes were written.

Finally, I wish to thank the students at *TASI 2009* for challenging me with their questions and for comments on a draft of these notes.

7.3 Classical Dynamics of Inflation by Baumann 1th lecture

Abstract

The aim of this lecture is a first-principles introduction to the classical dynamics of inflationary cosmology. After a brief review of basic FRW cosmology we show that the conventional Big Bang theory leads to an initial conditions problem: the universe as we know it can only arise for very special and finely-tuned initial conditions. We then explain how inflation (an early period of accelerated expansion) solves this initial conditions problem and allows our universe to arise from generic initial conditions. We describe the necessary conditions for inflation and explain how inflation modifies the causal structure of spacetime to solve the Big Bang puzzles. Finally, we end this lecture with a discussion of the physical origin of the inflationary expansion.

7.3.1 *Review*: The Homogeneous Universe

To set the stage, we review basic aspects of the homogeneous universe. Since this material was covered in Prof. Turner's lectures at *TASI 2009* and is part of any textbook treatment of cosmology (*e.g.* [8, 9, 10]), we will be brief and recall many of the concepts via exercises for the reader. We will naturally focus on the elements most relevant for the study of inflation.

FRW Spacetime

Cosmology describes the structure and evolution of the universe on the largest scales. Assuming *homogeneity* and *isotropy* [19] on large scales one is lead to the Friedmann-Robertson-Walker (FRW) metric for the spacetime of the universe (see **Problem 1**):

$$\boxed{\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2)\right)}. \tag{328}$$

[19] A *homogeneous* space is one which is translation invariant, or the same at every point. An *isotropic* space is one which is rotationally invariant, or the same in every direction. A space which is everywhere isotropic is necessarily homogeneous, but the converse is not true; *e.g.* a space with a uniform electric field is translationally invariant but not rotationally invariant.

Here, the scale factor $a(t)$ characterizes the relative size of spacelike hypersurfaces Σ at different times. The curvature parameter k is $+1$ for positively curved Σ, 0 for flat Σ, and -1 for negatively curved Σ. Eqn. (328) uses comoving coordinates – the universe expands as $a(t)$ increases, but galaxies/observers keep fixed coordinates r, θ, ϕ as long as there aren't any forces acting on them, *i.e.* in the absence of peculiar motion. The corresponding physical distance is obtained by multiplying with the scale factor, $R = a(t)r$, and is time-dependent even for objects with vanishing peculiar velocities. By a coordinate transformation the metric (328) may be written as

$$\boxed{\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\left(\mathrm{d}\chi^2 + \Phi_k(\chi^2)(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2)\right)}, \tag{329}$$

where

$$r^2 = \Phi_k(\chi^2) \equiv \begin{cases} \sinh^2\chi & k=-1 \\ \chi^2 & k=0 \\ \sin^2\chi & k=+1 \end{cases}. \tag{330}$$

For the FRW ansatz the evolution of the homogeneous universe boils down to the single function $a(t)$. Its form is dictated by the matter content of the universe via the Einstein field equations (see §7.3.1). An important quantity characterizing the FRW spacetime is the expansion rate

$$\boxed{H \equiv \frac{\dot{a}}{a}}. \tag{331}$$

The *Hubble parameter* H has unit of inverse time and is positive for an expanding universe (and negative for a collapsing universe). It sets the fundamental scale of the FRW spacetime, *i.e.* the characteristic time-scale of the homogeneous universe is the Hubble time, $t \sim H^{-1}$, and the characteristic length-scale is the Hubble length, $d \sim H^{-1}$ (in units where $c = 1$). The Hubble scale sets the scale for the age of the universe, while the Hubble length sets the size of the observable universe.

Kinematics: Conformal Time and Horizons

Having defined the metric for the average spacetime of the universe we can now study kinematical properties of the propagation of light and matter particles.

Conformal Time and Null Geodesics

The causal structure of the universe is determined by the propagation of light in the FRW spacetime (328). Massless photons follows null geodesics, $\mathrm{d}s^2 = 0$. These photon trajectories are studied most easily if we define conformal time[20]

$$\boxed{\tau = \int \frac{\mathrm{d}t}{a(t)}}, \tag{332}$$

for which the FRW metric becomes

$$\mathrm{d}s^2 = a(\tau)^2\left[-\mathrm{d}\tau^2 + \left(\mathrm{d}\chi^2 + \Phi_k(\chi^2)(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2)\right)\right]. \tag{333}$$

In an isotropic universe we may consider radial propagation of light as determined by the two-dimensional line element

$$\mathrm{d}s^2 = a(\tau)^2\left[-\mathrm{d}\tau^2 + \mathrm{d}\chi^2\right]. \tag{334}$$

[20] Conformal time may be interpreted as a "clock" which slows down with the expansion of the universe.

The metric has factorized into a static Minkowski metric multiplied by a time-dependent conformal factor $a(\tau)$. Expressed in conformal time the radial null geodesics of light in the FRW spacetime therefore satisfy

$$\chi(\tau) = \pm\tau + \text{const.}\,, \tag{335}$$

i.e. they correspond to straight lines at angles $\pm 45°$ in the τ–χ plane (see Fig. 56). If instead we had used physical time t to study light propagation, then the light cones for curved spacetimes would be curved.

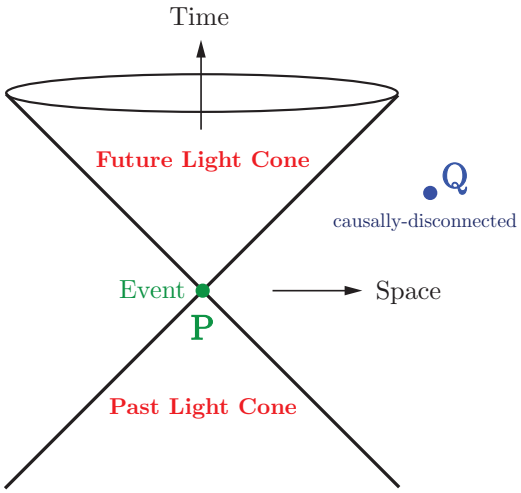


Figure 56: Light cones and causality. Photons travel along world lines of zero proper time, $\mathrm{d}s^2 = 0$, called *null geodesics*. Massive particles travel along world lines with real proper time, $\mathrm{d}s^2 > 0$, called *timelike geodescis*. Causally disconnected regions of spacetime are separated by *spacelike* intervals, $\mathrm{d}s^2 < 0$. The set of all null geodesics passing through a given point (or *event*) in spacetime is called the *light cone*. The interior of the light cone, consisting of all null and timelike geodesics, defined the region of spacetime causally related to that event.

Particle Horizon

The maximum comoving distance light can propagate between an initial time t_i and some later time t is

$$\chi_p(\tau) = \tau - \tau_\mathrm{i} = \int_{t_\mathrm{i}}^{t} \frac{\mathrm{d}t}{a(t)}\,. \tag{336}$$

This is called the (comoving) particle horizon. The initial time t_i is often taken to be the 'origin of the universe', $t_\mathrm{i} \equiv 0$, defined by the initial singularity, $a(t_\mathrm{i} \equiv 0) \equiv 0$.[21] The physical size of the particle horizon is

$$d_p(t) = a(t)\chi_p\,. \tag{337}$$

The particle horizon is of crucial importance to understanding the causal structure of the universe and it will be fundamental to our discussion of inflation. As we will see, the conventional Big Bang model 'begins' at a finite time in the past and at any time in the past the particle horizon was finite, limiting the distance over which spacetime region could have been in causal contact. This feature is at the heart of the 'Big Bang puzzles'.

[21] Whether $t_\mathrm{i} = 0$ also corresponds to $\tau_\mathrm{i} = 0$ depends on the evolution of the scale factor $a(t)$; *e.g.* for inflation $t_\mathrm{i} = 0$ will *not* be $\tau_\mathrm{i} = 0$.

Event Horizon

An event horizon defines the set of points from which signals sent at a given moment of time τ will never be received by an observer in the future. In comoving coordinates these points satisfy

$$\chi \; > \; \chi_e = \int_{\tau}^{\tau_{\rm max}} {\rm d}\tau = \tau_{\rm max} - \tau \,, \tag{338}$$

where $\tau_{\rm max}$ denotes the 'final moment of time' (this might be infinite or finite). The physical size of the event horizon is

$$d_e(t) = a(t)\chi_e \,. \tag{339}$$

Angular Diameter and Luminosity Distances

Conformal time (or comoving distance) relates in a simple way to angular diameter and luminosity distances which are important for the discussion of CMB anisotropies and supernova distances, respectively. These details won't concern us here, but may be found in the standard books [8, 9, 10].

Dynamics: Einstein Equations

The dynamics of the universe as characterized by the evolution of the scale factor of the FRW spacetime $a(t)$ is determined by the Einstein Equations

$$\boxed{G_{\mu\nu} = 8\pi G\, T_{\mu\nu}} \,. \tag{340}$$

We will often work in units where $8\pi G \equiv 1$.

Einstein Gravity

For convenience we here recall the definition of the Einstein tensor

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \,, \tag{341}$$

in terms of the Ricci tensor $R_{\mu\nu}$ and the Ricci scalar R,

$$R_{\mu\nu} = \Gamma^{\alpha}_{\ \mu\nu,\alpha} - \Gamma^{\alpha}_{\ \mu\alpha,\nu} + \Gamma^{\alpha}_{\ \beta\alpha}\Gamma^{\beta}_{\ \mu\nu} - \Gamma^{\alpha}_{\ \beta\nu}\Gamma^{\beta}_{\ \mu\alpha} \,, \qquad R \equiv g^{\mu\nu} R_{\mu\nu} \,, \tag{342}$$

where

$$\Gamma^{\mu}_{\ \alpha\beta} \equiv \frac{g^{\mu\nu}}{2} \left[g_{\alpha\nu,\beta} + g_{\beta\nu,\alpha} - g_{\alpha\beta,\nu} \right] \,. \tag{343}$$

Commas denote partial derivatives, *e.g.* $(\dots)_{,\mu} = \frac{\partial(\dots)}{\partial x^{\mu}}$. We will continue to follow this notation in the rest of these lectures.

Energy, Momentum and Pressure

To define the energy-momentum tensor of the universe, $T_{\mu\nu}$, we introduce a set of observers whose worldlines are tangent to the timelike velocity 4-vector

$$u^{\mu} \equiv \frac{dx^{\mu}}{d\tau} \,, \tag{344}$$

where τ is the proper time of the observers, so that $g_{\mu\nu}u^{\mu}u^{\nu} = -1$. We define the tensor $\gamma_{\mu\nu} \equiv g_{\mu\nu} + u_{\mu}u_{\nu}$ as the metric of the 3-dimensional spatial sections orthogonal to u_{μ}. We

use $\gamma_{\mu\nu}$ to project quantities orthogonal to the 4-velocity into the observers' instantaneous rest space. The energy-momentum tensor of a general (imperfect) fluid can then be written as

$$T_{\mu\nu} = \rho u_\mu u_\nu + p\gamma_{\mu\nu} + 2q_{(\mu}u_{\nu)} + \Sigma_{\mu\nu}\,, \tag{345}$$

where $\rho = T_{\mu\nu}u^\mu u^\nu$ is the matter energy density, $p = \frac{1}{3}T_{\mu\nu}\gamma^{\mu\nu}$ is the isotropic pressure, $q_\mu = -\gamma_\mu{}^\alpha T_{\alpha\beta}u^\beta$ is the energy-flux vector, and $\Sigma_{\mu\nu} = \gamma_{\langle\mu}{}^\alpha\gamma_{\nu\rangle}{}^\beta T_{\alpha\beta}$ is the symmetric and trace-free anisotropic stress tensor.[22] For a perfect fluid there exists a unique 4-velocity so that $q_\mu = \Sigma_{\mu\nu} = 0$, *i.e.* for the case of a perfect fluid the stress-energy tensor is

$$T^\mu{}_\nu = g^{\mu\alpha}T_{\alpha\nu} = (\rho + p)\,u^\mu u_\nu - p\,\delta^\mu_\nu\,, \tag{346}$$

where ρ and p are the proper energy density and pressure in the fluid rest frame and u^μ is the 4-velocity of the fluid. In a frame that is comoving with the fluid we may choose $u^\mu = (1,0,0,0)$, *i.e.*

$$T^\mu{}_\nu = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & -p & 0 & 0 \\ 0 & 0 & -p & 0 \\ 0 & 0 & 0 & -p \end{pmatrix}. \tag{347}$$

The Einstein Equations then take the form of two coupled, non-linear ordinary differential equations, also called the the *Friedmann Equations* (see **Problem 3**)

$$\boxed{H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{1}{3}\rho - \frac{k}{a^2}}\,, \tag{348}$$

and

$$\boxed{\dot{H} + H^2 = \frac{\ddot{a}}{a} = -\frac{1}{6}(\rho + 3p)}\,, \tag{349}$$

where overdots denote derivatives with respect to physical time t. Notice, that in an expanding universe (*i.e.* $\dot{a} > 0$) filled with ordinary matter (*i.e.* matter satisfying the strong energy condition: $\rho + 3p \geq 0$) Eqn. (349) implies $\ddot{a} < 0$. This indicates the existence of a singularity in the finite past: $a(t \equiv 0) = 0$. Of course, this conclusion relies on the assumption that General Relativity and the Friedmann Equations are applicable up to arbitrary high energies and that no exotic forms of matter become relevant at high energies. More likely the singularity signals the breakdown of General Relativity.

Eqns. (348) and (349) may be combined into the continuity equation

$$\boxed{\frac{d\rho}{dt} + 3H(\rho + p) = 0}\,. \tag{350}$$

This may also be written as

$$\frac{d\ln\rho}{d\ln a} = -3(1 + w)\,, \tag{351}$$

if we define the equation of state parameter

$$w \equiv \frac{p}{\rho}\,. \tag{352}$$

Eqn. (351) may be integrated to give

$$\rho \propto a^{-3(1+w)}\,. \tag{353}$$

[22] Here we use the notation $t_{\langle\mu\nu\rangle} = \gamma_{(\mu}{}^\alpha\gamma_{\nu)}{}^\beta t_{\alpha\beta} - \frac{1}{3}\gamma^{\alpha\beta}t_{\alpha\beta}\gamma_{\mu\nu}$ and $t_{(\mu\nu)} = \frac{1}{2}(t_{\mu\nu} + t_{\nu\mu})$.

Together with the Friedmann Equation (348) this leads to the time evolution of the scale factor

$$a(t) \propto \begin{cases} t^{2/3(1+w)} & w \neq -1\,, \\ e^{Ht} & w = -1\,, \end{cases} \tag{354}$$

i.e. $a(t) \propto t^{2/3}$, $a(t) \propto t^{1/2}$ and $a(t) \propto \exp(Ht)$, for the scale factor of a flat ($k = 0$) universe dominated by non-relativistic matter ($w = 0$), radiation or relativistic matter ($w = \frac{1}{3}$) and a cosmological constant ($w = -1$), respectively.

Table 2: FRW solutions for a flat universe dominated by radiation, matter or a cosmological constant.

	w	$\rho(a)$	$a(t)$	$a(\tau)$	$\tau_{\rm i}$
MD	0	a^{-3}	$t^{2/3}$	τ^2	0
RD	$\frac{1}{3}$	a^{-4}	$t^{1/2}$	τ	0
Λ	-1	a^0	e^{Ht}	$-\tau^{-1}$	$-\infty$

If more than one matter species (baryons, photons, neutrinos, dark matter, dark energy, etc.) contributes significantly to the energy density and the pressure, ρ and p refer to the sum of all components

$$\rho \equiv \sum_i \rho_i\,, \qquad p \equiv \sum_i p_i\,. \tag{355}$$

For each species 'i' we define the *present* ratio of the energy density relative to the *critical energy density* $\rho_{\rm crit} \equiv 3H_0^2$

$$\Omega_i \equiv \frac{\rho_0^i}{\rho_{\rm crit}}\,, \tag{356}$$

and the corresponding equations of state

$$w_i \equiv \frac{p_i}{\rho_i}\,. \tag{357}$$

Here and in the following the subscript '0' denotes evaluation of a quantity at the present time t_0. We normalize the scale factor such that $a_0 = a(t_0) \equiv 1$. This allows us to write the Friedmann Equation (348) as

$$\left(\frac{H}{H_0}\right)^2 = \sum_i \Omega_i a^{-3(1+w_i)} + \Omega_k a^{-2}, \tag{358}$$

with $\Omega_k \equiv -k/a_0^2H_0^2$ parameterizing curvature. Evaluating Eqn. (358) today implies the consistency relation

$$\sum_i \Omega_i + \Omega_k = 1\,. \tag{359}$$

The second Friedmann Equation (349) evaluated at $t = t_0$ becomes

$$\frac{1}{a_0 H_0^2}\frac{d^2 a_0}{dt^2} = -\frac{1}{2}\sum_i \Omega_i(1 + 3w_i). \tag{360}$$

This defines the condition for accelerated expansion today.

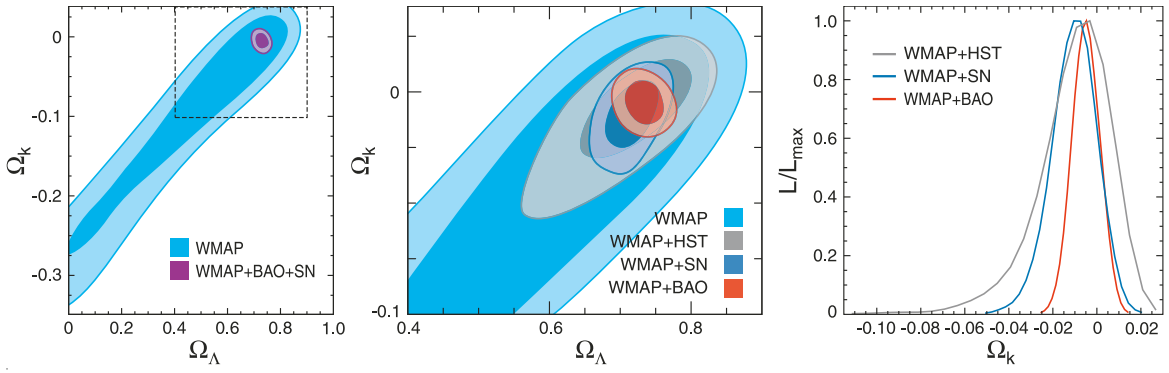


Figure 57: A combination CMB and LSS observations indicate that the spatial geometry of the universe is flat [11]. Note that the evidence for flatness cannot be obtained from CMB observations alone.

The Concordance Model

Observations of the cosmic microwave background and the large-scale structure find that the universe is flat (see Fig. 57)

$$\Omega_k \sim 0\,, \tag{361}$$

and composed of 4% atoms (or baryons, 'b'), 23% (cold) dark matter ('dm') and 73% dark energy (Λ) (see Fig. 58):

$$\Omega_b = 0.04\,, \quad \Omega_{dm} = 0.23\,, \quad \Omega_\Lambda = 0.72\,, \tag{362}$$

with $w_\Lambda \approx -1$ (see Fig. 59).

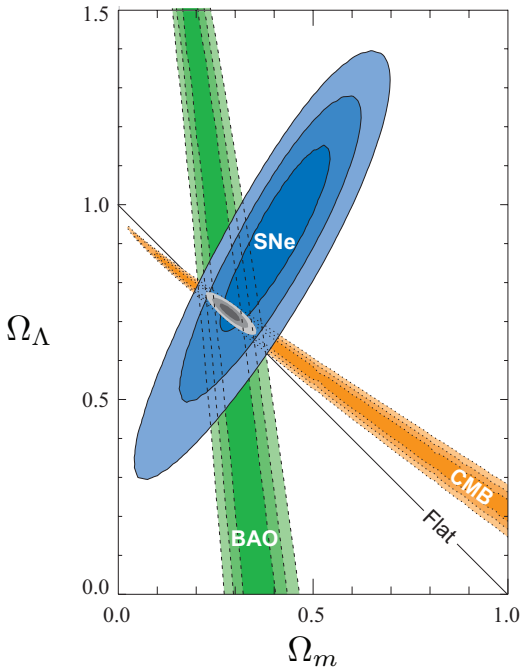


Figure 58: Evidence for dark energy. Shown are a combination of observations of the cosmic microwave background (CMB), supernovae (SNe) and baryon acoustic oscillations (BAO) [12].

It is also found that the universe has tiny ripples of adiabatic, scale-invariant, Gaussian density fluctuations. In the bulk of this lecture series I will describe how quantum fluctuations during inflation can explain the observed cosmological perturbations.

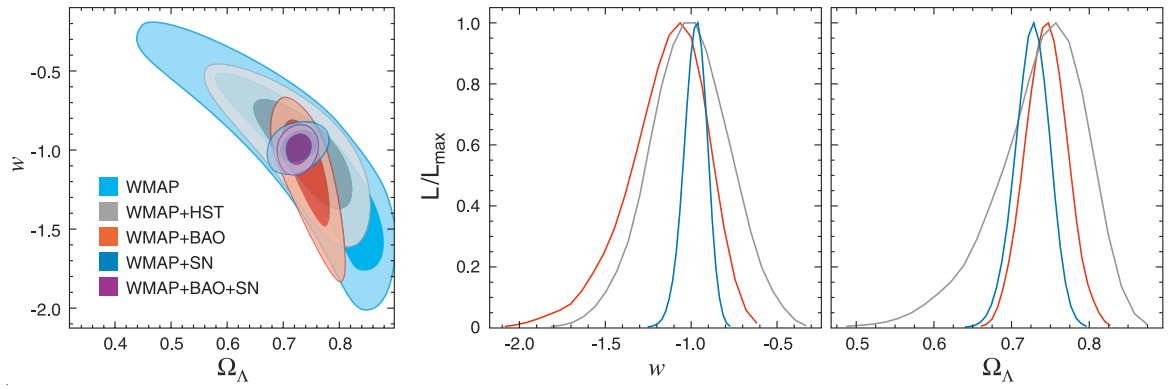


Figure 59: The properties of dark energy are close to a cosmological constant, $w_\Lambda \approx -1$ [11].

7.3.2 Big Bang Puzzles

It is somewhat of a philosophical questions whether initial conditions form part of a physical theory or should be considered separately. The purpose of physics is to predict the future evolution of a system given a set of initial conditions; *e.g.* Newton's laws of gravity will predict the path of a projectile if we define its initial position and velocity. It is therefore far from clear whether cosmology should predict or even just explain the initial conditions of the universe. On the other hand, it would be very disappointing if only very special and finely-tuned initial conditions would lead to the universe as we see it, making the observed universe an 'improbable accident'.

In this section we will explain that the conventional Big Bang theory requires precisely such a fine-tuned set of initial conditions to allow the universe to evolve to its current state. One of the major achievements of inflation is that it explains the initial conditions of the universe. Via inflation, the universe we know and love grew out of generic initial conditions.

The Cauchy Problem of the Universe

To specify the initial condition of the universe we consider a spatial slice of constant time Σ (we here won't worry about the gauge-dependence of the choice of Σ). On the 3-surface Σ we define the positions and velocities of all matter particles. The laws of gravity and fluid dynamics are then used to evolve the system forward in time.

- Initial Homogeneity

  We describe the spatial distribution of matter by its density and pressure as a function of coordinates $\mathbf{x}$, *i.e.* $\rho(\mathbf{x})$ and $p(\mathbf{x})$. In the previous section we assumed homogeneity and isotropy of the universe. Why is this a good assumption? Inhomogeneities are gravitationally unstable and therefore grow with time. Observations of the cosmic microwave background show that the inhomogeneities were much smaller in the past (at last-scattering) than today. One thus expects that these inhomogeneities were even smaller at yet earlier times. How do we explain the smoothness of the early universe?

  This is particularly surprising since we will show in §7.3.2 that in the conventional Big Bang picture the early universe (*e.g.* the CMB at last-scattering) consisted of a large number of causally-disconnected regions of space. In the Big Bang theory, there is no dynamical reason to explain why these causally-separated patches show such similar physical conditions. The homogeneity problem is therefore often called the *horizon problem*.

- Initial Velocities

In addition to specifying the initial density distribution, the complete characterization of the Cauchy problem of the universe requires the fluid velocities at every point in space. As we will see, to ensure that the universe remains homogeneous at late times requires the initial fluid velocities to take very precise values. If the initial velocities are just slightly too small, the universe recollapses within a fraction of a second. If they are just slightly too big, the universe expands too rapidly and quickly becomes nearly empty. The fine-tuning of initial velocities is made more dramatic by considering it in combination with the horizon problem. The fluid velocities need to be fine-tuned across causally-separated regions of space.

Since the difference between the potential energy and the kinetic energy defines the local curvature of a region of space (see **Exercise 1**), this fine-tuning of initial velocities is often called the *flatness problem*.

Horizon Problem

In the previous section, we defined the *comoving (particle) horizon*, τ, as the causal horizon or the maximum distance a light ray can travel between time 0 and time t

$$\boxed{\tau \equiv \int_0^t \frac{\mathrm{d}t'}{a(t')} = \int_0^a \frac{\mathrm{d}a}{Ha^2} = \int_0^a \mathrm{d}\ln a \left(\frac{1}{aH}\right)}. \tag{363}$$

Here, we have expressed the comoving horizon as an integral of the *comoving Hubble radius*, $(aH)^{-1}$, which plays a crucial role in inflation.

For a universe dominated by a fluid with equation of state w, we have

$$(aH)^{-1} = H_0^{-1} a^{\frac{1}{2}(1+3w)}\,. \tag{364}$$

Notice the dependence of the exponent on the combination $(1+3w)$. The qualitative behavior therefore depends on whether $(1+3w)$ is positive or negative. During the conventional Big Bang expansion ($w \gtrsim 0$) $(aH)^{-1}$ *grows* monotonically and the comoving horizon τ or the *fraction of the universe in causal contact increases with time*

$$\tau \propto a^{\frac{1}{2}(1+3w)}\,. \tag{365}$$

Again, the qualitative behavior depends on whether $(1+3w)$ is positive of negative. In particular, for radiation-dominated (RD) and matter-dominated (MD) universes we find

$$\tau = \int_0^a \frac{\mathrm{d}a}{Ha^2} \propto \begin{cases} a & \text{RD} \\ a^{1/2} & \text{MD} \end{cases}\,. \tag{366}$$

This means that the comoving horizon grows monotonically with time which implies that comoving scales entering the horizon today have been far outside the horizon at CMB decoupling.[23] But the near-homogeneity of the CMB tells us that the universe was extremely homogeneous at the time of last-scattering on scales encompassing many regions that *a priori* are causally independent. How is this possible?

[23] Recall that the comoving wavelength of a fluctuations is time-independent, while the comoving Hubble radius is time-dependent.

Flatness Problem

Exercise 1 (Flatness and Kinetic Energy)**.** *Show that the curvature parameter*

$$\Omega_k \equiv \Omega - 1 = \frac{\rho - \rho_{\rm crit}}{\rho_{\rm crit}}\,, \qquad \text{where} \quad \rho_{\rm crit} \equiv 3H^2\,,$$

may be interpreted as the difference between the average potential energy and the average kinetic energy of a region of space.

Spacetime in General Relativity is dynamical, curving in response to matter in the universe. Why then is the universe so closely approximated by flat Euclidean space? To quantify the problem we consider the Friedmann Equation

$$H^2 = \frac{1}{3}\rho(a) - \frac{k}{a^2}\,, \tag{367}$$

written as

$$\boxed{1 - \Omega(a) = \frac{-k}{(aH)^2}}\,, \tag{368}$$

where

$$\Omega(a) \equiv \frac{\rho(a)}{\rho_{\rm crit}(a)}\,, \qquad \rho_{\rm crit}(a) \equiv 3H(a)^2\,. \tag{369}$$

Notice that $\Omega(a)$ is now defined to be time-dependent, whereas the Ω's in the previous sections were constants, $\Omega(a_0)$. In standard cosmology the comoving Hubble radius, $(aH)^{-1}$, grows with time and from Eqn. (368) the quantity $|\Omega - 1|$ must thus diverge with time. The critical value $\Omega = 1$ is an *unstable fixed point*. Therefore, in standard Big Bang cosmology without inflation, the near-flatness observed today, $\Omega(a_0) \sim 1$, requires an extreme fine-tuning of Ω close to 1 in the early universe. More specifically, one finds that the deviation from flatness at Big Bang Nucleosynthesis (BBN), during the GUT era and at the Planck scale, respectively has to satisfy the following conditions

$$|\Omega(a_{\rm BBN}) - 1| \;\leq\; \mathcal{O}(10^{-16})\,, \tag{370}$$
$$|\Omega(a_{\rm GUT}) - 1| \;\leq\; \mathcal{O}(10^{-55})\,, \tag{371}$$
$$|\Omega(a_{\rm pl}) - 1| \;\leq\; \mathcal{O}(10^{-61})\,. \tag{372}$$

Another way of understanding the flatness problem is from the following differential equation

$$\frac{d\Omega}{d\ln a} = (1 + 3w)\Omega(\Omega - 1)\,. \tag{373}$$

Eqn. (373) is derived by differentiating Eqn. (368) and using the continuity equation (351). This makes it apparent that $\Omega = 1$ is an unstable fixed point if the strong energy condition is satisfied

$$\frac{d|\Omega - 1|}{d\ln a} > 0 \quad \Leftrightarrow \quad 1 + 3w > 0\,. \tag{374}$$

Again, why is $\Omega(a_0) \sim \mathcal{O}(1)$ and not much smaller or much larger?

On the Problem of Initial Conditions

We should emphasize that the flatness and horizon problems are *not* strict inconsistencies in the standard cosmological model. If one assumes that the initial value of Ω was extremely close to unity and that the universe began homogeneously over superhorizon distances (but with just the right level of inhomogeneity to explain structure formation) then the universe will continue

to evolve homogeneously in agreement with observations. The flatness and horizon problems are therefore really just severe shortcomings in the predictive power of the Big Bang model. The dramatic flatness of the early universe cannot be predicted by the standard model, but must instead be assumed in the initial conditions. Likewise, the striking large-scale homogeneity of the universe is not explained or predicted by the model, but instead must simply be assumed. A theory that explains these initial conditions dynamically seems very attractive.

7.3.3 A First Look at Inflation

The Shrinking Hubble Sphere

In §7.3.2 and §7.3.2 we emphasized the fundamental role of the comoving Hubble radius, $(aH)^{-1}$, in the horizon and flatness problems of the standard Big Bang cosmology. Both problems arise since in the conventional cosmology the comoving Hubble radius is strictly increasing. This suggest that all the Big Bang puzzles are solved by a beautifully simple idea: *invert the behavior of the comoving Hubble radius*, *i.e.* make is *decrease sufficiently in the very early universe.*

Comoving Horizon during Inflation

The evolution of the comoving horizon is of such crucial importance to the whole idea of inflation that it is worth being explicit about a few important points.

Recall the definition of the comoving horizon (= conformal time) as a logarithmic integral of the comoving Hubble radius

$$\boxed{\tau = \int_0^a \mathrm{d}\ln a' \, \frac{1}{a'H(a')}} \,. \tag{375}$$

Let us emphasize a subtle distinction between the comoving horizon τ and the comoving Hubble radius $(aH)^{-1}$ [8]:

> If particles are separated by distances greater than τ, they *never* could have communicated with one another; if they are separated by distances greater than $(aH)^{-1}$, they cannot talk to each other *now*! This distinction is crucial for the solution to the horizon problem which relies on the following: It is possible that τ is much larger than $(aH)^{-1}$ now, so that particles cannot communicate today but were in causal contact early on. From Eqn. (375) we see that this might happen if the comoving Hubble radius in the early universe was much larger than it is now so that τ got most of its contribution from early times. Hence, we require a phase of decreasing Hubble radius. Since H is approximately constant while a grows exponentially during inflation we find that the comoving Hubble radius decreases during inflation just as advertised.

Besides solving the Big Bang puzzles the decreasing comoving horizon during inflation is the key feature required for the quantum generation of cosmological perturbations described in the second lecture. I will describe how quantum fluctuations are generated on subhorizon scales, but exit the horizon once the Hubble radius becomes smaller than their comoving wavelength. In physical coordinates this corresponds to the superluminal expansion stretching perturbations to acausal distances. They become classical superhorizon density perturbations which re-enter the horizon in the subsequent Big Bang evolution and then gravitationally collapse to form the large-scale structure in the universe.

With this understanding of how the comoving horizon and the comoving Hubble radius evolve during inflation it is now almost trivial to explain how inflation solves the Big Bang puzzles.

Flatness Problem Revisited

Recall the Friedmann Equation (368) for a non-flat universe

$$|1 - \Omega(a)| = \frac{1}{(aH)^2} \,. \tag{376}$$

If the comoving Hubble radius decreases this *drives the universe toward flatness* (rather than away from it). This solves the flatness problem! The solution $\Omega = 1$ is an attractor during inflation.

Horizon Problem Revisited

A decreasing comoving horizon means that large scales entering the present universe were inside the horizon before inflation (see Figure 2). Causal physics before inflation therefore established spatial homogeneity. With a period of inflation, the uniformity of the CMB is not a mystery.

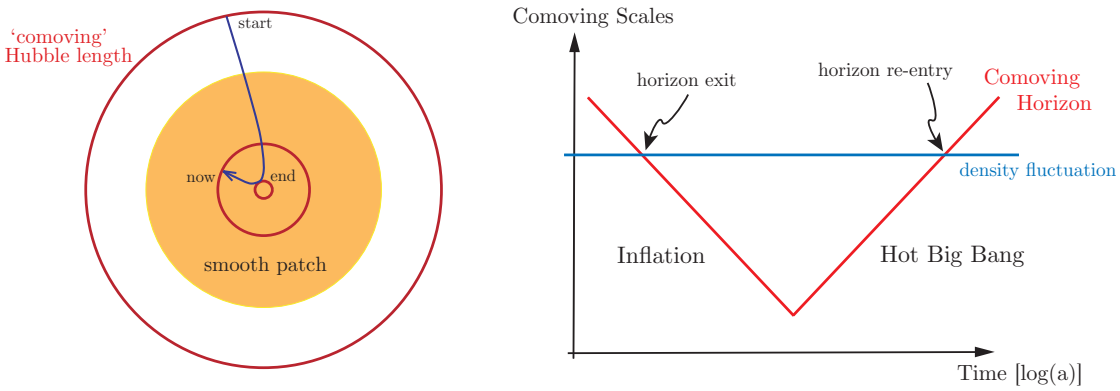


Figure 60: *Left:* Evolution of the comoving Hubble radius, $(aH)^{-1}$, in the inflationary universe. The comoving Hubble sphere shrinks during inflation and expands after inflation. Inflation is therefore a mechanism to 'zoom-in' on a smooth sub-horizon patch. *Right:* Solution of the horizon problem. All scales that are relevant to cosmological observations today were larger than the Hubble radius until $a \sim 10^{-5}$. However, at sufficiently early times, these scales were smaller than the Hubble radius and therefore causally connected. Similarly, the scales of cosmological interest came back within the Hubble radius at relatively recent times.

Conditions for Inflation

Via the Friedmann Equations a shrinking comoving Hubble radius can be related to the acceleration and the the pressure of the universe

$$\boxed{\frac{d}{dt}\left(\frac{H^{-1}}{a}\right) < 0 \quad \Rightarrow \quad \frac{d^2a}{dt^2} > 0 \quad \Rightarrow \quad \rho + 3p < 0} \,. \tag{377}$$

The three equivalent conditions for inflation therefore are:

- **Decreasing comoving horizon**

The shrinking Hubble sphere is defined as

$$\frac{d}{dt}\left(\frac{1}{aH}\right) < 0\,. \tag{378}$$

We used this as our fundamental definition of inflation since it most directly relates to the flatness and horizon problems and is key for the mechanism to generate fluctuations.

- **Accelerated expansion**

  From the relation

  $$\frac{d}{dt}(aH)^{-1} = \frac{-\ddot{a}}{(aH)^2}\,, \tag{379}$$

  we see immediately that a shrinking comoving Hubble radius implies accelerated expansion

  $$\frac{d^2a}{dt^2} > 0\,. \tag{380}$$

  This explains why inflation is often defined as a period of accelerated expansion. The second time derivative of the scale factor may of course be related to the first time derivative of the Hubble parameter H

  $$\frac{\ddot{a}}{a} = H^2(1-\varepsilon)\,, \qquad \text{where} \quad \varepsilon \equiv -\frac{\dot{H}}{H^2}\,. \tag{381}$$

  Acceleration therefore corresponds to

  $$\boxed{\varepsilon = -\frac{\dot{H}}{H^2} = -\frac{d\ln H}{dN} < 1} \tag{382}$$

  Here, we have defined $\mathrm{d}N = H\mathrm{d}t = \mathrm{d}\ln a$, which measures the number of e-folds N of inflationary expansion. Eqn. (382) therefore means that the fractional change of the Hubble parameter per e-fold is small.

- **Negative pressure**

  What stress-energy can source acceleration? Consulting Eqn. (349) we infer that $\ddot{a} > 0$ requires

  $$p < -\frac{1}{3}\rho\,, \tag{383}$$

  i.e. negative pressure or a violation of the strong energy condition (SEC). How this can arise in a physical theory will be explained in §7.3.4. We will see that there is nothing sacred about the SEC and it can easily be violated.

Conformal Diagram of Inflation

A truly illuminating way of visualizing inflation is with the aid of a conformal spacetime diagram. Recall from §7.3.1 the flat FRW metric in conformal time $\mathrm{d}\tau = \mathrm{d}t/a(t)$

$$\mathrm{d}s^2 = a^2(\tau)\left[-\mathrm{d}\tau^2 + \mathrm{d}\mathbf{x}^2\right]\,. \tag{384}$$

Also recall that in conformal coordinates null geodesics ($\mathrm{d}s^2 = 0$) are always at 45° angles, $\mathrm{d}\tau = \pm\sqrt{\mathrm{d}\mathbf{x}^2} \equiv \pm\mathrm{d}r$. Since light determines the causal structure of spacetime this provides a nice way to study horizons in inflationary cosmology.

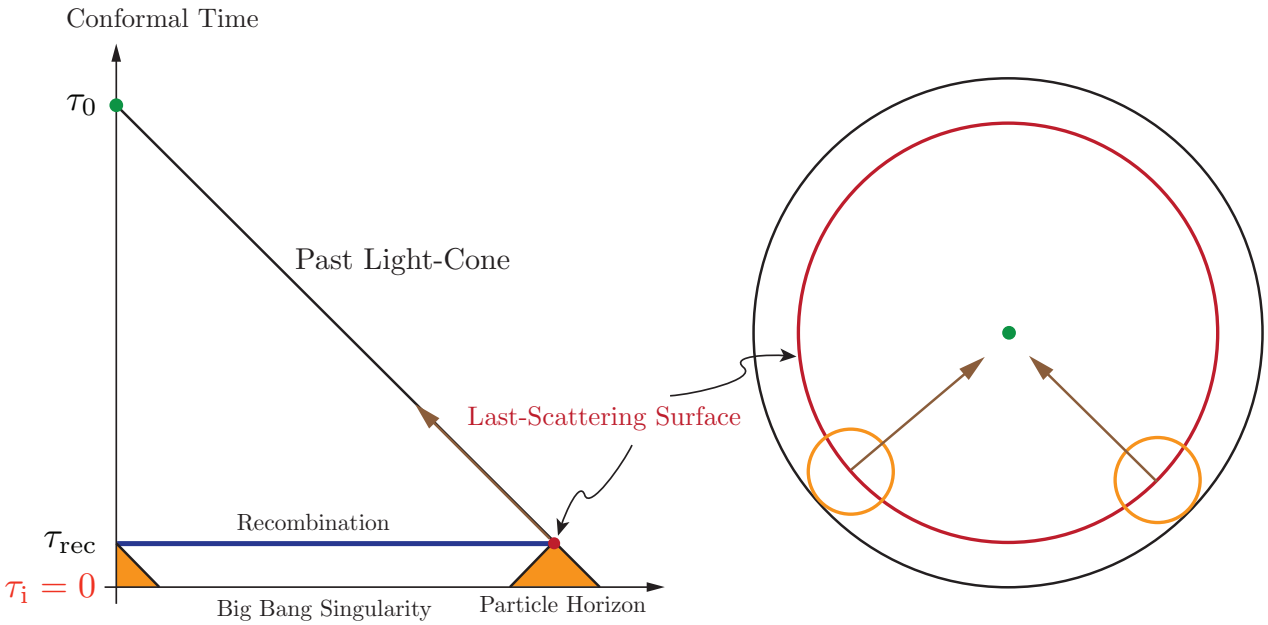


Figure 61: Conformal diagram of Big Bang cosmology. The CMB at last-scattering (recombination) consists of 10^5 causally disconnected regions!

During matter or radiation domination the scale factor evolves as

$$a(\tau) \propto \begin{cases} \tau & \text{RD} \\ \tau^2 & \text{MD} \end{cases} . \tag{385}$$

If and only if the universe had always been dominated by matter or radiation, this would imply the existence of the *Big Bang singularity* at $\tau_{\rm i} = 0$

$$a(\tau_{\rm i} \equiv 0) = 0 \, . \tag{386}$$

The conformal diagram corresponding to standard Big Bang cosmology is given in Figure 61. The horizon problem is apparent. Each spacetime point in the conformal diagram has an associated past light cone which defines its causal past. Two points on a given τ = constant surface are in causal contact if their past light cones intersect at the Big Bang, $\tau_{\rm i} = 0$. This means that the surface of last-scattering ($\tau_{\rm CMB}$) consisted of many causally disconnected regions that won't be in thermal equilibrium. The uniformity of the CMB on large scales hence becomes a serious puzzle.

During inflation ($H \approx$ const.), the scale factor is

$$a(\tau) = -\frac{1}{H\tau} \, , \tag{387}$$

and the singularity, $a = 0$, is pushed to the infinite past, $\tau_{\rm i} \to -\infty$. The scale factor (387) becomes infinite at $\tau = 0$! This is because we have assumed de Sitter space with $H =$ const., which means that inflation will continue forever with $\tau = 0$ corresponding to the infinite future $t \to +\infty$. In reality, inflation ends at some finite time, and the approximation (387) although valid at early times, breaks down near the end of inflation. So the surface $\tau = 0$ is not the Big Bang, but the end of inflation. The initial singularity has been pushed back arbitrarily far in conformal time $\tau \ll 0$, and light cones can extend through the apparent Big Bang so that apparently disconnected points are in causal contact. In other words, because of inflation, 'there was more (conformal) time before recombination than we thought'. This is summarized in the conformal diagram in Figure 62.

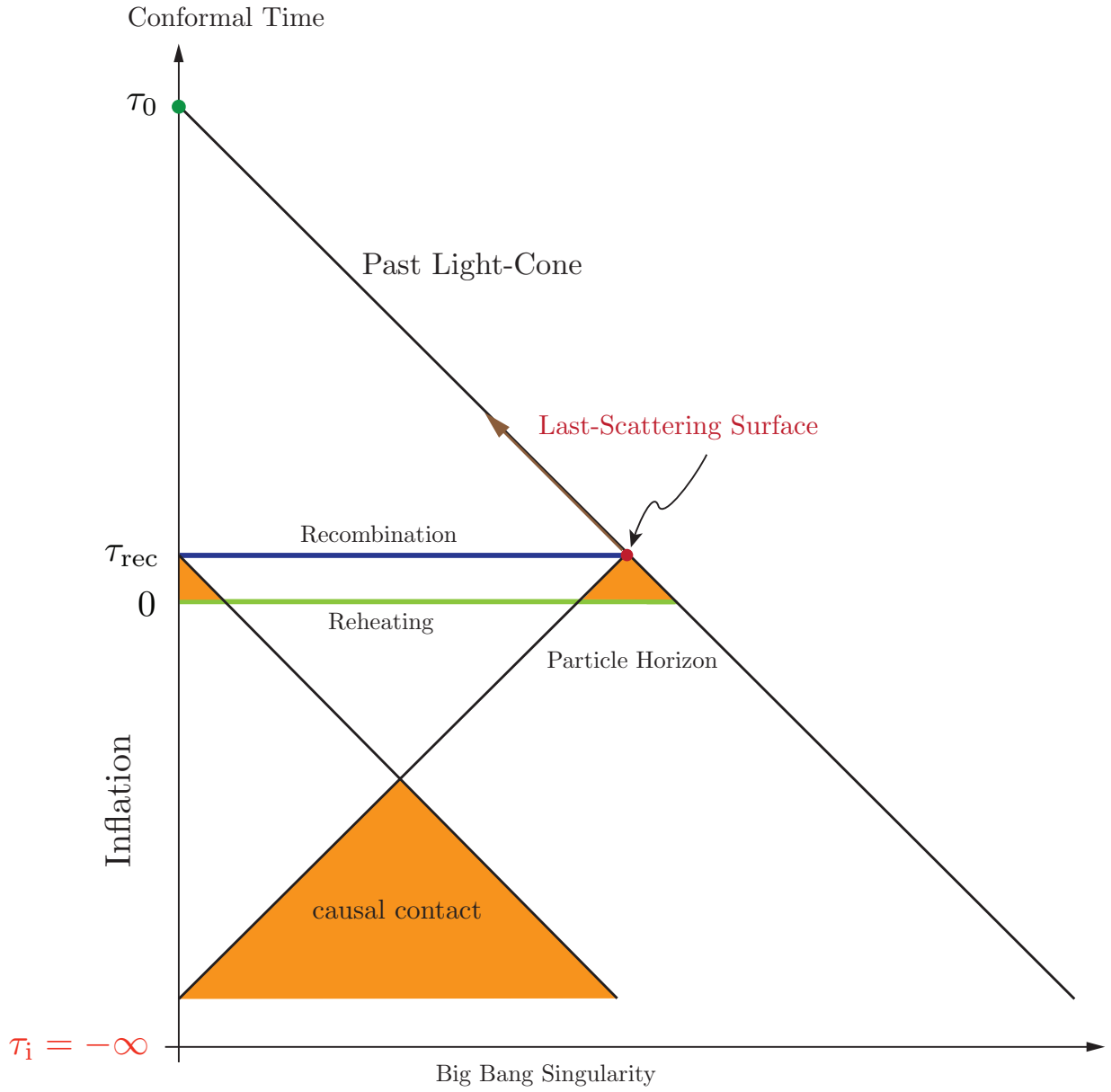


Figure 62: Conformal diagram of inflationary cosmology. Inflation extends conformal time to negative values! The end of inflation creates an "apparent" Big Bang at $\tau = 0$. There is, however, no singularity at $\tau = 0$ and the light cones intersect at an earlier time if inflation lasts for at least 60 e-folds.

7.3.4 The Physics of Inflation

Inflation is a very unfamiliar physical phenomenon: within a fraction a second the universe grew exponential at an accelerating rate. In Einstein gravity this requires a negative pressure source or equivalently a nearly constant energy density. In this section we describe the physical conditions under which this can arise.

Scalar Field Dynamics

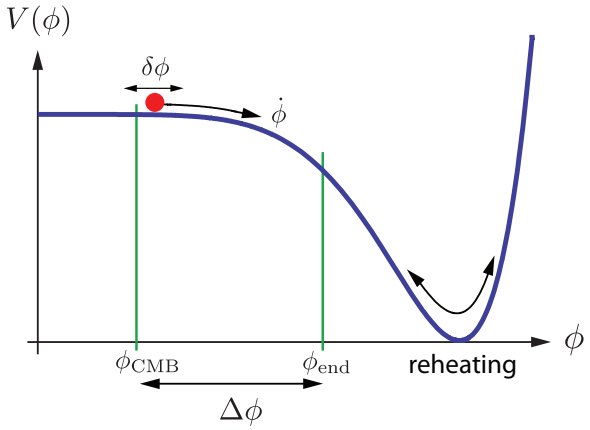


Figure 63: Example of an inflaton potential. Acceleration occurs when the potential energy of the field, $V(\phi)$, dominates over its kinetic energy, $\frac{1}{2}\dot{\phi}^2$. Inflation ends at $\phi_{\rm end}$ when the kinetic energy has grown to become comparable to the potential energy, $\frac{1}{2}\dot{\phi}^2 \approx V$. CMB fluctuations are created by quantum fluctuations $\delta\phi$ about 60 e-folds before the end of inflation. At reheating, the energy density of the inflaton is converted into radiation.

The simplest models of inflation involve a single scalar field ϕ, the *inflaton*. Here, we don't specify the physical nature of the field ϕ, but simply use it as an order parameter (or clock) to parameterize the time-evolution of the inflationary energy density. The dynamics of a scalar field (minimally) coupled to gravity is governed by the action

$$\boxed{S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2}R + \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\,\partial_\nu\phi - V(\phi)\right]} = S_{\rm EH} + S_\phi\,. \tag{388}$$

The action (388) is the sum of the gravitational Einstein-Hilbert action, $S_{\rm EH}$, and the action of a scalar field with canonical kinetic term, S_ϕ. The potential $V(\phi)$ describes the self-interactions of the scalar field. The energy-momentum tensor for the scalar field is

$$T^{(\phi)}_{\mu\nu} \equiv -\frac{2}{\sqrt{-g}}\frac{\delta S_\phi}{\delta g^{\mu\nu}} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left(\frac{1}{2}\partial^\sigma\phi\partial_\sigma\phi + V(\phi)\right). \tag{389}$$

The field equation of motion is

$$\frac{\delta S_\phi}{\delta\phi} = \frac{1}{\sqrt{-g}}\partial_\mu(\sqrt{-g}\partial^\mu\phi) + V_{,\phi} = 0\,, \tag{390}$$

where $V_{,\phi} = \frac{dV}{d\phi}$. Assuming the FRW metric (328) for $g_{\mu\nu}$ and restricting to the case of a homogeneous field $\phi(t,\mathbf{x}) \equiv \phi(t)$, the scalar energy-momentum tensor takes the form of a perfect fluid (347) with

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi)\,, \tag{391}$$

$$p_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi)\,. \tag{392}$$

The resulting equation of state

$$\boxed{w_\phi \equiv \frac{p_\phi}{\rho_\phi} = \frac{\frac{1}{2}\dot{\phi}^2 - V}{\frac{1}{2}\dot{\phi}^2 + V}}\,, \tag{393}$$

shows that a scalar field can lead to negative pressure ($w_\phi < 0$) and accelerated expansion ($w_\phi < -1/3$) if the potential energy V dominates over the kinetic energy $\frac{1}{2}\dot{\phi}^2$. The dynamics of the (homogeneous) scalar field and the FRW geometry is determined by

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = 0 \qquad \text{and} \qquad H^2 = \frac{1}{3}\left(\frac{1}{2}\dot{\phi}^2 + V(\phi)\right)\,. \tag{394}$$

For large values of the potential, the field experiences significant Hubble friction from the term $H\dot{\phi}$.

Slow-Roll Inflation

The acceleration equation for a universe dominated by a homogeneous scalar field can be written as follows

$$\frac{\ddot{a}}{a} = -\frac{1}{6}(\rho_\phi + 3p_\phi) = H^2(1-\varepsilon)\,, \tag{395}$$

where

$$\varepsilon \equiv \frac{3}{2}(w_\phi + 1) = \frac{1}{2}\frac{\dot{\phi}^2}{H^2}\,. \tag{396}$$

The so-called *slow-roll parameter* ε may be related to the evolution of the Hubble parameter

$$\boxed{\varepsilon = -\frac{\dot{H}}{H^2}} = -\frac{d\ln H}{dN}\,, \tag{397}$$

where $\mathrm{d}N = H\mathrm{d}t$. Accelerated expansion occurs if $\varepsilon < 1$. The de Sitter limit, $p_\phi \to -\rho_\phi$, corresponds to $\varepsilon \to 0$. In this case, the potential energy dominates over the kinetic energy

$$\dot{\phi}^2 \ll V(\phi)\,. \tag{398}$$

Accelerated expansion will only be sustained for a sufficiently long period of time if the second time derivative of ϕ is small enough

$$|\ddot{\phi}| \;\ll\; |3H\dot{\phi}|\,,\, |V_{,\phi}|\,. \tag{399}$$

This requires smallness of a second slow-roll parameter

$$\eta = -\frac{\ddot{\phi}}{H\dot{\phi}} = \varepsilon - \frac{1}{2\varepsilon}\frac{d\varepsilon}{dN}\,, \tag{400}$$

where $|\eta| < 1$ ensures that the fractional change of ε per e-fold is small. The slow-roll conditions, $\varepsilon, |\eta| < 1$, may also be expressed as conditions on the shape of the inflationary potential

$$\boxed{\epsilon_{\rm v}(\phi) \equiv \frac{M_{\rm pl}^2}{2}\left(\frac{V_{,\phi}}{V}\right)^2}\,, \tag{401}$$

and

$$\boxed{\eta_{\rm v}(\phi) \equiv M_{\rm pl}^2\frac{V_{,\phi\phi}}{V}}\,. \tag{402}$$

Here, we temporarily reintroduced the Planck mass to make $\epsilon_{\rm v}$ and $\eta_{\rm v}$ manifestly dimensionless. In the following we will set $M_{\rm pl}$ to one again. In the slow-roll regime

$$\epsilon_{\rm v}, |\eta_{\rm v}| \ll 1\,, \tag{403}$$

the background evolution is

$$H^2 \approx \frac{1}{3}V(\phi) \approx \text{const.}\,, \tag{404}$$

$$\dot\phi \approx -\frac{V_{,\phi}}{3H}\,, \tag{405}$$

and the spacetime is approximately *de Sitter*

$$a(t) \sim e^{Ht}\,. \quad H \approx \text{const}\,. \tag{406}$$

The parameters $\epsilon_{\rm v}$ and $\eta_{\rm v}$ are called the *potential slow-roll parameters* to distinguish them from the *Hubble slow-roll parameters* ε and η. In the slow-roll approximation the Hubble and potential slow-roll parameters are related as follows (see Appendix 7.8.4)

$$\varepsilon \approx \epsilon_{\rm v}\,, \qquad \eta \approx \eta_{\rm v} - \epsilon_{\rm v}\,. \tag{407}$$

Inflation ends when the slow-roll conditions are violated

$$\varepsilon(\phi_{\rm end}) \equiv 1\,, \qquad \epsilon_{\rm v}(\phi_{\rm end}) \approx 1\,. \tag{408}$$

The number of e-folds before inflation ends is

$$\begin{aligned} N(\phi) &\equiv \ln\frac{a_{\rm end}}{a} \\ &= \int_t^{t_{\rm end}} H{\rm d}t = \int_\phi^{\phi_{\rm end}} \frac{H}{\dot\phi}{\rm d}\phi \approx \int_{\phi_{\rm end}}^{\phi} \frac{V}{V_{,\phi}}{\rm d}\phi\,, \end{aligned} \tag{409}$$

where we used the slow-roll results (404) and (405). The result (409) may also be written as

$$\boxed{N(\phi) = \int_{\phi_{\rm end}}^{\phi} \frac{{\rm d}\phi}{\sqrt{2\varepsilon}} \approx \int_{\phi_{\rm end}}^{\phi} \frac{{\rm d}\phi}{\sqrt{2\epsilon_{\rm v}}}}\,. \tag{410}$$

To solve the horizon and flatness problems requires that the total number of inflationary e-folds exceeds about 60,

$$N_{\rm tot} \equiv \ln\frac{a_{\rm end}}{a_{\rm start}} \gtrsim 60\,. \tag{411}$$

The precise value depends on the energy scale of inflation and on the details of reheating after inflation. The fluctuations observed in the CMB are created $N_{\rm cmb} \approx 40-60$ e-folds before the end of inflation (the precise value again depending on the details of reheating and the post-inflationary thermal history of the universe). The following integral constraint gives the corresponding field value $\phi_{\rm cmb}$

$$\int_{\phi_{\rm end}}^{\phi_{\rm cmb}} \frac{{\rm d}\phi}{\sqrt{2\epsilon_{\rm v}}} = N_{\rm cmb} \approx 40-60\,. \tag{412}$$

Case Study: $m^2\phi^2$ Inflation

As an example, let us give the slow-roll analysis of arguably the simplest model of inflation: single field inflation driven by a mass term

$$V(\phi) = \frac{1}{2}m^2\phi^2\,. \tag{413}$$

The slow-roll parameters are

$$\epsilon_{\rm v}(\phi) = \eta_{\rm v}(\phi) = 2\left(\frac{M_{\rm pl}}{\phi}\right)^2\,. \tag{414}$$

To satisfy the slow-roll conditions $\epsilon_{\rm v}, |\eta_{\rm v}| < 1$, we need to consider super-Planckian values for the inflaton

$$\phi > \sqrt{2}M_{\rm pl} \equiv \phi_{\rm end}\,. \tag{415}$$

The relation between the inflaton field value and the number of e-folds before the end of inflation is

$$N(\phi) = \frac{\phi^2}{4M_{\rm pl}^2} - \frac{1}{2}\,. \tag{416}$$

Fluctuations observed in the CMB are created at

$$\phi_{\rm cmb} = 2\sqrt{N_{\rm cmb}}\,M_{\rm pl} \sim 15M_{\rm pl}\,. \tag{417}$$

In the next lecture we will come back to this example when we compute the fluctuation spectrum generate by $m^2\phi^2$ inflation.

Exercise 2 ($m^2\phi^2$ Inflation)**.** *Verify the above slow-roll results for $m^2\phi^2$ inflation.*

Reheating

After inflation ends the scalar field begins to oscillate around the minimum of the potential. During this phase of coherent oscillations the scalar field acts like pressureless matter

$$\frac{d\bar{\rho}_\phi}{dt} + 3H\bar{\rho}_\phi = 0\,. \tag{418}$$

Exercise 3 (Coherent Scalar Field Oscillations)**.** *Confirm Eqn. (418) from the equations of motion for ϕ.*

The coupling of the inflaton field to other particles leads to a decay of the inflaton energy

$$\frac{d\bar{\rho}_\phi}{dt} + (3H + \Gamma_\phi)\bar{\rho}_\phi = 0\,. \tag{419}$$

The coupling parameter Γ_ϕ depends on complicated and model-dependent physical processes that we do not have the time to review. Eventually, the inflationary energy density is converted into standard model degrees of freedom and the hot Big Bang commences.

Reheating is a rich and complicated subject to which we couldn't do justice to in these lectures. We refer the interested reader to the review by Bassett *et al.* [13] for more details.

Models of Inflation

The fundamental microscopic origin of inflation is still a mystery. Basic questions like: what is the inflaton? what is the shape of the inflationary potential? and why did the universe start in a high energy state? remain unanswered. The challenge to explain the physics of inflation is considerable. Inflation is believed to have occurred at an enormous energy scale (maybe as high as $\sim 10^{15}$ GeV), far out of reach of terrestrial particle accelerators. Any description of the inflationary era therefore requires a considerable extrapolation of the known laws of physics, and until recently, only a phenomenological parameterization of the inflationary dynamics was possible.[24] In this approach, a suitable inflationary potential function $V(\phi)$ is postulated (see Figures 63 and 64 for two popular examples) and the experimental predictions are computed from that. As we will see in the next lecture, details of the primordial fluctuation spectra will depend on the precise shape of the inflaton potential.

Single-Field Slow-Roll Inflation

The definition of an inflationary model amounts to a specification of the inflaton action (potential and kinetic terms) and its coupling to gravity. So far we have phrased our discussion of inflation in terms of the simplest models, single-field slow-roll inflation, characterized by the following action

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2}R + \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\,\partial_\nu\phi - V(\phi)\right] . \tag{420}$$

The dynamics of the inflaton field, from the time when CMB fluctuations were created (see **Lecture 2**) at $\phi_{\rm cmb}$ to the end of inflation at $\phi_{\rm end}$, is determined by the shape of the inflationary potential $V(\phi)$. The different possibilities for $V(\phi)$ can be classified in a useful way by determining whether they allow the inflaton field to move over a large or small distance $\Delta\phi \equiv \phi_{\rm cmb} - \phi_{\rm end}$, as measured in Planck units.

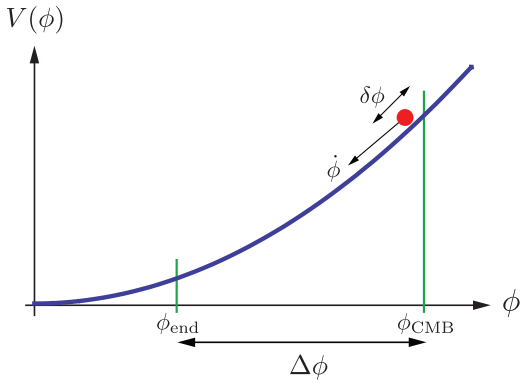


Figure 64: Large-field inflation. In an important class of inflationary models the inflationary dynamics is driven by a single monomial term in the potential, $V(\phi) \propto \phi^p$. In these models the inflaton field evolves over a super-Planckian range during inflation, $\Delta\phi > M_{\rm pl}$, and a large amplitude of gravitational waves is produced by quantum mechanical fluctuations (see **Lecture 2**).

1. *Small-Field Inflation*

   In small-field models the field moves over a small (sub-Planckian) distance: $\Delta\phi < M_{\rm pl}$. This is relevant for future observations because small-field models predict that the amplitude of the gravitational waves produced during inflation is too small to be detected

[24] Recently, progress has been made both in a systematic effective field theory description of inflation [14, 15] and in top-down derivations of inflationary potentials from string theory [16].

(see **Lecture 2**). The potentials that give rise to such small-field evolution often arise in mechanisms of *spontaneous symmetry breaking*, where the field rolls off an unstable equilibrium toward a displaced vacuum (see Fig. 63). A simple example is the Higgs-like potential

$$V(\phi) = V_0 \left[1 - \left(\frac{\phi}{\mu}\right)^2\right]^2 . \tag{421}$$

More generally, small-field models can be locally approximated by the following expansion

$$V(\phi) = V_0 \left[1 - \left(\frac{\phi}{\mu}\right)^p\right] + \cdots \tag{422}$$

where the dots represent higher-order terms that become important near the end of inflation and during reheating.

Historically, a famous inflationary potential is the *Coleman-Weinberg potential* [170, 15]

$$V(\phi) = V_0 \left[\left(\frac{\phi}{\mu}\right)^4 \left(\ln\left(\frac{\phi}{\mu}\right) - \frac{1}{4}\right) + \frac{1}{4}\right] , \tag{423}$$

which arises as the potential for radiatively-induced symmetry breaking in electroweak and grand unified theories. Although the original values of the parameters V_0 and μ based on the $SU(5)$ theory are incompatible with the small amplitude of inflationary fluctuations, the Coleman-Weinberg potential remains a popular phenomenological model (see *e.g.* [17]).

2. *Large-Field Inflation*

In large-field models the inflaton field starts a large field values and then evolves to a minimum at the origin $\phi = 0$. If the field evolution is super-Planckian, $\Delta\phi > M_{\rm pl}$, the gravitational waves produced by inflation should be observed in the near future.

The prototypical large-field model is *chaotic inflation* where a single monomial term dominates the potential (see Fig. 64)

$$V(\phi) = \lambda_p \phi^p . \tag{424}$$

For such a potential the slow-roll parameters are small for super-Planckian field values, $\phi \gg M_{\rm pl}$ (notice that the slow-roll conditions are independent of the coupling constant λ_p). However, to arrange for a small amplitude of density fluctuations (see **Lecture 2**) the inflaton self-coupling has to be very small, $\lambda_p \ll 1$. This condition automatically guarantees that the potential energy (density) is sub-Planckian, $V \ll M_{\rm pl}^4$, and quantum gravity effects are not necessarily important (but see §29.2 in **Lecture 5**).

One of the most elegant inflationary models is *natural inflation* where the potential takes the following form (see Fig. 65)

$$V(\phi) = V_0 \left[\cos\left(\frac{\phi}{f}\right) + 1\right] . \tag{425}$$

This potential often arises if the inflaton field is taken to be an axion. Depending on the parameter f the model can be of the small-field or large-field type. However, it is particularly attractive to consider natural inflation for large-field variations, $2\pi f > M_{\rm pl}$, since for axions a shift symmetry can be employed to protect the potential from correction terms even over large field ranges (see §29.2).

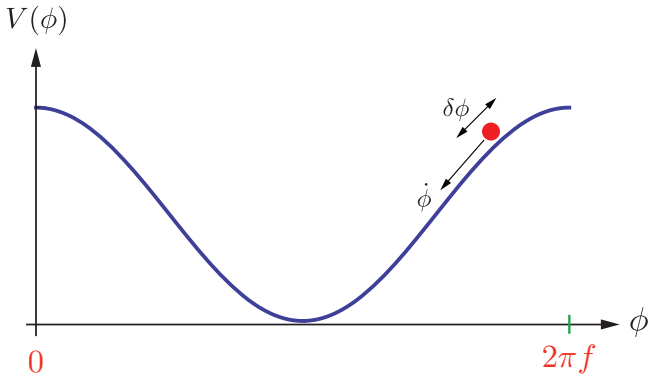


Figure 65: Natural Inflation. If the periodicity $2\pi f$ is super-Planckian the model can naturally support a large gravitational wave amplitude.

The Overshoot Problem

The Patch Problem

Beyond Single-Field Slow-Roll

The possibilities for getting inflationary expansion are (maybe frustratingly) varied. Inflation is a paradigm, a framework for a theory of the early universe, but it is not a unique theory. A large number of phenomenological models has been proposed with different theoretical motivations and observational predictions. For the majority of these lectures we will focus on the simplest single-field slow-roll models that we just described. However, in this short section we want to relieve ourselves from the sin of not mentioning the broader landscape of inflationary model-building (see also Ref. [18]).

The simplest inflationary actions (420) may be extended in a number of obvious ways:

1. *Non-minimal coupling to gravity.*

   The action (420) is called minimally coupled in the sense that there is no direct coupling between the inflaton field and the metric. In principle, we could imagine a non-minimal coupling between the inflaton and the graviton, however, in practice, non-minimally coupled theories can be written as minimally coupled theories by a field redefinition.

2. *Modified gravity.*

   Similarly, we could entertain the possibility that the Einstein-Hilbert part of the action is modified at high energies. However, the simplest examples for this UV modification of gravity, so-called $f(R)$ theories, can again be transformed into a minimally coupled scalar field with potential $V(\phi)$.

3. *Non-canonical kinetic term.*

   The action (420) has a canonical kinetic term

   $$\mathcal{L}_\phi = X - V(\phi)\,, \qquad X \equiv \frac{1}{2} g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi\,. \tag{426}$$

   Inflation can then only occur if the potential $V(\phi)$ is very flat. More generally, however, we could imagine that the high-energy theory has fields with non-canonical kinetic terms

   $$\mathcal{L}_\phi = F(\phi, X) - V(\phi)\,, \tag{427}$$

   where $F(\phi, X)$ is some function of the inflaton field and its derivatives. For actions such as (427) it is possible that inflation is driven by the kinetic term and occurs even in the presence of a steep potential.

4. *More than one field.*

   If we allow more than one field to be dynamically relevant during inflation, then the possibilities for the inflationary dynamics (and the mechanisms for the production of fluctuations) expand dramatically and the theory loses a lot of its predictive power. Some of the large number of possibilities of multi-field inflationary models are reviewed in Ref. [19].

Multi-Field Inflation

Non-Trival Kinetic Terms

7.4 Quantum Fluctuations during Inflation by Baumann 2th lecture

Abstract

In this lecture we present the famous calculation of the primordial fluctuation spectra generated by quantum fluctuations during inflation. We present the calculation in full detail and try to avoid 'cheating' and approximations. After a brief review of fundamental aspects of cosmological perturbation theory, we first give a qualitative summary of the basic mechanism by which inflation converts microscopic quantum fluctuations into macroscopic seeds for cosmological structure formation. As a pedagogical introduction to quantum field theory in curved spacetime we then review the quantization of the simple harmonic oscillator. We emphasize that a unique vacuum state is chosen by demanding that the vacuum is the minimum energy state. We then proceed by giving the corresponding calculation for inflation. We calculate the power spectra of both scalar and tensor fluctuations and discuss their dependence on scale.

In the last lecture we studied the classical ($\hbar = 0$) dynamics of a scalar field rolling down a potential with speed $\dot{\phi}$ (see Fig. 66). In this lecture we study the effects of quantum ($\hbar \neq 0$) fluctuations around the classical background evolution $\bar{\phi}(t)$. These fluctuations lead to a local time delay in the time at which inflation ends, *i.e.* different parts of the universe will end inflation at slightly different times. For instance, for the potential shown in Fig. 66 regions acquiring a negative frozen fluctuations $\delta\phi$ remain potential-dominated longer than regions with positive $\delta\phi$. Different parts of the universe therefore undergo slightly different evolutions. This induces relative density fluctuations $\delta\rho(t, \mathbf{x})$.

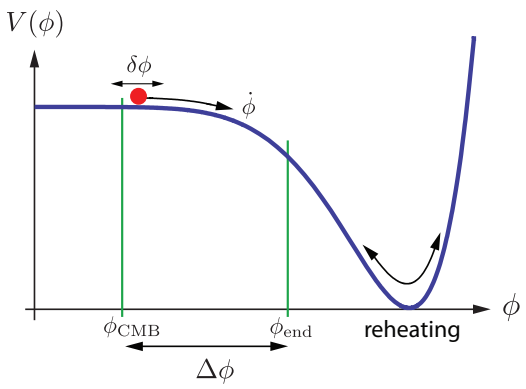


Figure 66: Quantum fluctuations $\delta\phi(t, \mathbf{x})$ around the classical background evolution $\bar{\phi}(t)$.

In this lecture we will discuss the technical details underlying this basic picture for the quantum origin of large-scale structure.

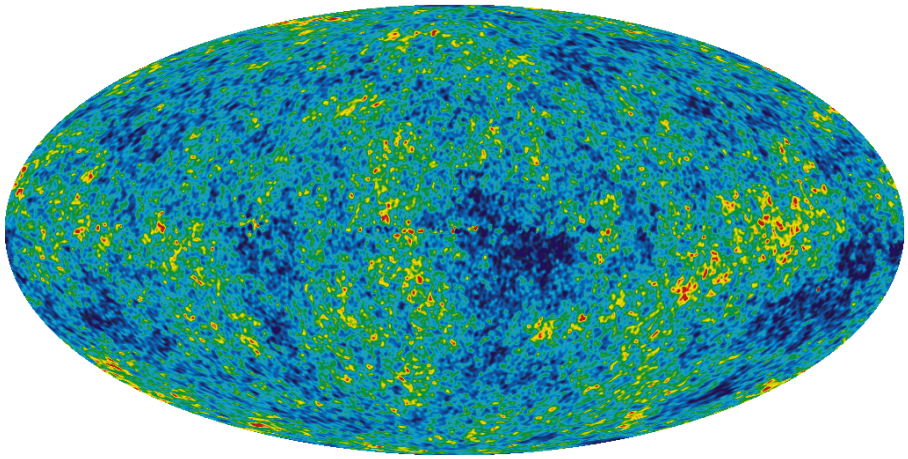


Figure 67: Observations of the CMB anisotropies prove that the early universe wasn't perfectly homogeneous. However, the observations also show that the inhomogeneities were small and can therefore be analyzed as linear perturbations around a homogenous background.

7.4.1 *Review*: Cosmological Perturbations

In this lecture we present in detail the generation of cosmological perturbations from quantum fluctuations during inflation. This discussion will require some background in cosmological perturbation theory which we now briefly review. More details may be found in Appendix A.

Generalities

Linear Perturbations

Observations of the CMB (Fig. 67) explain the success of cosmological perturbation theory. At the time of decoupling the universe was very nearly homogeneously with small inhomogeneities at the 10^{-5} level. A natural strategy therefore is to split all quantities $X(t, \mathbf{x})$ (metric $g_{\mu\nu}$ and matter fields $T_{\mu\nu} \to \phi$ ρ, p, etc.) into a homogeneous background $\bar{X}(t)$ that depends only on cosmic time and a spatially dependent perturbation

$$\delta X(t, \mathbf{x}) \equiv X(t, \mathbf{x}) - \bar{X}(t) \,. \tag{428}$$

Because the perturbations are small, $\delta X \ll \bar{X}$, expanding the Einstein Equations at linear order in perturbations approximates the full non-linear solution to very high accuracy

$$\delta G_{\mu\nu} = 8\pi G \, \delta T_{\mu\nu} \,. \tag{429}$$

Gauge Choice

A crucial subtlety in the study of cosmological perturbations is the fact that the split into background and perturbations, Eqn. (428), is not unique, but depends on the choice of coordinates or the *gauge choice*.[25] When we described the homogeneous universe in **Lecture 1** we introduced coordinates t and $\mathbf{x}$ to define the FRW metric. The spacelike hypersurfaces

[25]The perturbation δX in any relevent quantity, say represented by a tensor field X, is define as the difference between the value X has in the physical spacetime (the perturbed spacetime), and the value X_0 the same quantity has in the given (unperturbed) background spacetime. However, it is a basic fact of differential geometry that, in order to make the comparison of tensors meaningful, they can be compared only after a prescription for

of constant time t defined the *slicing* to the four-dimensional spacetime, while the timelike worldlines of constant $\mathbf{x}$ defined the *threading*. The FRW threading corresponds to the motion of comoving observers who see zero momentum density at their location. These observers a free-falling and the expansion defined by them is isotropic. The slicing is orthogonal to the threading with each spacelike slice corresponding to a homogeneous universe. These features made our coordinate choice so distinguished that we never worried about other coordinates (in which the universe would not look homogeneous and isotropic). However, now that we are considering perturbations it is important to realize that the slicing and threading of the perturbed spacetime is not unique. Furthermore, when describing an inhomogeneous spacetime there is often not a preferred coordinate choice. When we make a gauge choice to define the slicing and threading of the spacetime we implicitly also define the perturbations. If we aren't careful this gauge dependence of perturbations can lead to some confusion. To demonstrate this fact most dramatically consider an *unperturbed* homogeneous and isotropic universe, where the energy density is only a function of time, $\rho(t, \mathbf{x}) = \rho(t)$. We now show that a change of the time coordinate can introduce *fictitious* perturbations $\delta\rho$. Consider a new time coordinate $\tilde{t} = t + \delta t(t, \mathbf{x})$. In general, the energy density on the new time-slice will not be homogeneous, $\tilde{\rho}(\tilde{t}, \mathbf{x}) = \rho(t(\tilde{t}, \mathbf{x}))$. These perturbations in the energy density aren't physical, but entirely due to the choice of new time-slicing. Similarly, we can remove a *real* perturbation in the energy density by choosing the hypersurface of constant time to coincide with the hypersurface of constant energy density. Then $\delta\tilde{\rho} = 0$ although there are real inhomogeneities. To resolve ambiguities between real and fake perturbations in general relativity, we need to consider the complete set of perturbations, *i.e.* we need both the matter field perturbations and the metric perturbations and by a gauge transformation we can trade one for the other. To avoid misinterpretation of fictitious gauge modes it will also be useful to study gauge-invariant combinations of perturbations. By definition, fluctuations of gauge-invariant quantities cannot be removed by a coordinate transformation.

Scalars, Vectors and Tensors

The spatially flat, homogeneous and isotropic background spacetime possesses a great deal of symmetry. These symmetries allow a decomposition of the metric and the stress-energy perturbations into independent scalar (S), vector (V) and tensor (T) components. This *SVT decomposition* is most easily described in Fourier space

$$X_{\mathbf{k}}(t) = \int \mathrm{d}^3\mathbf{x}\; X(t, \mathbf{x})\, e^{i\mathbf{k}\cdot\mathbf{x}}\,, \qquad X \equiv \delta\phi,\, \delta g_{\mu\nu}, \text{etc}\,. \tag{430}$$

We note that *translation invariance* of the linear equations of motion for the perturbations means that the different Fourier modes do not interact (see Appendix 7.8.1 for the proof). Different Fourier modes can therefore be studied independently. This often simplifies the differential equations for the perturbations. Next we consider rotations around a single Fourier wavevector $\mathbf{k}$. A perturbation is said to have helicity m if its amplitude is multiplied by $e^{im\psi}$ under rotation of the coordinate system around the wavevector by an angle ψ

$$X_{\mathbf{k}} \to e^{im\psi} X_{\mathbf{k}}\,. \tag{431}$$

identifying points of these two different spacetimes is given. A *gauge choice* is precisely this, *i.e.* a one-to-one correspondence (map) between the background spacetime and the physical spacetime. A change of this map is then a gauge transformation, and the freedom one has in choosing it gives rise to an arbitrariness in the value of the perturbation of X at any given spacetime point, unless X is gauge-invariant.

Scalar, vector and tensor perturbations have helicity 0, ± 1 and ± 2, respectively.[26] The importance of the SVT decomposition is that the perturbations of each type evolve independently (at the linear level) and can therefore be treated separately (see Appendix 7.8.1 for the proof). This considerably simplifies the study of cosmological perturbations.

After these general remarks, let us now become more specific and explicitly define the perturbations around the homogenous and isotropic FRW universe.

The Inhomogeneous Universe

Metric Perturbations

During inflation we define perturbations around the homogeneous background solutions for the inflaton $\bar{\phi}(t)$ and the metric $\bar{g}_{\mu\nu}(t)$,

$$\phi(t, \mathbf{x}) = \bar{\phi}(t) + \delta\phi(t, \mathbf{x}) \,, \qquad g_{\mu\nu}(t, \mathbf{x}) = \bar{g}_{\mu\nu}(t) + \delta g_{\mu\nu}(t, \mathbf{x}) \,, \tag{432}$$

where

$$\begin{aligned} \mathrm{d}s^2 &= g_{\mu\nu}\, \mathrm{d}x^\mu \mathrm{d}x^\nu \\ &= -(1+2\Phi)\mathrm{d}t^2 + 2aB_i \mathrm{d}x^i \mathrm{d}t + a^2[(1-2\Psi)\delta_{ij} + E_{ij}]\mathrm{d}x^i \mathrm{d}x^j \,. \end{aligned} \tag{433}$$

In real space, the SVT decomposition of the metric perturbations (433) is[27]

$$B_i \equiv \partial_i B - S_i \,, \qquad \text{where} \quad \partial^i S_i = 0 \,, \tag{434}$$

and

$$E_{ij} \equiv 2\partial_{ij} E + 2\partial_{(i} F_{j)} + h_{ij} \,, \qquad \text{where} \quad \partial^i F_i = 0 \,, \quad h^i_i = \partial^i h_{ij} = 0 \,. \tag{435}$$

The vector perturbations S_i and F_i aren't created by inflation (and in any case decay with the expansion of the universe). For this reason we ignore vector perturbations in these lectures. Our focus will be on scalar and tensor fluctuations which are observed as density fluctuations and gravitational waves in the late universe.

Tensor fluctuations are gauge-invariant, but scalar fluctuations change under a change of coordinates. Consider the gauge transformation

$$t \;\to\; t + \alpha \tag{436}$$

$$x^i \;\to\; x^i + \delta^{ij}\beta_{,j} \,. \tag{437}$$

Under these coordinate transformations the scalar metric perturbations transform as

$$\Phi \;\to\; \Phi - \dot{\alpha} \tag{438}$$

$$B \;\to\; B + a^{-1}\alpha - a\dot{\beta} \tag{439}$$

$$E \;\to\; E - \beta \tag{440}$$

$$\Psi \;\to\; \Psi + H\alpha \,. \tag{441}$$

Exercise 4 (Linear Gauge Transformations)**.** *Derive the gauge transformations of the scalar metric perturbations (438)–(441). Hint: use invariance of the spacetime interval,*

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = \tilde{g}_{\mu\nu}\mathrm{d}\tilde{x}^\mu \mathrm{d}\tilde{x}^\nu \,. \tag{442}$$

[26] Should this abstract definition of scalar, vector and tensor perturbations in terms of their helicities be confusing, the reader may want to test those rules on the explicit metric and stress-energy perturbations introduced in the next section.

[27] SVT decomposition in real space corresponds to the distinct transformation properties of scalars, vectors and tensors on spatial hypersurfaces.

Matter Perturbations

During inflation the inflationary energy is the dominant contribution to the stress-energy of the universe, so that the inflaton perturbations $\delta\phi$ backreact on the spacetime geometry. This coupling between matter perturbations and metric perturbations is described by the Einstein Equations (see Appendix 7.8.1).

After inflation, the perturbations to the total stress-energy tensor of the universe are

$$\begin{aligned} T^0_0 &= -(\bar\rho + \delta\rho) && (443)\\ T^0_i &= (\bar\rho + \bar p)\, a v_i && (444)\\ T^i_0 &= -(\bar\rho + \bar p)(v^i - B^i)/a && (445)\\ T^i_j &= \delta^i_j(\bar p + \delta p) + \Sigma^i_j\,. && (446) \end{aligned}$$

The anisotropic stress Σ^i_j is gauge-invariant while the density, pressure and momentum density ($(\delta q)_{,i} \equiv (\bar\rho + \bar p)v_i$) transform as follows

$$\begin{aligned} \delta\rho &\to \delta\rho - \dot{\bar\rho}\,\alpha && (447)\\ \delta p &\to \delta p - \dot{\bar p}\,\alpha && (448)\\ \delta q &\to \delta q + (\bar\rho + \bar p)\,\alpha\,. && (449) \end{aligned}$$

Gauge-Invariant Variables

As we explained above, to avoid the pitfall of fictitious gauge modes, it useful to introduce gauge-invariant combinations of metric and matter perturbations [219]. An important gauge-invariant scalar quantity is the *curvature perturbation on uniform-density hypersurfaces* [21]

$$\boxed{-\zeta \equiv \Psi + \frac{H}{\dot{\bar\rho}}\delta\rho}\,. \tag{450}$$

Geometrically, ζ measures the spatial curvature of constant-density hypersurfaces, $R^{(3)} = 4\nabla^2\Psi/a^2$. An important property of ζ is that it remains constant outside the horizon for adiabatic matter perturbations, *i.e.* perturbations that satisfy

$$\boxed{\delta p_{en} \equiv \delta p - \frac{\dot{\bar p}}{\dot{\bar\rho}}\delta\rho} = 0\,. \tag{451}$$

Notice that the definition of δp_{en} is gauge-invariant. In the single-field inflation models studied in this lecture the condition (600) is always satisfy, so the perturbation $\zeta_{\mathbf{k}}$ doesn't evolve outside the horizon, $k \ll aH$.

In a gauge defined by spatially flat hypersurfaces, Ψ, the perturbations ζ is the dimensionless density perturbation $\frac{1}{3}\delta\rho/(\bar\rho+\bar p)$. Taking into account appropriate transfer functions to describe the sub-horizon evolution of the fluctuations, CMB and LSS observations can therefore be related to the primordial value of ζ (see **Lecture 3**). During slow-roll inflation

$$-\zeta \approx \Psi + \frac{H}{\dot{\bar\phi}}\delta\phi\,. \tag{452}$$

Another gauge-invariant scalar is the *comoving curvature perturbation*

$$\boxed{\mathcal{R} \equiv \Psi - \frac{H}{\bar\rho + \bar p}\delta q}\,, \tag{453}$$

where δq is the scalar part of the 3-momentum density $T^0_i = \partial_i \delta q$. During inflation $T^0_i = -\dot{\bar{\phi}}\,\partial_i \delta\phi$ and hence

$$\mathcal{R} = \Psi + \frac{H}{\dot{\bar{\phi}}}\delta\phi\,. \tag{454}$$

Geometrically, $\mathcal{R}$ measures the spatial curvature of comoving (or constant-ϕ) hypersurfaces.

The linearized Einstein equations relate ζ and $\mathcal{R}$ as follows (see Appendix 7.8.1)

$$-\zeta = \mathcal{R} + \frac{k^2}{(aH)^2}\frac{2\bar{\rho}}{3(\bar{\rho}+\bar{p})}\Psi_{\rm B}\,, \tag{455}$$

where

$$\Psi_{\rm B} \equiv \psi + a^2 H(\dot{E} - B/a)\,, \tag{456}$$

is one of the Bardeen potentials [219]. ζ and $\mathcal{R}$ are therefore equal on superhorizon scales, $k \ll aH$. ζ and $\mathcal{R}$ are also equal during slow-roll inflation, *cf.* Eqs. (452) and (454). The correlation functions of ζ and $\mathcal{R}$ are therefore equal at horizon crossing and both ζ and $\mathcal{R}$ are conserved on superhorizon scales. In this lecture we will compute the primordial spectrum of $\mathcal{R}$ at horizon crossing.

Finally, a gauge-invariant measure of inflaton perturbations is the *inflaton perturbation on spatially flat slices*

$$\boxed{Q \equiv \delta\phi + \frac{\dot{\bar{\phi}}}{H}\Psi}\,. \tag{457}$$

Exercise 5 (Gauge-Invariant Perturbations)**.** *Using the linear gauge transformations for the metric and matter perturbations, confirm that ζ, $\mathcal{R}$ and Q are gauge-invariant.*

Superhorizon (Non-)Evolution

The Einstein equations (see Appendix 7.8.1) give the evolution equation for the gauge-invariant curvature perturbation

$$\dot{\mathcal{R}} = -\frac{H}{\bar{\rho}+\bar{p}}\,\delta p_{en} + \frac{k^2}{(aH)^2}\Big(\dots\Big)\,. \tag{458}$$

Adiabatic matter perturbations satisfy $\delta p_{en} = 0$ and $\mathcal{R}$ is conserved on superhorizon scales, $k < aH$.

Exercise 6 (Separate Universe Approach)**.** *Read about the separate universe approach [22] for proving conservation of the curvature perturbation $\mathcal{R}$ on superhorizon scales.*

Statistics of Cosmological Perturbations

A crucial statistical measure of the primordial scalar fluctuations is the power spectrum of $\mathcal{R}$ (or ζ)[28]

$$\langle \mathcal{R}_{\mathbf{k}}\mathcal{R}_{\mathbf{k}'}\rangle = (2\pi)^3\,\delta(\mathbf{k}+\mathbf{k}')\,P_{\mathcal{R}}(k)\,, \qquad \Delta^2_{\rm s} \equiv \Delta^2_{\mathcal{R}} = \frac{k^3}{2\pi^2}P_{\mathcal{R}}(k)\,. \tag{459}$$

Here, $\langle\,\dots\,\rangle$ defines the ensemble average of the fluctuations. The scale-dependence of the power spectrum is defined by the scalar spectral index (or tilt)

$$n_{\rm s} - 1 \equiv \frac{d\ln\Delta^2_{\rm s}}{d\ln k}\,, \tag{460}$$

[28] The normalization of the dimensionless power spectrum $\Delta^2_{\mathcal{R}}(k)$ is chosen such that the variance of $\mathcal{R}$ is $\langle\mathcal{R}\mathcal{R}\rangle = \int_0^\infty \Delta^2_{\mathcal{R}}(k)\,{\rm d}\ln k$.

where scale-invariance corresponds to the value $n_{\rm s} = 1$. We may also define the running of the spectral index by

$$\alpha_{\rm s} \equiv \frac{dn_{\rm s}}{d\ln k}\,. \tag{461}$$

The power spectrum is often approximated by a power law form

$$\Delta^2_{\rm s}(k) = A_{\rm s}(k_\star)\left(\frac{k}{k_\star}\right)^{n_{\rm s}(k_\star)-1+\frac{1}{2}\alpha_{\rm s}(k_\star)\ln(k/k_\star)}\,, \tag{462}$$

where $k_\star$ is an arbitrary reference or pivot scale.

If $\mathcal{R}$ is Gaussian then the power spectrum contains all the statistical information. Primordial non-Gaussianity is encoded in higher-order correlation functions of $\mathcal{R}$. In single-field slow-roll inflation the non-Gaussianity is predicted to be small [23, 22], but non-Gaussianity can be significant in multi-field models or in single-field models with non-trivial kinetic terms and/or violation of the slow-roll conditions. We will return to primordial non-Gaussianity in **Lecture 4**. In this lecture we restrict our computation to Gaussian fluctuations and the associated power spectra.

The power spectrum for the two polarization modes of h_{ij}, *i.e.* $h \equiv h^+, h^\times$, is defined as

$$\langle h_{\mathbf{k}} h_{\mathbf{k}'}\rangle = (2\pi)^3\,\delta(\mathbf{k}+\mathbf{k}')\,P_h(k)\,, \qquad \Delta^2_h = \frac{k^3}{2\pi^2}P_h(k)\,. \tag{463}$$

We define the power spectrum of tensor perturbations as the sum of the power spectra for the two polarizations

$$\Delta^2_{\rm t} \equiv 2\Delta^2_h\,. \tag{464}$$

Its scale-dependence is defined analogously to Eqn. (460) but for historical reasons without the -1,

$$n_{\rm t} \equiv \frac{d\ln\Delta^2_{\rm t}}{d\ln k}\,, \tag{465}$$

i.e.

$$\Delta^2_{\rm t}(k) = A_{\rm t}(k_\star)\left(\frac{k}{k_\star}\right)^{n_{\rm t}(k_\star)}\,. \tag{466}$$

Aim of this Lecture

It will be the aim of this lecture to compute the power spectra of scalar and tensor fluctuations, $P_{\mathcal{R}}(k)$ and $P_h(k)$, from first principles. This is one of the most important calculations in modern theoretical cosmology, so to understand it will be well worth our efforts.

Uniform Density Gauge

Comoving Gauge

Spatially-Flat Gauge

7.4.2 *Preview*: The Quantum Origin of Structure

In the last lecture we discussed the classical evolution of the inflaton field. Something remarkable happens when one considers quantum fluctuations of the inflaton: inflation combined with quantum mechanics provides an elegant mechanism for generating the initial seeds of all structure in the universe. In other words, quantum fluctuations during inflation are the source of the primordial power spectra of scalar and tensor fluctuations, $P_{\rm s}(k)$ and $P_{\rm t}(k)$. In this section we sketch the mechanism by which inflation relates microscopic physics to macroscopic observables. In §7.4.4 we present the full calculation.

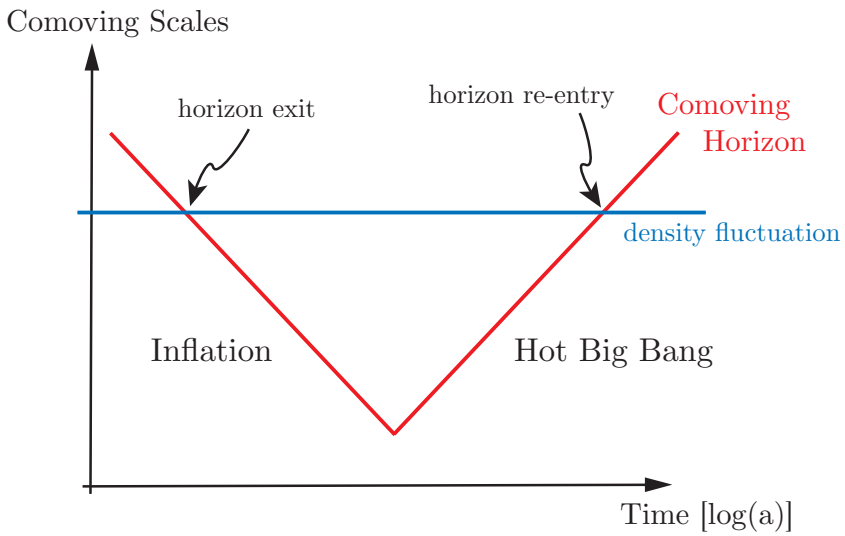


Figure 68: Creation and evolution of perturbations in the inflationary universe. Fluctuations are created quantum mechanically on subhorizon scales. While comoving scales, k^{-1}, remain constant the comoving Hubble radius during inflation, $(aH)^{-1}$, shrinks and the perturbations exit the horizon. Causal physics cannot act on superhorizon perturbations and they freeze until horizon re-entry at late times.

Quantum Zero-Point Fluctuations

As we will explain quantitatively in §7.4.4 quantum fluctuations during inflation induce a non-zero variance for fluctuations in all light fields (like the inflaton or the metric perturbations). This is very similar to the variance in the amplitude of a harmonic oscillator induced by zero-point fluctuations in the ground state; see §7.4.3. The amplitude of fluctuations scales with the expansion parameter H during inflation. This relates to the de Sitter horizon, H^{-1}, and the quantum fluctuations during inflation may also be interpreted as thermal fluctuations in de Sitter space in close analogy to the Hawking radiation for black holes.

Fluctuations are created on all length scales, *i.e.* with a spectrum of wavenumbers k. Cosmologically relevant fluctuations start their lives inside the horizon (Hubble radius),

$$\text{subhorizon}: \quad k \gg aH\,. \tag{467}$$

However, while the comoving wavenumber is constant the comoving Hubble radius shrinks during inflation (recall this is how we 'defined' inflation!), so eventually all fluctuations *exit the horizon*

$$\text{superhorizon}: \quad k < aH\,. \tag{468}$$

Horizon Exit and Re-Entry

Cosmological inhomogeneity is characterized by the intrinsic curvature of spatial hypersurfaces defined with respect to the matter, $\mathcal{R}$ or ζ. Both $\mathcal{R}$ and ζ have the attractive feature that they remain constant outside the horizon, *i.e.* when $k < aH$. In particular, their amplitude is *not* affected by the unknown physical properties of the universe shortly after inflation (recall that we know next to nothing about the details of reheating; it is the constancy of $\mathcal{R}$ and ζ outside the horizon that allows us to nevertheless predict cosmological observables). After inflation, the comoving horizon grows, so eventually all fluctuations will re-enter the horizon. After horizon re-entry, $\mathcal{R}$ or ζ determine the perturbations of the cosmic fluid resulting in the observed CMB anisotropies and the LSS.

In **Lecture 1** we explained the evolution of the comoving horizon during inflation and in the standard FRW expansion after inflation. In this lecture (**Lecture 2**) we will compute the primordial power spectrum of comoving curvature fluctuations $\mathcal{R}$ at horizon exit. In the next lecture (**Lecture 3**) we will compute the relation of curvature fluctuations $\mathcal{R}$ to fluctuations in cosmological observables after horizon re-entry. Together these three lectures therefore provide a complete account of both the generation and the observational consequences of the quantum fluctuations produced by inflation. It is a beautiful story. Let us begin to unfold it.

7.4.3 Quantum Mechanics of the Harmonic Oscillator

> "The career of a young theoretical physicist consists of treating the harmonic oscillator in ever-increasing levels of abstraction."
>
> *Sidney Coleman*

The computation of quantum fluctuations generated during inflation is algebraically quite intensive and it is therefore instructive to start with a simpler example which nevertheless contains most of the relevant physics. We therefore warm up by considering the quantization of a one-dimensional simple harmonic oscillator. Harmonic oscillators are one of the few physical systems that physicists know how to solve exactly. Fortunately, almost all more complicated physical systems can be represented by a collection of simple harmonic oscillators with different amplitudes and frequencies. This is of course what Fourier analysis is all about. We will show below that free fields in curved spacetime (and de Sitter space in particular) are similar to collections of harmonic oscillators with time-dependent frequencies. The detailed treatment of the quantum harmonic oscillator in this section will therefore not be in vain, but will provide important intuition for the inflationary calculation. This section is based on the excellent treatment of [26].

Action

The classical action of a harmonic oscillator with time-dependent frequency is

$$S = \int \mathrm{d}t \left(\frac{1}{2}\dot{x}^2 - \frac{1}{2}\omega^2(t)x^2 \right) \equiv \int \mathrm{d}t\, L\,, \tag{469}$$

where x is the deviation of the particle from its equilibrium state, $x \equiv 0$, and for convenience we have set the particle mass to one, $m \equiv 1$. For concreteness one may wish to consider a particle of mass m on a spring which is heated by an external source so that its spring constant depends on time, $k(t)$, where $\omega^2 = k/m$. The classical equation of motion follows from variation of the action with respect to the particle coordinate x

$$\frac{\delta S}{\delta x} = 0 \quad \Rightarrow \quad \boxed{\ddot{x} + \omega^2(t)\, x = 0}\,. \tag{470}$$

Canonical Quantization

Canonical quantization of the system proceeds in the standard way: We define the momentum conjugate to x

$$p \equiv \frac{dL}{d\dot{x}} = \dot{x}, \tag{471}$$

which agrees with the standard notion of the particle's momentum $p = mv$. We then promote the classical variables x, p to quantum operators $\hat{x}$, $\hat{p}$ and impose the canonical commutator

$$\boxed{[\hat{x}, \hat{p}] = i\hbar}\,, \tag{472}$$

where $[\hat{x}, \hat{p}] \equiv \hat{x}\hat{p} - \hat{p}\hat{x}$. The equation of motion implies that the commutator holds at all times if imposed at some initial time. In particular, for our present example

$$[x(t), \dot{x}(t)] = i\hbar\,. \tag{473}$$

Note that we are in the Heisenberg picture where operators vary in time while states are time-independent. The operator $\hat{x}$ is then expanded in terms of creation and annihilation operators

$$\boxed{\hat{x} = v(t)\,\hat{a} + v^*(t)\,\hat{a}^\dagger}\,, \tag{474}$$

where the (complex) mode function satisfies the classical equation of motion

$$\ddot{v} + \omega^2(t)v = 0\,. \tag{475}$$

The commutator (472) becomes

$$\langle v, v\rangle[\hat{a}, \hat{a}^\dagger] = 1\,, \tag{476}$$

where

$$\langle v, w\rangle \equiv \frac{i}{\hbar}\left(v^*\partial_t w - (\partial_t v^*)w\right)\,. \tag{477}$$

Without loss of generality, let us assume that the solution v is chosen so that the real number $\langle v, v\rangle$ is positive. The function v can then be rescaled such that $\langle v, v\rangle \equiv 1$ and hence

$$\boxed{[\hat{a}, \hat{a}^\dagger] = 1}\,. \tag{478}$$

Eqn. (478) is the standard relation for the raising and lowering operators of a harmonic oscillator. We have hence identified the following annihilation and creation operators

$$\hat{a} = \langle v, \hat{x}\rangle \tag{479}$$

$$\hat{a}^\dagger = -\langle v^*, \hat{x}\rangle\,, \tag{480}$$

and can define the vacuum state $|0\rangle$ via the prescription

$$\boxed{\hat{a}|0\rangle = 0}\,, \tag{481}$$

i.e. the vacuum is annihilated by $\hat{a}$. Excited states of the system are created by repeated application of creation operators

$$|n\rangle \equiv \frac{1}{\sqrt{n!}}(\hat{a}^\dagger)^n|0\rangle\,. \tag{482}$$

These states are eigenstates of the number operator $\hat{N} = \hat{a}^\dagger\hat{a}$ with eigenvalue n, *i.e.*

$$\hat{N}|n\rangle = n|n\rangle\,. \tag{483}$$

Non-Uniqueness of the Mode Functions

We haven't yet determined unique mode functions and hence we haven't fixed the vacuum state. Any change in $v(t)$ that keeps the solution $x(t)$ unchanged will lead to a change in the creating operator $\hat{a} = \langle v, \hat{x}\rangle$ and hence a change in the definition of the vacuum. For the simple harmonic oscillator with time-dependent frequency $\omega(t)$ (and for quantum fields in curved spacetime) there is in fact *no* unique choice for the mode function $v(t)$. Hence, there is *no* unique decomposition of $\hat{x}$ into annihilation and creation operators and *no* unique notion of the vacuum. Different choices for the solution $v(t)$ give different vacuum solutions. This problem

and its standard (but not uncontested) resolution in the case of inflation will be discussed in more detail below.

In the present case we can make progress by considering the special case of a constant-frequency harmonic oscillator[29] $\omega(t) = \omega$. In that case a preferred choice of $v(t)$ is the one that makes the vacuum state $|0\rangle$ the ground state of the Hamiltonian. First, we evaluate the Hamiltonian for a general mode function $v(t)$,

$$\begin{aligned} \hat{H} &= \frac{1}{2}\hat{p}^2 + \frac{1}{2}\omega^2\hat{x}^2 \\ &= \frac{1}{2}\left[(\dot{v}^2 + \omega^2 v^2)\hat{a}\hat{a} + (\dot{v}^2 + \omega^2 v^2)^*\,\hat{a}^\dagger\hat{a}^\dagger + (|\dot{v}|^2 + \omega^2|v|^2)(\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a})\right] . \end{aligned} \tag{484}$$

Using $\hat{a}|0\rangle = 0$ and $[\hat{a}, \hat{a}^\dagger] = 1$, we hence find the following action of the Hamiltonian operator on the vacuum state

$$\hat{H}|0\rangle = \frac{1}{2}(\dot{v}^2 + \omega^2 v^2)^*\,\hat{a}^\dagger\hat{a}^\dagger|0\rangle + \frac{1}{2}(|\dot{v}|^2 + \omega^2|v|^2)|0\rangle\,. \tag{485}$$

The requirement that $|0\rangle$ be an eigenstate of $\hat{H}$ means that the first term must vanish which implies the condition

$$\dot{v} = \pm i\omega v\,, \tag{486}$$

and hence

$$\langle v, v\rangle = \mp\frac{2\omega}{\hbar}|v|^2\,. \tag{487}$$

Positivity of the normalization condition $\langle v, v\rangle > 0$ selects the minus sign in Eqn. (486)

$$\dot{v} = -i\omega v\,. \tag{488}$$

Properly normalized ($\langle v, v\rangle = 1$) this gives the following *positive-frequency solution*

$$\boxed{v(t) = \sqrt{\frac{\hbar}{2\omega}}\,e^{-i\omega t}}\,. \tag{489}$$

With this choice of mode function v the Hamiltonian is

$$\hat{H} = \hbar\omega\left(\hat{N} + \frac{1}{2}\right)\,, \tag{490}$$

for which the vacuum $|0\rangle$ is the state of minimum energy $\hbar\omega/2$.

Exercise 7 (Non-Uniqueness for Time-Dependent Oscillators)**.** *What goes wrong with the above argument for the case of a simple harmonic oscillator with time-dependent frequency?*

Zero-Point Fluctuations in the Ground State

Consider the mean square expectation value of the position operator $\hat{x}$ in the ground state $|0\rangle$

$$\begin{aligned} \langle|\hat{x}|^2\rangle &\equiv \langle 0|\hat{x}^\dagger\hat{x}|0\rangle \\ &= \langle 0|(v^*\hat{a}^\dagger + v\hat{a})(v\hat{a} + v^*\hat{a}^\dagger)|0\rangle \\ &= |v(\omega,t)|^2\langle 0|\hat{a}\hat{a}^\dagger|0\rangle \\ &= |v(\omega,t)|^2\langle 0|[\hat{a}, \hat{a}^\dagger]|0\rangle \\ &= |v(\omega,t)|^2\,. \end{aligned} \tag{491}$$

[29] It will turn out that this is the relevant case for inflation at very early times when all modes are deep inside the horizon.

This characterizes the zero-point fluctuations of the position in the vacuum state as the square of the mode function

$$\boxed{\langle|\hat{x}|^2\rangle = |v(\omega,t)|^2} = \frac{\hbar}{2\omega}\,. \tag{492}$$

This is all we need to know about quantum mechanics to compute the fluctuation spectrum created by inflation. However, first we need to do quite a bit of work to derive the mode equation for the scalar mode of cosmological perturbations, *i.e.* the analogue of Eqn. (475).

7.4.4 Quantum Fluctuations in de Sitter Space

We have finally come to the highlight of this lecture: the full computation of the quantum-mechanical fluctuations generated during inflation and their relation to cosmological perturbations. Our calculation follows closely the treatment by Maldacena [22].

Summary of the Computational Strategy

The last two sections might have bored you, but they provided important background for the computation of inflationary fluctuations. We have defined the gauge-invariant curvature perturbation $\mathcal{R}$. It is conserved outside of the horizon, so we can compute it at horizon exit and remain ignorant about the subhorizon physics during and after reheating until horizon re-entry of a given $\mathcal{R}$-mode. We have recalled the quantization of the simple harmonic oscillator, so by writing the equation of motion for $\mathcal{R}$ in simple harmonic oscillator form we are in the position to study the quantization of scalar fluctuations during inflation.

Here is a summary of the steps we will perform in the following sections:

1. We expand the action for single-field slow-roll inflation to second order in fluctuations. Specially, we derive the second-order expansion of the action in terms of $\mathcal{R}$. The action approach guarantees the correct normalization for the quantization of fluctuations.

2. From the action we derive the equation of motion for $\mathcal{R}$ and show that it is of SHO form.

3. The mode equations for $\mathcal{R}$ will be hard to solve exactly so we consider several approximate solutions valid during slow-roll evolution.

4. We promote the classical field $\mathcal{R}$ to a quantum operator and quantize it. Imposing the canonical commutation relation for quantum operators will lead to a boundary condition on the mode functions. This doesn't fix the mode function completely.

5. We define the vacuum state by matching our solutions to the Minkowski vacuum in the ultraviolet, *i.e.* on small scales when the mode is deep inside the horizon. This fixes the mode functions completely and their large-scale limit is hence determined.

6. We then compute the power spectrum of curvature fluctuations at horizon crossing. In **Lecture 3** we will relate the power spectrum at horizon crossing during inflation to the angular power spectrum of CMB fluctuations at recombination.

Enough talking, let's compute!

Scalar Perturbations

We consider single-field slow-roll models of inflation defined by the action

$$S = \frac{1}{2}\int \mathrm{d}^4x\sqrt{-g}\,\left[R - (\nabla\phi)^2 - 2V(\phi)\right]\,, \tag{493}$$

in units where $M_{\rm pl}^{-2} \equiv 8\pi G = 1$. To fix time and spatial reparameterizations we choose the following gauge for the dynamical fields g_{ij} and ϕ

$$\delta\phi = 0\,, \qquad g_{ij} = a^2[(1-2\mathcal{R})\delta_{ij} + h_{ij}]\,, \qquad \partial_i h_{ij} = h^i_i = 0\,. \tag{494}$$

In this gauge the inflaton field is unperturbed and all scalar degrees of freedom are parameterized by the metric fluctuation $\mathcal{R}(t,\mathbf{x})$. An important property of $\mathcal{R}$ is that it remains constant outside the horizon. We can therefore restrict our computation to correlation functions of $\mathcal{R}$ at horizon crossing. The remaining metric perturbations Φ and B are related to $\mathcal{R}$ by the Einstein Equations; in the ADM formalism (see Appendix 7.8.2) these are pure constraint equations.

Free Field Action

With quite some effort (see Appendix 7.8.2) one may expand the action (493) to second order in $\mathcal{R}$

$$\boxed{S_{(2)} = \frac{1}{2}\int \mathrm{d}^4x\, a^3\frac{\dot\phi^2}{H^2}\left[\dot{\mathcal{R}}^2 - a^{-2}(\partial_i\mathcal{R})^2\right]}\,. \tag{495}$$

Defining the *Mukhanov variable*

$$v \equiv z\mathcal{R}\,, \qquad \text{where} \qquad z^2 \equiv a^2\frac{\dot\phi^2}{H^2} = 2a^2\varepsilon\,, \tag{496}$$

and transitioning to conformal time τ leads to the action for a canonically normalized scalar

$$\boxed{S_{(2)} = \frac{1}{2}\int \mathrm{d}\tau\mathrm{d}^3x\,\left[(v')^2 + (\partial_i v)^2 + \frac{z''}{z}v^2\right]}\,, \qquad (...)' \equiv \partial_\tau(...)\,. \tag{497}$$

Exercise 8 (Mukhanov Action)**.** *Confirm Eqn. (497). Hint: use integration by parts.*

We define the Fourier expansion of the field v

$$v(\tau,\mathbf{x}) = \int\frac{\mathrm{d}^3k}{(2\pi)^3}\,v_{\mathbf{k}}(\tau)e^{i\mathbf{k}\cdot\mathbf{x}}\,, \tag{498}$$

where

$$\boxed{v_k'' + \left(k^2 - \frac{z''}{z}\right)v_k = 0}\,. \tag{499}$$

Here, we have dropped to vector notation $\mathbf{k}$ on the subscript, since (499) depends only on the magnitude of k. The Mukhanov Equation (499) is hard to solve in full generality since the function z depends on the background dynamics. For a given inflationary background one may solve (499) numerically. However, to gain a more intuitive understanding of the solutions we will discuss approximate analytical solutions in the pure de Sitter limit (§7.4.4) and in the slow-roll approximation (**Problem 7**).

Quantization

The quantization of the field v is performed in completely analogy with our treatment of the quantum harmonic oscillator in §7.4.3.

As before we promote the field v and its conjugate momentum v' to quantum operator

$$v \quad \rightarrow \quad \hat{v} = \int \frac{\mathrm{d}\mathbf{k}^3}{(2\pi)^3} \left[v_k(\tau)\hat{a}_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} + v_k^*(\tau)\hat{a}_{\mathbf{k}}^\dagger e^{-i\mathbf{k}\cdot\mathbf{x}} \right] . \tag{500}$$

Alternatively, the Fourier components $v_{\mathbf{k}}$ are promoted to operators and expressed via the following decomposition

$$v_{\mathbf{k}} \quad \rightarrow \quad \hat{v}_{\mathbf{k}} = v_k(\tau)\hat{a}_{\mathbf{k}} + v_{-k}^*(\tau)\hat{a}_{-\mathbf{k}}^\dagger \,, \tag{501}$$

where the creation and annihilation operators $\hat{a}_{-\mathbf{k}}^\dagger$ and $\hat{a}_{\mathbf{k}}$ satisfy the canonical commutation relation

$$[\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}^\dagger] = (2\pi)^3 \delta(\mathbf{k} - \mathbf{k}') \,, \tag{502}$$

if and only if the mode functions are normalized as follows

$$\langle v_k, v_k \rangle \equiv \frac{i}{\hbar}(v_k^* v_k' - v_k^{*\prime} v_k) = 1 \,. \tag{503}$$

Eqn. (503) provides one of the boundary conditions on the solutions of Eqn. (499). The second boundary conditions that fixes the mode functions completely comes from *vacuum selection*.

Boundary Conditions and Bunch-Davies Vacuum

We must choose a vacuum state for the fluctuations,

$$\hat{a}_{\mathbf{k}}|0\rangle = 0 \,, \tag{504}$$

which corresponds to specifying an additional boundary conditions for v_k (see *e.g.* Chapter 3 in Birell and Davies [25]). The standard choice is the Minkowski vacuum of a comoving observer in the far past (when all comoving scales were far inside the Hubble horizon), $\tau \rightarrow -\infty$ or $|k\tau| \gg 1$ or $k \gg aH$. In this limit the mode equation (499) becomes

$$v_k'' + k^2 v_k = 0 \,. \tag{505}$$

This is the equation of a simple harmonic oscillator with time-independent frequency! For this case we showed that a unique solution (489) exists if we require the vacuum to be the minimum energy state. Hence we impose the initial condition

$$\boxed{\lim_{\tau \rightarrow -\infty} v_k = \frac{e^{-ik\tau}}{\sqrt{2k}}} \,. \tag{506}$$

The boundary conditions (503) and (506) completely fix the mode functions on *all* scales.

Solution in de Sitter Space

Consider the de Sitter limit $\varepsilon \rightarrow 0$ ($H = \text{const.}$) and

$$\frac{z''}{z} = \frac{a''}{a} = \frac{2}{\tau^2} \,. \tag{507}$$

In a de Sitter background we therefore wish to solve the mode equation

$$\boxed{v_k'' + \left(k^2 - \frac{2}{\tau^2} \right) v_k = 0} \,. \tag{508}$$

Exercise 9 (de Sitter Mode Functions)**.** *Verify by direct substitution that an exact solution to Eqn. (508) is*

$$v_k = \alpha \frac{e^{-ik\tau}}{\sqrt{2k}}\left(1 - \frac{i}{k\tau}\right) + \beta \frac{e^{ik\tau}}{\sqrt{2k}}\left(1 + \frac{i}{k\tau}\right). \tag{509}$$

The free parameters α and β characterize the non-uniqueness of the mode functions. However, we may fix α and β to unique values by considering the quantization condition (503) together with the subhorizon limit, $|k\tau| \gg 1$, Eqn. (506). This fixes $\alpha = 1$, $\beta = 0$ and leads to the unique Bunch-Davies mode functions

$$\boxed{v_k = \frac{e^{-ik\tau}}{\sqrt{2k}}\left(1 - \frac{i}{k\tau}\right)}. \tag{510}$$

Power Spectrum in Quasi-de Sitter

We then compute the power spectrum of the field $\hat{\psi}_{\mathbf{k}} \equiv a^{-1}\hat{v}_{\mathbf{k}}$,

$$\langle \hat{\psi}_{\mathbf{k}}(\tau)\hat{\psi}_{\mathbf{k}'}(\tau)\rangle = (2\pi)^3\delta(\mathbf{k}+\mathbf{k}')\frac{|v_k(\tau)|^2}{a^2} \tag{511}$$

$$= (2\pi)^3\delta(\mathbf{k}+\mathbf{k}')\frac{H^2}{2k^3}(1+k^2\tau^2)\,. \tag{512}$$

On superhorizon scales, $|k\tau| \ll 1$, this approaches a constant

$$\langle \hat{\psi}_{\mathbf{k}}(\tau)\hat{\psi}_{\mathbf{k}'}(\tau)\rangle \to (2\pi)^3\delta(\mathbf{k}+\mathbf{k}')\frac{H^2}{2k^3}\,. \tag{513}$$

or

$$\Delta^2_\psi = \left(\frac{H}{2\pi}\right)^2. \tag{514}$$

The de Sitter result for $\psi = v/a$, Eqn. (513), allows us to compute the power spectrum of $\mathcal{R} = \frac{H}{\dot{\phi}}\psi$ at horizon crossing, $a(t_\star)H(t_\star) = k$,

$$\langle \mathcal{R}_{\mathbf{k}}(t)\mathcal{R}_{\mathbf{k}'}(t)\rangle = (2\pi)^3\delta(\mathbf{k}+\mathbf{k}')\frac{H_\star^2}{2k^3}\frac{H_\star^2}{\dot{\phi}_\star^2}\,. \tag{515}$$

Here, $(...)_\star$ indicates that a quantity is to be evaluated at horizon crossing. We define the dimensionless power spectrum $\Delta^2_{\mathcal{R}}(k)$ by

$$\langle \mathcal{R}_{\mathbf{k}}\mathcal{R}_{\mathbf{k}'}\rangle = (2\pi)^3\delta(\mathbf{k}+\mathbf{k}')P_{\mathcal{R}}(k)\,, \qquad \Delta^2_{\mathcal{R}}(k) \equiv \frac{k^3}{2\pi^2}P_{\mathcal{R}}(k)\,, \tag{516}$$

such that the real space variance of $\mathcal{R}$ is $\langle \mathcal{R}\mathcal{R}\rangle = \int_0^\infty \Delta^2_{\mathcal{R}}(k)\,\mathrm{d}\ln k$. This gives

$$\boxed{\Delta^2_{\mathcal{R}}(k) = \frac{H_\star^2}{(2\pi)^2}\frac{H_\star^2}{\dot{\phi}_\star^2}}. \tag{517}$$

Since $\mathcal{R}$ approaches a constant on super-horizon scales the spectrum at horizon crossing determines the future spectrum until a given fluctuation mode re-enters the horizon.

The fact that we computed the power spectrum at a specific instant (horizon crossing, $a_\star H_\star = k$) implicitly extends the result for the pure de Sitter background to a slowly time-evolving quasi-de Sitter space. Different modes exit the horizon as slightly different times when $a_\star H_\star$ has a different value. This procedure gives the correct result for the power spectrum during slow-roll inflation (we prove this more rigorously in **Problem 7**.). For non-slow-roll inflation the background evolution will have to be tracked more precisely and the Mukhanov Equation typically has to be integrated numerically.

Spatially-Flat Gauge

In the previous sections we followed Maldacena and used the comoving gauge ($\delta\phi = 0$) to compute the scalar power spectrum. A popular alternative to obtain the same result is to use *spatially-flat gauge*. In spatially-flat gauge, perturbations in $\mathcal{R}$ are related to perturbations in the inflaton field value[30] $\delta\phi$, *cf.* Eqn. (454) with $\Psi = 0$

$$\mathcal{R} = H\frac{\delta\phi}{\dot{\phi}} \equiv -H\delta t\,. \tag{518}$$

The power spectrum of $\mathcal{R}$ and the power spectrum of inflaton fluctuations $\delta\phi$ are therefore related as follows

$$\langle \mathcal{R}_{\mathbf{k}}\mathcal{R}_{\mathbf{k}'}\rangle = \left(\frac{H}{\dot{\phi}}\right)^2 \langle \delta\phi_{\mathbf{k}}\,\delta\phi_{\mathbf{k}'}\rangle\,. \tag{519}$$

Finally, in the case of slow-roll inflation, quantum fluctuations of a light scalar field ($m_\phi \ll H$) in quasi-de Sitter space ($H \approx$ const.) scale with the Hubble parameter H, *cf.* Eqn. (514),

$$\langle \delta\phi_{\mathbf{k}}\,\delta\phi_{\mathbf{k}'}\rangle = (2\pi)^3\,\delta(\mathbf{k}+\mathbf{k}')\,\frac{2\pi^2}{k^3}\left(\frac{H}{2\pi}\right)^2, \qquad \Delta^2_{\delta\phi} = \left(\frac{H}{2\pi}\right)^2. \tag{520}$$

Inflationary quantum fluctuations therefore produce the following power spectrum for $\mathcal{R}$

$$\Delta^2_{\mathcal{R}}(k) = \frac{H^2_\star}{(2\pi)^2}\frac{H^2_\star}{\dot{\phi}^2_\star}\,. \tag{521}$$

This is consistent with our result (517).

Tensor Perturbations

Having discussed the quantization of scalar perturbation is some details, the corresponding calculation for tensor perturbations will appear almost trivial.

Action

By expansion of the Einstein-Hilbert action one may obtain the second-order action for tensor fluctuations is

$$S_{(2)} = \frac{M^2_{\rm pl}}{8}\int \mathrm{d}\tau\mathrm{d}x^3 a^2\left[(h'_{ij})^2 - (\partial_l h_{ij})^2\right]\,. \tag{522}$$

Here, we have reintroduced explicit factors of $M_{\rm pl}$ to make h_{ij} manifestly dimensionless. Up to a normalization factor of $\frac{M_{\rm pl}}{2}$ this is the same as the action for a massless scalar field in an FRW universe.

We define the following Fourier expansion

$$h_{ij} = \int \frac{\mathrm{d}^3k}{(2\pi)^3}\sum_{s=+,\times}\epsilon^s_{ij}(k)h^s_{\mathbf{k}}(\tau)e^{i\mathbf{k}\cdot\mathbf{x}}\,, \tag{523}$$

where $\epsilon_{ii} = k^i\epsilon_{ij} = 0$ and $\epsilon^s_{ij}(k)\epsilon^{s'}_{ij}(k) = 2\delta_{ss'}$. The tensor action (522) becomes

$$S_{(2)} = \sum_s \int \mathrm{d}\tau\mathrm{d}\mathbf{k}\frac{a^2}{4}M^2_{\rm pl}\left[h^{s\prime}_{\mathbf{k}}h^{s\prime}_{\mathbf{k}} - k^2 h^s_{\mathbf{k}}h^s_{\mathbf{k}}\right]\,. \tag{524}$$

[30] Intuitively, the curvature perturbation $\mathcal{R}$ is related to a spatially varying time-delay $\delta t(\mathbf{x})$ for the end of inflation. This time-delay is induced by the inflaton fluctuation $\delta\phi$.

We define the canonically normalized field

$$v_{\mathbf{k}}^s \equiv \frac{a}{2} M_{\rm pl} h_{\mathbf{k}}^s \,, \tag{525}$$

to get

$$\boxed{S_{(2)} = \sum_s \frac{1}{2} \int {\rm d}\tau {\rm d}^3\mathbf{k} \left[(v_{\mathbf{k}}^{s\prime})^2 - \left(k^2 - \frac{a''}{a} \right) (v_{\mathbf{k}}^s)^2 \right]}, \tag{526}$$

where

$$\frac{a''}{a} = \frac{2}{\tau^2} \tag{527}$$

holds in de Sitter space. This should be recognized as effectively two copies of the action (497).

Quantization

Each polarization of the gravitational wave is therefore just a renormalized massless field in de Sitter space

$$h_{\mathbf{k}}^s = \frac{2}{M_{\rm pl}} \psi_{\mathbf{k}}^s \,, \qquad \psi_{\mathbf{k}}^s \equiv \frac{v_{\mathbf{k}}}{a} \,. \tag{528}$$

Since we computed the power spectrum of $\psi = v/a$ in the previous section, $\Delta_\psi^2 = (H/2\pi)^2 {\rm m}$ we can simply right down the answer for Δ_h^2, the power spectrum for a single polarization of tensor perturbations,

$$\Delta_h^2(k) = \frac{4}{M_{\rm pl}^2} \left(\frac{H_\star}{2\pi} \right)^2 \,. \tag{529}$$

Again, the r.h.s. is to be evaluated at horizon exit. .

Power Spectrum

The dimensionless power spectrum of tensor fluctuations therefore is

$$\boxed{\Delta_{\rm t}^2 = 2\Delta_h^2(k) = \frac{2}{\pi^2} \frac{H_\star^2}{M_{\rm pl}^2}} \,. \tag{530}$$

The Tensor-to-Scalar Ratio

The Energy Scale of Inflation

Tensor fluctuations are often normalized relative to the (measured) amplitude of scalar fluctuations, $\Delta_{\rm s}^2 \equiv \Delta_{\mathcal{R}}^2 \sim 10^{-9}$. The *tensor-to-scalar ratio* is

$$r \equiv \frac{\Delta_{\rm t}^2(k)}{\Delta_{\rm s}^2(k)} \,. \tag{531}$$

Since $\Delta_{\rm s}^2$ is fixed and $\Delta_{\rm t}^2 \propto H^2 \approx V$, the tensor-to-scalar ratio is a direct measure of the energy scale of inflation

$$\boxed{V^{1/4} \sim \left(\frac{r}{0.01} \right)^{1/4} 10^{16}\,{\rm GeV}} \,. \tag{532}$$

Large values of the tensor-to-scalar ratio, $r \geq 0.01$, correspond to inflation occuring at GUT scale energies.

The Lyth Bound

Note from Eqns. (517) and (530) that the tensor-to-scalar ratio relates directly to the evolution of the inflaton as a function of e-folds N

$$r = \frac{8}{M_{\rm pl}^2}\left(\frac{d\phi}{dN}\right)^2 . \tag{533}$$

The total field evolution between the time when CMB fluctuations exited the horizon at $N_{\rm cmb}$ and the end of inflation at $N_{\rm end}$ can therefore be written as the following integral

$$\frac{\Delta\phi}{M_{\rm pl}} = \int_{N_{\rm end}}^{N_{\rm cmb}} {\rm d}N \sqrt{\frac{r}{8}} . \tag{534}$$

During slow-roll evolution, $r(N)$ doesn't evolve much and one may obtain the following approximate relation [26]

$$\boxed{\frac{\Delta\phi}{M_{\rm pl}} = \mathcal{O}(1) \times \left(\frac{r}{0.01}\right)^{1/2}}, \tag{535}$$

where $r(N_{\rm cmb})$ is the tensor-to-scalar ratio on CMB scales. Large values of the tensor-to-scalar ratio, $r > 0.01$, therefore correlate with $\Delta\phi > M_{\rm pl}$ or *large-field inflation*.

7.4.5 Primordial Spectra

The results for the power spectra of the scalar and tensor fluctuations created by inflation are

$$\Delta_{\rm s}^2(k) \equiv \Delta_{\mathcal{R}}^2(k) = \left.\frac{1}{8\pi^2}\frac{H^2}{M_{\rm pl}^2}\frac{1}{\varepsilon}\right|_{k=aH} , \tag{536}$$

$$\Delta_{\rm t}^2(k) \equiv 2\Delta_h^2(k) = \left.\frac{2}{\pi^2}\frac{H^2}{M_{\rm pl}^2}\right|_{k=aH} , \tag{537}$$

where

$$\varepsilon = -\frac{d\ln H}{dN} . \tag{538}$$

The horizon crossing condition $k = aH$ makes (536) and (537) functions of the comoving wavenumber k. The tensor-to-scalar ratio is

$$r \equiv \frac{\Delta_{\rm t}^2}{\Delta_{\rm s}^2} = 16\,\varepsilon_\star . \tag{539}$$

Scale-Dependence

The scale dependence of the spectra follows from the time-dependence of the Hubble parameter and is quantified by the spectral indices

$$n_{\rm s} - 1 \equiv \frac{d\ln\Delta_{\rm s}^2}{d\ln k} , \qquad n_{\rm t} \equiv \frac{d\ln\Delta_{\rm t}^2}{d\ln k} . \tag{540}$$

We split this into two factors

$$\frac{d\ln\Delta_{\rm s}^2}{d\ln k} = \frac{d\ln\Delta_{\rm s}^2}{dN} \times \frac{dN}{d\ln k} . \tag{541}$$

The derivative with respect to e-folds is

$$\frac{d\ln\Delta_{\rm s}^2}{dN} = 2\frac{d\ln H}{dN} - \frac{d\ln\varepsilon}{dN}\,. \tag{542}$$

The first term is just -2ε and the second term may be evaluated with the following result from Appendix 7.8.4

$$\frac{d\ln\varepsilon}{dN} = 2(\varepsilon-\eta)\,, \qquad \text{where} \quad \eta = -\frac{d\ln H_{,\phi}}{dN}\,. \tag{543}$$

The second factor in Eqn. (541) is evaluated by recalling the horizon crossing condition $k = aH$, or

$$\ln k = N + \ln H\,. \tag{544}$$

Hence

$$\frac{dN}{d\ln k} = \left[\frac{d\ln k}{dN}\right]^{-1} = \left[1 + \frac{d\ln H}{dN}\right]^{-1} \approx 1+\varepsilon\,. \tag{545}$$

To first order in the Hubble slow-roll parameters we therefore find

$$n_{\rm s} - 1 = 2\eta_\star - 4\varepsilon_\star\,. \tag{546}$$

Similarly, we find

$$n_{\rm t} = -2\varepsilon_\star\,. \tag{547}$$

Any deviation from perfect scale-invariance ($n_{\rm s} = 1$ and $n_{\rm t} = 0$) is an indirect probe of the inflationary dynamics as quantified by the parameters ε and η.

Slow-Roll Results

In the slow-roll approximation the Hubble and potential slow-roll parameters are related as follows

$$\varepsilon \approx \epsilon_{\rm v}\,, \qquad \eta \approx \eta_{\rm v} - \epsilon_{\rm v}\,. \tag{548}$$

The scalar and tensor spectra are then expressed purely in terms of $V(\phi)$ and $\epsilon_{\rm v}$ (or $V_{,\phi}$)

$$\Delta_{\rm s}^2(k) \approx \frac{1}{24\pi^2}\frac{V}{M_{\rm pl}^4}\frac{1}{\epsilon_{\rm v}}\bigg|_{k=aH}\,, \qquad \Delta_{\rm t}^2(k) \approx \frac{2}{3\pi^2}\frac{V}{M_{\rm pl}^4}\bigg|_{k=aH}\,. \tag{549}$$

The scalar spectral index is

$$\boxed{n_{\rm s} - 1 = 2\eta_{\rm v}^\star - 6\epsilon_{\rm v}^\star}\,. \tag{550}$$

The tensor spectral index is

$$\boxed{n_{\rm t} = -2\epsilon_{\rm v}^\star}\,, \tag{551}$$

and the tensor-to-scalar ratio is

$$\boxed{r = 16\epsilon_{\rm v}^\star}\,. \tag{552}$$

We see that single-field slow-roll models satisfy a consistency condition between the tensor-to-scalar ratio r and the tensor tilt $n_{\rm t}$

$$\boxed{r = -8n_{\rm t}}\,. \tag{553}$$

In the slow-roll approximation measurements of the scalar and tensor spectra relate directly to the shape of the potential $V(\phi)$, *i.e.* H is a measure of of the scale of the potential, $\epsilon_{\rm v}$ of its first derivative $V_{,\phi}$, $\eta_{\rm v}$ of its second derivative $V_{,\phi\phi}$, etc. Measurements of the amplitude and the scale-dependence of the cosmological perturbations therefore encode information about the potential driving the inflationary expansion. This allows to reconstruct a power series expansion of the potential around $\phi_{\rm cmb}$ (corresponding to the time when CMB fluctuations exited the horizon).

Case Study: $m^2\phi^2$ Inflation

Recall from **Lecture 1** the slow-roll parameters for $m^2\phi^2$ inflation evaluated at $\phi_\star = \phi_{\rm cmb}$, *i.e.* $N_{\rm cmb} \sim 60$ e-folds before the end of inflation

$$\epsilon_{\rm v}^\star = \eta_{\rm v}^\star = 2\left(\frac{M_{\rm pl}}{\phi_{\rm cmb}}\right)^2 = \frac{1}{2N_{\rm cmb}}\,. \tag{554}$$

To satisfy the normalization of scalar fluctuations, $\Delta_{\rm s}^2 \sim 10^{-9}$, we need to fix the inflaton mass to $m \sim 10^{-6}M_{\rm pl}$. To see this note that Eqn. (549) implies

$$\Delta_{\rm s}^2 = \frac{m^2}{M_{\rm pl}^2}\frac{N_{\rm cmb}^2}{3}\,. \tag{555}$$

The scalar spectral index $n_{\rm s}$ and the tensor-to-scalar ratio r evaluated at CMB scales are

$$\boxed{n_{\rm s}} = 1 + 2\eta_{\rm v}^\star - 6\epsilon_{\rm v}^\star = 1 - \frac{2}{N_{\rm cmb}} \approx \boxed{0.96}\,, \tag{556}$$

and

$$\boxed{r} = 16\epsilon_{\rm v}^\star = \frac{8}{N_{\rm cmb}} \approx \boxed{0.1}\,. \tag{557}$$

These predictions of one of the simplest inflationary models are something to look out for in the near future.

7.5 Contact with Observations by Baumann 3rd lecture

Abstract

In this lecture we describe the inverse problem of extracting information on the inflationary perturbation spectra from observations of the cosmic microwave background and the large-scale structure. We define the precise relations between the scalar and tensor power spectra computed in the previous lecture and the observed CMB anisotropies and the galaxy power spectrum. We describe the transfer functions that relate the primordial fluctuations to the late-time observables. We then use these results to discuss the current observational evidence for inflation. Finally, we indicate opportunities for future tests of inflation.

7.5.1 Connecting Observations to the Early Universe

> "It doesn't matter how beautiful your theory is, it doesn't matter how smart you are or what your name is. If it doesn't agree with experiment, it's wrong."
>
> *Richard Feynman*

In the last lecture we computed the power spectra of the primordial scalar and tensor fluctuations $\mathcal{R}$ and h at horizon exit. In this lecture we relate these results to observations of the cosmic microwave background (CMB) and the large-scale structure (LSS). Making this correspondence explicit is crucial for constraining the inflationary predictions.

The curvature perturbation $\mathcal{R}$ and the gravitational wave amplitude h both freeze at constant values once the mode exits the horizon, $k = a(\tau_\star)H(\tau_\star)$. In the previous lecture we therefore computed the primordial perturbations at the time of horizon exit, $\tau_\star$. To relate this to a cosmological observable (like the CMB temperature or the density of galaxies) we need to

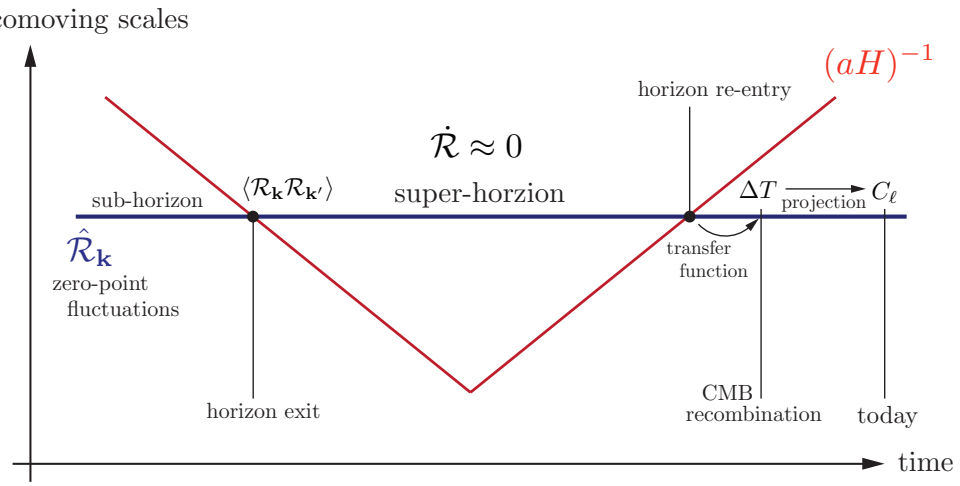


Figure 69: Creation and evolution of perturbations in the inflationary universe. Fluctuations are created quantum mechanically on subhorizon scales (see **Lecture 2**). While comoving scales, k^{-1}, remain constant the comoving Hubble radius during inflation, $(aH)^{-1}$, shrinks and the perturbations exit the horizon and freeze until horizon re-entry at late times. After horizon re-entry the fluctuations evolve into anisotropies in the CMB and perturbations in the LSS. This time-evolution has to be accounted for to relate cosmological observations to the primordial perturbations laid down by inflation (see **Lecture 3**).

i) relate $\mathcal{R}$ (or h) to the quantity $\mathcal{Q}$ that is actually measured in an experiment and

ii) take into account the time evolution of $\mathcal{R}$ (and $\mathcal{Q}$) once it re-enters the horizon.

Schematically, we may write

$$\boxed{\mathcal{Q}_{\mathbf{k}}(\tau) = T_{\mathcal{Q}}(k, \tau, \tau_\star)\, \mathcal{R}_{\mathbf{k}}(\tau_\star)}\,, \tag{558}$$

where $T_{\mathcal{Q}}$ is the *transfer function* between $\mathcal{R}$ fluctuations at time $\tau_\star$ and $\mathcal{Q}$ fluctuations at some later time τ. As we have indicated the transfer function may depend on scale. The quantity $\mathcal{Q}$ may be the temperature fluctuations measured by a CMB satellite such as the Wilkinson Microwave Anisotropy Probe (WMAP) or the galaxy density inferred in a galaxy survey such as the Sloan Digital Sky Survey (SDSS).

CMB anisotropies

The main result of §7.5.2 will be the following relation between the inflationary input spectra $P(k) \equiv \{P_{\mathcal{R}}(k), P_h(k)\}$ and the angular power spectra of CMB temperature fluctuations and polarization

$$\boxed{C_\ell^{XY} = \frac{2}{\pi} \int k^2 \mathrm{d}k \underbrace{P(k)}_{\text{Inflation}} \underbrace{\Delta_{X\ell}(k) \Delta_{Y\ell}(k)}_{\text{Anisotropies}}}\,, \tag{559}$$

where

$$\Delta_{X\ell}(k) = \int_0^{\tau_0} \mathrm{d}\tau \underbrace{S_X(k, \tau)}_{\text{Sources}} \underbrace{P_{X\ell}(k[\tau_0 - \tau])}_{\text{Projection}}\,. \tag{560}$$

The labels X, Y refer to temperature T and polarization modes E and B (see §7.5.2). The integral (559) relates the inhomogeneities predicted by inflation, $P(k)$, to the anisotropies observed in the CMB, C_ℓ^{XY}. The correlations between the different X and Y modes are related by the transfer functions $\Delta_{X\ell}(k)$ and $\Delta_{Y\ell}(k)$. The transfer functions may be written as the line-of-sight integral (560) which factorizes into physical source terms $S_X(k, \tau)$ and geometric

projection factors $P_{X\ell}(k[\tau_0 - \tau])$ (combinations of Bessel functions). A derivation of the source terms and the projection factors is beyond the scope of this lecture, but may be found in Dodelson's book [8]. An intuitive explanation for these results may be found in the animations on Wayne Hu's website [27].

Our interest in this lecture lies in experimental constraints on the primordial power spectra $P_{\mathcal{R}}(k)$ and $P_h(k)$. To measure the primordial spectra the observed CMB anisotropies C_ℓ^{XY} need to be *deconvolved* by taking into account the appropriate transfer functions and projection effects, *i.e.* for a given background cosmology we can compute the evolution and projection effects in Eqn. (559) and therefore extract the inflationary initial conditions $P(k)$. By this deconvolution procedure, the CMB provides a fascinating probe of the early universe.

Large-scale structure

To study fluctuations in the matter distribution (as measured *e.g.* by the distribution of galaxies on the sky) we define the density contrast $\delta \equiv \delta\rho/\bar{\rho}$. We distinguish between fluctuations in the density of galaxies δ_g and the dark matter density δ. A common assumption is that galaxies are (biased) tracers of the underlying dark matter distribution, $\delta_g = b\,\delta$. If we have an independent way of determining the *bias* parameter b, we can use observations of the galaxy density contrast δ_g to infer the underlying dark matter distribution δ. The late-time power spectrum of dark matter density fluctuations is related to the primordial spectrum of curvature fluctuations as follows

$$\boxed{P_\delta(k,\tau) = \frac{4}{25}\left(\frac{k}{aH}\right)^4 T_\delta^2(k,\tau) P_{\mathcal{R}}(k)}\,. \tag{561}$$

The numerical factor and the k-scaling that have been factored out from the transfer function is conventional. The transfer function T_δ reflects the relative growth of fluctuations during matter domination, $\delta \sim a$, and radiation domination, $\delta \sim \ln a$. It usually has to be computed numerically using codes such as CMBFAST [28] or CAMB [166], however, in §7.5.3 we will cite useful fitting functions for T_δ. Again, since for a fixed background cosmology the transfer function can be assumed as given, observations of the matter power spectrum can be a probe of the initial fluctuations from the early universe.

7.5.2 *Review*: The Cosmic Microwave Background

We give a very brief review of the physics and the statistical interpretation of CMB fluctuations. More details may be found in Dodelson's book [8] or Prof. Pierpaoli's lectures at *TASI 2009*.

Temperature Anisotropies

Harmonic Expansion

Figure 70 shows a map of the measured CMB temperature fluctuations $\Delta T(\hat{n})$ relative to the background temperature $T_0 = 2.7$ K. Here the unit vector $\hat{n}$ denotes the direction in sky. The harmonic expansion of this map is

$$\Theta(\hat{n}) \equiv \frac{\Delta T(\hat{n})}{T_0} = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{n})\,, \tag{562}$$

where

$$a_{\ell m} = \int \mathrm{d}\Omega\, Y^*_{\ell m}(\hat{n})\Theta(\hat{n})\,. \tag{563}$$

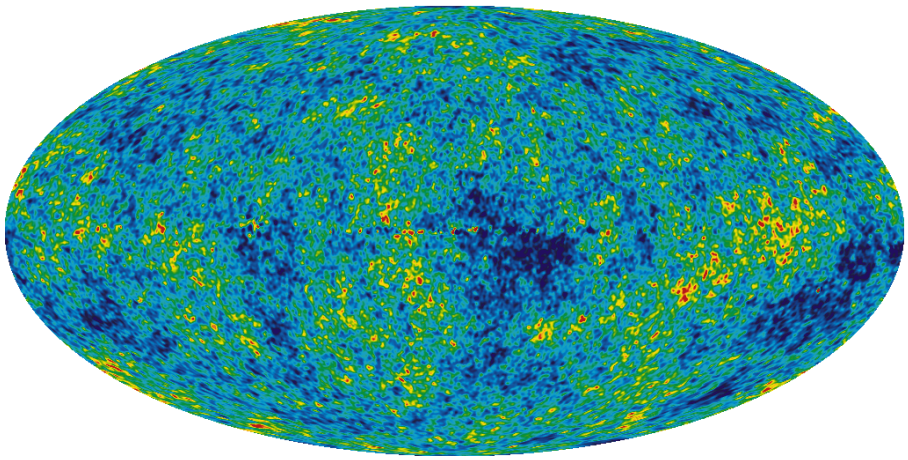


Figure 70: Temperature fluctuations in the CMB. Blue spots represent directions on the sky where the CMB temperature is $\sim 10^{-5}$ below the mean, $T_0 = 2.7$ K. This corresponds to photons losing energy while climbing out of the gravitational potentials of overdense regions in the early universe. Yellow and red indicate hot (underdense) regions. The statistical properties of these fluctuations contain important information about both the background evolution and the initial conditions of the universe.

Here, $Y_{\ell m}(\hat{n})$ are the standard spherical harmonics on a 2-sphere with $\ell = 0$, $\ell = 1$ and $\ell = 2$ corresponding to the monopole, dipole and quadrupole, respectively. The magnetic quantum numbers satisfy $m = -\ell, \ldots, +\ell$. The multipole moments $a_{\ell m}$ may be combined into the rotationally-invariant angular power spectrum

$$C_\ell^{TT} = \frac{1}{2\ell+1} \sum_m \langle a_{\ell m}^* a_{\ell m} \rangle \,, \qquad \text{or} \qquad \langle a_{\ell m}^* a_{\ell' m'} \rangle = C_\ell^{TT} \delta_{\ell\ell'} \delta_{mm'} \,. \tag{564}$$

The angular power spectrum is an important tool in the statistical analysis of the CMB. It describes the cosmological information contained in the millions of pixels of a CMB map in terms of a much more compact data representation. Figure 71 shows the most recent measurements of the CMB angular power spectrum. The figure also shows a fit of the theoretical prediction for the CMB spectrum to the data. The theoretical curve depends both on the background cosmological parameters and on the spectrum of initial fluctuations. We hence can use the CMB as a probe of both.

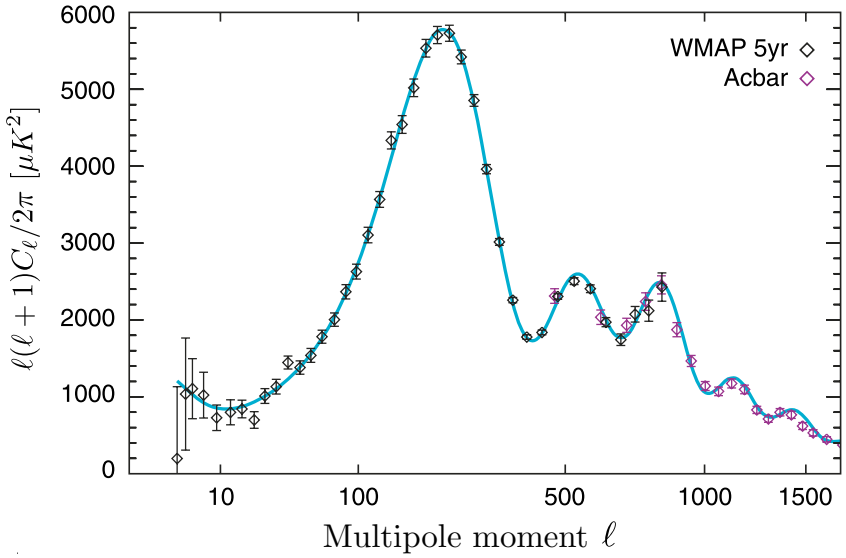


Figure 71: Angular power spectrum of CMB temperature fluctuations.

CMB temperature fluctuations are dominated by the *scalar* modes $\mathcal{R}$ (at least for the values of the tensor-to-scalar ratio now under consideration, $r < 0.3$). The linear evolution which relates $\mathcal{R}$ and ΔT is mediated by the transfer function $\Delta_{T\ell}(k)$ through the k-space integral [8]

$$\boxed{a_{\ell m} = 4\pi(-i)^\ell \int \frac{\mathrm{d}^3k}{(2\pi)^3}\, \Delta_{T\ell}(k)\, \mathcal{R}_{\mathbf{k}}\, Y_{\ell m}(\hat{\mathbf{k}})}\,. \tag{565}$$

Substituting (565) into (564) and using the identity

$$\sum_{m=-\ell}^{\ell} Y_{\ell m}(\hat{\mathbf{k}}) Y_{\ell m}(\hat{\mathbf{k}}') = \frac{2\ell+1}{4\pi} P_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}')\,, \tag{566}$$

we find

$$C_\ell^{TT} = \frac{2}{\pi}\int k^2 \mathrm{d}k \underbrace{P_{\mathcal{R}}(k)}_{\text{Inflation}} \underbrace{\Delta_{T\ell}(k)\Delta_{T\ell}(k)}_{\text{Anisotropies}}\,. \tag{567}$$

The transfer functions $\Delta_{T\ell}(k)$ generally have to be computed numerically using Boltzmann-codes such as CMBFAST [28] or CAMB [166]. They depend on the parameters of the background cosmology. Assuming a fixed background cosmology the shape of the power spectrum C_ℓ^{TT} contains information about the initial conditions as described by the primoridial power spectrum $P_{\mathcal{R}}(k)$.[31] Of course, learning from observations about $P_{\mathcal{R}}(k)$ and hence about inflation is the primary objective of this lecture.

Large Scales

On large scales, modes were still outside of the horizon at recombination. The large-scale CMB spectrum has therefore not been affected by subhorizon evolution and is simply the geometric projection of the primordial spectrum from recombination to us today. In this *Sachs-Wolfe regime* the transfer function $\Delta_{T\ell}(k)$ is simply a Bessel function [8]

$$\Delta_{T\ell}(k) = \frac{1}{3} j_\ell(k[\tau_0 - \tau_{\mathrm{rec}}])\,. \tag{568}$$

The angular power spectrum on large scales (small ℓ) therefore is

$$C_\ell^{TT} = \frac{2}{9\pi}\int k^2 \mathrm{d}k\, P_{\mathcal{R}}(k)\, j_\ell^2(k[\tau_0 - \tau_{\mathrm{rec}}])\,. \tag{569}$$

The Bessel projection function is peaked at $k[\tau_0 - \tau_{\mathrm{rec}}] \approx \ell$ and so effectively acts like a δ-function mapping between k and ℓ. Given that modes with wavenumber $k \approx \ell/(\tau_0 - \tau_{\mathrm{rec}})$ domintate the integral in Eqn. (569), we can write

$$C_\ell^{TT} \propto k^3 P_{\mathcal{R}}(k)\big|_{k\approx\ell/(\tau_0-\tau_{\mathrm{rec}})} \underbrace{\int \mathrm{d}\ln x\; j_\ell^2(x)}_{\propto \ell(\ell+1)}\,. \tag{570}$$

Hence,

$$\ell(\ell+1)C_\ell^{TT} \propto \Delta_{\mathrm{s}}^2(k)\big|_{k\approx\ell/(\tau_0-\tau_{\mathrm{rec}})} \propto \ell^{n_{\mathrm{s}}-1}\,. \tag{571}$$

For a scale-invariant input spectrum, $n_{\mathrm{s}} = 1$, the quantity

$$\mathcal{C}_\ell \equiv \frac{\ell(\ell+1)}{2\pi} C_\ell^{TT} \tag{572}$$

is independent of ℓ (except for a rise at very low ℓ due to the integrated Sachs-Wolfe effect arising from the late-time evolution of the gravitational potential in a dark energy dominated universe). This explains why the CMB power spectrum is often plotted for $\mathcal{C}_\ell$ instead of C_ℓ^{TT}.

[31] In practice, the CMB data is fit simultaneously to the background cosmology and a spectrum of fluctuations.

Non-Gaussianity

So far we have shown that the angular power spectrum of CMB fluctuations essentially is a measure of the primordial power spectrum $P_{\mathcal{R}}(k)$ if we take into account subhorizon evolution and geometric projection effects

$$\langle a^*_{\ell m} a_{\ell' m'} \rangle = C_\ell^{TT} \delta_{\ell\ell'} \delta_{mm'} \quad \Leftrightarrow \quad \langle \mathcal{R}_{\mathbf{k}} \mathcal{R}_{\mathbf{k}'} \rangle = (2\pi)^3 P_{\mathcal{R}}(k)\, \delta(\mathbf{k} + \mathbf{k}') \,. \tag{573}$$

If the primordial fluctuations are Gaussian then $P_{\mathcal{R}}(k)$ contains all the information. Single-field slow-roll inflation in fact predicts that $\mathcal{R}$ should be Gaussian to a very high degree [22]. However, as we explain in **Lecture 4**, even a small amount of non-Gaussianity would provide crucial information about the inflationary action as it would require to go beyond the simplest single-field slow-roll models.

The primary measure of non-Gaussianity is the *three-point function* or equivalently the *bispectrum*

$$\langle \mathcal{R}_{\mathbf{k}_1} \mathcal{R}_{\mathbf{k}_2} \mathcal{R}_{\mathbf{k}_3} \rangle = (2\pi)^3 B_{\mathcal{R}}(k_1, k_2, k_3)\, \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \,. \tag{574}$$

In the CMB a non-zero bispectrum $B_{\mathcal{R}}(k_1, k_2, k_3)$ leaves a signature in the angular bispectrum

$$B^{\ell_1 \ell_2 \ell_3}_{m_1 m_2 m_3} = \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle \,. \tag{575}$$

Substituting (565) into (575) we may relate the primordial bispectrum to the observed CMB bispectrum [30].[32]

Note that the primordial bispectrum $B_{\mathcal{R}}(k_1, k_2, k_3)$ is a function of three momenta subject only to momentum conservation (*i.e.* the three vectors $\mathbf{k}_i$ form a closed triangle). This makes observational constraints on non-Gaussianity challenging (there are many different forms of non-Gaussianity to consider), but also means that if detected non-Gaussianity contains a lot of information about the physics of the early universe.

A simple model of primoridal non-Gaussianity is *local non-Gaussianity* defined by a Taylor expansion of the curvature perturbation around the Gaussian part $\mathcal{R}_g$

$$\mathcal{R}(\mathbf{x}) = \mathcal{R}_g(\mathbf{x}) + \frac{3}{5} f^{\mathcal{R}}_{\rm NL} \star \mathcal{R}_g^2(\mathbf{x}) \,. \tag{576}$$

This is local in real space and the parameter $f^{\mathcal{R}}_{\rm NL}$ characterizes the level of non-Gaussianity. The reader is invited to show that Eqn. (576) implies the following simple form for the bispectrum

$$B_{\mathcal{R}}(k_1, k_2, k_3) = \frac{6}{5} f^{\mathcal{R}}_{\rm NL} \Big[P_{\mathcal{R}}(k_1) P_{\mathcal{R}}(k_2) + P_{\mathcal{R}}(k_2) P_{\mathcal{R}}(k_3) + P_{\mathcal{R}}(k_3) P_{\mathcal{R}}(k_1) \Big] \,. \tag{577}$$

Present observational constraints on non-Gaussianity are therefore often phrased as constraints on the parameter $f^{\mathcal{R}}_{\rm NL}$ (see §7.5.4).

Polarization

CMB polarization is likely to become one of the most important tools to probe the physics governing the early universe. In addition to anisotropies in the CMB temperature, we expect the CMB to become polarized via Thomson scattering [8]. As we now explain, this polarization contains crucial information about the primordial fluctuations and hence about inflation [7].

[32] Non-linear evolution can lead to additional non-Gaussianity (see **Lecture 4**).

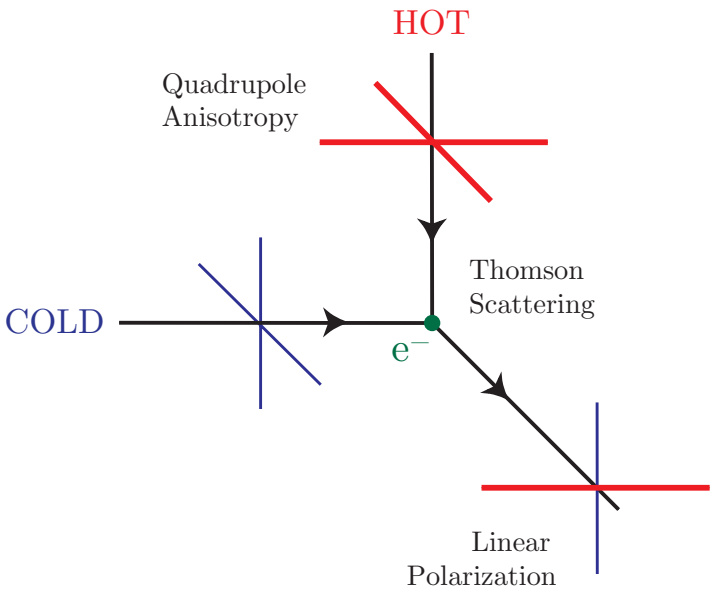


Figure 72: Linear polarization is generation via Thomson scattering of radiation with a quadrupolar anisotropy. Here, the red (thick) lines represent hot radiation and the blue (thin) line cold radiation.

Polarization via Thomson Scattering

Let us see how polarization is generated by the scattering between photons and free electrons. If the incident radiation pattern is isotropic (in the rest frame of the electron), then the outgoing radiation remains unpolarized because orthogonal polarization directions cancel out. A net linear polarization only arises if the incoming radiation field has a quadrupole component (see Fig. 72). Such a quadrupole moment is generated when photons decouple from the electrons and protons just before recombination. Since the temperature anisotropies are created by primordial density fluctuations, a component of the polarization should be correlated with the temperature anisotropy.

Characterization of the Radiation Field

The mathematical characterization of CMB polarization anisotropies is slightly more involved than that the description of temperature fluctuations because polarization is not a scalar field so the standard expansion in terms of spherical harmonics is not applicable.

The anisotropy field is defined in terms of a 2×2 intensity tensor $I_{ij}(\hat{n})$, where as before $\hat{n}$ denotes the direction on the sky. The components of I_{ij} are defined relative to two orthogonal basis vectors $\hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_2$ perpendicular to $\hat{n}$. Linear polarization is then described by the Stokes parameters $Q = \frac{1}{4}(I_{11} - I_{22})$ and $U = \frac{1}{2}I_{12}$, while the temperature anisotropy is $T = \frac{1}{4}(I_{11} + I_{22})$. The polarization magnitude and angle are $P = \sqrt{Q^2 + U^2}$ and $\alpha = \frac{1}{2}\tan^{-1}(U/Q)$. The quantity T is invariant under a rotation in the plane perpendicular to $\hat{n}$ and hence may be expanded in terms of scalar (spin-0) spherical harmonics (562). The quantities Q and U, however, transform under rotation by an angle ψ as a spin-2 field $(Q \pm iU)(\hat{n}) \to e^{\mp 2i\psi}(Q \pm iU)(\hat{n})$. The harmonic analysis of $Q \pm iU$ therefore requires expansion on the sphere in terms of tensor (spin-2) spherical harmonics [197, 1888, 33]

$$(Q \pm iU)(\hat{n}) = \sum_{\ell,m} a_{\pm 2,\ell m}\, {}_{\pm 2}Y_{\ell m}(\hat{n})\,. \tag{578}$$

A description of the mathematical properties of these tensor spherical harmonics, ${}_{\pm 2}Y_{\ell m}$, would

take us too far off the main track of this lecture, so we refer the reader to the classic papers [197, 1888] or Dodelson's book [8].

E and B-modes

Instead of the moments $a_{\pm 2,\ell m}$ it is convenient to introduce the linear combinations

$$a_{E,\ell m} \equiv -\frac{1}{2}\left(a_{2,\ell m} + a_{-2,\ell m}\right) , \qquad a_{B,\ell m} \equiv -\frac{1}{2i}\left(a_{2,\ell m} - a_{-2,\ell m}\right) . \tag{579}$$

Then one can define two scalar (spin-0) fields instead of the spin-2 quantities Q and U

$$E(\hat{n}) = \sum_{\ell,m} a_{E,\ell m}\, Y_{\ell m}(\hat{n}) , \qquad B(\hat{n}) = \sum_{\ell,m} a_{B,\ell m}\, Y_{\ell m}(\hat{n}) . \tag{580}$$

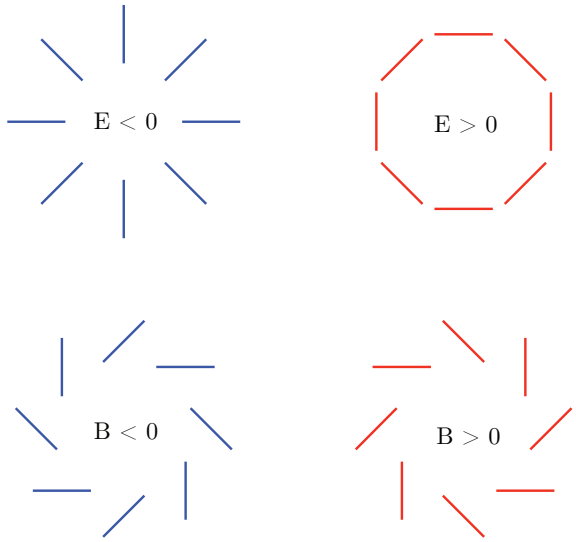


Figure 73: Examples of E-mode and B-mode patterns of polarization. Note that if reflected across a line going through the center the E-mode patterns are unchanged, while the positive and negative B-mode patterns get interchanged.

The scalar quantities E and B completely specify the linear polarization field. E-mode polarization is often also characterized as a *curl-free* mode with polarization vectors that are radial around cold spots and tangential around hot spots on the sky. In contrast, B-mode polarization is *divergence-free* but has a *curl*: its polarization vectors have vorticity around any given point on the sky.[33] Fig. 73 gives examples of E- and B-mode patterns. Although E and B are both invariant under rotations, they behave differently under parity transformations. Note that when reflected about a line going through the center, the E-mode patterns remain unchanged, while the B-moe patterns change sign.

The symmetries of temperature and polarization (E- and B-mode) anisotropies allow four types of correlations: the autocorrelations of temperature fluctuations and of E- and B-modes denoted by TT, EE, and BB, respectively, as well as the cross-correlation between temperature fluctuations and E-modes: TE. All other correlations (TB and EB) vanish for symmetry reasons.

The angular power spectra are defined as before

$$C_\ell^{XY} \equiv \frac{1}{2\ell+1}\sum_m \langle a^*_{X,\ell m} a_{Y,\ell m}\rangle , \qquad X, Y = T, E, B . \tag{581}$$

[33] Evidently the E and B nomenclature reflects the properties familiar from electrostatics, $\nabla \times \mathbf{E} = 0$ and $\nabla \cdot \mathbf{B} = 0$.

In Fig. 74 we show the latest measurement of the TE cross-correlation [11]. The EE spectrum has now begun to be measured, but the errors are still large. So far there are only upper limits on the BB spectrum, but no detection.

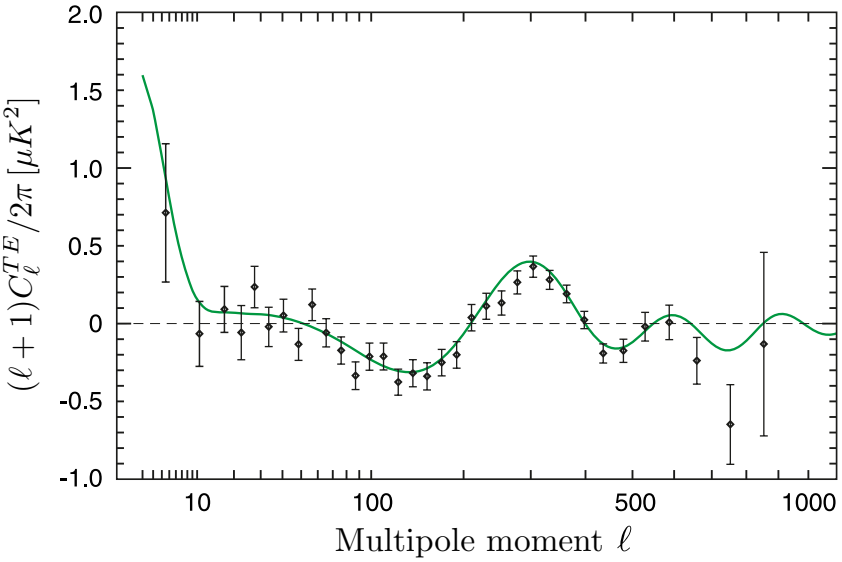


Figure 74: Power spectrum of the cross-correlation between temperature and E-mode polarization anisotropies [11]. The anti-correlation for $\ell = 50 - 200$ (corresponding to angular separations $5^\circ > \theta > 1^\circ$) is a distinctive signature of adiabatic fluctuations on superhorizon scales at the epoch of decoupling [34, 35], confirming a fundamental prediction of the inflationary paradigm.

The cosmological significance of the E/B decomposition of CMB polarization was realized by the authors of Refs. [197, 1888], who proved the following remarkable facts:

i) scalar (density) perturbations create only E-modes and *no* B-modes.

ii) vector (vorticity) perturbations create mainly B-modes.[34]

iii) tensor (gravitational wave) perturbations create both E-modes and B-modes.

The fact that *scalars do not produce B-modes* while *tensors do* is the basis for the statement that detection of B-modes is a smoking gun of tensor modes, and therefore of inflation.

E-modes and Scalars

The power spectrum of E-modes and the TE cross-correlation is dominated by inflationary *scalar* modes, *i.e.*

$$C_\ell^{EE} \approx (4\pi)^2 \int k^2 \mathrm{d}k \overbrace{P_{\mathcal{R}}(k)}^{\text{Inflation}} \Delta_{E\ell}^2(k)\,, \tag{582}$$

$$C_\ell^{TE} \approx (4\pi)^2 \int k^2 \mathrm{d}k \underbrace{P_{\mathcal{R}}(k)}_{\text{Inflation}} \Delta_{T\ell}(k)\Delta_{E\ell}(k)\,. \tag{583}$$

Like C_ℓ^{TT}, the spectra C_ℓ^{EE} and C_ℓ^{TE} provide information about $P_{\mathcal{R}}(k)$. However, since the primordial spectrum is convolved with different transfer functions in each case (polarization is generated only by scattering from free electrons), the signals are usefully complementary.

[34] However, vectors decay with the expansion of the universe and are therefore believed to be subdominant at recombination. We therefore do not consider them here.

B-modes and Tensors

B-modes are only generated by *tensor* modes, *i.e.*

$$C_\ell^{BB} = (4\pi)^2 \int k^2 \mathrm{d}k \underbrace{P_h(k)}_{\text{Inflation}} \Delta^2_{B\ell}(k)\,. \tag{584}$$

Measuring C_ℓ^{BB} is therefore a unique opportunity to access information about primordial tensor fluctuations.

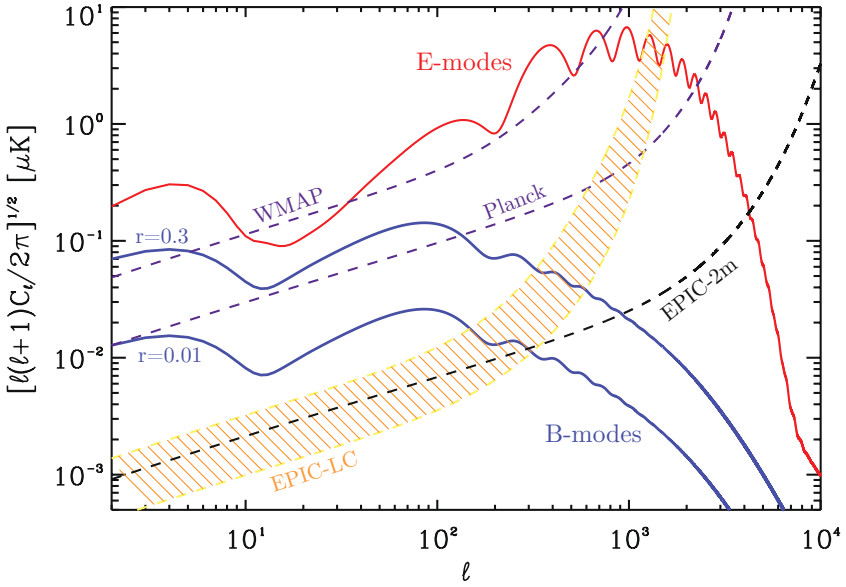


Figure 75: *E*- and *B*-mode power spectra for a tensor-to-scalar ratio saturating current bounds, $r = 0.3$, and for $r = 0.01$. Shown are also the experimental sensitivities for *WMAP*, *Planck* and two different realizations of a future CMB satellite (*CMBPol*) (EPIC-LC and EPIC-2m) [36].

7.5.3 *Review*: Large-Scale Structure

The galaxy (or dark matter) power spectrum is a measure of the spectrum of primordial curvature fluctuations

$$P_\delta(k, z) \quad \Rightarrow \quad P_{\mathcal{R}}(k)\,, \tag{585}$$

if the effects of subhorizon evolution are accounted for. This is done by the dark matter transfer function.

Dark Matter Transfer Functions

Density fluctuations evolve under the competing influence of pressure and gravity. During radiation domination the large radiation pressure prevents the rapid growth of fluctuations; the density contrast only grows logarithmically, $\delta \sim \ln a$. During matter domination the background pressure is negligible and gravitational collapse operates more effectively, $\delta \sim a$.

Under the simplifying assumption that there is no significant growth of perturbations between the time of horizon entry and matter domination one may derive the following approximate transfer function

$$T_\delta(k) \approx \begin{cases} 1 & k < k_{\rm eq} \\ (k_{\rm eq}/k)^2 & k > k_{\rm eq} \end{cases}\,. \tag{586}$$

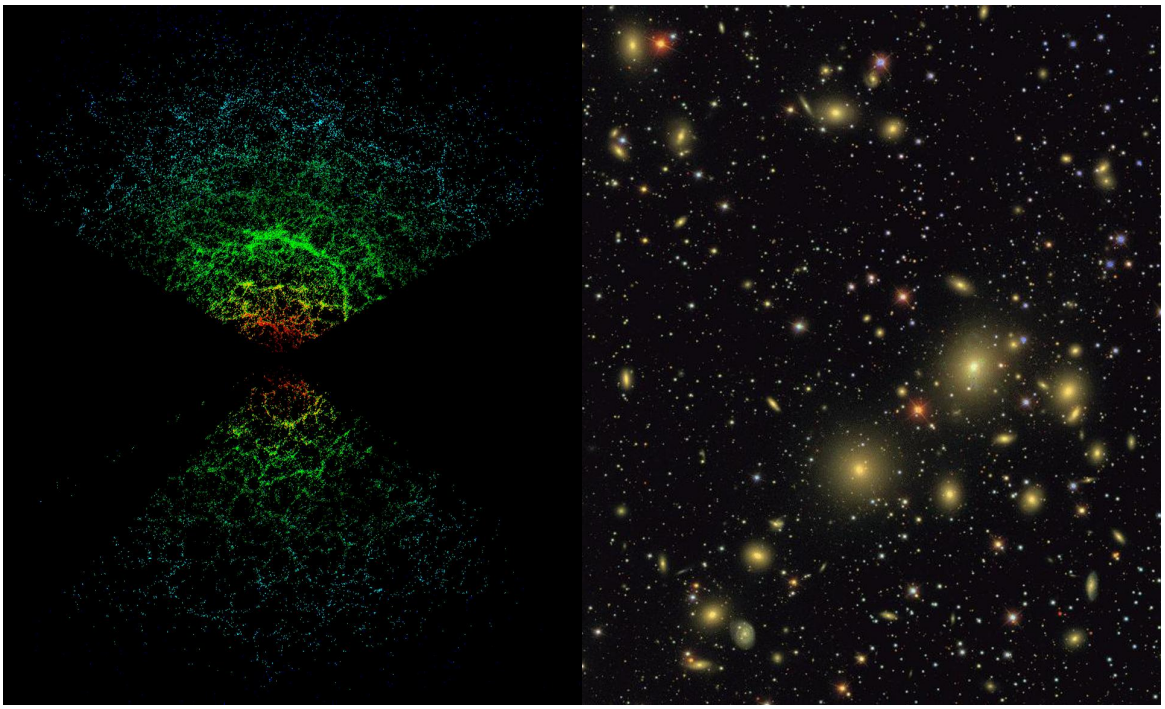


Figure 76: Distribution of galaxies. The Sloan Digital Sky Survey (SDSS) has measured the positions and distances (redshifts) of nearly a million galaxies. Galaxies first identified on 2d images, like the one shown above on the right, have their distances measured to create the 3d map. The left image shows a slice of such a 3d map. The statistical properties of the measured distribution of galaxies reveal important information about the structure and evolution of the late time universe.

Although Eqn. (586) is intuitively appealing for understanding the qualitative shape of the spectrum (*i.e.* the break in the spectrum at $k \approx k_{\rm eq}$), it is not accurate enough for most applications. A famous fitting function for the matter transfer function was given by Bardeen *et al.* (BBKS) [37]

$$T_\delta(q) = \frac{\ln(1+2.34q)}{2.34q}\left[1+3.89q+(1.61q)^2+(5.46q)^3+(6.71q)^4\right]^{-1/4}\,, \tag{587}$$

where $q = k/\Gamma h\,\mathrm{Mpc}^{-1}$ and we defined the shape parameter

$$\Gamma \equiv \Omega h \exp(-\Omega_b - \sqrt{2h}\Omega_b/\Omega)\,. \tag{588}$$

More accurate transfer functions may be found in Eisenstein and Hu [38]. Finally, exact transfer functions may be computed numerically with CMBFAST [28] or CAMB [166].

For our purposes it is only important to note that (give the background cosmological parameters) the dark matter transfer function can be computed and used to relate the dark matter power spectrum $P_\delta(k,z)$ to the inflationary spectrum $P_{\mathcal{R}}(k)$.

Galaxy Bias

With the exception of gravitational lensing we unfortunately can't observe the dark matter directly. What we observe (*e.g.* in galaxy surveys like the Sloan Digital Sky Survey (SDSS)) is luminous or baryonic matter. On large scales the following phenomenological ansatz for relating the galaxy distribution and the dark matter has proven useful

$$\delta_g = b\,\delta\,, \tag{589}$$

or

$$P_{\delta_g} = b^2 P_\delta \,. \tag{590}$$

Here, b is called the (linear) bias parameter. It may be viewed as a parameter describing the ill-understood physics of galaxy formation. The bias parameter b can be obtained by measuring the galaxy bispectrum B_{δ_g}.

Modulo these complications the galaxy power spectrum $P_{\delta_g}(k)$ is an additional probe of inflationary scalar fluctuations $P_{\mathcal{R}}(k)$. As it probes smaller scales it is complementary to observations of the CMB fluctuations.

7.5.4 Current Evidence for Inflation

Inflation is a hypothesis. In order to increase our confidence that inflation describes the physical reality of the early universe, we compare the predictions of inflation to cosmological observations. In this section we describe the current observational evidence for inflation, before discussing future tests in the next section.

Flatness

The universe is filled with baryons, dark matter, photons, neutrinos and dark energy

$$\Omega_{\rm tot} = \Omega_b + \Omega_{\rm cdm} + \Omega_\gamma + \Omega_\nu + \Omega_\Lambda \,. \tag{591}$$

The value of $\Omega_{\rm tot}$ determines the spatial geometry of the universe with $\Omega_{\rm tot} = 1$ corresponding to a flat universe, $\Omega_{\rm tot} < 1$ to a negatively curved universe and $\Omega_{\rm tot} > 1$ to a positively curved universe. Inflation predicts

$$\Omega_{\rm tot} = 1 \pm 10^{-5} \,, \tag{592}$$

while the data shows [11]

$$\Omega_{\rm tot} = 1 \pm 0.02 \,. \tag{593}$$

Although this agreement between theory and data is impressive, one could argue that inflation achieves the flatness of the universe somewhat 'by design'.[35] We should therefore search for additional tests of the inflationary idea.

Coherent Phases and Superhorizon Fluctuations

As we have repeatedly emphasized in these lectures, the observations of the inhomogeneous universe allow detailed test of the inflationary dynamics. In this subsection, we discuss non-trivial qualitative features of the observations that inflation explains naturally, before giving quantitative results in the next subsection.

The following is a trivialization of arguments that have been explained beautifully by Dodelson in [35].

The Peaks of the TT Spectrum

Inflation produces a nearly scale-invariant spectrum of perturbations,

$$\langle \mathcal{R}_{\mathbf{k}} \mathcal{R}_{\mathbf{k}'} \rangle = (2\pi)^3 \delta(\mathbf{k} + \mathbf{k}') P_{\mathcal{R}}(k) \,, \tag{594}$$

where $k^3 P_{\mathcal{R}}(k) \propto k^{n_{\rm s}-1}$ with $n_{\rm s} \approx 1$.

[35] However, it is worth pointing out that when Guth introduced inflation in 1980, the flatness of the universe was a non-trivial prediction that at the time was inconsistent with observations!

> You might think then that the shape of the power spectrum can be measured in observations, and this is what convinces us that inflation is right. Well, it is true that we can measure the power spectrum, both of the matter and of the radiation, and it is true that the observations agree with the theory. But this is not what tingles our spines when we look at the data. Rather, the truly striking aspect of perturbations generated during inflation is that *all Fourier modes have the same phase* [35].

Consider a Fourier mode with physical wavelength λ. While the mode is inside the horizon during inflation it oscillates with a frequency given by $k = 2\pi/\lambda$. However, before inflation ends, the mode exits the horizon, *i.e.* its physical wavelength gets stretched to a length greater than the instantaneous Hubble radius, $\lambda > H^{-1}$. After that its amplitude remains constant. Only at a much later time when the mode re-enters the horizon can causal physics affect it and lead to a time-evolution. Since the fluctuation amplitude was constant outside the horizon, $\dot{\mathcal{R}}$ is very small at horizon re-entry. In general, each Fourier mode could be a linear combination of a sine and a cosine mode. However, the special feature of inflation is that excites only the cosine mode (defining horizon re-entry as $t \equiv 0$).

Once inside the horizon the curvature perturbation $\mathcal{R}$ sources density fluctuations δ which evolve under the influence of gravity and pressure

$$\ddot{\delta} - c_s^2 \nabla^2 \delta = F_g[\mathcal{R}]\,, \tag{595}$$

where c_s is the sound speed and F_g is the gravitational source term. This leads to oscillations in the density field. In the plasma of the early universe, fluctuations in the matter density were strongly coupled to fluctuations in the radiation. The CMB fluctuations therefore provide a direct snapshot of the conditions of the underlying density field at the time of recombination. Imagine that recombination happens instantaneously (this is not a terrible approximation). Fluctuations with different wavelengths would be captured at different phases in their oscillations. Modes of a certain wavelength would be captured at maximum or minimum amplitude, while others would be captured at zero amplitude. If all Fourier modes of a given wavelength had the same phases they would interfere coherently and the spectrum of all Fourier would produce a series of peaks and troughs in the CMB power spectrum as seen on the last-scattering surface. This is of course what we see in Fig. 70. However, in order for the theory of initial fluctuations to explain this it needs to involve a mechanism that produces coherent initial phases for all Fourier modes. Inflation does precisely that! Because fluctuations freeze when the exit the horizon the phases for the Fourier modes were set well before the modes of interest entered the horizon. When were are admiring the peak structure of the CMB power spectrum we are really admiring the ability of the primordial mechanism for generating flucutations to coordinate the phases of all Fourier modes. Without this coherence, the CMB power spectrum would simply be white noise with no peaks and troughs (in fact, this is precisely why cosmic strings or topological defects are ruled out at the primary sources for the primordial fluctuations.).

$\ell < 100$ in the TE Spectrum

The skeptic might not be convinced by the above argument. The peaks and troughs of the CMB temperature fluctuation spectrum are at $\ell > 200$ corresponding to angular scales $\theta < 1^\circ$. All of these scales were within the horizon at the time of recombination. So it is in principle possible (and people have tried in the 90s) to engineer a theory of structure formation which obeys causality and still manages to produce only the 'cosine mode'. Such a theory would explain the CMB peaks without appealing to inflation. This doesn't sound like the most elegant thing in the world but it can't be excluded as a logical possibility.

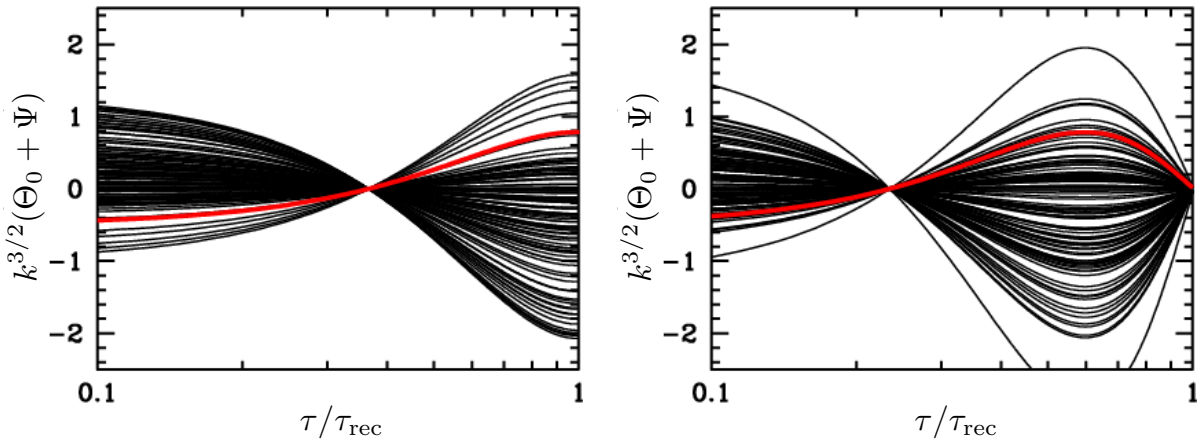


Figure 77: Evolution of modes with the same wavelength. Recombination is at $\tau = \tau_{\rm rec}$. The left figure illustrates the wavelength corresponding to the first peak in the CMB angular power spectrum, while the right figure shows the wavelength corresponding to the first trough. Since all modes start with the same phase, the ones on the left all reach maximum amplitude at recombination, while the ones on the right all go to zero at recombination. This explains to the peaks of the CMB power spectrum.

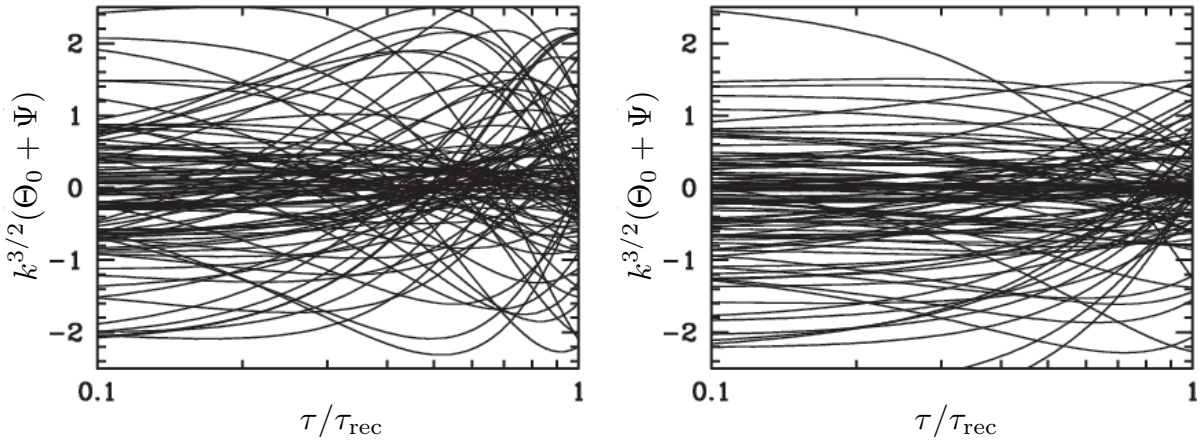


Figure 78: Modes corresponding to the same two wavelengths as in Fig. 77, but this time with random initial phases. We see that the angular peak structure of the CMB would be washed away.

However, we now show that when considering CMB polarization, then even these highly-tuned alternatives to inflation can be ruled out. Looking at Fig. 74 we see that the cross-correlation between CMB temperature fluctuations and the E-mode polarization has a negative peak around $100 < \ell < 200$. This anti-correlation signal is also the result of phase coherence, but now the scales involved were *not* within the horizon at recombination. Hence, there is *no* causal mechanism (after $\tau = 0$) that could have produced this signal. One is almost forced to consider something like inflation with its shrinking comoving horizon leading to horizon exit and re-entry.[36]

As Dodelson explains [35]

> At recombination, [the phase difference between the monopole (sourcing T) and the dipole (sourcing E) of the density field] causes the product of the two to be negative

[36]It should be mentioned here that there are two ways to get a shrinking comoving Hubble radius, $1/(aH)$. During inflation H is nearly constant and the scale factor a grows exponentially. However, in a *contracting* spacetime a shrinking horizon can be achieved if H grows with time. This is the mechanism employed by ekpyrotic/cyclic cosmology [39, 40, 41]. When viewed in terms of the evolution of the comoving Hubble scale inflation and ekpyrosis appear very similar, but there are important differences, *e.g.* in ekyprosis it is a challenge to match the contracting phase to our conventional Big Bang expansion.

for $100 < \ell < 200$ and positive on smaller scales until $\ell \sim 400$. But this is precisely what WMAP has observed! We have clear evidence that the monopole and the dipole were out of phase with each other at recombination.

This evidence is exciting for the small scale modes ($\ell > 200$). Just as the acoustic peaks bear testimony to coherent phases, the cross-correlation of polarization and temperatures speaks to the coherence of the dipole as well. It solidifies our picture of the plasma at recombination. The evidence from the larger scale modes ($\ell < 200$) though is positively stupendous. For, these modes were not within the horizon at recombination. So the *only* way they could have their phases aligned is if some primordial mechanism did the job, when they were in causal contact. Inflation is just such a mechanism.

Scale-Invariant, Gaussian and Adiabatic

We now describe quantitative constraints on the primordial fluctuations. The simplest versions of inflation predict that the scalar perturbations are nearly *scale-invariant*, *Gaussian* and *adiabatic*. In this section we give the latest quantitative constraints on these fundamental predictions of the theory.

Parameter	5-year *WMAP*	*WMAP*+BAO+SN
n_s	$0.963^{+0.014}_{-0.015}$	$0.960^{+0.013}_{-0.013}$
n_s	0.986 ± 0.022	0.970 ± 0.015
r	< 0.43	< 0.22
n_s	$1.031^{+0.054}_{-0.055}$	$1.017^{+0.042}_{-0.043}$
α_s	-0.037 ± 0.028	$-0.028^{+0.020}_{-0.020}$
n_s	$1.087^{+0.072}_{-0.073}$	$1.089^{+0.070}_{-0.068}$
r	< 0.58	< 0.55
α_s	-0.050 ± 0.034	-0.058 ± 0.028

Table 3: 5-year *WMAP* constraints on the primordial power spectra in the power law parameterization [11]. We present results for (n_s), (n_s, r), (n_s, α_s) and (n_s, r, α) marginalized over all other parameters of a flat Λ CDM model.

Spectral Index

As we explained in detail above, observations of the CMB relate to the inflationary spectrum of curvature perturbations $\mathcal{R}$

$$\boxed{C_\ell^{TT}, C_\ell^{TE}, C_\ell^{EE} \quad \Rightarrow \quad P_{\mathcal{R}}(k)}\,. \tag{596}$$

Here, we present the latest quantitative constraints on $P_{\mathcal{R}}(k)$ in the standard power-law parameterization

$$\Delta^2_{\rm s}(k) \equiv \frac{k^3}{2\pi^2} P_{\mathcal{R}}(k) = A_{\rm s} \left(\frac{k}{k_\star}\right)^{n_{\rm s}-1}\,. \tag{597}$$

Measurements of $n_{\rm s}$ are degenerate with the tensor-to-scalar ratio r so constraints on $n_{\rm s}$ are often shown as confidence contours in the $n_{\rm s}$-r plane. The latest WMAP 5-year constraints on the scalar spectral index are shown in Fig. 79 and Table 3. Two facts may be noted: i) the spectrum is nearly scale-invariant, $n_{\rm s} \approx 1$, just as inflation predicts and ii) there are already interesting indications that the spectrum is not perfectly scale-invariant, but slightly red, $n_{\rm s} < 1$. This

deviation from scale-invariance provides the first test of the detailed time-dependence of the inflationary expansion. In fact, as we have seen in **Lecture 2**, inflation predicts this percent level deviation from scale-invariance.[37]

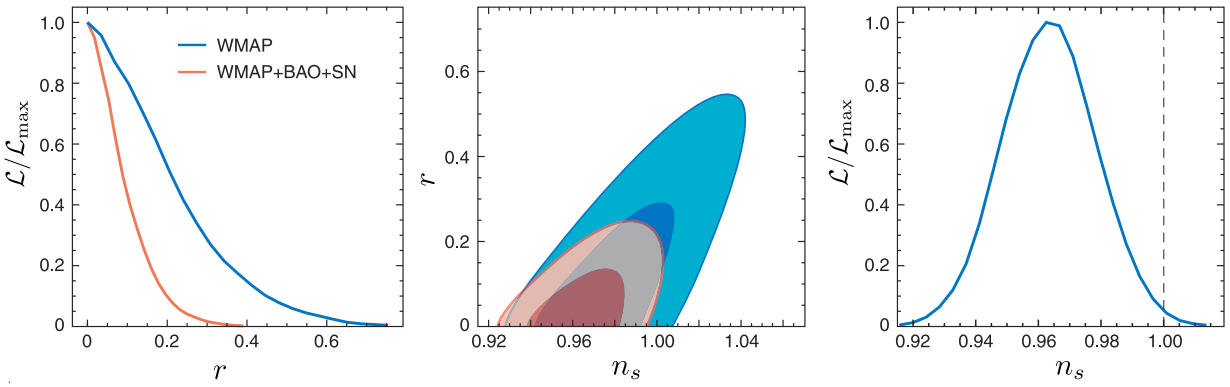


Figure 79: *WMAP* 5-year constraints on the inflationary parameters n_s and r [11]. The *WMAP*-only results are shown in blue, while constraints from *WMAP* plus other cosmological observations are in red. The third plot assumes that r is negligible.

Gaussianity

If $\mathcal{R}$ is Gaussian then the power spectrum $P_{\mathcal{R}}(k)$ (two-point correlations in real space) is the end of the story. However, if $\mathcal{R}$ is *non-Gaussian* then the fluctuations have a non-zero bispectrum $B_{\mathcal{R}}(k_1, k_2, k_3)$ (corresponding to three-point correlations in real space). There is only one way to be Gaussian but many ways to be non-Gaussian, so constraints on non-Gaussianity are a bit hard to describe. One of the simplest forms of non-Gaussianity is described by the parameterization

$$\mathcal{R}(\mathbf{x}) = \mathcal{R}_g(\mathbf{x}) + \frac{3}{5} f_{\rm NL} \star \mathcal{R}_g^2(\mathbf{x})\,. \tag{598}$$

where $\mathcal{R}_g$ is Gaussian. In this *local model* for non-Gaussianity, the information is reduced to a single number $f_{\rm NL}$. The latest constraint on $f_{\rm NL}$ by Smith, Senatore, and Zaldarriaga [42] is

$$-4 \;<\; f_{\rm NL} \;<\; 80 \quad \text{at} \quad 95\%\ \text{CL}\,. \tag{599}$$

Notice that an $f_{\rm NL}$ value of order 100 corresponds to a 0.1% correction to $\mathcal{R}_g \sim 10^{-5}$ in Eq. (598). The constraint (599) therefore implies that the CMB is Gaussian to 0.1%! This is better than our constraint on the curvature of the universe which is usually celebrated as the triumph of inflation. The CMB is highly Gaussian and it didn't have to be that way. However, if inflation is correct then the observed Gaussianity is a rather natural consequence.[38]

Adiabaticity

In single-field inflation, the fluctuations of the inflaton field on large scales (where spatial gradients can be neglected) can be identified with a local shift backwards or forwards along the the trajectory of the homogeneous background field. These shifts along the inflaton trajectory affect the total density in different parts of the universe after inflation, but cannot give rise

[37] For inflation to end, the Hubble parameter H has to change in time. This time-dependence changes the conditions at the time when each fluctuation mode exits the horizon and therefore gets translated into a scale-dependence of the fluctuations.

[38] Non-Gaussianity is a measure of *interactions* of the inflaton field. However, for slow-roll dynamics to occur, the inflaton has to be very weakly self-interacting (the potential is very flat) and the non-Gaussianity is necessarily small, $f_{\rm NL} \sim \mathcal{O}(0.01)$ [22].

to variations in the relative density between different components. Hence, single-field inflation produces purely *adiabatic* primordial density perturbations characterized by an overall curvature perturbations, $\mathcal{R}$. This means that all perturbations of the cosmological fluid (photons, neutrinos, baryons and cold dark matter (CDM) particles) originate from the same curvature perturbation $\mathcal{R}$ and satisfy the adiabaticity property, $\delta(n_m/n_r) = 0$, or

$$\frac{\delta\rho_m}{\rho_m} = \frac{3}{4}\frac{\delta\rho_r}{\rho_r}\,, \tag{600}$$

where the index m collectively stands for non-relativistic species (baryons or CDM) and r for relativistic species (photons or neutrinos). The latest data shows no violation of the condition (600) [11]. If such a violation were to be found this would be a clear signature of multi-field inflation (see §7.5.5).

Testing Slow-Roll Models

In Figure 80 we present current observational constraints on some of the simplest single-field slow-roll models of inflation (see **Lecture 1**). Future measurements will significantly cut into the parameter space of allowed models.

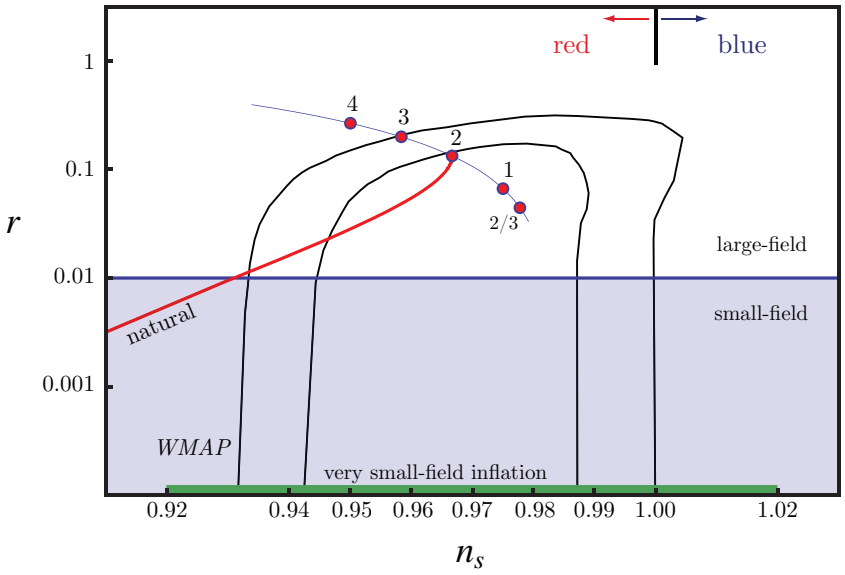


Figure 80: Constraints on single-field slow-roll models in the n_s-r plane. The value of r determines whether the models involve large or small field variations. The value of n_s classifies the scalar spectrum as red or blue. Combinations of the values of r and n_s determine whether the curvature of the potential was positive ($\eta_{\rm v} > 0$) or negative ($\eta_{\rm v} < 0$) when the observable universe exited the horizon. Also shown are the *WMAP* 5-year constraints on n_s and r [11] as well as the predictions of a few representative models of single-field slow-roll inflation: *chaotic inflation*: $\lambda_p\,\phi^p$, for general p (thin solid line) and for $p = 4, 3, 2, 1, \frac{2}{3}(\bullet)$; models with $p = 2$ [43], $p = 1$ [44] and $p = \frac{2}{3}$ [251] have recently been obtained in string theory; *natural inflation*: $V_0[1-\cos(\phi/\mu)]$ (solid line); *very small-field inflation*: models of inflation with a very small tensor amplitude, $r \ll 10^{-4}$ (green bar); examples of such models in string theory include warped D-brane inflation [46, 47, 48], Kähler inflation [49], and racetrack inflation [50].

7.5.5 Future Tests of Inflation

We are only at the beginning of really testing the inflationary paradigm. The flatness of the universe and the near scale-invariance, Gaussianity and adiabaticity of the density fluctuations

are encouraging evidence for inflation[39], but they are not proof that inflation really occurred.[40] Let us therefore look into the future and describe how future experiments can provide further tests of inflationary physics.

Amplitude of Tensor Modes

Probably the single most important piece of evidence for inflation would come from a measurement of a primordial tensor amplitude. We showed above that a detection of primordial CMB B-modes would be virtually impossible to explain by anything other inflationary gravitational waves [51]

$$\boxed{C_\ell^{BB} \quad \Rightarrow \quad P_h(k)}\,, \tag{601}$$

where

$$\Delta_{\rm t}^2(k) \equiv \frac{k^3}{2\pi^2} P_h(k) = A_{\rm t} \left(\frac{k}{k_\star}\right)^{n_{\rm t}} . \tag{602}$$

We also explained in **Lecture 2** that the tensor amplitude A_t is directly linked with the energy scale of inflation. As a single clue about the physics of inflation, what could be more important and higher on the wish-list of inflationary theorists? In addition, a detection of tensor modes would imply that the inflaton field moved over a super-Planckian distance in field space, making string theorists and quantum gravity affectionatos think hard about Planck-suppressed corrections to the inflaton potential (see **Lecture 5**).

Scale Dependence of Scalar Modes

The variation of the spectral index $n_{\rm s}$ with scale (also called the "running" of the spectral index) arises only at second-order in slow-roll and is therefore expected to be small

$$\alpha_{\rm s} \equiv \frac{dn_{\rm s}}{d\ln k} \sim \mathcal{O}(\varepsilon^2)\,. \tag{603}$$

On the other hand, an unexpected large positive or negative running would force us to rethink some of our simplest notions about inflation and the generation of perturbations.

Scale Dependence of Tensor Modes

Measuring the amplitude of primordial tensor fluctuations from inflation will be a significant observational challenge. Hoping to measure its dependence on scale seems unrealistic unless the tensor amplitude is near its current upper limit. In single-field slow-roll models the tensor-to-scalar ratio r and the tensor spectral index $n_{\rm t}$ are related by the *consistency relation*

$$r = -8n_{\rm t}\,. \tag{604}$$

Measuring (604) would offer another way to falsify single-field slow-roll inflation.

Non-Gaussianity

The primordial fluctuations are to a high degree Gaussian. However, as we now describe, even a small non-Gaussianity would encode a tremendous amount of information about the inflationary action. We mentioned that the three-point function of inflationary fluctuations is

[39] Note that at any stage we could have made measurements that would have falsified the whole idea of inflation.

[40] I know, I used the word 'proof' when Karl Popper taught us that we can never prove a theory.

the prime diagnostic of non-Gaussian statistics. In momentum space, the three-point correlation function can be written generically as

$$\langle \mathcal{R}_{\mathbf{k}_1} \mathcal{R}_{\mathbf{k}_2} \mathcal{R}_{\mathbf{k}_3} \rangle = (2\pi)^3 \, \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \; f_{\rm NL} \; F(k_1, k_2, k_3) \, . \tag{605}$$

Here, $f_{\rm NL}$ is a dimensionless parameter defining the amplitude of non-Gaussianity, while the function $F(k_1, k_2, k_3)$ captures the momentum dependence. The amplitude and sign of $f_{\rm NL}$, as well as the shape and scale dependence of $F(k_1, k_2, k_3)$, depend on the details of the interaction generating the non-Gaussianity, making the three-point function a powerful discriminating tool for probing models of the early universe [30].

Two simple and distinct shapes $F)k_1, k_2, k_3)$ are generated by two very different mechanisms [52]: The *local shape* is a characteristic of multi-field models and takes its name from the expression for the primordial curvature perturbation $\mathcal{R}$ in real space,

$$\mathcal{R}(\mathbf{x}) = \mathcal{R}_g(\mathbf{x}) + \frac{3}{5} f_{\rm NL}^{\rm local} \mathcal{R}_g(\mathbf{x})^2 \, , \tag{606}$$

where $\mathcal{R}_g(\mathbf{x})$ is a Gaussian random field. Fourier transforming this expression shows that the signal is concentrated in "squeezed" triangles where $k_3 \ll k_1, k_2$. Local non-Gaussianity arises in multi-field models where the fluctuations of an isocurvature field (see below) are converted into curvature perturbations. As this conversion happens outside of the horizon, when gradients are irrelevant, one generates non-linearities of the form (606). Specific models of this type include multi-field inflation [53, 54, 55, 56, 57, 58, 52, 60, 61, 62, 63, 64, 65], the curvaton scenario [173, 67], inhomogeneous reheating [68, 69], and New Ekpyrotic models [70, 71, 72, 73, 74, 75, 76].

The second important shape is called *equilateral* as it is largest for configurations with $k_1 \sim k_2 \sim k_3$. The equilateral form is generated by single-field models with non-canonical kinetic terms such as DBI inflation [77], ghost inflation [78, 79] and more general models with small sound speed [80, 81].

Isocurvature Fluctuations

In inflationary models with more than one field the perturbations are not necessarily adiabatic. With more than one field, fluctuations orthogonal to the background trajectory can affect the relative density between different matter components even if the total density and therefore the spatial curvature is unperturbed [111]. There are various different possibilities for such *isocurvature* perturbations (also called non-adiabitic or entropic perturbations), *e.g.* we may define relative perturbations between CDM and photons

$$\mathcal{S}_m \equiv \frac{\delta\rho_m}{\rho_m} - \frac{3}{4} \frac{\delta\rho_\gamma}{\rho_\gamma} \, . \tag{607}$$

Adiabatic and isocurvature perturbations lead to a different peak structure in the CMB fluctuations. CMB measurements can therefore distinguish between the different types of fluctuations and in fact already show that isocurvature perturbations have to be a subdominant component (if at all present).

Isocurvature perturbations could be correlated with the adiabatic perturbations. To capture this we define the following correlation parameter

$$\beta \equiv \frac{P_{\mathcal{S}\mathcal{R}}}{\sqrt{P_{\mathcal{S}} P_{\mathcal{R}}}} \, , \tag{608}$$

where $P_{\mathcal{R}}$ and $P_{\mathcal{S}}$ are the power spectra of adiabatic and isocurvature fluctuations and $P_{\mathcal{S}\mathcal{R}}$ is their cross-correlation. Parameterizing the relative amplitude between the two types of perturbations by a coefficient α

$$\frac{P_{\mathcal{S}}}{P_{\mathcal{R}}} \equiv \frac{\alpha}{1-\alpha} \, , \tag{609}$$

the present constraints on the isocurvature contribution are $\alpha_0 < 0.067$ (96% CL) in the uncorrelated case ($\beta = 0$) and $\alpha_{-1} < 0.0037$ (95% CL) in the totally anti-correlated case ($\beta = -1$).

Theoretical predictions for the amplitude of isocurvature perturbations are complicated by the fact that they are strongly model-dependent: the isocurvature amplitude does not depend entirely on the multi-field inflationary dynamics, but also on the post-inflationary evolution. If all particle species are in thermal equilibrium after inflation and their local densities are uniquely given by their temperature (with vanishing chemical potential) then the primordial perturbations are adiabatic [176, 84]. Thus, it is important to note that the existence of primordial isocurvature modes requires at least one field to decay into some species whose abundance is not determined by thermal equilibrium (*e.g.* CDM after decoupling) or respects some conserved quantum numbers, like baryon or lepton numbers.

7.6 Primordial Non-Gaussianity by Baumann 4th lecture

Abstract

In this lecture we summarize key theoretical results in the study of primordial non-Gaussianity. Most results are stated without proof, but their significance for constraining the fundamental physical origin of inflation is explained. After introducing the bispectrum as a basic diagnostic of non-Gaussian statistics, we show that its momentum dependence is a powerful probe of the inflationary action. Large non-Gaussianity can only arise if inflaton interactions are significant during inflation. In single-field slow-roll inflation non-Gaussianity is therefore predicted to be unobservably small, while it can be significant in models with multiple fields, higher-derivative interactions or non-standard initial states. Finally, we end the lecture with a discussion of the observational prospects for detecting or constraining primordial non-Gaussianity.

7.6.1 Preliminaries

Non-Gaussianity, *i.e.* the study of non-Gaussian contributions to the correlations of cosmological fluctuations, is emerging as an important probe of the early universe [85]. Being a direct measure of inflaton *interactions*, constraints on primordial non-Gaussianities will teach us a great deal about the inflationary dynamics. It also puts strong constraints on alternatives to the inflationary paradigm [70, 71, 72, 73, 74, 75, 76].

In **Lecture 2** we expanded the inflationary action to second order in the comoving curvature perturbation $\mathcal{R}$. This free-field action allowed us to compute the power spectrum $P_{\mathcal{R}}(k)$. As we mentioned in **Lecture 3**, if the fluctuations $\mathcal{R}$ are drawn from a Gaussian distribution, then the power spectrum (or two-point correlation function) contains all the information.[41] However, for non-Gaussian fluctuations higher-order correlation functions beyond the two-point function contain additional information about inflation. Computing the leading non-Gaussian effects requires expansion of the action to third order in order to capture the leading non-trivial interaction terms. These computations can be algebraically quite challenging, so we will limit this lecture to a review of the main results and their physical interpretations. For more details and derivations we refer the reader to the comprehensive review by Bartolo *et al.* [30] and the references cited therein.

[41] The three-point function and all odd higher-point correlation functions vanish for Gaussian fluctuations, while all even higher-point functions can be expressed in terms of the two-point function. In other words, all *connected* higher-point functions vanish for Gaussian fluctuations.

The Bispectrum and Local Non-Gaussianity

Bispectrum

The Fourier transform of the two-point function is the power spectrum

$$\langle \mathcal{R}_{\mathbf{k}_1} \mathcal{R}_{\mathbf{k}_2} \rangle = (2\pi)^3 \delta(\mathbf{k}_1 + \mathbf{k}_2) P_{\mathcal{R}}(k_1) \,. \tag{610}$$

Similarly, the Fourier equivalent of the three-point function is the *bispectrum*

$$\boxed{\langle \mathcal{R}_{\mathbf{k}_1} \mathcal{R}_{\mathbf{k}_2} \mathcal{R}_{\mathbf{k}_3} \rangle = (2\pi)^3 \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_{\mathcal{R}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)} \,. \tag{611}$$

Here, the delta function (enforcing momentum conservation) is a consequence of translation invariance of the background. The function $B_{\mathcal{R}}$ is symmetric in its arguments and for scale-invariant fluctuations it is a homogeneous function of degree -6

$$B_{\mathcal{R}})\lambda \mathbf{k}_1, \lambda \mathbf{k}_2, \lambda \mathbf{k}_3) = \lambda^{-6} B_{\mathcal{R}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \,. \tag{612}$$

Rotational invariance further reduces the number of independent variables to just two, *e.g.* the two ratios k_2/k_1 and k_3/k_1.

To compute the three-point function for a specific inflationary model requires a careful treatment of the time-evolution of the vacuum in the presence of interactions (while for the two-point function this effect is higher-order). In Appendix C we describe the "in-in" formalism for computing cosmological correlation functions [86, 87, 88, 22, 89]. In practice, computing three-point functions can be algebraically very cumbersome, so in the lecture we restrict us to citing the final results. The details on how to compute these three-point functions deserves a review of its own.

Local Non-Gaussianity

One of the first ways to parameterize non-Gaussianity phenomenologically was via a non-linear correction to a Gaussian perturbation $\mathcal{R}_g$ [90],

$$\boxed{\mathcal{R}(\mathbf{x}) = \mathcal{R}_g(\mathbf{x}) + \frac{3}{5} f_{\rm NL}^{\rm local} \left[\mathcal{R}_g(\mathbf{x})^2 - \langle \mathcal{R}_g(\mathbf{x})^2 \rangle \right]} \,. \tag{613}$$

This definition is local in real space and therefore called *local non-Gaussianity*. Experimental constraints on non-Gaussianity (see **Lecture 3**) are often set on the parameter $f_{\rm NL}^{\rm local}$ defined via Eqn. (613).[42] Using Eqn. (613) the bispectrum of local non-Gaussianity may be derived

$$B_{\mathcal{R}}(k_1, k_2, k_3) = \frac{6}{5} f_{\rm NL}^{\rm local} \times \left[P_{\mathcal{R}}(k_1) P_{\mathcal{R}}(k_2) + P_{\mathcal{R}}(k_2) P_{\mathcal{R}}(k_3) + P_{\mathcal{R}}(k_3) P_{\mathcal{R}}(k_1) \right] \,. \tag{614}$$

Exercise 10 (Local Bispectrum)**.** *Derive Eqn. (614) from Eqns. (611) and (613).*

For a scale-invariant spectrum, $P_{\mathcal{R}}(k) = A k^{-3}$, this is

$$B_{\mathcal{R}}(k_1, k_2, k_3) = \frac{6}{5} f_{\rm NL}^{\rm local} \times A^2 \left[\frac{1}{(k_1 k_2)^3} + \frac{1}{(k_2 k_3)^3} + \frac{1}{(k_3 k_1)^3} \right] \,. \tag{615}$$

Without loss of generality, let us order the momenta such that $k_3 \leq k_2 \leq k_1$. The bispectrum for local non-Gaussianity is then largest when the smallest k (*i.e.* k_3) is very small, $k_3 \ll k_1 \sim k_2$. The other two momenta are then nearly equal. In this *squeezed* limit, the bispectrum for local non-Gaussianity becomes

$$\lim_{k_3 \ll k_1 \sim k_2} B_{\mathcal{R}}(k_1, k_2, k_3) = \frac{12}{5} f_{\rm NL}^{\rm local} \times P_{\mathcal{R}}(k_1) P_{\mathcal{R}}(k_3) \,. \tag{616}$$

[42] The factor of $3/5$ in Eqn. (613) is conventional since non-Gaussianity was first defined in terms of the Newtonian potential, $\Phi(\mathbf{x}) = \Phi_g(\mathbf{x}) + f_{\rm NL}^{\rm local} \left[\Phi_g(\mathbf{x})^2 - \langle \Phi_g(\mathbf{x})^2 \rangle \right]$, which during the matter era is related to $\mathcal{R}$ by a factor of $3/5$.

Shapes of Non-Gaussianity

The delta function in Eqn. (611) enforces that the three Fourier modes of the bispectrum form a closed triangle. Different inflationary models predict maximal signal for different triangle configurations. This *shape of non-Gaussianity* [52] is potentially a powerful probe of the mechanism that laid down the primordial perturbations.

It will be convenient to define the shape function

$$\mathcal{S}(k_1, k_2, k_3) \equiv N(k_1 k_2 k_3)^2 B_{\mathcal{R}}(k_1, k_2, k_3)\,, \tag{617}$$

where N is an appropriate normalization factor. Two commonly discussed shapes are the *local* model, *cf.* Eqn. (615),

$$\mathcal{S}^{\rm local}(k_1, k_2, k_3) \propto \frac{K_3}{K_{111}}\,, \tag{618}$$

and the *equilateral* model,

$$\mathcal{S}^{\rm equil}(k_1, k_2, k_3) \propto \frac{\tilde{k}_1 \tilde{k}_2 \tilde{k}_3}{K_{111}}\,. \tag{619}$$

Here, we have introduced a notation first defined by Fergusson and Shellard [87],

$$K_p = \sum_i (k_i)^p \quad \text{with} \quad K = K_1 \tag{620}$$

$$K_{pq} = \frac{1}{\Delta_{pq}} \sum_{i \neq j} (k_i)^p (k_j)^q \tag{621}$$

$$K_{pqr} = \frac{1}{\Delta_{pqr}} \sum_{i \neq j \neq l} (k_i)^p (k_j)^q (k_l)^q \tag{622}$$

$$\tilde{k}_{ip} = K_p - 2(k_i)^p \quad \text{with} \quad \tilde{k}_i = \tilde{k}_{i1}\,, \tag{623}$$

where $\Delta_{pq} = 1 + \delta_{pq}$ and $\Delta_{pqr} = \Delta_{pq}(\Delta_{qr} + \delta_{pr})$ (no summation). This notation significantly compresses the increasingly complex expressions for the bispectra discussed in the literature.

We have argued above that for scale-invariant fluctuations the bispectrum is only a function of the two ratios k_2/k_1 and k_3/k_1. We hence define the rescaled momenta

$$x_i \equiv \frac{k_i}{k_1}\,. \tag{624}$$

We have ordered the momenta such that $x_3 \leq x_2 \leq 1$. The triangle inequality implies $x_2 + x_3 > 1$. In the following we plot $\mathcal{S}(1, x_2, x_3)$ (see Figs. 81, 83, and 84). We use the normalization, $\mathcal{S}(1, 1, 1) \equiv 1$. To avoid showing equivalent configurations twice $\mathcal{S}(1, x_2, x_3)$ is set to zero outside the triangular region $1 - x_2 \leq x_3 \leq x_2$. We see in Fig. 81 that the signal for the local shape is concentrated at $x_3 \approx 0$, $x_2 \approx 1$, while the equilateral shape peaks at $x_2 \approx x_3 \approx 1$. Fig. 82 illustrates how the different triangle shapes are distributed in the x_2-x_3 plane.

Physically motivated models for producing non-Gaussian perturbations often produce signals that peak at special triangle configurations. Three important special cases are:

i) *squeezed triangle* ($k_1 \approx k_2 \gg k_3$)

This is the dominant mode of models with multiple light fields during inflation [53, 54, 55, 56, 57, 58, 52, 60, 61, 62, 63, 64, 65], the curvaton scenario [173, 67], inhomogeneous reheating [68, 69], and New Ekpyrotic models [70, 71, 72, 73, 74, 75, 76].

ii) *equilateral triangle* ($k_1 = k_2 = k_3$)

Signals that peak at equilateral triangles arise in models with higher-derivative interactions and non-trivial speeds of sound [92, 81].

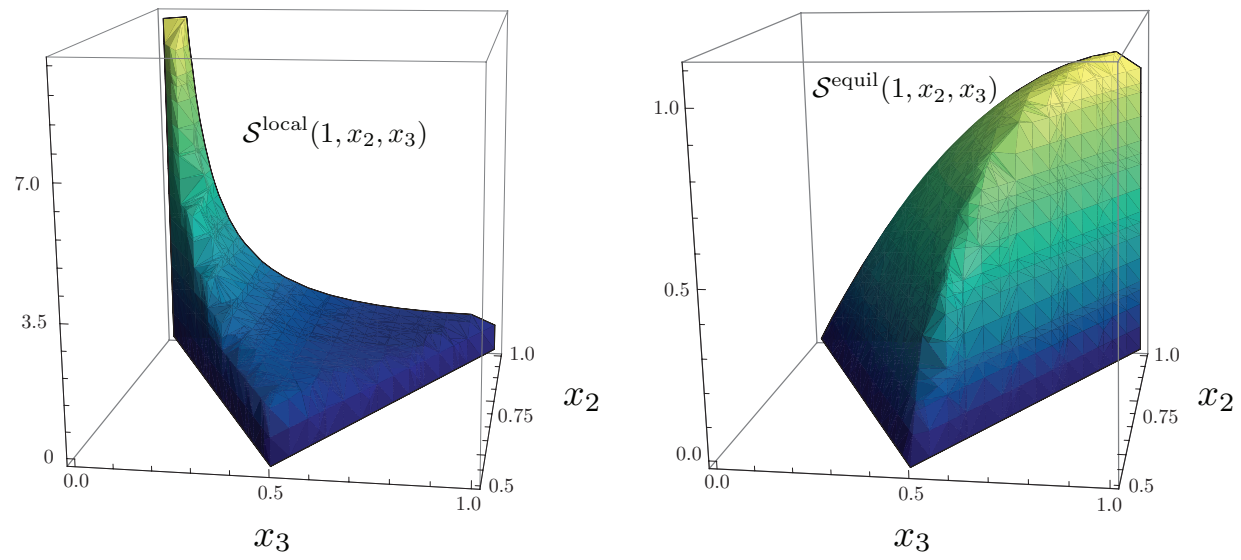


Figure 81: 3D plots of the *local* and *equilateral* bispectra. The coordinates x_2 and x_3 are the rescaled momenta k_2/k_1 and k_3/k_1, respectively. Momenta are order such that $x_3 < x_2 < 1$ and satsify the triangle inequality $x_2 + x_3 > 1$.

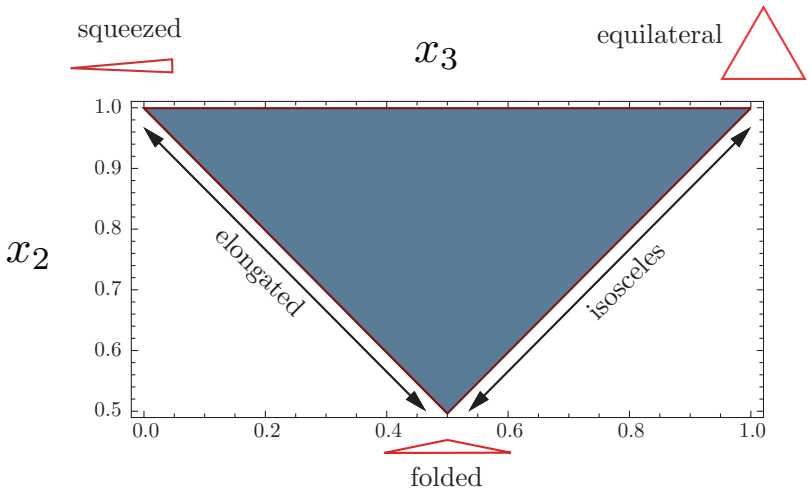


Figure 82: Shapes of Non-Gaussianity. The coordinates x_2 and x_3 are the rescaled momenta k_2/k_1 and k_3/k_1, respectively. Momenta are order such that $x_3 < x_2 < 1$ and satsify the triangle inequality $x_2 + x_3 > 1$.

iii) *folded triangle* ($k_1 = 2k_2 = 2k_3$)

Folded triangles arise in models with non-standard initial states [93, 81].

In addition, there are the intermediate cases: *elongated triangles* ($k_1 = k_2 + k_3$) and *isosceles triangles* ($k_1 > k_2 = k_3$).

$f_{\rm NL}$: The Amplitude of Non-Gaussianity

For arbitrary shape functions we measure the magnitude of non-Gaussianity by defining the generalized $f_{\rm NL}$ parameter

$$\boxed{f_{\rm NL} \equiv \frac{5}{18}\frac{B_{\mathcal{R}}(k,k,k)}{P_{\mathcal{R}}(k)^2}}\,. \tag{625}$$

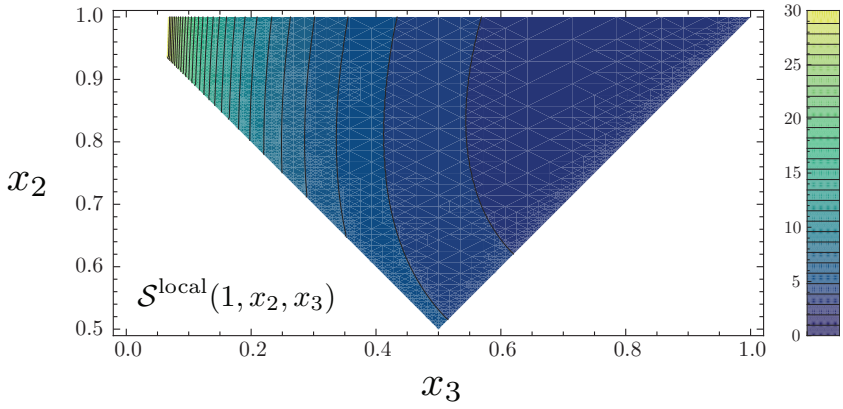


Figure 83: Contour plot of the *local* bispectrum.

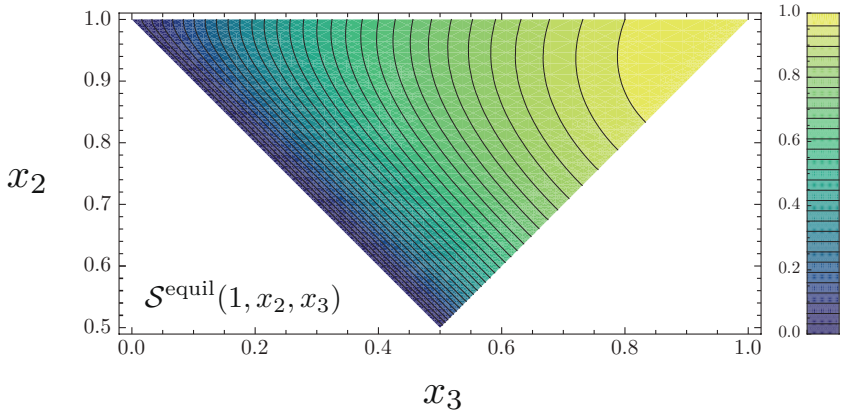


Figure 84: Contour plot of the *equilateral* bispectrum.

In this definition the amplitude of non-Gaussianity is normalized in the equilateral configuration.

Exercise 11 ($f_{\rm NL}$)**.** *Show from Eqn. (614) that the definition (625) is consistent with our definition of* $f_{\rm NL}^{\rm local}$*, Eqn. (613).*

7.6.2 Theoretical Expectations

Single-Field Slow-Roll Inflation

Successful slow-roll inflation demands that the interactions of the inflaton field are weak. Since the wave function of free fields in the ground state is Gaussian, the fluctuations created during slow-roll inflation are expected to be Gaussian. Maldacena [22] first derived the bispectrum for slow-roll (SR) inflation

$$\mathcal{S}^{\rm SR}(k_1, k_2, k_3) \quad \propto \quad)\varepsilon - 2\eta)\frac{K_3}{K_{111}} + \varepsilon\left(K_{12} + 8\frac{K_{22}}{K}\right) \tag{626}$$

$$\approx \quad (4\varepsilon - 2\eta)\,\mathcal{S}^{\rm local}(k_1, k_2, k_3) + \frac{5}{3}\varepsilon\,\mathcal{S}^{\rm equil}(k_1, k_2, k_3)\,, \tag{627}$$

where $\mathcal{S}^{\rm local}$ and $\mathcal{S}^{\rm equil}$ are normalized so that $\mathcal{S}^{\rm local}(k, k, k) = \mathcal{S}^{\rm equil}(k, k, k)$. The bispectrum for slow-roll inflation peaks at squeezed triangles and has an amplitude that is suppressed by slow-roll parameters [22]

$$f_{\rm NL}^{\rm SR} = \mathcal{O}(\varepsilon, \eta)\,. \tag{628}$$

This makes intuitive sense since the slow-roll parameters characterize deviations of the inflaton from a free field.

The Maldacena Theorem

Under the assumption of single-field inflation, but *no* other assumptions about the inflationary action, Creminelli and Zaldarriaga [94] were able to prove a powerful theorem:

$$\boxed{\lim_{k_3\to 0}\langle \mathcal{R}_{\mathbf{k}_1}\mathcal{R}_{\mathbf{k}_2}\mathcal{R}_{\mathbf{k}_3}\rangle = (2\pi)^3\delta(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\,\underline{\underline{(1-n_{\rm s})}}\,P_{\mathcal{R}}(k_1)P_{\mathcal{R}}(k_3)}\,, \tag{629}$$

where

$$\langle \mathcal{R}_{\mathbf{k}_i}\mathcal{R}_{\mathbf{k}_j}\rangle = (2\pi)^3\delta(\mathbf{k}_i+\mathbf{k}_j)P_{\mathcal{R}}(k_i)\,. \tag{630}$$

Eqn. (629) states that for single-field inflation, the squeezed limit of the three-point function is suppressed by $(1-n_{\rm s})$ and vanishes for perfectly scale-invariant perturbations. A detection of non-Gaussianity in the squeezed limit can therefore rule out single-field inflation! In particular, this statement is independent of: the form of the potential, the form of the kinetic term (or sound speed) and the initial vacuum state.

Proof:

The squeezed triangle correlates one long-wavelength mode, $k_{\rm L}=k_3$ to two short-wavelength modes, $k_{\rm S}=k_1\approx k_2$,

$$\langle \mathcal{R}_{\mathbf{k}_1}\mathcal{R}_{\mathbf{k}_2}\mathcal{R}_{\mathbf{k}_3}\rangle \approx \langle (\mathcal{R}_{\mathbf{k}_{\rm S}})^2\mathcal{R}_{\mathbf{k}_{\rm L}}\rangle\,. \tag{631}$$

Modes with longer wavelengths freeze earlier. Therefore, $k_{\rm L}$ will be already frozen outside the horizon when the two smaller modes freeze and acts as a background field for the two short-wavelength modes.

Why should $(\mathcal{R}_{\mathbf{k}_{\rm S}})^2$ be correlated with $\mathcal{R}_{\mathbf{k}_{\rm L}}$? The theorem says that "it isn't correlated if $\mathcal{R}_{\rm k}$ is precisely scale-invariant". The proof is simplest in real-space (see Creminelli and Zaldarriaga [94]): The long-wavelength curvature perturbation $\mathcal{R}_{\mathbf{k}_{\rm L}}$ rescales the spatial coordinates (or changes the effective scale factor) within a given Hubble patch

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a(t)^2e^{2\mathcal{R}}\mathrm{d}\mathbf{x}^2\,. \tag{632}$$

The two-point function $\langle \mathcal{R}_{\mathbf{k}_1}\mathcal{R}_{\mathbf{k}_2}\rangle$ will depend on the value of the background fluctuations $\mathcal{R}_{\mathbf{k}_{\rm L}}$ already frozen outside the horizon. In position space the variation of the two-point function given by the long-wavelength fluctuations $\mathcal{R}_{\rm L}$ is at linear order

$$\frac{\partial}{\partial\mathcal{R}_{\rm L}}\langle\mathcal{R}(x)\mathcal{R}(0)\rangle\cdot\mathcal{R}_{\rm L} = x\frac{d}{dx}\langle\mathcal{R}(x)\mathcal{R}(0)\rangle\cdot\mathcal{R}_{\rm L}\,. \tag{633}$$

To get the three-point function Creminelli and Zaldarriaga multiply Eqn. (633) by $\mathcal{R}_{\rm L}$ and average over it. Going to Fourier space gives Eqn. (629).[43] QED.

Models with Large Non-Gaussianity

Higher-Derivative Interactions

Although Maldacena proved that for single-field *slow-roll* inflation non-Gaussianity is always small, single-field models can still give large non-Gaussianity if higher-derivative terms are important during inflation (as opposed to assuming a canonical kinetic term and no higher-derivative corrections as in slow-roll inflation). Consider the following action

$$S = \frac{1}{2}\int \mathrm{d}^4x\sqrt{-g}\,[R-P(X,\phi)]\;,\quad \text{where}\quad X\equiv(\partial_\mu\phi)^2\,. \tag{634}$$

[43] For more details see Cheung *et al.* [95].

Here, $P(X,\phi)$ is an arbitrary function of the kinetic term $X = (\partial_\mu\phi)^2$ and hence can contain higher-derivative interactions. These models in general have a non-trivial sound speed for the propagation of fluctuations

$$c_s^2 \equiv \frac{P_{,X}}{P_{,X} + 2XP_{,XX}}\,. \tag{635}$$

The second-order action for $\mathcal{R}$ (giving $P_\mathcal{R}$) is [81]

$$S_{(2)} = \int \mathrm{d}^4x\, \varepsilon \left[a^3(\dot{\mathcal{R}})^2/\underline{\underline{c_s^2}} - a(\partial_i\mathcal{R})^2\right] + \mathcal{O}(\varepsilon^2) \tag{636}$$

The third-order action for $\mathcal{R}$ (giving $B_\mathcal{R}$; see Appendix 7.8.3 and Ref. [81]) is

$$S_{(3)} = \int \mathrm{d}^4x\, \underline{\underline{\varepsilon^2}} \left[\ldots a^3(\dot{\mathcal{R}})^2\mathcal{R}/\underline{\underline{c_s^2}} + \ldots a(\partial_i\mathcal{R})^2\mathcal{R} + \ldots a^3(\dot{\mathcal{R}})^3/\underline{\underline{c_s^2}}\right] + \mathcal{O}(\varepsilon^3) \tag{637}$$

We notice that the third-order action is surpressed by an extra factor of ε relative to the second-order action. This is a reflection of the fact that non-Gaussianity is small in the slow-roll limit: $P(X,\phi) = X - V(\phi)$, $c_s^2 = 1$. However, away from the slow-roll limit, for small sound speeds, $c_s^2 \ll 1$, a few interaction terms in Eqn. (637) get boosted and non-Gaussianity can become significant. The signal is peaked at equilateral triangles, with

$$f_{\rm NL}^{\rm equil} = -\frac{35}{108}\left(\frac{1}{c_s^2} - 1\right) + \frac{5}{81}\left(\frac{1}{c_s^2} - 1 - 2\Lambda\right)\,, \tag{638}$$

where

$$\Lambda \equiv \frac{X^2P_{,XX} + \frac{2}{3}X^3P_{,XXX}}{XP_{,X} + 2X^2P_{,XX}}\,. \tag{639}$$

Whether actions with arbitrary $P(X,\phi)$ exist in consistent high-energy theories is an important challenge for these models. It is encouraging that one of the most interesting models of inflation in string theory, DBI inflation [92] (see **Lecture 5**), has precisely such a structure with

$$P_{\rm DBI}(X,\phi) = -f^{-1}(\phi)\sqrt{1 - 2f(\phi)X} + f^{-1}(\phi) - V(\phi)\,. \tag{640}$$

In this case, the second term in Eqn. (641) is identically zero and we find

$$f_{\rm NL}^{\rm DBI} = -\frac{35}{108}\left(\frac{1}{c_s^2} - 1\right)\,. \tag{641}$$

The shape function for DBI inflation is

$$\mathcal{S}^{\rm DBI}(k_1,k_2,k_3) \propto \frac{1}{K_{111}K^2}(K_5 + 2K_{14} - 3K_{23} + 2K_{113} - 8K_{122})\,. \tag{642}$$

Multiple Fields

In single-field slow-roll inflation interactions of the inflaton are constrained by the requirement that inflation should occur. However, if more than one field was relevant during inflation this constraint may be circumvented. Models like the *curvaton mechanism* [173, 67] or *inhomogeneous reheating* [68, 69] exploit this to create non-Gaussian fluctuations via fluctuations is a second field that is not the inflaton. The signal is peaked at squeezed triangles.

For more details on these mechanisms to produce local-type non-Gaussianity we refer the reader to the review by Bartolo *et al.* [30].

Non-Standard Vacuum

If inflation started in an excited state rather than in the Bunch-Davies vacuum, remnant non-Gaussianity may be observable [93] (unless inflation lasted much more than the minimal number of e-folds, in which case the effect is exponentially diluted). The signal is peaked at folded triangles with a shape function

$$\mathcal{S}^{\rm folded}(k_1,k_2,k_3) \propto \frac{1}{K_{111}}(K_{12}-K_3)+4\frac{K_2}{(\tilde{k}_1\tilde{k}_2\tilde{k}_3)^2}\,. \tag{643}$$

For a more detailed discussion of this effect the reader may consult the paper by Holman and Tolley [93].

7.6.3 Observational Prospects

Observational constraints on primordial non-Gaussianity are beginning to reach interesting levels. Precision CMB experiments now probe the regime of parameter space where some inflationary models [173, 67, 68, 69] and most models of New Ekpyrosis [70, 71, 72, 73, 74, 75, 76] predict a signal.

Cosmic Microwave Background

The latest constraint on $f_{\rm NL}^{\rm local}$ and $f_{\rm NL}^{\rm equil}$ by Smith, Senatore, and Zaldarriaga [42, 96] are

$$-4 \;<\; f_{\rm NL}^{\rm local} \;<\; +80 \quad \text{at} \;\; 95\%\ \text{CL}\,, \tag{644}$$

$$-125 \;<\; f_{\rm NL}^{\rm equil} \;<\; +435 \quad \text{at} \;\; 95\%\ \text{CL}\,. \tag{645}$$

The Planck satellite and the proposed CMBPol mission are projected to give $\sigma(f_{\rm NL}^{\rm local})\sim 5$ and $\sigma(f_{\rm NL}^{\rm local})\sim 2$, respectively. At the level of $f_{\rm NL}\sim\mathcal{O}(1)$ we, in fact, expect to see a signal from secondary effects not associated with inflation. In order, not to confuse these effects with the primordial signal, one needs to compute in detail how the non-linear evolution of fluctuations can induce its own non-Gaussianity. To date, the effects haven't been fully computed (but see, *e.g.* Refs.[97, 98, 99]). Often only their order of magnitude is estimated. A systematic characterization of all effects inducing observable levels of non-Gaussianity is clearly timely.

Large-Scale Structure

Non-Gaussianity also leaves signatures in the large-scale structure in the universe. In general, extracting primordial non-Gaussianity from large-scale structure observations is complicated by the fact that non-linear fluctuations produce a non-Gaussianity that completely dominates over the signal from primordial origin. However, recently, the concept of a *scale-dependent bias* has been introduced as a promising probe of primordial non-Gaussianity [91, 101]. It has been shown [91, 101] that for highly biased tracers of the underlying density field, the bias parameter depends on scale and on $f_{\rm NL}$

$$P_{\delta_g}(k) = \left[b+\Delta b(k,f_{\rm NL}^{\rm local})\right]^2 P_\delta(k)\,. \tag{646}$$

The details of the method are beyond the scope of this lecture but may be found *e.g.* in Ref. [101]. Application of the method to the luminous red galaxies (LRGs) sample of SDSS yields [101]

$$-29 \;<\; f_{\rm NL}^{\rm local} \;<\; +70 \quad \text{at} \;\; 95\%\ \text{CL}\,. \tag{647}$$

Note that this limit is competitive with the constraints obtained from the CMB. Although more work is needed to make this a truly robust test of primordial non-Gaussianity, the preliminary results by Slosar *et al.* [101] provide an encouraging proof-of-principle demonstration of the method.

7.6.4 *Summary*: Lecture 4

The study of non-Gaussian contributions to the correlations of cosmological fluctuations, is emerging as an important probe of the early universe [85]. Being a direct measure of inflaton interactions non-Gaussianity can potentially teach us a great deal about the inflationary dynamics.

The basic diagnostic for non-Gaussian fluctuations is the three-point function or bispectrum

$$\langle \mathcal{R}_{\mathbf{k}_1} \mathcal{R}_{\mathbf{k}_2} \mathcal{R}_{\mathbf{k}_3} \rangle = (2\pi)^3 \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_{\mathcal{R}}(k_1, k_2, k_3) \,.$$

Physically motivated models for producing non-Gaussian perturbations often produce signals that peak at special triangle configurations. Three special cases are:

i) *squeezed triangle* ($k_1 \approx k_2 \gg k_3$)

This is the dominant mode of models with multiple light fields during inflation [53, 54, 55, 56, 57, 58, 52, 60, 61, 62, 63, 64, 65], the curvaton scenario [173, 67], inhomogeneous reheating [68, 69], and New Ekpyrotic models [70, 71, 72, 73, 74, 75, 76].

ii) *equilateral triangle* ($k_1 = k_2 = k_3$)

Signals that peak at equilateral triangles arise in models with higher-derivative interactions [92].

iii) *folded triangle* ($k_1 = 2k_2 = 2k_3$)

Folded triangles arise in models with non-standard initial states [93].

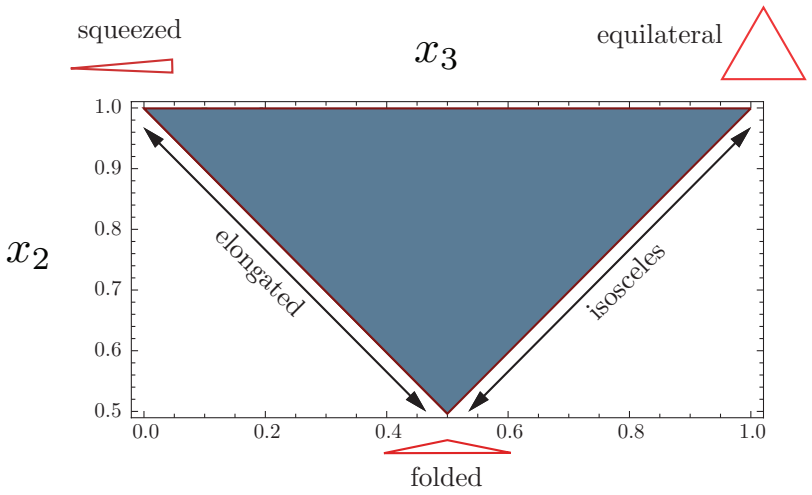


Figure 85: Shapes of Non-Gaussianity. The triangle shapes are parameterized by the rescaled momenta, $x_2 = k_2/k_1$, $x_3 = k_3/k_1$.

The single-field consistency relation is

$$\lim_{k_3 \to 0} \langle \mathcal{R}_{\mathbf{k}_1} \mathcal{R}_{\mathbf{k}_2} \mathcal{R}_{\mathbf{k}_3} \rangle = (2\pi)^3 \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \underline{\underline{(1 - n_{\rm s})}} P_{\mathcal{R}}(k_1) P_{\mathcal{R}}(k_3) \,.$$

This states that the squeezed limit of the bispectrum for single-field inflation is proportional to the deviation from scale-invariance, $1 - n_{\rm s}$.

7.7 Conclusions

7.7.1 *Recap*: TASI Lectures on Inflation

Fig. 86 summarizes many of the key concepts described in these lectures.

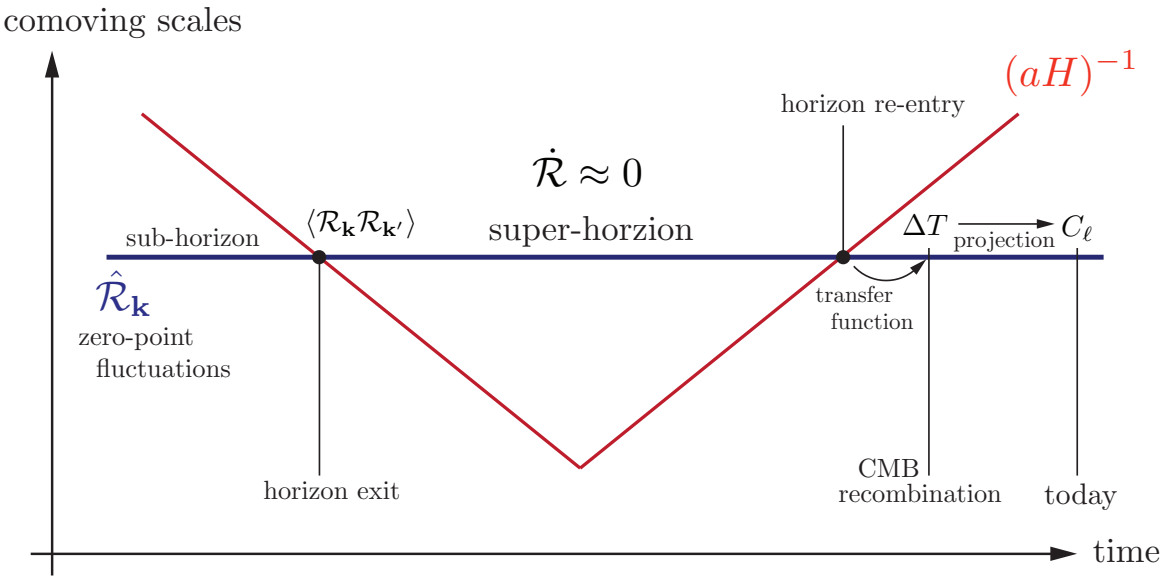


Figure 86: Evolution of the horizon and generation of perturbations in the inflationary universe.

- **Lecture 1**: We defined inflation as a phase in the very early universe when the comoving Hubble radius, $(aH)^{-1}$, was decreasing. We explained that this key characteristic of inflation was at the heart of the solution to the horizon and flatness problems. The apparent acausal correlations of CMB fluctuations on super-horizon scales at recombination are explained by those scales being inside the horizon during inflation (and hence causally-connected).

- **Lecture 2**: Modes exit the horizon during inflation and re-enter at later times during the conventional FRW expansion. We described scalar fluctuations during inflation in terms of the comoving curvature perturbation $\mathcal{R}$. A crucial feature of $\mathcal{R}$ is that it freezes on super-horizon scales, $\dot{\mathcal{R}} \approx 0$. The initial conditions for $\mathcal{R}$ can therefore be computed at horizon exit during inflation and translated without change to horizon re-entry (under fairly weak assumptions this is independent of the unknown physics of reheating). In Lecture 2 we computed the power spectrum of curvature perturbations, $\langle \mathcal{R}_{\mathbf{k}} \mathcal{R}_{\mathbf{k}'} \rangle = (2\pi)^3 \delta(\mathbf{k} + \mathbf{k}') P_{\mathcal{R}}(k)$, at horizon exit.

- **Lecture 3**: After horizon re-entry, the curvature perturbation $\mathcal{R}$ evolves into fluctuations of the CMB temperature ΔT at recombination. This sub-horizon evolution is captured by the transfer functions discussed in Lecture 3. Finally, today we see a projection of the CMB fluctuations from the last-scattering-surface to us. Experiments measure the angular power spectrum of CMB temperature fluctuations, C_ℓ. In Lecture 3 we explained how to relate the observed angular power spectrum of CMB anisotropies to the power spectrum of primordial curvature fluctuations, $P_{\mathcal{R}}(k)$, generated during inflation. Inverting the sub-horizon evolution and removing projection effects, CMB observations therefore provide a powerful probe of the inflationary perturbations.

- **Lecture 4**: The three-point function of primordial curvature perturbations, $\langle \mathcal{R}_{\mathbf{k}_1} \mathcal{R}_{\mathbf{k}_2} \mathcal{R}_{\mathbf{k}_3} \rangle = (2\pi)^3 \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_{\mathcal{R}}(k_1, k_2, k_3)$, can be an additional probe of the physics of inflation if the primordial fluctuations are sufficiently non-Gaussian.

7.7.2 Future Prospects and Open Problems

We have described the present observational evidence for inflation and highlighted future observational opportunities for further tests of the physics of inflation. Two of the most direct probes of inflation are primordial tensor modes and primordial non-Gaussianity:

- **B-modes**

  Detecting primordial B-modes (the CMB polarization signature of inflationary tensor modes) is clearly the most distinctive observation we could make to confirm inflation. We would measure the energy scale of inflation and learn that the inflaton field moved over a super-Planckian distance. The European *Planck* satellite [142], many ground-based or balloon experiments [143, 144, 145, 146, 147, 148, 149, 150], as well as the planned *CMBPol* mission [7, 151], all hope to detect this signal from the inflationary era. The theoretical community is awaiting the results from these experiments with great anticipation.

- **Non-Gaussianity**

  A slightly more model-dependent signature of the physics of inflation is the possible existence of non-Gaussianity in the primordial fluctuations. While predicted to be small for single-field slow-roll models, models with multiple fields, higher-derivative interactions or non-trivial vacuum states may leave non-Gaussian signatures. The momentum dependence of the Fourier-space signal is a powerful diagnostic of the mechanism that laid down the primordial fluctuations. The *Planck* satellite will be a sensitive probe of primordial non-Gaussianity.

In these lectures we have presented a rather optimistic view on inflation. While this illustrates the significant theoretical and observational advances that have been made in recent years in understanding and constraining the physics of inflation, it ignores important conceptual problems that the theory still faces. Here we mention some of these theoretical challenges and point to the relevant literature for more details:

- **Initial Conditions**

  The lectures have mentioned the initial conditions required to start inflation only in a very superficial way. Partly this is a reflection of the fact that the inflationary initial conditions aren't very well understood.

  Our simple slow-roll analysis of inflation has *assumed* that the initial inflaton velocities are small and that initial inhomogeneities in the inflaton field aren't large enough to prevent inflation:

  - *The overshoot problem*

    If the initial inflaton velocity near the region of the potential where inflation is supposed to occur is non-negligible, it is possible that the field will *overshoot* that region without sourcing accelerated expansion. This problem is stronger for small-field models where Hubble friction is often not efficient enough to slow the field before it reaches the region of interest.

  - *The patch problem*

    Initial inhomogeneities in the inflaton field provide gradient energy that also hinders accelerated expansion. Numerical analysis for specific examples shows that typically the inflaton field has to be smooth over a few times the horizon size at that time to start inflation.

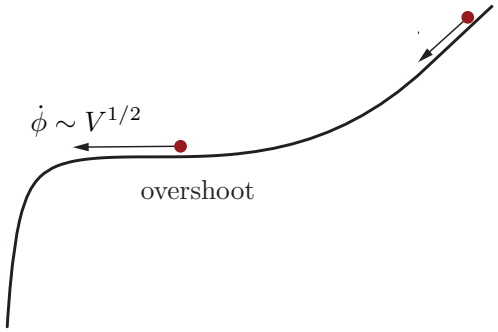


Figure 87: Graphical illustration of the overshoot problem.

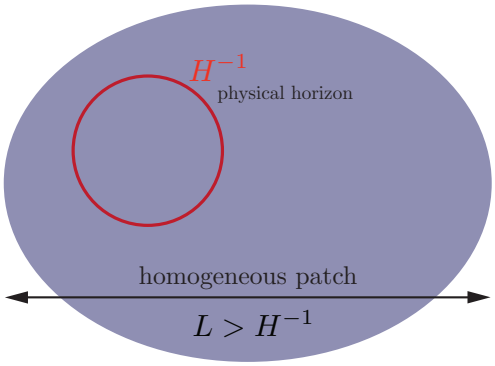


Figure 88: Graphical illustration of the patch problem.

How severe the fine-tuning of initial conditions really is for inflation cannot be discussed outside of the incompletely understood topic of eternal inflation and the measure problem.

- **Eternal Inflation and the Measure Problem**

  The modern view of inflation is that globally it never ends! Inflation ends locally to produce pockets of FRW universes, but there are always region where quantum fluctuations keep the field at high values of the potential energy. Those regions keep expanding exponentially and produce more volume of inflationary regions.

  How likely the initial conditions for inflation are and even what the inflationary predictions themselves are depends on the relative probabilities of the inflationary and non-inflationary patches of the universe (or multiverse). This is the measure problem. Different probability measures can significantly affect the probability of inflationary initial conditions and the likelihood of FRW universes with certain observable characteristics (like flatness, scale-invariant fluctuations, etc.)

  For more on eternal inflation and the measure problem see Refs. [152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165].

These problems illustrate that there is still room for increasing our theoretical understanding of inflation and cosmological initial conditions. At the same time, the advent of high-precision measurements of CMB polarization and small-scale temperature fluctuations promises real experimental test of the inflationary hypothesis.

7.7.3 Guide to Further Reading

The following textbooks, reviews and papers have been useful to me in the preparation of these lectures. The student will find valuable further details about inflation in those works.

Textbooks

- Mukhanov, *Physical Foundations of Cosmology*

  A nice treatment of early universe cosmology and the theory of cosmological perturbations.

- Dodelson, *Modern Cosmology*

  An excellent book about cosmology with a strong focus on the cosmic microwave background. Very readable, *i.e.* you can read it while lying down.

- Weinberg, *Cosmology*

  It is by Steven Weinberg!

- Liddle and Lyth, *Cosmological Inflation and Large-Scale Structure*

  A comprehensive review of inflationary cosmology.

- Longair, *Galaxy Formation*

  A more astrophysical perspective of cosmology.

- Birrell and Davies, *Quantum Field Theory in Curved Spacetime*

  The classic treatment of quantum field theory in curved spacetime.

Reviews

- Baumann et al., *Probing Inflation with CMB Polarization*

  White paper of the Inflation Working Group of the CMBPol Mission Concept Study. More than 60 experts on inflation combined to write this very comprehensive review.

- Baumann and Peiris, *Cosmological Inflation: Theory and Observations*

  In this review Hiranya Peiris and I summarize the basics of inflation and CMB observations for a non-expert audience. The level might be too elementary for the readers of these lectures, but could be of interest to readers looking for some bedtime reading.

- Baumann and McAllister, *Advances in String Inflation*

  In this review Liam McAllister and I describe the challenge of realizing inflation in string theory.

- Lyth and Riotto, *Particle Physics Models of Inflation*

  What these lectures lack on inflationary model-building may be found here.

- Bassett et al., *Inflation Dynamics and Reheating*

  What these lectures lack on reheating may be found here.

- Kinney, *TASI Lectures on Inflation*

  Will Kinney's lectures at *TASI 2008* are perfect as a first read on inflation. It is hoped that these *TASI 2009* lectures make a good second read. I tried to complement Will's lectures by giving more technical details.

- Malik and Wands, *Cosmological Perturbations*

  A nice review of first and second-order perturbation theory. Many useful formulas.

- Komatsu, *The Pursuit of Non-Gaussian Fluctuations in the Cosmic Microwave Background*

  Eiichiro Komatsu's PhD thesis contains a useful review of non-Gaussian fluctuations from inflation.

Papers

Some of the original papers on inflation are very accessible and well worth reading:

- Guth, *Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems*

  This classic is of course a must-read. It provides a very clear explanation of the Big Bang puzzles.

- Maldacena, *Non-Gaussian Features of Primordial Fluctuations in Single Field Inflationary Models*

  This paper provided the first rigorous computation of the three-point function for slow-roll inflation. It also gives one of the clearest and most elegant expositions of the calculation of the power spectra of inflationary fluctuations. My treatment in these lectures was heavily inspired by Maldacena's paper.

7.8 Appendix

7.8.1 Cosmological Perturbation Theory

In this appendix we summarize basic facts of cosmological perturbation theory. This is based on unpublished lecture notes of a course at Princeton University by Uros Seljak and Chris Hirata as well as a review by Malik and Wands [166].

The Perturbed Universe

We consider perturbations to the homogeneous background spacetime and the stress-energy of the universe.

Metric Perturbations

The most general first-order perturbation to a spatially flat FRW metric is

$$\mathrm{d}s^2 = -(1+2\Phi)\mathrm{d}t^2 + 2a(t)B_i\mathrm{d}x^i\mathrm{d}t + a^2(t)[(1-2\Psi)\delta_{ij} + 2E_{ij}]\mathrm{d}x^i\mathrm{d}x^j \tag{A.1}$$

where Φ is a 3-scalar called the *lapse*, B_i is a 3-vector called the *shift*, Ψ is a 3-scalar called the spatial *curvature* perturbation, and E_{ij} is a spatial *shear* 3-tensor which is symmetric and traceless, $E^i_i = \delta^{ij}E_{ij} = 0$. 3-surfaces of constant time t are called *slices* and curves of constant spatial coordinates x^i but varying time t are called *threads*.

Stress-Energy Perturbations

The stress-energy tensor may be described by a density ρ, a pressure p, a 4-velocity u^μ (of the frame in which the 3-momentum density vanishes), and an anisotropic stress $\Sigma^{\mu\nu}$.

Density and pressure perturbations are defined in an obvious way

$$\delta\rho(t,x^i) \equiv \rho(t,x^i) - \bar{\rho}(t)\,, \qquad \text{and} \qquad \delta p(t,x^i) \equiv p(t,x^i) - \bar{p}(t)\,. \tag{A.2}$$

Here, the background values have been denoted by overbars. The 4-velocity has only three independent components (after the metric is fixed) since it has to satisfy the constraint $g_{\mu\nu}u^\mu u^\nu = -1$. In the perturbed metric (A.1) the perturbed 4-velocity is

$$u_\mu \equiv (-1-\Phi, a v_i)\,, \qquad \text{or} \qquad u^\mu \equiv (1-\Phi, a^{-1}(v^i - B^i))\,. \tag{A.3}$$

Here, u_0 is chosen so that the constraint $u_\mu u^\mu = -1$ is satisfied to first order in all perturbations. Anisotropic stress vanishes in the unperturbed FRW universe, so $\Sigma^{\mu\nu}$ is a first-order perturbation. Furthermore, $\Sigma^{\mu\nu}$ is constrained by

$$\Sigma^{\mu\nu}u_\nu = \Sigma^\mu_\mu = 0\,. \tag{A.4}$$

The orthogonality with u_μ implies $\Sigma^{00} = \Sigma^{0j} = 0$, *i.e.* only the spatial components Σ^{ij} are non-zero. The trace condition then implies $\Sigma^i_i = 0$. Anisotropic stress is therefore a traceless symmetric 3-tensor.

Finally, with these definitions the perturbed stress-tensor is

$$\begin{aligned} T^0_0 &= -(\bar\rho + \delta\rho) && \text{(A.5)}\\ T^0_i &= (\bar\rho + \bar p)\, a v_i && \text{(A.6)}\\ T^i_0 &= -(\bar\rho + \bar p)(v^i - B^i)/a && \text{(A.7)}\\ T^i_j &= \delta^i_j(\bar p + \delta p) + \Sigma^i_j\,. && \text{(A.8)} \end{aligned}$$

If there are several contributions to the stress-energy tensor (*e.g.* photons, baryons, dark matter, etc.), they are added: $T_{\mu\nu} = \sum_I T^I_{\mu\nu}$. This implies

$$\begin{aligned} \delta\rho &= \sum_I \delta\rho_I && \text{(A.9)}\\ \delta p &= \sum_I \delta p_I && \text{(A.10)}\\ (\bar\rho + \bar p)v^i &= \sum_I (\bar\rho_I + \bar p_I) v^i_I && \text{(A.11)}\\ \Sigma^{ij} &= \sum_I \Sigma^{ij}_I\,. && \text{(A.12)} \end{aligned}$$

Density, pressure and anisotropic stress perturbations simply add. However, velocities do not add, which motivates defining the 3-momentum density

$$\delta q^i \equiv (\bar\rho + \bar p)\, a v^i\,, \tag{A.13}$$

such that

$$\delta q^i = \sum_I \delta q^i_I\,. \tag{A.14}$$

Scalars, Vectors and Tensors

The Einstein Equations relate metric perturbations to the stress-energy perturbations. Einstein's Equations are both complicated (coupled second-order partial differential equations) and non-linear. Fortunately, the symmetries of the flat FRW background spacetime allow perturbations to be decomposed into independent scalar, vector and tensor components. This reduces the Einstein Equations to a set of uncoupled ordinary differential equations.

Helicity and SVT-Decomposition in Fourier Space

The decomposition into scalar, vector and tensor perturbations is most elegantly explained in Fourier space. We define the Fourier components of a general perturbation $\delta Q(t, \mathbf{x})$ as follows

$$\delta Q(t, \mathbf{k}) = \int \mathrm{d}^3\mathbf{x}\, \delta Q(t, \mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}\,. \tag{A.15}$$

First note that as a consequence of translation invariance different Fourier modes (different wavenumbers k) evolve independently.[44]

Proof:

Consider the linear evolution of N perturbations δQ_I, $I = 1, \ldots, N$ from an initial time t_1 to a final time t_2

$$\delta Q_I(t_2, \mathbf{k}) = \sum_{J=1}^{N} \int \mathrm{d}^3\bar{\mathbf{k}}\; T_{IJ}(t_2, t_1, \mathbf{k}, \bar{\mathbf{k}}) \delta Q_J(t_1, \bar{\mathbf{k}})\,, \tag{A.16}$$

where the transfer matrix $T_{IJ}(t_2, t_1, \mathbf{k}, \bar{\mathbf{k}})$ follows from the Einstein Equations and we have allowed for the possibility of a mixing of k-modes. We now show that translation invariance in fact forbids such couplings. Consider the coordinate transformation

$$x^{i'} = x^i + \Delta x^i\,, \qquad \text{where} \quad \Delta x^i = const. \tag{A.17}$$

You may convince yourself that the Fourier amplitude gets shifted as follows

$$\delta Q'_I(t, \mathbf{k}) = e^{-ik_j \Delta x^j} \delta Q_I(t, \mathbf{k})\,. \tag{A.18}$$

Thus the evolution equation in the primed coordinate system becomes

$$\delta Q'_I(t_2, \mathbf{k}) = \sum_{J=1}^{N} \int \mathrm{d}^3\bar{\mathbf{k}}\; e^{-ik_j \Delta x^j} T_{IJ}(t_2, t_1, \mathbf{k}, \bar{\mathbf{k}}) e^{i\bar{k}_j \Delta x^j}\, \delta Q'_J(t_1, \bar{\mathbf{k}}) \tag{A.19}$$

$$\equiv \sum_{J=1}^{N} \int \mathrm{d}^3\bar{\mathbf{k}}\; T'_{IJ}(t_2, t_1, \mathbf{k}, \bar{\mathbf{k}}) \delta Q_J(t_1, \bar{\mathbf{k}})\,. \tag{A.20}$$

By translation invariance the equations of motion must be the same in both coordinate systems, *i.e.* the transfer matrices T_{IJ} and T'_{IJ} must be the same

$$T_{IJ}(t_2, t_1, \mathbf{k}, \bar{\mathbf{k}}) = e^{i(\bar{k}_j - k_j)\Delta x^j} T_{IJ}(t_2, t_1, \mathbf{k}, \bar{\mathbf{k}})\,. \tag{A.21}$$

This must hold for all Δx^j. Hence, either $\bar{\mathbf{k}} = \mathbf{k}$ or $T_{IJ}(t_2, t_1; \mathbf{k}, \bar{\mathbf{k}}) = 0$, *i.e.* the perturbation $\delta Q_I(t_2, \mathbf{k})$ of wavevector $\mathbf{k}$ depends only on the initial perturbations of wavevector $\mathbf{k}$. At linear order there is no coupling of different k-modes. QED.

Now consider rotations around the Fourier vector $\mathbf{k}$ by an angle ψ. We classify perturbations according to their *helicity* m: a perturbation of helicity m has its amplitude multiplied by $e^{im\psi}$ under the above rotation. We define scalar, vector and tensor perturbations as having helicities 0, ± 1, ± 2, respectively.

Consider a Fourier mode with wavevector $\mathbf{k}$. Without loss of generality we may assume that $\mathbf{k} = (0, 0, k)$ (or use rotational invariance of the background). The spatial dependence of any perturbation then is

$$\delta Q \propto e^{ikx^3}\,. \tag{A.22}$$

[44] The following proof was related to me by Uros Seljak and Chris Hirata.

To study rotations around $\mathbf{k}$ it proves convenient to switch to the helicity basis

$$\mathbf{e}_\pm \equiv \frac{\mathbf{e}_1 \pm i\mathbf{e}_2}{\sqrt{2}}\,, \qquad \mathbf{e}_3\,, \tag{A.23}$$

where $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is the Cartesian basis. A rotation around the 3-axis by an angle ψ has the following effect

$$\begin{pmatrix} x^{1'} \\ x^{2'} \end{pmatrix} = \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix}\,, \qquad x^{3'} = x^3\,, \tag{A.24}$$

and

$$\mathbf{e}'_\pm = e^{\pm i\psi}\mathbf{e}_\pm\,, \qquad \mathbf{e}'_3 = \mathbf{e}_3\,. \tag{A.25}$$

The contravariant components of any tensor $T_{i_1 i_2 \ldots i_n}$ transform as

$$T'_{i_1 i_2 \ldots i_n} = e^{i(n_+ - n_-)\psi} T_{i_1 i_2 \ldots i_n} \equiv e^{im\psi} T_{i_1 i_2 \ldots i_n} \tag{A.26}$$

where n_+ and n_- count the number of plus and minus indices in $i_1 \ldots i_n$, respectively. Helicity is defined as the difference $m \equiv n_+ - n_-$.

In the helicity basis $\{\mathbf{e}_\pm, \mathbf{e}_3\}$, a 3- scalar α has a single component with no indicies and is therefore obviously of helicity 0; a 3-vector β_i has 3 components $\beta_+, \beta_-, \beta_3$ of helicity ± 1 and 0; a symmetric and traceless 3-tensor γ_{ij} has 5 components $\gamma_{--}, \gamma_{++}, \gamma_{-3}, \gamma_{+3}, \gamma_{33}$ (the tracelessness condition makes γ_{-+} redundant), of helicity ± 2, ± 1 and 0.

Rotational invariance of the background implies that helicity scalars, vectors and tensors evolve independently.[45]

Proof:

Consider N perturbations δQ_I, $I = 1, \ldots, N$ of helicity m_I. The linear evolution is

$$\delta Q_I(t_2, \mathbf{k}) = \sum_{J=1}^{N} T_{IJ}(t_2, t_1, \mathbf{k}) \delta Q_J(t_1, \mathbf{k})\,, \tag{A.27}$$

where the transfer matrix $T_{IJ}(t_2, t_1, \mathbf{k})$ follows from the Einstein Equations. Under rotation the perturbations transform as

$$\delta Q'_I(t, \mathbf{k}) = e^{im_I\psi} \delta Q_I(t, \mathbf{k}) \tag{A.28}$$

and

$$\delta Q'_I(t_2, \mathbf{k}) = \sum_{J=1}^{N} e^{im_I\psi}\, T_{IJ}(t_2, t_1, \mathbf{k})\, e^{-im_J\psi} \delta Q'_J(t_1, \mathbf{k})\,. \tag{A.29}$$

By rotational invariance of the equations of motion

$$T_{IJ}(t_2, t_1, \mathbf{k}) = e^{im_I\psi}\, T_{IJ}(t_2, t_1, \mathbf{k})\, e^{-im_J\psi} = e^{i(m_I - m_J)\psi} T_{IJ}(t_2, t_1, \mathbf{k})\,, \tag{A.30}$$

which has to hold for any angle ψ; it follows that eithers $m_I = m_J$, *i.e.* δQ_I and δQ_J have the same helicity or $T_{IJ}(t_2, t_1, \mathbf{k}) = 0$. This proves that the equations of motion don't mix modes of different helicity. QED.

[45] The following proof was related to me by Uros Seljak and Chris Hirata.

Real Space SVT-Decomposition

In the last section we have seen that 3-scalars correspond to helicity scalars, 3-vectors decompose into helicity scalars and vectors, and 3-tensors decompose into helicity scalars, vectors and tensors. We now look at this from a different perspective.

A 3-scalar is obviously also a helicity scalar

$$\alpha = \alpha^S \,. \tag{A.31}$$

Consider a 3-vector β_i. We argue that it can be decomposed as

$$\beta_i = \beta_i^S + \beta_i^V \,, \tag{A.32}$$

where

$$\beta_i^S = \nabla_i \hat{\beta} \,, \qquad \nabla^i \beta_i^V = 0 \,, \tag{A.33}$$

or, in Fourier space,

$$\beta_i^S = -\frac{ik_i}{k}\beta \,, \qquad k_i \beta_i^V = 0 \,. \tag{A.34}$$

Here, we have defined $\beta \equiv k\hat{\beta}$. .

Exercise 12 (Helicity Vector)**.** *Show that β_i^V is a helicity vector.*

Similarly, a traceless, symmetric 3-tensor can be written as

$$\gamma_{ij} = \gamma_{ij}^S + \gamma_{ij}^V + \gamma_{ij}^T \,, \tag{A.35}$$

where

$$\gamma_{ij}^S = \left(\nabla_i\nabla_j - \frac{1}{3}\delta_{ij}\nabla^2\right)\hat{\gamma} \tag{A.36}$$

$$\gamma_{ij}^V = \frac{1}{2}\left(\nabla_i\hat{\gamma}_j + \nabla_j\hat{\gamma}_i\right) \,, \quad \nabla_i\hat{\gamma}_i = 0 \tag{A.37}$$

$$\nabla_i\gamma_{ij}^T = = 0 \,. \tag{A.38}$$

or

$$\gamma_{ij}^S = \left(-\frac{k_ik_j}{k^2} + \frac{1}{3}\delta_{ij}\right)\gamma \tag{A.39}$$

$$\gamma_{ij}^V = -\frac{i}{2k}\left(k_i\gamma_j + k_j\gamma_i\right) \,, \quad k_i\gamma_i = 0 \tag{A.40}$$

$$k_i\gamma_{ij}^T = = 0 \,. \tag{A.41}$$

Here, we have defined $\gamma \equiv k^2\hat{\gamma}$ and $\gamma_i \equiv k\hat{\gamma}_i$.

Exercise 13 (Helicity Vectors and Tensors)**.** *Show that γ_{ij}^V and γ_{ij}^T are a helicity vector and a helicity tensor, respectively.*

Choosing $\mathbf{k}$ along the 3-axis, *i.e.* $\mathbf{k} = (0, 0, k)$ we find

$$\gamma_{ij}^S = \frac{1}{3}\begin{pmatrix} \gamma & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & -2\gamma \end{pmatrix} \tag{A.42}$$

$$\gamma_{ij}^V = -\frac{i}{2}\begin{pmatrix} 0 & 0 & \gamma_1 \\ 0 & 0 & \gamma_2 \\ \gamma_1 & \gamma_2 & 0 \end{pmatrix} \tag{A.43}$$

$$\gamma_{ij}^T = \begin{pmatrix} \gamma^\times & \gamma^+ & 0 \\ \gamma^+ & -\gamma^\times & 0 \\ 0 & 0 & 0 \end{pmatrix} \,. \tag{A.44}$$

Scalars

Metric Perturbations

Four scalar metric perturbations Φ, $B_{,i}$, $\Psi\delta_{ij}$ and $E_{,ij}$ may be constructed from 3-scalars, their derivatives and the background spatial metric, *i.e.*

$$\mathrm{d}s^2 = -(1+2\Phi)\mathrm{d}t^2 + 2a(t)B_{,i}\mathrm{d}x^i\mathrm{d}t + a^2(t)[(1-2\Psi)\delta_{ij} + 2E_{,ij}]\mathrm{d}x^i\mathrm{d}x^j \tag{A.45}$$

Here, we have absorbed the $\nabla^2 E\,\delta_{ij}$ part of the helicity scalar E^S_{ij} in $\Psi\,\delta_{ij}$.

The intrinsic Ricci scalar curvature of constant time hypersurfaces is

$$R_{(3)} = \frac{4}{a^2}\nabla^2\Psi\,. \tag{A.46}$$

This explains why Ψ is often referred to as the curvature perturbation.

There are two scalar gauge transformations

$$t \;\rightarrow\; t+\alpha \tag{A.47}$$
$$x^i \;\rightarrow\; x^i + \delta^{ij}\beta_{,j}\,. \tag{A.48}$$

Under these coordinate transformations the scalar metric perturbations transform as

$$\Phi \;\rightarrow\; \Phi - \dot{\alpha} \tag{A.49}$$
$$B \;\rightarrow\; B + a^{-1}\alpha - a\dot{\beta} \tag{A.50}$$
$$E \;\rightarrow\; E - \beta \tag{A.51}$$
$$\Psi \;\rightarrow\; \Psi + H\alpha\,. \tag{A.52}$$

Note that the combination $\dot{E} - B/a$ is independent of the spatial gauge and only depends on the temporal gauge. It is called the scalar potential for the anisotropic shear of world lines orthogonal to constant time hypersurfaces. To extract physical results it is useful to define gauge-invariant combinations of the scalar metric perturbations. Two important gauge-invariant quantities were introduced by Bardeen [219]

$$\Phi_{\mathrm{B}} \;\equiv\; \Phi - \frac{d}{dt}[a^2(\dot{E} - B/a)] \tag{A.53}$$
$$\Psi_{\mathrm{B}} \;\equiv\; \Psi + a^2 H(\dot{E} - B/a)\,. \tag{A.54}$$

Matter Perturbations

Matter perturbations are also gauge-dependent, *e.g.* density and pressure perturbations transform as follows under temporal gauge transformations

$$\delta\rho \rightarrow \delta\rho - \dot{\bar{\rho}}\,\alpha\,, \qquad \delta p \rightarrow \delta p - \dot{\bar{p}}\,\alpha\,. \tag{A.55}$$

Adiabatic pressure perturbations are defined as

$$\delta p_{ad} \equiv \frac{\dot{\bar{p}}}{\dot{\bar{\rho}}}\,\delta\rho\,. \tag{A.56}$$

The non-adiabiatic, or entropic, part of the pressure perturbations is then gauge-invariant

$$\delta p_{en} \equiv \delta p - \frac{\dot{\bar{p}}}{\dot{\bar{\rho}}}\,\delta\rho\,. \tag{A.57}$$

The scalar part of the 3-momentum density, $(\delta q)_{,i}$, transforms as

$$\delta q \to \delta q + (\bar{\rho} + \bar{p})\,\alpha\,. \tag{A.58}$$

We may then define the gauge-invariant comoving density perturbation

$$\delta\rho_m \equiv \delta\rho - 3H\delta q\,. \tag{A.59}$$

Finally, two important gauge-invariant quantities are formed from combinations of matter and metric perturbations. The *curvature perturbation on uniform density hypersurfaces* is

$$-\zeta \equiv \Psi + \frac{H}{\dot{\bar{\rho}}}\delta\rho\,. \tag{A.60}$$

The *comoving curvature perturbation* is

$$\mathcal{R} = \Psi - \frac{H}{\bar{\rho}+\bar{p}}\delta q\,. \tag{A.61}$$

We will show that ζ and $\mathcal{R}$ are equal on superhorizon scales, where they become time-independent. The computation of the inflationary perturbation spectrum is most clearly phrased in terms of ζ and $\mathcal{R}$.

Einstein Equations

To relate the metric and stress-energy perturbations, we consider the perturbed Einstein Equations

$$\delta G_{\mu\nu} = 8\pi G\,\delta T_{\mu\nu}\,. \tag{A.62}$$

We work at linear order. This leads to the *energy and momentum constraint equations*

$$3H(\dot{\Psi} + H\Phi) + \frac{k^2}{a^2}\left[\Psi + H(a^2\dot{E} - aB)\right] = -4\pi G\,\delta\rho \tag{A.63}$$

$$\dot{\Psi} + H\Phi = -4\pi G\,\delta q\,. \tag{A.64}$$

These can be combined into the gauge-invariant *Poisson Equation*

$$\frac{k^2}{a^2}\Psi_{\rm B} = -4\pi G\delta\rho_m\,. \tag{A.65}$$

The Einstein equation also yield two *evolution equations*

$$\ddot{\Psi} + 3H\dot{\Psi} + H\dot{\Phi} + (3H^2 + 2\dot{H})\Phi = 4\pi G\left(\delta p - \frac{2}{3}k^2\delta\Sigma\right) \tag{A.66}$$

$$(\partial_t + 3H)(\dot{E} - B/a) + \frac{\Psi - \Phi}{a^2} = 8\pi G\,\delta\Sigma\,. \tag{A.67}$$

The last equation may be written as

$$\Psi_{\rm B} - \Phi_{\rm B} = 8\pi G\,a^2\delta\Sigma\,. \tag{A.68}$$

In the absence of anisotropic stress this implies, $\Psi_{\rm B} = \Phi_{\rm B}$.

Energy-momentum conservation, $\nabla_\mu T^{\mu\nu} = 0$, gives the *continuity equation* and the *Euler Equation*

$$\dot{\delta\rho} + 3H(\delta\rho + \delta p) = \frac{k^2}{a^2}\delta q + (\bar{\rho}+\bar{p})[3\dot{\Psi} + k^2(\dot{E} + B/a)]\,, \tag{A.69}$$

$$\dot{\delta q} + 3H\delta q = -\delta p + \frac{2}{3}k^2\delta\Sigma - (\bar{\rho}+\bar{p})\Phi\,. \tag{A.70}$$

Expressed in terms of the curvature perturbation on uniform-density hypersurfaces, ζ, Eqn. (A.69) reads

$$\dot{\zeta} = -H\frac{\delta p_{en}}{\bar{\rho}+\bar{p}} - \Pi\,, \tag{A.71}$$

where δp_{en} is the non-adiabatic component of the pressure perturbation, and Π is the scalar shear along comoving worldlines

$$\frac{\Pi}{H} \equiv -\frac{k^2}{3H}\left[\dot{E} - B/a + \frac{\delta q}{a^2(\bar{\rho}+\bar{p})}\right] \tag{A.72}$$

$$= -\frac{k^2}{3a^2H^2}\left[\zeta - \Psi_{\rm B}\left(1 - \frac{2\bar{\rho}}{9(\bar{\rho}+\bar{p})}\frac{k^2}{a^2H^2}\right)\right]\,. \tag{A.73}$$

For adiabative perturbations, $\delta p_{en} = 0$ on superhorizon scales, $k/(aH) \ll 1$ (*i.e.* $\Pi/H \to 0$ for finite ζ and $\Psi_{\rm B}$), the curvature perturbation ζ is constant. This is a crucial result for our computation of the inflationary spectrum of ζ in **Lecture 2**. It justifies computing ζ at horizon exit and ignoring superhorizon evolution.

Popular Gauges

For reference we now give the Einstein Equations and the conservation equations is various popular gauges:

- **Synchronous gauge**

  A popular gauge, especially for numerical implementation of the perturbation equations (*cf.* CMBFAST [28] or CAMB [166]), is synchronous gauge. It is defined by

  $$\Phi = B = 0\,. \tag{A.74}$$

  The Einstein Equations become

  $$3H\dot{\Psi} + \frac{k^2}{a^2}\left[\Psi + Ha^2\dot{E}\right] = -4\pi G\,\delta\rho \tag{A.75}$$

  $$\dot{\Psi} = -4\pi G\,\delta q \tag{A.76}$$

  $$\ddot{\Psi} + 3H\dot{\Psi} = 4\pi G\left(\delta p - \frac{2}{3}k^2\delta\Sigma\right) \tag{A.77}$$

  $$(\partial_t + 3H)\dot{E} + \frac{\Psi}{a^2} = 8\pi G\,\delta\Sigma\,. \tag{A.78}$$

  The conservation equation are

  $$\dot{\delta\rho} + 3H(\delta\rho + \delta p) = \frac{k^2}{a^2}\delta q + (\bar{\rho}+\bar{p})[3\dot{\Psi} + k^2\dot{E}] \tag{A.79}$$

  $$\dot{\delta q} + 3H\delta q = -\delta p + \frac{2}{3}k^2\delta\Sigma\,. \tag{A.80}$$

- **Newtonian gauge**

  The Newtonian gauge has its name because it reduces to Newtonian gravity in the small-scale limit. It is popular for analytic work since it leads to algebraic relations between metric and stress-energy perturbations.

  Newtonian gauge is defined by

  $$B = E = 0\,, \tag{A.81}$$

and

$$\mathrm{d}s^2 - (1+2\Phi)\mathrm{d}t^2 + a^2(t)(1-2\Psi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\,. \tag{A.82}$$

The Einstein Equations are

$$\begin{aligned} 3H(\dot{\Psi} + H\Phi) + \frac{k^2}{a^2}\Psi &= -4\pi G\,\delta\rho & \text{(A.83)}\\ \dot{\Psi} + H\Phi &= -4\pi G\,\delta q & \text{(A.84)}\\ \ddot{\Psi} + 3H\dot{\Psi} + H\dot{\Phi} + (3H^2 + 2\dot{H})\Phi &= 4\pi G\left(\delta p - \frac{2}{3}k^2\delta\Sigma\right) & \text{(A.85)}\\ \frac{\Psi - \Phi}{a^2} &= 8\pi G\,\delta\Sigma\,. & \text{(A.86)} \end{aligned}$$

The continuity equations are

$$\begin{aligned} \dot{\delta\rho} + 3H(\delta\rho + \delta p) &= \frac{k^2}{a^2}\delta q + 3(\bar{\rho} + \bar{p})\dot{\Psi}\,, & \text{(A.87)}\\ \dot{\delta q} + 3H\delta q &= -\delta p + \frac{2}{3}k^2\delta\Sigma - (\bar{\rho} + \bar{p})\Phi\,. & \text{(A.88)} \end{aligned}$$

- **Uniform density gauge**

  The uniform density gauge is useful for describing the evolution of perturbations on superhorizon scales. As its name suggests it is defined by

$$\delta\rho = 0\,. \tag{A.89}$$

  In addition, it is convenient to take

$$E = 0\,, \qquad -\Psi \equiv \zeta\,. \tag{A.90}$$

  The Einstein Equations are

$$\begin{aligned} 3H(-\dot{\zeta} + H\Phi) - \frac{k^2}{a^2}\left[\zeta + aHB\right] &= 0 & \text{(A.91)}\\ -\dot{\zeta} + H\Phi &= -4\pi G\,\delta q & \text{(A.92)}\\ -\ddot{\zeta} - 3H\dot{\zeta} + H\dot{\Phi} + (3H^2 + 2\dot{H})\Phi &= 4\pi G\left(\delta p - \frac{2}{3}k^2\delta\Sigma\right) & \text{(A.93)}\\ (\partial_t + 3H)B/a + \frac{\zeta + \Phi}{a^2} &= -8\pi G\,\delta\Sigma\,. & \text{(A.94)} \end{aligned}$$

  The continuity equations are

$$\begin{aligned} 3H\delta p &= \frac{k^2}{a^2}\delta q + (\bar{\rho} + \bar{p})[-3\dot{\zeta} + k^2 B/a]\,, & \text{(A.95)}\\ \dot{\delta q} + 3H\delta q &= -\delta p + \frac{2}{3}k^2\delta\Sigma - (\bar{\rho} + \bar{p})\Phi\,. & \text{(A.96)} \end{aligned}$$

- **Comoving gauge**

  Comoving gauge is defined by the vanishing of the scalar momentum density,

$$\delta q = 0\,, \qquad E = 0\,. \tag{A.97}$$

  It is also conventional to set $-\Psi \equiv \mathcal{R}$ in this gauge.

The Einstein Equations are

$$3H(-\dot{\mathcal{R}} + H\Phi) + \frac{k^2}{a^2}\left[-\mathcal{R} - aHB\right] = -4\pi G\,\delta\rho \tag{A.98}$$

$$-\dot{\mathcal{R}} + H\Phi = 0 \tag{A.99}$$

$$-\ddot{\mathcal{R}} - 3H\dot{\mathcal{R}} + H\dot{\Phi} + (3H^2 + 2\dot{H})\Phi = 4\pi G\left(\delta p - \frac{2}{3}k^2\delta\Sigma\right) \tag{A.100}$$

$$(\partial_t + 3H)B/a + \frac{\mathcal{R} + \Phi}{a^2} = -8\pi G\,\delta\Sigma\,. \tag{A.101}$$

The continuity equations are

$$\dot{\delta\rho} + 3H(\delta\rho + \delta p) = (\bar{\rho} + \bar{p})[-3\dot{\mathcal{R}} + k^2 B/a]\,. \tag{A.102}$$

$$0 = -\delta p + \frac{2}{3}k^2\delta\Sigma - (\bar{\rho} + \bar{p})\Phi\,. \tag{A.103}$$

Equations (A.103) and (A.99) may be combined into

$$\Phi = \frac{-\delta p + \frac{2}{3}\Sigma}{\bar{\rho} + \bar{p}}\,, \qquad kB = \frac{4\pi G a^2\delta\rho - k^2\mathcal{R}}{aH}\,. \tag{A.104}$$

- **Spatially-flat gauge**

  A convenient gauge for computing inflationary perturbation is spatially-flat gauge

$$\Psi = E = 0\,. \tag{A.105}$$

  During inflation all scalar perturbations are then described by $\delta\phi$.

  The Einstein Equations are

$$3H^2\Phi + \frac{k^2}{a^2}\left[-aHB)\right] = -4\pi G\,\delta\rho \tag{A.106}$$

$$H\Phi = -4\pi G\,\delta q \tag{A.107}$$

$$H\dot{\Phi} + (3H^2 + 2\dot{H})\Phi = 4\pi G\left(\delta p - \frac{2}{3}k^2\delta\Sigma\right) \tag{A.108}$$

$$(\partial_t + 3H)B/a + \frac{\Phi}{a^2} = -8\pi G\,\delta\Sigma\,. \tag{A.109}$$

  The continuity equations are

$$\dot{\delta\rho} + 3H(\delta\rho + \delta p) = \frac{k^2}{a^2}\delta q + (\bar{\rho} + \bar{p})[k^2 B/a]\,, \tag{A.110}$$

$$\dot{\delta q} + 3H\delta q = -\delta p + \frac{2}{3}k^2\delta\Sigma - (\bar{\rho} + \bar{p})\Phi\,. \tag{A.111}$$

Vectors

Metric Perturbations

Vector type metric perturbations are defined as

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + 2a(t)S_i\mathrm{d}x^i\mathrm{d}t + a^2(t)[\delta_{ij} + 2F_{(i,j)}]\mathrm{d}x^i\mathrm{d}x^j\,, \tag{A.112}$$

where $S_{i,i} = F_{i,i} = 0$. The vector gauge transformation is

$$x^i \to x^i + \beta^i\,, \qquad \beta_{i,i} = 0\,. \tag{A.113}$$

They lead to the transformations

$$S_i \to S_i + a\dot{\beta}_i\,, \tag{A.114}$$

$$F_i \to F_i - \beta_i\,. \tag{A.115}$$

The combination $\dot{F}_i + S_i/a$ is called the gauge-invariant vector shear perturbation.

Matter Perturbations

We define the vector part of the anisotropic stress by

$$\delta\Sigma_{ij} = \partial_{(i}\Sigma_{j)}\,, \tag{A.116}$$

where Σ_i is divergence-free, $\Sigma_{i,i} = 0$.

Einstein Equations

For vector perturbations there are only two Einstein Equations,

$$\dot{\delta q_i} + 3H\delta q_i = k^2\delta\Sigma_i\,, \tag{A.117}$$

$$k^2(\dot{F}_i + S_i/a) = 16\pi G\,\delta q_i\,. \tag{A.118}$$

In the absence of anisotropic stress ($\delta\Sigma_i = 0$) the divergence-free momentum δq_i decays with the expansion of the universe; see Eqn. (A.117). The shear perturbation $\dot{F}_i + S_i/a$ then vanishes by Eqn. (A.118). Under most circumstances vector perturbations are therefore subdominant. They won't play an important role in these lectures. In particular, vector perturbations aren't created by inflation.

Tensors

Metric Perturbations

Tensor metric perturbations are defined as

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)[\delta_{ij} + h_{ij}]\mathrm{d}x^i\mathrm{d}x^j\,, \tag{A.119}$$

where $h_{ij,i} = h^i_i = 0$. Tensor perturbations are automatically gauge-invariant (at linear order). It is conventional to decompose tensor perturbations into eigenmodes of the spatial Laplacian, $\nabla^2 e_{ij} = -k^2 e_{ij}$, with comoving wavenumber k and scalar amplitude $h(t)$,

$$h_{ij} = h(t)e_{ij}^{(+,\times)}(x)\,. \tag{A.120}$$

Here, $+$ and $\times$ denote the two possible polarization states.

Matter Perturbations

Tensor perturbations are sourced by anisotropic stress Σ_{ij}, with $\Sigma_{ij,i} = \Sigma^i_i = 0$. It is typically a good approximation to assume that the anisotropic stress is negligible, although a small amplitude is induced by neutrino free-streaming.

Einstein Equations

For tensor perturbations there is only one Einstein Equation. In the absence of anisotropic stress this is

$$\ddot{h} + 3H\dot{h} + \frac{k^2}{a^2}h = 0\,. \tag{A.121}$$

This is a wave equation describing the evolution of gravitational waves in an expanding universe. Gravitational waves are produced by inflation, but then decay with the expansion of the universe. However, at recombination their amplitude may still be large enough to leave distinctive signatures in B-modes of CMB polarization.

Statistics

We recall some basic facts about statistics. More details may be found in Licia Verde's notes [167].

Fourier Conventions

Different conventions exist for the normalization of Fourier transforms. Defining

$$\mathcal{R}_{\mathbf{k}} = A \int \mathrm{d}^3x\, \mathcal{R}(\mathbf{x})\, \mathbf{e}^{-\mathbf{i}\mathbf{k}\cdot\mathbf{x}}\,, \tag{A.122}$$

$$\mathcal{R}(\mathbf{x}) = B \int \mathrm{d}^3k\, \mathcal{R}_{\mathbf{k}}\, e^{i\mathbf{k}\cdot\mathbf{x}}\,, \tag{A.123}$$

implies that the Dirac delta function is

$$\delta(\mathbf{k}) = BA \int \mathrm{d}^3x\, e^{\pm i\mathbf{k}\cdot\mathbf{x}}\,, \qquad BA = \frac{1}{(2\pi)^3}\,. \tag{A.124}$$

Except for the constraint $BA = 1/(2\pi)^3$ different conventions are possible for the values of A and B. These conventions can lead to some confusion about factors of 2π in the normalization of the power spectrum. In the main text we follow the convention $A = 1$, $B = 1/(2\pi)^3$ (the other common convention is $A = B = 1/(2\pi)^{3/2}$; it is nice, since it makes the basis function $e^{i\mathbf{k}\mathbf{x}}$ orthonormal rather than just orthogonal.), but in this appendix we will keep things general in order to help identifying normalization errors in the literature.

Two-Point Correlation Function

We make frequent use of the two-point correlation function

$$\xi_{\mathcal{R}}(r) \equiv \langle \mathcal{R}(\mathbf{x})\mathcal{R}(\mathbf{x}+\mathbf{r})\rangle\,. \tag{A.125}$$

Here, we have made the assumption that by isotropy ξ depends only on $r \equiv |\mathbf{r}|$ (distance not orientation).

Power Spectrum

Consider the following ensemble average

$$\langle \mathcal{R}_{\mathbf{k}}\, \mathcal{R}_{\mathbf{k}'}\rangle\,, \tag{A.126}$$

where $\mathcal{R}^*_{\mathbf{k}} = \mathcal{R}_{-\mathbf{k}}$ because $\mathcal{R}(\mathbf{x})$ is real. Substituting (A.122) gives

$$\langle \mathcal{R}_{\mathbf{k}}\, \mathcal{R}_{\mathbf{k}'}\rangle = A^2 \int \mathrm{d}^3x\, e^{-i(\mathbf{k}+\mathbf{k}')\mathbf{x}} \int \mathrm{d}^3r\, \xi_{\mathcal{R}}(r) e^{-i\mathbf{k}\mathbf{r}} \tag{A.127}$$

$$= \frac{A}{B}\delta(\mathbf{k}+\mathbf{k}') \int \mathrm{d}^3r\, \xi_{\mathcal{R}}(r) e^{-i\mathbf{k}\mathbf{r}}\,. \tag{A.128}$$

If we *define* the power spectrum as the Fourier transform of the two-point correlation function

$$P_{\mathcal{R}}(k) \equiv A \int \mathrm{d}^3r\, \xi_{\mathcal{R}}(r)\, e^{-i\mathbf{k}\cdot\mathbf{r}}\,, \tag{A.129}$$

then we get

$$\langle \mathcal{R}_{\mathbf{k}}\, \mathcal{R}_{\mathbf{k}'}\rangle = \frac{1}{B} P_{\mathcal{R}}(k)\delta(\mathbf{k}+\mathbf{k}')\,. \tag{A.130}$$

Notice that often the power spectrum is defined as

$$\langle \mathcal{R}_{\mathbf{k}}\, \mathcal{R}_{\mathbf{k}'} \rangle = (2\pi)^3 P_{\mathcal{R}}(k) \delta(\mathbf{k} + \mathbf{k}') \,. \tag{A.131}$$

In the present discussion we realize that this implies a fixed Fourier convention, $B = 1/(2\pi)^3$, if we mean by the power spectrum really the Fourier transform of the two-point function; this is often not done correctly in the literature.

Consider the variance

$$\sigma^2_{\mathcal{R}} \equiv \langle \mathcal{R}^2(x) \rangle = \xi_{\mathcal{R}}(0) = B \int \mathrm{d}^3k\, P_{\mathcal{R}}(k) \,. \tag{A.132}$$

This is often defined as

$$\sigma^2_{\mathcal{R}} \equiv \int \mathrm{d} \ln k\, \Delta^2_{\mathcal{R}}(k) \,, \tag{A.133}$$

where

$$\Delta^2_{\mathcal{R}}(k) \equiv 4\pi B\, k^3 P_{\mathcal{R}}(k) \,. \tag{A.134}$$

In the common Fourier convention $B = 1/(2\pi)^3$ this becomes

$$\Delta^2_{\mathcal{R}}(k) \equiv \frac{k^3}{2\pi^2} P_{\mathcal{R}}(k) \,. \tag{A.135}$$

For other Fourier conventions the relation between $\Delta^2_{\mathcal{R}}(k)$ and $P_{\mathcal{R}}(k)$ will differ by a numerical factor.

Bispectrum

For Gaussian perturbations the power spectrum contains all the information (all higher-order correlation functions can be expressed in terms of the two-point function). Non-Gaussianity is measured by a non-zero three-point function, or equivalently in Fourier space the bispectrum

$$\langle \mathcal{R}_{\mathbf{k}_1} \mathcal{R}_{\mathbf{k}_2} \mathcal{R}_{\mathbf{k}_3} \rangle = (2\pi)^3 B_{\mathcal{R}}(k_1, k_2, k_3) \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \,. \tag{A.136}$$

7.8.2 Free Field Action for $\mathcal{R}$

In this appendix we compute the second-order action for the comoving curvature perturbation $\mathcal{R}$. This is a basic element for the quantization of cosmological scalar perturbations in **Lecture 2**.

We consider slow-roll models of inflation which are described by a canonical scalar field ϕ minimally coupled to gravity

$$S = \frac{1}{2} \int \mathrm{d}^4x \sqrt{-g}\, \left[R - (\nabla\phi)^2 - 2V(\phi) \right] \,, \tag{A.137}$$

in units where $M_{\rm pl}^{-2} \equiv 8\pi G = 1$. We will study perturbations of this action due to fluctuations in the scalar field $\delta\phi(t, x^i) \equiv \phi(t, x^i) - \bar{\phi}(t)$ and the metric. We will treat metric fluctuations in the ADM formalism (Arnowitt-Deser-Misner) [168].

Slow-Roll Background

We consider a flat background metric

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a(t)^2\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j = a^2(\tau)(-\mathrm{d}\tau^2 + \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j)\,, \tag{A.138}$$

with scale factor $a(t)$ and Hubble parameter $H(t) \equiv \partial_t \ln a$ satisfying the Friedmann Equations

$$3H^2 = \frac{1}{2}\dot{\phi}^2 + V(\phi)\,, \qquad \dot{H} = -\frac{1}{2}\dot{\phi}^2\,. \tag{A.139}$$

The scalar field satisfies the Klein-Gordon Equation

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = 0\,. \tag{A.140}$$

The standard slow-roll parameters are

$$\epsilon_{\rm v} = \frac{1}{2}\Big(\frac{V_{,\phi}}{V}\Big)^2 \approx \frac{1}{2}\frac{\dot{\phi}^2}{H^2}\,, \qquad \eta_{\rm v} = \frac{V_{,\phi\phi}}{V} \approx -\frac{\ddot{\phi}}{H\dot{\phi}} + \frac{1}{2}\frac{\dot{\phi}^2}{H^2}\,. \tag{A.141}$$

ADM Formalism

We treat fluctuations in the ADM formalism [168] where spacetime is sliced into three-dimensional hypersurfaces

$$\mathrm{d}s^2 = -N^2\mathrm{d}t^2 + g_{ij}(\mathrm{d}x^i + N^i\mathrm{d}t)(\mathrm{d}x^j + N^j\mathrm{d}t)\,. \tag{A.142}$$

Here, g_{ij} is the three-dimensional metric on slices of constant t. The lapse function $N(\mathbf{x})$ and the shift function $N_i(\mathbf{x})$ contain the same information as the metric perturbations Φ and B in (A.45). However, they were chosen in such a way that they appear as non-dynamical Lagrange multipliers in the action, *i.e.* their equations of motion are purely algebraic. The action (A.137) becomes

$$S = \frac{1}{2}\int \mathrm{d}^4x\sqrt{-g}\Big[NR^{(3)} - 2NV + N^{-1}(E_{ij}E^{ij} - E^2) + N^{-1}(\dot{\phi} - N^i\partial_i\phi)^2 - Ng^{ij}\partial_i\phi\partial_j\phi - 2V\Big]\,, \tag{A.143}$$

where

$$E_{ij} \equiv \frac{1}{2}(\dot{g}_{ij} - \nabla_i N_j - \nabla_j N_i)\,, \qquad E = E^i_i\,. \tag{A.144}$$

E_{ij} is related to the extrinsic curvature of the three-dimensional spatial slices $K_{ij} = N^{-1}E_{ij}$.

Exercise 14 (ADM Action)**.** *Confirm Eqn. (A.143).*

Comoving Gauge

To fix time and spatial reparameterizations we choose the following gauge for the dynamical fields g_{ij} and ϕ

$$\delta\phi = 0\,, \qquad g_{ij} = a^2[(1-2\mathcal{R})\delta_{ij} + h_{ij}]\,, \qquad \partial_i h_{ij} = h^i_i = 0\,. \tag{A.145}$$

In this gauge the inflaton field is unperturbed and all scalar degrees of freedom are parameterized by the metric fluctuation $\mathcal{R}(t,\mathbf{x})$. Geometrically, $\mathcal{R}$ measures the spatial curvature of constant-ϕ hypersurfaces, $R^{(3)} = 4\nabla^2\mathcal{R}/a^2$. An important property of $\mathcal{R}$ is that it remains constant outside the horizon. This allows us in **Lecture 2** to restrict our computation to correlation functions at horizon crossing.

Constraint Equations

The ADM action (A.143) implies the following constraint equations for the Lagrange multipliers N and N^i

$$\nabla_i[N^{-1}(E^i_j - \delta^i_j E)] = 0\,, \tag{A.146}$$
$$R^{(3)} - 2V - N^{-2}(E_{ij}E^{ij} - E^2) - N^{-2}\dot{\phi}^2 = 0\,. \tag{A.147}$$

Exercise 15 (Constraint Equations)**.** *Derive the constraint equations (A.146) and (A.147) from the ADM action (A.143).*

To solve the constraints, we split the shift vector N_i into irrotational (scalar) and incompressible (vector) parts

$$N_i \equiv \psi_{,i} + \tilde{N}_i\,, \quad \text{where} \quad \tilde{N}_{i,i} = 0\,, \tag{A.148}$$

and define the lapse perturbation as

$$N \equiv 1 + \alpha\,. \tag{A.149}$$

The quantities α, ψ and $\tilde{N}_i$ then admit expansions in powers of $\mathcal{R}$,

$$\begin{aligned} \alpha &= \alpha_1 + \alpha_2 + \dots\,, \\ \psi &= \psi_1 + \psi_2 + \dots\,, \\ \tilde{N}_i &= \tilde{N}_i^{(1)} + \tilde{N}_i^{(2)} + \dots\,, \end{aligned} \tag{A.150}$$

where, *e.g.* $\alpha_n = \mathcal{O}(\mathcal{R}^n)$. The constraint equations may then be set to zero order-by-order.

Exercise 16 (First-Order Solution of Constraint Equations)**.** *Show that at first order (A.147) implies*

$$\alpha_1 = \frac{\dot{\mathcal{R}}}{H}\,, \qquad \partial^2 \tilde{N}_i^{(1)} = 0\,. \tag{A.151}$$

With an appropriate choice of boundary conditions one may set $\tilde{N}_i^{(1)} \equiv 0$. Show that at first order Eqn. (A.146) implies

$$\psi_1 = -\frac{\mathcal{R}}{H} + \frac{a^2}{H}\epsilon_{\rm v}\,\partial^{-2}\dot{\mathcal{R}}\,, \tag{A.152}$$

where ∂^{-2} is defined via $\partial^{-2}(\partial^2\phi) = \phi$.

The Free Field Action

Substituting the first-order solutions for N and N_i back into the action, one finds the following second-order action [22]

$$\boxed{S_2 = \frac{1}{2}\int \mathrm{d}^4x\, a^3 \frac{\dot{\phi}^2}{H^2}\left[\dot{\mathcal{R}}^2 - a^{-2}(\partial_i\mathcal{R})^2\right]}\,. \tag{A.153}$$

Exercise 17 (Second-Order Action)**.** *Confirm Eqn. (A.153). Hint: use integration by parts and the background equations of motion.*

The quadratic action (A.153) for $\mathcal{R}$ is the main result of this appendix and forms the basis for the quantization of cosmological perturbations in **Lecture 2**.

7.8.3 A Brief Review of the In-In Formalism

The problem of computing correlation functions in cosmology differs in important ways from the corresponding analysis of quantum field theory applied to particle physics. In particle physics the central object is the S-matrix describing the transition probability for a state in the far past $|\psi\rangle$ to become some state $|\psi'\rangle$ in the far future, $\langle\psi'|S|\psi\rangle = \langle\psi'(+\infty)|\psi(-\infty)\rangle$. Imposing asymptotic conditions at very early *and* very late times makes sense in this case, since in Minkowski space, states are assumed to non-interacting in the far past and the far future, *i.e.* the asymptotic state are taken to be vacuum state of the free Hamiltonian H_0.

In cosmology, however, we evaluate the expectation values of products of fields *at a fixed time*. Conditions are *not* imposed on the fields at both very early and very late times, but only at very early times, when the wavelength is deep inside the horizon. As we argued in **Lecture 2**, in this limit (according to the equivalence principle) the interaction picture fields should have the same firm as in Minkowski space. This lead us to the definition of the Bunch-Davies vacuum (the free vacuum in Minkowski space).

In this appendix we describe the Schwinger-Keldysh "in-in" formalism [86] to compute cosmological correlation functions. After pioneering work by Calzetta and Hu [87] and Jordan [88] the application of the "in-in" formalism to cosmological problems was recently revived by Maldacena [22] and Weinberg [89] (see also [169, 170]).

Time Evolution in the Interaction Picture

To describe the time evolution of cosmological perturbations we split the Hamiltonian into a free part and an interacting part

$$\boxed{H = H_0 + H_{\rm int}}\,. \tag{A.154}$$

The free-field Hamiltonian H_0 is quadratic in perturbations. Quadratic order was sufficient to compute the two-point correlations of **Lecture 2**. However, the higher-order correlations that concerned us in our study of non-Gaussianity in **Lecture 4** require going beyond quadratic order and defining the interaction Hamiltonian $H_{\rm int}$. The interaction Hamiltonian defines the evolution of states via the well-known time-evolution operator

$$U(\tau_2,\tau_1) = T\exp\left(-i\int_{\tau_1}^{\tau_2}{\rm d}\tau' H_{\rm int}(\tau')\right)\,, \tag{A.155}$$

where T denotes the time-ordering operator. The time-evolution operator U may be used to relate the interacting vacuum at arbitrary time $|\Omega(\tau)\rangle$ to the free (Bunch-Davies) vacuum $|0\rangle$. We first expand $\Omega(\tau)$ in eigenstates of the free Hamiltonian,

$$|\Omega\rangle = \sum_n |n\rangle\langle n|\Omega(\tau)\rangle\,. \tag{A.156}$$

Then we evolve $|\Omega(\tau)\rangle$ as

$$|\Omega(\tau_2)\rangle = U(\tau_2,\tau_1)|\Omega(\tau_1)\rangle = |0\rangle\langle 0|\Omega\rangle + \sum_{n\geq 1} e^{+iE_n(\tau_2-\tau_1)}|n\rangle\langle n|\Omega(\tau_1)\rangle\,. \tag{A.157}$$

$|{\rm in}\rangle$ Vacuum

From Eqn. (A.157) we see that the choice $\tau_2 = -\infty(1-i\epsilon)$ projects out all excited states. Hence, we have the following relation between the interacting vacuum at $\tau = -\infty(1-i\epsilon)$ and the free vacuum $|0\rangle$

$$|\Omega(-\infty(1-i\epsilon))\rangle = |0\rangle\langle 0|\Omega\rangle\,. \tag{A.158}$$

Finally, the interacting vacuum at an arbitrary time τ is

$$|\text{in}\rangle \equiv |\Omega(\tau)\rangle = U(\tau, -\infty(1-i\epsilon))|\Omega(-\infty(1-i\epsilon))\rangle \tag{A.159}$$

$$= T\exp\left(-i\int_{-\infty(1-i\epsilon)}^{\tau} \mathrm{d}\tau' H_{\rm int}(\tau')\right)|0\rangle\langle 0|\Omega\rangle\,. \tag{A.160}$$

Expectation Values

In the "in-in" formalism, the expectation value $\langle W(\tau)\rangle$, of a product of operators $W(\tau)$ at time τ, is evaluated as[46]

$$\langle W(\tau)\rangle \equiv \frac{\langle \text{in}|W(\tau)|\text{in}\rangle}{\langle \text{in}|\text{in}\rangle} \tag{A.161}$$

$$= \left\langle 0\right|\left(\bar{T}e^{-i\int_{-\infty^+}^{\tau} H_{\rm int}(\tau')\mathrm{d}\tau'}\right)^{\dagger} W(\tau)\left(Te^{-i\int_{-\infty^+}^{\tau} H_{\rm int}(\tau'')\mathrm{d}\tau''}\right)\left|0\right\rangle, \tag{A.162}$$

or

$$\boxed{\langle W(\tau)\rangle = \left\langle 0\right|\left(\bar{T}e^{-i\int_{-\infty^-}^{\tau} H_{\rm int}(\tau')\mathrm{d}\tau'}\right) W(\tau)\left(Te^{-i\int_{-\infty^+}^{\tau} H_{\rm int}(\tau'')\mathrm{d}\tau''}\right)\left|0\right\rangle,} \tag{A.163}$$

where we defined the anti-time-ordering operator $\bar{T}$ and the notation $-\infty^{\pm} \equiv -\infty(1\mp i\epsilon)$. This definition of the correlation functions $\langle W(\tau)\rangle$ in terms of the interaction Hamiltonian $H_{\rm int}$ is the main result of the "in-in" formalism. The interaction Hamiltonian is computed in the ADM approach to General Relativity [22] and $\langle W(\tau)\rangle$ is then evaluated perturbatively.

In **Lecture 4** this formalism was implicitly used to compute the three-point functions for various inflationary models,

$$\boxed{\langle \mathcal{R}_{\mathbf{k}_1}\mathcal{R}_{\mathbf{k}_2}\mathcal{R}_{\mathbf{k}_3}\rangle(\tau) = \left\langle 0\right|\left(\bar{T}e^{-i\int_{-\infty^-}^{\tau} H_{\rm int}(\tau')\mathrm{d}\tau'}\right)\mathcal{R}_{\mathbf{k}_1}(\tau)\mathcal{R}_{\mathbf{k}_2}(\tau)\mathcal{R}_{\mathbf{k}_3}(\tau)\left(Te^{-i\int_{-\infty^+}^{\tau} H_{\rm int}(\tau'')\mathrm{d}\tau''}\right)\left|0\right\rangle.} \tag{A.164}$$

Interaction Hamiltonian

Let us sketch how the interaction Hamiltonian is computed:[47] The inflationary action is expanded perturbatively

$$S = S_0[\bar{\phi}, \bar{g}_{\mu\nu}] + S_2[\mathcal{R}^2] + S_3[\mathcal{R}^3] + \cdots\,. \tag{A.165}$$

Here, we have defined a background part S_0, a quadratic free-field part S_2 and a non-linear interaction term S_3. The background action S_0 defines the Hubble parameter H and the slow-roll parameters ε and η. The free-field action S_2 defines the time-evolution of the mode functions $\mathcal{R}(\tau)$ in the interaction picture (often denoted by $\mathcal{R}_{\rm I}(\tau)$). The non-linear part of the action defines the interaction Hamiltonian, *e.g.* at cubic order $S_3 = -\int \mathrm{d}\tau H_{\rm int}(\mathcal{R}_{\rm I})$. Schematically, the interaction Hamiltonian takes the following form

$$H_{\rm int} = \sum_i f_i(\varepsilon, \eta, \ldots)\mathcal{R}_{\rm I}^3(\tau)\,. \tag{A.166}$$

[46] For a derivation of this result see Weinberg [89].

[47] For a sample calculation that shows the full (painful) details see Maldacena [22].

Perturbative Expansion

In **Lecture 2** we defined the expansion of the operator corresponding to the Mukhanov variable, $v = 2a^2\varepsilon\,\mathcal{R}$, in terms of creation and annihilation operators

$$\hat{v}_{\mathbf{k}}(\tau) = v_k(\tau)\hat{a}_{\mathbf{k}} + v_k^*(\tau)\hat{a}^\dagger_{-\mathbf{k}}\,. \tag{A.167}$$

The mode functions $v_k(\tau)$ were defined uniquely by initial state boundary conditions when all modes were deep inside the horizon

$$v_k(\tau) = \frac{e^{-ik\tau}}{\sqrt{2k}}\left(1 - \frac{i}{k\tau}\right)\,. \tag{A.168}$$

The free two-point correlation function is

$$\langle 0|\hat{v}_{\mathbf{k}_1}(\tau_1)\hat{v}_{\mathbf{k}_2}(\tau_2)|0\rangle = (2\pi)^3\delta(\mathbf{k}_1 + \mathbf{k}_2)G_{k_1}(\tau_1,\tau_2)\,, \tag{A.169}$$

with

$$G_{k_1}(\tau_1,\tau_2) \equiv v_k(\tau_1)v_k^*(\tau_2)\,. \tag{A.170}$$

Expansion of Eqn. (A.164) in powers of $H_{\rm int}$ gives:

- at zeroth order
$$\langle W(\tau)\rangle^{(0)} = \langle 0|W(\tau)|0\rangle\,, \tag{A.171}$$
where $W(\tau) \equiv \mathcal{R}_{\mathbf{k}_1}(\tau)\mathcal{R}_{\mathbf{k}_2}(\tau)\mathcal{R}_{\mathbf{k}_3}(\tau)$.
- at first order
$$\langle W(\tau)\rangle^{(1)} = 2\,\mathrm{Re}\left[-i\int_{-\infty^+}^{\tau}\mathrm{d}\tau'\langle 0|W(\tau)H_{\rm int}(\tau')|0\rangle\right]\,. \tag{A.172}$$
- at second order
$$\begin{aligned}\langle W(\tau)\rangle^{(2)} \;=\; & -2\,\mathrm{Re}\left[\int_{-\infty^+}^{\tau}\mathrm{d}\tau'\int_{-\infty^+}^{\tau'}\mathrm{d}\tau''\langle 0|W(\tau)H_{\rm int}(\tau')H_{\rm int}(\tau'')|0\rangle\right] \\ & + \int_{-\infty^-}^{\tau}\mathrm{d}\tau'\int_{-\infty^+}^{\tau}\mathrm{d}\tau''\langle 0|H_{\rm int}(\tau')W(\tau)H_{\rm int}(\tau'')|0\rangle\,.\end{aligned} \tag{A.173}$$

In the bispectrum calculations of **Lecture 2** the zeroth-order term (A.171) vanishes for Gaussian initial conditions. The leading result therefore comes from Eqn. (A.172). Evaluating Eqn. (A.172) makes use of Wick's theorem to expresses the result as products of two-point functions (A.170).

7.8.4 Slow-Roll Inflation in the Hamilton-Jacobi Approach

In these lectures we have defined *exact* slow-roll conditions via the parameters

$$\varepsilon = -\frac{\dot{H}}{H^2}\,, \qquad \eta = -\frac{\ddot{\phi}}{H\dot{\phi}}\,, \tag{B.1}$$

and *approximate* conditions via

$$\epsilon_{\rm v} = \frac{M_{\rm pl}^2}{2}\left(\frac{V_{,\phi}}{V}\right)^2\,, \qquad \eta_{\rm v} = M_{\rm pl}^2\frac{V_{,\phi\phi}}{V}\,. \tag{B.2}$$

In this appendix we explore their relationship in more detail.

Hamilton-Jacobi Formalism

The Hamilton-Jacobi approach treats the Hubble expansion rate $H(\phi) = \mathcal{H}/a$ as the fundamental quantity, considered as a function of time. Consider

$$H_{,\phi} = \frac{H'}{\phi'} = \frac{-(\mathcal{H}^2 - \mathcal{H}')/a}{\phi'} = -\frac{\phi'}{2a}\,, \tag{B.3}$$

where we used $\mathcal{H}^2 - \mathcal{H}' = a^2(\rho + p)/2 = (\phi')^2/2$ and primes are derivatives with respect to conformal time. This gives the master equation

$$\boxed{\frac{d\phi}{dt} = \frac{\phi'}{a} = -2H_{,\phi}}\,. \tag{B.4}$$

This allows us to rewrite the Friedmann Equation

$$H^2 = \frac{1}{3}\left[\frac{1}{2}\left(\frac{d\phi}{dt}\right)^2 + V(\phi)\right] \tag{B.5}$$

in the following way

$$\boxed{[H_{,\phi}]^2 - \frac{3}{2}H^2 = -\frac{1}{2}V(\phi)}\,. \tag{B.6}$$

Notice the following important consequence of the Hamilton-Jacobi Equation (B.6): For any specified function $H(\phi)$, it produces a potential $V(\phi)$ which admits the given $H(\phi)$ as an exact inflationary solution. Integrating Eqn. (B.4)

$$\int \mathrm{d}t = -\frac{1}{2}\int \frac{\mathrm{d}\phi}{H'(\phi)} \tag{B.7}$$

relates ϕ to proper time t. This enables us to obtain $H(t)$, which can be integrated to give $a(t)$. The Hamilton-Jacobi formalism can therefore be used to generate infinitely many inflationary models with exactly known analytic solutions for the background expansion. However, here we are more concerned with the fact that it allows an elegant and intuitive definition of the slow-roll parameters.

Hubble Slow-Roll Parameters

During slow-roll inflation the background spacetime is approximately de Sitter. Any deviation of the background equation of state

$$w = \frac{p}{\rho} = \frac{(\phi')^2/2a^2 - V}{(\phi')^2/2a^2 + V} \tag{B.8}$$

from the perfect de Sitter limit $w = -1$ may be defined by the parameter

$$\boxed{\varepsilon \equiv \frac{3}{2}(1 + w)}\,. \tag{B.9}$$

We can express the Friedmann Equations

$$\mathcal{H}^2 = \frac{1}{3}a^2\rho \tag{B.10}$$

$$\mathcal{H}' = -\frac{1}{6}a^2(\rho + 3p) \tag{B.11}$$

in terms of ε

$$\mathcal{H}^2 = \frac{1}{3}\frac{(\phi')^2}{\varepsilon} \tag{B.12}$$

$$\mathcal{H}' = \mathcal{H}^2(1-\varepsilon)\,. \tag{B.13}$$

Hence,

$$\varepsilon = 1 - \frac{\mathcal{H}'}{\mathcal{H}^2} = \frac{d(H^{-1})}{dt} = -\frac{\dot{H}}{H^2}\,. \tag{B.14}$$

Note that this can be interpreted as the rate ot change of the Hubble parameter H with respect to the number of e-foldings $\mathrm{d}N = H\mathrm{d}t = -\frac{1}{2}\frac{H(\phi)}{H_{,\phi}}\mathrm{d}\phi$

$$\boxed{\varepsilon = -\frac{d\ln H}{dN} = 2\left(\frac{H_{,\phi}}{H}\right)^2}\,. \tag{B.15}$$

Analogously we define the second slow-roll parameter as the rate of change of $H_{,\phi}$

$$\boxed{\eta = -\frac{d\ln|H_{,\phi}|}{dN} = 2\frac{H_{,\phi\phi}}{H}}\,. \tag{B.16}$$

Using Eqn. (B.4) this is also

$$\eta = \frac{d\ln|\dot{\phi}|}{dN}\,. \tag{B.17}$$

Slow-Roll Inflation

By definition, slow-roll corresponds to a regime where all dynamical characteristics of the universe, measured in physical (proper) units, change little over a single e-folding of expansion. This ensures that the primordial perturbations are generated with approximately equal power on all scales, leading to a scale-invariant perturbation spectrum.

Since ε and η characterize the rate of change of H and $H_{,\phi}$ with e-foldings, slow-roll is naturally defined by

$$\varepsilon \ll 1\,, \quad |\eta| \ll 1\,. \tag{B.18}$$

The first slow-roll condition implies

$$\varepsilon \ll 1 \qquad \Rightarrow \qquad \mathcal{H}^2 = \frac{1}{3}\frac{(\phi')^2}{\varepsilon} \gg (\phi')^2\,, \tag{B.19}$$

so that the slow-roll limit of the first Friedmann Equation is

$$\mathcal{H}^2 \approx \frac{1}{3}a^2V\,. \tag{B.20}$$

The second slow-roll condition implies

$$\eta = \frac{d\ln|\dot{\phi}|]}{dN} = \frac{\ddot{\phi}}{H|\dot{\phi}|} \ll 1 \qquad \Rightarrow \qquad |\ddot{\phi}| \ll H|\dot{\phi}|\,, \tag{B.21}$$

so that the Klein-Gordon Equation reduces to

$$\dot{\phi} \approx -\frac{a^2V'}{3\mathcal{H}}\,. \tag{B.22}$$

In **Lecture 1** we defined a second set of common slow-roll parameters in terms of the local shape of the potential $V(\phi)$

$$\epsilon_{\rm v} \equiv \frac{1}{2}\left(\frac{V_{,\phi}}{V}\right)^2 \tag{B.23}$$

$$\eta_{\rm v} \equiv \frac{V_{,\phi\phi}}{V}\,. \tag{B.24}$$

We note that $\varepsilon(\phi_{\rm end}) \equiv 1$ is an exact definition of the end of inflation, while $\epsilon_{\rm v}(\phi_{\rm end}) = 1$ is only an approximation. In the slow-roll regime the following relations hold

$$\varepsilon \approx \epsilon_{\rm v} \tag{B.25}$$

$$\eta \approx \eta_{\rm v} - \epsilon_{\rm v}\,. \tag{B.26}$$

Inflationary Attractor Solution

We now show that the slow-roll condition, $\varepsilon < 1$, also implies that inflation is an attractor solution.

Let $\bar{H}(\phi)$ be a solution of the Hamilton-Jacobi Equation (B.6) (at this point we don't demand that this is an inflationary solution). Now consider a small perturbation $\delta H(\phi)$, *i.e.*

$$H(\phi) = \bar{H}(\phi) + \delta H(\phi)\,. \tag{B.27}$$

We linearize Eqn. (B.6) to find

$$\bar{H}_{,\phi}\,\delta H_{,\phi} \approx \frac{3}{2}\bar{H}\delta H\,, \tag{B.28}$$

or

$$\frac{d}{d\phi}(\ln \delta H) = \frac{3}{2}\frac{\bar{H}}{\bar{H}_{,\phi}}\,. \tag{B.29}$$

This has the solution

$$\delta H(\phi) = \delta H(\phi_{\rm i}) \exp\left[\frac{3}{2}\int_{\phi_{\rm i}}^{\phi} \frac{\bar{H}}{\bar{H}_{,\phi}} {\rm d}\phi\right]\,. \tag{B.30}$$

Recalling that

$${\rm d}N = -\frac{1}{2}\frac{H}{H_{,\phi}}{\rm d}\phi = \frac{|{\rm d}\phi|}{\sqrt{2\varepsilon}} > 0\,, \tag{B.31}$$

this may be written as

$$\delta H(\phi) = \delta H(\phi_i) \exp\left[-3(N - N_i)\right]\,. \tag{B.32}$$

During inflation, $\varepsilon < 1$, the number of e-folds of expansion N rapidly becomes large and any perturbation to the inflationary solution δH gets diluted exponentially. $H(\phi)$ then approaches $\bar{H}(\phi)$.

8 General Fundamental Questions

8.1 What Cosmology Teaches Us About Our Life?

(??? actually, this is a very important question. it is better to write an answer to it, and not pretend "oh, we don't want even to ask this question". I'll write some ideas later)

8.2 What Cosmology Teaches Us About Physical Nature of the Word?

(also, it is a good question, and these ideas will tell us something about)

9 Alternative Fundamental Ideas

(not valuable - at the end of this note. here, there will be only very fundamentally valuable ones.)

9.0.1 Dicke's Explanation of the Redshift

(????вопрос - верная эта теория или нет? подумаю, хороший вопрос)

10 Cosmology by Weinberg

10.1 The Expansion of the Universe

The visible universe seems the same in all directions around us, at least if we look out to distances larger than about 300 million light years. [1] The isotropy is much more precise (to about one part in 10^{-5}) in the cosmic microwave background, to be discussed in Chapters 2 and 7. As we will see there, this radiation has been traveling to us for about 14 billion years, supporting the conclusion that the universe at sufficiently large distances is nearly the same in all directions.

It is difficult to imagine that we are in any special position in the universe, so we are led to conclude that the universe should appear isotropic to observers throughout the universe. But not to all observers. The universe does not seem at all isotropic to observers in a spacecraft whizzing through our galaxy at half the speed of light. Such observers will see starlight and the cosmic microwave radiation background coming toward them from the direction toward which they are moving with much higher intensity than from behind. In formulating the assumption of isotropy, one should specify that the universe seems the same in all directions to a family of "typical" freely falling observers: those that move with the average velocity of typical galaxies in their respective neighborhoods. That is, conditions must be the same at the same time (with a suitable definition of time) at any points that can be carried into each other by a rotation about any typical galaxy. But any point can be carried into any other by a sequence of such rotations about various typical galaxies, so the universe is then also homogeneous - observers in all typical galaxies at the same time see conditions pretty much the same. [2]

The assumption that the universe is isotropic and homogeneous will lead us in Section 1.1 to choose the spacetime coordinate system so that the metric takes a simple form, first worked out by Friedmann [3] as a solution of the Einstein field equations, and then derived on the basis of isotropy and homogeneity alone by Robertson [4] and Walker. [5] Almost all of modern cosmology is based on this Robertson-Walker metric, at least as a first

[471] K. K. S. Wu, O. Lahav, and M. J. Rees, Nature 397, 225 (January 21, 1999). For a contrary view, see P. H. Coleman, L. Pietronero, and R. H. Sanders, Astron. Astrophys. 200, L32 (1988): L. Pietronero, M. Montuori, and F. Sylos-Labini, in Critical Dialogues in Cosmology, (World Scientific, Singapore, 1997): 24; F. Sylos-Labini, F. Montuori, and L. Pietronero, Phys. Rep. 293, 61 (1998).

[2] The Sloan Digital Sky Survey provides evidence that the distribution of galaxies is homogeneous on scales larger than about 300 light years; see J. Yadav, S. Bharadwaj, B. Pandey, and T. R. Seshadri, Mon. Not. Roy. Astron. Soc. 364, 601 (2005) [astro-ph/0504315].

[3] A. Friedmann, Z. Phys. 10, 377 (1922); ibid. 21, 326 (1924).

approximation. The observational implications of these assumptions are discussed in Sections 1.2-1.4, without reference to any dynamical assumptions. The Einstein field equations are applied to the Robertson-Walker metric in Section 1.5, and their consequences are then explored in Sections 1.6-1.13.

10.1.1 1.1 Spacetime geometry

As preparation for working out the spacetime metric, we first consider the geometry of a three-dimensional homogeneous and isotropic space. As discussed in Appendix B, geometry is encoded in a metric $g_{ij}(\mathbf{x})$ (with i and j running over the three coordinate directions), or equivalently in a line element $ds^2 \equiv g_{ij}dx^i dx^j$, with summation over repeated indices understood. (We say that ds is the proper distance between $\mathbf{x}$ and $\mathbf{x} + d\mathbf{x}$, meaning that it is the distance measured by a surveyor who uses a coordinate system that is Cartesian in a small neighborhood of the point $\mathbf{x}$.) One obvious homogeneous isotropic three-dimensional space with positive definite lengths is flat space, with line element

$$ds^2 = d\mathbf{x}^2. \tag{1.1.1}$$

The coordinate transformations that leave this invariant are here simply ordinary three-dimensional rotations and translations. Another fairly obvious possibility is a spherical surface in four-dimensional Euclidean space with some radius a, with line element

$$ds^2 = d\mathbf{x}^2 + dz^2, \quad z^2 + \mathbf{x}^2 = a^2. \tag{1.1.2}$$

Here the transformations that leave the line element invariant are fourdimensional rotations; the direction of $\mathbf{x}$ can be changed to any other direction by a four-dimensional rotation that leaves z unchanged (that is, an ordinary three-dimensional rotation), while $\mathbf{x}$ can be carried into any other point by a four-dimensional rotation that does change z. It can be proved [6] that the only other possibility (up to a coordinate transformation) is a hyperspherical surface in four-dimensional pseudo-Euclidean space, with line element

$$ds^2 = d\mathbf{x}^2 - dz^2, \quad z^2 - \mathbf{x}^2 = a^2, \tag{1.1.3}$$

where a^2 is (so far) an arbitrary positive constant. The coordinate transformations that leave this invariant are four-dimensional pseudo-rotations, just like Lorentz transformations, but with z instead of time.

We can rescale coordinates

$$\mathbf{x}' \equiv a\mathbf{x}, \quad z' \equiv az \tag{1.1.4}$$

Dropping primes, the line elements in the spherical and hyperspherical cases are

$$ds^2 = a^2\left[d\mathbf{x}^2 \pm dz^2\right], \quad z^2 \pm \mathbf{x}^2 = 1 \tag{1.1.5}$$

The differential of the equation $z^2 \pm \mathbf{x}^2 = 1$ gives $zdz = \mp\mathbf{x}\cdot d\mathbf{x}$ so

$$ds^2 = a^2\left[d\mathbf{x}^2 \pm \frac{(\mathbf{x}\cdot d\mathbf{x})^2}{1 \mp \mathbf{x}^2}\right] \tag{1.1.6}$$

We can extend this to the case of Euclidean space by writing it as

[4] H. P. Robertson, Astrophys. J. 82, 284 (1935); ibid., 83, 187, 257 (1936).
[5] A. G. Walker, Proc. Lond. Math. Soc. (2) 42, 90 (1936).
[476] See S. Weinberg, Gravitation and Cosmology (John Wiley & Sons, New York, 1972) [quoted below as G&Cl, Sec. 13.2.

$$ds^2 = a^2 \left[d\mathbf{x}^2 + K \frac{(\mathbf{x} \cdot d\mathbf{x})^2}{1 - K\mathbf{x}^2} \right] \tag{1.1.7}$$

where

$$K = \begin{cases} +1 & \text{spherical} \\ -1 & \text{hyperspherical} \\ 0 & \text{Euclidean} \end{cases} \tag{1.1.8}$$

(The constant K is often written as k, but we will use upper case for this constant throughout this book to avoid confusion with the symbols for wave number or for a running spatial coordinate index.) Note that we must take $a^2 > 0$ in order to have ds^2 positive at $\mathbf{x} = 0$, and hence everywhere.

There is an obvious way to extend this to the geometry of spacetime: just include a term (1.1.7) in the spacetime line element, with a now an arbitrary function of time (known as the Robertson-Walker scale factor):

$$d\tau^2 \equiv -g_{\mu\nu}(x) dx^\mu dx^\nu = dt^2 - a^2(t) \left[d\mathbf{x}^2 + K \frac{(\mathbf{x} \cdot d\mathbf{x})^2}{1 - K\mathbf{x}^2} \right] \tag{1.1.9}$$

Another theorem [7] tells us that this is the unique metric (up to a coordinate transformation) if the universe appears spherically symmetric and isotropic to a set of freely falling observers, such as astronomers in typical galaxies. The components of the metric in these coordinates are:

$$g_{ij} = a^2(t) \left(\delta_{ij} + K \frac{x^i x^j}{1 - K\mathbf{x}^2} \right), \quad g_{i0} = 0, \quad g_{00} = -1 \tag{1.1.10}$$

with i and j running over the values 1,2 , and 3 , and with $x^0 \equiv t$ the time coordinate in our units, with the speed of light equal to unity. Instead of the quasi-Cartesian coordinates x^i, we can use spherical polar coordinates, for which

$$d\mathbf{x}^2 = dr^2 + r^2 d\Omega, \quad d\Omega \equiv d\theta^2 + \sin^2\theta d\phi^2$$

so

$$d\tau^2 = dt^2 - a^2(t) \left[\frac{dr^2}{1 - Kr^2} + r^2 d\Omega \right] \tag{1.1.11}$$

in which case the metric becomes diagonal, with

$$g_{rr} = \frac{a^2(t)}{1 - Kr^2}, \quad g_{\theta\theta} = a^2(t) r^2, \quad g_{\phi\phi} = a^2(t) r^2 \sin^2\theta, \quad g_{00} = -1 \tag{1.1.12}$$

We will see in Section 1.5 that the dynamical equations of cosmology depend on the overall normalization of the function $a(t)$ only through a term $K/a^2(t)$, so for $K = 0$ this normalization has no significance; all that matters are the ratios of the values of $a(t)$ at different times.

The equation of motion of freely falling particles is given in Appendix B by Eq. (B.12):

$$\frac{d^2 x^\mu}{du^2} + \Gamma^\mu_{\nu\kappa} \frac{dx^\nu}{du} \frac{dx^\kappa}{du} = 0 \tag{1.1.13}$$

where $\Gamma^\mu_{\nu\kappa}$ is the affine connection, given in Appendix B by Eq. (B.13),

$$\Gamma^\mu_{\nu\kappa} = \frac{1}{2} g^{\mu\lambda} \left[\frac{\partial g_{\lambda\nu}}{\partial x^\kappa} + \frac{\partial g_{\lambda\kappa}}{\partial x^\nu} - \frac{\partial g_{\nu\kappa}}{\partial x^\lambda} \right] \tag{1.1.14}$$

[477] G&C, Sec. 13.5.

and u is a suitable variable parameterizing positions along the spacetime curve, proportional to τ for massive particles. (A spacetime path $x^\mu = x^\mu(u)$ satisfying Eq. (1.1.13) is said to be a geodesic, meaning that the integral $\int d\tau$ is stationary under any infinitesimal variation of the path that leaves the endpoints fixed.) Note in particular that the derivatives $\partial_i g_{00}$ and $\dot{g}_{0i}$ vanish, so $\Gamma^i_{00} = 0$. A particle at rest in these coordinates will therefore stay at rest, so these are co-moving coordinates, which follow the motion of typical observers. Because $g_{00} = -1$, the proper time interval $(-g_{\mu\nu}dx^\mu dx^\nu)^{1/2}$ for a co-moving clock is just dt, so t is the time measured in the rest frame of a co-moving clock.

The meaning of the Robertson-Walker scale factor $a(t)$ can be clarified by calculating the proper distance at time t from the origin to a co-moving
object at radial coordinate r :

$$d(r,t) = a(t)\int_0^r \frac{dr}{\sqrt{1-Kr^2}} = a(t)\times\begin{cases}\sin^{-1} r & K=+1\\ \sinh^{-1} r & K=-1\\ r & K=0\end{cases} \tag{1.1.15}$$

In this coordinate system a co-moving object has r time-independent, so the proper distance from us to a co-moving object increases (or decreases) with $a(t)$. Since there is nothing special about our own position, the proper distance between any two co-moving observers anywhere in the universe must also be proportional to $a(t)$. The rate of change of any such proper distance $d(t)$ is just

$$\dot{d} = d\dot{a}/a \tag{1.1.16}$$

We will see in the following section that in fact $a(t)$ is increasing.
We also need the non-zero components of the affine connection, given by Eq. (1.1.14) as:

$$\begin{aligned}\Gamma^0_{ij} =& -\frac{1}{2}\left(g_{0i,j}+g_{0j,i}-g_{ij,0}\right) = a\dot{a}\left(\delta_{ij}+K\frac{x^i x^j}{1-K\mathbf{x}^2}\right)\\ =& \, a\dot{a}\tilde{g}_{ij} \qquad (1.1.17)\\ \Gamma^i_{0j} =& \frac{1}{2}g^{il}\left(g_{l0,j}+g_{lj,0}-g_{0j,l}\right) = \frac{\dot{a}}{a}\delta_{ij} \qquad (1.1.18)\\ \Gamma^i_{jl} =& \frac{1}{2}\tilde{g}^{im}\left(\frac{\partial\tilde{g}_{jm}}{\partial x^l}+\frac{\partial\tilde{g}_{lm}}{\partial x^j}-\frac{\partial\tilde{g}_{jl}}{\partial x^m}\right)\equiv\tilde{\Gamma}^i_{jl} \qquad (1.1.19)\end{aligned}$$

Here $\tilde{g}_{ij}$ and $\tilde{\Gamma}^i_{jl}$ are the purely spatial metric and affine connection, and $\tilde{g}^{ij}$ is the reciprocal of the 3×3 matrix $\tilde{g}_{ij}$, which in general is different from the ij component of the reciprocal of the 4×4 matrix $g_{\mu\nu}$. In quasi-Cartesian coordinates,

$$\tilde{g}_{ij} = \delta_{ij}+K\frac{x^i x^j}{1-K\mathbf{x}^2}, \quad \tilde{\Gamma}^i_{jl} = K\tilde{g}_{jl}x^i \tag{1.1.20}$$

We can use these components of the affine connection to find the motion of a particle that is not at rest in the co-moving coordinate system. First, let's calculate the rate of change of the momentum of a particle of non-zero mass m_0. Consider the quantity

$$P \equiv m_0\sqrt{g_{ij}\frac{dx^i}{d\tau}\frac{dx^j}{d\tau}} \tag{1.1.21}$$

where $d\tau^2 = dt^2 - g_{ij}dx^i dx^j$. In a locally inertial Cartesian coordinate system, for which $g_{ij} = \delta_{ij}$, we have $d\tau = dt\sqrt{1-\mathbf{v}^2}$ where $v^i = dx^i/dt$,
so Eq. (1.1.21) is the formula given by special relativity for the magnitude of the momentum.

On the other hand, the quantity (1.1.21) is evidently invariant under arbitrary changes in the spatial coordinates, so we can evaluate it just as well in co-moving Robertson-Walker coordinates. This can be done directly, using Eq. (1.1.13), but to save work, suppose we adopt a spatial coordinate system in which the particle position is near the origin $x^i = 0$, where $\tilde{g}_{ij} = \delta_{ij} + O(\mathbf{x}^2)$, and we can therefore ignore the purely spatial components Γ^i_{jk} of the affine connection. General relativity gives the equation of motion

$$\frac{d^2x^i}{d\tau^2} = -\Gamma^i_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = -\frac{2}{a}\frac{da}{dt}\frac{dx^i}{d\tau}\frac{dt}{d\tau}$$

Multiplying with $d\tau/dt$ gives

$$\frac{d}{dt}\left(\frac{dx^i}{d\tau}\right) = -\frac{2}{a}\frac{da}{dt}\frac{dx^i}{d\tau}$$

whose solution is

$$\frac{dx^i}{d\tau} \propto \frac{1}{a^2(t)} \tag{1.1.22}$$

Using this in Eq. (1.1.21) with a metric $g_{ij} = a^2(t)\delta_{ij}$, we see that

$$P(t) \propto 1/a(t) \tag{1.1.23}$$

This holds for any non-zero mass, however small it may be compared to the momentum. Hence, although for photons both m_0 and $d\tau$ vanish, Eq. (1.1.23) is still valid.

It is important to characterize the paths of photons and material particles in interpreting astronomical observations (especially of gravitational lenses, in Chapter 9). Photons and particles passing through the origin of our spatial coordinate system obviously travel on straight lines in this coordinate system, which are spatial geodesics, curves that satisfy the condition

$$\frac{d^2x^i}{ds^2} + \tilde{\Gamma}^i_{jl}\frac{dx^j}{ds}\frac{dx^l}{ds} = 0 \tag{1.1.24}$$

where ds is the three-dimensional proper length

$$ds^2 \equiv \tilde{g}_{ij}dx^i dx^j \tag{1.1.25}$$

But the property of being a geodesic is invariant under coordinate transformations (since it states the vanishing of a vector), so the path of the photon
or particle will also be a spatial geodesic in any spatial coordinate system, including those in which the photon or particle's path does not pass through the origin. (This can be seen in detail as follows. Using Eqs. (1.1.17) and (1.1.18), the equations of motion (1.1.13) of a photon or material particle are

$$\begin{aligned} 0 &= \tfrac{d^2x^i}{du^2} + \Gamma^i_{jl}\tfrac{dx^j}{du}\tfrac{dx^l}{du} + \tfrac{2\dot{a}}{a}\tfrac{dx^i}{du}\tfrac{dt}{du} \\ 0 &= \tfrac{d^2t}{du^2} + a\dot{a}\tilde{g}_{ij}\tfrac{dx^i}{du}\tfrac{dx^j}{du} \end{aligned} \tag{1.1.27}$$

Eq. (1.1.26) can be written

$$0 = \left(\frac{ds}{du}\right)^2\left[\frac{d^2x^i}{ds^2} + \Gamma^i_{jl}\frac{dx^j}{ds}\frac{dx^l}{ds}\right] + \left[\frac{d^2s}{du^2} + \frac{2\dot{a}}{a}\frac{dt}{du}\frac{ds}{du}\right]\frac{dx^i}{ds} \tag{1.1.28}$$

where s is so far arbitrary. If we take s to be the proper length (1.1.25) in the spatial geometry, then as we have seen

$$du^2 \propto d\tau^2 \propto dt^2 - a^2ds^2$$

Dividing by du^2, differentiating with respect to u, and using Eq. (1.1.27) shows that

$$\frac{d^2s}{du^2} + \frac{2\dot{a}}{a}\frac{dt}{du}\frac{ds}{du} = 0$$

so that Eq. (1.1.28) gives Eq. (1.1.24).)

There are various smoothed-out vector and tensor fields, like the current of galaxies and the energy-momentum tensor, whose mean values satisfy the requirements of isotropy and homogeneity. Isotropy requires that the mean value of any three-vector v^i must vanish, and homogeneity requires the mean value of any three-scalar (that is, a quantity invariant under purely spatial coordinate transformations) to be a function only of time, so the current of galaxies, baryons, etc. has components

$$J^i = 0, \quad J^0 = n(t) \tag{1.1.29}$$

with $n(t)$ the number of galaxies, baryons, etc. per proper volume in a comoving frame of reference. If this is conserved, in the sense of Eq. (B.38), then

$$0 = J^\mu_{;\mu} = \frac{\partial J^\mu}{\partial x^\mu} + \Gamma^\mu_{\mu\nu}J^\nu = \frac{dn}{dt} + \Gamma^i_{i0}n = \frac{dn}{dt} + 3\frac{da}{dt}\frac{n}{a}$$

so

$$n(t) = \frac{\text{constant}}{a^3(t)} \tag{1.1.30}$$

This shows the decrease of number densities due to the expansion of the co-moving coordinate mesh for increasing $a(t)$.

Likewise, isotropy requires the mean value of any three-tensor t^{ij} at $\mathbf{x} = 0$ to be proportional to δ_{ij} and hence to g^{ij}, which equals $a^{-2}\delta_{ij}$ at $\mathbf{x} = 0$. Homogeneity requires the proportionality coefficient to be some function only of time. Since this is a proportionality between two three-tensors t^{ij} and g^{ij} it must remain unaffected by an arbitrary transformation of space coordinates, including those transformations that preserve the form of g^{ij} while taking the origin into any other point. Hence homogeneity and isotropy require the components of the energy-momentum tensor everywhere to take the form

$$T^{00} = \rho(t), \quad T^{0i} = 0, \quad T^{ij} = \tilde{g}^{ij}(\mathbf{x})a^{-2}(t)p(t) \tag{1.1.31}$$

(These are the conventional definitions of proper energy density ρ and pressure p, as given by Eq. (B.43) in the case of a velocity four-vector with $u^i = 0$, $u^0 = 1$.) The momentum conservation law $T^{i\mu}{}_{;\mu} = 0$ is automatically satisfied for the Robertson-Walker metric and the energy-momentum tensor (1.1.31), but the energy conservation law gives the useful information

$$\begin{aligned} 0 = T^{0\mu}_{;\mu} &= \frac{\partial T^{0\mu}}{\partial x^\mu} + \Gamma^0_{\mu\nu}T^{\nu\mu} + \Gamma^\mu_{\mu\nu}T^{0\nu} \\ &= \frac{\partial T^{00}}{\partial t} + \Gamma^0_{ij}T^{ij} + \Gamma^i_{i0}T^{00} = \frac{d\rho}{dt} + \frac{3\dot{a}}{a}(p+\rho), \end{aligned}$$

so that

$$\frac{d\rho}{dt} + \frac{3\dot{a}}{a}(p+\rho) = 0 \tag{1.1.32}$$

This can easily be solved for an equation of state of the form

$$p = w\rho \tag{1.1.33}$$

with w time-independent. In this case, Eq. (1.1.32) gives

$$\rho \propto a^{-3-3w}. \tag{1.1.34}$$

In particular, this applies in three frequently encountered extreme cases:

- Cold Matter (e.g. dust): $p = 0$

$$\rho \propto a^{-3} \tag{1.1.35}$$

- Hot Matter (e.g. radiation): $p = \rho/3$

$$\rho \propto a^{-4} \tag{1.1.36}$$

10.1.2 1.1 Spacetime geometry

- Vacuum energy: As we will see in Section 1.5, there is another kind of energy-momentum tensor, for which $T^{\mu\nu} \propto g^{\mu\nu}$, so that $p = -\rho$, in which case the solution of Eq. (1.1.32) is that ρ is a constant, known (up to conventional numerical factors) either as the cosmological constant or the vacuum energy.

These results apply separately for coexisting cold matter, hot matter, and a cosmological constant, provided that there is no interchange of energy between the different components. They will be used together with the Einstein field equations to work out the dynamics of the cosmic expansion in Section 1.5.

So far, we have considered only local properties of the spacetime. Now let us look at it in the large. For $K = +1$ space is finite, though like any spherical surface it has no boundary. The coordinate system used to derive Eq. (1.1.7) with $K = +1$ only covers half the space, with $z > 0$, in the same way that a polar projection map of the earth can show only one hemisphere. Taking account of the fact that z can have either sign, the circumference of the space is $2\pi a$, and its volume is $2\pi^2 a^3$.

The spaces with $K = 0$ or $K = -1$ are usually taken to be infinite, but there are other possibilities. It is also possible to have finite spaces with the same local geometry, constructed by imposing suitable conditions of periodicity. For instance, in the case $K = 0$ we might identify the points $\mathbf{x}$ and $\mathbf{x} + n_1\mathbf{L}_1 + n_2\mathbf{L}_2 + n_3\mathbf{L}_3$, where n_1, n_2, n_3 run over all integers, and $\mathbf{L}_1, \mathbf{L}_2$, and $\mathbf{L}_3$ are fixed non-coplanar three-vectors that characterize the space. This space is then finite, with volume $a^3\mathbf{L}_1 \cdot (\mathbf{L}_2 \times \mathbf{L}_3)$. Looking out far enough, we should see the same patterns of the distribution of matter and radiation in opposite directions. There is no sign of this in the observed distribution of galaxies or cosmic microwave background fluctuations, so any periodicity lengths such as $|\mathbf{L}_i|$ must be larger than about 10^{10} light years. [8]

There are an infinite number of possible periodicity conditions for $K = -1$ as well as for $K = +1$ and $K = 0$.[9] We will not consider these possibilities further here, because they seem ill-motivated. In imposing conditions of periodicity we give up the rotational (though not translational) symmetry that led to the Robertson-Walker metric in the first place, so there seems little reason to impose these periodicity conditions while limiting the local spacetime geometry to that described by the Robertson-Walker metric.

10.1.3 1.2 The cosmological redshift

The general arguments of the previous section gave no indication whether the scale factor $a(t)$ in the Robertson-Walker metric (1.1.9) is increasing, decreasing, or constant. This information comes to us from the observation of a shift in the frequencies of spectral lines from distant galaxies as compared with their values observed in terrestrial laboratories.

To calculate these frequency shifts, let us adopt a Robertson-Walker coordinate system in which we are at the center of coordinates, and consider a light ray coming to us along the radial direction. A ray of light obeys the equation $d\tau^2 = 0$, so for such a light ray Eq. (1.1.11) gives

$$dt = \pm a(t)\frac{dr}{\sqrt{1-Kr^2}} \tag{1.2.1}$$

For a light ray coming toward the origin from a distant source, r decreases as t increases, so we must choose the minus sign in Eq. (1.2.1). Hence if light leaves a source at co-moving coordinate r_1 at time t_1, it arrives at the origin $r = 0$ at a later time t_0, given by

$$\int_{t_1}^{t_0}\frac{dt}{a(t)} = \int_0^{r_1}\frac{dr}{\sqrt{1-Kr^2}} \tag{1.2.2}$$

Taking the differential of this relation, and recalling that the radial coordinate r_1 of co-moving sources is time-independent, we see that the interval δt_1 between departure of subsequent light signals is related to the interval δt_0 between arrivals of these light signals by

$$\frac{\delta t_1}{a\left(t_1\right)} = \frac{\delta t_0}{a\left(t_0\right)} \tag{1.2.3}$$

If the "signals" are subsequent wave crests, the emitted frequency is $v_1 = 1/\delta t_1$, and the observed frequency is $v_0 = 1/\delta t_0$, so

$$v_0/v_1 = a\left(t_1\right)/a\left(t_0\right) \tag{1.2.4}$$

If $a(t)$ is increasing, then this is a redshift, a decrease in frequency by a factor $a\left(t_1\right)/a\left(t_0\right)$, equivalent to an increase in wavelength by a factor conventionally called $1+z$:

$$1 + z = a\left(t_0\right)/a\left(t_1\right) \tag{1.2.5}$$

Alternatively, if $a(t)$ is decreasing then we have a blueshift, a decrease in wavelength given by the factor Eq. (1.2.5) with z negative. These results are frequently interpreted in terms of the familiar Doppler effect; Eq. (1.1.15)

10.1.4 1.2 The cosmological redshift

shows that for an increasing or decreasing $a(t)$, the proper distance to any comoving source of light like a typical galaxy increases or decreases with time, so that such sources are receding from us or approaching us, which naturally produces a redshift or blueshift. For this reason, galaxies with redshift (or blueshift) z are often said to have a cosmological radial velocity cz. (The meaning of relative velocity is clear only for $z \ll 1$, so the existence of distant sources with $z > 1$ does not imply any violation of special relativity.) However, the interpretation of the cosmological redshift as a Doppler shift can only take us so far. In particular, the increase of wavelength from emission to absorption of light does not depend on the rate of change of $a(t)$ at the times of emission or absorption, but on the increase of $a(t)$ in the whole period from emission to absorption.

We can also understand the frequency shift (1.2.4) by reference to the quantum theory of light: The momentum of a photon of frequency v is hv/c (where h is Planck's constant), and we saw in the previous section that this momentum varies as $1/a(t)$.

For nearby sources, we may expand $a(t)$ in a power series, so

$$a(t) \simeq a\left(t_0\right)\left[1+\left(t-t_0\right)H_0+\ldots\right] \tag{1.2.6}$$

where H_0 is a coefficient known as the Hubble constant:

$$H_0 \equiv \dot{a}\left(t_0\right)/a\left(t_0\right) \tag{1.2.7}$$

Eq. (1.2.5) then gives the fractional increase in wavelength as

$$z = H_0\left(t_0-t_1\right)+\ldots \tag{1.2.8}$$

Note that for close objects, $t_0 - t_1$ is the proper distance d (in units with $c = 1$). We therefore expect a redshift (for $H_0 > 0$) or blueshift (for $H_0 < 0$) that increases linearly with the proper distance d for galaxies close enough to use the approximation (1.2.6):

$$z = H_0 d + \ldots \tag{1.2.9}$$

The redshift of light from other galaxies was first observed in the 1910s by Vesto Melvin Slipher at the Lowell Observatory in Flagstaff, Arizona. In 1922, he listed 41 spiral nebulae, of which 36 had positive z up to 0.006 , and only 5 had negative z, the most negative being the Andromeda nebula M31, with $z = -0.001$.[1] From 1918 to 1925 C. Wirtz and K. Lundmark [2]

discovered a number of spiral nebulae with redshifts that seemed to increase with distance. But until 1923 it was only possible to infer the relative distances of the spiral nebulae, using observations of their apparent luminosity or angular diameter. With the absolute luminosity and physical dimensions unknown, it was even possible that the spiral nebulae were outlying parts of our own galaxy, as was in fact believed by many astronomers. Edwin Hubble's 1923 discovery of Cepheid variable stars in the Andromeda nebula M31 (discussed in the next section) allowed him to estimate its distance and size, and made it clear that the spiral nebulae are galaxies like our own, rather than objects in our own galaxy.

No clear linear relation between redshift and distance could be seen in the early data of Slipher, Wirtz, and Lundmark, because of a problem that has continued to bedevil measurements of the Hubble constant down to the present. Real galaxies generally do not move only with the general expansion or contraction of the universe; they typically have additional "peculiar" velocities of hundreds of kilometers per second, caused by gravitational fields of neighboring galaxies and intergalactic matter. To see a linear relation between redshift and distance, it is necessary to study galaxies with $|z| \gg 10^{-3}$, whose cosmological velocities zc are thousands of kilometers per second.

In 1929 Hubble [3] announced that he had found a "roughly linear" relation between redshift and distance. But at that time redshifts and distances had been measured only for galaxies out to the large cluster of galaxies in the constellation Virgo, whose redshift indicates a radial velocity of about $1,000$ km/sec, not much larger than typical peculiar velocities. His data points were therefore spread out widely in a plot of redshift versus distance, and did not really support a linear relation. But by the early 1930s he had measured redshifts and distances out to the Coma cluster, with redshift $z \simeq 0.02$, corresponding to a recessional velocity of about

[471] V. M. Slipher, table prepared for A. S. Eddington, The Mathematical Theory of Relativity, 2nd ed. (Cambridge University Press, London, 1924): 162.

[2] C. Wirtz, Astr. Nachr. 206, 109 (1918); ibid. 215, 349 (1921); ibid. 216, 451 (1922); ibid. 222, 21 (1924); Scientia 38, 303 (1925); K. Lundmark, Stock. Hand. 50, No. 8 (1920); Mon. Not. Roy. Astron. Soc. 84, 747 (1924); ibid. 85, 865 (1925).

7,000 km/sec, and a linear relation between redshift and distance was evident. The conclusion was clear (at least, to some cosmologists): the universe really is expanding. The correctness of this interpretation of the redshift is supported by observations to be discussed in Section 1.7.

From Hubble's time to the present galaxies have been discovered with ever larger redshifts. Galaxies were found with redshifts of order unity, for which expansions such as Eq. (1.2.9) are useless, and we need formulas that take relativistic effects into account, as discussed in Sections 1.4 and 1.5. At the time of writing, the largest accurately measured redshift is for a galaxy observed with the Subaru telescope. [4] The Lyman alpha line from

this galaxy (emitted in the transition from the $2p$ to $1s$ levels of hydrogen), which is normally at an ultraviolet wavelength of $1,215$Å, is observed at the infrared wavelength of $9,682$Å, indicating a redshift $1+z = 9682/1215$, or $z = 6.96$.

It may eventually become possible to measure the expansion rate $H(t) \equiv \dot{a}(t)/a(t)$ at times t earlier than the present, by observing the change in very accurately measured redshifts of individual galaxies over times as short as a decade. [5] By differentiating Eq. (1.2.5) we see that the rate of change of redshift with the time of observation is

$$\frac{dz}{dt_0} = \frac{\dot{a}\,(t_0)}{a\,(t_1)} - \frac{a\,(t_0)\,\dot{a}\,(t_1)}{a^2\,(t_1)}\frac{dt_1}{dt_0} = \left[H_0 - H\,(t_1)\,\frac{dt_1}{dt_0}\right](1+z)$$

From the same argument that led to Eq. (1.2.3) we have $dt_1/dt_0 = 1/(1+z)$, so if we measure dz/dt_0 we can find the expansion rate at the time of light emission from the formula

$$H\,(t_1) = H_0(1+z) - \frac{dz}{dt_0} \tag{1.2.10}$$

10.1.5 1.3 Distances at small redshift: The Hubble constant

We must now think about how astronomical distances are measured. In this section we will be considering objects that are relatively close, say with z not much greater than 0.1 , so that effects of the spacetime curvature and cosmic expansion on distance determinations can be neglected. These measurements are of cosmological importance in themselves, as they are used to learn the value of the Hubble constant H_0. Also, distance measurements at larger redshift, which are used to find the shape of the function $a(t)$, rely on the observations of "standard candles," objects of known intrinsic luminosity, that must be identified and calibrated by studies at these relatively small redshifts. Distance determinations at larger redshift will be discussed in Section 1.6, after we have had a chance to lay a foundation in Sections 1.4 and 1.5 for an analysis of the effects of expansion and spacetime geometry on measurements of distances of very distant objects.

It is conventional these days to separate the objects used to measure distances in cosmology into primary and secondary distance indicators. The absolute luminosities of the primary distance indicators in our local group

of galaxies are measured either directly, by kinematic methods that do not depend on an a priori knowledge of absolute luminosities, or indirectly, by observation of primary distance indicators in association with other primary distance indicators whose distance is measured by kinematic methods. The sample of these relatively close primary distance indicators is large enough to make it possible to work out empirical rules that give their absolute luminosities as functions of various observable properties. Unfortunately, the primary distance indicators are

[473] E. P. Hubble, Proc. Nat. Acad. Sci. 15, 168 (1929).

[4] M. Iye et al., Nature 443, 186 (2006) [astro-ph/0609393].

[475] A. Loeb, Astrophys. J. 499, L111 (1998) [astro-ph/9802122]; P-S. Corasaniti, D. Huterer, and A. Melchiorri, Phys. Rev. D 75, 062001 (2007) [astro-ph/0701433]. For an earlier suggestion along this line, see A. Sandage, Astrophys. J. 139, 319 (1962).

not bright enough for them to be studied at distances at which z is greater than about 0.01 , redshifts at which cosmological velocities cz would be greater than typical random departures of galactic velocities from the cosmological expansion, a few hundred kilometers per second. Thus they cannot be used directly to learn about $a(t)$. For this purpose it is necessary to use secondary distance indicators, which are bright enough to be studied at these large distances, and whose absolute luminosities are known through the association of the closer ones with primary distance indicators.

A. Primary distance indicators [1]

Almost all distance measurements in astronomy are ultimately based on measurements of the distance of objects within our own galaxy, using one or the other of two classic kinematic methods.

1. Trigonometric parallax

The motion of the earth around the sun produces an annual motion of the apparent position of any star around an ellipse, whose maximum angular radius π is given in radians (for $\pi \ll 1$, which is the case for all stars) by

$$\pi = \frac{d_E}{d} \tag{1.3.1}$$

where d is the star's distance from the solar system, and d_E is the mean distance of the earth from the sun, [2] defined as the astronomical unit,

$1\text{AU} = 1.496 \times 10^8$ km. A parsec (pc) is defined as the distance at which $\pi = 1''$; there are 206,264.8 seconds of arc per radian so

$$1\text{pc} = 206,264.8\text{AU} = 3.0856 \times 10^{13} \text{ km} = 3.2616 \text{ light years.}$$

The parallax in seconds of arc is the reciprocal of the distance in parsecs.
The first stars to have their distances found by measurement of their trigonometric parallax were α Centauri, by Thomas Henderson in 1832, and 61 Cygni, by Friedrich Wilhelm Bessel in 1838. These stars are at distances 1.35 pc and 3.48 pc , respectively. The earth's atmosphere makes it very difficult to measure trigonometric parallaxes less than about $0.03''$ from ground-based telescopes, so that for many years this method could be used to find the distances of stars only out to about 30 pc , and at these distances only for a few stars and with poor accuracy.

This situation has been improved by the launching of a European Space Agency satellite known as Hipparcos, used to measure the apparent positions and luminosities of large numbers of stars in our galaxy. [3] For stars of sufficient brightness, parallaxes could be measured with an accuracy (standard deviation) in the range of 7 to 9×10^{-4} arc seconds. Of the 118,000 stars in the Hipparcos Catalog, it was possible in this way to find distances with a claimed uncertainty of no more than 10% for about 20,000 stars, some at distances over 100 pc.

[471] For a survey, see M. Feast, in Nearby Large-Scale Structures and the Zone of Avoidance, eds. A. P. Fairall and P. Woudt (ASP Conference Series, San Francisco, 2005) [astro-ph/0405440].

[2] The history of measurements of distances in the solar system goes back to Aristarchus of Samos (circa 310 BC-230 BC). From the ratio of the breadth of the earth's shadow during a lunar eclipse to the angular diameter of the moon he estimated the ratio of the diameters of the moon and earth; from the angular diameter of the moon he estimated the ratio of the diameter of the moon to its distance from the earth; and from the angle between the lines of sight to the sun and moon when the moon is half full he estimated the ratio of the distances to the sun and moon; and in this way he was able to measure the distance to the sun in units of the diameter of the earth. Although the method of Aristarchus was correct, his observations were poor, and his result for the distance to the sun was far too low. [For an account of Greek astronomy before Aristarchus and a translation of his work, see T. L. Heath, Aristarchus of

2. Proper motions

A light source at a distance d with velocity $v_\perp$ perpendicular to the line of sight will appear to move across the sky at a rate μ in radians/time given by

$$\mu = v_\perp/d \tag{1.3.2}$$

This is known as its proper motion. Of course, astronomers generally have no way of directly measuring the transverse velocity $v_\perp$, but they can measure the component v_r of velocity along the line of sight from the Doppler shift of the source's spectral lines. The problem is to infer $v_\perp$ from the measured value of v_r. This can be done in a variety of special cases:

- Moving clusters are clusters of stars that were formed together and hence move on parallel tracks with equal speed. (These are open

clusters, in the sense that they are not held together by gravitational attraction, in distinction to the much larger globular clusters whose spherical shape indicates a gravitationally bound system.) The most important such cluster is the Hyades (called by Tennyson's Ulysses the "rainy Hyades"), which contains over 100 stars. The velocities of these stars along the line of sight are measured from their Doppler shifts, and if we knew the distance to the cluster then the velocities of its stars at right angles to the line of sight could be measured from their proper motions. The distance to the cluster was determined long ago to be about 40 pc by imposing the constraint that all these velocities are parallel. Distances measured in this way are often expressed as moving cluster parallaxes. Since the advent of the Hipparcos satellite, the moving cluster method has been supplemented with a direct measurement of the trigonometric parallax of some of these clusters, including the Hyades.

- A second method is based on the statistical analysis of the Doppler shifts and proper motions of stars in a sample whose relative distances are all known, for instance because they all have the same (unknown) absolute luminosity, or because they all at the same (unknown) distance. The Doppler shifts give the velocities along the line of sight, and the proper motions and the relative distances give the velocities transverse to the line of sight, up to a single overall factor related to the unknown absolute luminosity or distance. This factor can be determined by requiring that the distribution of velocities transverse to the line of sight is the same as the distribution of velocities along the line of sight. Distances measured in this way are often called statistical parallaxes, or dynamical distances.

- The distance to the Cepheid variable star ζ Geminorum has been measured [4] by comparing the rate of change of its physical diameter, as found from the Doppler effect, with the rate of change of its angular diameter, measured using an optical interferometer. (About Cepheids, more below.) The distance was found to be 336 ± 44pc, much greater than could have been found from a trigonometric parallax. This method has subsequently been extended to eight other Cepheids. [5]

- It is becoming possible to measure distances by measuring the proper motion of the material produced by supernovae, assuming a

[47]Samos (Oxford University Press, Oxford, 1913).] The first reasonably accurate determination of the distance of the earth to the sun was made by the measurement of a parallax. In 1672 Jean Richer and Giovanni Domenico Cassini measured the distance from the earth to Mars, from which it was possible to infer the distance from the earth to the sun, by observing the difference in the apparent direction to Mars as seen from Paris and Cayenne, which are separated by a known distance of 6,000 miles. Today distances within the solar system are measured very accurately by measurement of the timing of radar echoes from planets and of radio signals from transponders carried by spacecraft.

[3] M. A. C. Perryman et al., Astron. Astrophys. 323, L49 (1997).

more-or-less cylindrically symmetric explosion, so that the transverse velocity $v_\perp$ can be inferred from the radial velocity v_r measured by Doppler shifts. This method has been applied [6] to the ring around the supernova SN1987A, observed in 1987 in the Large Magellanic Cloud, with the result that its distance is 52 kpc (thousand parsecs).

- The measurement of the time-varying Doppler shift and proper motion of an object in orbit around a central mass can be used to find the distance to the object. For instance, if the line of sight happens to be in the plane of the orbit, and if the orbit is circular, then the Doppler shift is a maximum when the object is moving along the line of sight, and hence gives the orbital velocity ν, while the proper motion μ is a maximum when the object is moving with the same velocity at right angles to the line of sight, and gives the distance as ν/μ. This method can also be used for orbits that are inclined to the line of sight and not circular, by studying the time-variation of the Doppler shift and proper motion. The application of this method to the star S2, which orbits the massive black hole in the galactic center, gives what is now the best value for the distance of the solar system from the galactic center, [7] as 8.0 ± 0.4kpc. This method also allows the measurement of some distances outside our galaxy, by using the motion of masers - point microwave sources - in the accretion disks of gas and dust in orbit around black holes at the centers of galaxies. The orbital velocity can be judged from the Doppler shifts of masers at the edge of the accretion disk, which are moving directly toward us or away from us, and if this is the same as the orbital velocities of masers moving transversely to the line of sight, then the ratio of this orbital velocity to their observed proper motion gives the distance to the galaxy. So far, this method has been used to measure the distance to the galaxy NGC 4258, [8] as 7.2 ± 0.5Mpc (million parsecs), and to the galaxy M33, [9] as 0.730 ± 0.168 Mpc .

These kinematic methods have limited utility outside the solar neighborhood. We need a different method to measure larger distances.

3. Apparent luminosity

The most common method of determining distances in cosmology is based on the measurement of the apparent luminosity of objects of known (or

supposedly known) absolute luminosity. The absolute luminosity L is the energy emitted per second, and the apparent luminosity ℓ is the energy received per second per square centimeter of receiving area. If the energy is emitted isotropically, then we can find the relation between the absolute and apparent luminosity in Euclidean geometry by imagining the luminous object to be surrounded with a sphere whose radius is equal to the distance d between the object and the earth. The total energy per second passing through the sphere is $4\pi d^2 \ell$, so

$$\ell = \frac{L}{4\pi d^2} \tag{1.3.3}$$

This relation is subject to corrections due to interstellar and/or intergalactic absorption, as well as possible anisotropy of the source, which though important in practice involve too many technicalities to go into here.

[474] B. F. Lane, M. J. Kuchner, A. F. Boden, M. Creech-Eakman, and S. B. Kulkarni, Nature 407, 485 (September 28, 2000).

[5] P. Kervella et al., Astron. Astrophys. 423, 327 (2004) [astro-ph/0404179].

[476] N. Panagia, Mem. Soc. Astron. Italiana 69, 225 (1998).

[7] F. Eisenhauer et al., Astrophys. J. Lett. 597, L121 (2003) [astro-ph/0306220].

[8] J. Herrnstein et al., Nature 400, 539 (3 August 1999).

[9] A. Brunthaler, M. J. Reid, H. Falcke, L. J. Greenhill, and C. Henkel, Science 307, 1440 (2005) [astro-ph/0503058].

Astronomers unfortunately use a traditional notation for apparent and absolute luminosity in terms of apparent and absolute magnitude. [10] In the second century A.D., the Alexandrian astronomer Claudius Ptolemy published a list of 1,022 stars, labeled by categories of apparent brightness, with bright stars classed as being of first magnitude, and stars just barely visible being of sixth magnitude. [11] This traditional brightness scale was made quantitative in 1856 by Norman Pogson, who decreed that a difference of five magnitudes should correspond to a ratio of a factor 100 in apparent luminosities, so that $\ell \propto 10^{-2m/5}$. With the advent of photocells at the beginning of the twentieth century, it became possible to fix the constant of proportionality: the apparent bolometric luminosity (that is, including all wavelengths) is given in terms of the apparent bolometric magnitude m by

$$\ell = 10^{-2m/5} \times 2.52 \times 10^{-5} \text{erg cm}^{-2} \text{ s}^{-1} \tag{1.3.4}$$

For orientation, Sirius has a visual magnitude $m_{\rm vis} = -1.44$, the Andromeda nebula M31 has $m_{\rm vis} = 0.1$, and the large galaxy M87 in the nearest large cluster of galaxies has $m_{\rm vis} = 8.9$. The absolute magnitude in any wavelength band is defined as the apparent magnitude an object would have at a distance of 10 pc , so that the absolute bolometric luminosity is given in terms of the absolute bolometric magnitude M by

$$L = 10^{-2M/5} \times 3.02 \times 10^{35} \text{erg s}^{-1} \tag{1.3.5}$$

For comparison, in the visual wavelength band the absolute magnitude $M_{\rm vis}$ is +4.82 for the sun, +1.45 for Sirius, and -20.3 for our galaxy. Eq. (1.3.3) may be written as a formula for the distance in terms of the distance-modulus $m - M$:

$$d = 10^{1+(m-M)/5} \text{pc} \tag{1.3.6}$$

There are several different kinds of star that have been used in measurements of distance through the observation of apparent luminosity:

- Main Sequence: Stars that are still burning hydrogen at their cores obey a characteristic relation between absolute luminosity and color, both depending on mass. This is known as the main sequence, discovered in the decade before the First World War by Ejnar Hertzsprung and Henry Norris Russell. The luminosity is greatest for blue-white stars, and then steadily decreases for colors tending toward yellow and red. The shape of the main sequence is found by observing the apparent luminosities and colors of large numbers of stars in clusters, all of which in each cluster may be assumed to be at the same distance from us, but we need to know the distances to the clusters to calibrate absolute luminosities on the main sequence. For many years the calibration of the main sequence absolute luminosities was based on observation of a hundred or so main sequence stars in the Hyades cluster, whose distance was measured by the moving cluster method described above. With the advent of the Hipparcos satellite, the calibration of the main sequence has been greatly improved through the observation of colors and apparent luminosities of nearly 100,000 main sequence stars whose distance is known through measurement of their trigonometric parallax. Including in this sample are stars in open clusters such as the Hyades, Praesepe, the Pleiades, and NGC 2516; these clusters yield consistent main sequence absolute magnitudes if care is taken to take proper account of the varying chemical compositions of the stars in different clusters. [12] With the main sequence calibrated

[4710] For the history of the apparent magnitude scale, see J. B. Hearnshaw, The Measurement of Starlight: Two centuries of astronomical photometry (Cambridge University Press, Cambridge, 1996); K. Krisciunas, astro-ph/0106313.

[11] For the star catalog of Ptolemy, see M. R. Cohen and I. E. Drabkin, A Source Book in Greek Science (Harvard University Press, Cambridge, MA, 1948): p. 131.

in this way, we can use Eq. (1.3.3) to measure the distance of any star cluster or galaxy in which it is possible to observe stars exhibiting the main sequence relation between apparent luminosity and color. Distances measured in this way are sometimes known as photometric parallaxes.
The analysis of the Hipparcos parallax measurements revealed a discrepancy between the distances to the Pleiades star cluster measured by observations of main sequence stars and by measurements of

trigonometric parallax. [13] The traditional method, using a main sequence calibration based on the application of the moving cluster method to the closer Hyades cluster, gave a distance to the Pleiades [14] of 132 ± 4pc. Then trigonometric parallaxes of a number of stars in the Pleiades measured by the Hipparcos satellite gave a distance [15] of 118 ± 4pc, in contradiction with the results of main sequence fitting. More recently, these Hipparcos parallaxes have been contradicted by more accurate measurements of the parallaxes of three stars in the Pleiades with the Fine Guidance Sensor of the Hubble Space Telescope, [16] which gave a distance of 133.5 ± 1.2pc, in good agreement with the main sequence results. At the time of writing, the balance of astronomical opinion seems to be favoring the distances given by main sequence photometry. [17]

- Red Clump Stars: The color-magnitude diagram of clusters in metalrich [18] parts of the galaxy reveals distinct clumps of red giant stars in a small region of the diagram, with a spread of only about 0.2 in visual magnitude. These are stars that have exhausted the hydrogen at their cores, with helium taking the place of hydrogen as the fuel for nuclear reactions at the stars' cores. The absolute magnitude of the red clump stars in the infrared band (wavelengths around 800 nm) has been determined [19] to be $M_I = -0.28 \pm 0.2$mag, using the distances and apparent magnitudes measured with the Hipparcos satellite and in an earlier survey. [20] In this band there is little dependence of absolute magnitude on color, but it has been argued that even the infrared magnitude may depend significantly on metallicity. [21]

- RR Lyrae Stars: These are variable stars that have been used as distance indicators for many decades. [22] They can be recognized by their periods, typically 0.2 to 0.8 days. The use of the statistical parallax, trigonometric parallax and moving cluster methods (with data

from both ground-based observatories and the Hipparcos satellite) give respectively [23] an absolute visual magnitude for RR Lyrae stars in our galaxy's halo of $0.77 \pm 0.13, 0.71 \pm 0.15$, and 0.67 ± 0.10, in good agreement with an earlier result [24] $M_{\rm vis} = 0.71 \pm 0.12$ for halo RR Lyrae stars and 0.79 ± 0.30 for RR Lyrae stars in the thick disk of the galaxy. RR Lyrae stars are mostly too

[4712] S. M. Percival, M. Salaris, and D. Kilkenny, Astron. Astrophys. 400, 541 (2003) [astro-ph/0301219].
[4713] B. Paczynski, Nature 227, 299 (22 January, 2004).
[14] G. Meynet, J.-C. Mermilliod, and A. Maeder, Astron. Astrophys. Suppl. Ser. 98, 477 (1993).
[15] J.-C. Mermilliod, C. Turon, N. Robichon, F. Arenouo, and Y. Lebreton, in ESA SP-402 HipparcosVenice '97, eds. M.A.C. Perryman and P. L. Bernacca (European Space Agency, Paris, 1997), 643; F. van Leeuwen and C. S. Hansen Ruiz, ibid, 689; F. van Leeuwen, Astron. Astrophys. 341, L71 (1999).
[16] D. R. Soderblom et al., Astron. J. 129, 1616 (2005) [astro-ph/0412093].
[17] A new reduction of the raw Hipparcos data is given by F. van Leeuwen and E. Fantino, Astron. Astrophys. 439, 791 (2005) [astro-ph/0505432].
[18] Astronomers use the word "metal" to refer to all elements heavier than helium.
[19] B. Paczyński and K. Z. Stanek, Astrophys. J. 494, L219 (1998).
[20] A. Udalski et al., Acta. Astron. 42, 253 (1992).
[21] L. Girardi, M. A. T. Groenewegen, A. Weiss, and M. Salaris, astro-ph/9805127.
[22] For a review, see G. Bono, Lect. Notes Phys. 635, 85 (2003) [astro-ph/0305102].

far for a measurement of their trigonometric parallax, but recently measurements [25] with the Hubble Space Telescope have given a value of 3.82×10^{-3} arcsec for the trigonometric parallax of the eponymous star RR Lyr itself, implying an absolute visual magnitude of $0.61^{-0.11}_{+0.10}$.

- Eclipsing Binaries: In favorable cases it is possible to estimate the intrinsic luminosity of a star that is periodically partially eclipsed by a smaller companion, without the use of any intermediate distance indicators. The velocity of the companion can be inferred from the Doppler shift of its spectral lines (with the ellipticity of the orbit inferred from the variation of the Doppler shift with time), and the radius of the primary star can then be calculated from the duration of the eclipse. The temperature of the primary can be found from measurement of its spectrum, typically from its apparent luminosity in various wavelength bands. Knowing the radius, and hence the area, and the temperature of the primary, its absolute luminosity can then be calculated from the Stefan-Boltzmann law for black body radiation. This method has been applied to measure distances to two neighboring dwarf galaxies, the Large Magellanic Cloud (LMC) [26] and the Small Magellanic Cloud (SMC), [27] and to the Andromeda galaxy M31 [28] and its satellite M33. [29]

- Cepheid variables: Because they are so bright, these are by far the most important stars used to measure distances outside our galaxy. Named after the first such star observed, δ Cephei, they can be

recognized from the characteristic time dependence of their luminosity, with periods ranging from 2 to 45 days. (Cepheids in other galaxies have been observed with periods extending up to 100 days.) In 1912 Henrietta Swan Leavitt discovered that the Cepheid variables in the Small Magellanic Cloud (SMC) have apparent luminosities given by a smooth function of the period of the variation in luminosity, but the distance to the SMC was not known. Having measured the distances and apparent luminosities of several Cepheids in open clusters, and hence their absolute luminosities, it became possible to calibrate the relation between period and luminosity. Cepheid variables thus became a "standard candle" that could be used to measure the distance to any galaxy close enough for Cepheids to be seen. It was the discovery of Cepheids in M31, together with Leavitt's calibration of the Cepheid period-luminosity relation, that allowed Edwin Hubble in 1923 to measure the distance of M31, and show that it was far outside our own galaxy, and hence a galaxy in its own right.

Today the form of the Cepheid period-luminosity relation is derived more from the Large Magellanic Cloud (LMC), where there are many Cepheids, and the dependence of the absolute luminosity on color is also taken into account. The calibration of Cepheid absolute luminosities can therefore be expressed as (and often in fact amounts to) a measurement of the distance to the LMC. Main sequence photometry and other methods gave what for some years was a generally accepted LMC distance modulus of 18.5 mag , corresponding according to Eq. (1.3.6) to a distance of 5.0×10^4pc. The use of red clump stars [30] has given a distance modulus of

[4][7][23] P. Popowski and A. Gould, Astrophys. J. 506, 259, 271 (1998); also astro-ph/9703140, astroph/9802168; and in Post-Hipparcos Cosmic Candles, eds. A. Heck and F. Caputo (Kluwer Academic Publisher, Dordrecht) [astro-ph/9808006]; A. Gould and P. Popowski, Astrophys. J. 568, 544 (1998) [astro-ph/9805176]; and references cited therein.

[24] A. Layden, R. B. Hanson, S. L. Hawley, A. R. Klemola, and C. J. Hanley, Astron. J. 112, 2110 (1996).

[25] G. F. Benedict et al., Astrophys. J. 123, 473 (2001) [astro-ph/0110271]

[26] E. F. Guinan et al., Astrophys. J. 509, L21 (1998); E. L. Fitzpatrick et al. Astrophys. J. 587, 685 (2003).

[27] T. J. Harries, R. W. Hilditch, and I. D. Howarth, Mon. Not. Roy. Astron. Soc. 339, 157 (2003); R. W. Hilditch, I. D. Howarth, and T. J. Harries, Mon. Not. Roy. Astron. Soc. 357, 304 (2005).

[28] I. Ribas et al., Astrophys. J. 635, L37 (2005).

[29] A. Z. Bonanos et al., Astrophys. Space Sci. 304, 207 (2006) [astro-ph/0606279].

18.47 , with a random error ±0.01, and a systematic error $^{+0.05}_{-0.06}$. A large catalog [31] of Cepheids in the LMC has been interpreted by the members of the Hubble Space Telescope Key Project [32] to give the Cepheid visual and infrared absolute magnitudes as functions of the period P in days:
$M_V = -2.760 \log_{10} P - 1.458, \quad M_I = -2.962 \log_{10} P - 1.942,$
under the assumption that the LMC distance modulus is 18.5 .
This result was challenged in two distinct ways, which illustrate the difficulty of this sort of distance measurement:

First, there have been discordant measurements of the distance to the LMC. Under the assumption that red clump stars in the LMC

have the same infrared luminosity as those in the local galactic disk, a distance modulus was found [33] that was 0.45 magnitudes smaller, giving a distance to the LMC that is smaller by a factor 0.8 . This has in turn been challenged on the grounds that the stars in the LMC have distinctly lower metallicity than in the local disk; two groups taking this into account [34] have given LMC distance moduli of 18.36± 0.17 mag and 18.28 ± 0.18mag, in fair agreement with the previously accepted value. This also agrees with the measurement of the distance to the LMC inferred [35] from observations of RR Lyrae stars, which gives a distance modulus of 18.33 ± 0.06mag. This distance modulus for the LMC is further confirmed by the measurement of the distance of the eclipsing binary HV2274; taking account of its distance from the center of the LMC gives [36] a distance modulus for the LMC of 18.30 ± 0.07.

Second, there have been new calibrations of the Cepheid periodluminosity relation, that do not rely on Cepheids in the LMC, which together with observations of Cepheids in the LMC can be used to give an independent estimate of the LMC distance. [37] In recent years the satellite Hipparcos [38] has measured trigonometric parallaxes for 223 Cepheid variables in our galaxy, of which almost 200 can be used to calibrate the period-luminosity relation, without relying on main sequence photometry, red clump stars, or RR Lyrae stars. The nearest Cepheids are more than 100 pc away from us (the distance to Polaris is about 130 pc), so the parallaxes are just a few milliarcseconds, and individual measurements are not very accurate, but with about 200 Cepheids measured it has been possible to get pretty good accuracy. One early result [39] gave the relation between the absolute visual magnitude M_V and the period P (in days) as

$$M_V = -2.81 \log_{10} P - 1.43 \pm 0.10$$

This was a decrease of about 0.2 magnitudes from previous results, i.e., an increase of the intrinsic luminosity of Cepheids by a factor $10^{0.2 \times 2/5} = 1.20$ leading to a 10% increase in all cosmic distances based

directly or indirectly on the Cepheid period-luminosity relation. With this value of Cepheid absolute luminosity, the LMC distance modulus would be 18.66, or slightly less with corrections for the metallicity of the LMC (though with the absolute luminosity of Cepheids calibrated

[4730] M. Salaris, S. Percival, and L. Girardi, Mon. Not. Roy. Astron. Soc. 345, 1030 (2003) [astroph/0307329].
[31] A. Udalski et al., Acta Astr. 49, 201 (1999): Table 1.
[32] W. L. Freedman et al., Astrophys. J. 553, 47 (2001).
[4733] K. Z. Stanek, D. Zaritsky, and J. Harris, Astrophys. J. 500, L141 (1998) [astro-ph/9803181].
[34] A. A. Cole, Astrophys. J. 500, L137 (1998) [astro-ph/9804110]; L. Girardi et al., op. cit..
[35] P. Popowski and A. Gould, op. cit..
[36] E. F. Guinan et al., op. cit..
[37] For a review, see M. Feast, Odessa Astron. Publ. 14 [astro-ph/0110360].
[38] M. A. C. Perryman, Astron. Astrophys. 323, L49 (1997).
[39] M. W. Feast and R. M. Catchpole, Mon. Not. Roy. Astron. Soc. 286, L1 (1997); also see F. Pont, in Harmonizing Cosmic Distances in a Post-Hipparcos Era, eds. D. Egret and A. Heck (ASP Conference Series, San Francisco, 1998) [astro-ph/9812074]; H. Baumgardt, C. Dettbarn, B. Fuchs, J. Rockmann, and R. Wielen, in Harmonizing Cosmic Distance Scales in a Post-Hipparcos Era, ibid [astro-ph/9812437].

by Hipparcos observations, the only relevance of the LMC for the Cepheid period-luminosity relation is to determine its shape.) This result for the Cepheid absolute luminosities has in turn been contradicted. [40]

These uncertainties may now have been resolved by measurements of the trigonometric parallax of Cepheids in our galaxy with the Fine Guidance Sensor of the Hubble Space Telescope. First, the trigonometric parallax of δ Cephei [41] gave a distance of 273 ± 11pc, corresponding to an LMC distance modulus of 18.50 ± 0.13. More recently, trigonometric parallaxes have been measured for nine Galactic Cepheids, giving an LMC distance modulus of 18.50 ± 0.03, or with metallicity corrections, 18.40 ± 0.05.[42]

There has also been an independent calibration of the Cepheid period-luminosity relation through observations [43] of Cepheids in the galaxy NGC 4258, whose distance 7.2 ± 0.5Mpc has been measured using the observations of proper motions of masers in this galaxy mentioned above. This distance is in satisfactory agreement with the distance 7.6 ± 0.3Mpc obtained from the Cepheids in NGC 4258 under the assumption that these Cepheids have the period-luminosity relation (1.3.7) obtained under the assumption that the LMC distance modulus is 18.5 , which tends to confirm this period-luminosity relation. But there are differences in the metallicity of the Cepheids in NGC 4258 and in the LMC, which makes this conclusion somewhat controversial. [44] A 2006 calibration of the Cepheid period-luminosity relation based on the study of 281 Cepheids in NGC 4258[45] (whose distance, as we have seen, is known from observations of maser Doppler shifts and proper motions) gave an LMC distance modulus 18.41 ± 0.10 (stat.) ± 0.13 (syst.). This study includes both a field that is metal rich, like our Galaxy, and a field that is metal poor, like the LMC, so

it provides a calibration of the metallicity dependence of the Cepheid period-luminosity relation.

In a 2003 survey [46] the LMC distance modulus measured using a variety of distance indicators other than Cepheid variables (including RR Lyrae stars, red clump stars, etc.) was found to be 18.48 ± 0.04, in very good agreement with the earlier value 18.52 ± 0.05 found by observation of Cepheids, with the corrections adopted by the Hubble Space Telescope group.

B. Secondary distance indicators

None of the above distance indicators are bright enough to be used to measure distances at redshifts large enough so that peculiar velocities can be neglected compared with the expansion velocity, say, $z > 0.03$. For this we need what are called secondary distance indicators that are brighter than Cepheids, such as whole galaxies, or supernovae, which can be as bright as whole galaxies.

For many years Cepheids could be used as distance indicators only out to a few million parsecs (Mpc), which limited their use to the Local Group (which consists of our galaxy and the Andromeda nebula M31, and a dozen or so smaller galaxies like M33 and the LMC and SMC) and some other nearby groups (the M81, M101, and Sculptor groups). This was not enough to calibrate distances to an adequate population of galaxies or supernovae, and so it was necessary to use a variety of intermediate distance indicators: globular clusters, HII regions, brightest stars in galaxies, etc. Now the Hubble Space Telescope allows us to observe

[4740] See, e.g., B. F. Madore and W. L. Freedman, Astrophys. J. 492, 110 (1998) For a recent survey of the theory underlying the Cepheid period-luminosity relation, see A. Gautschy, in Recent Results on $H_0 - 19$ th Texas Symposium on Relativistic Astrophysics [astro-ph/9901021].

[41] G. F. Benedict et al., Astrophys. J. 124, 1695 (2002).

[42] G. F. Benedict et al., Astron. J. 133, 1810 (2007), [astro-ph/0612465].

[43] J. A. Newman et al., Astrophys. J. 553, 562 (2001) [astro-ph-0012377].

[44] For instance, see B. Paczynski, Nature 401, 331 (1999); F. Caputo, M. Marconi, and I. Musella, Astrophys. J. [astro-ph/0110526].

[45] L. M. Macri et al., Astrophys. J. 652, 1133 (2006) [astro-ph/0608211].

Cepheids in a great many galaxies at much greater distances, out to about 30 Mpc , and so the secondary distance indicators can now be calibrated directly, without the use of intermediate distance indicators. Four chief secondary distance indicators have been developed:

1. The Tully-Fisher relation

Although whole galaxies can be seen out to very large distances, it has not been possible to identify any class of galaxies with the same absolute luminosity. However, in 1977 Tully and Fisher [47] developed a method for estimating the absolute luminosity of suitable spiral galaxies. The 21 cm absorption line in these galaxies (arising in transitions of hydrogen atoms from lower to the higher of their two hyperfine states) is widened by the

Doppler effect, caused by the rotation of the galaxy. The line width W gives an indication of the maximum speed of rotation of the galaxy, which is correlated with the mass of the galaxy, which in turn is correlated with the galaxy's absolute luminosity. [48] (It is also possible to apply the Tully-Fisher relation using the width of other lines, such as a radio frequency transition in the carbon monoxide molecule. [49])

In one application of this approach [50] the shape of the function $L_I(W)$ that gives the infrared band absolute luminosity as a function of 21 cm line width (that is, the absolute luminosity up to a common constant factor) was found from a sample of 555 spiral galaxies in 24 clusters, many with redshifts less than 0.01 . (The relative distances to these galaxies were found from the ratios of their redshifts, using Eq. (1.2.9), so that the peculiar velocities of these galaxies introduced considerable errors into the estimated ratios of absolute luminosities of individual pairs of galaxies, but with 555 galaxies in the sample it could be assumed that these errors would cancel in a leastsquares fit of the measured relative values of absolute luminosity to a smooth curve.) Roughly speaking, $L_I(W)$ turned out to be proportional to W^3. The overall scale of the function $L_I(W)$ was then found by fitting it to the absolute luminosities of 15 spiral galaxies whose distances were accurately known from observations of Cepheid variables they contain. (These 15 galaxies extended out only to 25 Mpc , not far enough for them to be used to measure the Hubble constant directly.) The Hubble constant could then be found by using the function $L_I(W)$ calibrated in this way to find the distances to galaxies in 14 clusters with redshifts ranging from 0.013 to 0.03 , and comparing the results obtained with Eq. (1.2.8). (These redshifts may not be large enough to ignore peculiar velocities altogether, but again, this problem is mitigated by the use of a fairly large number of galaxies.) The Hubble constant found in this way was 70 ± 5 km s^{-1}Mpc^{-1}. More recently, the Hubble Space Telescope Key Project to Measure the Hubble Constant has used Cepheid variables to recalibrate the Tully-Fisher relation (assuming an LMC distance of 50 kpc) and then found H_0 by plotting distances found from the Tully-Fisher relation against redshift for a sample of 19 clusters with redshifts from 0.007 to 0.03,[51] taken from the G97 survey of Giovanelli et al. [52] The result was $H_0 = 71 \pm 3 \pm 7$ km s^{-1}Mpc^{-1}, with the first quoted uncertainty statistical and the second systematic.

[4746] M. Feast, Lect. Notes Phys. 635, 45 (2003) [astro-ph/0301100].

[47] R. B. Tully and J. R. Fisher, Astron. Astrophys. 54, 661 (1977)

[4748] M. Aaronson, J. R. Mould, and J. Huchra, Astrophys. J. 229, 1 (1979).

[49] Y. Tutui et al., Publ. Astron. Soc. Japan 53, 701 (2001) [astro-ph/0108462].

[50] R. Giovanelli, in The Extragalactic Distance Scale - Proceedings of the Space Telescope Science Institute Symposium held in Baltimore, MD, May, 1996 (Cambridge University Press, Cambridge, 1997): 113; R. Giovanelli et al., Astron. J. 113, 22 (1997).

[51] S. Sakai et al., Astrophys. J. 529, 698 (2000); W. L. Freedman et al., Astrophys. J. 553, 47 (2001); and references cited therein.

[52] Giovanelli et al., op. cit.

2. Faber-Jackson relation

Just as the Tully-Fisher method is based on a correlation of orbital velocities with absolute luminosities in spiral galaxies, the Faber-Jackson method is based on a correlation of random velocities with absolute luminosities in elliptical galaxies. [53] An advantage of this method over the Tully-Fisher method is that it has a firmer theoretical foundation, provided by the virial theorem to be discussed in Section 1.9, which directly relates the mean square random velocity to the galaxy mass.

3. Fundamental plane

The Faber-Jackson method was improved by the recognition that the correlation between orbital velocity and absolute luminosity depends also on the surface brightness of the cluster, and hence on its area. [54] (The term "fundamental plane" refers to the way that data on elliptical galaxies are displayed graphically.) This method has been used [55] to estimate that $H_0 = 78 \pm 5$ (stat.) ± 9 (syst.) kmsec^{-1}Mpc^{-1}.

4. Type Ia supernovae

Supernovae of Type Ia are believed to occur when a white dwarf star in a binary system accretes sufficient matter from its partner to push its mass close to the Chandrasekhar limit, the maximum possible mass that can be supported by electron degeneracy pressure. [56] When this happens the white dwarf becomes unstable, and the increase in temperature and density allows the conversion of carbon and oxygen into ^{56}Ni, triggering a thermonuclear explosion that can be seen at distances of several thousand megaparsecs. The exploding star always has a mass close to the Chandrasekhar limit, so there is little variation in the absolute luminosity of these explosions, making them nearly ideal distance indicators. [57] What variation there is seems

to be well correlated with the rise time and decline time of the supernova light: the slower the decline, the higher the absolute luminosity. [58]

This relation has been calibrated by measurements of Type Ia supernovae in several galaxies of known distance. From 1937 to 1999 there were ten supernovae in galaxies whose distance had been measured by observation of Cepheid variables they contain. [59] Of these, six Type Ia supernovae were used by the HST Key H_0 Group [60] to calibrate the relation between absolute luminosity and decline time. This relation was then used to calculate distances to a sample

[4753] S. M. Faber and R. E. Jackson, Astrophys. J. 204, 668 (1976).

[54] A. Dressler et al., Astrophys. J. 313, 42 (1987).

[55] D. D. Kelson et al., Astrophys. J. 529, 768 (2000) [astro-ph/9909222]; J. P. Blakeslee, J. R. Lucey, J. L. Tonry, M. J. Hudson, V. K. Narayan, and B. J. Barris, Mon. Not. Roy. Astron. Soc. 330, 443 (2002) [astro-ph/011183].

[56] W. A. Fowler and F. Hoyle, Astrophys. J. 132, 565 (1960). Calling a supernova Type I simply means that hydrogen lines are not observed in its spectrum. In addition to Type Ia supernovae, there are other Type I supernovae that occur in the collapse of the cores of stars much more massive than white dwarfs, whose outer layer of hydrogen has been lost in stellar winds, as well as Type II supernovae, produced by core collapse in massive stars that have not lost their outer layer of hydrogen. For a discussion of the Chandrasekhar limit, see G&C, Section 11.3.

[57] The use of Type Ia supernovae as distance indicators was pioneered by A. Sandage and G. A. Tammann, Astrophys. J. 256, 339 (1982), following an earlier observation that they had fairly uniform luminosity by C. T. Kowal, Astron. J. 73, 1021 (1968). In 1982 it was necessary to use brightest supergiant stars as intermediate distance indicators, to bridge the gap between the distances that could then be measured using Cepheids and the distances at which the Type Ia supernova could be found. For reviews of the use of type Ia supernovae as standard candles, see D. Branch, Ann. Rev. Astron. & Astrophys. 36, 17 (1998); P. Höflich, C. Gerardy, E. Linder, and H. Marion, in Stellar Candles, eds. W. Gieren et al. (Lecture Notes in Physics) [astro-ph/0301334].

of 29 Type Ia supernovae in galaxies with redshifts extending from 0.01 to 0.1 , observed at the Cerro Tololo Inter-American Observatory. [61] Plotting these distances against measured redshifts gave a Hubble constant [62] of 71 ± 2 (statistical) ± 6 (systematic) $\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$. This agrees well with an older determination using Type Ia supernovae by a Harvard group, [63] which found $H_0 = 67 \pm 7$ km $\mathrm{s}^{-1}\mathrm{Mpc}^{-1}$. Members of this group have superceded this result, [64] now giving a Hubble constant $H_0 = 73 \pm 4$ (stat.) ± 5 (syst.) $\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$. On the other hand, a group headed by Sandage using Type Ia supernovae and the Tully-Fisher relation has consistently found lower values of H_0.[65] The gap seems to be narrowing; in 2006, this group quoted [66] a value $H_0 = 62.3 \pm 1.3$ (stat.) ± 5.0 (syst.) $\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$. (According to Sandage et al., the difference between these results is due to a difference in the Cepheid period-luminosity relation used to measure distances to the galaxies that host the supernovae that are used to calibrate the relation between supernova absolute luminosity and decline time. Sandage et al. use a metallicity-dependent period-luminosity relation. However Macri et al. [45] subsequently reported no difference in the period-luminosity relation for Cepheids in a metal-rich and a metal-poor region of NGC 4258.)

It is an old hope that with a sufficient theoretical understanding of supernova explosions, it might be possible to measure their distance

without use of primary distance indicators. A 2003 comparison [67] of observed light curves (apparent magnitude as a function of time) and spectra with theory for 26 Type Ia supernovae with redshifts extending up to 0.05 , plus one with redshift 0.38 , gave $H_0 = 67$ $\mathrm{kmsec}^{-1}\mathrm{Mpc}^{-1}$, with a two standard deviation uncertainty of 8 $\mathrm{kmsec}^{-1}\mathrm{Mpc}^{-1}$. It is too soon for this method to replace the older method based on the use of primary distance indicators to calibrate the supernova absolute luminosities, but the agreement between the values of H_0 found in these two ways provides some reassurance that no large error is being made with the older method.

It is instructive to consider a fifth secondary distance indicator that is also used to measure the Hubble constant:

5. Surface brightness fluctuations

In 1988 Tonry and Schneider [68] suggested using the fluctuations in the observed surface brightness of a galaxy from one part of the image to another as a measure of the galaxy's distance. Suppose that the stars in a galaxy can be classified in luminosity classes, all the stars in a luminosity class i having the same absolute luminosity L_i. The rate of receiving energy per unit area of telescope aperture in a small part of the galactic image (as for instance, a single pixel in a charge-coupled device) is

$$\ell = \sum_i \frac{N_i L_i}{4\pi d^2} \tag{1.3.8}$$

[4758] M. Phillips, Astrophys. J. 413, L105 (1993); M. Hamuy et al., Astron. J. 109, 1 (1995); A. Reiss, W. Press, and R. Kirshner, Astrophys. J. 438, L17 (1996); S. Jha, A. Riess, & R. P. Kirshner, Astrophys. J. 659, 122 (20007). A dependence of absolute luminosity on color as well as decline time has been considered by R. Tripp and D. Branch, Astrophys. J. [astro-ph/9904347].

[59] For a list, see Tripp and Branch, op. cit..

[60] B. Gibson et al., Astrophys. J. 529, 723 (2000) [astro-ph/9908149].

[61] M. Hamuy et al., Astron. J. 112, 2398 (1996).

[62] L. Ferrarese et al., in Proceedings of the Cosmic Flows Workshop, eds. S. Courteau et al. (ASP Conference Series) [astro-ph/9909134]; W. L. Freedman et al., Astrophys. J. 553, 47 (2001).

[63] A. G. Riess, W. H. Press, And R. P. Kirshner, Astrophys. J. 438, L17 (1995)

[64] A. Riess et al., Astrophys. J. 627, 579 (2005) [astro-ph/0503159].

[65] For a 1996 summary, see G. A. Tammann and M. Federspeil, in The Extragalactic Distance Scale, eds. M. Livio, M. Donahue, and N. Panagia (Cambridge University Press, 1997): 137.

[66] A. Sandage et al., Astrophys. J. 653, 843 (2006) [astro-ph/0603647].

where N_i the number of stars of class i in this part of the galaxy's image, and d is the distance of the galaxy. Usually only the brightest stars can be resolved, so it is not possible to measure all the N_i directly, but one can measure the fluctuations in ℓ from one part of the image to another due to the finite values of the N_i. Suppose that the different N_i fluctuate independently from one small part of the galaxy's image to another, and obey the rules of Poisson statistics, so that

$$\langle (N_i - \langle N_i \rangle)\,(N_j - \langle N_j \rangle) \rangle = \delta_{ij}\,\langle N_i \rangle\,, \tag{1.3.9}$$

with brackets denoting an average over small parts of the central portion of the galaxy's image. It follows then that

$$\frac{\langle (\ell - \langle \ell \rangle)^2 \rangle}{\langle \ell \rangle} = \frac{\bar{L}}{4\pi d^2} \tag{1.3.10}$$

where $\bar{L}$ is a luminosity-weighted mean stellar luminosity

$$\bar{L} \equiv \frac{\sum_i \langle N_i \rangle\, L_i^2}{\sum_i \langle N_i \rangle\, L_i} \tag{1.3.11}$$

which is expected to vary much less from one galaxy to another than the luminosities of the galaxies themselves. Eq. (1.3.10) can be used to measure distances once this relation is calibrated by measuring $\bar{L}$. By studying surface brightness fluctuations in a survey of galaxies whose distances were found by observations of Cepheids they contain, Tonry et al. [69] found an absolute magnitude $\bar{M}_I$ that in the infrared band is equivalent to the absolute luminosity $\bar{L}$:

$$\bar{M}_I = (-1.74 \pm 0.07) + (4.5 \pm 0.25)\,[m_V - m_I - 1.15] \tag{1.3.12}$$

where $m_V - m_I$ is a parameter characterizing the color of the galaxy, equal to the difference of its apparent magnitudes in the infrared and visual bands, assumed here to lie between 1.0 and 1.5. Using Eq. (1.3.10) to find distances of galaxies of higher redshift, they obtained a Hubble constant 81 ± 6 km s^{-1}Mpc^{-1}.

There are other phenomena that are used to measure the Hubble constant, including the comparison of apparent and absolute luminosity of supernovae of other types, novae, globular clusters, and planetary nebulae, the diameter-velocity dispersion relation for elliptical galaxies, gravitational lenses (discussed in Section 1.12), the Sunyaev-Zel'dovich effect (discussed in Section 2.3), etc. [70] The HST Key H_0 Group have put together their results of measurements of the Hubble constant using the Tully-Fisher relation, Type Ia supernovae, and several of these other secondary distance indicators, and conclude that [71]

$$H_0 = 71 \pm 6 \text{ km s}^{-1}\text{Mpc}^{-1}$$

As we will see in Section 7.2, the study of anisotropies in the cosmic microwave background has given a value $H_0 = 73 \pm 3$ km s^{-1}Mpc^{-1}. This does not depend on any of the tools discussed in this section, but it does depend on some far-reaching cosmological assumptions: including flat spatial geometry, time-independent vacuum energy, and cold dark matter. For this reason, the increasingly precise measurement of H_0 provided by the

[4767] P. Höflich, C. Gerardy, E. Linder, and H. Marion, op. cit.

[68] J. Tonry and D. P. Schneider, Astron. J. 96, 807 (1988).

[4769] J. L. Tonry, J. P. Blakeslee, E. A. Ajhar, and A. Dressler, Astrophys. J. 473, 399 (1997). For a more recent survey, see J. L. Tonry et al., Astrophys. J. 546, 681 (2001) [astro-ph/0011223].

[70] For a survey of most of these methods, with references, see G. H. Jacoby, D. Branch, R. Ciardullo, R. L. Davies, W. E. Harris, M. J. Pierce, C. J. Pritchet, J. L. Tonry, and D. L. Welch, Publ. Astron. Soc. Pacific 104, 599 (1992).

[71] L. Ferrarese et al., op. cit.; W. L. Freedman et al., Astrophys. J. 553, 47 (2001).

cosmic microwave background will not supplant the older measurements discussed in this section - rather, the agreement (or possible future disagreement) between the values of H_0 provided by these very different methods will serve to validate (or possibly invalidate) the cosmological assumptions made in the analysis of the cosmic microwave background.

To take account of the remaining uncertainty in the Hubble constant, it is usual these days to take

$$H_0 = 100h \text{ km s}^{-1}\text{Mpc}^{-1} \tag{1.3.13}$$

with the dimensionless parameter h assumed to be in the neighborhood of 0.7. This corresponds to a Hubble time

$$1/H_0 = 9.778 \times 10^9 h^{-1} \text{ years} \tag{1.3.14}$$

10.1.6 1.4 Luminosity distances and angular diameter distances

We must now consider the measurement of distances at large redshifts, say $z > 0.1$, where the effects of cosmological expansion on the determination of distance can no longer be neglected. It is these measurements that can tell us whether the expansion of the universe is accelerating or decelerating, and how fast. Before we can interpret these measurements, we will need to consider in this section how to define distance at large redshifts, and we will have to apply Einstein's field equations to the Robertson-Walker metric in the following section. After that, we will return in Section 1.6 to the measurements of distances for large redshift, and their interpretation.

In the previous section we derived the familiar relation $\ell = L/4\pi d^2$ for the apparent luminosity ℓ of a source of absolute luminosity L at a distance d. At large distances this derivation needs modification for three reasons:

1. At the time t_0 that the light reaches earth, the proper area of a sphere drawn around the luminous object and passing through the earth is given by the metric (1.1.10) as $4\pi r_1^2 a^2(t_0)$, where r_1 is the coordinate distance of the earth as seen from the luminous object, which is just the same as the coordinate distance of the luminous object as seen from the earth. The fraction of the light received in a telescope of aperture A on earth is therefore $A/4\pi r_1^2 a^2(t_0)$, and so the factor $1/d^2$ in the formula for ℓ must be replaced with $1/r_1^2 a^2(t_0)$.

2. The rate of arrival of individual photons is lower than the rate at which they are emitted by the redshift factor $a(t_1)/a(t_0) = 1/(1+z)$.

3. The energy hv_0 of the individual photons received on earth is less than the energy hv_1 with which they were emitted by the same redshift factor $1/(1+z)$.

Putting this together gives the correct formula for apparent luminosity of a source at radial coordinate r_1 with a redshift z of any size:

$$\ell = \frac{L}{4\pi r_1^2 a^2(t_0)(1+z)^2} \tag{1.4.1}$$

It is convenient to introduce a "luminosity distance" d_L, which is defined so that the relation between apparent and absolute luminosity and luminosity distance is the same as Eq. (1.3.3):

$$\ell = \frac{L}{4\pi d_L^2} \tag{1.4.2}$$

Eq. (1.4.1) can then be expressed as

$$d_L = a\left(t_0\right) r_1 (1+z) \tag{1.4.3}$$

For objects with $z \ll 1$, we can usefully write the relation between luminosity distance and redshift as a power series. The redshift $1+z \equiv a\left(t_0\right)/a\left(t_1\right)$ is related to the "look-back time" $t_0 - t_1$ by

$$z = H_0\left(t_0 - t_1\right) + \frac{1}{2}\left(q_0+2\right) H_0^2 \left(t_0 - t_1\right)^2 + \ldots \tag{1.4.4}$$

where H_0 is the Hubble constant (1.2.7) and q_0 is the deceleration parameter

$$q_0 \equiv \frac{-1}{H_0^2 a\left(t_0\right)} \frac{d^2 a(t)}{dt^2}\bigg|_{t=t_0} \tag{1.4.5}$$

This can be inverted, to give the look-back time as a power series in the redshift

$$H_0\left(t_0 - t_1\right) = z - \frac{1}{2}\left(q_0+2\right) z^2 + \ldots \tag{1.4.6}$$

The coordinate distance r_1 of the luminous object is given by Eq. (1.2.2) as

$$\frac{t_0 - t_1}{a\left(t_0\right)} + \frac{H_0\left(t_0 - t_1\right)^2}{2a\left(t_0\right)} + \cdots = r_1 + \ldots \tag{1.4.7}$$

with the dots on the right-hand side denoting terms of third and higher order in r_1. Using Eq. (1.4.6), the solution is

$$r_1 a\left(t_0\right) H_0 = z - \frac{1}{2}\left(1+q_0\right) z^2 + \cdots \tag{1.4.8}$$

This gives the luminosity distance (1.4.3) as a power series

$$d_L = H_0^{-1}\left[z + \frac{1}{2}\left(1-q_0\right) z^2 + \cdots\right]. \tag{1.4.9}$$

We can therefore measure q_0 as well as H_0 by measuring the luminosity distance as a function of redshift to terms of order z^2. The same reasoning has been used to extend the expression (1.4.9) to fourth order in z :[1]

$$\begin{aligned} d_L(z) =& H_0^{-1}\left[z + \frac{1}{2}\left(1-q_0\right) z^2 - \frac{1}{6}\left(1 - q_0 - 3q_0^2 + j_0 + \frac{K}{H_0^2 a_0^2}\right) z^3 \right. \\ &+ \frac{1}{24}\left(2 - 2q_0 - 15q_0^2 - 15q_0^3 + 5j_0 + 10q_0 j_0 \right. \\ &\left.\left. + s_0 + \frac{2K\left(1+3q_0\right)}{H_0^2 a_0^2}\right) z^4 + \cdots\right], \end{aligned}$$

where j_0 and s_0 are parameters known as the jerk and snap:

$$j_0 \equiv \frac{1}{H_0^3 a\left(t_0\right)} \frac{d^3 a(t)}{dt^3}\bigg|_{t=t_0}, \quad s_0 \equiv \frac{1}{H_0^4 a\left(t_0\right)} \frac{d^4 a(t)}{dt^4}\bigg|_{t=t_0}.$$

Years ago cosmology was called "a search for two numbers," H_0 and q_0. The determination of H_0 is still a major goal of astronomy, as discussed in the previous section. On the other hand, there is less interest now in q_0. Instead of high-precision distance determinations at moderate redshifts, of order 0.1 to 0.2 , which would give an accurate value of q_0, we now have distance determinations of only moderate precision at high redshifts, of order unity, which depend on

the whole form of the function $a(t)$ over the past few billion years. For redshifts of order unity, it is not very useful to expand in powers of redshift. In order to interpret these measurements, we will need a dynamical theory of the expansion, to be developed in the next section. As we will see there, modern observations suggest strongly that there are not two but at least three parameters that need to be measured to calculate $a(t)$.

Before turning to this dynamical theory, let's pause a moment to clarify the distinction between different measures of distance. So far, we have encountered the proper distance (1.1.15) and the luminosity distance (1.4.3). There is another sort of distance, which is what we measure when we compare angular sizes with physical dimensions. Inspection of the metric

(1.1.12) shows that a source at co-moving radial coordinate r_1 that emits light at time t_1 and is observed at present to subtend a small angle θ will extend over a proper distance s (normal to the line of sight) equal to $a(t_1)\, r_1\theta$. The angular diameter distance d_A is defined so that θ is given by the usual relation of Euclidean geometry

$$\theta = s/d_A \tag{1.4.10}$$

and we see that

$$d_A = a(t_1)\, r_1 \tag{1.4.11}$$

Comparison of this result with Eq. (1.4.3) shows that the ratio of the luminosity and angular-diameter distances is simply a function of redshift:

$$d_A/d_L = (1+z)^{-2} \tag{1.4.12}$$

Therefore if we have measured the luminosity distance at a given redshift (and if we are convinced of the correctness of the Robertson-Walker metric), then we learn nothing additional about $a(t)$ if we also measure the angular diameter distance at that redshift. Neither galaxies nor supernovas have well-defined edges, so angular diameter distances are much less useful in studying the cosmological expansion than are luminosity distances. However, as we shall see, they play an important role in the theoretical analysis of both gravitational lenses in Chapter 9 and of the fluctuations in the cosmic microwave radiation background in Chapters 2 and 7 . We will see in Section 8.1 that the observation of acoustic oscillations in the matter density may allow a measurement of yet another distance, a structure distance, equal to $a(t_0)\, r_1 = (1+z)d_A$.

10.1.7 1.5 Dynamics of expansion

All our results up to now have been very general, not depending on assumptions about the dynamics of the cosmological expansion. To go further we will need now to apply the gravitational field equations of Einstein, with various tentative assumptions about the cosmic energy density and pressure.

The expansion of the universe is governed by the Einstein field equations (B.71), which can be put in the convenient form

$$R_{\mu\nu} = -8\pi G S_{\mu\nu} \tag{1.5.1}$$

where $R_{\mu\nu}$ is the Ricci tensor:

$$R_{\mu\nu} = \frac{\partial \Gamma^\lambda_{\lambda\mu}}{\partial x^\nu} - \frac{\partial \Gamma^\lambda_{\mu\nu}}{\partial x^\lambda} + \Gamma^\lambda_{\mu\sigma}\Gamma^\sigma_{\nu\lambda} - \Gamma^\lambda_{\mu\nu}\Gamma^\sigma_{\lambda\sigma} \tag{1.5.2}$$

[471] M. Visser, Class. Quant. Grav. 21, 2603 (2004) [gr-qc/0309109]. The term of third order in z was previously calculated by T. Chiba and T. Nakamura, Prog. Theor. Phys. 100, 1077 (1998).

and $S_{\mu\nu}$ is given in terms of the energy-momentum tensor $T_{\mu\nu}$ by

$$S_{\mu\nu} \equiv T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T^{\lambda}_{\lambda} \tag{1.5.3}$$

As we saw in Section 1.1, for the Robertson-Walker metric the components of the affine connection with two or three time indices all vanish, so

$$\begin{aligned} R_{ij} =& \frac{\partial \Gamma^k_{ki}}{\partial x^j} - \left[\frac{\partial \Gamma^k_{ij}}{\partial x^k} + \frac{\partial \Gamma^0_{ij}}{\partial t} \right] \\ &+ \left[\Gamma^0_{ik} \Gamma^k_{j0} + \Gamma^k_{i0} \Gamma^0_{jk} + \Gamma^l_{ik} \Gamma^k_{jl} \right] \\ &- \left[\Gamma^k_{ij} \Gamma^l_{kl} + \Gamma^0_{ij} \Gamma^l_{0l} \right] \end{aligned} \tag{1.5.4}$$

$$R_{00} = \frac{\partial \Gamma^i_{i0}}{\partial t} + \Gamma^i_{0j} \Gamma^j_{0i} \tag{1.5.5}$$

We don't need to calculate $R_{i0} = R_{0i}$, because it is a three-vector, and therefore must vanish due to the isotropy of the Robertson-Walker metric. Using the formulas (1.1.17)-(1.1.19) for the non-vanishing components of the affine connection gives

$$\begin{array}{ll} \frac{\partial \Gamma^0_{ij}}{\partial t} = \tilde{g}_{ij} \frac{d}{dt}(a\dot{a}), & \Gamma^0_{ik} \Gamma^k_{j0} = \tilde{g}_{ij} \dot{a}^2, \quad \Gamma^0_{ij} \Gamma^l_{0l} = 3 \tilde{g}_{ij} \dot{a}^2 \\ \frac{\partial \Gamma^i_{i0}}{\partial t} = 3 \frac{d}{dt} \left(\frac{\dot{a}}{a} \right), & \Gamma^i_{0j} \Gamma^j_{i0} = 3 \left(\frac{\dot{a}}{a} \right)^2 \end{array} \tag{1.5.6}$$

where dots denote time derivatives. Using this in Eqs. (1.5.4) and (1.5.5), we find that the non-vanishing components of the Ricci tensor are

$$R_{ij} = \tilde{R}_{ij} - 2\dot{a}^2 \tilde{g}_{ij} - a\ddot{a} \tilde{g}_{ij} \tag{1.5.7}$$

$$R_{00} = 3 \frac{d}{dt} \left(\frac{\dot{a}}{a} \right) + 3 \left(\frac{\dot{a}}{a} \right)^2 = 3 \frac{\ddot{a}}{a} \tag{1.5.8}$$

where $\tilde{R}_{ij}$ is the purely spatial Ricci tensor

$$\tilde{R}_{ij} = \frac{\partial \Gamma^k_{ki}}{\partial x^j} - \frac{\partial \Gamma^k_{ij}}{\partial x^k} + \Gamma^l_{ik} \Gamma^k_{jl} - \Gamma^l_{ij} \Gamma^k_{kl}. \tag{1.5.9}$$

According to Eq. (1.1.19), the spatial components Γ^i_{jk} of the four-dimensional affine connection are here the same as those of the affine connection that would be calculated in three dimensions from the three-metric $\tilde{g}_{ij}$:

$$\Gamma^k_{ij} = K x^k \tilde{g}_{ij} \tag{1.5.10}$$

To calculate $\tilde{R}_{ij}$, we use a trick used earlier in calculating particle trajectories: we calculate $\tilde{R}_{ij}$ where the calculation is simplest, at $\mathbf{x} = 0$, and express the result as a relation that is invariant under all transformations of the spatial coordinates, so that the homogeneity of the three-dimensional metric insures that this relation is valid everywhere. The spatial Ricci tensor at $\mathbf{x} = 0$ is

$$\tilde{R}_{ij} = \frac{\partial \Gamma^l_{li}}{\partial x^j} - \frac{\partial \Gamma^l_{ji}}{\partial x^l} = K\delta_{ij} - 3K\delta_{ij} = -2K\delta_{ij} \tag{1.5.11}$$

At $\mathbf{x} = 0$ the spatial metric $\tilde{g}_{ij}$ is just δ_{ij}, so this can be rewritten as

$$\tilde{R}_{ij} = -2K \tilde{g}_{ij} \tag{1.5.12}$$

which, since it is an equality between two three-tensors, is then true in all spatial coordinate systems, including systems in which the point $\mathbf{x} = 0$ is transformed into any other point. Hence Eq. (1.5.12) is true everywhere, and together with Eq. (1.5.7) gives

$$R_{ij} = -\left[2K + 2\dot{a}^2 + a\ddot{a}\right]\tilde{g}_{ij} \tag{1.5.13}$$

We also need the values of S_{ij} and S_{00}. For this, we use Eq. (1.1.31) in the form

$$T_{00} = \rho, \quad T_{i0} = 0, \quad T_{ij} = a^2 p\tilde{g}_{ij} \tag{1.5.14}$$

where $\rho(t)$ and $p(t)$ are the proper energy density and pressure. Eq. (1.5.3) gives then

$$S_{ij} = T_{ij} - \frac{1}{2}\tilde{g}_{ij}a^2\left(T^k{}_k + T^0_0\right) = a^2 p\tilde{g}_{ij} - \frac{1}{2}a^2\tilde{g}_{ij}(3p - \rho) = \frac{1}{2}(\rho - p)a^2\tilde{g}_{ij} \tag{1.5.15}$$

$$S_{00} = T_{00} + \frac{1}{2}\left(T^k{}_k + T^0_0\right) = \rho + \frac{1}{2}(3p - \rho) = \frac{1}{2}(\rho + 3p) \tag{1.5.16}$$

and $S_{i0} = 0$. The Einstein equations are therefore

$$-\frac{2K}{a^2} - \frac{2\dot{a}^2}{a^2} - \frac{\ddot{a}}{a} = -4\pi G(\rho - p) \tag{1.5.17}$$

$$\frac{3\ddot{a}}{a} = -4\pi G(3p + \rho) \tag{1.5.18}$$

We can eliminate the second derivative terms by adding three times the first equation to the second, and find

$$\dot{a}^2 + K = \frac{8\pi G\rho a^2}{3} \tag{1.5.19}$$

This is the fundamental Friedmann equation [1] governing the expansion of the universe.

The remaining information in Eqs. (1.5.17) and (1.5.18) just reproduces the conservation law (1.1.32):

$$\dot{\rho} = -\frac{3\dot{a}}{a}(\rho + p) \tag{1.5.20}$$

(This should come as no surprise. Under all circumstances, the energymomentum conservation law may be derived as a consequence of the Einstein field equations.) Given p as a function of ρ, we can solve Eq. (1.5.20) to find ρ as a function of a, and then use this in Eq. (1.5.19) to find a as a function of t.

There is another way of deriving Eq. (1.5.19), at least for the case of nonrelativistic matter. Imagine a co-moving ball cut out from the expanding universe, with some typical galaxy at its center, and suppose it then emptied of the matter it contains. According to Birkhoff's theorem, [2] in any system that is spherically symmetric around some point, the metric in an empty ball centered on this point must be that of flat space. This holds whatever is happening outside the empty ball, as long as it is spherically symmetric. Now imagine putting the matter back in the ball, with a velocity proportional to distance from the center of symmetry, taken as $\mathbf{X} = 0$:

$$\dot{\mathbf{X}} = H(t)\mathbf{X} \tag{1.5.21}$$

(Here the components X^i of $\mathbf{X}$ are ordinary Cartesian coordinates, not the co-moving coordinates x^i used in the Robertson-Walker metric. Note that this is the one pattern of velocities consistent with the principle of homogeneity: The velocity of a co-moving particle at $\mathbf{X}_1$ relative to a co-moving particle at $\mathbf{X}_2$ is $\dot{\mathbf{X}}_1 - \dot{\mathbf{X}}_2 = H(t)\left(\mathbf{X}_1 - \mathbf{X}_2\right)$.) The solution of Eq. (1.5.21) is

$$\mathbf{X}(t) = \left(\frac{a(t)}{a\left(t_0\right)} \right) \mathbf{X}\left(t_0\right), \tag{1.5.22}$$

where $a(t)$ is the solution of the equation

$$\dot{a}(t)/a(t) = H(t) \tag{1.5.23}$$

As long as the radius of the ball is chosen to be not too large, the expansion velocity (1.5.21) of the matter we put into it will be non-relativistic, and the gravitational field will be weak, so that we can follow its motion using Newtonian mechanics. The kinetic energy of a co-moving particle of mass m at $\mathbf{X}$ is

$$K.E. = \frac{1}{2} m \dot{\mathbf{X}}^2 = \frac{m \dot{a}^2 \mathbf{X}^2}{2a^2} \tag{1.5.24}$$

The mass interior to the position of the particle is $M(\mathbf{X}) = 4\pi\rho|\mathbf{X}|^3/3$, so the potential energy of the particle is

$$P.E. = -\frac{GmM(\mathbf{X})}{|\mathbf{X}|} = -\frac{4\pi G m \rho |\mathbf{X}|^2}{3} \tag{1.5.25}$$

The condition of energy conservation thus tells us that

$$E = K.E. + P.E. = \frac{m\left|\mathbf{X}\left(t_0\right)\right|^2}{a^2\left(t_0\right)} \left[\frac{\dot{a}^2}{2} - \frac{4\pi G \rho a^2}{3} \right] = \text{ constant} \tag{1.5.26}$$

This is the same as Eq. (1.5.19), providing we identify the particle energy as

$$E = -\frac{Km\left|\mathbf{X}\left(t_0\right)\right|^2}{2a^2\left(t_0\right)} \tag{1.5.27}$$

Particles will be able to escape to infinity if and only if $E \geq 0$, which requires $K = 0$ or $K = -1$. For $K = +1$ they have less than escape velocity, so the expansion eventually stops, and particles fall back toward each other.

Returning now to the relativistic formalism and an arbitrary dependence of ρ on a, even without knowing this dependence we can use Eq. (1.5.19) to draw important consequences about the general features of the expansion. First, as long as ρ remains positive, it is only possible for the expansion of the universe to stop if $K = +1$, the case of spherical geometry. Also, for any value of the Hubble constant $H_0 \equiv \dot{a}\left(t_0\right)/a\left(t_0\right)$, we may define a critical present density

$$\rho_{0,\text{ crit}} \equiv \frac{3H_0^2}{8\pi G} = 1.878 \times 10^{-29} h^2 \text{ g/cm}^3 \tag{1.5.28}$$

where h is the Hubble constant in units of 100 km s^{-1}Mpc^{-1}. According to Eq. (1.5.19), whatever we assume about the constituents of the universe, the curvature constant K will be $+1$ or 0 or -1 according to whether the present density ρ_0 is greater than, equal to, or less than $\rho_{0,\text{ crit}}$. If the quantity $3p + \rho$ is positive (as it is for any mixture of matter and radiation, in the absence of a vacuum energy density) then Eq. (1.5.18) shows that $\ddot{a}/a \leq 0$, so the expansion must have started with $a = 0$ at some moment in the past; the present age of the universe t_0 is less than the Hubble time

$$t_0 < H_0^{-1} \tag{1.5.29}$$

[471] A. Friedmann, Z. Phys. 16, 377 (1922); ibid 21, 326 (1924).

[2] G&C, Section 11.7.

Also, if $K = +1$ and the expansion stops, then with $\ddot{a}/a \leq 0$ the universe will again contract to a singularity at which $a = 0$.

We can use Eq. (1.5.18) to give a general formula for the deceleration parameter $q_0 \equiv -\ddot{a}\left(t_0\right) a\left(t_0\right) / \dot{a}^2\left(t_0\right)$:

$$q_0 = \frac{4\pi G\left(\rho_0 + 3p_0\right)}{3H_0^2} = \frac{\rho_0 + 3p_0}{2\rho_{0,\text{crit}}} \tag{1.5.30}$$

with a subscript 0 denoting a present value. If the present density of the universe were dominated by non-relativistic matter then $p_0 \ll \rho_0$, and the curvature constant K would be $+1$ or 0 or -1 according to whether $q_0 > \frac{1}{2}$ or $q_0 = \frac{1}{2}$ or $q_0 < \frac{1}{2}$. On the other hand, if the present density of the universe were dominated by relativistic matter then $p_0 = \rho_0/3$, and the critical value of the deceleration parameter at which $K = 0$ would be $q_0 = 1$. Finally, if the present density of the universe were dominated by vacuum energy then $p_0 = -\rho_0$, and the value of the deceleration parameter at which $K = 0$ would be $q_0 = -1$.

There is a peculiar aspect to these results. The contribution of nonrelativistic and relativistic matter to the quantity ρa^2 in Eq. (1.5.19) grows as a^{-1} and a^{-2}, respectively, as $a \to 0$, so at sufficiently early times in the expansion we may certainly neglect the constant K, and Eq. (1.5.19) gives

$$\frac{\dot{a}^2}{a^2} \to \frac{8\pi G\rho}{3} \tag{1.5.31}$$

That is, at these early times the density becomes essentially equal to the critical density $3H^2/8\pi G$, where $H \equiv \dot{a}/a$ is the value of the Hubble "constant" at those times. On the other hand, we will see later that the total energy density of the present universe is still a fair fraction of the critical density. How is it that after billions of years, ρ is still not very different from ρ_{crit} ? This is sometimes called the flatness problem.

The simplest solution to the flatness problem is just that we are in a spatially flat universe, in which $K = 0$ and ρ is always precisely equal to ρ_{crit} . A more popular solution is provided by the inflationary theories discussed in Chapter 4. In these theories K may not vanish, and ρ may not start out close to ρ_{crit} , but there is an early period of rapid growth in which ρ/ρ_{crit} rapidly approaches unity. In inflationary theories it is expected though not required that ρ should now be very close to ρ_{crit} , in which case it is a good approximation to take $K = 0$.

For $K = 0$ we get very simple solutions to Eq. (1.5.19) in the three special cases listed in Section 1.1:

Non-relativistic matter: Here $\rho = \rho_0\left(a/a_0\right)^{-3}$, and the solution of Eq. (1.5.19) with $K = 0$ is

$$a(t) \propto t^{2/3} \tag{1.5.32}$$

This gives $q_0 \equiv -a\ddot{a}/\dot{a}^2 = 1/2$, and a simple relation between the age of the universe and the Hubble constant

$$t_0 = \frac{2}{3H_0} = 6.52 \times 10^9 h^{-1}\text{yr} \tag{1.5.33}$$

Eqs. (1.5.32) and (1.5.18) show that for $K = 0$, the energy density at time t is $\rho = 1/6\pi G t^2$. This is known as the Einstein-de Sitter model. It was for many years the most popular cosmological model, though as we shall see, the age (1.5.33) is uncomfortably short compared with the ages of certain stars.

Relativistic matter: Here $\rho = \rho_0\left(a/a_0\right)^{-4}$, and the solution of Eq. (1.5.19) with $K = 0$ is

$$a(t) \propto \sqrt{t} \tag{1.5.34}$$

This gives $q_0 = +1$, while the age of the universe and the Hubble constant are related by

$$t_0 = \frac{1}{2H_0} \tag{1.5.35}$$

The energy density at time t is $\rho = 3/32\pi G t^2$.
Vacuum energy: Lorentz invariance requires that in locally inertial coordinate systems the energy-momentum tensor $T_V^{\mu\nu}$ of the vacuum must be proportional to the Minkowski metric $\eta^{\mu\nu}$ (for which $\eta^{ij} = \eta_{ij} = \delta_{ij}, \eta^{i0} = \eta_{i0} = \eta^{0i} = \eta_{0i} = 0, \eta^{00} = \eta_{00} = -1$), and so in general coordinate systems $T_V^{\mu\nu}$ must be proportional to $g^{\mu\nu}$. Comparing this with Eq. (B.43) shows that the vacuum has $p_V = -\rho_V$, so that $T_V^{\mu\nu} = -\rho_V g^{\mu\nu}$. In the absence of any other form of energy this would satisfy the conservation law $0 = T_{V\ ;\mu}^{\mu\nu} = g^{\mu\nu}\partial\rho_V/\partial x^\mu$, so that ρ_V would be a constant, independent of spacetime position. Eq. (1.5.19) for $K = 0$ requires that $\rho_V > 0$, and has the solutions

$$a(t) \propto \exp(Ht) \tag{1.5.36}$$

where H is the Hubble constant, now really a constant, given by

$$H = \sqrt{\frac{8\pi G\rho_V}{3}} \tag{1.5.37}$$

Here $q_0 = -1$, and the age of the universe in this case is infinite. This is known as the de Sitter model. [3] Of course, there is some matter in the universe, so even if the energy density of the universe is now dominated by a constant vacuum energy, there was a time in the past when matter and/or radiation were more important, and so the expansion has a finite age, although greater than it would be without a vacuum energy.

More generally, for arbitrary K and a mixture of vacuum energy and relativistic and non-relativistic matter, making up fractions Ω_Λ, Ω_M, and Ω_R of the critical energy density, [4] we have

$$\rho = \frac{3H_0^2}{8\pi G}\left[\Omega_\Lambda + \Omega_M\left(\frac{a_0}{a}\right)^3 + \Omega_R\left(\frac{a_0}{a}\right)^4\right], \tag{1.5.38}$$

where the present energy densities in the vacuum, non-relativistic matter, and and relativistic matter (i.e., radiation) are, respectively,

$$\rho_{V0} \equiv \frac{3H_0^2\Omega_\Lambda}{8\pi G}, \quad \rho_{M0} \equiv \frac{3H_0^2\Omega_M}{8\pi G}, \quad \rho_{R0} \equiv \frac{3H_0^2\Omega_R}{8\pi G} \tag{1.5.39}$$

and, according to Eq. (1.5.19),

$$\Omega_\Lambda + \Omega_M + \Omega_R + \Omega_K = 1, \quad \Omega_K \equiv -\frac{K}{a_0^2H_0^2} \tag{1.5.40}$$

Using this in Eq. (1.5.19) gives

$$\begin{aligned} dt &= \frac{dx}{H_0 x\sqrt{\Omega_\Lambda + \Omega_K x^{-2} + \Omega_M x^{-3} + \Omega_R x^{-4}}} \\ &= \frac{-dz}{H_0(1+z)\sqrt{\Omega_\Lambda + \Omega_K(1+z)^2 + \Omega_M(1+z)^3 + \Omega_R(1+z)^4}} \end{aligned} \tag{1.5.41}$$

where $x \equiv a/a_0 = 1/(1+z)$. Therefore, if we define the zero of time as corresponding to an infinite redshift, then the time at which light was emitted that reaches us with redshift z is given by

$$t(z) = \frac{1}{H_0}\int_0^{1/(1+z)} \frac{dx}{x\sqrt{\Omega_\Lambda + \Omega_K x^{-2} + \Omega_M x^{-3} + \Omega_R x^{-4}}} \tag{1.5.42}$$

In particular, by setting $z = 0$, we find the present age of the universe:

$$t_0 = \frac{1}{H_0}\int_0^{1} \frac{dx}{x\sqrt{\Omega_\Lambda + \Omega_K x^{-2} + \Omega_M x^{-3} + \Omega_R x^{-4}}} \tag{1.5.43}$$

In order to calculate luminosity or angular diameter distances, we also need to know the radial coordinate $r(z)$ of a source that is observed now with redshift z. According to Eqs. (1.2.2) and (1.5.41), this is

$$\begin{aligned} r(z) &= S\left[\int_{t(z)}^{t_0} \frac{dt}{a(t)}\right] \\ &= S\left[\frac{1}{a_0 H_0}\int_{1/(1+z)}^{1} \frac{dx}{x^2\sqrt{\Omega_\Lambda + \Omega_K x^{-2} + \Omega_M x^{-3} + \Omega_R x^{-4}}}\right] \end{aligned}$$

where

$$S[y] \equiv \begin{cases} \sin y & K = +1 \\ y & K = 0 \\ \sinh y & K = -1 \end{cases}$$

This can be written more conveniently by using Eq. (1.5.40) to express $a_0 H_0$ in terms of Ω_K. We then have a single formula

$$\begin{aligned} a_0 r(z) =& \frac{1}{H_0 \Omega_K^{1/2}} \\ & \times \sinh\left[\Omega_K^{1/2}\int_{1/(1+z)}^{1} \frac{dx}{x^2\sqrt{\Omega_\Lambda + \Omega_K x^{-2} + \Omega_M x^{-3} + \Omega_R x^{-4}}}\right] \end{aligned} \tag{1.5.44}$$

which can be used for any curvature. (Eq. (1.5.43) has a smooth limit for $\Omega_K \to 0$, which gives the result for zero curvature. Also, for $\Omega_K < 0$, the argument of the hyperbolic sine is imaginary, and we can use $\sinh ix = i \sin x$.) Using Eq. (1.5.44) in Eq. (1.4.3) gives the luminosity distance of a source observed with redshift z as

$$\begin{aligned} d_L(z) =& a_0 r(z)(1+z) = \frac{1+z}{H_0 \Omega_K^{1/2}} \\ & \times \sinh\left[\Omega_K^{1/2}\int_{1/(1+z)}^{1} \frac{dx}{x^2\sqrt{\Omega_\Lambda + \Omega_K x^{-2} + \Omega_M x^{-3} + \Omega_R x^{-4}}}\right] \end{aligned} \tag{1.5.45}$$

For $K = 0$ we have $\Omega_K = 0$ and Eq. (1.5.45) becomes

$$d_L(z) = a_0 r_1(1+z) = \frac{1+z}{H_0}\int_{1/(1+z)}^{1} \frac{dx}{x^2\sqrt{\Omega_\Lambda + \Omega_M x^{-3} + \Omega_R x^{-4}}} \tag{1.5.46}$$

[473] W. de Sitter, Proc. Roy. Acad. Sci. (Amsterdam), 19, 1217 (1917); ibid. 20, 229 (1917); ibid. 20, 1309 (1917); Mon. Not. Roy. Astron. Soc., 78, 2 (1917).

[4] The use of the symbol Ω_Λ instead of Ω_V for the ratio of the vacuum energy density to the critical energy density has become standard, because of a connection with the cosmological constant discussed in a historical note below.

As we will see in Section 2.1, Ω_R is much less than Ω_M, and the integral (1.5.46) converges at its lower bound for $z \to \infty$ whether or not Ω_R vanishes, so it is a good approximation to take $\Omega_R = 0$ here.

It is of some interest to express the deceleration parameter q_0 in terms of the Ω s. The p/ρ ratio w for vacuum, matter, and radiation is $-1, 0$, and $1/3$, respectively, so Eq. (1.5.39) gives the present pressure as

$$p_0 = \frac{3H_0^2}{8\pi G}\left(-\Omega_\Lambda + \frac{1}{3}\Omega_R\right). \tag{1.5.47}$$

Eq. (1.5.30) then gives

$$q_0 = \frac{4\pi G\left(3p_0 + \rho_0\right)}{3H_0^2} = \frac{1}{2}\left(\Omega_M - 2\Omega_\Lambda + 2\Omega_R\right) \tag{1.5.48}$$

One of the reasons for our interest in the values of Ω_K, Ω_M, etc. is that they tell us whether the present expansion of the universe will ever stop. According to Eq. (1.5.38), the expansion can only stop if there is a real root of the cubic equation

$$\Omega_\Lambda u^3 + \Omega_K u + \Omega_M = 0 \tag{1.5.49}$$

where $u \equiv a(t)/a\left(t_0\right)$ is greater than one. (We are ignoring radiation here, since it will become even less important as the universe expands.) This expression has the value $+1$ for $u = 1$. If $\Omega_\Lambda < 0$ then the left-hand side of Eq. (1.5.49) becomes negative for sufficiently large u, so it must take the value zero at some intermediate value of u, and the expansion will stop when this value of u is reached. Even for $\Omega_\Lambda \geq 0$ it is still possible for the expansion to stop, provided $\Omega_K = 1 - \Omega_\Lambda - \Omega_M$ is sufficiently negative (which, among other things, requires that $K = +1$).

Historical Note 1: If we express the total energy momentum tensor $T_{\mu\nu}$ as the sum of a possible vacuum term $-\rho_V g_{\mu\nu}$ and a term $T_{\mu\nu}^M$ arising from matter (including radiation), then the Einstein equations take the form

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R_\lambda^\lambda = -8\pi G T_{\mu\nu}^M + 8\pi G\rho_V g_{\mu\nu}. \tag{1.5.50}$$

Thus the effect of a vacuum energy is equivalent to modifying the Einstein field equations to read

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R - \Lambda g_{\mu\nu} = -8\pi G T_{\mu\nu}^M \tag{1.5.51}$$

where

$$\Lambda = 8\pi G\rho_V \tag{1.5.52}$$

The quantity Λ is known as the cosmological constant. It was introduced into the field equation by Einstein in 1917 in order to satisfy a condition that at the time was generally regarded as essential, that the universe should be static. [5] According to Eqs. (1.5.18) and (1.5.19), a static universe is only possible if $3p + \rho = 0$ and $K = 8\pi G\rho a^2/3$. If the contents of the universe are limited to vacuum energy and non-relativistic matter, then $\rho = \rho_M + \rho_V$, $p = -\rho_V$, and $\rho_M \geq 0$. It follows that $\rho_M = 2\rho_V \geq 0$, so $K > 0$, which by convention means $K = +1$, so that a takes the value $a_E = 1/\sqrt{8\pi G\rho_V} = 1/\sqrt{\Lambda}$. This is known as the Einstein model.

Einstein did not realize it, but his cosmology was unstable: If a is a little less than a_E then ρ_M is a little larger than $2\rho_V$, so Eq. (1.5.18) shows that $\ddot{a}/a < 0$, and a thus begins to decrease. Likewise, if a is a little greater than a_E then it begins to increase. The models with $K = +1$

and $\Lambda > 0$ in which a starts at the Einstein radius $a = a_E$ with $\rho_M = 2\rho_V$ and then expands to infinity (or starts at $a = 0$ and approaches a_E as $t \to \infty$ with just enough matter so that $\rho_M = 2\rho_V$ at the Einstein radius), are known as Eddington-Lemaître models. [6] There are also models with $K = +1$ and a little more matter, that start at $a = 0$, spend a long time near the Einstein radius, and then expand again to infinity, approaching a de Sitter model. These are known as Lemaître models. [7]

Oddly, de Sitter also invented his cosmological model (with $a \propto \exp(Ht)$) in order to satisfy a supposed need for a static universe. He originally proposed a time-independent metric, given by

$$d\tau^2 = \left(1 - r^2/R^2\right) dt^2 - \frac{dr^2}{1 - r^2/R^2} - r^2 d\theta^2 - r^2 \sin^2\theta d\phi^2 \tag{1.5.53}$$

or equivalently, setting $r = R \sin\chi$,

$$d\tau^2 = \cos^2\chi dt^2 - R^2 \left(d\chi^2 + \sin^2\chi d\theta^2 + \sin^2\chi \sin^2\theta d\phi^2\right) \tag{1.5.54}$$

with $R = \sqrt{3/\Lambda}$ constant. De Sitter did not realize at first that this metric has $\Gamma^i_{00} \neq 0$, so that his coordinate system was not co-moving. [8] Only later was it noticed that, using co-moving spatial coordinates and cosmological standard time, de Sitter's model is equivalent to a Robertson-Walker metric with $K = 0$ and $a \propto \exp(t/R)$.

After the discovery of the expansion of the universe, cosmologists lost interest in a static universe, and Einstein came to regret his introduction of a cosmological constant, calling it his greatest mistake. But as we shall see in the next section, there are theoretical reasons to expect a non-vanishing vacuum energy, and there is observational evidence that in fact it does not vanish. Einstein's mistake was not that he introduced the cosmological constant - it was that he thought it was a mistake.

Historical Note 2: There is a cosmological model due to Bondi and Gold [9] and in a somewhat different version to Hoyle, [10] known as the steady state theory. In this model nothing physical changes with time, so the Hubble constant really is constant, and hence $a(t) \propto \exp(Ht)$, just as in the de Sitter model. To keep the curvature constant, it is necessary to take $K = 0$. In this model new matter must be continually created to keep ρ constant as the universe expands. Since the discovery of the cosmic microwave background (discussed in Chapter 2) the steady state theory in its original form has been pretty well abandoned.

10.1.8 1.6 Distances at large redshift: Accelerated expansion

We now return to our account of the measurement of distances as a function of redshift, considering now redshifts $z > 0.1$, which are large enough so that we can ignore the peculiar motions of the light sources, and also large enough so that we need to take into account the effects of cosmological expansion on distance determination.

For many years, the chief "standard candles" used at large redshift were the brightest galaxies in rich clusters. It is now well established that the

[475] A. Einstein, Sitz. Preuss. Akad. Wiss. 142 (1917). For an English translation, see The Principle of Relativity (Methuen, 1923; reprinted by Dover Publications, New York, 1952), p. 35.

[6] A. S. Eddington, Mon. Not. Roy. Astron. Soc. 90, 668 (1930); G. Lemaître, Ann. Soc. Sci. Brux. A47, 49 (1927); Mon. Not. Roy. Astron. Soc. 91, 483 (1931). The interpretation of the cosmological constant in terms of vacuum energy was stated by Lemaître in Proc. Nat. Acad. Sci. 20, 12L (1934).

[7] G. Lemaître, op. cit.

[478] A. S. Eddington, The Mathematical Theory of Relativity, 2nd ed. (Cambridge University Press, Cambridge, 1924), Section 70. It is interesting that Eddington interpreted Slipher's observation that most spiral nebulae exhibit redshifts rather than blueshifts in terms of the de Sitter model, rather than Friedmann's models.

[9] H. Bondi and T. Gold, Mon. Not. Roy. Astron. Soc. 108, 252 (1948).

absolute luminosity of these brightest galaxies evolves significantly over cosmological time scales. There are also severe selection effects: there is a tendency to pick out larger clusters with brightest galaxies of higher absolute luminosity at large distances. The evolution of brightest galaxies is of interest in itself, and continues to be the object of astronomical study, [1] but the use of these galaxies as distance indicators has been pretty well abandoned. Similarly, although the Tully-Fisher relation discussed in Section 1.3 has been applied to galaxies with redshifts of order unity, at these redshifts it is used to study galactic evolution, rather than to measure cosmological parameters. [2]

Fortunately, the Type Ia supernovae discussed in Section 1.3 provide an excellent replacement as standard candles. [3] They are very bright; the peak blue absolute magnitude averages about -19.2 , which compares well with the absolute magnitude -20.3 estimated for our own galaxy. Also, as described in Section 1.3, a Type Ia supernova typically occurs when a white dwarf member of a binary pair has accreted just enough mass to push it over the Chandrasekhar limit, so that the nature of the explosion does not depend much on when in the history of the universe this happens, or on the mass with which the white dwarf started or the nature of the companion star. But it might depend somewhat on the metallicity (the proportion of elements heavier than helium) of the white dwarf, which can depend on the epoch of the explosion. The absolute luminosity of Type Ia supernovae is observed to vary with environmental conditions, but fortunately in the use of supernovae as distance indicators the bulk of this variation is correctable empirically.

Observations of Type Ia supernovae have been compared with theoretical predictions (equivalent to Eq. (1.5.45)) for luminosity distance as a function of redshift at about the same time by two groups: The Supernova Cosmology Project [4] and the High- z Supernova Search Team. [5]

The Supernova Cosmology Project analyzed the relation between apparent luminosity and redshift for 42 Type Ia supernovae, with redshifts z ranging from 0.18 to 0.83 , together with a set of closer supernovae from another supernova survey, at redshifts below 0.1 . Their original results are shown in Figure 1.1.

With a confidence level of 99%, the data rule out the case $\Omega_\Lambda = 0$ (or $\Omega_\Lambda < 0$). For a flat cosmology with $\Omega_K = \Omega_R = 0$, so that $\Omega_\Lambda + \Omega_M = 1$, the data indicate a value

$$\Omega_M = 0.28^{+0.09}_{-0.08}(1\sigma \text{ statistical })^{+0.05}_{-0.04} \text{ (identified systematics)}$$

(These results are independent of the Hubble constant or the absolute calibration of the relation between supernova absolute luminosity and time scale, though they do depend on the shape of this relation.) This gives the age (1.5.43) as

[10] F. Hoyle, Mon. Not. Roy. Astron. Soc. 108, 372 (1948), ibid. 109, 365 (1949).

[471] See, e.g., D. Zaritsky et al., in Proceedings of the Sesto 2001 Conference on Tracing Cosmic Evolution with Galaxy Clusters [astro-ph/0108152]; S. Brough et al., in Proceedings of the Sesto 2001 Conference on Tracing Cosmic Evolution with Galaxy Clusters [astro-ph/0108186].

[2] N. P. Vogt et al., Astrophys. J. 465, 115 (1996). For a review and more recent references, see A. Aragón-Salmanca, in Galaxy Evolution Across the Hubble Time - Proceedings of I.A. U. Symposium 235, eds. F. Combes and J. Palous [astro-ph/0610587].

[3] For reviews, see S. Perlmutter and B. P. Schmidt, in Supernovae & Gamma Ray Bursts, ed. K. Weiler (Springer, 2003) [astro-ph/0303428]; P. Ruiz-Lapuente, Astrophys. Space Sci. 290, 43 (2004) [astroph/0304108]; A. V. Filippenko, in Measuring and Modeling of the Universe (Carnegie Observatories Astrophysics Series, Vol 2., Cambridge University Press) [astro-ph/0307139]; Lect. Notes Phys. 645, 191 (2004) [astro-ph/0309739]; N. Panagia, Nuovo Cimento B 210, 667 (2005) [astro-ph/0502247].

[4] S. Perlmutter et al., Astrophys. J. 517, 565 (1999) [astro-ph/9812133]. Also see S. Perlmutter et al., Nature 391, 51 (1998) [astro-ph/9712212].

[5] A. G. Riess et al., Astron. J. 116, 1009 (1998) [astro-ph/9805201]. Also see B. Schmidt et al., Astrophys. J. 507, 46 (1998) [astro-ph/9805200].

$$t_0 = 13.4^{+1.3}_{-1.0} \times 10^9 \left(\frac{70 \text{ km s}^{-1}\text{Mpc}^{-1}}{H_0} \right) \text{yr}.$$

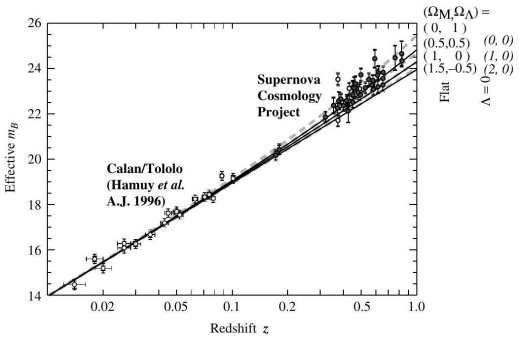


Figure 1.1: Evidence for dark energy, found in 1998 by the Supernova Cosmology Project, from S. Perlmutter et al., Astrophys. J. 517, 565 (1999) [astro-ph/9812133]. Here the effective blue apparent magnitude (corrected for variations in absolute magnitude, as indicated by supernova light curves) are plotted versus redshift for 42 high redshift Type Ia supernovae observed by the Supernova Cosmology Project, along with 18 lower redshift Type Ia supernovae from the Calán-Tololo Supernovae Survey. Horizontal bars indicate the uncertainty in cosmological redshift due to an assumed peculiar velocity uncertainty of \$300 \mathrm{˜km} \mathrm{sec}ˆ{-1}\$. Dashed and solid curves give the theoretical effective apparent luminosities for cosmological models with $\Omega_K = 0$ or $\Omega_\Lambda = 0$, respectively, and various possible values of Ω_M.

For $\Omega_M = 0.28$ and $\Omega_\Lambda = 1 - \Omega_M$, Eq. (1.5.48) gives a negative deceleration parameter, $q_0 = -0.58$, indicating that the expansion of the universe is accelerating.

The High-z Supernova Search Team originally studied 16 Type Ia supernovae of high redshift (with redshifts ranging from 0.16 to 0.97), including 2 from the Supernova Cosmology Project, together with 34 nearby supernovae, and conclude that $\Omega_\Lambda > 0$ at the 99.7% confidence level, with no assumptions about spatial curvature. Their original results are shown in Figure 1.2.

Their best fit for a flat cosmology is $\Omega_M = 0.28 \pm 0.10$ and $\Omega_\Lambda = 1 - \Omega_M$, giving an age of about $(14.2 \pm 1.5) \times 10^9$ years, including uncertainties in the Cepheid distance scale. Assuming only $\Omega_M \geq 0$, and with a conservative

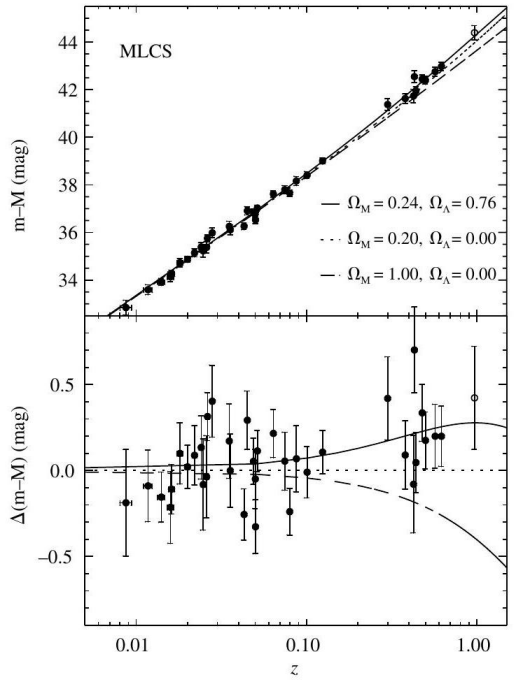


Figure 1.2: Evidence for dark energy, found in 1998 by the High-z Supernova Search Team, from A. G. Riess et al., Astron. J. 116, 1009 (1998) [astro-ph/9805201]. In the upper panel distance modulus is plotted against redshift for a sample of Type Ia supernovae. The curves give the theoretical results for two cosmologies with $\Omega_{\Lambda}=0$ and a good-fit flat cosmology with $\Omega_M = 0.24$ and $\Omega_\Lambda = 0.76$. The bottom panel shows the difference between data and a formerly popular Einstein-de Sitter model with $\Omega_M = 0.2$ and $\Omega_\Lambda = 0$, represented by the horizontal dotted line.

fitting method, with 99.5% confidence they conclude that $q_0 < 0$, again strongly indicating an accelerated expansion. Including 8 new supernovae in a sample of 230 supernovae of Type Ia gave [6] $1.4\Omega_M - \Omega_\Lambda = -0.35 \pm 0.14$, providing further evidence that $\Omega_\Lambda > 0$. The case for vacuum energy was then strenghtened when the Supernova Cosmology Project, [7] including a new set of supernova, found for a flat universe that $\Omega_\Lambda = 0.75^{+0.06}_{-0.07}$ (stat.) $\pm$ 0.032(syst.).

Both groups agree that their results are chiefly sensitive to a linear combination of Ω_Λ and Ω_M, given as $0.8\Omega_M - 0.6\Omega_\Lambda$ by the Supernova Cosmology Project and $\Omega_M - \Omega_\Lambda$ or $1.4\Omega_M - \Omega_\Lambda$ by the High- z Supernova Search Team. The minus sign in these linear combinations, as in Eq. (1.5.48), reflects the fact that matter and vacuum energy have opposite effects on the cosmological acceleration: Matter causes it to slow down, while a positive vacuum energy causes it to accelerate. The negative values found for these linear combinations shows the presence of a component of energy something like vacuum energy, with $p \simeq -\rho$. This is often called dark energy.

Incidentally, these linear combinations of Ω_Λ and Ω_M are quite different from the expression $\Omega_M/2 - \Omega_\Lambda$, which according to Eq. (1.5.48) gives the deceleration parameter q_0 that was the target of much cosmological work of the past. Thus the observations of Type Ia supernovae at cosmological distances should not be regarded as simply measurements of q_0.

The High-z Supernova Search Team subsequently began to use the same survey observations to follow the time development of supernovae that were used to find them. [8] They discovered 23 new high redshift supernovae of Type Ia, including 15 with $z > 0.7$. Using these new supernovae along with the 230 used earlier by Tonry et al., and with the assumption that $\Omega_M + \Omega_\Lambda = 1$, they found the best-fit values $\Omega_M = 0.33$ and $\Omega_\Lambda = 0.67$.

The crucial feature of the supernova data that indicates that $\Omega_\Lambda > \Omega_M$ is that the apparent

luminosity of Type Ia supernovae falls off more rapidly with redshift than would be expected in an Einstein-de Sitter cosmology with $\Omega_M = 1$ and $\Omega_\Lambda = 0$. We can see the effect of vacuum energy on apparent luminosity by comparing the luminosity distance calculated in two extreme cases, both with no matter or radiation. For a vacuumdominated flat model with $\Omega_\Lambda = 1$ and $\Omega_K = \Omega_M = \Omega_R = 0$, Eq. (1.5.46) gives

$$d_L(z) = \frac{z + z^2}{H_0} \quad \text{(vacuum dominated)} , \tag{1.6.1}$$

while for an empty model with $\Omega_K = 1$ and $\Omega_\Lambda = \Omega_M = \Omega_R = 0$, Eq, (1.5.46) gives

$$d_L(z) = \frac{z + z^2/2}{H_0} \quad \text{(empty)}, \tag{1.6.2}$$

Evidently for all z, vacuum energy increases the luminosity distance. The same increase is seen if we compare the more realistic case $\Omega_\Lambda = 0.7, \Omega_M = 0.3$, $\Omega_K = \Omega_R = 0$ with the corresponding case without vacuum energy and $\Omega_K = 0.7, \Omega_M = 0.3, \Omega_\Lambda = \Omega_R = 0$, as can be seen in Figure 1.3.

Both the Supernova Cosmology Project and the High-z Supernova Search Team found that curve of measured luminosity distances vs. redshift of Type Ia supernovae was closer to the upper than the lower curve in Figure 1.3. Indeed, according to Eq. (1.4.9), the negative value of q_0 found by all groups corresponds to the fact that the apparent luminosity of the type Ia supernovae seen at moderate redshifts is less than in the empty model, for which $q_0 = 0$, in contrast with what had been expected, that the expansion is dominated by matter, in which case we would have had $q_0 > 0$, and the apparent luminosities at moderate redshifts would have been larger than for $q_0 = 0$.

The connection between an accelerating expansion and a reduced apparent luminosity can be understood on the basis of the naive Newtonian cosmological model discussed in Section 1.5. In this model, the redshift we observe from a distant galaxy depends on the speed the galaxy had when the light we observe was emitted, but the apparent luminosity is inversely proportional to the square of the distance of this galaxy now, because the galaxy's light is now spread over an area equal to 4π times this squared distance. If the galaxies we observe have been traveling at constant speed since the beginning, as in the empty model, then the distance of any galaxy from

[476] J. L. Tonry et al., Astrophys. J. 594, 1 (2003) [astro-ph/0305008].
[7] R. Knop et al., Astrophys. J. 598, 102 (2003) [astro-ph/0309368].
[8] B. J. Barris et al., Astrophys. J. 502, 571 (2004) [astro-ph/0310843].

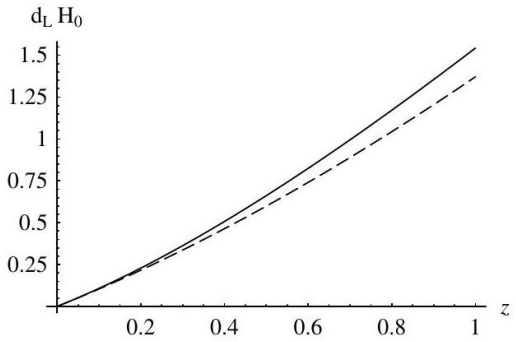


Figure 1.3: Luminosity distance versus redshift for two cosmological models. The upper solid curve is for the case \$\Omega_{\Lambda}=0.7, \Omega_{M}=0.3, \Omega_{K}=\Omega_{R}=0\$; the lower dashed curve is for an empty model, with $\Omega_K = 1, \Omega_\Lambda = \Omega_M = \Omega_R = 0$. The vertical axis gives the luminosity distance times the Hubble constant.

us now would be proportional to its speed when the light was emitted. In the absence of a vacuum energy, we would expect the galaxies to be slowing down under the influence of their mutual gravitational attraction, so that the speed we observe would be greater than the speed they have had since the light was emitted, and their distances now would therefore be smaller than they would be if the speeds were constant. Thus in the absence of vacuum energy we would expect an enhanced apparent luminosity of the supernovae in these galaxies. In fact, it seems that the luminosity distances of supernovae are larger than they would be if the speeds of their host galaxies were constant, indicating that these galaxies have not been slowing down, but speeding up. This is just the effect that would be expected from a positive vacuum energy.

Of course, it is also possible that the reduction in apparent luminosity is due to absorption or scattering of light by intervening material rather than an accelerated expansion. It is possible to distinguish such effects from a true increase in luminosity distance by the change in the apparent color produced by such absorption or scattering, but this is a complicated business. [9] This concern has been allayed by careful color measurements. [10] But it is still possible to invent intergalactic media (so-called gray dust) that would reduce the apparent luminosity while leaving the color unchanged.

This concern has been largely put to rest, first by the study [11] of the supernovae SN1997ff found in the Hubble Deep Field [12] in a galaxy with a redshift $z = 1.7 \pm 0.1$, the greatest yet found for any supernova, and then by the discovery and analysis by a new team, the Higher-z Supernova Team, [13] of 16 new Type Ia supernovae, of which six have $z > 1.25$. These redshifts are so large that during a good part of the time that the light from these supernovae has been on its way to us, the energy density of the universe would have been dominated by matter rather than by a cosmological constant, and so the expansion of the universe would have been decelerating rather than accelerating as at present. Thus if the interpretation of the results of the two groups at smaller redshifts in terms of Ω_M and Ω_Λ is correct, then the apparent luminosity of these supernovae should be larger than would be given by a linear relation between luminosity distance and redshift, a result that could not be produced by absorption or scattering of light. We see this in Figure 1.4, which shows the difference between the luminosity distance (in units H_0^{-1}) for the realistic case with $\Omega_\Lambda = 0.7, \Omega_M = 0.3, \Omega_K = \Omega_R = 0$ and for

[479] See e.g., A. Aguirre, Astrophys. J. 525, 583 (1999) [astro-ph/9904319].

[10] R. Knop et al., ref. 7; also see M. Sullivan et al., Mon. Not. Roy. Astron. Soc. 340, 1057 (2003) [astro-ph/0211444].

[11] A. G. Riess et al., Astrophys. J. 560, 49 (2001) [astro-ph/0104455].

[12] R. L. Gilliland, P. E. Nugent, and M. M. Phillips, Astrophys. J. 521, 30 (1999).

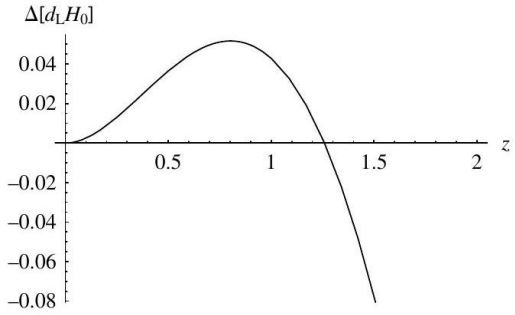


Figure 1.4: The luminosity distance times H_{0} for the realistic case $\Omega_\Lambda = 0.7, \Omega_M = 0.3$, $\Omega_K = \Omega_R = 0$, minus its value for the empty case $\Omega_K = 1, \Omega_\Lambda = \Omega_M = \Omega_R = 0$, plotted against redshift.

the empty model with $\Omega_K = 1, \Omega_M = \Omega_K = \Omega_R = 0$. We see that luminosity distances for the realistic model are greater than for the empty model for moderate redshift, but become less than for the empty model for $z > 1.25$. This is just what is seen. The apparent luminosity of all supernovae is consistent with the parameters $\Omega_M \approx 0.3, \Omega_\Lambda \approx 0.7$ found in the 1998 studies, but not consistent with what would be expected for gray dust and $\Omega_\Lambda = 0$. These conclusions have subsequently been strengthened by the measurement of luminosity distances of additional Type Ia supernovae with redshifts near 0.5. [14] In 2006 Riess et al. [15] announced the discovery with the Hubble Space Telescope of 21 new Type Ia supernovae, which included 13 supernovae with redshifts $z \geq 1$ measured spectroscopically (not just photometrically). Their measured luminosity distances and redshifts, together with data on previously discovered Type Ia supernovae, gave further evidence of a transition from a matter-dominated to a vacuum energy-dominated expansion, and showed that the pressure/density ratio of the vacuum energy for $z > 1$ is consistent with $w = -1$, and not rapidly evolving.

Another serious concern arises from the possibility that the absolute luminosity of Type Ia supernovae may depend on when the supernovae occur. Because Type Ia supernovae occur at a characteristic moment in the history of a star, evolution effects on the luminosities of these supernovae are not expected to be as important as for whole galaxies, which at great distances are seen at an earlier stage in their history. [16] Even so, the absolute luminosity of a Type Ia supernova is affected by the chemical composition

of the two progenitor stars of the supernova, which is in turn affected by the evolution of the host galaxy. [17] Such effects are mitigated by taking account of the correlation of supernova absolute luminosity with decay time and with intrinsic color, both of which presumably depend on the progenitor's chemical composition. Also, evidence for dark energy has been found in studies of subsets of Type Ia supernovae found in very different environments with very different histories. [18] The study of the seven supernovae with $z > 1.25$ mentioned above rules out models with $\Omega_\Lambda = 0$ and any sort of dramatic monotonic evolution of supernovae absolute luminosities that would mimic the effects of dark energy.

There are other effects that might possibly impact the observed relation between supernova apparent luminosities and redshifts:

1. The effect of weak gravitational lensing on the implications of the supernova observations is expected to be small, [19] except perhaps for small area surveys. [20] (Gravitational lensing

[13] A. G. Riess et al., Astrophys. J. 607, 665 (2004) [astro-ph/0402512].
[4714] A. Clocchiatti et al., Astrophys. J. 642, 1 (2006) [astro-ph/0510155].
[15] A. Riess et al., Astrophys. J. 659, 98 (2007) [astro-ph/0611572].
[16] D. Branch, S. Perlmutter, E. Baron, and P. Nugent, contribution to the Supernova Acceleration Probe Yellow Book (Snowmass, 2001) [astro-ph/0109070].

is discussed in Chapter 9.) It had been thought that the apparent luminosity of the most distant observed supernova, SN1997ff, may be enhanced by gravitational lensing, [21] conceivably reopening the possibility that the reduction of the apparent luminosity of the nearer supernovae is due to gray dust. However, a subsequent analysis by the same group [22] reported that the magnification of this supernova due to gravitational lensing is less than had been thought, and that the effect of the corrections due to gravitational lensing on current cosmological studies is small. Members of the High-z Supernova project have reported that instead this effect is likely to improve agreement with the estimate that $\Omega_M = 0.35$ and $\Omega_\Lambda = 0.65$.[23]

2. It has been argued that inhomogeneities in the cosmic distribution of matter could produce an accelerating expansion, without the need for any sort of exotic vacuum energy. [24] Given the high degree of

homogeneity of the universe when averaged over sufficiently large scales, this seems unlikely. [25]

3. There is some evidence for two classes of type Ia supernovae, [26] with the minority associated perhaps with merging white dwarfs, or with a variation in explosion physics. The effect on cosmological studies remains to be evaluated.
4. Other uncertainties that can degrade the accuracy of measurements of dark energy (without casting doubt on its existence) arise from the circumstance that the shape of the curve of luminosity distance versus redshift is found by numerous observatories, both ground-based and space-based, and there are various flux calibration errors that can arise between these different observatories.
5. The measurement of luminosity distance of any source of light at large redshift has historically been plagued by the fact that measurements are not "bolometric," that is, equally sensitive to all wavelengths, but are rather chiefly sensitive to wavelengths in a limited range. The cosmological redshift changes the apparent colors of sources, and thereby changes the sensitivity with which apparent luminosity is measured. To take this into account, the observed apparent magnitude is corrected with a so-called K-correction. [27] The K -correction for supernovae were worked out before the discovery of dark energy, [28] and has been refined subsequently. [29] As the precision of supernovae observations improves, further improvements may also be needed in the K-correction.

These observations of an accelerated expansion are consistent with the existence of a constant vacuum energy, but do not prove that this energy density really is constant. According to Eq. (1.5.18), the existence of an accelerated expansion does however require that a large part of the energy density of the universe is in a form that has $\rho + 3p < 0$, unlike ordinary matter or radiation. This has come to be called dark energy. [30]

[4717] P. Podsiadlowski et al., astro-ph/0608324. Evolution may also affect the extinction of light by dust in the host galaxy; see T. Totani and C. Kobayashi, Astrophys. J. 526, 65 (1999).
[18] M. Sullivan et al., ref. 10.
[19] A. J. Barber, Astron. Soc. Pacific Conf. Ser. 237, 363 (2001) [astro-ph/0109043].
[20] A. Cooray, D. Huterer, and D. E. Holz, Phys. Rev. Lett. 96, 021301 (2006).
[21] E. Mörtstell, C. unnarsson, and A. Goobar, Astrophys. J. 561, 106 (2001); C. Gunnarsson, in Proceedings of a Conference on New Trends in Theoretical and Observational Cosmology - Tokyo, 2001 [astro-ph/0112340].
[22] J. Jönsson et al., Astrophys. J. 639, 991 (2006) [astro-ph/0506765].
[23] N. Benítez et al., Astrophys. J. 577, L1 (2002) [astro-ph/0207097].
[24] E. W. Kolb, S. Matarrese, A. Notari, and A. Riotto, Astrophys. J. 626, 195 (2005) [hep-th/0503117]; E. W. Kolb, S. Matarrese, and A. Riotto, New J. Phys. 8, 322 (2006) [astro-ph/0506534]; E. Barausse, S. Matarrese, and A. Riotto, Phys. Rev. D 71, 063537 (2005).
[4725] É. É. Flanagan, Phys. Rev. D71, 103521 (2005)[hep-th/0503202]; G. G. Geshnizjani, D. J. H. Chung, and N. Afshordi, Phys. Rev. D 72, 023517 (2005) [astro-ph/0503553]; C. M. Hirata and U. Seljak, Phys. Rev.

To take into account the possibility that the dark energy density is not constant, it has become conventional to analyze observations in terms of its pressure/density ratio $p_{\rm D.E.}/\rho_{\rm D.E.} \equiv w$. Except in the case of a constant vacuum energy density, for which $w = -1$, there is no special reason why w should be time-independent. (A different, more physical, possibility is explored at the end of Section 1.12.) Still, it is popular to explore cosmological models with w constant but not necessarily equal to -1 . As long as the dark energy density and Ω_K are non-negative, the expansion of the universe will continue, with $\dot{a}$ always positive. As shown in Eq. (1.1.34), the dark energy density in this case goes as a^{-3-3w}, so if w is negative (as indicated by the supernova observations) the energy density of radiation and matter must eventually become negligible compared with the dark energy density. For $w < -1/3$, the effect of a possible curvature in the Friedmann equation (1.5.19) also eventually becomes negligible. The solution of this equation for $w > -1$ with $\dot{a} > 0$ then becomes $t - t_1 \to Ca^{(3+3w)/2}$, with $C > 0$, and t_1 an integration constant. This is a continued expansion, with a decreasing expansion rate. But for $w < -1$, sometimes known as the case of phantom energy, the solution with $\dot{a} > 0$ is instead $t_1 - t \to Ca^{(3+3w)/2}$, again with $C > 0$. This solution has the remarkable feature that $a(t)$ becomes infinite at the time t_1. In contrast with the case $w \geq -1$, for $w < -1$ all structures - galaxy clusters, galaxy clusters, stars, atoms, atomic nuclei, protons and neutrons - eventually would be ripped apart by the repulsive forces associated with dark energy. [31]

To further study the time dependence of the dark energy, a five year supernova survey, the Supernova Legacy Survey, [32] was begun in 2003 at the Canada-France-Hawaii telescope on Mauna Kea. At the end of the first year, 71 high redshift Type Ia supernovae had been discovered and studied, with the result that $\Omega_M = 0.263 \pm 0.042$ (stat) ± 0.032 (sys). Combining this supernova data with data from the Sloan Digital Sky Survey (discussed in Chapter 8), and assuming that the dark energy has $w \equiv p/\rho$ timeindependent, it is found that if w is constant then $w = -1.023 \pm 0.09$ (stat) $\pm$ 0.054 (sys), consistent with the value $w = -1$ for a constant vacuum energy. At the time of writing, results have just become available for 60 Type Ia supernovae from another supernova survey, ESSENCE. [33] (The acronym is for Equation of State: Supernovae trace Cosmic Expansion). Combining these with the results of the Supernova Legacy Survey, the ESSENCE group found that if w is constant then $w = -1.07 \pm 0.09$ (stat, 1σ) ± 0.13 (syst), and $\Omega_M = 0.267^{0.028}_{-0.018}$ (stat, 1σ).

The conclusion that dark energy makes up a large fraction of the energy of the universe has been confirmed by observations of the cosmic microwave background, as discussed in Section 7.2. This conclusion has also received support from the use of a different sort of secondary distance indicator, the emission of X-rays from hot gas in galaxy clusters. In Section 1.9 we will see that the measurement of redshift, temperature, apparent X-ray luminosity and angular diameter of a cluster allows a determination of the ratio of hot gas ("baryons") to all matter in the cluster, with this ratio proportional to $d_A^{-3/2}$, where d_A is the assumed angular-diameter distance of the cluster. This can be turned around: under the assumption that the ratio of hot gas to all matter is the same for all clusters in a sample, X-ray observations can be used to find the dependence of

D 72, 083501 (2005) [astro-ph/0503582]; A. Ishibashi and R. M. Wald, Class. Quant. Grav. 23, 235 (2006) [gr-qc/0509108].

[26] D. Howell et al., Nature 443, 308 (2006); S. Jha, A. Riess, & R. P. Kirshner, Astrophys. J. 654, 122 (2007); R. Quimby, P. Höflich, and J. C. Wheeler, 0705.4467.

[27] For a discussion of the K-correction applied to observations of whole galaxies, and original references, see G&C, p. 443.

[28] A. Kim, A. Goobar, and S. Perlmutter, Proc. Astron. Soc. Pacific 108, 190 (1995) [astroph/9505024].

[29] P. Nugent, A. Kim, and S. Perlmutter, Proc. Astron. Soc. Pacific 114, 803 (2002) [astro-ph/0205351].

[30] For a general review, see P. J. E. Peebles and B. Ratra, Rev. Mod. Phys. 75, 559 (2003).

[4731] R. R. Caldwell, M. Kamionkowski, and N. N. Weinberg, Phys. Rev. Lett. 91, 071301 (2003) [astro-ph/0302506].

[32] P. Astier, et al., Astron. Astrophys. 447, 31 (2006) [astro-ph/0510447].

[33] M. Wood-Vesey et al., astro-ph/0701041

the cluster angular diameter distances on redshift. [34] In this way, observations by the Chandra satellite of X-rays from 26 galaxy clusters with redshifts in the range $0.07 < z < 0.9$ have been used to determine that in a cosmology with a constant vacuum energy and cold dark matter, $\Omega_\Lambda = 0.94^{+0.21}_{-0.25}$, within 68% confidence limits. [35] Relaxing the assumption that the cosmological dark energy density is constant, but assuming $\Omega_K = 0$ and a constant w, and taking the baryon density to have the value indicated by cosmological nucleosynthesis (discussed in Section 3.2), this analysis of the Chandra data yields $1-\Omega_M = 0.76\pm0.04$ and a dark energy pressure/density ratio $w = -1.20^{+0.24}_{-0.28}$.

It is possible that measurements of luminosity distance can be pushed to much larger redshifts by the use of long gamma ray bursts as secondary distance indicators. These bursts definitely do not have uniform absolute luminosity, but there are indications that the absolute gamma ray luminosity is correlated with the peak gamma ray energy and a characteristic time scale. [36]

The discovery of dark energy is of great importance, both in interpreting other observations and as a challenge to fundamental theory. It is profoundly puzzling why the dark energy density is so small. The contribution of quantum fluctuations in known fields up to 300 GeV , roughly the highest energy at which current theories have been verified, gives a vacuum energy

density of order $(300\text{GeV}),$ [4] or about 10^{27} g/cm^3. This of course is vastly larger than the observed dark energy density, $\Omega_V \rho_{0,\,\text{crit}} \simeq 10^{-29}$ g/cm^3, by a factor of order 10^{56}. There are other unknown contributions to the vacuum energy that might cancel this contribution, coming from fluctuations in fields at higher energies or from the field equations themselves, but this cancelation would have to be precise to about 56 decimal places. There is no known reason for this remarkable cancelation. [37] The discovery of dark energy now adds a second problem: why is the dark energy density comparable to the matter energy density at this particular moment in the history of the universe?

In thinking about these problems, it is crucial to know whether the vacuum energy is really time-independent, or varies with time, a question that may be answered by future studies of distant Type Ia supernovae or other measurements at large redshift. The possibility of a varying dark energy (known as quintessence) will be considered further in Section 1.12.

10.1.9 1.7 Cosmic expansion or tired light?

In comparing observations of redshifts and luminosity distances with theory, we rely on the general understanding of redshifts and luminosities outlined in Sections 1.2 and 1.4. One thing that might invalidate this understanding is absorption or scattering, which reduces the number of photons reaching us from distant sources. This possibility is usually taken into account by measuring the color of the source, which would be affected by absorption or scattering, though as mentioned in the previous section there is a possibility of gray dust, which could not be detected in this way. Another possible way that apparent luminosities could be reduced is through the conversion of photons into particles called axions by intergalactic magnetic fields.

[4734] S. Sasaki, Publ. Astron. Soc. Japan 48, 119 (1996) [astro-ph/9611033]; U.-L. Pen, New Astron. 2, 309 (1997) [astro-ph/9610147].

[35] S. W. Allen, R. W. Schmidt, H. Ebeling, A. C. Fabian, and L. van Speybroeck, Mon. Not. Roy. Astron. Soc. 353, 457 (2004) [astro-ph/0405340]. For earlier applications of this technique, see K. Rines et al., Astrophys. J. 517, 70 (1999); S. Ettori and A. Fabian, Astron. Soc. Pac. Conf. Ser. 200, 369 (2000); S. W. Allen, R. W. Schmidt, and A. C. Fabian, Mon. Not. Roy. Astron. Soc. 334, L11 (2002); S. Ettori, P, Tozzi, and P. Rosati, Astron. & Astrophys. 398, 879 (2003). The possibility of a variable ratio of hot gas to all matter is explored by R. Sadat et al., Astron. & Astrophys. 437, 310 (2005); L. D. Ferramacho and A. Blanchard, Astron. & Astrophys. 463, 423 (2007) [astro-ph/0609822].

[36] C. Firmani, V. Avila-Reese, G. Ghisellini, and G. Ghirlanda, Mon. Not. Roy. Astron. Soc. 372, 28 (2006) [astro-ph/0605430]; G. Ghirlanda, G. Ghisellini, and C. Firmani, New J. Phys. 8, 123 (2006) [astro-ph/0610248].

There is also a more radical possibility. Ever since the discovery of the cosmological redshift, there has been a nagging doubt about its interpretation as evidence of an expanding universe. Is it possible that the universe is really static, and that photons simply suffer a loss of energy and hence a decrease in frequency as they travel to us, the loss of energy and hence the redshift naturally increasing with the distance that the photons have to travel?

It is possible to rule out all these possibilities by comparing the luminosity distance $d_L(z)$ with the angular diameter distance $d_A(z)$ of the same distant source. None of the possibilities mentioned above can affect the angular diameter distance, while the conventional interpretation of

redshifts and luminosities provides the model-independent result (1.4.12), that $d_L(z)/d_A(z) = (1+z)^2$, so a verification of this ratio can confirm the conventional understanding of $d_L(z)$.

We can check this formula for $d_L(z)/d_A(z)$ by a "surface brightness test" suggested long ago by Tolman. [1] If a light source has an absolute luminosity per proper area $\mathcal{L}$, then the apparent luminosity of a patch of area A will be $\ell = \mathcal{L}A/4\pi d_L^2$. This patch will subtend a solid angle $\Omega = A/d_A^2$. The surface brightness B is defined as the apparent luminosity per solid angle, so

$$B \equiv \frac{\ell}{\Omega} = \frac{\mathcal{L} d_A^2}{4\pi d_L^2} \tag{1.7.1}$$

In the conventional big bang cosmology the ratio d_A/d_L is given by Eq. (1.4.12), so

$$B = (1+z)^{-4} \left(\frac{\mathcal{L}}{4\pi}\right) \tag{1.7.2}$$

If we can find a class of light sources with a common value for the absolute luminosity per proper area $\mathcal{L}$, then their surface brightness should be found to decrease with redshift precisely as $(1+z)^{-4}$.

For instance, one important difference between "tired light" theories and the conventional big bang theory is that in the conventional theory all rates at the source are decreased by a factor $(1+z)^{-1}$, while in tired light theories there is no such slowing down. One rate that is slowed down at large redshifts in the conventional theory is the rate at which photons are emitted by the source. This is responsible for one of the two factors of $(1+z)^{-1}$ in formula (1.4.1) for apparent luminosity, the other factor being due to the reduction of energy of individual photons. On the other hand, if the rate of photon emission is not affected by the redshift, then in a static Euclidean universe in which photons lose energy as they travel to us, the apparent luminosity of a distant source L at a distance d will be given by $L/4\pi(1+z)d^2$, with only a single factor $1+z$ in the denominator to take account of the photon energy loss. That is, the luminosity distance will be $(1+z)^{1/2}d$, while the angular diameter distance in a Euclidean universe is just d, so here $d_L/d_A = (1+z)^{1/2}$, and the surface brightness of distant galaxies should decrease as $(1+z)^{-1}$.

Lubin and Sandage [2] have used the Hubble Space Telescope to compare the surface brightness of galaxies in three distant clusters with redshifts

$0.76, 0.90$, and 0.92 with the surface brightness measured in closer galaxies. They detect a decrease of B with increasing z that is consistent with Eq. (1.7.2) with reasonable corrections for the effects of galaxy evolution, and is quite inconsistent with the behavior $B \propto (1+z)^{-1}$ expected in a static universe with "tired light."

[4737] For a survey of efforts to answer this question, see S. Weinberg, Rev. Mod. Phys. 61, 1 (1989).

[471] R. C. Tolman, Proc. Nat. Acad. Sci 16, 5111 (1930); Relativity, Thermodynamics, and Cosmology (Oxford Press, Oxford, 1934): 467.

[2] L. M. Lubin and A. Sandage, Astron. J. 122, 1084 (2001) [astro-ph/0106566]. Their earlier work is described in A. Sandage and L. M. Lubin, Astron. J. 121, 2271 (2001); L. M. Lubin and A. Sandage, ibid, 2289 (2001) and Astron. J. 122, 1071 (2001) [astro-ph/0106563.]

In the standard big bang cosmology all rates observed from a distant source are slowed by a factor $1/(1+z)$, not just the rate at which photons are emitted. This slowing has been confirmed [3] for the rate of decline of light from some of the Type Ia supernovae used by the Supernova Cosmology Project discussed in the previous section. The hypothesis that the absolute luminosity is simply correlated with the intrinsic decline time is found to work much better if the observed decline time is taken to be the intrinsic decline time stretched out by a factor $1+z$. Nothing like this would be expected in a static Euclidean universe with redshifts attributed to tired light.

10.1.10 1.8 Ages

As we have seen, a knowledge of the Hubble constant and of the matter and vacuum density parameters Ω_M and Ω_Λ allows us to estimate the age of the universe. In this section we will discuss independent estimates of the age of the universe, based on calculations of the ages of some of the oldest things it contains.

Since metals (elements heavier than helium) found in the outer parts of stars arise chiefly from earlier generations of stars, the oldest stars are generally those whose spectra show small abundances of metals. These are the so-called Population II stars, found in the halo of our galaxy, and in particular in globular clusters. There are two main ways of estimating ages of old stars:

A. Heavy element abundances

If a nucleus decays with decay rate λ, and has an initial abundance $A_{\rm init}$, then the abundance A after a time T is $A = A_{\rm init}\, \exp(-\lambda T)$. Hence if we knew $A_{\rm init}$ and could measure A, we could determine T from $T = \lambda^{-1} \ln\left(A_{\rm init}\,/A\right)$. Unfortunately neither condition is likely to be satisfied. On the other hand, it is often possible to calculate the ratio of the initial abundances $A_{1\ {\rm init}}$ and $A_{2\ {\rm init}}$ of two nuclei, and to measure their relative present abundance A_1/A_2.

This relative abundance is given by

$$\frac{A_1}{A_2} = \left(\frac{A_{1\ {\rm init}}}{A_{2\ {\rm init}}}\right) \exp\left(\left(\lambda_2 - \lambda_1\right) T\right),$$

so

$$T = \frac{1}{\lambda_2 - \lambda_1}\left[\ln\left(\frac{A_1}{A_2}\right) - \ln\left(\frac{A_{1\ {\rm init}}}{A_{2\ {\rm init}}}\right)\right] \tag{1.8.1}$$

If the initial abundances are similar and the observed abundances are very different, then the estimated value of T will be insensitive to the precise value of the initial abundance ratio.

The initial relative abundances of heavy, radioactive elements are estimated on the well-founded assumption that these elements are made in the so-called r-process, the rapid addition of neutrons to lighter elements such as iron in core-collapse supernova explosions, after which the neutron-rich isotopes formed in this way undergo multiple beta decays, transforming them to the most deeply bound nuclei with the same number of nucleons. This method has been used to put a lower bound on the age of our galaxy from the terrestrial abundance of ^{235}U, which has a decay rate of 0.971×10^{-9}/yr. To avoid uncertainties in the distribution of ^{235}U in earth, its abundance is measured relative to the isotope ^{238}U, which has a slower decay rate of 0.154×10^{-9}/yr, but behaves the same chemically and is presumably distributed in the same way. The initial abundance ratio of ^{235}U to ^{238}U is estimated to be 1.65 ± 0.15; it is larger than

[473] B. Leibundgut et al., Astrophys. J. 466, L21 (1996); G. Goldhaber et al. (Supernova Cosmology Project), Astrophys. J. 558, 359 (2001) [astro-ph/0104382].

one because three additional neutrons must be added to the progenitor of ^{235}U to form the progenitor of ^{238}U. On the other hand, the larger decay rate of ^{235}U makes it (fortunately) less abundant than ^{238}U now. The present abundance ratio of uranium isotopes on earth is 0.00723 , so this uranium has been decaying for a time

$$\frac{\ln(1.65) - \ln(.00723)}{0.971 \times 10^{-9}/\text{yr} - 0.154 \times 10^{-9}/\text{yr}} = 6.6\text{Gyr} \quad \left[1\text{Gyr} = 10^9\text{yr}\right]$$

But the sun is a second (or perhaps third) generation (called "Population I") star, and presumably its uranium was being produced over a long time interval before the formation of the solar system. The uranium abundance ratio has been supplemented with measurements of other abundance ratios on the earth and meteorites, such as ^{232}Th/^{238}U and ^{187}Re/^{187}Os ratios, and analyzed with the length of the period of heavy element formation left as a free parameter. This gives a more stringent (but less certain) lower bound of 9.6Gyr[1] on the age of the heavy elements in the neighborhood of the solar system.

A much more stringent lower bound on the age of the galaxy is given by applying these methods to heavy elements in metal-poor stars beyond the solar system. First thorium was observed spectroscopically in a very metal-poor star (and hence presumably old) K giant star, CS 22892-052. [2] The relative abundances in this star of the more stable elements produced in the r-process, as measured from the intensity of absorption lines in the star's spectrum, matches those of the same elements in the solar system, except for a much lower abundance of the heaviest detected element thorium, which (for ^{232}Th) has a half life 14 Gyr . Attributing the decrease in thorium to its radioactive decay, the age of the thorium in this star is estimated as 14.1 ± 3 Gyr. Other estimates of the ages of CS 22892-052 and other metalpoor stars have been made using the measured abundance ratios of thorium to europium and lanthanum. [3]

Uranium-238 decays more rapidly than ^{232}Th, so we can get a more sensitive estimate of the age of a star by using both its uranium and its thorium abundances, providing of course that uranium as well as thorium lines can be observed in the star's spectrum. No uranium absorption lines were observed in the spectrum of CS 22892-052, but absorption lines from singly ionized uranium were subsequently observed in two other metal-poor star with an abundance of r-process elements, CS31082-001 and BD + 17°3248. The observed abundance ratio of uranium to thorium in CS31082-001 is $10^{-0.74\pm0.15}$, while the initial abundance ratio has been variously estimated as $10^{-0.255}$ or $10^{-0.10}$. Using these numbers in Eq. (1.8.1) gives this star an age of 12.5 ± 3 Gyr. [4] Subsequent observations indicated ages of 14 ± 2Gyr, [5] 15.5 ± 3.2Gyr, [6] and 14.1 ± 2.5Gyr.[7] In a similar way, the age of BD + 17°3248 has been calculated as 13.8 ± 4Gyr.[8] (See Figure 1.5.) More recently, both uranium and thorium lines have been found in the spectrum of the newly discovered metal-poor star HE 1523-0903; the ratio of thorium and uranium abundance to the abundances of other r-process elements, and to each other, was used to give an age of the star as 13.2 Gyr. [9]

[471] B. S. Meyer and D. N. Schramm, Astrophys. J. 311, 406 (1986).

[472] C. Sneden et al., Astrophys. J. 467, 819 (1996); Astrophys. J. 591, 936 (2003) [astro-ph/0303542]. A review with references to earlier work on thorium abundances was given by C. Sneden and J. J. Cowan, Astronomia y Astrofisica (Serie de Conferencia) 10, 221 (2001) [astro-ph/0008185].

[3] I. I Ivans et al., Astrophys. J. 645, 613 (2006) [astro-ph/0604180], and earlier references cited therein.

[4] R. Cayrel et al., Nature 409, 691 (2001).

[5] V. Hill et al., Astron. Astrophys. 387, 580 (2002).

[6] H. Schatz et al., Astrophys. J. 579, 626 (2002).

[7] S. Wanajo, Astrophys. J. 577, 853 (2002).

[8] J. J. Cowan, et al., Astrophys. J. 572, 861 (2002) [astro-ph/0202429].

[9] A. Frebel et al., Astrophys. 660, L117 (2007). [astro-ph/0703414].

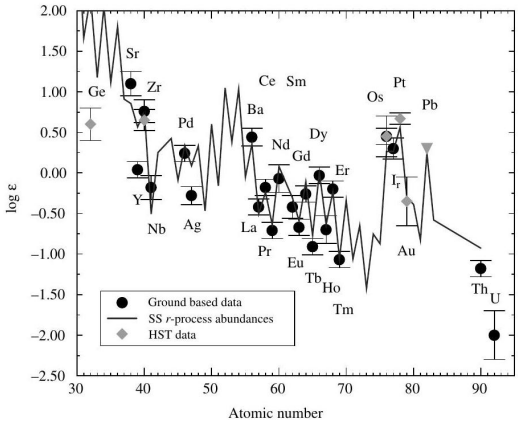


Figure 1.5: Abundances of elements produced by the r-process in the star $\mathrm{BD}+17^{\circ} 3248$, obtained by ground-based and Hubble Space Telescope spectroscopic observations. For comparison, the solid curve gives theoretical initial abundances, based on solar system data. Note the low observed abundances of thorium and uranium, compared with the theoretical initial abundances, which indicate an age for BD + 17°3248 of 13.8 ± 4 Gyr. From J. J. Cowan et al., Astrophys. J. 572, 861 (2002) [astro-ph/0202429].

B. Main sequence turn-off

The stars that satisfy the main sequence relation between absolute luminosity and surface temperature are still burning hydrogen at their core. When the hydrogen is exhausted at the core, hydrogen-burning continues in a shell around a (temporarily) inert helium core. The star then moves off the main sequence, toward higher luminosity and lower surface temperature. The heavier a star is, the more luminous it is on the main sequence, and the faster it evolves. Thus as time passes, the main sequence of a cluster of stars of different masses but the same age shows a turn-off that moves to lower and lower luminosities. (See Figure 1.6). Roughly, the absolute luminosity of stars at the turn-off point is inversely proportional to the age of the cluster. In particular, observations of the main sequences of a number of globular clusters gave ages variously calculated [10] as 11.5 ± 1.3Gyr, 12 ± 1Gyr, 11.8 ± 1.2Gyr, 14.0 ± 1.2Gyr, 12 ± 1Gyr, and 12.2 ± 1.8Gyr. A summary by Schramm [11] found that most of the discrepancies disappear when

[4710] For references, see B. Chaboyer, Phys. Rep. 307, 23 (1998) [astro-ph/9808200].

[11] D. Schramm, in Critical Dialogues in Cosmology, N. Turok, ed. (World Scientific, Singapore, 1997): 81

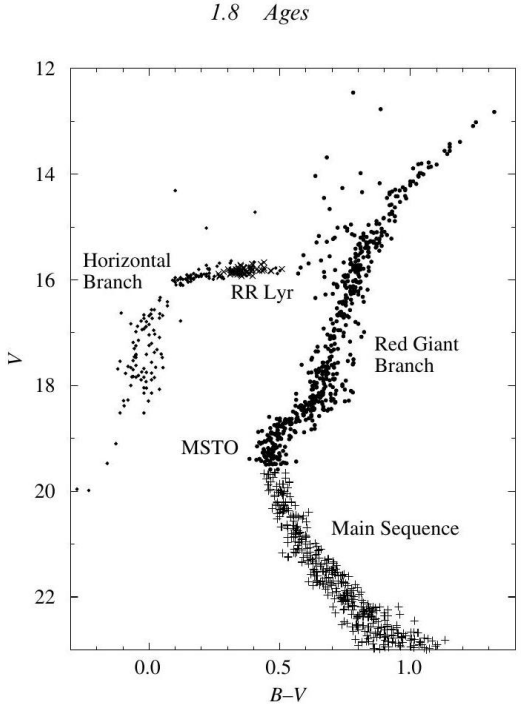


Figure 1.6: Color-magnitude diagram for the globular cluster M15. Visual apparent magnitudes of M15 stars are plotted on the vertical axis. Since all stars in M15 are at about the same distance from earth, the apparent visual magnitude differs from the absolute visual magnitude by a constant term, with absolute luminosities increasing upwards. The difference of apparent blue and visual magnitudes is plotted on the horizontal axis. This is a measure of surface temperature, with temperature decreasing to the right. If M15 were young, the main sequence would continue upwards and to the left; the position of the main sequence turn-off (MSTO) and other features of the diagram indicate that the age of the cluster is 15 ± 3 Gyr. Diagram from B. Chaboyer, Phys. Rep. 307, 23 (1998), based on data of P. R. Durrell and W. E. Harris, Astron. J. 105, 1420 (1993) [astro-ph/9808200].

the various calculations are done with the same input values for parameters like the initial abundance of helium, oxygen, and iron, and gave a consensus value as 14 ± 2 (statistical) ± 2 (systematic) Gyr. Note that all these ages are sensitive to the distance scale; a fractional change $\delta d/d$ in estimates of distances would produce a fractional change $\delta L/L = -2\delta d/d$ in estimates of absolute luminosities, and hence a fractional change $\delta t/t \approx +2\delta d/d$ in estimates of ages. Using measurements of distances to nine globular clusters with the Hipparcos satellite yields an estimated galactic age [12] of 13.2 ± 2.0 Gyr.

These ages would pose a problem for what used to be the most popular model, with $\Omega_M = 1$ and no vacuum energy. In this case, the age of the universe is

$$t_0 = \frac{2}{3H_0} = 9.3 \left(\frac{70 \text{ km/sec/Mpc}}{H_0} \right) \text{Gyr}$$

which is somewhat younger than the oldest objects in the galaxy, though not by many standard deviations. Inclusion of a constant vacuum energy helps to avoid this problem; as remarked in Section 1.5, with nothing else in the universe we would have $a(t) \propto \exp(Ht)$, and the age of the universe would be infinite. As we saw in Section 1.6, the supernovae distance-redshift relation indicates that the vacuum energy is now roughly twice the matter energy,

[4712] E. Carretta, R. G. Gratton, G. Clementini, and F. F. Pecci, Astrophys. J. 533, 215 (2000) [astro-ph/9902086].

giving an age much longer than $2/3H_0$:

$$t_0 = 13.4^{+1.3}_{-1.0}\left(\frac{70\ \text{km/sec/Mpc}}{H_0}\right)\text{Gyr}$$

This removes the danger of a conflict, provided that the globular clusters in our galaxy are not much younger than the universe itself. In fact, there is now a truly impressive agreement between the age of the oldest stars and star clusters on one hand and the cosmic age calculated using values of H_0, Ω_M, and Ω_Λ found from the redshift-distance relation. As we will see in Section 7.2, there is also an excellent agreement between these ages and the age calculated using parameters measured in observations of anisotropies in the cosmic microwave background.

So far in this section, we have considered only the present age of our own galaxy. It is also possible to estimate the ages of other galaxies at high redshift, at the time far in the past when the light we now observe left these galaxies. Of course, it is not possible to distinguish individual stars or globular clusters in these galaxies, but the spectrum of the galaxy gives a good idea of the age. We need the whole spectrum to separate the effects of metallicity, scattering, etc., but roughly speaking, the redder the galaxy, the more of its bright bluer stars have left the main sequence, and hence the older it is. In this way, it has been found [13] that the radio galaxies 53 W091($z = 1.55$) and 53 W069($z = 1.43$) have ages $\simeq 3.5$Gyr and 3 to 4 Gyr , respectively. This sets useful lower bounds on the vacuum energy. In a model with non-relativistic matter and a constant vacuum energy, the age of the universe at the time of emission of light that is seen at present with

redshift z is given by Eq. (1.5.42) as

$$t(z) = \frac{1}{H_0}\int_0^{\frac{1}{1+z}} \frac{dx}{x\sqrt{\Omega_\Lambda + \Omega_K x^{-2} + \Omega_M x^{-3} + \Omega_R x^{-4}}} \tag{1.8.2}$$

Any galaxy observed with redshift z must have been younger than this at the time that its light was emitted. For instance, for a flat universe with $\Omega_K = \Omega_R = 0$, so that $\Omega_M = 1 - \Omega_\Lambda$, the existence of a galaxy at $z = 1.55$ with age $\simeq 3.5$Gyr sets a lower bound [14] on Ω_Λ of about 0.6 for $H_0 = 70\text{kms}^{-1}\text{Mpc}^{-1}$.

Eventually the accuracy of these age determinations may become good enough to allow us to measure at least the dependence of redshift on the cosmic age. Of course, galaxies form at various times in the history of the universe, so the age of any one galaxy does not allow us to infer the age of the universe at the time light we now see left that galaxy. However, the homogeneity of the universe implies that the distribution of cosmic times of formation for any one variety of galaxy is the same anywhere in the universe. From differences in the distributions of ages of a suitable species of galaxy at different redshifts, we can then infer the difference of cosmic age t at these redshifts. The Robertson-Walker scale factor $a(t)$ is related to the redshift $z(t)$ observed now of objects that emitted light when the cosmic age was t by $1 + z(t) = a\left(t_0\right)/a(t)$, so $\dot{z} = -H(t)(1+z)$. To calculate $\ddot{z}$, we note that for $K = 0, H^2(t) = 8\pi G\rho(t)/3$, and $\dot{\rho} = -3H(\rho + p)$,

$$\dot{H}(t) = -4\pi G(\rho(t) + p(t)) \tag{1.8.3}$$

Then for $K = 0$

$$\ddot{z} = \frac{\dot{z}^2}{1+z}\left(\frac{5}{2} + \frac{3p}{2\rho}\right) \tag{1.8.4}$$

Thus measurements of differences in t for various differences in redshift may allow a measurement of the ratio p/ρ at various times in the recent history of the universe. [15]

[4713] J. S. Dunlop et al., Nature 381, 581 (1996); J. S. Dunlop, in The Most Distant Radio Galaxies KNAW Colloquium, Amsterdam, October 1997, eds. Best et al. [astro-ph/9801114].

10.1.11 1.9 Masses

We saw in Section 1.6 that the observed dependence of luminosity distance on redshift suggests that the fraction Ω_M of the critical density provided by

non-relativistic matter is roughly 30%. In this section we will consider other independent ways that Ω_M is measured.

A. Virialized clusters of galaxies

The classic approach [1] to the measurement of Ω_M is to use the virial theorem to estimate the masses of various clusters of galaxies, calculate a mean ratio of mass to absolute luminosity, and then use observations of the total luminosity of the sky to estimate the total mass density, under the assumption that the mass-to-light ratio of clusters of galaxies is typical of the universe as a whole.

To derive the virial theorem, consider a non-relativistic gravitationally bound system of point masses m_n (either galaxies, or stars, or single particles) with positions relative to the center of mass (in an ordinary Cartesian coordinate system) $\mathbf{X}_n$. The equations of motion are

$$m_n \ddot{X}_n^i = -\frac{\partial V}{\partial X_n^i}, \tag{1.9.1}$$

where the potential energy V is

$$V = -\frac{1}{2}\sum_{n\neq\ell} \frac{G m_n m_\ell}{|\mathbf{X}_n - \mathbf{X}_\ell|} \tag{1.9.2}$$

Multiplying Eq. (1.9.1) with X_n^i and summing over n and i gives

$$-\sum_n X_n^i \frac{\partial V}{\partial X_n^i} = \sum_n m_n \mathbf{X}_n \cdot \ddot{\mathbf{X}}_n = \frac{1}{2}\frac{d^2}{dt^2}\sum_n m_n \mathbf{X}_n^2 - 2T, \tag{1.9.3}$$

where T is the internal kinetic energy (not counting any motion of the center of mass)

$$T = \frac{1}{2}\sum_n m_n \dot{\mathbf{X}}_n^2 \tag{1.9.4}$$

Let us assume that the system has reached a state of equilibrium ("become virialized"), so that although the individual masses are moving there is no further statistical evolution, and in particular that

$$0 = \frac{d^2}{dt^2}\sum_n m_n \mathbf{X}_n^2 \tag{1.9.5}$$

(This is why it was important to specify that $\mathbf{X}_n$ is measured relative to the center of mass; otherwise a motion of the whole cluster would give the sum a term proportional to t^2, invalidating Eq. (1.9.5).) But V is of order -1 in the coordinates, so the left-hand side of Eq. (1.9.3) is just V, giving the virial theorem:

$$2T + V = 0 \tag{1.9.6}$$

[4714] L. M. Krauss, Astrophys. J. 489, 486 (1997); J. S. Alcaniz and J. A. S. Lima, Astrophys. J. 521, L87 (1999) [astro-ph/9902298].

[15] R. Jiminez and A. Loeb, Astrophys. J. 573, 37 (2002) [astro-ph/0106145].

[471] F. Zwicky, Astrophys. J. 86, 217 (1937); J. H. Oort, in La Structure et l'Evolution de l'Universe (Institut International de Physique Solvay, R. Stoops, Brussels, 1958): 163.

We may express T and V as

$$T = \frac{1}{2}M\left\langle v^2\right\rangle, \quad V = -\frac{1}{2}GM^2\left\langle\frac{1}{r}\right\rangle \tag{1.9.7}$$

where $\langle v^2\rangle$ is the mean (mass weighted) square velocity relative to the center of mass, $\langle 1/r\rangle$ is the mean inverse separation, and $M = \sum_n m_n$ is the total mass. Eq. (1.9.6) thus gives the virial formula for M :

$$M = \frac{2\left\langle v^2\right\rangle}{G\langle 1/r\rangle} \tag{1.9.8}$$

This derivation does not apply to irregular clusters of galaxies, like the nearby one in Virgo. Clusters like this do not seem to have settled into a configuration in which the condition (1.9.5) is satisfied, and therefore probably do not satisfy the virial theorem. On the other hand, the virial theorem probably does apply at least approximately to other clusters of galaxies, like the one in Coma, which appear more or less spherical. According to general ideas of statistical equilibrium, we may expect the rms velocity dispersion $\sqrt{\langle v^2\rangle}$ of the dominant masses in such clusters to equal the velocity dispersion of the visible galaxies in the cluster, which can be measured from the spread of their Doppler shifts, and also to equal the velocity dispersion of the ionized intergalactic gas in the cluster, which since the advent of X-ray astronomy can be measured from the X-ray spectrum of the gas. The values obtained in these ways for $\langle v^2\rangle$ are independent of the distance scale. On the other hand, values for $\langle 1/r\rangle$ are obtained from angular separations: the true transverse proper distance d is given in terms of the angular separation θ by $d = \theta d_A$, where d_A is the angular diameter distance (1.4.11). For clusters with $z \ll 1$, Eqs. (1.4.9) and (1.4.11) give $d_A \simeq z/H_0$, so $d \simeq \theta z/H_0$. Thus the estimated values of $\langle 1/r\rangle$ for galaxy clusters with $z \ll 1$ scale as H_0, and the values of M inferred from Eq. (1.9.8) scale as $1/H_0$. The absolute luminosity L of a cluster of galaxies with redshift z and apparent luminosity ℓ is given for $z \ll 1$ by Eqs. (1.4.2) and (1.4.9) as $L = 4\pi z^2\ell/H_0^2$, so the values of L scale as H_0^{-2}, and the mass-to-light ratios obtained in this way therefore scale as $H_0^{-1}/H_0^{-2} = H_0$.

Estimates of M/L for rich clusters have generally given results of order 200 to $300hM_\odot/L_\odot$, where h is the Hubble constant in units of 100 km s^{-1}Mpc^{-1}, and $M_\odot$ and $L_\odot$ are the mass and absolute luminosity of the sun. For instance, a 1996 study [2] of 16 clusters of galaxies with redshifts between 0.17 and 0.55 gave $M/L = (295 \pm 53)hM_\odot/L_\odot$. Some of the same group [3] have corrected this result for various biases, and now find $M/L = (213 \pm 59)hM_\odot/L_\odot$. A more recent application [4] of the virial theorem to 459 clusters has found a value $M/L \simeq 348hM_\odot/L_\odot$.

All these values of M/L for clusters of galaxies are very much larger than the mass-to-light ratios of the visible regions of individual galaxies. [5] The mass-to-light ratios of individual elliptical galaxies can be measured using the virial theorem, with $\sqrt{\langle v^2\rangle}$ taken as the velocity dispersion of stars contained in the galaxy; this gives mass-to-light ratios generally in the range of 10 to $20hM_\odot/L_\odot$.[6] All of the visible light from clusters comes from their galaxies, so we must conclude that most of the mass in clusters of galaxies is in some non-luminous form, either in the outer non-luminous parts of galaxies or in intergalactic space. It has been argued that this mass is in large dark halos surrounding galaxies, extending to 200 kpc for bright galaxies. [7] The nature of this dark matter is an outstanding problem of cosmology, to which we will frequently return.

Incidentally, the large value of M/L given by the virial theorem for elliptical galaxies shows that most of the mass of these galaxies is not in the form of stars as bright as the sun. It is harder to estimate M/L for spiral galaxies, but since the work of Vera Rubin [8] it has been known that most of their mass is also not in luminous stars. [9] If most of the mass of a spiral galaxy were in the luminous central regions of the galaxy, then the rotational speeds of stars

outside this region would follow the Kepler law, $v \propto r^{-1/2}$. Instead, it is observed that v outside the central region is roughly constant, even beyond the visible disk of the galaxy, which is what would be expected for a spherical halo with a mass density that decreases only as $1/r^2$, in which case most of the mass of the galaxy would be in the dark outer

parts of the halo. There is some evidence from the absence of gravitational microlensing by the halo (discussed in Section 9.2) that this mass is not in the form of dark stars either, but it is still possible that most of the matter in galaxies is baryonic. We will not go into this in detail, because the formation of galaxies involves cooling processes that requires baryonic matter of the same sort as in stars, so that we would only expect the value of M/L for galaxies to be similar to the value for the universe as a whole if the matter of the universe were mostly baryonic.

In using the value of M/L derived from the virial theorem for clusters of galaxies to find the mass density of the universe, we cannot just add up the luminosity per volume of clusters, because most of the light of the universe comes from "field" galaxies that are not in clusters. Instead, if we assume that the field galaxies are accompanied by the same amount of dark matter as the galaxies in clusters, as argued in ref. 7 , then we can find Ω_M by using the value of M/L for clusters together with an estimate of the total luminosity density $\mathcal{L}$ to estimate the total mass density as

$$\rho_M = (M/L)\mathcal{L} \tag{1.9.9}$$

Since values of absolute luminosities inferred from apparent luminosities and redshifts scale as H_0^{-2}, and distances inferred from redshifts scale as H_0^{-1}, the total luminosity density of the universe calculated by adding up the absolute luminosities of galaxies per volume scales as $H_0^{-2}/\left(H_0^{-1}\right)^3 = H_0$. For example, a 1999 estimate [10] gave $\mathcal{L} = 2 \pm 0.2 \times 10^8 h L_\odot \mathrm{Mpc}^{-3}$. For the purpose of calculating Ω_M it is more convenient to write this as a ratio of the critical mass density to the luminosity density:

$$\rho_{0,\,\mathrm{crit}}\,/\mathcal{L} = (1390 \pm 140) h M_\odot/L_\odot$$

(Here we use $M_\odot = 1.989 \times 10^{33}$ g, $1\mathrm{Mpc} = 3.0857 \times 10^{24}$ cm, and $\rho_{0,\,\mathrm{crit}} = 1.878 \times 10^{-29} h^2$ g/cm^3.) Taking $M/L = (213 \pm 53) h M_\odot/L_\odot$ gives then

$$\Omega_M = \frac{M/L}{\rho_{0,\,\mathrm{crit}}\,/\mathcal{L}} = 0.15 \pm 0.02 \pm .04$$

with the first uncertainty arising from $\mathcal{L}$ and the second from M/L. It is important to note that this is independent of the Hubble constant, as both $\mathcal{L}M/L$ and $\rho_{0,\,\mathrm{crit}}$ scale as H_0^2.

This estimate of Ω_M is somewhat lower than those derived from the redshift-luminosity relation of supernovae and from the anisotropies in the

[472] R. G. Carlberg et al., Astrophys. J. 462, 32 (1996).

[3] R. G. Carlberg, H. K. C. Yee, and E. Ellingson, Astrophys. J. 478, 462 (1997).

[4] H. Andernach, M. Plionis, O. López-Cruz, E. Tago, and S. Basilakos, Astron. Soc. Pacific Conf. Ser. 329, 289 (2005) [astro-ph/0407098].

[5] This conclusion was first reached in a study of the Coma cluster by F. Zwicky, Helv. Phys. Acta 6, 110 (1933).

[6] T. R. Lauer, Astrophys. J. 292, 104 (1985); J. Binney and S. Tremaine, Galactic Dynamics (Princeton University Press, Princeton, 1987).

[7] N. A. Bahcall, L. M. Lubin, and V. Dorman, Astrophys. J. 447, L81 (1995).

[8] V. C. Rubin, W. K. Ford, and N. Thonnard, Astrophys. J. 225, L107; 238, 471 (1980).

[9] M. Persic and P. Salucci, Astrophys. J. Supp. 99, 501 (1995); M. Persic, P. Salucci, and F. Stel, Mon. Not. Roy. Astron. Soc. 281, 27P (1996).

[4710] S. Folkes et al., Mon. Not. Roy. Astron. Soc. 308, 459 (1999); M. L. Blanton et al., Astron. J. 121, 2358 (2001).

cosmic microwave background, to be discussed in Section 2.6 and Chapter 7. But all these estimates agree that Ω_M is distinctly less than unity.

B. X-ray luminosity of clusters of galaxies

Does the dark matter in clusters of galaxies consist of ordinary nuclei and electrons? We can find the ratio of the fraction Ω_B of the critical density provided by baryonic matter (nuclei and electrons) to the fraction Ω_M provided by all forms of non-relativistic matter by studying the X-rays from clusters of galaxies, for it is only the collisions of ordinary baryonic particles that produces these X-rays. Because these collision processes involve pairs of particles of baryonic matter, the absolute X-ray luminosity per unit proper volume takes the form

$$\mathcal{L}_X = \Lambda\left(T_B\right)\rho_B^2 \tag{1.9.10}$$

where T_B and ρ_B are the temperature and density of the baryonic matter, and $\Lambda(T)$ is a known function of temperature and fundamental constants. The baryonic density satisfies the equation of hydrostatic equilibrium, which (assuming spherical symmetry) follows from the balance of pressure and gravitational forces acting on the baryons in a small area A and between radii r and $r+\delta r$:

$$A\left(p_B(r+\delta r)-p_B(r)\right) = -\frac{A\delta r\rho_B(r)G}{r^2}\int_0^r 4\pi r^2\rho_M(r)dr$$

or, canceling factors of A and δr and using the ideal gas law $p_B = k_{\mathcal{B}}T_B\rho_B/m_B$,

$$\frac{d}{dr}\left(\frac{k_{\mathcal{B}}T_B(r)\rho_B(r)}{m_B}\right) = -\frac{G\rho_B(r)}{r^2}\int_0^r 4\pi r^2\rho_M(r)dr$$

where $\rho_M(r)$ is the total mass density, $k_{\mathcal{B}}$ is Boltzmann's constant, m_B is a characteristic mass of the baryonic gas particles, and r is here the proper distance to the center of the cluster. Multiplying by $r^2/\rho_B(r)$ and differentiating with respect to r yields

$$\frac{d}{dr}\left[\frac{r^2}{\rho_B(r)}\frac{d}{dr}\left(\frac{k_{\mathcal{B}}T_B(r)\rho_B(r)}{m_B}\right)\right] = -4\pi G r^2\rho_M(r) \tag{1.9.11}$$

If we make the assumption that cold dark matter particles, or whatever particles dominate the dark intergalactic matter, have an isotropic velocity
distribution (which is not very well motivated), then the same derivation applies to these particles, and their density $\rho_D = \rho_M - \rho_B$ satisfies the non-linear differential equation

$$\frac{d}{dr}\left[\frac{r^2}{\rho_D(r)}\frac{d}{dr}\left(\frac{k_{\mathcal{B}}T_D(r)\rho_D(r)}{m_D}\right)\right] = -4\pi G r^2\rho_M(r) \tag{1.9.12}$$

where $T_D(r)$ and m_D are the temperature and mass of the dark matter particles. With perfect X-ray data and a knowledge of the distance of the source, one could measure the X-ray luminosity density $\mathcal{L}_X(r)$ and (using the Xray spectrum) the baryon temperature $T_B(r)$ at each point in the cluster, then use Eq. (1.9.10) to find the baryon density $\rho_B(r)$ at each point, and then use Eq. (1.9.11) to find the total mass density at each point. We could then calculate the fractional baryon density ρ_B/ρ_M, and if we were interested we could also use Eq. (1.9.12) to find the velocity dispersion $k_{\mathcal{B}}T_D(r)/m_D$ of the dark matter.

In practice, it is usually necessary to use some sort of cluster model. In the simplest sort of model, one assumes an isothermal sphere: the temperatures T_B and T_D are taken to be independent of position, at least near the center of the cluster where most of the X-rays come from. It is also often assumed that the same gravitational effects that causes the

concentration of the hot intergalactic gas in the cluster is also responsible for the concentration of the dark matter, so that the densities $\rho_B(r)$ and $\rho_M(r)$ are the same, up to a constant factor, which represents the cosmic ratio Ω_B/Ω_M of baryons to all non-relativistic matter. (These gravitational effects are believed to be a so-called "violent relaxation," [11] caused by close encounters of clumps of matter whose gravitational attraction cannot be represented as an interaction with a smoothed average gravitational field. The condensation of galaxies out of this mixture requires quite different cooling processes that can affect only the baryonic gas, which is why galaxies have a lower proportion of dark matter and a lower mass-to-light ratio.) Comparison of Eqs. (1.9.11) and (1.9.12) shows that $\rho_B(r)$ and $\rho_D(r)$ will be proportional to each other, and hence also to $\rho_M(r)$ if the velocity dispersions of the dark matter and hot baryonic gas are the same:

$$k_{\mathcal{B}}T_M/m_M = k_{\mathcal{B}}T_D/m_D \equiv \sigma^2 \tag{1.9.13}$$

Equations (1.9.11) and (1.9.12) both then tell us that

$$\rho_M(r) = \rho_M(0)F\left(r/r_0\right) \tag{1.9.14}$$

where $F(0) \equiv 1; r_0$ is a core radius, defined conventionally by

$$r_0 \equiv \sqrt{\frac{9\sigma^2}{4\pi G\rho_M(0)}}; \tag{1.9.15}$$

and $F(u)$ is a function satisfying the differential equation

$$\frac{d}{du}\left(\frac{u^2}{F(u)}\frac{dF(u)}{du}\right) = -9u^2F(u). \tag{1.9.16}$$

We must also impose the boundary condition that ρ_M is analytic in the coordinate $\mathbf{X}$ at $\mathbf{X} = 0$, which for a function only of r means that it is given near $r = 0$ by a power series in r^2, so that $F(u)$ is given near $u = 0$ by a power series in $u^2, F(u) = 1 + O\left(u^2\right)$. Together with this boundary condition, Eq. (1.9.16) defines a unique function [12] that for small u has the approximate behavior [13]

$$F(u) \simeq \left(1 + u^2\right)^{-3/2}. \tag{1.9.17}$$

The solution to Eq. (1.9.16) is shown together with the approximation (1.9.17) in Figure 1.7.

For large $u, F(u)$ approaches the exact solution $2/9u^2$. Taken literally, this would make the integral for the total mass diverge at large r, which shows that the assumption of constant σ^2 must break down at some large radius. Often the function $F(u)$ is taken simply as [14]

$$F(u) = \left(1 + u^2\right)^{-3\beta/2},$$

where β is an exponent of order unity.

[4711] D. Lynden-Bell, Mon. Not. Roy. Astron. Soc. 136, 101 (1967).

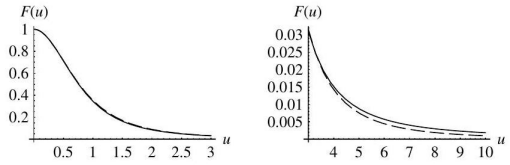


Figure 1.7: The solution to Eq. (1.9.16) (solid line) and the approximation (1.9.17) (dashed line). For the lower values of u in the figure at the left, the two curves are indistinguishable.

Also, Eq. (1.9.11) has the solution

$$\rho_B(r) = \rho_B(0) F\left(r/r_0\right) \tag{1.9.18}$$

with the same function $F(u)$ and the same core radius r_0. We can measure the core radius from the X-ray image of the cluster, and measure σ^2 from the X-ray spectrum, so that Eq. (1.9.15) can be used to find the central density $\rho_M(0)$ of all non-relativistic matter. The central density $\rho_B(0)$ of the baryonic matter can then be found from the total X-ray luminosity, which with these approximations (and using Eq. (1.9.10)) is

$$L_X \equiv \int d^3x \mathcal{L}_X = 4\pi\Lambda\left(T_B\right) r_0^3 \rho_B(0)^2 \mathcal{I} \tag{1.9.19}$$

where

$$\mathcal{I} \equiv \int_0^\infty u^2 F^2(u) du \tag{1.9.20}$$

Even though the solution of Eq. (1.9.16) gives an infinite mass, it gives a finite total X-ray luminosity, with $\mathcal{I} = 0.1961$. (The approximation (1.9.17) would give $\mathcal{I} = \pi/16 = 0.1963$.)

For a cluster at redshift z, the core radius r_0 inferred from observation of the angular size of the cluster will be proportional to the angular diameter distance $d_A(z)$, while the temperature and velocity dispersion found from the X-ray spectrum will not depend on the assumed distance. Thus the value of the central total matter density $\rho_M(0)$ given by Eq. (1.9.15) will be proportional to $1/d_A^2(z)$. On the other hand, the absolute X-ray luminosity L_X inferred from the apparent X-ray luminosity will (like all absolute luminosities) be proportional to the value assumed for $d_L^2(z)$, the square of the luminosity distance, so with $r_0 \propto d_A$, the central baryon density $\rho_B(0)$ given by Eq. (1.9.19) will be proportional to $[d_L^2(z)/d_A^3(z)]^{1/2}$. The value of the ratio of central densities inferred from observations of a given cluster at redshift z will therefore have a dependence on the distance assumed for the cluster given by

$$\frac{\rho_B(0)}{\rho_M(0)} \propto d_L(z) d_A^{1/2}(z) = (1+z)^2 d_A^{3/2}(z) \tag{1.9.21}$$

in which we have used the relation (1.4.12) between luminosity and angular diameter distances.

For $z \ll 1$, we have $d_A(z) \simeq d_L(z) \simeq z/H_0$, and so according to Eq. (1.9.21) the value of $\rho_B(0)/\rho_M(0)$ obtained from observations of clusters of small redshift will be proportional to the assumed value of $H_0^{-3/2}$. It is believed that most of the baryonic mass in a cluster of galaxies is in the

[4712] For a tabulation of values of $F(u)$, see e.g. J. Binney and S. Tremaine, Galactic Dynamics (Princeton University, Princeton, 1987): Table 4.1.
[13] I. R. King, Astron. J. 67, 471 (1962).
[14] A. Cavaliere and R. Fusco-Fermiano, Astron. Astrophys. 49, 137 (1976).

hot gas outside the galaxies, and if we suppose that this mass is the same fraction of the total mass as in the universe as a whole, [15] then we should get the same value of $\rho_B(0)/\rho_M(0)$, equal to Ω_B/Ω_M, for all clusters, whatever value we assume for H_0, but this value of Ω_B/Ω_M will be proportional to the assumed value of $H_0^{-3/2}$. For example, Schindler [16] quotes various studies that give $\rho_B(0)/\rho_M(0)$ as $0.14, 0.11, 0.12$, and 0.12 for $H_0 = 65$ km s^{-1}Mpc^{-1}, so if we take the average 0.12 of these values as the cosmic value of Ω_B/Ω_M for $H_0 = 65$ km s^{-1}Mpc^{-1}, then for a general Hubble constant we find

$$\Omega_B/\Omega_M \simeq 0.06h^{-3/2} \tag{1.9.22}$$

where h as usual is Hubble's constant in units of 100 km s^{-1}Mpc^{-1}. We can thus conclude pretty definitely that only a small fraction of the mass in clusters of galaxies is in a baryonic form that can emit X-rays.

On the other hand, when we study clusters with a range of redshifts that are not all small, we will not get a uniform value of $\rho_B(0)/\rho_M(0)$ unless we use values of $d_A(z)$ with the correct dependence on z. As remarked in Section 1.6, observations of clusters have been used in this way to learn about the z-dependence of $d_A(z)$.

It should be mentioned that computer simulations that treat galaxy clusters as assemblages of collisionless particles do not show evidence for a central core, [17] but instead indicate that the dark matter density at small distances r from the center should diverge as r^{-1} to $r^{-3/2}$. On the other hand, it has been shown [18] that the density of a baryonic gas in hydrostatic equilibrium in the gravitational field of such a distribution of dark matter does exhibit the core expected from Eq. (1.9.18). In any case, the dark matter and baryonic gas densities do have the same distributions at distances from the center that are larger than r_0.

As we will see in Section 3.2, it is possible to infer a value for $\Omega_B h^2$ from the abundances of deuterium and other light isotopes, which together with Eq. (1.9.22) can be used to derive a value for $\Omega_M h^{1/2}$. There are several other methods for estimating Ω_M or $\Omega_M h^2$ that will be discussed elsewhere

in this book, using gravitational lenses (Section 9.3), the Sunyaev-Zel'dovich effect (Section 2.5), and anisotropies in the cosmic microwave background (Sections 2.6 and 7.2), the last of which also gives a value for $\Omega_B h^2$. In addition to these, there are methods [19] based on the evolution of clusters of galaxies, cosmic flows, cluster correlations, etc., that depend on detailed dynamical theories of structure formation.

10.1.12 1.10 Intergalactic absorption

Some of the cosmic gas of nuclei and electrons from which the first galaxies and clusters of galaxies condensed must be still out there in intergalactic space. Atoms or molecules in this gas could be observed through the resonant absorption of light or radio waves from more distant galaxies or quasars, but it is believed that most of the gas was ionized by light from a first generation of hot massive stars, now long gone, that are sometimes called stars of Population III. It now appears that some quasars formed before this ionization was complete, giving us

[4715] This is argued by S. D. M. White, J. F. Navarro, A. E. Evrard, and C. S. Frenk, Nature 366, 429 (1993). Calculations supporting this assumption are described in Section 8.3.

[16] S. Schindler, in Space Science Reviews 100, 299 (2002), ed. P. Jetzer, K. Pretzl, and R. von Steiger (Kluwer) [astro-ph/0107028].

[17] J. F. Navarro, C. S. Frenk, and S. D. M. White, Astrophys. J. 462, 563 (1996) [astro-ph/9508025]; 490, 493 (1997) [astro-ph/9610188]; T. Fukushige and J. Makino, Astrophys. J. 477, L9 (1997) [astroph/9610005]; B. Moore et al., Mon. Not. Roy. Astron. Soc. 499, L5 (1998).

[18] N. Makino, S. Sasaki, and Y. Suto, Astrophys. J. 497, 555 (1998). Also see Y. Suto, S. Sasaki, and M. Makino, Astrophys. J. 509, 544 (1998); E. Komatsu and U. Seljak, Mon. Not. Roy. Astron. Soc. 327, 1353 (2001).

the opportunity to observe the intergalactic gas through resonant absorption of the light from these very distant quasars.

Let us suppose that an atomic transition in a distant source produces a ray of light of frequency v_1 that leaves the source at time t_1 and arrives at the Earth with frequency v_0 at time t_0. At time t along its journey the light will have frequency redshifted to $v_1 a(t_1)/a(t)$, so if the intergalactic medium absorbs light of frequency ν at a rate (per proper time) $\Lambda(\nu, t)$, and does not emit light, then the intensity $I(t)$ of the light ray will decrease according to the equation

$$\dot{I}(t) = -\Lambda\left(v_1 a(t_1)/a(t), t\right) I(t)$$

But if the intergalactic gas is at a non-zero temperature $T(t)$, then photons will also be added to the light ray through the process of stimulated emission, as a rate per photon given by the Einstein formula [1] $\exp(-hv/k_{\mathcal{B}}T)\,\Lambda(v,t)$, so the intensity of the light ray will satisfy

$$\dot{I}(t) = -\left[1 - \exp\left(-\frac{hv_1 a(t_1)}{k_{\mathcal{B}} T(t) a(t)}\right)\right] \Lambda\left(v_1 a(t_1)/a(t), t\right) I(t) \tag{1.10.1}$$

The intensity observed at the earth will then be

$$I(t_0) = \exp(-\tau) I(t_1) \tag{1.10.2}$$

where τ is the optical depth:

$$\tau = \int_{t_1}^{t_0} \left[1 - \exp\left(-\frac{hv_1 a(t_1)}{k_{\mathcal{B}} T(t) a(t)}\right)\right] \Lambda\left(v_1 a(t_1)/a(t), t\right) dt \tag{1.10.3}$$

The absorption rate is given by

$$\Lambda(v,t) = n(t)\sigma(v) \tag{1.10.4}$$

where $\sigma(v)$ is the absorption cross section at frequency v, and $n(t)$ is the number density (per proper volume) of absorbing atoms. Often the absorption cross section is sharply peaked at some frequency v_R, so the absorption takes place only close to a time t_R, given by

$$a(t_R) = v_1 a(t_1)/v_R \tag{1.10.5}$$

Therefore the optical depth can be approximated as

$$\tau \simeq n(t_R)\left[1 - \exp\left(-hv_R/k_{\mathcal{B}}T(t_R)\right)\right] \int \sigma\left(v_1 a(t_1)/a(t)\right) dt$$

By changing the variable of integration from time to frequency, we can write this as

$$\tau \simeq n(t_R)\left[1 - \exp\left(-hv_R/k_{\mathcal{B}}T(t_R)\right)\right]\left[a(t_R)/\dot{a}(t_R)\right]\mathcal{I}_R \tag{1.10.6}$$

where

$$\mathcal{I}_R \equiv \frac{1}{v_R}\int \sigma(v) dv \tag{1.10.7}$$

[4719] For surveys, see N. A. Bahcall, Astrophys. J. 535, 593 (2000) [astro-ph/0001076]; M. Turner, Astrophys. J. 576, L101 (2002) [astro-ph/0106035]; S. Schindler, op. cit.; K. A. Olive, lectures given at Theoretical Advanced Study Institute on Elementary Particle Physics, Boulder, June 2002 [astro-ph/0301505].

[1] A. Einstein, Phys. Z. 18, 121 (1917).

the integral being taken over a small range of frequencies containing the absorption line. The only thing in the formula for τ that depends on a cosmological model is the Hubble expansion rate $\dot{a}(t_R)/a(t_R)$ at the time of absorption, given by Eq. (1.5.19) and (1.5.38) as

$$\frac{\dot{a}(t_R)}{a(t_R)} = H_0\sqrt{\Omega_\Lambda + \Omega_K(1+z_R)^2 + \Omega_M(1+z_R)^3 + \Omega_R(1+z_R)^4} \tag{1.10.8}$$

where $z_R = a(t_0)/a(t_R) - 1 = v_R/v_0 - 1$ is the redshift of the location of the resonant absorption. For a source at redshift z, the absorption takes place over a range of observed frequencies $v_0 = v_1/(1+z)$ given by the condition that the time t_R defined by Eq. (1.10.5) should be between t_1 and t_0 :

$$v_R/(1+z) \leq v_0 \leq v_R. \tag{1.10.9}$$

For example, in 1959 Field [2] suggested looking for the effects of absorption of radio frequencies in the 21 cm transition in hydrogen atoms, caused in transitions from the spin zero to spin one hyperfine states in the $1s$ state of intergalactic hydrogen. Here $v_R = 1420$MHz, so the radio spectrum of the galaxy Cygnus A at a redshift $z = 0.056$ should show an absorption trough (1.10.9) from 1342 MHz to 1420 MHz . Unfortunately, the temperature of neutral hydrogen in intergalactic space is much larger than $hv_R/k_{\mathcal{B}} = 0.068$ K, so the optical depth (1.10.6) is suppressed by a factor $\simeq 0.068\text{ K}/T(t_R)$. No sign of this absorption trough has been discovered. It is hoped that in the future a new generation of low frequency radio telescopes with good angular resolution may be able to use the emission and absorption of 21 cm radiation at large redshifts to study both the growth of structure and primordial density perturbations from which they grew. [3] For instance, by 2010 the Low Frequency Array (LOFAR) should be able to study 21 cm radiation from sources at redshift between 5 and 15 with good sensitivity and high angular resolution. [4]

For the present, a much better probe of intergalactic hydrogen atoms is provided by absorption of photons in the Lyman α transition from the $1s$ ground state to the $2p$ excited state, known as the Gunn-Peterson effect. [5] This has a resonant frequency in the ultraviolet, $v_R = 2.47\times10^{15}$ Hz, corresponding to a wavelength $1,215$Å, but for a source of redshift $z > 1.5$ the lower part or the absorption trough (1.10.9) will be observable on the Earth's surface at wavelengths greater than $3,000$Å, in the visible or infrared part of the spectrum. Here $hv_R/k_{\mathcal{B}} = 118,000$ K, which is likely to be larger than the temperature of the intergalactic medium, in which case the factor $1-\exp(-hv_R/k_{\mathcal{B}}T(t_R))$ in Eq. (1.10.6) can be set equal to unity. The integral (1.10.7) here has the value 4.5×10^{-18} cm^2, so Eq. (1.10.6) gives the optical depth just above the lower end of the absorption trough (1.10.9) as

$$\tau_{v_0=v_R/(1+z)+} = \left(\frac{n(t_R)}{2.4h\times10^{-11}\text{ cm}^{-3}}\right)\left(\Omega_\Lambda + \Omega_K(1+z)^2 + \Omega_M(1+z)^3 + \Omega_R(1+z)^4\right)^{-1/2} \tag{1.10.10}$$

where again h is Hubble's constant in units of 100 km s^{-1}Mpc^{-1}. For instance, if a fraction f of the baryons of the universe at a time corresponding

[472] G. Field, Astrophys. J. 129, 525 (1959).

[3] A. Loeb and M. Zaldarriaga, Phys. Rev. Lett. 92, 211301 (2004) [astro-ph/0312134]; S. Furlanetto, S. P. Oh, and F. Briggs, Phys. Rep. 433, 181 (2006) [astro-ph/0608032].

[4] H. J. A. Röttgering et al., in Cosmology, Galaxy Formation, and Astroparticle Physics on the Pathway to the SKA, eds. H.-R. Klöckner et al. [astro-ph/0610596].

[5] J. E. Gunn and B. A. Peterson, Astrophys. J. 142, 1633 (1965). Also see I. S. Shklovsky, Astron. Zh. 41, 408 (1964); P. A. G. Scheuer, Nature 207, 963 (1965).

to $z = 5$ were in the form of neutral intergalactic hydrogen atoms, and $\Omega_B h^2 = 0.02$, then the number density of hydrogen atoms at $z = 5$ would be $4.8f \times 10^{-5}$ cm^{-3}. Taking $h = 0.65, \Omega_\Lambda = 0.7, \Omega_M = 0.3$, and $\Omega_K = \Omega_R = 0$, the optical depth (1.10.10) would be $3.8f \times 10^5$. Thus with these parameters intergalactic neutral hydrogen that makes up a fraction of baryonic matter $f \gg 2.6 \times 10^{-6}$ would completely block any light with a frequency above the redshifted Lyman α line from sources beyond $z = 5$. Evidently the Gunn-Peterson effect provides a very sensitive probe of even a small proportion of neutral hydrogen atoms.

For many years the search for the Lyman α absorption trough was unsuccessful. Quasar spectra show numerous Lyman α absorption lines, forming what are sometimes called "Lyman α forests," which are believed to arise from clouds of neutral hydrogen atoms along the line of sight, but for quasars out to $z \approx 5$ there was no sign of a general suppression of frequencies above the redshifted Lyman α frequency, [6] that would be produced by even a small fraction f of the baryons in the universe in the form of neutral intergalactic hydrogen atoms. Then in 2001 the spectrum of the quasar SDSSp J103027.10+052455.0 with redshift $z = 6.28$ discovered by the Sloan Digital Sky Survey was found to show clear signs of a complete suppression of light in the wavelength range from just below the redshifted Lyman α wavelength at $8,845$Å down to $8,450$Å, indicating a significant fraction f of baryons in the form of neutral intergalactic hydrogen atoms at redshifts greater than $8,450/1,215 - 1 = 5.95$.[7] (See Figure 1.8.) Thus a redshift of order 6 may mark the end of a "dark age," in which the absorption of light by neutral hydrogen atoms made the universe opaque to light with frequencies above the redshifted Lyman α frequency. Further evidence for this conclusion is supplied by the spectrum of intense gamma ray sources, known as gamma ray bursters, at large redshifts. [8]

This does not mean that all or even most of the hydrogen in the universe was in the form of neutral atoms at $z > 6$. As we have seen, even small concentrations of neutral hydrogen could have produced an absorption trough in the spectrum of distant quasars. In fact, we shall see in Chapter 7 that there is now some evidence from the study of the cosmic microwave background that hydrogen became mostly ionized at redshifts considerably larger than $z \approx 6$, perhaps around $z \approx 10$.

The clouds of neutral hydrogen at redshifts $z < 6$ which produce the Lyman α forest can provide an independent means of measuring Ω_M and Ω_Λ. The idea goes back to a 1979 paper of Alcock and Paczyński. [9] Suppose we observe a luminous object at a redshift z that extends a proper distance $D_\perp$ perpendicular to the line of sight and a proper distance $D_\parallel$along the line of sight. According to the definition of the angular diameter distance, the object will subtend an angle

$$\Delta\theta = D_\perp / d_A(z) \tag{1.10.11}$$

Also, when we observe light from the whole object at the same time t_0, the difference in the time t_1 that the light was emitted from the far and near points of the object will be $\Delta t_1 = D_\parallel$. The redshift is $a(t_0)/a(t_1) - 1$, so the absolute value of the difference of redshift from the far and near points of the object will be

$$\Delta z = \frac{a(t_0)}{a^2(t_1)} \dot{a}(t_1) \Delta t_1 = (1+z) H(z) D_\parallel \tag{1.10.12}$$

where $H(z) \equiv \dot{a}(t_1)/a(t_1)$ is the Hubble constant at the time of emission. Taking the ratio, we have

[476] A. Songalia, E. Hu, L. Cowie, and R. McMahon, Astrophys. J. 525, L5 (1999).

[7] R. H. Becker et al., Astron. J. 122, 2850 (2001) [astro-ph/0108097]. See S. G. Djorgovski et al., Astrophys. J. 560, L5 (2001) [astro-ph/0108069], for a hint of absorption by neutral hydrogen at slightly smaller redshifts. Also see X. Fan et al., Astrophys. J. 123, 1247 (2002) [astro-ph/0111184].

[8] T. Totani et al., Publ. Astron. Soc. Pacific 58, 485 (2006) [astro-ph/0512154].

$$\frac{\Delta z}{\Delta\theta} = (1+z)H(z)d_A(z)\left(D_\parallel/D_\perp\right) \tag{1.10.13}$$

It is then only necessary to use Eq. (1.5.19) to write $H(z)$ as

$$\begin{aligned} H(z) &= \sqrt{\left(\frac{8\pi G}{3}\right)(\rho_{M0}(1+z)^3 + \rho_V + \rho_{R0}(1+z)^4) - \frac{K}{a_0^2}(1+z)^2} \\ &= H_0\sqrt{\Omega_M(1+z)^3 + \Omega_\Lambda + \Omega_R(1+z)^4 + \Omega_K(1+z)^2} \end{aligned} \tag{1.10.14}$$

and use Eqs. (1.4.12) and (1.5.45) to write $d_A(z)$ as

$$\begin{aligned} d_A(z) =& \frac{1}{(1+z)H_0\Omega_K^{1/2}} \\ &\times \sinh\left[\Omega_K^{1/2}\int_{1/(1+z)}^1 \frac{dx}{x^2\sqrt{\Omega_\Lambda + \Omega_K x^{-2} + \Omega_M x^{-3} + \Omega_R x^{-4}}}\right] \end{aligned} \tag{1.10.15}$$

The Hubble constant H_0 cancels in the product, and we find a result that depends only on $z, D_\parallel/D_\perp$, and the Ω s:

$$\begin{aligned} \frac{\Delta z}{\Delta\theta} =& \left(D_\parallel/D_\perp\right)\Omega_K^{-1/2}\sqrt{\Omega_M(1+z)^3 + \Omega_\Lambda + \Omega_R(1+z)^4 + \Omega_K(1+z)^2} \\ &\times \sinh\left[\Omega_K^{1/2}\int_{1/(1+z)}^1 \frac{dx}{x^2\sqrt{\Omega_\Lambda + \Omega_K x^{-2} + \Omega_M x^{-3} + \Omega_R x^{-4}}}\right] \end{aligned} \tag{1.10.16}$$

For instance, if the object is known to be a sphere, such as a spherical cluster of galaxies, then $D_\parallel/D_\perp = 1$, and we can use a measurement of Δz and $\Delta\theta$ to set a model-independent constraint on the Ωs, with no need to worry about effects of evolution or intergalactic absorption.

Unfortunately, it is not so easy to find spherical objects at large redshift. But there are various objects whose distribution functions are spherically symmetric. For instance, the distribution of field galaxies is presumably spherically symmetric about any point in space, and it has been proposed that the application of the Alcock-Paczyński method to galaxies might allow a determination of the cosmological constant. [10] This method has been applied [11] instead to the distribution of quasars measured in the 2 dF QSO Redshift Survey. [12] Assuming $K = 0$, this analysis gives $\Omega_\Lambda = 0.71^{0.09}_{-0.17}$.

Recently the Alcock-Paczyński idea has been applied to the distribution function of Lyman α clouds. [13] As already mentioned, these are intergalactic clouds containing neutral hydrogen atoms, which absorb light from more distant quasars along the line of sight in $1s \to 2p$ transitions, showing up as dark lines in the spectrum of the quasar at wavelengths $1215(1+z)$Å for clouds at redshift z. Suppose we measure the number density $N(z,\hat{n})$ of Lyman α clouds at various redshifts z in various directions $\hat{n}$. Assuming a spherically symmetric distribution of Lyman α clouds, the mean value of the product of the number densities of these clouds at two nearby points with redshifts z and $z+\Delta z$ (with $\Delta z \ll 1$) and directions $\hat{n}$ and $\hat{n}+\Delta\hat{n}$ separated by a small angle $\Delta\theta$ will be a function only of z and the proper distance between the points, and will be analytic in the components of the vector

[479] C. Alcock and B. Paczyński, Nature 281, 358 (1979).

[4710] W. E. Ballinger, J. A. Peacock, and A. F. Heavens, Mon. Not. Roy. Astron. Soc. 281, 877 (1996).

[11] P. J. Outram et al., Mon. Not. Roy. Astron. Soc. 348, 745 (2004) [astro-ph/0310873].

[12] S. M. Croom et al., Mon. Not. Roy. Astron. Soc. 349, 1397 (2004); available at www.2df quasar.org.

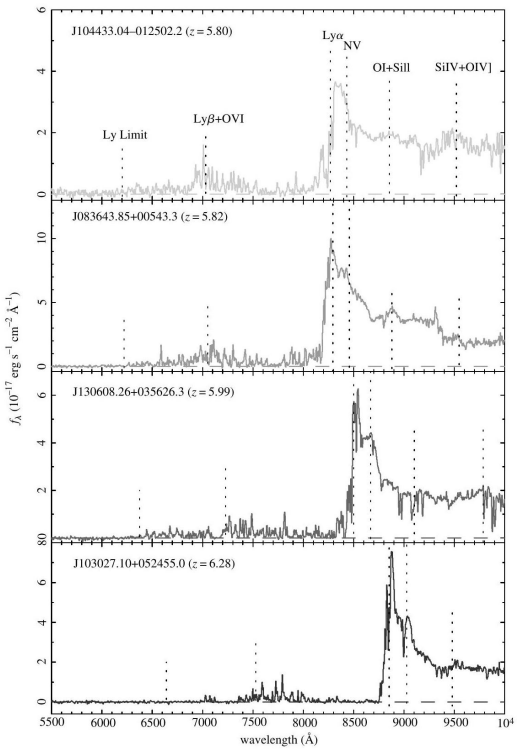


Figure 1.8: Observed intensity versus wavelength for four high-redshift quasars, from R. H. Becker et al., Astron. J. 122, 2850 (2001) [astro-ph/0108097]. Vertical dashed lines indicate the redshifted wavelengths for various spectral lines. In the direction of the quasar with $z = 6.28$ the intensity drops to zero within experimental accuracy just to the left of the Lyman α line at 8845Å, a feature not seen for the quasar with $z = 5.99$, indicating the onset of patches of nearly complete ionization at a redshift between 5.99 and 6.28.

separating these components, so for small separations it can be written

$$\langle N(z,\hat{n})N(z+\Delta z,\hat{n}+\Delta\hat{n})\rangle \simeq \left\langle N^2(z,\hat{n})\right\rangle \left[1-\frac{D_\perp^2+D_\parallel^2}{L^2(z)}\right] \tag{1.10.17}$$

where $D_\perp$ and $D_\parallel$are given by Eqs. (1.10.11) and (1.10.12), and L is some correlation length. This can be written in terms of the observed Δz and $\Delta\theta$, as

$$\langle N(z,\hat{n})N(z+\Delta z,\hat{n}+\Delta\hat{n})\rangle \simeq \left\langle N^2(z,\hat{n})\right\rangle \left[1-\frac{\Delta z^2}{L_z^2(z)}-\frac{\Delta\theta^2}{L_\theta^2(z)}\right] \tag{1.10.18}$$

where L_z and L_θ are correlation lengths for redshift and angle

$$L_\theta(z)=\frac{L(z)}{d_A(z)},\quad L_z(z)=L(z)(1+z)H(z) \tag{1.10.19}$$

By measuring this product for various redshifts and directions, we can infer a value for the ratio of correlation lengths, which is independent of L :

[13] L. Hui, A. Stebbins, and S. Burles, Astrophys. J. 511, L5 (1999); P. McDonald and J. MiraldaEscudeé, Astrophys. J. 518, 24 (1999); W-C. Lin and M. L. Norman, talk at the Theoretical Astrophysics in Southern California meeting, Santa Barbara, October 2002 [astro-ph/0211177]; P. McDonald, Astrophys. J. 585, 34 (2003).

$$\frac{L_z(z)}{L_\theta(z)} = \Omega_K^{-1/2}\sqrt{\Omega_M(1+z)^3+\Omega_\Lambda+\Omega_R(1+z)^4+\Omega_K(1+z)^2} \times \sinh\left[\Omega_K^{1/2}\int_{1/(1+z)}^{1}\frac{dx}{x^2\sqrt{\Omega_\Lambda+\Omega_K x^{-2}+\Omega_M x^{-3}+\Omega_R x^{-4}}}\right] \tag{1.10.20}$$

This method has been applied [14] to five pairs of close quasars, with redshifts in the range from 2.5 to 3.5 and separations ranging from 33 to 180 arcseconds. Use of this limited sample sets only weak constraints on the Ωs, but it rules out $\Omega_\Lambda = 0$ at the level of 2 standard deviations.

10.1.13 1.11 Number counts

A uniform distribution of sources with a smooth distribution of absolute luminosity leads in ordinary Euclidean space to a unique distribution in apparent luminosity. If there are $N(L)dL$ sources per unit volume with absolute luminosity between L and $L + dL$, then the number $n(> \ell)$ of

sources observed with apparent luminosity greater than ℓ is given by

$$\begin{aligned} n(>\ell) &= \int_0^\infty N(L)dL\int_0^{\sqrt{L/4\pi\ell}} 4\pi r^2 dr \\ &= \frac{1}{3\sqrt{4\pi}\ell^{3/2}}\int_0^\infty L^{3/2}N(L)dL \end{aligned} \tag{1.11.1}$$

Thus whatever the distribution in absolute luminosity, we expect that $n(> \ell) \propto \ell^{-3/2}$.

This analysis needs several changes in a cosmological setting:

1. Instead of the volume element $r^2\sin\theta dr d\theta d\phi$, the proper volume element here is $\left(\mathrm{Det}\, g^{(3)}\right)^{1/2} dr d\theta d\phi$, where $g^{(3)}_{ij} \equiv a^2\tilde{g}_{ij}$ is the three-dimensional metric, with non-vanishing components $g^{(3)}_{rr} = a^2/\left(1-Kr^2\right), g^{(3)}_{\theta\theta} = a^2r^2, g^{(3)}_{\phi\phi} = a^2r^2\sin^2\theta$, so

$$dV = \frac{a^3(t)r^2\sin\theta dr d\theta d\phi}{\sqrt{1-Kr^2}} \tag{1.11.2}$$

2. The apparent luminosity is related to the absolute luminosity by

$$\ell = \frac{L}{4\pi d_L^2(z)} \tag{1.11.3}$$

where $d_L(z)$ is the luminosity distance (1.4.3).

3. Except in the steady state cosmology, the number density of sources changes with time, even if only through the cosmic expansion.

4. We can often measure the redshift z as well as the apparent luminosity.

Eq. (1.11.2) gives the number of sources with redshift between z and $z + dz$ and apparent luminosity between ℓ and $\ell + d\ell$ as

$$n(z,\ell)dzd\ell = 4\pi\mathcal{N}(t,L)dL\frac{a^3(t)r^2dr}{\sqrt{1-Kr^2}} \tag{1.11.4}$$

[4714] A. Lidz, L. Hui, A. P. S. Crotts, and M. Zaldarriaga, astro-ph/0309204 (unpublished).

where $\mathcal{N}(t, L)dL$ is the number of sources per proper volume at time t with absolute luminosity between L and $L + dL$; t and z are related by $1 + z = a\,(t_0)\,/a(t)$, and t and r are related by Eq. (1.2.2):

$$\int_t^{t_0} \frac{dt'}{a\,(t')} = \int_0^r \frac{dr'}{\sqrt{1 - Kr'^2}} \tag{1.11.5}$$

We use (1.11.5) to express the differential dr in terms of dt, and then express dt in terms of dz :

$$\frac{dr}{\sqrt{1 - Kr^2}} = -\frac{dt}{a(t)} = \frac{dz}{H(z)a_0}$$

where $H(z) \equiv \dot{a}(t)/a(t)$ and $a_0 \equiv a\,(t_0)$. As a reminder, for a universe containing radiation, matter, and a constant vacuum energy, Eq. (1.5.41) gives

$$H(z) = H_0\sqrt{\Omega_\Lambda + \Omega_K(1 + z)^2 + \Omega_M(1 + z)^3 + \Omega_R(1 + z)^4}$$

Canceling dz in Eq. (1.11.4), we then have

$$n(z, \ell)d\ell = \frac{4\pi\mathcal{N}(t(z), L)r^2(z)a_0^2 dL}{(1 + z)^3 H(z)}$$

We next use Eq. (1.11.3) to write (with z now held fixed):

$$dL = 4\pi d_L^2(z)d\ell$$

so that canceling $d\ell$ gives

$$n(z, \ell) = \frac{16\pi^2\mathcal{N}\,(t(z), 4\pi d_L^2(z)\ell)\,d_L^4(z)}{H(z)(1 + z)^5} \tag{1.11.6}$$

in which we have used Eq. (1.4.3) to express $a_0 r$ in terms of d_L.
In particular, for a sample of sources that are not evolving at a time $t(z)$, the time dependence of the number density $\mathcal{N}$ is just proportional to $a^{-3} \propto (1 + z)^3$:

$$\mathcal{N}(t(z), L) = (1 + z)^3\mathcal{N}_0(L) \tag{1.11.7}$$

If all members of this sample are bright enough to be visible at a redshift z, then the total number of sources observed with redshifts between z and $z + dz$ will be $n(z)dz$, where

$$n(z) \equiv \int_0^\infty n(z, \ell)d\ell = \frac{4\pi\mathcal{N}_0 d_L^2(z)}{H(z)(1 + z)^2} \tag{1.11.8}$$

where $d_L(z)$ is given by Eq. (1.5.45), and

$$\mathcal{N}_0 \equiv \int_0^\infty \mathcal{N}_0(L)dL \tag{1.11.9}$$

In principle, even without knowing $\mathcal{N}_0$ or H_0, if $n(z)$ were accurately measured we could compare the observed shape of this function with Eq. (1.11.8) to find the Ω s.

There are several obvious dangers in using Eq. (1.11.8) in this way. For one thing, it is necessary to avoid missing sources that have high redshift and hence low apparent luminosity. Also, evolution in the number of sources can introduce an additional dependence on the light emission time t, and hence on z. In 1986 Loh and Spillar [1] carried out a survey of galaxy numbers as a function of redshift. The redshifts were measured photometrically (i. e., from their luminosities at various colors rather than by the shift of specific spectral lines), which

generally gives less reliable results. Comparing their results with Eq. (1.11.8) in the case $\Omega_K = \Omega_R = 0$ (so that $\Omega_\Lambda + \Omega_M = 1$), they found that $\Omega_\Lambda/\Omega_M = 0.1^{-0.4}_{+0.2}$. By now it has been realized that the evolution of sources cannot be neglected at redshifts large enough for $n(z)$ to be sensitive to cosmological parameters, and this result for Ω_Λ/Ω_M has been abandoned. [2]

Useful results can be obtained when evolution is taken into account. One group [3] used number counts of very faint galaxies [4] as a function of apparent luminosity to estimate the free parameters in a model of galactic luminosity evolution (assuming the number of galaxies per coordinate volume to be constant), and then used this model together with a redshift survey [5] extending to $z \simeq 0.47$ to conclude that Ω_M is small and that Ω_Λ is in the range of 0.5 to 1 . More recently, several surveys [6] of numbers of galaxies at different redshifts that yield important results about galactic evolution, and with the use of dynamical models they can yield information about Ω_M and Ω_Λ.[7] But it appears that number counts of galaxies will be more useful in learning about galactic evolution than in making precise determinations of cosmological parameters. In a dramatic application of this approach, [8] a

search at the Lick Observatory for galaxies with redshifts in the range $z \approx 7$ to 8 found at most just one galaxy, while it is estimated that if Eq. (1.11.7) were valid then, on the basis of the number of galaxies observed (with the same conservative selection criteria) at redshifts $z \approx 6$, ten galaxies should have been found with $z \approx 7$ to 8 . The implication is that there must have been a spurt in the formation of luminous galaxies at a redshift in the range 6 to 7. This fits in well with the conclusion discussed in Section 1.10, that the ionization of intergalactic hydrogen became essentially complete at a redshift of order 6, presumably due to ultraviolet radiation from massive stars formed around that time.

Historically the first important application of number counts was in radio source surveys, where redshifts are not generally available. These surveys take place at a fixed receiving frequency ν, corresponding to a variable emitted frequency $v(1+z)$, so the source counts are affected by the frequency dependence of the distribution of intrinsic source powers.

If a source with a redshift z emits a power [9]$P(v)dv$ between frequencies v and $v + dv$, then the power received at the origin per unit antenna area between frequencies ν and $\nu + d\nu$ is

$$S(v)dv = \frac{P(v(1+z))dv(1+z)}{4\pi d_L^2(z)} \tag{1.11.10}$$

Many radio sources have a "straight" spectrum, i.e.

$$P(v) \propto v^{-\alpha} \tag{1.11.11}$$

with the spectral index α typically about 0.7 to 0.8 . This allows a great simplification in Eq. (1.11.10):

[471] E. D. Loh, Phys. Rev. Lett. 57, 2865 (1986); E. D. Loh and E. J. Spillar, Astrophys. J. 284, 439 (1986).
[2] For a discussion of future prospects for measuring Ω_Λ in redshift surveys, see W. E. Ballinger, J. A. Peacock, and A. F. Heavens, Mon. Not. Roy. Astron. Soc. 282, 877 (1996).
[3] M. Fukugita, F. Takahara, K. Yamashita, and Y. Yoshii, Astrophys. J. 361, L1 (1990).
[4] J. A. Tyson, Astron. J. 96, 1 (1988).
[5] T. J. Broadhurst, R. S. Ellis, and T. Shanks, Mon. Not. Roy. Astron. Soc. 235, 827 (1988).
[6] G. Efstathiou, R. S. Ellis, B. A. Peterson, Mon. Not. Roy. Astron. Soc. 232, 431 (1988); J. Loveday, B. A. Peterson, G. Efstathiou, and S. J. Maddox, Astrophys. J. 390, 338 (1992); L. da Costa, in Proceedings of the Conference on Evolution of Large Scale Structure, Garching, August 1998 [astroph/9812258]; S. Borgani, P. Rosati, P. Tozzi, and C. Norman, Astrophys. J. 517, 40 (1999) [astroph/9901017]; S. J. Oliver, in Highlights of the ISO Mission: Special Scientific Session of the IAU General Assembly. eds. D. Lemke et al. (Kluwer) [astro-ph/9901272]; M. Colless, in Publ. Astron. Soc. Australia [astro-ph/9911326]; S. Rawlings, astro-ph/0008067.
[7] W. J. Percival et al., Mon. Not. Roy. Astron. Soc. 327, 1297 (2001) [astro-ph/0105252]; S. Borgnani et al., Astrophys. J. 561, 13 (2001) [astro-ph/0106428].
[8] R. J. Bouwens and G. D. Illingworth, Nature 443, 189 (2006).

$$S(v)dv = \frac{P(v)dv}{4\pi d_L^2(z)(1+z)^{\alpha-1}} \tag{1.11.12}$$

From now on we will take the observed frequency v as fixed, and write $S(v) = S$ and $P(v) = P$. Canceling dv, Eq. (1.11.12) then reads

$$S = \frac{P}{4\pi d_L^2(z)(1+z)^{\alpha-1}} \tag{1.11.13}$$

If at time t there are $N(P,t)dP$ sources per proper volume with power between P and $P + dP$, then the number of sources observed with power per antenna area greater than S is

$$n(>S) = \int_0^\infty dP \int N(P,t)\frac{4\pi r^2 a^3(t)dr}{\sqrt{1-Kr^2}} \tag{1.11.14}$$

with the upper limit on the integral over r set by the condition that

$$a_0^2 r^2 (1+z)^{1+\alpha} < \frac{P}{4\pi S} \tag{1.11.15}$$

Of course, r, z, and t are related by the familiar formulas

$$\int_0^r \frac{dr'}{\sqrt{1-Kr'^2}} = \int_t^{t_0} \frac{dt'}{a\left(t'\right)}, \quad 1 + z = a\left(t_0\right)/a(t) \tag{1.11.16}$$

This becomes much simpler if we assume that the time-dependence of the source number density can be parameterized as

$$N(P,t) = N(P)\left(\frac{a(t)}{a_0}\right)^\beta \tag{1.11.17}$$

For instance, if sources do not evolve and are neither created nor destroyed, then $\beta = -3$, while in the steady-state model $\beta = 0$. Eq. (1.11.14) now reads

$$n(>S) = a_0^3 \int_0^\infty N(P)dP \int \frac{4\pi r^2 (1+z)^{-\beta-3} dr}{\sqrt{1-Kr^2}} \tag{1.11.18}$$

with the same P/S-dependent upper limit (1.11.15) on r.
The coordinate r is given in terms of z by the power series (1.4.8)

$$a_0 H_0 r = z - \frac{1}{2}\left(1 + q_0\right) z^2 + \ldots \tag{1.11.19}$$

We can then convert the integral over r to one over z, with

$$a_0 H_0 dr = dz\left[1 - \left(1 + q_0\right) z + \ldots\right] \tag{1.11.20}$$

and the upper limit on z is given by

$$z^2\left[1 + z\left(\alpha - q_0\right) + \ldots\right] < \frac{PH_0^2}{4\pi S}$$

or, in other words,

$$z < \sqrt{\frac{PH_0^2}{4\pi S}}\left[1 - \frac{1}{2}\left(\alpha - q_0\right)\sqrt{\frac{PH_0^2}{4\pi S}} + \ldots\right] \tag{1.11.21}$$

[479] In G&C, P was defined as the power emitted per solid angle, while here it is the power emitted in all directions.

Then Eq. (1.11.18) becomes

$$n(>S) = \frac{4\pi}{H_0^3}\int_0^\infty N(P)dP \times\left[\frac{1}{3}\left(\frac{PH_0^2}{4\pi S}\right)^{3/2}\left(1-\frac{3}{2}(\alpha-q_0)\left(\frac{PH_0^2}{4\pi S}\right)^{1/2}+\ldots\right) -\frac{1}{4}\left(\frac{PH_0^2}{4\pi S}\right)^2(\beta+5+2q_0)+\ldots\right]$$

or, collecting terms,

$$n(>S) = \frac{1}{3\sqrt{4\pi}S^{3/2}}\int_0^\infty P^{3/2}N(P)dP \times\left[1-\frac{3}{4}(5+\beta+2\alpha)\left(\frac{PH_0^2}{4\pi S}\right)^{1/2}+\ldots\right] \tag{1.11.22}$$

We see that $n(>S)$ has a term with the familiar $S^{-3/2}$ dependence, plus a correction proportional to S^{-2} with a coefficient proportional to $5+\beta+2\alpha$. It is noteworthy that this coefficient is independent of q_0 or K. For the standard cosmology with no evolution of sources $\beta=-3$, and we have mentioned that $\alpha\approx 0.75$, so $5+\beta+2\alpha=3.5$. Although the precise value is uncertain, this coefficient is definitely positive, which means that for faint sources $n(>S)$ should fall off more slowly than $S^{-3/2}$. This is definitely not what is observed. [10] It has been known for many years that for $S>5\times10^{-26}\mathrm{Wm}^{-2}/\mathrm{Hz}$, the source count function $N(>S)$ falls off more rapidly than $S^{-3/2}$. The conclusion is inevitable that the number of radio sources per co-moving volume is decreasing, with $\beta<-6.5$. Radio source counts are useful in studying this evolution, but not for measuring cosmological parameters.

On the other hand, for the steady state cosmology (discussed in Section 1.5) we have $\beta=0$, so the coefficient $5+\beta+2\alpha\approx 6.5$, and the predicted number count $N(>S)$ decreases even more slowly with S, making the disagreement with experiment even worse than for the standard cosmology with no evolution of sources. Here it is not possible to save the situation by appealing to evolution, because the essence of the steady state model is that on the average there is no evolution. This observation

discredited the steady state model even before the discovery of the cosmic microwave radiation background.

10.1.14 1.12 Quintessence

So far, we have taken into account only non-relativistic matter, radiation, and a constant vacuum energy in calculating the rate of expansion of the universe. It appears that the vacuum energy is not only much smaller than would be expected from order-of-magnitude estimates based on the quantum theory of fields, but is only a few times greater than the present matter density. This has led to a widespread speculation that the vacuum energy is not in fact constant; it may now be small because the universe is old. A time-varying vacuum energy is sometimes called quintessence. [1]

The natural way to introduce a varying vacuum energy is to assume the existence of one or more scalar fields, on which the vacuum energy depends, and whose cosmic expectation values change with time. Scalar fields of this sort play a crucial part in the modern theory of weak

[4][7][10] For a list of major radio source surveys, and references to the original literature, see G&C, Sec. 14.8.

and electromagnetic interactions, and are also introduced in theories of inflation, as discussed in Chapters 4 and 10.

For simplicity, let us consider a single real scalar field $\varphi(\mathbf{x}, t)$. We will be concerned here with fields that are vary little on elementary particle spacetime scales, so the action of these field is taken to have a minimum number of spacetime derivatives:

$$I_\varphi = -\int d^4x \sqrt{-\operatorname{Det} g}\left[\frac{1}{2} g^{\lambda\kappa}\frac{\partial\varphi}{\partial x^\lambda}\frac{\partial\varphi}{\partial x^\kappa} + V(\varphi)\right] \tag{1.12.1}$$

with an unspecified potential function $V(\varphi)$. We are interested here in the case of a Robertson-Walker metric, and a scalar field that depends only on time, not position. In this case the formulas (B.66) and (B.67) for the scalar field energy density and pressure become

$$\rho_\varphi = \frac{1}{2}\dot{\varphi}^2 + V(\varphi) \tag{1.12.2}$$

$$p_\varphi = \frac{1}{2}\dot{\varphi}^2 - V(\varphi). \tag{1.12.3}$$

It follows immediately that $(1 + w)\rho_\varphi \geq 0$, where $w \equiv p_\varphi/\rho_\varphi$, so as long as $\rho_\varphi \geq 0$ this model has $w \geq -1$, and the phantom energy disaster discussed in Section 1.6 does not occur.

The equation (1.1.32) of energy conservation here reads

$$\ddot{\varphi} + 3H\dot{\varphi} + V'(\varphi) = 0, \tag{1.12.4}$$

(where as usual $H(t) \equiv \dot{a}(t)/a(t)$), which is the same as the field equation derived from the action (1.12.1). This is the equation of a particle of unit mass with one-dimensional coordinate φ, moving in a potential $V(\varphi)$ with a frictional force $-3H\dot{\varphi}$. The field will run toward lower values of $V(\varphi)$, finally coming to rest if it can reach any field value where $V(\varphi)$ is at least a local minimum. Unfortunately, we do not know any reason why the value of $V(\varphi)$ where it is stationary should be small.

Nevertheless, there are potentials that have some attractive properties once we adjust an additive constant in the potential to make them vanish at their stationary point. The original and simplest example is provided by a potential [2]

$$V(\varphi) = M^{4+\alpha}\varphi^{-\alpha} \tag{1.12.5}$$

where α is positive but otherwise arbitrary, and M is a constant with the units of mass (taking $\hbar = c = 1$), which gives $V(\varphi)$ the dimensions of an energy density. There is no special reason to believe that the potential has this form, and in particular there is no known reason for excluding an additive constant (including effects of quantum fluctuations in all other fields), which would give the potential a non-zero value at its stationary point, at $\varphi = \infty$. Nevertheless, it may be illuminating to work out the consequences of this one specific model of quintessence.

For any potential it is necessary to assume that at sufficiently early times ρ_φ was much less than the energy density ρ_R of radiation because, as we will see in Section 3.2, any appreciable increase in the energy density at the time of cosmological nucleosynthesis would lead to a helium abundance exceeding what is observed. At these early times the energy density of radiation (including particles like neutrinos with masses less than $k_B T$) is also greater than that of non-relativistic matter, so Eq. (1.5.34) gives $a(t) \propto t^{1/2}$, and therefore $H = 1/2t$. The field equation (1.12.4) with potential (1.12.5) then reads

[471] For reviews with references to the original literature, see B. Ratra and P. J. E. Peebles, Rev. Mod. Phys. 75, 559 (2003); E. V. Linder, 0704.2064.

$$\ddot{\varphi} + \frac{3}{2t}\dot{\varphi} - \alpha M^{4+\alpha}\varphi^{-\alpha-1} = 0 \tag{1.12.6}$$

This has a solution

$$\varphi = \left(\frac{\alpha(2+\alpha)^2 M^{4+\alpha} t^2}{6+\alpha}\right)^{\frac{1}{2+\alpha}} \tag{1.12.7}$$

Both $\dot{\varphi}^2$ and $V(\varphi)$ then go as $t^{-2\alpha/(2+\alpha)}$, and therefore at very early times ρ_φ must have been less than ρ_R, which goes as t^{-2}. This solution is not unique, but it is an attractor, in the sense that any other solution that comes close to it will approach it as t increases. (To see this, note that a small perturbation $\delta\varphi$ of the solution (1.12.7) will satisfy
$0 = \delta\ddot{\varphi} + \frac{3}{2t}\delta\dot{\varphi} + \alpha(1+\alpha)M^{4+\alpha}\varphi^{-\alpha-2}\delta\varphi = \delta\ddot{\varphi} + \frac{3}{2t}\delta\dot{\varphi} + \frac{(6+\alpha)(1+\alpha)}{(2+\alpha)^2 t^2}\delta\varphi$.
This has two independent solutions of the form

$$\delta\varphi \propto t^\gamma, \quad \gamma = -\frac{1}{4} \pm \sqrt{\frac{1}{16} - \frac{(6+\alpha)(1+\alpha)}{(2+\alpha)^2}}.$$

The square root is imaginary for $\alpha > 0$, so both solutions for $\delta\varphi$ decay as $t^{-1/4}$ for increasing t, while φ itself is increasing.) For this reason, the particular solution of Eq. (1.12.6) that goes as Eq. (1.12.7) for $t \to 0$ is known as the tracker solution. There is no particular physical reason to require that the initial conditions for the scalar field are such that the scalar field has approached the tracker solution by the present moment (the set of such initial conditions is called the "basin of attraction"), but since this requirement would make the present evolution of the scalar field insensitive to the initial conditions, it has the practical advantage of providing a model of quintessence with just two free parameters: M and α.

Nothing much changes when the radiation energy density drops below the energy density of non-relativistic matter. The tracker solution for the scalar field continues to grow as $t^{2/(2+\alpha)}$ (though with a different constant factor), so $\dot{\varphi}^2$ and $V(\varphi)$ continue to fall as $t^{-2\alpha/(2+\alpha)}$. But ρ_M and ρ_R are decreasing faster, like t^{-2} and $t^{-8/3}$, respectively, so eventually ρ_M and ρ_R will fall below ρ_φ. It is interesting that the value of φ where ρ_φ becomes equal to ρ_M is independent of the unknown constant M. When the expansion is dominated by matter ρ_M is given by Eq. (1.5.31) as $1/6\pi G t^2$, while (1.1.2), (1.12.5) and (1.12.7) give $\rho_\varphi \approx M^{2(4+\alpha)/(2+\alpha)} t^{-2\alpha/(2+\alpha)}$, so the time t_c at which $\rho_\varphi = \rho_M$ is of order

$$t_c \approx M^{-(4+\alpha)/2} G^{-(2+\alpha)/4} \tag{1.12.8}$$

Using this in Eq. (1.12.7) then gives

$$\varphi(t_c) \approx G^{-1/2} \tag{1.12.9}$$

Once ρ_M falls well below ρ_φ, the equation of motion of $\varphi(t)$ becomes

$$\ddot{\varphi} + \sqrt{24\pi G\rho_\varphi}\,\dot{\varphi} - \alpha M^{4+\alpha}\varphi^{-\alpha-1} = 0 \tag{1.12.10}$$

with ρ_φ given by Eq. (1.12.2). The tracker solution in this era has a complicated time dependence, but it becomes simple again at sufficiently late times, times that may be later than the present. We can guess that the damping term proportional to $\dot{\varphi}$ in this equation will eventually slow the growth of φ, so that $\dot{\varphi}^2$ will become less than $V(\varphi)$, and also guess that the

[472] P. J. E. Peebles and B. Ratra, Astrophys. J. 325, L17 (1988); B. Ratra and P. J. E. Peebles, Phys. Rev. D 37, 3406 (1988); C. Wetterich, Nucl. Phys. B302, 668 (1988). Quintessence models with this potential were intensively studied by I. Zlatev, L. Wang, and P. J. Steinhardt, Phys. Rev. Lett. 82, 896 (1999); P. J. Steinhardt, L. Wang, and I. Zlatev, Phys. Rev. D 59, 123504 (1999).

inertial term proportional to $\ddot{\varphi}$ will become negligible compared to the damping and potential terms. (Similar "slow roll" conditions will play an important role in the theory of inflation, described in Chapters 4 and 10.) Equation (1.12.10) then becomes

$$\sqrt{24\pi G M^{4+\alpha}\varphi^{-\alpha}}\dot{\varphi} = \alpha M^{4+\alpha}\varphi^{-\alpha-1}$$

and so

$$\dot{\varphi} = \frac{\alpha M^{2+\alpha/2}\varphi^{-\alpha/2-1}}{\sqrt{24\pi G}} \tag{1.12.11}$$

The solution is

$$\varphi = M\left(\frac{\alpha(2+\alpha/2)t}{\sqrt{24\pi G}}\right)^{1/(2+\alpha/2)} \tag{1.12.12}$$

(In general this involves a redefinition of the zero of time, to avoid a possible integration constant that might be added to t.) We can now check the approximations used in deriving Eq. (1.12.11), of which this is the solution. From Eq. (1.12.12) we see that $\dot{\varphi}^2 \propto t^{-(2+\alpha)/(2+\alpha/2)}$ while $V(\varphi) \propto t^{-\alpha/(2+\alpha/2)}$, so the kinetic energy term in Eq. (1.12.2) does become small compared with the potential term at late times. Also, $\ddot{\varphi} \propto t^{-(3+\alpha)/(2+\alpha/2)}$ while $V'(\varphi) \propto t^{-(1+\alpha)/(2+\alpha/2)}$, so the inertial term in Eq. (1.12.10) does become small compared with the potential term at late times. Eq. (1.12.12) is therefore a valid asymptotic solution of Eq. (1.12.10) for $t \to \infty$. Numerical calculations show that it is not only a solution for $t \to \infty$; it is the asymptotic form approached for $t \to \infty$ by the tracker solution.

With $\rho_\varphi \propto t^{-\alpha/(2+\alpha/2)}$ dominating the expansion rate at late times, we have $\dot{a}/a \propto t^{-\alpha/2(2+\alpha/2)}$, so

$$\ln a \propto t^{2/(2+\alpha/2)} \tag{1.12.13}$$

This is a similar but less rapid growth of a than would be produced by a cosmological constant, for which $\ln a \propto t$. The difference between the deceleration parameter q_0 and the value -1 for an expansion dominated by a cosmological constant vanishes as $t^{-(2+\alpha)/(2+\alpha/2)}$. Note that the radiation and matter densities decrease as $1/a^4$ and $1/a^3$ respectively, and the curvature decreases as $1/a^2$, all of which have a much faster rate of decrease with time than the power-law decrease of ρ_φ, so the expansion rate is indeed dominated by ρ_φ at late times, justifying the derivation of Eq. (1.12.10).

We have found that, at least for a range of initial conditions, the potential (1.12.5) leads to an expansion that is dominated by radiation and then matter at early times, but becomes dominated by the scalar field energy at late times. But to get agreement with observation it is necessary arbitrarily to exclude a large constant term that might be added to (1.12.5), and also to adjust the value of M to make the critical time (1.12.8) at which the values of ρ_φ and ρ_M cross be close to the present moment $t_0 \approx 1/H_0$. Specifically, Eq. (1.12.8) shows that we need the constant factor in $V(\varphi)$ to take the value

$$M^{4+\alpha} \approx G^{-1-\alpha/2}H_0^2. \tag{1.12.14}$$

There is no known reason why this should be the case.

Several groups of observers are now planning programs to discover whether the vacuum energy density is constant, as in the case of a cosmological constant, or changing with time. In such programs, one would compare the observed luminosity distance (or angular diameter distance) with a formula obtained by replacing the term Ω_Λ in the argument of the square root in Eq. (1.5.45) with a time-varying dark energy term. These observations will not actually measure

the value w_0 of w at the present time, much less the present time derivatives $\dot{w}_0, \ddot{w}_0$, etc., because for that purpose it would be necessary to have extremely precise measurements of the luminosity distance or angular-diameter distance for small redshifts. Instead, measurements will be made with only moderate precision, but over a fairly large range of redshifts. To compare such measurements with theory, one needs a model of the time-variation of the dark energy. One model is simply to assume that w is constant, or perhaps varying linearly with time or redshift, but there is no physical model that entails such behavior. [3] It seems preferable to compare observation with the model of a scalar field rolling down a potential, which (whatever reservations may have about its naturalness) at least provides a physically possible model of varying dark

energy. [4] Because these observations are difficult, it pays to adopt scalar field models with just two parameters, which can if we like be expressed in terms of $\Omega_\Lambda = 1 - \Omega_M$ (assuming flatness and neglecting the radiation energy density) and w_0.

One possibility is to suppose that over the latest e-folding of cosmic expansion the scalar field φ has taken values for which $V(\varphi)$ is only slowly varying. If $V(\varphi)$ were constant, we would have a constant vacuum energy, with $w = -1$, and the only parameter to measure would be Ω_V. For a two-parameter fit, we can take $V(\varphi)$ to vary linearly with φ :

$$V(\varphi) = V_0 + (\varphi - \varphi_0)\, V_0' \tag{1.12.15}$$

This is valid if the fractional change in $V'(\varphi)$ in a time interval of order $1/H_0$ is small; that is, if $|V_0''\dot{\varphi}_0| \ll H_0\, |V_0'|$.

The field equation (1.12.4) for $\varphi(t)$ can be put in a convenient dimensionless form by replacing the dependent variable t and independent variable φ with dimensionless variables x and ω, defined by

$$x \equiv H_0\sqrt{\Omega_M}t, \quad \omega \equiv \frac{8\pi G V(\varphi)}{3\Omega_M H_0^2} \tag{1.12.16}$$

Because V is linear in φ, we have

$$\dot{\varphi} = \frac{3\Omega_M H_0^2 \dot{\omega}}{8\pi G V_0'} = \frac{3\Omega_M^{3/2} H_0^3}{8\pi G V_0'}\frac{d\omega}{dx}$$

Then Eq. (1.12.4) becomes

$$\frac{d^2\omega}{dx^2} + 3\mathcal{H}\frac{d\omega}{dx} + \lambda = 0 \tag{1.12.17}$$

where λ is the dimensionless parameter

$$\lambda \equiv \frac{8\pi G V_0'^2}{3H_0^4\Omega_M^2} \tag{1.12.18}$$

and $\mathcal{H}$ is a function of ω and $d\omega/dx$:

$$\mathcal{H} \equiv \frac{H}{H_0\sqrt{\Omega_M}} = \sqrt{(1+z)^3 + \omega + \frac{1}{2\lambda}\left(\frac{d\omega}{dx}\right)^2} \tag{1.12.19}$$

We will also need the differential equation for the redshift:

[473] Other assumptions about the form of w as a function of redshift that can mimic scalar field models have been considered by J. Weller and A. Albrecht, Phys. Rev. D 65, 103512 (2002); E. V. Linder, Phys. Rev. Lett. 90, 091301 (2003) [astro-ph/0208512].

[474] This approach is followed by D. Huterer and H. V. Peiris, Phys. Rev. D 75, 083502 (2007) [astroph/0610427]; R. Crittenden, E. Majerotto, and F. Piazza, astro-ph/0702003.

$$\frac{dz}{dx} = -\mathcal{H}(1+z) \tag{1.12.20}$$

In general, even if we wrote all derivatives with respect to x in terms of derivatives with respect to z, to solve these equations we would need initial conditions for ω and $d\omega/dz$ at some initial z, which with λ would give a threeparameter set of solutions. However, assuming that for large redshift the energy density is dominated by matter rather than vacuum energy (which as we shall see is the case), the derivative $d\omega/dx$ sufficiently late in the matterdominated era becomes quite insensitive to initial conditions. [5] For $z \gg 1$, Eq. (1.12.19) gives

$$\mathcal{H} \to (1+z)^{3/2} \tag{1.12.21}$$

and (1.12.17) and (1.12.20) then have the solution

$$1 + z \to \left(\frac{3x}{2}\right)^{-2/3}, \quad \frac{d\omega}{dx} \to -\frac{\lambda x}{3} \tag{1.12.22}$$

(An integration constant in the solution for z has been absorbed into the definition of x, setting the zero of time. An integration constant in the solution for $d\omega/dx$ has been dropped, because it gives a term in $d\omega/dx$ that dies away with increasing time as $x^{-2} \propto t^{-2}$.) The free parameters in our solution are then λ, together with the value of ω at some arbitrary initial value x_1 of x, taken sufficiently small so that at x_1 the energy density is dominated by matter rather than vacuum energy. (Note that the constant V_0 appears nowhere in these equations; it contributes a term to $\omega(x_1)$, but there is no need to isolate this term.) One must adopt various trial values of λ and $\omega(x_1)$; use Eq. (1.12.22) to calculate $1+z$ and $d\omega/dx$ at $x = x_1$; with these initial conditions, integrate the differential equations (1.12.17) and (1.12.20) numerically from x_1 to a value x_0 where $z = 0$; and then if we like calculate the values of $\Omega_V = 1 - \Omega_M$ and the present value w_0 of the ratio p_φ/ρ_φ for this particular solution, [6] using

$$\frac{\Omega_\Lambda}{\Omega_M} = \omega(x_0) + \frac{1}{2\lambda}\left(\frac{d\omega}{dx}\right)^2_{x=x_0}, \quad w_0 = \frac{(d\omega/dx)^2_{x=x_0} - 2\lambda\omega(x_0)}{(d\omega/dx)^2_{x=x_0} + 2\lambda\omega(x_0)} \tag{1.12.23}$$

The ratio of the dark energy at a given time to its value at the present is

$$\xi \equiv \frac{\rho_V(t)}{\rho_V(t_0)} = \frac{(d\omega(x)/dx)^2 + 2\lambda\omega(x)}{(d\omega/dx)^2_{x=x_0} + 2\lambda\omega(x_0)} \tag{1.12.24}$$

For instance, if we take $\Omega_\Lambda = 1 - \Omega_M = 0.76$ and $w_0 = -0.777$, the ratio ξ of the dark energy density to its value at present rises to 1.273 at $z = 1$ and to 1.340 at infinite redshift. [7] The leveling off of $\xi(z)$ for large z occurs because the growth of the matter density for increasing redshift makes the expansion rate grow, so that the friction term $3H\dot{\varphi}$ in Eq. (1.12.4) freezes the value of the scalar field at early times.

It should not be thought that the leveling off of the dark energy for large z for the potential (1.12.15) means that in analyzing dark energy observations with this potential one must give up the idea motivating theories of quintessence, that the vacuum energy is now small because the universe is old. In fact, for the potential $V(\varphi) \propto \varphi^{-\alpha}$, for typical initial conditions the quintessence energy drops at first precipitously, and then levels off while the scalar field rolls slowly down the potential until the field approaches the tracker solution, with the tracker

[475] R. Cahn, private communication. Cahn has also shown that the approximation of neglecting the second derivative term in the field equation does not work well in this context.

[6] As already mentioned, with models of this sort one can only have $w_0 > -1$. To compare the case $w_0 < -1$ with observation, it is necessary to adopt a model with the opposite sign for the derivative term in the action (1.12.1). The analysis given here can then be applied, with only obvious sign changes here and there.

solution not reached by the present time if α is small. [8] The condition $|V_0''\dot{\varphi}_0| \ll H_0\,|V_0'|$ for treating this potential as linear over a time of order $1/H_0$ is satisfied if $\alpha(1+\alpha)\varphi_0^{-2} \ll 8\pi G$, which in light of Eq. (1.12.9) is likely to be satisfied if $\alpha < 1$.

Another possible two-parameter model is provided by the same potential, $V(\varphi) \propto \varphi^{-\alpha}$, but now under the assumption that the tracker solution is reached by some early time (say, for $z \leq 10$) in the matter-dominated era. With this assumption the observable history of dark energy is insensitive to initial conditions, so the model has just two parameters: M and α. The equations of this model can be put in dimensionless form by writing the coupling constant of this potential in terms of a dimensionless parameter β as

$$M^{4+\alpha} \equiv \beta\Omega_M H_0^2 (8\pi G)^{-1-\alpha/2} \tag{1.12.25}$$

and replacing the dependent variable t and independent variable φ with dimensionless variables x and f, defined by

$$t \equiv x/H_0\sqrt{\Omega_M}, \quad \varphi(t) \equiv f(x)/\sqrt{8\pi G} \tag{1.12.26}$$

The field equation (1.12.4) (with no slow roll approximation) in the era dominated by matter and vacuum energy then takes the form

$$\frac{d^2 f}{dx^2} + 3\mathcal{H}\frac{df}{dx} - \alpha\beta f^{-\alpha-1} = 0 \tag{1.12.27}$$

where

$$\mathcal{H} \equiv H/\sqrt{\Omega_M}H_0 = \sqrt{\frac{1}{6}\left(\frac{df}{dx}\right)^2 + \frac{\beta}{3}f^{-\alpha} + (1+z)^3} \tag{1.12.28}$$

in which we also need

$$\frac{dz}{dx} = -\mathcal{H}(1+z) \tag{1.12.29}$$

Because the large z solution (1.12.7) is an attractor, the initial conditions introduce no new free parameters; in terms of these dimensionless variables, the initial conditions are that, for $x \to 0$,

$$f \to \left[\frac{\alpha\beta(\alpha+2)^2 x^2}{2(\alpha+4)}\right]^{1/(\alpha+2)}, \quad 1+z \to \left(\frac{2}{3x}\right)^{2/3} \tag{1.12.30}$$

We need to integrate the equations (1.12.27) and (1.12.29) from some small x (say, $x = 0.01$) to a value x_0 at which $z = 0$, with the initial conditions (1.12.30), and then evaluate $\Omega_M = 1 - \Omega_\Lambda$ from the condition that $\mathcal{H}(x_0) = 1/\sqrt{\Omega_M}$. We can also evaluate the present value w_0 of $w \equiv p_\varphi/\rho_\varphi$ from the formula

$$w_0 = \frac{f'^2(x_0)\, f(x_0)^\alpha/2\beta - 1}{f'^2(x_0)\, f(x_0)^\alpha/2\beta + 1} \tag{1.12.31}$$

and then replace the parameters α and β with Ω_M and w_0. For instance, if we arbitrarily take $\alpha = 1$, then to get the realistic value $\Omega_M = 0.24$ we must take $\beta = 9.93$, in which case $w_0 = -0.777$. Of course, we can get any other values of w_0 greater than -1 by choosing different values of α and re-adjusting β to give the same value of Ω_M (though for small α, the range

[477] Numerical results for various values of redshift are given in Table 1.1. These results for the linear potential were calculated by R. Cahn, private communication.

[8] Steinhardt, Wang, and Zlatev, ref. 2.

of initial conditions that allow the tracker solution to be reached well before the present is relatively small.) For instance, for $\alpha = 1/2$ we must take $\beta = 7.82$ to have $\Omega_M = 0.24$, and in this case we calculate that $w_0 = -0.87$. (For the case $w < -1$, see footnote 6.) The ratios of dark energy to its value at present calculated in this way for $\Omega_M = 0.24$ and $w_0 = -0.777$ are shown in Table 1.1, along with the values calculated with the same choice of Ω_M and w_0 for both the case of constant w and for the linear potential (1.12.15). The tracker and linear models evidently represent opposite extreme assumptions about the time-dependence of dark

Table 1.1: Ratio of dark energy to its present value, for the tracker solution with the potential (1.12.5), and for the linear potential (1.12.15), calculated for $\Omega_{M}=1-\Omega_{\Lambda}=0.24$ and $w_0 = -0.777$, compared with the results for a constant $w = -0.777$.

z	tracker	linear	constant w
0	1	1	1
0.1	1.067	1.062	1.066
0.5	1.347	1.200	1.312
1	1.712	1.273	1.590
2	2.469	1.318	2.086
3	3.224	1.331	2.528
$\gg 1$	$\gg 1$	1.340	$\gg 1$

energy, but both are better motivated physically than the assumption of a constant w.

10.1.15 1.13 Horizons

Modern cosmological theories can exhibit horizons of two different types, which limit the distances at which past events can be observed or at which it will ever be possible to observe future events. These are called by Rindler [1] particle horizons and event horizons, respectively.

According to Eq. (1.2.2), if the big bang started at a time $t = 0$, then the greatest value $r_{\rm max}\,(t)$ of the Robertson-Walker radial coordinate from which an observer at time t will be able to receive signals traveling at the speed of light is given by the condition

$$\int_0^t \frac{dt'}{a\,(t')} = \int_0^{r_{\rm max}(t)} \frac{dr}{\sqrt{1-Kr^2}} \tag{1.13.1}$$

Thus there is a particle horizon unless the integral $\int dt/a(t)$ does not converge at $t = 0$. It does converge in conventional cosmological theories; whatever the contribution of matter or vacuum energy at the present, it is likely that the energy density will be dominated by radiation at early times, in which case $a(t) \propto t^{1/2}$, and the integral converges. The proper distance of the horizon is given by Eq. (1.1.15) and (1.13.1) as

$$d_{\rm max}(t) = a(t)\int_0^{r_{\rm max}(t)} \frac{dr}{\sqrt{1-Kr^2}} = a(t)\int_0^t \frac{dt'}{a\,(t')} \tag{1.13.2}$$

For instance, during the radiation-dominated era $a(t) \propto t^{1/2}$, so $d_{\rm max}\,(t) = 2t = 1/H$. Well into the matter-dominated era most of the integral over time in Eq. (1.13.1) comes from a time when $a \propto t^{2/3}$, so that $d_{\rm max}(t) \simeq 3t = 2/H$. At present most of the integral over t' comes

[471] W. Rindler, Mon. Not. Roy. Astron. Soc. 116, 663 (1956).

from a period when the expansion is dominated by matter and the vacuum energy, and perhaps curvature as well. According to Eq. (1.5.41), the particle horizon distance at present is

$$d_{\max}(t_0) = \frac{1}{H_0}\int_0^1 \frac{dx}{x^2\sqrt{\Omega_\Lambda + \Omega_K x^{-2} + \Omega_M x^{-3}}} \tag{1.13.3}$$

We will see in Chapter 4 that there may have been a time before the radiationdominated era in which there was nothing in the universe but vacuum energy, in which case the particle horizon distance would actually be infinite. But as far as telescopic observations are concerned, Eq. (1.13.3) gives the proper distance beyond which we cannot now see.

Just as there are past events that we cannot now see, there may be events that we never will see. Again returning to Eq. (1.2.2), if the universe re-collapses at a time T, then the greatest value $r_{\rm MAX}$ of r from which an observer will be able to receive signals traveling at the speed of light emitted at any time later than t is given by the condition

$$\int_t^T \frac{dt'}{a(t')} = \int_0^{r_{\rm MAX}(t)} \frac{dr}{\sqrt{1-Kr^2}} \tag{1.13.4}$$

Even if the future is infinite, if the integral $\int dt/a(t)$ converges at $t = \infty$ there will be an event horizon given by

$$\int_t^\infty \frac{dt'}{a(t')} = \int_0^{r_{\rm MAX}(t)} \frac{dr}{\sqrt{1-Kr^2}} \tag{1.13.5}$$

Since co-moving sources are labeled with a fixed value of r, the condition $r < r_{\rm MAX}$ limits the events occurring at time t that we can ever observe. In the case where the universe does not recollapse, the proper distance to the event horizon is given by

$$d_{\rm MAX}(t) = a(t)\int_0^{r_{\rm MAX}(t)} \frac{dr}{\sqrt{1-Kr^2}} = a(t)\int_t^\infty \frac{dt'}{a(t')} \tag{1.13.6}$$

In the absence of a cosmological constant, $a(t)$ grows like $t^{2/3}$, and the integral diverges, so that there is no event horizon. But with a cosmological constant $a(t)$ will eventually grow as $\exp(Ht)$ with $H = H_0\Omega_\Lambda^{1/2}$ constant, and there really is an event horizon, which approaches the value $d_{\rm MAX}(\infty) = 1/H$. As time passes all sources of light outside our gravitationally bound Local Group will move beyond this distance, and become unobservable. The same is true for the quintessence theory described in the previous section. In that case $a(t)$ eventually grows as $\exp(\text{ constant } \times t^{2/(2+\alpha/2)})$, so for any $\alpha \geq 0$ the integral (1.13.6) again converges.

If a source is at a radial coordinate r in a Robertson-Walker coordinate system based on us, then we are at a radial coordinate r in a RobertsonWalker coordinate system based on the source. Hence Eq. (1.13.4) or (1.13.5) also gives the greatest radial coordinate to which, starting at time t, we will ever be able to travel.

10.2 The Cosmic Microwave Radiation Background

Before the mid-1960s by far the greatest part of our information about the structure and evolution of the universe came from observations of the redshifts and distances of distant galaxies, discussed in the previous chapter. In 1965 a nearly isotropic background of microwave radiation was discovered, which has provided a wealth of new cosmological data. After reviewing the expectations and discovery of this radiation, this chapter will explore some of its implications. We will only be able to give a first look at the anisotropies in this radiation in this chapter. In Chapter 7 we will return to this very important topic, applying the analysis of the evolution of cosmological perturbations presented in Chapters 5 and 6, and in Chapter 10 we will consider the origin of these perturbations in the very early universe.

10.2.1 2.1 Expectations and discovery of the microwave background

The work done by pressure in an expanding fluid uses heat energy drawn from the fluid. The universe is expanding, so we expect that in the past matter was hotter as well as denser than at present. If we look far enough backward in time we come to an era when it was too hot for electrons to be bound into atoms. At sufficiently early times the rapid collisions of photons with free electrons would have kept radiation in thermal equilibrium with the hot dense matter. The number density of photons in equilibrium with matter at temperature T at photon frequency between ν and $\nu + d\nu$ is given by the black-body spectrum:

$$n_T(v)dv = \frac{8\pi v^2 dv}{\exp\left(hv/k_{\mathcal{B}}T\right) - 1} \tag{2.1.1}$$

where h is the original Planck's constant (which first made its appearance in a formula equivalent to this one), and $k_{\mathcal{B}}$ is Boltzmann's constant. (Recall that we are using units with $c = 1$.)

As time passed, the matter became cooler and less dense, and eventually the radiation began a free expansion, but its spectrum has kept the same form. We can see this most easily under an extreme assumption, that there was a time t_L when radiation suddenly went from being in thermal equilibrium with matter to a free expansion. (The subscript L stands for "last scattering.") Under this assumption, a photon that has frequency v at some later time t when photons are traveling freely would have had frequency
$va(t)/a\left(t_L\right)$ at the time the radiation went out of equilibrium with matter, and so the number density at time t of photons with frequency between v and $v + dv$ would be

$$n(v,t)dv = \left(a\left(t_L\right)/a(t)\right)^3 n_{T(t_L)}\left(va(t)/a\left(t_L\right)\right) d\left(va(t)/a\left(t_L\right)\right) \tag{2.1.2}$$

with the factor $\left(a\left(t_L\right)/a(t)\right)^3$ arising from the dilution of photons due to the cosmic expansion. Using Eq. (2.1.1) in (2.1.2), we see that the redshift factors $a(t)/a\left(t_L\right)$ all cancel except in the exponential, so that the number density at time t is given by

$$n(v,t)dv = \frac{8\pi v^2 dv}{\exp\left(hv/k_{\mathcal{B}}T(t)\right) - 1} = n_{T(t)}(v)dv \tag{2.1.3}$$

where

$$T(t) = T\left(t_L\right) a\left(t_L\right)/a(t) \tag{2.1.4}$$

Thus the photon density has been given by the black-body form even after the photons went out of equilibrium with matter, but with a redshifted temperature (2.1.4).

This conclusion is obviously unchanged if the transition from opacity to transparency occupied a finite time interval, as long as the interactions of photons with matter during this interval are limited to elastic scattering processes in which photon frequencies are not changed. This is a very good approximation. We will see in Section 2.3 that the last interaction of photons with matter (until near the present) took place at a time when the cosmic temperature T was of order $3,000$ K, when by far the most important interaction was the elastic scattering of photons with electrons, in which the fractional shift of photon frequency was of order $k_{\mathcal{B}}T/m_e c^2 \approx 3 \times 10^{-7}$. In the following section we shall show that, because of the large photon entropy, even the small shift of photon frequency in elastic scattering and the relatively infrequent inelastic interactions of photons with hydrogen atoms had almost no effect on the photon spectrum.

It was George Gamow and his collaborators who first recognized in the late 1940s that the universe should now be filled with black-body radiation. [1] The first plausible estimate of the

present temperature of this radiation was

made in 1950 by Ralph Alpher and Robert Herman. [2] On the basis of considerations of cosmological nucleosynthesis, to be discussed in Section 3.2, they found a present temperature of 5 K . This work was largely forgotten in subsequent decades, until in 1965 a group at Princeton started to search for a cosmic radiation background left over from the early universe. They had only a rough idea of the temperature to be expected, based on a nucleosynthesis calculation of P. J. E. Peebles, which suggested a value of 10 K.[3] Before they could complete their experiment the radiation was discovered in a study of noise backgrounds in a radio telescope by Arno Penzias and Robert Wilson, [4] who published their work along with a companion article [5] by the Princeton group explaining its possible cosmological significance. [6]

Originally Penzias and Wilson could only report that the antenna temperature at a wavelength 7.5 cm was 3.5 ± 1.0 K, meaning that the intensity of the radiation at this one wavelength agreed with Eq. (2.1.1) for this temperature. This of course did not show that they were observing black-body radiation. Then Roll and Wilkinson [7] measured the radiation intensity at a wavelength of 3.2 cm , finding an antenna temperature of 3.0 ± 0.5 K, in agreement with what would be expected for black-body radiation at the temperature measured by Penzias and Wilson. In the following few years a large number of measurements were made by other radio astronomers at other wavelengths. These measurements also gave antenna temperatures at the wavelengths being studied around 3 K , with uncertainties that gradually improved to of order 0.2 K . But this also did not establish the black-body nature of the radiation, because these measurements were all at wavelengths greater than about 0.3 cm , where the black-body energy distribution $h\nu n_T(\nu)$ with $T \approx 3$ K has its maximum. For these long wavelengths the argument of the exponential is small, and Eq. (2.1.1) gives

$$h\nu n_T(\nu) \simeq 8\pi\nu^2 k_{\mathcal{B}} T, \tag{2.1.5}$$

This is the Rayleigh-Jeans formula of classical statistical mechanics, but it describes the long-wavelength distribution of radiant energy under various circumstances more general than black-body radiation. For instance, black-body radiation diluted by an expansion that preserves the energy of individual photons also has $h\nu n(\nu) \propto \nu^2$ for low frequencies. [8] To confirm that the cosmic microwave background radiation is really described by the black-body formula, it was necessary to see at least the beginning of the exponential fall-off of $n_T(\nu)$ for wavelengths shorter than about 0.3 cm .

This was difficult because the earth's atmosphere becomes increasingly opaque at wavelengths shorter than about 0.3 cm . However, there had been a measurement of the radiation temperature at a wavelength 0.264 cm in 1941, long before the discovery by Penzias and Wilson. In between the star ζ Oph and the earth there is a cloud of cold molecular gas, whose absorption of light produces dark lines in the spectrum of the star. In 1941 W. S. Adams, [9]

[471] G. Gamow, Phys. Rev. 70, 572 (1946); R. A. Alpher, H. A. Bethe, and G. Gamow, Phys. Rev. 73, 803 (1948); G. Gamow, Phys. Rev. 74, 505 (1948); R. A. Alpher and R. C. Herman, Nature 162, 774 (1948); R. A. Alpher, R. C. Herman, and G. Gamow, Phys. Rev. 74, 1198 (1948); ibid 75, 332A (1949); ibid 75, 701 (1949); G. Gamow, Rev. Mod. Phys. 21, 367 (1949); R. A. Alpher, Phys. Rev. 74, 1577 (1948); R. A. Alpher and R. C. Herman, Phys. Rev. 75, 1089 (1949).

[472] R. A. Alpher and R. C. Herman, Rev. Mod. Phys. 22, 153 (1950)

[3] This work was never published. According to A. Guth, The Inflationary Universe (Perseus Books, Reading, MA, 1997), Peebles' paper was rejected by The Physical Review, apparently because of the issue of the credit to be given to earlier work by R. Alpher, G. Gamow, and R. Herman. This earlier work and the subsequent work of Peebles and others is briefly described here in Section 3.2.

[4] A. A. Penzias and R. W. Wilson, Astrophys. J. 142, 419 (1965).

[5] R. H. Dicke, P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson, Astrophys. J. 142, 414 (1965).

[6] For a more detailed history of these developments, see A. Guth, op. cit., and S. Weinberg, The First Three Minutes (Basic Books, New York, 1977; second edition 1993).

[7] P. G. Roll and D. T. Wilkinson, Phys. Rev. Lett. 16, 405 (1966).

following a suggestion of Andrew McKellar, found two dark lines in the spectrum of ζ Oph that could be identified as due to absorption of light by cyanogen (CN) in the molecular cloud. The first line, observed at a wavelength of $3,874.62$Å, could be attributed to absorption of light from the CN ground state, with rotational angular momentum $J = 0$, leading to the component of the first vibrationally excited state with $J = 1$. But the second line, at $3,874.00$Å, represented absorption from the $J = 1$ rotationally excited vibrational ground state, leading to the $J = 2$ component of the first vibrationally excited state. [10] From this, McKellar concluded [11] that a fraction of the CN molecules in the cloud were in the first excited rotational component of the vibrational ground state, which is above the true $J = 0$ ground state by an energy $hc/(0.264$ cm). and from this fraction he estimated an equivalent molecular temperature of 2.3 K. Of course, he did not know that the CN molecules were being excited by radiation, much less by black-body radiation. After the discovery by Penzias and Wilson several astrophysicists [12] independently noted that the old Adams-McKellar result could be explained by radiation with a

black-body temperature at wavelength 0.264 cm in the neighborhood of 3 K . Theoretical analysis showed that nothing else could explain the excitation of this rotational state. [13] This interpretation was then borne out by continuing observations on this and other absorption lines in CN as well as CH and CH^+in the spectrum of ζOph and other stars. [14]

The black-body spectrum of the cosmic microwave radiation background began to be established by balloon-borne and rocket-borne observations above the earth's atmosphere at wavelengths below 0.3 cm . For some years there were indications of an excess over the black-body formula at these short wavelengths. It was clearly necessary to do these observations from space, but this is difficult; to measure the absolute value of the microwave radiation intensity it is necessary to compare the radiation received from space with that emitted by a "cold load" of liquid helium, which rapidly evaporates. Finally, the Planck spectrum of the microwave background was settled in the 1990s by observations with the FIRAS radiometer carried by the Cosmic Background Explorer Satellite (COBE), launched in November 1989. [15] When a slide showing the agreement of the observed spectrum with the Planck black-body spectrum was shown by J. C. Mather at a meeting of the American Astronomical Society in January 1990, it received a standing ovation. It was found that the background radiation has a nearly exact black-body spectrum in the wavelength range from 0.5 cm to 0.05 cm.[16] The comparison of observation with the black-body spectrum is shown in Figure 2.1. After six years of further analysis, the temperature was given as 2.725 ± 0.002 K (95% confidence). [17] Other observations at wavelengths between 3 cm and 75 cm and at 0.03 cm are all consistent with a Planck distribution at this temperature. [18]

The energy density in this radiation is given by

$$\int_0^\infty h\nu n(\nu)d\nu = a_{\mathcal{B}}T^4 \tag{2.1.6}$$

where $a_{\mathcal{B}}$ is the radiation energy constant; in c.g.s. units,

[478] The sunlight falling on the earth's surface provides a pretty good example of dilute black-body radiation; it is described by the Planck formula (2.1.1), with $T \approx 6,000$ K the temperature of the sun's surface, but with the right-hand side of Eq. (2.1.1) multiplied by a factor $(R_\odot/r)^2$, where $R_\odot$ is the radius of the sun and r is the distance from the sun to the earth.

[9] W. S. Adams, Astrophys. J. 93, 11 (1941)

[10] Today the wavelengths of these two lines are more accurately known to be $3,874.608$ and $3,873.998$Å. There is another line at 3875.763Å, produced by transitions from the $J = 1$ rotationally excited vibrational ground state to the $J = 0$ component of the first vibrationally excited state.

[11] A. McKellar, Publs. Dominion Astrophys. Observatory (Victoria, B.C.) 7, 251 (1941).

[12] G. Field, G. H. Herbig, and J. L. Hitchcock, Astron. J. 71, 161 (1966); G. Field and J. L. Hitchcock, Phys. Rev. Lett. 16. 817 (1966); Astrophys. J. 146, 1 (1966); N. J. Woolf, quoted by P. Thaddeus and J. F. Clauser, Phys. Rev. Lett. 16, 819 (1966); I. S. Shklovsky, Astronomicheskii Tsircular No. 364 (1966).

$$a_{\mathcal{B}} = \frac{8\pi^5 k_{\mathcal{B}}^4}{15h^3c^3} = 7.56577(5) \times 10^{-15}\text{erg cm}^{-3}\text{deg}^{-4} \tag{2.1.7}$$

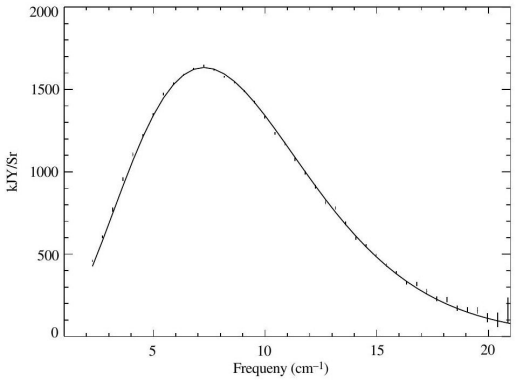


Figure 2.1: Comparison of the intensity of radiation observed with the FIRAS radiometer carried by COBE with a black-body spectrum with temperature 2.728 K , from D . J . Fixsen et al., Astrophys. J. 473, 576 (1996) [astro-ph/9605054]. The vertical axis gives the intensity in kiloJansky per steradian (one Jansky equals \$10ˆ{-23} \mathrm{erg} \mathrm{cm}ˆ{-2} \mathrm{˜s}ˆ{-1} \mathrm{˜Hz}ˆ{-1}\$); the horizontal axis gives the reciprocal wavelength in cm^{-1}. The 1σ experimental uncertainty in intensity is indicated by the tiny vertical bars; the uncertainty in wavelength is negligible.

Using $T = 2.725$ K, this gives an equivalent mass density (reverting to $c = 1$)

$$\rho_{\gamma 0} = a_{\mathcal{B}} T_{\gamma 0}^4 = 4.64 \times 10^{-34}\text{ g cm}^{-3}.$$

Taking the ratio of this with the critical density (1.5.28) gives

$$\Omega_\gamma \equiv \frac{\rho_{\gamma 0}}{\rho_{0\text{ crit}}} = 2.47 \times 10^{-5} h^{-2} \tag{2.1.8}$$

We will see in Section 3.1 that the photons are accompanied with neutrinos and antineutrinos of three different types, giving a total energy density in radiation (that is, in massless or nearly massless particles):

$$\rho_{R0} = \left[1 + 3\left(\frac{7}{8}\right)\left(\frac{4}{11}\right)^{4/3}\right]\rho_{\gamma 0} = 7.80 \times 10^{-34}\text{ g cm}^{-3}, \tag{2.1.9}$$

or in other words, using Eq. (1.5.28),

$$\Omega_R \equiv \frac{\rho_{R0}}{\rho_{0,\text{ crit}}} = 4.15 \times 10^{-5} h^{-2} \tag{2.1.10}$$

[4][7][13] Field et al., ref. 12; Thaddeus and Clauser, ref. 12.

[14] For a summary of this early work with references to the original literature, see G&C, Table 15.1.

[15] J. C. Mather et al., Astrophys. J. 354, 237 (1990).

[16] J. C. Mather et al., Astrophys. J. 420, 439 (1994).

[17] J. C. Mather, D. J. Fixsen, R. A. Shafer, C. Mosier, and D. T. Wilkinson, Astrophys. J. 512, 511 (1999). A 1999 review by G. F. Smoot, in Proc. 3K Cosmology Conf., eds. A. Melchiorri et al. [astroph/9902027], gave a temperature 2.7377 ± 0.0038 K (95% confidence), but the result of Mather et al. seems to be the one usually quoted.

[18] For a review, see G. Sironi et al., in Proc. Third Sakharov Conf. - Moscow 2002 [astro-ph/0301354].

We see that ρ_{R0} is much less than the critical mass density needed to give $K = 0$, and much less even than the mass density of ordinary matter seen in stars. It is for this reason that we have generally neglected Ω_R in calculating luminosity distances as a function of redshift.

On the other hand, even at present the number density of photons is relatively very large. Eq. (2.1.1) gives

$$\begin{aligned} n_{\gamma 0} &= \int_0^\infty \frac{8\pi v^2 dv}{\exp\left(hv/k_{\mathcal{B}}T\right) - 1} = \frac{30\zeta(3)}{\pi^4}\frac{a_{\mathcal{B}}T^3}{k_{\mathcal{B}}} \\ &= 0.3702\frac{a_{\mathcal{B}}T^3}{k_{\mathcal{B}}} = 20.28[T(\deg K)]^3 \text{ cm}^{-3} \end{aligned} \tag{2.1.11}$$

where $\zeta(3) = 1.202057\ldots$ For $T = 2.725$ K this gives a present number density

$$n_{\gamma 0} = 410 \text{ photons } /\text{cm}^3 \tag{2.1.12}$$

This is much larger than the present number density n_{B0} of nucleons, given by

$$n_{B0} = \frac{3\Omega_B H_0^2}{8\pi G m_N} = 1.123 \times 10^{-5}\Omega_B h^2 \text{ nucleons } /\text{cm}^3 \tag{2.1.13}$$

Both n_γ and n_B vary with time as $a^{-3}(t)$, so the ratio n_γ/n_B has been the same at least during the whole period that photons have been traveling freely.

There is an effect of the cosmic microwave background that has long been expected but has been difficult to observe. A cosmic ray proton of moderate energy striking a photon in the cosmic microwave background can only scatter the photon, a process whose rate is proportional to the square of the fine structure constant $\alpha \simeq 1/137$. However, if the proton has sufficiently high energy then it is also possible for the photon to be converted into a π meson in the reactions $\gamma + p \to \pi^0 + p$ or $\gamma + p \to \pi^+ + n$, processes whose rate is proportional to α, not α^2. Assuming that high energy cosmic rays come to us from outside our galaxy, we therefore expect a dip in the spectrum of cosmic ray protons at an energy where the cross section for these processes becomes appreciable. [19] Although some pions can be produced at lower energy, the effective threshold is at a value of the total energy W of the

initial proton and photon in the center-of-mass system equal to $m_\Delta = 1232$ MeV , the mass of the pion-nucleon resonance with spin-parity $3/2^+$and isospin 3/2. The center of mass energy is

$$\begin{aligned} W &= \left(\left(q + \sqrt{p^2 + m_p^2}\right)^2 - |\mathbf{q} + \mathbf{p}|^2\right)^{1/2} \\ &\simeq \left(2qp(1 - \cos\theta) + m_p^2\right)^{1/2}, \end{aligned} \tag{2.1.14}$$

where $\mathbf{q}$ and $\mathbf{p}$ are the initial photon and proton momenta (with $p \gg m_p \gg q$) and θ is the angle between them. The threshold condition that $W > m_\Delta$ thus requires that

$$qp(1 - \cos\theta) > m_\Delta^2 - m_p^2. \tag{2.1.15}$$

The typical energy of photons in black-body radiation at temperature $T_{\gamma 0} = 2.725$ K is $\rho_{\gamma 0}/n_{\gamma 0} \simeq 6 \times 10^{-4}$eV, while the largest value for $1 - \cos\theta$ is 2 , so the effective threshold is roughly at a proton energy

[4719] K. Greisen, Phys. Rev. Lett. 16, 748 (1966); G. T. Zatsepin and V. A. Kuzmin, Pis'ma Sh. Exsp. Teor. Fiz. 4, 114 (1966) [transl. Sov. Phys. JETP Lett. 4, 78 (1966)]; F. W. Stecker, Phys. Rev. Lett. 21, 1016 (1968).

$$p_{\rm threshold} \approx \frac{m_\Delta^2 - m_p^2}{2\rho_{\gamma 0}/n_{\gamma 0}} \simeq 10^{20}{\rm eV} \tag{2.1.16}$$

This effect is not easy to see. The flux of cosmic ray protons with energies between E and $E + dE$ goes roughly as $E^{-3}dE$, so there are few protons at these very high energies, roughly one per square kilometer per year above 10^{19}eV and 0.01 per square kilometer per year above 10^{20}eV. At these rates, direct observation is clearly impossible, and the cosmic rays have had to be studied indirectly by observation at ground level of the large showers of photons and charged particles that they produce. Also, there is a smooth distribution of photon energies and directions, so one is not looking for a sharp cut-off at 10^{20}eV, but rather for a dip below the E^{-3} curve at around this energy. [20] No such effect was observed by the Akeno Giant Air Shower Array, [21] but a subsequent analysis of this and other observations showed the effect. [22] More recently signs of a drop appeared in the High Resolution Fly's Eye experiment. [23] In 2006 this group announced the observation of a "sharp suppression" of the primary cosmic ray spectrum at an energy of 6×10^{19}GeV, just about where expected. [24]

10.2.2 2.2 The equilibrium era

As already remarked in the previous section, if we look back in time sufficiently far, we surely must come to an era when the temperature and density were sufficiently high so that radiation and matter were in thermal equilibrium. We will now consider this era, jumping over the intermediate time when radiation was going out of equilibrium with matter, which will be the subject of the next section.

As we saw in the previous section, in a free expansion of photons, the frequency distribution preserves the Planck black-body form (2.1.1), but with a temperature that falls as $1/a(t)$. On the other hand, in a free expansion of non-relativistic particles such as electrons or nuclei, the momentum distribution preserves the Maxwell-Boltzmann form, $n(p)dp \propto \exp\left(-p^2/2mk_{\mathcal{B}}T\right)$, but since (as shown in Eq. (1.1.23)) the momentum of any particle decreases as the universe expands with $p \propto 1/a$, the temperature of the Maxwell-Boltzmann distribution decreases as $1/a^2(t)$. So if radiation is in equilibrium with matter, who wins? Does the temperature decrease as $1/a(t)$, or $1/a^2(t)$, or something more complicated?

The issue is settled democratically, on the basis of one particle, one vote. Since there are so many more photons than electrons or nucleons, the photons win, and the temperature decreases almost exactly as $1/a(t)$.

We can see this in more detail by applying the second law of thermodynamics. In equilibrium both the entropy and the baryon number (that is, at temperatures $\ll 10^{13}$ K, the number of protons and neutrons) in any co-moving volume were constant, and so their ratio, the entropy per baryon, was also constant. It is convenient to write the entropy per baryon as $k_{\mathcal{B}}\sigma$, with $k_{\mathcal{B}}$ the Boltzmann constant and σ dimensionless. The second law of thermodynamics tells us that this satisfies

$$d\left(k_{\mathcal{B}}\sigma\right) = \frac{d\left(\epsilon/n_B\right) + pd\left(1/n_B\right)}{T}, \tag{2.2.1}$$

[4720] For instance, see I. F. M. Albuquerque and G. F. Smoot, Astroparticle Phys. 25, 375 (2006) [astro-ph/0504088].

[21] M. Takeda et al. Phys. Rev. Lett. 81, 1163 (1998).

[22] J. N. Bahcall and E. Waxman, Phys. Lett. B556, 1 (2003) [hep-ph/0206217].

[23] R. U. Abbasi et al., Phys. Rev. Lett. 92, 151101 (2004) [astro-ph/0208243]; Phys. Lett. B619, 271 (2005) [astro-ph/0501317].

[24] G. B. Thompson, for HiRes Collaboration, in Proc. Quarks '06 Conf. [astro-ph/0609403]; R. U. Abbasi et al., astro-ph/0703099.

where n_B is the baryon number density (so that $1/n_B$ is the volume per baryon), ϵ is the thermal energy density and p is the pressure. For simplicity, let us consider an ideal gas of photons and non-relativistic particles (mostly protons, helium nuclei, and electrons), with a fixed number (of order unity) $\mathcal{N}$ of the non-relativistic particles per baryon. Then

$$\epsilon = a_{\mathcal{B}} T^4 + \frac{3}{2} n_B \mathcal{N} k_{\mathcal{B}} T, \quad p = \frac{1}{3} a_{\mathcal{B}} T^4 + n_B \mathcal{N} k_{\mathcal{B}} T, \tag{2.2.2}$$

with $a_{\mathcal{B}}$ the radiation energy constant (2.1.7). The solution of Eq. (2.2.1) is here

$$\sigma = \frac{4 a_{\mathcal{B}} T^3}{3 n_B k_{\mathcal{B}}} + \mathcal{N} \ln \left(\frac{T^{3/2}}{n_B C} \right) \tag{2.2.3}$$

where C is an arbitrary constant of integration. It is the quantity σ that remains constant in thermal equilibrium.

We saw in the previous section that the first term in Eq. (2.2.3) is larger than 10^8 at present. (Compare Eqs. (2.1.12) and (2.1.13).) Let us tentatively assume that this quantity was also much larger than unity when photons were in equilibrium with matter. Since σ was constant in this era, this means that the ratio T^3/n_B would also have been very close to constant during the era of equilibrium, unless the quantity $T^{3/2}/n_B$ changed by an enormous amount. For instance, if at some time in the equilibrium era the first term in Eq. (2.2.3) were of order 10^8, then in order for the first term in Eq. (2.2.3) to have changed by even 0.01%, to keep σ constant the value of $T^{3/2}/n_B$ would have had to change by a factor $e^{10000/\mathcal{N}}$. We are not going to be considering such enormous density or temperature ratios here, so we can conclude that T^3/n_B was essentially constant in thermal equilibrium. We saw in the previous section that this ratio was also constant when the photons were traveling freely, and also when photons were interacting only by purely elastic Thomson scattering, so it has been close to constant from the beginning of the era considered here to the present. If we define the constant C in Eq. (2.2.3) to equal the value of $T^{3/2}/n_B$ at some typical time during the era of equilibrium, then the entropy per baryon throughout this era can be taken as

$$\sigma = \frac{4 a_{\mathcal{B}} T^3}{3 n_B k_{\mathcal{B}}} = \frac{3.60 n_{\gamma 0}}{n_{B0}} = 1.31 \times 10^8 h^{-2} \Omega_B^{-1} \tag{2.2.4}$$

The conservation of baryon number tells us that $n_B a^3$ is constant, so the constancy of (2.2.3) has the consequence that $T \propto 1/a$. (Note incidentally that if the first term in Eq. (2.2.3) were of order 10^{-8} instead of greater than 10^8, then it would be the logarithm in the second term that would have to remain constant during thermal equilibrium, in which case $T \propto n_B^{2/3} \propto a^{-2}$, as expected if non-relativistic particles dominate the thermal evolution.)

We can now see why the black-body spectrum with $T \propto 1/a$ is preserved as photons go out of equilibrium with matter, even when we take into account small inelastic effects, like the loss of energy to electron recoil in photon-electron scattering. Photons effectively stop gaining or losing energy to matter when $\Gamma_\gamma < H$, where $H \equiv \dot{a}/a$ and Γ_γ is the rate at which an individual photon loses or gains an energy $k_{\mathcal{B}} T$ through scattering on electrons (to be calculated later). But the rate Γ_e at which an individual electron would gain or lose an energy $k_{\mathcal{B}} T$ by scattering on photons is greater than Γ_γ by a factor $n_\gamma/n_e > 10^8$, so at the time that Γ_γ drops below H we still have $\Gamma_e \gg H$. Thus instead of the electron kinetic energies decreasing like $1/a^2$, as they would in a free expansion, they continue to remain in thermal equilibrium with the photons. While in equilibrium with matter
the photon temperature goes as $1/a$, and we saw in the previous section that when they stop interacting with matter the temperature continues to drop like $1/a$, so through this whole period the electron temperature also drops as $1/a$, and the last few exchanges of energy that photons have with electrons do not affect the photon energy distribution.

According to results quoted in the previous section, the present ratio of the equivalent photon and neutrino mass density and the total matter mass density $\rho_{M0} = \Omega_M \rho_{0,\,\mathrm{crit}}$ is

$$\frac{\rho_{R0}}{\rho_{M0}} = \frac{\Omega_R}{\Omega_M} = 4.15 \times 10^{-5} \Omega_M^{-1} h^{-2} \tag{2.2.5}$$

As we have seen, the photon and neutrino energy density varied as $T^4 \propto a^{-4}$ even before the photons began their free expansion, while the density of pressureless matter varied as a^{-3}, so ρ_R/ρ_M varied as $1/a \propto T$. Therefore the energy density of photons and neutrinos was equal to that of matter when the temperature was

$$T_{\mathrm{EQ}} = \frac{T_{\gamma 0} \Omega_M}{\Omega_R} = 6.56 \times 10^4 \text{ K} \times \Omega_M h^2 \tag{2.2.6}$$

and ρ_R was greater than ρ_M at earlier times. For $\Omega_M h^2 \simeq 0.15, T_{\mathrm{EQ}}$ is about 10^4 K.

Although collisions cannot change the distribution of photon energies as long as the photon number is much greater than the number of charged particles, at sufficiently high temperatures collisions can drastically change the energy of an individual photon. It is of some interest to work out when photons stopped exchanging energies of order $k_{\mathcal{B}} T$ with electrons. The rate at which any individual photon is scattered by electrons is $\Lambda_\gamma = \sigma_{\mathcal{T}} n_e c$, where n_e is the number density of electrons, and $\sigma_{\mathcal{T}} = 0.66525 \times 10^{-24} \mathrm{cm}^2$ is the cross section for Thomson scattering, the elastic scattering of photons by non-relativistic electrons. As we will see in Section 3.2, about 76% of the matter of the universe in this era was ionized hydrogen, with the rest helium, completely ionized at temperatures above $20,000$ K, so with one electron per nucleon for hydrogen and half an electron per nucleon for helium, the net number of electrons per nucleon is $0.76 + (1/2)0.24 = 0.88$, and the number density of electrons at temperature T is $n_e \simeq 0.88 n_B = 0.88 n_{B0} \left(T/T_{\gamma 0}\right)^3$, with a subscript zero as usual indicating the present moment. Using Eq. (2.1.13), the rate at which a photon is scattered by electrons is

$$\Lambda_\gamma = 0.88 n_{B0} \left(\frac{T}{T_{\gamma 0}}\right)^3 \sigma_{\mathcal{T}} c = 1.97 \times 10^{-19} \text{ s}^{-1} \times \Omega_B h^2 \left(\frac{T}{T_{\gamma 0}}\right)^3 \tag{2.2.7}$$

But this is not the rate that governs the effectiveness of the energy exchange between matter and radiation. A photon with energy much less than $m_e c^2$ that strikes a non-relativistic electron will transfer a momentum to the electron of the order of its own momentum, typically about $k_{\mathcal{B}} T$ (in units with $c = 1$), and so it will gain or lose an energy of order $\left(k_{\mathcal{B}} T\right)^2 / m_e$. (This agrees with Eq. (C.20) for $\omega \approx k_{\mathcal{B}} T \ll m_e$.) Hence the rate for energy transfer of order $k_{\mathcal{B}} T$ between a given photon and the electrons equals the rate of collisions times the fraction of the energy $k_{\mathcal{B}} T$ that is transferred per collision:

$$\Gamma_\gamma \simeq \left(\frac{k_{\mathcal{B}} T}{m_e}\right) \Lambda_\gamma \approx 9.0 \times 10^{-29} \text{ s}^{-1} \Omega_B h^2 \left(\frac{T}{T_{\gamma 0}}\right)^4 . \tag{2.2.8}$$

We have to compare this with the cosmic expansion rate. Let's tentatively assume that at the time that concerns us here the density of the universe was dominated by photons and neutrinos, not matter, in which case the cosmic expansion rate was

$$H \equiv \frac{\dot{a}}{a} = H_0 \sqrt{\Omega_R T^4 / T_{\gamma 0}^4} = 2.1 \times 10^{-20} \text{ s}^{-1} \left(\frac{T}{T_{\gamma 0}}\right)^2 . \tag{2.2.9}$$

Therefore Γ_γ was greater than H until the temperature dropped to a value of order

$$T_{\mathrm{freeze}} = 1.5 \times 10^4 \text{ K} \left(\Omega_B h^2\right)^{-1/2} \tag{2.2.10}$$

For $\Omega_B h^2 \simeq 0.02$, this is about 10^5 K. (Comparison with Eq. (2.2.6) shows that for plausible values of $\Omega_M h^2$ and $\Omega_B h^2$, as for instance $\Omega_M h^2 \simeq 0.15$ and $\Omega_B h^2 \simeq 0.02$, we have $T_{\rm freeze} > T_{\rm EQ}$, so the temperature (2.2.10) was reached while the expansion rate was still dominated by radiation, as we have been assuming.)

After the temperature dropped below about 10^5 K photons no longer exchanged appreciable energy with electrons, but at this temperature the rate (2.2.7) of elastic scattering was still roughly 10^5 times greater than H. If Eq. (2.2.7) remained valid then Λ_γ would remain larger than H until the temperature dropped to much lower temperature. For 3 K $\ll T \ll 10^4$ K the universe was matter-dominated, with

$$H = H_0 \sqrt{\Omega_M T^3/T_{\gamma 0}^3} = 3.3 \times 10^{-18} \text{ s}^{-1} \sqrt{\Omega_M h^2} \left(\frac{T}{T_{\gamma 0}}\right)^{3/2} .$$

This became equal to the rate given by Eq. (2.2.7) at a temperature $T \simeq 18 \text{ K} \left(\Omega_M h^2\right)^{1/3} / \left(\Omega_B h^2\right)^{2/3}$, or about 130 K for $\Omega_M h^2 = 0.15$ and $\Omega_B h^2 =$
0.02 , and until then each photon would be scattered many times by electrons in each doubling of $a(t)$. This is not what actually happens, because when the temperature became low enough for electrons and nuclei to hold together as neutral atoms, the elastic scattering rate Λ_γ dropped sharply below the value (2.2.7). As we shall see in the next section, this was at a temperature of about 3,000 K, which marked the end of the era of rapid scattering of photons by electrons.

10.2.3 2.3 Recombination and last scattering

We saw in the previous section that photons stopped exchanging energy effectively with electrons when the temperature of the expanding universe dropped to about 10^5 K. After that, photons continued to be scattered by free electrons, but without appreciable gain or loss of energy. This terminated when the free electrons became bound into hydrogen and helium atoms, ending the scattering of photons. This is called recombination. [1] Let's consider when this happened.

We start our calculation at a time early enough so that protons, electrons, and hydrogen and helium atoms were in thermal equilibrium at the temperature of the radiation. In a gas in equilibrium at temperature T, the number density of any non-relativistic non-degenerate particle of type i is given by the Maxwell-Boltzmann formula:

$$n_i = g_i (2\pi\hbar)^{-3} e^{\mu_i/k_{\mathcal{B}}T} \int d^3q \exp\left[-\left(m_i + \frac{q^2}{2m_i}\right)/k_{\mathcal{B}}T\right] \tag{2.3.1}$$

where m_i is the particle mass, g_i is the number of its spin states, and μ_i is a characteristic of the gas known as the chemical potential of particles of type i. (The generalization of Eq. (2.3.1) to include the effects of relativity and/or degeneracy is given in the following chapter.) The property of the chemical potentials that make this a useful formula is that they are conserved in any reaction that is occurring rapidly in the gas. In our case, the particles are protons, electrons, and hydrogen atoms in any bound state, for which we take i as $p, e, 1s, 2s, 2p$, etc. (As already mentioned, about 24% of the mass of the early universe was in the form of helium nuclei, but helium atoms are more tightly bound than hydrogen atoms, so that at the time that concerns us now, say for $T < 4,400$ K, almost all the helium was locked up in the form of neutral atoms, and therefore played no role here.) The electron and proton

have spin one-half, so $g_p = g_e = 2$, while the $1s$ ground state of the hydrogen atom has two hyperfine states with spins 0 and 1 , so $g_{1s} = 1 + 3 = 4$. At first the recombination and

[471] The "re" in "recombination" may be misleading; before this time electrons and protons had never been combined into atoms. This has become the standard term throughout astrophysics for the capture of electrons into atoms.

ionization reactions $p + e \rightleftharpoons H_{1s}$ occurred rapidly by cascades of radiative transitions through excited states, so the chemical potentials satisfied

$$\mu_p + \mu_e = \mu_{1s} \tag{2.3.2}$$

(Photons can be freely created and destroyed in reactions like $e + p \rightleftharpoons e + p + \gamma$, so their chemical potential vanishes.) The integrals are

$$(2\pi\hbar)^{-3} \int d^3p \exp\left(-\frac{p^2}{2mk_{\mathcal{B}}T}\right) = \left(\frac{mk_{\mathcal{B}}T}{2\pi\hbar^2}\right)^{3/2}$$

so the conservation law (2.3.2) gives

$$\frac{n_{1s}}{n_p n_e} = \left(\frac{m_e k_{\mathcal{B}}T}{2\pi\hbar^2}\right)^{-3/2} \exp\left(B_1/k_{\mathcal{B}}T\right) \tag{2.3.3}$$

where $B_1 \equiv m_p + m_e - m_H = 13.6\text{eV}$ is the binding energy of the $1s$ ground state of hydrogen. (Here we ignore the difference between the hydrogen mass and the proton mass, except in the exponential.) Also, the charge neutrality of cosmic matter requires that

$$n_e = n_p \tag{2.3.4}$$

Further, in equilibrium the number density of hydrogen atoms in any one excited state is less than the number density in the ground states by a factor of order $\exp\left(-\Delta E/k_{\mathcal{B}}T\right)$, where ΔE is the excitation energy, which is necessarily not less than the difference 10.6 eV in the binding energies of the $n = 1$ and $n = 2$ states of hydrogen. (Excited states had the same chemical potential as the ground state, since in equilibrium the atom could go rapidly from one to the other by emitting or absorbing photons.) For temperatures below $4,200$ K this exponential factor is less than 6×10^{-13}, so that to a good approximation we can neglect the presence of excited hydrogen atoms as long as thermal equilibrium persisted. As we will see in Section 3.2, the matter at the time of recombination was about 76% by weight neutral or ionized hydrogen, so we can take

$$n_p + n_{1s} = 0.76 n_B \tag{2.3.5}$$

where n_B is the number density of baryons (i.e. at the temperatures of interest here, of neutrons and protons.) The fractional hydrogen ionization $X \equiv n_p/\left(n_p + n_{1s}\right)$ therefore satisfies the Saha equation:

$$X(1 + SX) = 1 \tag{2.3.6}$$

where

$$S \equiv \frac{\left(n_p + n_{1s}\right) n_{1s}}{n_p^2} = 0.76 n_B \left(\frac{m_e k_{\mathcal{B}}T}{2\pi\hbar^2}\right)^{-3/2} \exp\left(B_1/k_{\mathcal{B}}T\right). \tag{2.3.7}$$

One might think that the fractional ionization in equilibrium would become small when the temperature drops below the value $B_1/k_{\mathcal{B}} = 157,894$ K , but even in equilibrium the recombination is considerably delayed, because of the small value of the coefficient of the exponential in Eq. (2.3.7). With $n_B = n_{B0}\left(T/T_{\gamma 0}\right)^3$ and n_{B0} given by Eq. (2.1.13), we have

$$S = 1.747 \times 10^{-22} e^{157894/T} T^{3/2} \Omega_B h^2, \tag{2.3.8}$$

where h is here again the Hubble constant in units of 100 km $s^{-1}Mpc^{-1}$, and T is the temperature in degrees Kelvin. This function of temperature is extremely rapidly varying where it is of order unity, so the Saha equation gives a quite sharp temperature of recombination, as shown in Table 2.1. We see from Table 2.1 that the equilibrium value of the ionization dropped from over 97% for $T = 4,200$ K to less than 1% for $T = 3,000$ K.

This gives the correct order of magnitude of the temperature of the steep decline in fractional ionization, but it is not correct in detail, because

Table 2.1: Equilibrium hydrogen ionization X for various values of the temperature T and $\Omega_{B} h^{2}$.

T(K)	$\Omega_B h^2 = 0.01$	$\Omega_B h^2 = 0.02$	$\Omega_B h^2 = 0.03$
4,500	0.999	0.998	0.997
4,200	0.990	0.981	0.971
4,000	0.945	0.900	0.863
3,800	0.747	0.634	0.565
3,600	0.383	0.290	0.244
3,400	0.131	0.094	0.078
3,200	0.0337	0.0240	0.0196
3,000	0.00693	0.00491	0.00401
2,800	0.00112	0.00079	0.00065
2.725	2.8×10^{-12571}	2.0×10^{-12571}	1.6×10^{-12571}

equilibrium was not actually maintained for low ionization levels. The photon that is emitted when a free electron is captured by a proton into the ground state has more than enough energy to ionize another hydrogen atom, so this process produces no net decrease in ionization. Similarly, the photon emitted when an electron in a high orbit of the hydrogen atom with principal quantum number $n \geq 3$ falls into the ground state has more than enough energy to lift an electron in the ground state of some other hydrogen atom to the $n = 2$ excited state, so this process also produces no net increase in the number of electrons in the ground state. The ground state of hydrogen is typically reached by formation of excited states H^* in the reaction $e + p \rightarrow H^* + \gamma$, followed by a cascade of radiative decays down to the $n = 2$ excited state. The final transition from the $2p$ excited state to the $1s$ ground state by emission of a single Lyman α photon is impeded by the same effect that impedes the transition of free electrons or electrons in higher excited states to the ground state: the Lyman α photon does not simply merge into the thermal radiation background; it has a large resonant cross section for exciting another hydrogen atom from the ground state to the first excited state, from which it is most often reionized (as shown by the small equilibrium numbers of hydrogen atoms in excited states). But in cosmology this process is not entirely ineffective, because the Lyman α photon emitted by one atom will have just barely enough energy to excite another hydrogen atom in its ground state to the $2p$ state, so if it does not interact very soon with another atom then the cosmological redshift will take its energy outside the resonant line, after which it no longer has enough energy to excite a hydrogen atom in its ground state. Even so, the formation of the ground state by radiative decay from the $2p$ state is so inefficient that we also have to consider slower pathways to the ground state, such as the formation of the $2s$ excited state, which can only decay into the ground state by emitting two photons, neither of which has enough energy to re-excite a hydrogen atom in its ground state. When the ionization became small the rate of these reactions could no longer compete with the cosmic expansion rate, and

the ionization no longer fell as fast as it would in thermal equilibrium. [2]

This is a complicated business, but it is not hard to see the main outlines of the recombination process. We make the following approximations:

1. Collisions between hydrogen atoms and radiative transitions between the states of these atoms are sufficiently rapid so that all states of the atoms are in equilibrium with each other at the temperature T of the radiation, except for the $1s$ ground state, which as already mentioned is reached only through slow or inefficient processes. This has the consequence that the number density of states $n\ell$ of hydrogen having principal quantum number n (with $n > 1$) and orbital angular momentum ℓ can be expressed in terms of the number density of any one state, say the $2s$ state: [3]

$$n_{n\ell} = (2\ell + 1) n_{2s} \exp\left(\left(B_2 - B_n\right)/k_{\mathcal{B}}T\right) \tag{2.3.9}$$

where B_n is the binding energy [4] of the state with principal quantum number n.

2. The net rate of change in the population of hydrogen atoms in their $1s$ state is given by the rate of radiative decay from the $2s$ and $2p$ states, minus the rate of excitation of these states from the $1s$ state. In accordance with the discussion above, all other processes leading to the ground state are assumed to be canceled by the reionization or re-excitation of other atoms by the emitted photon. Because recombination occurs in collisions of electrons and protons, it decreases the number $n_e a^3$ of free electrons in a co-moving volume a^3 at a rate $\alpha(T) n_p n_e a^3$ that is proportional both to $n_p = n_e$ and $n_e a^3$, with the coefficient $\alpha(T)$ depending only on temperature, not on n_e or a. (We do not include recombination directly to the ground state here, since it is canceled by the ionization of other atoms by the emitted photon, so $\alpha(t)$ is what in astrophysics is called the "case B recombination coefficient.") Also, leaving aside ionization from the ground state which just cancels recombination to the ground state, the ionization from excited states of hydrogen increases $n_e a^3$ at a rate given by a sum of terms proportional to the $n_{n\ell} a^3$ with $n > 1$, with coefficients depending only on temperature. Eq. (2.3.9) gives all the $n_{n\ell}$ with $n > 1$ as proportional to n_{2s}, with coefficients that also depend only on temperature, so ionization increases the number of electrons in a co-moving volume a^3 at a rate that can be written as $\beta(T) n_{2s} a^3$, with $\beta(T)$

a function only of temperature, not of a or n_e or any of the $n_{n\ell}$. Putting these rates together, we have

$$\frac{d}{dt}\left(n_e a^3\right) = -\alpha n_e^2 a^3 + \beta n_{2s} a^3$$

Dividing by the constant na^3 (where $n \equiv n_p + n_H = n_p + \sum_{n\ell} n_{n\ell} = 0.76 n_B$), this gives

[472] The classic paper on this subject is by P. J. E. Peebles, Astrophys. J. 153, 1 (1968). Also see Ya. B. Zel'dovich, V. G. Kurt, and R. A. Sunyaev, Soviet Physics JETP 28, 146 (1969). At the time of writing the most thorough analysis known to me is that of S. Seager, D. D. Sasselov, and D. Scott , Astrophys. J. 523, L1 (1999); Astrophys. J. Suppl. Ser. 128, 407 (2000) [astro-ph/9909275]. Corrections of the order of a percent of the calculated value due to additional transitions and stimulated emission were subsequently found by V. K. Dubrovich and S. I. Grachev, Astron. Lett. 31, 359 (2005) [astro-ph/051672]; J. Chluba and R. A. Sunyaev, Astron. Astrophys. 446, 39 (2006) [astro-ph/0508144]; W. Y. Wong and D. Scott, Mon. Not. Roy. Astron. Soc. 375, 1441 (2007) [astro-ph/0610691]. The possible use of observations of the cosmic microwave background to resolve these uncertainties is discussed by A. Lewis, J. Weller, and R. Battye, Mon. Not. Roy. Astron. Soc. 373, 561 (2006) [astro-ph/0606552]. For the sake of simplicity here I will make the same approximations as Peebles, but will use the "escape probability"

[47]method discussed by Seager et al., which seems to me more direct and easier to justify.

[3] This formula does not apply to hydrogen atoms in very high excited states, which have such large radii and small binding energies that they cannot be treated as free particles. See, e.g., D. Mihalas, Stellar Atmospheres, 2nd edition (Freeman, San Francisco, 1978).

[4] We are here neglecting the fine structure, hyperfine structure, and Lamb energy shifts, which give the binding energies a very small dependence on ℓ and on the total (including spin) angular momentum j.

$$\frac{d}{dt}\left(\frac{n_e}{n}\right) = -\frac{\alpha n_e^2}{n} + \frac{\beta n_{2s}}{n} \tag{2.3.10}$$

Further, this must vanish under conditions of equilibrium, where (since the transitions $e + p \rightleftharpoons 2s$ occur rapidly) instead of Eq. (2.3.3) we would have

$$\frac{n_{2s}}{n_e^2} = \left(\frac{m_e k_{\mathcal{B}} T}{2\pi\hbar^2}\right)^{-3/2} \exp\left(B_2/k_{\mathcal{B}}T\right)$$

so the coefficients in Eq. (2.3.10) are related by

$$\beta/\alpha = \left(\frac{n_e^2}{n_{2s}}\right)_{\text{equilibrium}} = \left(\frac{m_e k_{\mathcal{B}} T}{2\pi\hbar^2}\right)^{3/2} \exp\left(-B_2/k_{\mathcal{B}}T\right) \tag{2.3.11}$$

In order for Eq. (2.3.10) to be useful, we need to relate n_{2s} to the total number density n of protons and hydrogen atoms.

3. The total number of excited hydrogen atoms in a co-moving volume $1/n$ changes so much more slowly than individual radiative processes that the net increase in this number due to recombination and reionization of hydrogen is balanced by the net decrease in this number by transitions to and from the $1s$ state. That is,

$$\alpha n_e^2 - \beta n_{2s} = \left(\Gamma_{2s} + 3P\Gamma_{2p}\right) n_{2s} - \mathcal{E} n_{1s} \tag{2.3.12}$$

where Γ_{2s} and Γ_{2p} are the rates for the radiative decay processes $2s \to 1s + \gamma + \gamma$ and $2p \to 1s + \gamma$, respectively; [5]P is the probability that the Lyman α photon emitted in the decay $2p \to 1s + \gamma$ will escape to infinity without exciting some other hydrogen atom in the $1s$ state back to the $2p$ state (to be calculated below); and $\mathcal{E}$ is the rate at which hydrogen atoms in the $1s$ state are excited to the $2s$ or $2p$ state, not including those that are excited to the $2p$ state by Lyman α photons from the decays $2p \to 1s + \gamma$, which we take into account with the factor P. (The factor 3 in Eq. (2.3.12)

enters because Eq. (2.3.9) gives $n_{2p} = 3n_{2s}$.) From this, we obtain the needed formula for n_{2s} :

$$n_{2s} = \frac{\alpha n_e^2 + \mathcal{E} n_{1s}}{\Gamma_{2s} + 3P\Gamma_{2p} + \beta} \tag{2.3.13}$$

We will be concerned here with temperatures $T \ll \left(B_2 - B_3\right)/k_{\mathcal{B}} = 21,900$ K, so according to Eq. (2.3.9) all $n_{n\ell}$ with $n > 2$ are much less than n_{2s}, and hence to a very good approximation the total number of hydrogen atoms is

$$n_H = n_{1s} + n_{2s} + n_{2p} = n_{1s} + 4n_{2s}. \tag{2.3.14}$$

We can therefore eliminate n_{1s} in favor of n_H in Eq. (2.3.13), so

$$n_{2s} = \frac{\alpha n_e^2 + \mathcal{E} n_H}{\Gamma_{2s} + 3P\Gamma_{2p} + \beta + 4\mathcal{E}} \tag{2.3.15}$$

Further, in equilibrium the number of hydrogen atoms in the $1s$ state would be constant, so the coefficients on the right-hand side of Eq. (2.3.12) (which gives the net rate of increase of the number density of hydrogen atoms in the ground state) must have the ratio

$$\frac{\mathcal{E}}{\Gamma_{2s} + 3P\Gamma_{2p}} = \left(\frac{n_{2s}}{n_{1s}}\right)_{\text{equilibrium}} = \exp\left(-\left(B_1 - B_2\right)/k_{\mathcal{B}}T\right). \tag{2.3.16}$$

[475] We neglect the process $2p \to 1s + \gamma + \gamma$, which is much slower than $2s \to 1s + \gamma + \gamma$.

Using Eqs. (2.3.15) and (2.3.16) in Eq. (2.3.10) now gives the rate equation in a more useful form

$$\frac{d}{dt}\left(\frac{n_e}{n}\right) = \frac{\Gamma_{2s} + 3P\Gamma_{2p}}{\left(\Gamma_{2s} + 3P\Gamma_{2p}\right)\left[1 + 4\exp\left(-\left(B_1 - B_2\right)/k_{\mathcal{B}}T\right)\right] + \beta} \times \left(-\left[1 + 4\exp\left(-\left(B_1 - B_2\right)/k_{\mathcal{B}}T\right)\right]\frac{\alpha n_e^2}{n} + \exp\left(-\left(B_1 - B_2\right)/k_{\mathcal{B}}T\right)\frac{\beta n_H}{n}\right)$$

At the temperatures of interest the factor $1 + 4\exp\left(-\left(B_1 - B_2\right)/k_{\mathcal{B}}T\right)$ can be replaced with unity. Using Eq. (2.3.11) and the definition of fractional ionization $X \equiv n_e/n = n_p/n = 1 - n_H/n$, we have at last

$$\frac{dX}{dt} = \left(\frac{\Gamma_{2s} + 3P\Gamma_{2p}}{\Gamma_{2s} + 3P\Gamma_{2p} + \beta}\right)\alpha n\left(-X^2 + S^{-1}(1 - X)\right), \tag{2.3.17}$$

where $S(T)$ is the function (2.3.7) appearing in the Saha equation. Note that if the temperature were constant then this would be satisfied by any solution of the Saha equation (2.3.6). In fact, X is always larger than the value given by the Saha equation, so Eq. (2.3.17) gives a monotonically decreasing fractional ionization. The first factor in Eq. (2.3.17) represents the suppression of recombination that occurs when the transitions of the $2s$ and $2p$ states to the ground state are slower than the reionization of the atom.

It remains to calculate the photon survival probability P. This is given in general by

$$P(t) = \int_{-\infty}^{+\infty} d\omega \mathcal{P}(\omega) \exp\left[-\int_t^{\infty} dt' n_{1s}\left(t'\right) c\sigma\left(\omega a(t)/a\left(t'\right)\right)\right] \tag{2.3.18}$$

where $\mathcal{P}(\omega)d\omega$ is the probability that a photon emitted in the transition $2p \to 1s$ has energy between $\hbar\omega$ and $\hbar(\omega + d\omega)$, normalized so that $\int \mathcal{P}(\omega)d\omega = 1$, and $\sigma(\omega)$ is the cross section for the excitation $1s \to 2p$ by a photon of energy $\hbar\omega$. The factor $a(t)/a\left(t'\right)$ arises from the cosmological redshift of the photon. The cross section is given by the Breit-Wigner formula [6]

$$\sigma(\omega) = \left(\frac{3}{2}\right)\left(\frac{2\pi^2\Gamma_{2p}}{k_\alpha^2}\right)\mathcal{P}(\omega) \tag{2.3.19}$$

where k_α is the mean wave number $\left(B_1 - B_2\right)/\hbar c$ of the Lyman α photon emitted in the transition $2p \to 1s$. The probability density here is

$$\mathcal{P}(\omega) = \frac{\Gamma_{2p}}{2\pi}\frac{1}{\left(\omega - \omega_\alpha\right)^2 + \Gamma_{2p}^2/4} \tag{2.3.20}$$

where ω_α is the circular frequency ck_α corresponding to the wave number k_α.

Now, if a photon is captured at all then it is captured in a time much less than the characteristic expansion time of the universe, so we can set the density $n_{1s}\left(t'\right)$ in Eq. (2.3.18) equal to $n_{1s}(t)$. The cross section $\sigma(\omega)$ varies very rapidly with ω, so we cannot neglect the time-dependence of $a(t)$, but we can take the expansion rate to be constant in the integral over t', so that $a(t)/a\left(t'\right) = 1 - H(t)\left(t' - t\right)$, where $H(t) = \dot{a}(t)/a(t)$. It is convenient then

[476] See, e.g., S. Weinberg, The Quantum Theory of Fields, (Cambridge University Press, Cambridge, UK, 1995, 1996, 2000) [referred to henceforth as QTF]: Vol. I, Eq. (3.8.16). The factor 3/2 arises from the number $2\ell + 1 = 3$ of $2p$ angular momentum states, and the number 2 of photon helicity states. If the spin of the electron and proton are taken into account then the numerator is 12 and the denominator is 8 , but the ratio is the same.

to change the variable of integration in the exponential in Eq. (2.3.18) from t' to $\omega' = (1 - H(t)\,(t' - t))\,\omega$, so that Eq. (2.3.18) reads

$$P(t) = \int_{-\infty}^{+\infty} d\omega \mathcal{P}(\omega) \exp\left[-\frac{3\pi^2 \Gamma_{2p} n_{1s}(t) c}{\omega H(t) k_\alpha^2} \int_{-\infty}^{\omega} d\omega' \mathcal{P}\left(\omega'\right)\right] \tag{2.3.21}$$

The function $\mathcal{P}(\omega)$ is negligible except for ω very near ω_α, so we can replace ω with ω_α in the factor $1/\omega$ in the argument of the exponential (but not in the upper limit of the integral over ω', which rises steeply from zero to one for ω in the neighborhood of ω_α.) The integral over ω is now trivial, and gives

$$P(t) = F\left(\frac{3\pi^2 \Gamma_{2p} n_{1s}(t) c}{\omega_\alpha H(t) k_\alpha^2}\right) \tag{2.3.22}$$

where

$$F(x) \equiv \left(1 - e^{-x}\right)/x \tag{2.3.23}$$

The Lyman α transition rate [7] is $\Gamma_{2p} = 4.699 \times 10^8\ \mathrm{s}^{-1}$, so that the argument of the function $F(x)$ in Eq. (2.3.22) is very large. For instance, if we take $\Omega_M h^2 = 0.15$ and $\Omega_B h^2 = 0.01$, then for $T < 6,000$ K the argument x is greater than 10. Hence for the temperatures of interest here we can drop the exponential in Eq. (2.3.23), and find

$$P = \frac{\omega_\alpha H k_\alpha^2}{3\pi^2 \Gamma_{2p} n_{1s} c} = \frac{8\pi H}{3\lambda_\alpha^3 \Gamma_{2p} n_{1s}}, \tag{2.3.24}$$

where $\lambda_\alpha = 1215.682 \times 10^{-8}$ cm is the wavelength of Lyman α photons, and the argument t is now understood. We see that the quantity $3P\Gamma_{2p}$ in the rate equation (2.3.17) is independent of Γ_{2p} :

$$3P\Gamma_{2p} = \frac{8\pi H}{\lambda_\alpha^3 n_{1s}} \tag{2.3.25}$$

With this result for P, equation (2.3.17) is the same as the rate equation found in a different way by Peebles. [2] In this equation we may use the approximation that at the temperatures of interest, which are much less than $(B_1 - B_2)/k_{\mathcal{B}} = 118,420$ K, even though n_{1s}/n_{2s} is less than it would be in thermal equilibrium it is still much larger than unity, so that we can replace n_{1s} with $n_H = (1 - X)n$. We also replace time with temperature as the independent variable, using

$$\frac{dt}{dT} = -\frac{1}{HT} \tag{2.3.26}$$

The rate equation (2.3.17) then becomes

$$\frac{dX}{dT} = \frac{\alpha n}{HT}\left(1 + \frac{\beta}{\Gamma_{2s} + 8\pi H/\lambda_\alpha^3 n(1 - X)}\right)^{-1} \left[X^2 - (1 - X)/S\right] \tag{2.3.27}$$

Now we must put in some numbers. In calculating the expansion rate we need to include both the energy density of non-relativistic matter and that of neutrinos and photons, but at temperatures where neutrinos are important their mass is negligible, and for temperatures $T > 30$ K we may neglect vacuum energy, so

[477] W. L. Wiese, M. W. Smith, and R. M. Glennon, Atomic Transition Probabilities, Vol. I, National Standard Reference Data Series NBS 4 (1966).

$$H = H_0 \left[\Omega_M \left(\frac{T}{T_{\gamma 0}} \right)^3 + \Omega_R \left(\frac{T}{T_{\gamma 0}} \right)^4 \right]^{1/2}$$
$$= 7.204 \times 10^{-19} T^{3/2} \sqrt{\Omega_M h^2 + 1.523 \times 10^{-5} T} \text{ s}^{-1} \quad (2.3.28)$$

where again in all numerical expressions T is the temperature in degrees Kelvin. The number density of ionized and un-ionized hydrogen is

$$n = 0.76 \times \frac{3H_0^2 \Omega_B}{8\pi G m_p} \left(\frac{T}{T_{\gamma 0}} \right)^3 = 4.218 \times 10^{-7} \Omega_B h^2 T^3 \text{ cm}^{-3} \quad (2.3.29)$$

The two-photon decay rate of the $2s$ state is [8]

$$\Gamma_{2s} = 8.22458 \text{ s}^{-1} \quad (2.3.30)$$

The coefficient of the proton number density in the electron recombination rate used by Peebles is [9]

$$\alpha = 2.84 \times 10^{-11} T^{-1/2} \text{ cm}^3 \text{ s}^{-1}$$

The factor $T^{-1/2}$ here is what would be expected from a factor velocity $^{-1}$ that generally appears in the cross sections for exothermic reactions like $e+p \to H+\gamma$.[10] Actually α represents a chain of reactions more complicated than just the radiative capture of an electron, including the cascade of radiative decays down to the $2s$ and $2p$ states, and so it has a more complicated temperature dependence. A variety of detailed numerical calculations of the effective rate of recombination to the $2s$ and $2p$ states can be fit

with the simple formula: [11]

$$\alpha = \frac{1.4377 \times 10^{-10} T^{-0.6166} \text{ cm}^3 \text{ s}^{-1}}{1 + 5.085 \times 10^{-3} T^{0.5300}}. \quad (2.3.31)$$

The value of β is given in terms of α by Eq. (2.3.11):
$\beta = \left(\frac{m_e k_{\mathcal{B}} T}{2\pi\hbar^2}\right)^{3/2} \exp\left(-B_2/k_{\mathcal{B}} T\right) \alpha = 2.4147 \times 10^{15} \text{ cm}^{-3} T^{3/2} e^{-39474/T} \alpha.$

Finally, the function $S[T]$ is given by Eq. (2.3.8).

The values of $X[T]$ calculated [12] from the differential equation (2.3.27) with the inputs (2.3.28)-(2.3.32) (and Eq. (2.3.8) for $S[T]$) are given in Table 2.2 for $\Omega_M h^2 = 0.15$ and a range of values of $\Omega_B h^2$. The initial condition here is taken to be that $X[T]$ is given by the solution of the Saha equation (2.3.6) for thermal equilibrium at the highest temperature considered, which we take to be $T = 4,226$ K (i.e. $z = 1,550$), high enough so that thermal equilibrium should be a good approximation (because the equilibrium value of X is very close to one, while the true value must be between one and the equilibrium value), but low enough so that hardly any of the helium is still ionized. (Almost all helium was doubly ionized until the temperature dropped below $20,000$ K, and there was still appreciable singly ionized helium until the temperature dropped below about $4,400$ K.) The actual value of $X[4226]$ is slightly higher than the equilibrium value, but to three significant figures there would be no effect on the results at lower temperatures if we increased the assumed value of $X[4226]$ by a small amount, as say from 0.984 to 0.99 for $\Omega_B h^2 = 0.02$.

[478] S. P. Goldman, Phys. Rev. A40, 1185 (1989). Peebles used an earlier value, 8.227 s^{-1}, given by L. Spitzer and J. L. Greenstein, Astrophys. J. 114, 407 (1951).
[9] W. J. Boardman, Astrophys. J. Suppl. 9, 185 (1964).
[10] See, e.g., QTF, Vol. I, p. 157.

Comparison of Tables 2.1 and 2.2 shows that for the small values of $\Omega_B h^2$ considered here, the Saha equation stopped giving a good approximation to the fractional ionization as soon as the equilibrium ionization dropped appreciably below unity. In particular, although in equilibrium the fractional ionization would have dropped to vanishingly small values for temperatures below $2,000$ K, the ionization calculated from Eq. (2.3.27) leveled off at low temperatures to a small but non-zero asymptotic value,

Table 2.2: Hydrogen ionization X calculated from Eq. (2.3.27) and time t for the temperature to drop to T from \$10\hat{}\{6\{ \}\^{}\{\circ\} \mathrm\{K\}\$ calculated from Eq. (2.3.26). The fourth through sixth columns give results for various values of $\Omega_B h^2$, with $\Omega_M h^2 = 0.15$. The last column gives results for parameters $\Omega_B h^2 = 0.02238, \Omega_M h^2 = 0.13229$ used in Section 7.2 to compare the analytic calculation of microwave background anisotropies with a numerical calculation.

z	T(K)	t (yrs)	$X_{\Omega_B h^2=0.01}$	$X_{\Omega_B h^2=0.02}$	$X_{\Omega_B h^2=0.03}$	$X_{\text{Sec.7.2}}$
1550	4226	202,600	0.992	0.984	0.977	0.982
1500	4090	213,200	0.976	0.958	0.943	0.954
1450	3954	225,900	0.935	0.902	0.878	0.895
1400	3818	239,800	0.861	0.815	0.780	0.805
1350	3681	255,200	0.759	0.703	0.659	0.690
1300	3545	272,000	0.645	0.580	0.529	0.564
1250	3409	290,600	0.526	0.456	0.402	0.437
1200	3273	311,300	0.409	0.339	0.289	0.321
1150	3136	334,600	0.299	0.236	0.194	0.220
1100	3000	360,400	0.205	0.154	0.122	0.142
1050	2864	389,600	0.129	0.0928	0.0721	0.0846
1000	2728	422,600	0.0752	0.0520	0.0396	0.0470
950	2591	460,500	0.0405	0.0270	0.0203	0.0243
900	2455	503,600	0.0210	0.0136	0.0101	0.0121
800	2183	611,400	0.00662	0.00387	0.00276	0.00339
700	1910	761,300	0.00319	0.00174	0.00120	0.00150
600	1638	977,700	0.00203	0.00107	0.000731	0.00920
500	1365	1.312×10^6	0.00147	0.000762	0.000517	0.000653
400	1093	1.872×10^6	0.00114	0.000585	0.000395	0.000499
250	684	3.922×10^6	0.000829	0.000423	0.000285	0.000361
100	275	1.604×10^7	0.000632	0.000321	0.000216	0.000273
50	139	4.535×10^7	0.000579	0.000294	0.000197	0.000250
10	30.0	4.568×10^8	0.000537	0.000272	0.000183	0.000231

[4711] D. Péquignot, P. Petijean, and C. Boisson, Astron. Astrophys. 251, 680 (1991). This includes an over-all "fudge factor" of 1.14 recommended by S. Seager, D. D. Sasselov, and D. Scott, Astrophys. J. 523, L1 (1999). Eq. (2.3.31) agrees well with subsequent calculations of D. G. Hummer, Mon. Not. Roy. Astron. Soc. 268, 109 (1994).

[12] I thank D. Dicus for the calculation of these numerical results. The calculation is stopped at $z = 10$, because at later times hydrogen is reionized by the first generation of stars, and also the rate of expansion is affected by the vacuum energy.

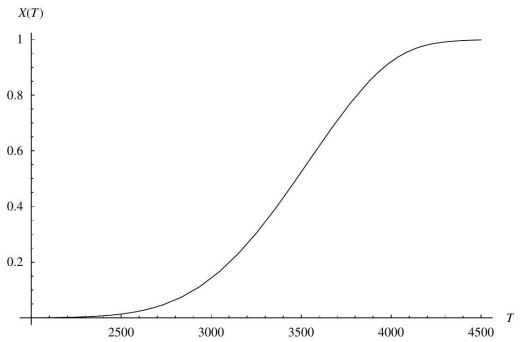


Figure 2.2: The fractional ionization given by the rate equation (2.3.17) as a function of temperature in degrees Kelvin, for $\Omega_{M} h^{2}=0.132, \Omega_{B} h^{2}=0.0224$.

due to the increasing rarity of encounters of the few remaining free protons and electrons. This residual ionization played an important role in the formation of the first stars.

The fractional ionization given by Eq. (2.3.17) for a currently favored set of cosmological parameters is also shown in Figure 2.2. On the scale of this figure, the results are indistinguishable from those given by the more elaborate calculations of Seager, Sasselov, and Scott, [2] except for temperatures above $T > 4,300$ K, where the contribution of electrons from the ionization of helium (ignored in Eq. (2.3.17)) was still significant.

The most important application of these results for the fractional ionization is the calculation of the opacity $\mathcal{O}(T)$, the probability that a photon present at a time $t(T)$ when the temperature is T will undergo at least one more scattering before the present, given by

$$\mathcal{O}(T) = 1 - \exp\left[-\int_{t(T)}^{t_0} c\sigma_{\mathcal{T}} n_e(t) dt\right] \tag{2.3.33}$$

This rises from near zero at low temperature, where the integral in the exponent is small, to near one at high temperature, where the integral becomes large.

It is convenient to convert the integral over time to an integral over temperature, again using Eq. (2.3.26):

$$\mathcal{O}(T) = 1 - \exp\left[-c\sigma_{\mathcal{T}} \int_{2.725}^{T} n_e\left(T'\right) dT' / H\left(T'\right) T'\right] \tag{2.3.34}$$

We use Eq. (2.3.28) for H and take $n_e = Xn$ with n given by Eq. (2.3.29), so

$$\begin{aligned}\mathcal{O}(T) &= 1 - \exp\left[-c\sigma_{\mathcal{T}} \int_{2.725}^{T} n_e\left(T'\right) dT' / H\left(T'\right) T'\right] \\ &= 1 - \exp\left[-\int_{2.725}^{T} \frac{0.01168 \Omega_B h^2 T'^{1/2} X\left(T'\right) dT'}{\sqrt{\Omega_M h^2 + 1.523 \times 10^{-5} T'}}\right]\end{aligned} \tag{2.3.35}$$

In studies of anisotropies in the cosmic microwave background we are particularly interested in when the photons observed today were last scattered. The probability that the last scattering of a photon was before the temperature dropped to T is $1 - \mathcal{O}(T)$, and the probability that the last scattering was after the temperature dropped further to $T - dT$ is $\mathcal{O}(T - dT)$, so the probability that the last scattering of a photon was at a temperature between T and $T - dT$ is

$$1 - (1 - \mathcal{O}(T)) - \mathcal{O}(T - dT) = \mathcal{O}'(T) dT \tag{2.3.36}$$

The opacity function $\mathcal{O}(T)$ increases monotonically with temperature from $\mathcal{O} = 0$ at $T = T_0$ to $\mathcal{O} \to 1$ for $T \to \infty$, so $\mathcal{O}'(T)$ is a positive normalized probability distribution, with unit

integral. Values [12] of $\mathcal{O}'(T)$ for $\Omega_M h^2 = 0.15$ and a range of values of $\Omega_B h^2$ are given in Table 2.3. The distribution $\mathcal{O}'(T)$ is peaked at a temperature T_L, with a standard deviation σ, given in the two bottom rows of Table 2.3.

It may be useful to note an approximation [13] for the fractional ionization, even though it was not used in the calculations of Tables 2.2 or 2.3. In equilibrium the fractional ionization depends only on the temperature and $\Omega_B h^2$. But as soon as the Saha equation stopped giving a good approximation to the fractional ionization, the ionization then depended not only on $\Omega_B h^2$, but also on $\Omega_M h^2$. The approximation derived below shows that at sufficiently low temperatures the fractional ionization depends on $\Omega_B h^2$ and $\Omega_M h^2$ chiefly through a multiplicative factor $(\Omega_M h^2)^{1/2}/\Omega_B h^2$, while the opacity $\mathcal{O}'(T)$ is nearly independent of these parameters.

Table 2.3: The normalized probability distribution $\mathcal{O}'(T)$ of the temperature of last scattering, as a function of the temperature T (in degrees Kelvin), calculated from Eq. (2.3.35). The second through fourth columns give results for various values of $\Omega_B h^2$, with $\Omega_M h^2 = 0.15$. The last column gives results for parameters $\Omega_B h^2 = 0.02238$, $\Omega_M h^2 = 0.13229$ used in Section 7.2 to compare the analytic calculation of microwave background anisotropies with a numerical calculation. The bottom two rows give the parameters for a fit of $\mathcal{O}'(T)$ to the Gaussian $\exp\left[-(T-T_L)^2/2\sigma^2\right]/\sigma\sqrt{2\pi}$, found by setting T_L equal to temperature at which $\mathcal{O}'$ is a maximum, and $1/\sigma\sqrt{2\pi}$ equal to the value of $\mathcal{O}'$ at that maximum.

T(K)	$\mathcal{O}'_{\Omega_B h^2=0.01}$	$\mathcal{O}'_{\Omega_B h^2}=0.02$	$\mathcal{O}'_{\Omega_B h^2=0.03}$	$\mathcal{O}'_{\text{Sec.7.2}}$
4000	6.80×10^{-7}	5.39×10^{-10}	1.08×10^{-12}	5.75×10^{-11}
3500	2.26×10^{-4}	4.12×10^{-5}	1.15×10^{-5}	2.53×10^{-5}
3400	0.000451	0.000152	0.000069	0.000112
3300	0.000759	0.000412	0.000262	0.000345
3200	0.00109	0.000826	0.000664	0.000759
3100	0.00132	0.00127	0.00118	0.00124
3000	0.00139	0.00155	0.00157	0.00156
2900	0.00127	0.00154	0.00164	0.00158
2800	0.00102	0.00130	0.00142	0.00135
2700	0.000746	0.000965	0.00107	0.00101
2600	0.000502	0.000650	0.000721	0.000680
2500	0.000320	0.000411	0.000455	0.000429
2000	4.66×10^{-5}	5.16×10^{-5}	5.39×10^{-5}	5.25×10^{-5}
1000	9.50×10^{-6}	9.76×10^{-6}	9.87×10^{-6}	9.84×10^{-6}
T_L(K)	3017	2954	2930	2941
σ(K)	287	253	241	248

To derive this approximation, note first that as soon as the fractional ionization dropped well below its equilibrium value, the term $(1-X)/S$ in the square brackets in Eq. (2.3.27) (which would equal X^2 in equilibrium) became much less than the term X^2. Comparison of Tables 2.1 and 2.2 shows that for a plausible range of cosmological parameters, this is the case for $T < 3,400$ K. At such temperatures we can also neglect the radiation
energy density compared with the mass density, so that the expansion rate (2.3.28) becomes simply $7.2\times10^{-19}\sqrt{\Omega_M h^2}$ s^{-1}. Also, for temperatures in the range from 3,400 K down to 2000 K the effective $2p$ decay rate $3P\Gamma_{2p}$ was less than the $2s$ decay rate, while for $T < 2,000$ K the reionization rate β was less than either $3P\Gamma_{2p}$ or Γ_{2s}, so for all temperatures below $T < 3,400$ K, Eq. (2.3.27) took the form

[4713] B. J. T. Jones and R. F. Wyse, Astron. Astrophys. 149, 144 (1985).

$$\frac{dX}{dT} = f(T)X^2 \tag{2.3.37}$$

where

$$f(T) = \frac{\alpha n}{HT}\left(1 + \frac{\beta}{\Gamma_{2s}}\right)^{-1} = \frac{\Omega_B h^2}{(\Omega_M h^2)^{1/2}} g(T), \tag{2.3.38}$$

with $g(T)$ a function of temperature alone

$$g(T) = \frac{84.2 T^{-0.1166}}{1 + 5.085 \times 10^{-3} T^{0.53} + 4.22 \times 10^4 T^{.8834} e^{-39474/T}} \tag{2.3.39}$$

The solution is

$$X(T) \simeq \left[X(3400)^{-1} + \frac{\Omega_B h^2}{(\Omega_M h^2)^{1/2}} \int_T^{3400} g\left(T'\right) dT'\right]^{-1} \tag{2.3.40}$$

Comparison of this formula with the results shown in Table 2.2 for $\Omega_M = 0.15$ and $\Omega_B = 0.02$ shows that the error introduced by these approximations rises to about 25% as the temperature drops from $3,400$ K to $\simeq 2,500$ K, and then drops to less than 10% for $T < 1,400$ K. For temperatures less than about $2,600$ K the fractional ionization is so much less than at $3,400$ K that we can neglect the term $X(3400)^{-1}$ in the denominator of Eq. (2.3.36) without introducing any appreciable additional error, giving

$$X(T) \simeq \frac{(\Omega_M h^2)^{1/2}}{\Omega_B h^2} \left[\int_T^{3400} g\left(T'\right) dT'\right]^{-1} \tag{2.3.41}$$

so that $X(T)$ is proportional at sufficiently low temperature to $(\Omega_M h^2)^{1/2} / \Omega_B h^2$, as was to be shown. Inspection of Table 2.2 confirms that for $T < 1,700$ K and with $\Omega_M h^2 = 0.15$ fixed, the fractional ionization is indeed inversely proportional to $\Omega_B h^2$ with fair accuracy, but this is not a good approximation in the neighborhood of $3,000$ K.

It is striking that the values of $\mathcal{O}'(T)$ depend only weakly on $\Omega_B h^2$ for $T < 3,200$ K. The reason, as noted by Jones and Wyse, [13] is that at temperatures low enough to neglect the radiation energy density, Eq. (2.3.35) gives
$\ln \mathcal{O}(T)$ as proportional to $\Omega_B h^2 / \sqrt{\Omega_M h^2}$ times an integral of $T^{1/2} X(T)$, while the approximate formula (2.3.41) gives $X(T)$ proportional at low temperature to $\sqrt{\Omega_M h^2} / \Omega_B h^2$, so at sufficiently low temperature $\mathcal{O}(T)$ and hence $\mathcal{O}'(T)$ are independent of $\Omega_B h^2$ and also of $\Omega_M h^2$. Table 2.3 shows that for plausible values of cosmological parameters, the temperature T_L of last scattering is always close to $3,000$ K, with a spread (in the sense of standard deviation) of about 10%.

10.2.4 2.4 The dipole anisotropy

In the previous sections of this chapter we have treated the cosmic microwave background as perfectly isotropic and homogeneous. This is certainly a good approximation. Indeed, in the discovery of the cosmic microwave background in 1965, the one thing that enabled Penzias and Wilson to distinguish the background radiation from radiation emitted by earth's atmosphere was that the microwave background did not seem to vary with direction in the sky.

Of course, the cosmic microwave background does have small variations in direction that are too small to have been detected by Penzias and Wilson. The departures from perfect isotropy provides some of the most important information we have about the evolution of the universe. This section deals with the simplest and earliest detected departure from isotropy of the observed cosmic microwave background, arising from the earth's motion. The following

two sections deal with anisotropies due to scattering of photons by intergalactic electrons in clusters of galaxies, and with the primary anisotropies left over from the early universe.

To the extent that the cosmic microwave background is itself perfectly homogeneous and isotropic, it provides a frame of reference for the whole universe, with respect to which we can measure the peculiar velocities of individual galaxies. To analyze the measurement of our own galaxy through the cosmic microwave background, it is useful to consider the density $N_\gamma(\mathbf{p})$ of photons in phase space, defined by specifying that there are $N_\gamma(\mathbf{p})d^3p$ photons of each polarization (right or left circularly polarized) per unit spatial volume in a momentum-space volume d^3p centered at $\mathbf{p}$. Since $|\mathbf{p}| = hv/c$ and the momentum-space volume between frequencies v and dv is $4\pi h^3 v^2 dv/c^3$, Eq. (2.1.1) gives

$$N_\gamma(\mathbf{p}) = \frac{1}{2}\frac{n_T(c|\mathbf{p}|/h)}{4\pi h^3 v^2/c^3} = \frac{1}{h^3}\frac{1}{\exp(|\mathbf{p}|c/kT) - 1}, \tag{2.4.1}$$

(The factor 1/2 takes account of the fact that n_T includes both possible photon polarization states.) This is of course the density that would be measured
by an observer at rest in the radiation background. The phase space volume is Lorentz invariant, and the number of photons is also Lorentz invariant, so N_γ is a scalar, in the sense that a Lorentz transformation to a coordinate system moving with respect to the radiation background that takes $\mathbf{p}$ to $\mathbf{p}'$ also takes N_γ to N'_γ, where

$$N'_\gamma(\mathbf{p}') = N_\gamma(\mathbf{p}) \tag{2.4.2}$$

If the earth is moving in the three-direction with a velocity (in units of c) of β, and we take $\mathbf{p}$ to be the photon momentum in the frame at rest in the cosmic microwave background and $\mathbf{p}'$ to be the photon momentum measured on the earth, then

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ |\mathbf{p}| \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \gamma & \beta\gamma \\ 0 & 0 & \beta\gamma & \gamma \end{pmatrix} \begin{pmatrix} p'_1 \\ p'_2 \\ p'_3 \\ |\mathbf{p}'| \end{pmatrix} \tag{2.4.3}$$

where as usual $\gamma \equiv (1-\beta^2)^{-1/2}$. In particular

$$|\mathbf{p}| = \gamma(1 + \beta\cos\theta)\,|\mathbf{p}'| \tag{2.4.4}$$

where θ is the angle between $\mathbf{p}'$ and the three-axis. Thus

$$N'_\gamma(\mathbf{p}') = \frac{1}{h^3}\frac{1}{\exp\left(|\mathbf{p}'|\,c/kT'\right) - 1} \tag{2.4.5}$$

where the temperature is a function of the angle between the direction of the photon and the earth's velocity

$$T' = \frac{T}{\gamma(1 + \beta\cos\theta)} \tag{2.4.6}$$

Since the galaxy can be expected to be moving at a velocity of several hundred kilometers per second, comparable to the peculiar velocities observed for other galaxies relative to the mean Hubble flow, and the solar system is moving with a similar velocity within the galaxy, we expect β to be roughly of order 10^{-3}, in which case γ is essentially unity. The apparent temperature is greatest if we observe photons coming from the direction toward which the earth is moving, for which $\cos\theta = -1$, where it is greater than the intrinsic temperature by a fractional amount β_{earth} . It is least if we observe photons moving in the same direction as the earth, for which $\cos\theta = +1$, and the temperature is decreased by the same fractional amount.

This effect was first observed in 1969 with a ground-based radiometer, but at the time it was only possible to measure the component of the earth's velocity in the earth's equatorial plane, found to be 350 km/sec in a direction corresponding to right ascension 11 h20 m.[1] The full velocity vector of the earth was measured in 1977 by a Berkeley group, [2] using measurements from a U2 aircraft flying above most of the earth's atmosphere. Our knowledge of this effect has been greatly improved by measurements from the COBE satellite. The Far Infrared Absolute Spectrophotometer group [3] found a maximum temperature increase ΔT of 3.372 ± 0.014mK (95% confidence level) in a direction with galactic coordinates [4] $\ell = 264^\circ.14 \pm 0^\circ.30, b = 48^\circ.26 \pm 0^\circ.30$; the Differential Microwave Interferometer group [5] found $\Delta T = 3.353 \pm 0.024$mK (95% confidence level) in a direction with galactic coordinates $\ell = 264^\circ.26 \pm 0^\circ.33, b = 48^\circ.22 \pm 0^\circ.13$, corresponding to right ascension $11^{\rm h}12^{\rm m}.2 \pm 0^{\rm m}.8$, declination $-7^\circ.06 \pm 0^\circ.16$.

More recently, the WMAP satellite experiment [6] (discussed in detail in Chapter 7) has given a maximum temperature increase of 3.346 ± 0.017mK in a direction $\ell = 263^\circ.85 \pm 0^\circ.1, b = 48^\circ.25 \pm 0^\circ.04$. These results indicate a motion of the solar system with a velocity $(0.00335)c/(2.725) = 370$ km/sec, not quite in the direction of the Virgo cluster, which has $\ell \approx 284$ and $b \approx 74$. For comparison, the rotation of the galaxy gives the earth a velocity relative to the center of the galaxy of about 215 km/sec, more or less in the opposite direction. Taking this into account gives a net velocity of the local group of galaxies [7] relative to the microwave background of 627 ± 22 km/sec in a direction ($\ell = 276^\circ \pm 3^\circ, b = 30^\circ \pm 3^\circ$) between the Hydra and Centaurus clusters of galaxies.

Expanding Eq. (2.4.6) in powers of β, the temperature shift can be expressed as a sum of Legendre polynomials

$$\Delta T \equiv T' - T = T\left[-\frac{\beta^2}{6} - \beta P_1(\cos\theta) + \frac{2\beta^2}{3}P_2(\cos\theta) + \dots\right] \tag{2.4.7}$$

Because $\beta = 370$ km/sec$/c = 0.0013$ is small, the temperature shift is primarily a dipole, but Eq. (2.4.7) also exhibits a "kinematic quadrupole"

term that is not much smaller than the intrinsic quadrupole term in the temperature, to be discussed in Section 2.6.

10.2.5 2.5 The Sunyaev-Zel'dovich effect

There is another contribution to the anisotropy of the cosmic microwave radiation background, due to the scattering of this radiation by electrons in intergalactic space within clusters of galaxies along the line of sight. This is known as the Sunyaev-Zel'dovich effect. [1] Eq. (C.26) (the Kompaneets equation) of Appendix C shows that scattering of the cosmic microwave background by a non-relativistic [2] electron gas changes the observed photon occupation number $N(\omega)$ (defined so that $4\pi\omega^2 N(\omega)d\omega$ is the number of photons of each of the two polarization states with energy between $\hbar\omega$ and $\hbar(\omega + d\omega)$) at photon energy $\hbar\omega \ll m_e c^2$ at a rate (here in cgs units)

[471] E. K. Conklin, Nature 222, 971 (1969).

[2] G. F. Smoot, M. V. Gorenstein, and R. A. Muller, Phys. Rev. Lett. 39, 898 (1977).

[3] D. J. Fixsen et al., Astrophys. J. 473, 576 (1996).

[4] The galactic coordinate b is the angle between the line of sight and the plane of our galaxy, so that the north galactic pole is at $b = 90^\circ$; the galactic coordinate ℓ is the azimuthal angle around the axis of rotation of our galaxy, with the center of the galaxy at $\ell = 0^\circ$.

[5] C. L. Bennett et al., Astrophys. J. 464, L1 (1996).

[6] C. L. Bennett et al., Astrophys. J. Suppl. 148, 1 (2003).

[7] G. F. Smoot, C. L. Bennett, A. Kogut, J. Aymon, C. Backus et al., Astrophys. J. 371, L1 (1991); A. Kogut, C. Lineweaver, G. F. Smoot, C. L. Bennett, A. Banday et al., Astrophys. J. 419, 1 (1993).

$$\dot{N}(\omega) = \frac{n_e \sigma_{\mathcal{T}}}{m_e c \omega^2} \frac{\partial}{\partial \omega} \left[k_{\mathcal{B}} T_e \omega^4 \frac{\partial N(\omega)}{\partial \omega} + \hbar \omega^4 N(\omega)(1 + N(\omega)) \right],$$

where $\sigma_{\mathcal{T}}$ is the Thomson scattering cross section, n_e is the electron number density, and T_e is the electron temperature. This formula can be used directly in calculating the rate of change of the occupation number $N(\omega)$ of a homogeneous isotropic photon gas through interactions with a MaxwellBoltzmann distribution of electrons at temperature T_e. In the application that concerns us now, we are interested instead in the change of the appearance to us of the cosmic microwave background due to scattering by a cloud of electrons along the line of sight. In this context, we rewrite the Kompaneets equation as

$$\begin{aligned} \frac{\partial}{\partial \ell} N(\omega, \ell) =& \frac{n_e(\ell) \sigma_{\mathcal{T}}}{m_e c^2 \omega^2} \frac{\partial}{\partial \omega} \\ & \times \left[k_{\mathcal{B}} T_e(\ell) \omega^4 \frac{\partial N(\omega, \ell)}{\partial \omega} + \hbar \omega^4 N(\omega, \ell)(1 + N(\omega, \ell)) \right] \end{aligned} \tag{2.5.1}$$

where ℓ is the proper distance coordinate along the line of sight through the cloud.

The ionized plasma in clusters of galaxies is typically at temperatures greater than 10^6 degrees, so that $k_{\mathcal{B}} T_e$ is very much greater than the typical

value of the photon energy $\hbar\omega$, which for the cosmic microwave background is 10^{-4}eV to 10^{-3}eV. In this case Eq. (2.5.1) simplifies to

$$\frac{\partial}{\partial \ell} N(\omega, \ell) = \frac{n_e(\ell) \sigma_{\mathcal{T}} k_{\mathcal{B}} T_e(\ell)}{m_e c^2 \omega^2} \frac{\partial}{\partial \omega} \left[\omega^4 \frac{\partial N(\omega, \ell)}{\partial \omega} \right]. \tag{2.5.2}$$

This is now a linear differential equation, so it can be solved exactly, [3] but in the usual case this is unnecessary, because the cloud of electrons is optically thin. Eq. (2.5.2) then tells us that the change in the occupation number due to passage of the radiation through the cloud is simply

$$\Delta N(\omega) = \frac{y}{\omega^2} \frac{\partial}{\partial \omega} \left[\omega^4 \frac{\partial N(\omega)}{\partial \omega} \right] \tag{2.5.3}$$

where y is the dimensionless parameter

$$y \equiv \frac{\sigma_{\mathcal{T}}}{m_e c^2} \int d\ell n_e(\ell) k_{\mathcal{B}} T_e(\ell) \tag{2.5.4}$$

the integral being taken along the line of sight through the cloud. For black-body radiation at temperature T_γ we have

$$N(\omega) = \frac{1}{\exp\left(\hbar\omega / k_{\mathcal{B}} T_\gamma\right) - 1} \tag{2.5.5}$$

Using this in Eq. (2.5.3) gives

$$\Delta N = y \left(\frac{-x + (x^2/4) \coth(x/2)}{\sinh^2(x/2)} \right), \tag{2.5.6}$$

[471] Ya. B. Zel'dovich and R. A. Sunyaev, Astrophys. Space Sci. 4, 301 (1969); R. A. Sunyaev and Ya. B. Zel'dovich, Comments Astrophys. and Space Physics 2, 66 (1970); 4, 173 (1972).

[2] The ionized gas in galaxy clusters is hot, so that relativistic corrections, though small, are not negligible; see Y. Rephaeli, Astrophys. J. 445, 33 (1995); A. Challinor and A. Lasenby, Astrophys. J. 499, 1 (1998); N. Itoh, U. Kohyama, and S. Nozawa, Astrophys. J. 502, 7 (1998).

where $x \equiv \hbar\omega/k_{\mathcal{B}}T_\gamma$. We see that in general the scattering changes the shape of the photon energy distribution, not just the effective temperature. The characteristic dependence on ω of Eq. (2.5.6) makes it possible in principle for radio astronomers to distinguish between anisotropies due to the Sunyaev-Zel'dovich effect and the primary anisotropies discussed in Section 2.6 and Chapter 7.

But in the Rayleigh-Jeans part of the spectrum, where $x \ll 1$, Eq. (2.5.6) gives $\Delta N \to -2y/x$, while Eq. (2.5.5) gives $N \to 1/x$, so the shape of the spectrum is preserved, with a fractional change in photon temperature equal to

$$\frac{\Delta T_\gamma}{T_\gamma} = \frac{\Delta N}{N} = -2y. \tag{2.5.7}$$

It may seem surprising that scattering of photons by a cloud of electrons with much higher temperature would lower the photon temperature, but this is only in the Rayleigh-Jeans part of the spectrum. Far beyond the peak in the black-body spectrum, where $x \gg 1$, Eqs. (2.5.5) and (2.5.6) give $\Delta N/N$ the positive value yx^2; the cooling observed in the Rayleigh-Jeans region is due to the transfer of photons from low to high energy. Indeed, Eq. (2.5.3) shows that $\int \Delta N(\omega)\omega^2 d\omega$ vanishes, so scattering by the electron cloud preserves the total number of photons we receive.

Eq. (2.5.7) gives the fractional change in photon temperature observed at low frequency due to scattering in a cloud of electrons, typically associated with a cluster of galaxies. If the cluster is at a redshift z, then both T_γ and ΔT_γ are reduced by a factor $(1+z)^{-1}$, but the ratio is independent of z, and independent also of the Hubble constant and other cosmological parameters, except in so far as these quantities are needed in estimating n_e and proper lengths. The electron temperature T_e can be measured from observations of the luminosity of the cluster as a function of photon wavelength, with a result that is independent of the cluster redshift or H_0, but the proper length along the line of sight has to be inferred from the angular size of the cluster and the angular diameter distance d_A, which is inversely proportional to H_0, so the electron density calculated from measurements of the SunyaevZel'dovich parameter y is proportional to the value assumed for H_0.

Measurements of the Sunyaev-Zel'dovich effect can usefully be combined with measurements of the X-ray luminosity of the cluster of galaxies. [4] The cross section [5] $d\sigma$ for the bremsstrahlung of photons of energy between $\hbar\omega$ and $\hbar(\omega + d\omega)$ (with $\hbar\omega$ comparable to but not very close to $m_e v^2/2$) in collisions between non-relativistic electrons of velocity v (with $1/137 \ll v \ll 1$) and atomic nuclei of small velocity is proportional to $v^{-2}d\omega/\omega$; the density of protons equals the density of electrons; and v is typically of order $(k_{\mathcal{B}}T/m_e)^{1/2}$; so the rate at which an antenna of radius R receives bremsstrahlung photons of wavelength λ in this energy range and in the resolution solid angle $(\lambda/R)^2$ from an optically thin plasma cloud with electron temperature T_e and number density n_e is

$$\left[R^2/d_A^2\right]\left[d_A^2(\lambda/R)^2\right]\int d\ell\langle v d\sigma\rangle n_e^2 \propto \lambda^2\left(\frac{d\omega}{\omega}\right)\int d\ell T_e^{-1/2} n_e^2 \tag{2.5.8}$$

Hence the value of n_e inferred from measurements of the X-ray luminosity and the angular size of the cluster is proportional to the assumed value of $H_0^{1/2}$. The requirement that the value of n_e inferred from the

[473] Ya. B. Zel'dovich and I. D. Novikov, Relativistic Astrophysics, Volume 2: The Structure and Evolution of the Universe, transl. L. Fishbone (University of Chicago Press, Chicago, 1983): Eq. (8.7.3).

[474] J. Silk and S. D. M. White, Astrophys. J. 226, L103 (1978); A. Cavaliere, L. Danese, and G. De Zotti, Astron. Astrophys. 75, 322 (1979). For a review, see M. Birkinshaw, Phys. Rep. 310, 97 (1999).

[5] See, e.g., V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, Quantum Electrodynamics, 2nd edition, transl. by J. B. Sykes and J. S. Bell (Pergamon, Oxford, 1982).

X-ray luminosity should agree with the value obtained from the SunyaevZel'dovich effect thus allows an estimate of H_0. For example, study of two galaxy clusters with $z \simeq 0.55$ (with d_A calculated assuming $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$) has given a Hubble constant of $63^{+12}_{-9} \pm 21$ km/sec/Mpc, with the first uncertainty statistical and the second systematic, both with 68% confidence. [6] Any such measurement depends so much on assumptions about the shape of the galaxy cluster and about the clumpiness of the electron distribution, that it seems likely that measurements of the SunyaevZel'dovich effect and X-ray luminosities will be more useful in providing information about galaxy clusters than in fixing the Hubble constant. In one respect, the Sunyaev-Zel'dovich effect is a nuisance - it will interfere with future measurements of the correlation between the primary temperature fluctuations at very small angular separation. [7] But if combined with a model of structure formation, such as discussed in Chapter 8, observations of the Sunyaev-Zel'dovich effect in small angle correlations of the microwave background temperature fluctuations can provide useful cosmological information. [8] Observations of such small-angle correlations [9] have been interpreted as due to the Sunyaev-Zel'dovich effect. [10]

10.2.6 2.6 Primary fluctuations in the microwave background: A first look

In the two previous sections we have considered anisotropies in the cosmic microwave background that arise from effects in the recent universe: the motion of the earth relative to the cosmic microwave background, and the scattering of light by intergalactic electrons in clusters of galaxies along the line of sight. Now we turn to general anisotropies, including the highly revealing primary anisotropies that have their origin in the early universe. [1]

It is convenient to expand the difference $\Delta T(\hat{n})$ between the microwave radiation temperature observed in a direction given by the unit vector $\hat{n}$ and the present mean value T_0 of the temperature in spherical harmonics $Y_\ell^m(\hat{n})$:

$$\Delta T(\hat{n}) \equiv T(\hat{n}) - T_0 = \sum_{\ell m} a_{\ell m} Y_\ell^m(\hat{n}), \quad T_0 \equiv \frac{1}{4\pi} \int d^2\hat{n} T(\hat{n}) \tag{2.6.1}$$

the sum over ℓ running over all positive-definite integers, and the sum over m running over integers from $-\ell$ to ℓ. Since $\Delta T(\hat{n})$ is real, the coefficients $a_{\ell m}$ must satisfy the reality condition

$$a^*_{\ell m} = a_{\ell -m} \tag{2.6.2}$$

(We are defining the spherical harmonics so that $Y_\ell^m(\hat{n})^* = Y_\ell^{-m}(\hat{n})$.) As we saw in Section 2.4, the earth's motion contributes to $\Delta T(\hat{n})$ an anisotropy that to a good approximation is proportional to $P_1(\cos\theta) \propto Y_1^0(\theta, \phi)$ (with the z-axis taken in the direction of the earth's motion), so the main $a_{\ell m}$ produced by this effect is that with $\ell = 1$ and $m = 0$.

The coefficients $a_{\ell m}$ reflect not only what was happening at the time of last scattering, but also the particular position of the earth in the universe. No cosmological theory can tell us this. The quantities of greatest cosmological interest are averages, which may be regarded either as

[476] E. D. Reese et al., Astrophys. J. 533, 38 (2000).

[7] Y. Rephaeli, Astrophys. J. 245, 351 (1981); S. Cole and N. Kaiser, Mon. Not. Roy. Astron. Soc. 233, 637 (1988).

[8] E. Komatsu and T. Kitayam, Astrophys. J. 526, L1 (1999).

[9] B. S. Mason et al., Astrophys. J. 591, 540 (2003).

[10] E. Komatsu and U. Seljak, Mon. Not. Roy. Astron. Soc. 336, 1256 (2002); J. R. Bond et al., Astrophys. J. 626, 12 (2005).

[1] P. J. E. Peebles and J. T. Yu, Astrophys. J. 162, 815 (1970); R. A. Sunyaev and Ya. B. Zel'dovich, Astrophys. Space Sci. 7, 20 (1970); Ya. B. Zel'dovich, Mon. Not. Roy. Astron. Soc. 160, 1 (1972).

averages over the possible positions from which the radiation could be observed, or averages over the historical accidents that produced a particular pattern of fluctuations. The ergodic theorem described in Appendix D shows that, under reasonable assumptions, these two kinds of average are the same. These averages will be denoted $\langle \cdots \rangle$. As discussed in Chapter 10, for anisotropies that arise from quantum fluctuations during inflation, it is these averages over historical accidents that are related to quantum mechanical expectation values. We will return shortly to the question of how to use observations from one position in a universe produced by one specific sequence of accidents to learn about these averages, but first we must introduce some notation.

We assume that the universe is rotationally invariant on the average, so all averages $\langle \Delta T\left(\hat{n}_1\right) \Delta T\left(\hat{n}_2\right) \Delta T$ are rotationally invariant functions of the directions $\hat{n}_1, \hat{n}_2$, etc. In particular, $\langle \Delta T(\hat{n}) \rangle$ is independent of $\hat{n}$, Since $\Delta T(\hat{n})$ is defined as the departure of the temperature from its angular average, its angular average $\int \Delta T(\hat{n}) d^2\hat{n}/4\pi$ vanishes. Averaging over the position of the observer, we have $\int \langle \Delta T(\hat{n}) \rangle d^2\hat{n} = 0$, so since $\langle \Delta T(\hat{n}) \rangle$ is independent of $\hat{n}$, it too vanishes.

The simplest non-trivial quantity characterizing the anisotropies in the microwave background is the average of a product of two ΔT s. Rotational invariance requires that the product of two a s takes the form

$$\langle a_{\ell m} a_{\ell' m'} \rangle = \delta_{\ell \ell'} \delta_{m -m'} C_\ell \tag{2.6.3}$$

for in this case the average of the product of two ΔT s is rotationally invariant:

$$\langle \Delta T(\hat{n}) \Delta T\left(\hat{n}'\right) \rangle = \sum_{\ell m} C_\ell Y_\ell^m(\hat{n}) Y_\ell^{-m}\left(\hat{n}'\right) = \sum_\ell C_\ell \left(\frac{2\ell+1}{4\pi} \right) P_\ell\left(\hat{n} \cdot \hat{n}'\right) \tag{2.6.4}$$

where P_ℓ is the usual Legendre polynomial. Given the left-hand side, we can find C_ℓ by inverting the Legendre transformation

$$C_\ell = \frac{1}{4\pi} \int d^2\hat{n} d^2\hat{n}' P_\ell\left(\hat{n} \cdot \hat{n}'\right) \langle \Delta T(\hat{n}) \Delta T\left(\hat{n}'\right) \rangle \tag{2.6.5}$$

Instead of Eq. (2.6.3), we could equivalently define the multipole coefficients C_ℓ by

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = \delta_{\ell \ell'} \delta_{m m'} C_\ell$$

which shows that the C_ℓ are real and positive. For perturbations ΔT that are Gaussian in the sense described in Appendix E, a knowledge of the C_ℓ tells us all we need to know about averages of all products of ΔT s.

Of course, we cannot average over positions from which to view the microwave background. What is actually observed is a quantity averaged over m but not position:

$$C_\ell^{\rm obs} \equiv \frac{1}{2\ell+1} \sum_m a_{\ell m} a_{\ell -m} = \frac{1}{4\pi} \int d^2\hat{n} d^2\hat{n}' P_\ell\left(\hat{n} \cdot \hat{n}'\right) \Delta T(\hat{n}) \Delta T\left(\hat{n}'\right) \tag{2.6.6}$$

The fractional difference between the cosmologically interesting C_ℓ and the observed $C_\ell^{\rm obs}$ is known as the cosmic variance. Fortunately, for Gaussian perturbations, the mean square cosmic variance decreases with ℓ. The mean square fractional difference is

$$\left\langle \left(\frac{C_\ell - C_\ell^{\rm obs}}{C_\ell} \right)^2 \right\rangle = 1 - 2 + \frac{1}{(2\ell+1)^2 C_\ell^2} \sum_{m m'} \langle a_{\ell m} a_{\ell -m} a_{\ell m'} a_{\ell -m'} \rangle \,. \tag{2.6.7}$$

If ΔT is governed by a Gaussian distribution, then so are its multipole coefficients $a_{\ell m}$ (but not quantities quadratic in the $a_{\ell m}$, such as C_ℓ.) It follows then that [2]

$$\begin{aligned}\langle a_{\ell m}a_{\ell -m}a_{\ell m'}a_{\ell -m'}\rangle = &\langle a_{\ell m}a_{\ell -m}\rangle\langle a_{\ell m'}a_{\ell -m'}\rangle \\ &+\langle a_{\ell m}a_{\ell m'}\rangle\langle a_{\ell -m}a_{\ell -m'}\rangle \\ &+\langle a_{\ell m}a_{\ell -m'}\rangle\langle a_{\ell -m}a_{\ell m'}\rangle \end{aligned} \tag{2.6.8}$$

Using Eq. (2.6.3), we find that the first term on the right-hand side of Eq. (2.6.8) contributes $(2\ell+1)^2C_\ell^2$ to the sum in Eq. (2.6.7), while the second and third terms each contribute $(2\ell+1)C_\ell^2$ to the sum, so that

$$\left\langle \left(\frac{C_\ell - C_\ell^{\rm obs}}{C_\ell}\right)^2 \right\rangle = \frac{2}{2\ell+1} \tag{2.6.9}$$

This sets a limit on the accuracy with which we can measure C_ℓ for small values of ℓ. On the other hand, the same analysis shows that for $\ell \neq \ell'$,

$$\left\langle \left(\frac{C_\ell - C_\ell^{\rm obs}}{C_\ell}\right)\left(\frac{C_{\ell'} - C_{\ell'}^{\rm obs}}{C_{\ell'}}\right) \right\rangle = 0, \tag{2.6.10}$$

so the fluctuations of $C_\ell^{\rm obs}$ away from the smoothly varying quantity C_ℓ are uncorrelated for different values of ℓ. This means that when $C_\ell^{\rm obs}$ is measured for all ℓ in some range $\Delta\ell$ in which C_ℓ actually varies little, the uncertainty due to cosmic variance in the value of C_ℓ obtained in this range is reduced by a factor $1/\sqrt{\Delta\ell}$. Even so, measurements of C_ℓ for $\ell < 5$ probably tell us little about cosmology. Also, measurements for $\ell > 2,000$ are corrupted by foreground effects, such as the Sunyaev-Zel'dovich effect discussed in the previous section. Fortunately there is lots of structure in C_ℓ at values of ℓ between 5 and 2,000 that provides invaluable cosmological information.

The primary anisotropies in the cosmic microwave background arise from several sources:

1. Intrinsic temperature fluctuations in the electron-nucleon-photon plasma at the time of last scattering, [3] at a redshift of about 1,090 .

2. The Doppler effect due to velocity fluctuations in the plasma at last scattering.

3. The gravitational redshift or blueshift due to fluctuations in the gravitational potential at last scattering. This is known as the SachsWolfe effect. [4]

[472] Non-Gaussian terms in the probability distribution of anisotropies would show up as non-vanishing averages of products of odd numbers of the $a_{\ell m}$, as well as corrections to formulas like Eq. (2.6.8) for the averages of products of even numbers of the $a_{\ell m}$. Such non-Gaussian terms are produced both in the early universe and at relatively late times. For a review, see N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, Phys. Rep. 402, 103 (2004). Non-Gaussian terms produced by quantum fluctuations during inflation were calculated in the tree graph approximation by J. Maldacena, J. High Energy Phys.

[47]05 (2003) 013. The effect of loop graphs is considered by S. Weinberg, Phys. Rev. D 72, 043514 (2005) [hep-ph/0506236]; Phys. Rev. D. 74, 023508 (2006) [hep-ph/0605244]; K. Chaicherdsakul, Phys. Rev. D 75, 063522 (2007) [hep-th/0611352]. For late-time contributions, see M. Liguori, F. K. Hansen, E. Komatsu, S. Matarrese, and A. Riotto, Phys. Rev. D 73, 043505 (2006) [astro-ph/0509098]. The weakness of microwave background anisotropies indicates that any non-Gaussian terms are likely to be quite small. So far, there is no observational evidence of such terms.

[3] Strictly speaking, in the approximation of a sudden drop in opacity at a fixed temperature $T_L \simeq 3,000$ K, it is not the intrinsic fluctuation in temperature we observe, but the consequent fluctuation in the redshift z_L of last scattering. Since the unperturbed temperature $\bar{T}(t)$ goes as $1/a(t)$, the value a_L of $a(t)$ at which the total temperature $\bar{T}(t) + \delta T(t)$ reaches a fixed value T_L is shifted by an amount δa_L such that $-(T_L/a_L)\,\delta a_L + \delta T(t_L) = 0$. The observed temperature (leaving aside other effects) is $T_L a(t_L)/a_0$, so to first order this is shifted by a fractional amount $T_L\delta a_L/a_0T_0 = \delta T(t_L)\,a_L/a_0T_0 = \delta T(t_L)/T_L$, just as if it were the intrinsic temperature fluctuation that we observe.

4. Gravitational redshifts or blueshifts due to time-dependent fluctuations in the gravitational potential between the time of last scattering and the present. (It is necessary that the fluctuations be time-dependent; a photon falling into a time-independent potential well will lose the energy it gains when it climbs out of it.) This is known, somewhat confusingly, as the integrated Sachs-Wolfe effect. [4]

A proper treatment of these effects requires the use of general relativity. This will be the subject of Chapters 5-7. On the other hand, from the time the temperature dropped below 10^4 K until vacuum energy became important at a redshift of order unity, the gravitational field of the universe was dominated to a fair approximation by cold dark matter, which can be treated by the methods of Newtonian physics. Therefore in this introductory section we will concentrate on the Sachs-Wolfe and integrated Sachs-Wolfe effect, which turn out to dominate the multipole coefficients C_ℓ for relatively small ℓ, less than about 40 . We will make only a few tentative remarks in this section about the contribution of intrinsic temperature fluctuations and of the Doppler effect.

In considering the Sachs-Wolfe and integrated Sachs-Wolfe effects, we return to the Newtonian approach to cosmology outlined in Eqs. (1.5.21)(1.5.27). The treatment of perturbations to a homogeneous isotropic cosmology in this approach is presented in Appendix F. For the moment, we need only one result of this analysis, that the perturbation to the gravitational potential, when expressed as a function of the co-moving coordinate $\mathbf{x}$, is a time-independent function $\delta\phi(\mathbf{x})$. This perturbation has two effects. First, there is the usual gravitational redshift: A photon emitted at a point $\mathbf{x}$ at the time of last scattering will have its frequency and hence its energy shifted by a fractional amount $\delta\phi(\mathbf{x})$, so the temperature seen when we look in a direction $\hat{n}$ will be shifted from the average over the whole sky by an amount

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)_1 = \delta\phi\left(\hat{n}r_L\right) \tag{2.6.11}$$

Here r_L is the radial coordinate of the surface of last scattering, given by Eq. (1.5.44) as

$$r_L = \frac{1}{\Omega_K^{1/2} H_0 a\left(t_0\right)} \sinh\left[\Omega_K^{1/2}\int_{1/(1+z_L)}^{1}\frac{dx}{\sqrt{\Omega_\Lambda x^4+\Omega_K x^2+\Omega_M x+\Omega_R}}\right] \tag{2.6.12}$$

where $\Omega_K = 1-\Omega_\Lambda-\Omega_M-\Omega_R$; $z_L \simeq 1090$ is the redshift of last scattering; and t_0 is the present. The perturbation to the gravitational potential also has the effect of changing the rate at which the universe expands by a fractional amount $\delta\phi(\mathbf{x})$, and since the temperature in a matter-dominated universe is falling like $a^{-1} \propto t^{-2/3}$, this shifts the value of the redshift at which the universe reaches the temperature $\simeq 3,000$ K of last scattering in direction $\hat{n}$ by a fractional amount

$$\left(\frac{\delta z}{1+z}\right)_{T=3000\text{ K}} = -\left(\frac{\delta a(t)}{a(t)}\right)_{T=3000\text{ K}} = \left(\frac{\dot{a}}{a}\right)_{T=3000\text{ K}}$$
$$\delta\phi\left(r_L\hat{n}\right)t_L = \frac{2}{3}\delta\phi\left(r_L\hat{n}\right)$$

With all other effects neglected, the temperature observed in direction $\hat{n}$ would be $3,000$ K divided by $1+z$, so this fractional shift in $1+z$ changes the observed temperature by a fractional amount

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)_2 = -\frac{2}{3}\delta\phi\left(\hat{n}r_L\right) \tag{2.6.13}$$

The sum of the fractional shifts (2.6.11) and (2.6.13) gives the Sachs-Wolfe effect:

[474] R. K. Sachs and A. M. Wolfe, Astrophys. J. 147, 73 (1967).

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)_{\rm SW} = \frac{1}{3}\delta\phi\left(\hat{n}r_L\right) \tag{2.6.14}$$

The factor 1/3 will be obtained in Chapter 7 as a result of a better-grounded relativistic treatment.

It is convenient to write $\delta\phi(\mathbf{x})$ as a Fourier transform

$$\delta\phi(\mathbf{x}) = \int d^3q e^{i\mathbf{q}\cdot\mathbf{x}}\delta\phi_{\mathbf{q}} \tag{2.6.15}$$

We make use of the well-known Legendre expansion of the exponential

$$e^{i\mathbf{q}\cdot\mathbf{x}} = \sum_\ell (2\ell+1) i^\ell P_\ell(\hat{q}\cdot\hat{n}) j_\ell(qr) \tag{2.6.16}$$

where j_ℓ is the spherical Bessel function, defined in terms of the usual Bessel function $J_\nu(z)$ by $j_\ell(z) \equiv (\pi/2z)^{1/2} J_{\ell+1/2}(z)$. Eq. (2.6.14) then
gives

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)_{\rm SW} = \frac{1}{3}\sum_{\ell=0}^{\infty}(2\ell+1)i^\ell \int d^3q \delta\phi_{\mathbf{q}} j_\ell\left(qr_L\right) P_\ell(\hat{q}\cdot\hat{n}) \tag{2.6.17}$$

Now we must consider how to calculate the average of a product of two of these fractional temperature shifts in two different directions. Although $\delta\phi(\mathbf{x})$ depends on position, the probability distribution of $\delta\phi(\mathbf{x})$ as seen by observers in different parts of the universe is invariant under spatial rotations and translations. This implies among other things that

$$\langle \delta\phi_{\mathbf{q}} \delta\phi_{\mathbf{q}'}\rangle = \mathcal{P}_\phi(q)\delta^3\left(\mathbf{q}+\mathbf{q}'\right) \tag{2.6.18}$$

where $\mathcal{P}_\phi(q)$ is a function only of the magnitude of $\mathbf{q}$. (The delta function is needed so that $\langle\delta\phi(\mathbf{x})\delta\phi(\mathbf{y})\rangle$ should be only a function of $\mathbf{x}-\mathbf{y}$.) Because $\delta\phi(\mathbf{x})$ is real, its Fourier transform satisfies the reality condition $\delta\phi^*_{\mathbf{q}} = \delta\phi_{-\mathbf{q}}$, which together with Eq. (2.6.18) tells us that $\mathcal{P}_\phi(q)$ is real and positive.

Now, using Eqs. (2.6.17) and (2.6.18), together with the reflection property $P_\ell(-z) = (-1)^\ell P_\ell(z)$ and the orthogonality property of Legendre polynomials

$$\int d\Omega_{\hat{q}} P_\ell(\hat{n}\cdot\hat{q}) P_{\ell'}\left(\hat{n}'\cdot\hat{q}\right) = \left(\frac{4\pi}{2\ell+1}\right)\delta_{\ell\ell'} P_\ell\left(\hat{n}\cdot\hat{n}'\right) \tag{2.6.19}$$

we have
$\langle\Delta T(\hat{n})\Delta T\left(\hat{n}'\right)\rangle_{\rm SW} = \frac{4\pi T_0^2}{9}\sum_\ell (2\ell+1) P_\ell\left(\hat{n}\cdot\hat{n}'\right)\int_0^\infty q^2 dq \mathcal{P}_\phi(q) j_\ell^2\left(qr_L\right)$,
or, comparing with Eq. (2.6.4),

$$C_{\ell,\rm SW} = \frac{16\pi^2 T_0^2}{9}\int_0^\infty q^2 dq \mathcal{P}_\phi(q) j_\ell^2\left(qr_L\right) \tag{2.6.21}$$

To the extent that the gravitational potential is produced by pressureless cold dark matter, the differential equation for $\delta\phi$ does not involve gradients, and so the differential equation for its Fourier transform does not involve the wave vector $\mathbf{q}$. (See Eqs. (F.12) and (F.18).) The dependence of $\delta\phi_{\mathbf{q}}$ on $\boldsymbol{q}$ then can arise only from the initial conditions for these differential equations. [5] It is therefore natural to try the hypothesis that the function

[475] This is not true even for the Sachs-Wolfe effect if q is so large that q/a became greater than H before the density of the universe became dominated by cold matter. As we will see in Chapter 7, this qualification affects C_ℓ SW only for $\ell > 100$.

$\mathcal{P}_\phi(q)$ has a simple form, such as a power of q. This power is conventionally written as $n-4$:

$$\mathcal{P}_\phi(q) = N_\phi^2 q^{n-4} \tag{2.6.22}$$

where N_ϕ^2 is a positive constant. Then we can use a standard formula:

$$\int_0^\infty j_\ell^2(s) s^{n-2} ds = \frac{2^{n-4}\pi\Gamma(3-n)\Gamma\left(\ell+\frac{n-1}{2}\right)}{\Gamma^2\left(\frac{4-n}{2}\right)\Gamma\left(\ell+2-\frac{n-1}{2}\right)} \tag{2.6.23}$$

and find that for $\ell < 100$, Eq. (2.6.21) gives

$$C_{\ell \mathrm{SW}} \to \frac{16\pi^3 2^{n-4}\Gamma(3-n) r_L^{1-n} N_\phi^2 T_0^2}{9\Gamma^2\left(\frac{4-n}{2}\right)} \frac{\Gamma\left(\ell+\frac{n-1}{2}\right)}{\Gamma\left(\ell+2-\frac{n-1}{2}\right)}. \tag{2.6.24}$$

In particular, even before the discovery of primary fluctuations in the cosmic microwave background there was a general expectation about the values of n and N, based not on the microwave background, but on the large scale structure of matter observed relatively close to the present. The perturbation $\delta\rho$ in the total mass density is related to the Sachs-Wolfe effect through the Poisson equation, which gives

$$a^{-2}\nabla^2\delta\phi = 4\pi G\delta\rho \tag{2.6.25}$$

with the factor $a^{-2}(t)$ inserted because it is $\mathbf{X} = a(t)\mathbf{x}$ that measures proper distances. (See Eq. (F.12).) Expressing the Fourier transform of $\delta\rho$ in terms of the Fourier transform of $\delta\phi$, we find the correlation function of the density fluctuations to be

$$\langle \delta\rho(\mathbf{x},t)\delta\rho\left(\mathbf{x}',t'\right)\rangle = \left(4\pi G a(t) a\left(t'\right)\right)^{-2} \int d^3q q^4 \mathcal{P}_\phi(q) e^{i\mathbf{q}\cdot(\mathbf{x}-\mathbf{x}')} \tag{2.6.26}$$

The use of this formula to measure $\mathcal{P}_\phi$ is discussed in Chapter 8. For the present, it is enough to note that observations of the density correlation function long ago led to the expectation that $\mathcal{P}_\phi$ takes the so-called Harrison-Zel'dovich form [6] with $n = 1$, and that $N_\phi \approx 10^{-5}$. As we will see in Chapter 10, inflationary theories generally predict that n is close but not

precisely equal to unity. For $n = 1$, we obtain a result that is scale-invariant, in the sense of being independent of r_L :

$$C_{\ell,\mathrm{SW}} \to \frac{8\pi^2 N_\phi^2 T_0^2}{9\ell(\ell+1)} \tag{2.6.27}$$

This is why experimental data on the cosmic microwave background anisotropies is usually presented as a plot of $\ell(\ell+1)C_\ell$ versus ℓ.

What about the other contributions to C_ℓ ? Pressure gradients are important in the dynamics of the photon-nucleon-electron plasma, so in estimating the contributions of the Doppler effect and intrinsic temperature fluctuations we must deal with differential equations in which the wave number q enters in an important way. Whatever sort of perturbation we are considering, we can always write it as a Fourier integral like Eq. (2.6.15) and use Eq. (2.6.16) to express $e^{i\mathbf{q}\cdot\hat{n}r_L}$ as a series of Legendre polynomials $P_\ell(\hat{q}\cdot\hat{n})$ with coefficients proportional to $j_\ell\left(qr_L\right)$. The integral over q is then dominated by values of q of order ℓ/r_L in the case $\ell \gg 1$. (This is the most interesting case because Eq. (2.6.9) shows that it is only for $\ell \gg 1$ that cosmic variance can be neglected in measurements of C_ℓ.) This is because for $\ell \gg 1$, the spherical

[476] E. R. Harrison, Phys. Rev. D1, 2726 (1970); Ya. B. Zel'dovich, Mon. Not. Roy. Astron. Soc. 160, 1P (1972).

Bessel function $j_\ell(z)$ is peaked at $z \simeq \ell$. Specifically, for $v \equiv \ell + 1/2 \to \infty$ and $z \to \infty$ with v/z fixed at a value other than unity, we have

$$j_\ell(z) \to \begin{cases} 0 & z < v \\ z^{-1/2}\left(z^2 - v^2\right)^{-1/4} \cos\left(\sqrt{z^2 - v^2} - v \arccos(v/z) - \frac{\pi}{4}\right) & z > v. \end{cases} \tag{2.6.28}$$

Hence C_ℓ for large ℓ chiefly reflects the behavior of the Fourier components of perturbations for $q \approx \ell/r_L$. To put this another way, the physical wave number at time t_L is $\mathbf{k}_L \equiv \mathbf{q}/a\left(t_L\right)$ (because it is $a\left(t_L\right)\mathbf{x}$ that measures proper distances at this time) so C_ℓ for large ℓ reflects the behavior of the perturbations for $k_L \approx \ell/d_A$, where d_A is the angular diameter distance of the surface of last scattering

$$\begin{aligned} d_A \equiv r_L a\left(t_L\right) = & \frac{1}{\Omega_K^{1/2} H_0\left(1 + z_L\right)} \\ & \times \sinh\left[\Omega_K^{1/2} \int_{\frac{1}{1+z_L}}^1 \frac{dx}{\sqrt{\Omega_\Lambda x^4 + \Omega_K x^2 + \Omega_M x + \Omega_R}}\right] \end{aligned} \tag{2.6.29}$$

with $1 + z_L \equiv a\left(t_0\right)/a\left(t_L\right)$.

For physical wave numbers q/a that are much less than the expansion rate H the differential equation governing any one perturbation is essentially independent of the wave number q, so that the whole dependence of the perturbation on q comes from the initial conditions. (Ratios of different perturbations, such as $\delta\phi_q$ and $\delta\rho_q$, may of course depend on q.) Such perturbations are said to be "outside the horizon" because the wavelength $2\pi a/q$ is much larger than the horizon distance (strictly speaking, the "particle horizon" distance), which we saw in Section 1.13 is roughly of order $1/H$. During the radiation- or matter-dominated eras q/a decreased like $t^{-1/2}$ or $t^{-2/3}$, while H decreased faster, like $1/t$, so perturbations that were outside the horizon at early times subsequently came within the horizon, those with high wave number entering the horizon earlier than those with lower wave numbers.

We need to be a little more precise about the horizon distance. At the time of last scattering the universe was largely matter dominated, so as shown in Section 1.13 the horizon distance at that time was approximately $2/H\left(t_L\right)$. But this is the maximum proper distance that light could have traveled since the beginning of the present phase of the expansion of the universe. As we will see in Chapters 5 and 6, the dominant perturbations to the plasma of nucleons, electrons, and photons that are relevant to the anisotropies in the cosmic microwave background are sound waves, so we need to take into account the smaller speed of sound. During the era before recombination, when radiation and matter were in thermal equilibrium, the speed of sound was $v_s = (\delta p/\delta\rho)^{1/2}$, where δp and $\delta\rho$ are infinitesimal variations in the pressure and density in an adiabatic fluctuation. In such a perturbation, the entropy per baryon remains unperturbed:

$$0 = \delta\sigma \propto \delta\left(\frac{\epsilon}{n_B}\right) + p\delta\left(\frac{1}{n_B}\right),$$

where ϵ is the thermal energy density, n_B is the baryon number density, and p is the pressure. As we saw in Section 2.2, there are so many more photons than baryons that both ϵ and p are dominated by radiation:

$$\epsilon = a_{\mathcal{B}} T^4, \quad p = \frac{1}{3} a_{\mathcal{B}} T^4,$$

and therefore for adiabatic perturbations

$$3\frac{\delta T}{T} = \frac{\delta n_B}{n_B} = \frac{\delta \rho_B}{\rho_B}, \tag{2.6.30}$$

where ρ_B is the baryonic mass density. This gives a sound speed

$$v_s = \left(\frac{4a_{\mathcal{B}}T^3\delta T/3}{\delta\rho_B + 4a_{\mathcal{B}}T^3\delta T}\right)^{1/2} = \frac{1}{\sqrt{3(1+R)}}, \tag{2.6.31}$$

where $R \equiv 3\rho_B/4a_{\mathcal{B}}T^4$. Since $\rho_B \propto a^{-3}$ and $T \propto a^{-1}$, we have $R \propto a$, and hence $dt = dR/HR$, or in more detail,

$$dt = \frac{dR}{RH_0\sqrt{\Omega_M\left(R_0/R\right)^3 + \Omega_R\left(R_0/R\right)^4}} = \frac{RdR}{H_0\sqrt{\Omega_M}R_0^{3/2}\sqrt{R_{\rm EQ}+R}}$$

where $R_{\rm EQ} \equiv \Omega_R R_0/\Omega_M = 3\Omega_R\Omega_B/4\Omega_M\Omega_\gamma$ is the value of R at matterradiation equality. The acoustic horizon distance is then

$$\begin{aligned} d_H &\equiv a_L\int_0^{t_L}\frac{dt}{a\sqrt{3(1+R)}} = R_L\int_0^{t_L}\frac{dt}{R\sqrt{3(1+R)}} \\ &= \frac{2}{H_0\sqrt{3R_L\Omega_M}\left(1+z_L\right)^{3/2}}\ln\left(\frac{\sqrt{1+R_L}+\sqrt{R_{\rm EQ}+R_L}}{1+\sqrt{R_{\rm EQ}}}\right) \end{aligned} \tag{2.6.32}$$

where $R_L = 3\Omega_B/4\Omega_\gamma\left(1+z_L\right)$ is the value of R at last scattering. Gradients become important when $k_L \approx 1/d_H$, and since in C_ℓ the integral over wave number is dominated by $k_L \approx \ell/d_A$, gradients become important when ℓ reaches the value ℓ_H, where $\ell_H \equiv d_A/d_H$. Both d_A and d_H are proportional to $1/H_0$, so H_0 cancels in the ratio.

For a crude estimate of ℓ_H, we note that d_A is proportional to $\left(1+z_L\right)^{-1}$, while d_H is proportional to $\left(1+z_L\right)^{-3/2}$, so ℓ_H is of the order of $\left(1+z_L\right)^{1/2} = 33$. To get a closer estimate, we will take cosmological parameters to have sample values suggested by supernova observations and cosmological nucleosynthesis (discussed in Section 3.2): $\Omega_M = 0.26, \Omega_\Lambda = 0.74, \Omega_B = 0.043$. This gives $R_L = 0.62, R_{\rm EQ} = 0.21, d_A = 3.38H_0^{-1}\left(1+z_L\right)^{-1}$ and $d_H = 1.16H_0^{-1}\left(1+z_L\right)^{-3/2}$, and hence $\ell_H = 2.9\sqrt{1+z_L} = 96$.

We can now estimate the relative magnitude of the contributions to C_ℓ other than the Sachs-Wolfe effect:

- Doppler effect: Like any vector field, the plasma velocity can be decomposed into a term given by the gradient of a scalar, plus a "vector" term whose divergence vanishes. Appendix F shows that the vector term decays as $1/a$, so the dominant perturbations are compressional modes, for which the velocity is the gradient of a scalar. We can therefore expect that in the integral over wave numbers q for $\Delta T/T_0$, the contribution of the Doppler effect will be suppressed for small wave numbers by a factor of order $\mathbf{k}_L d_H$. We will see in Chapter 7 that, because it is proportional to the vector $\mathbf{k}_L$, the Doppler effect contribution does not interfere with the contribution of the SachsWolfe effect, so for a multipole order $\ell \ll \ell_H$, the contribution of the Doppler effect will be less than that of the Sachs-Wolfe effect by a factor of order $\left[\left(\ell/d_A\right)d_H\right]^2 = \ell^2/\ell_H^2$.

- Intrinsic temperature fluctuations: As we have seen in Eq. (2.6.30) the intrinsic fractional temperature fluctuation at the time of last scattering will be just one-third the intrinsic fractional perturbation in the plasma density. As discussed in Chapters 5 and 6, the particular perturbations that are believed to dominate outside the horizon are adiabatic

in the further sense that the fractional perturbation in the plasma density is equal to the fractional perturbation $\delta\rho/\bar{\rho}$ in the total matter density. But the perturbation to the total matter density is related to the perturbation to the gravitational potential by Poisson's equation (2.6.25), which if evaluated at the time of last scattering gives for the Fourier transform:

$$\delta\rho_{\mathbf{q}}(t_L) = -\frac{q^2}{4\pi G a^2(t_L)}\delta\phi_{\mathbf{q}} = -\frac{k_L^2}{4\pi G}\delta\phi_{\mathbf{q}}$$

where, as before, $k = q/a(t_L)$. Also, the mean total mass density $\bar{\rho}(t_L)$ at last scattering is related to the horizon size d_H by

$$\bar{\rho}(t_L) \simeq \frac{3H^2(t_L)}{8\pi G} \approx \frac{1}{2\pi G d_H^2}$$

so the order of magnitude of the intrinsic fractional temperature perturbation is related to the perturbation to the gravitational potential by

$$\frac{\delta T(t_L)}{\bar{T}(t_L)} = \frac{\delta\rho(t_L)}{3\bar{\rho}(t_L)} \approx k_L^2 d_H^2 \delta\phi_q \tag{2.6.33}$$

Thus we expect that in the integral over wave numbers q for $\Delta T/T_0$, the contribution of intrinsic temperature fluctuations will be suppressed for small wave numbers by a factor $\mathbf{k}_L^2 d_H^2$. The interference of this contribution with the Sachs-Wolfe term then makes a contribution to C_ℓ for $\ell \ll \ell_{\text{horizon}}$ that, like the contribution of the Doppler effect, is smaller than the Sachs-Wolfe contribution by a factor of order $[(\ell/d_A)\, d_H]^2 \approx \ell^2/\ell^2_{\text{horizon}}$. (It should be noted that the distinction between the Sachs-Wolfe effect and the effect of intrinsic temperature fluctuations depends on how the time coordinate is defined. The nonrelativistic estimates made here correspond to what in the relativistic treatment of Chapters $5-7$ would be the use of what are called Newtonian gauge coordinates.)

- Integrated Sachs-Wolfe effect: As already mentioned, to the extent that $\delta\phi$ is truly time independent, there is no integrated Sachs-Wolfe effect. The blueshift caused by a photon falling into a time-independent gravitational potential well along the line of sight would be canceled by
  the redshift caused when the photon climbs out of the well. [7] In fact, the perturbation to the gravitational field is not strictly time-independent, both because radiation continues to make a non-negligible contribution to the gravitational field for some time after last scattering, and also because at late times vacuum energy requires modifications to the Newtonian treatment presented in Appendix F. The late-time integrated Sachs-Wolfe effect chiefly affects C_ℓ for ℓ less than about 10 .

It is reasonable then to assume that for $10 < \ell < 50$, the dominant contribution to C_ℓ is from the Sachs-Wolfe effect.

Apart from observation of the $\ell = 1$ anisotropy due to the earth's motion through the microwave background, the first detection of an anisotropy in the cosmic microwave radiation background was achieved by the COBE satellite in 1992. [8] This experiment scanned the sky with two microwave antennae separated by 60°, at frequencies $31.5, 53$, and 90 GHz near the minimum of emission from our galaxy and the maximum of the Planck distribution for 2.7 K . The 1992 data showed an rms fluctuation in the temperature with angle of $30 \pm 5\mu$ K, with an angular distribution consistent with $n = 1$. After four years, values of C_ℓ had been measured [9] with the same instruments at values of ℓ ranging from $\ell = 2$ to $\ell = 40$. For $\ell \geq 4$ the results

were fit to the ℓ-dependence given by Eq. (2.6.24), with the result that $n = 1.13^{+0.3}_{-0.4}$, which is consistent with the value $n = 1$ for what is called a Harrison-Zel'dovich spectrum. This result is often written as

$$C_\ell = \frac{24\pi Q^2}{5\ell(\ell+1)} \tag{2.6.34}$$

with Q known as the quadrupole moment. Fitting the values of C_ℓ for $10 \leq \ell \leq 40$, the 1996 COBE results gave $Q = 18 \pm 1.4\mu$ K. Comparing Eq. (2.6.34) with (2.6.27), we see that

$$N_\phi = \sqrt{\frac{27}{5\pi}}\frac{Q}{T_0} = (8.7 \pm 0.7) \times 10^{-6} \tag{2.6.35}$$

Surprisingly, the multipole coefficients for $\ell = 2$ and $\ell = 3$ were found to be much less than would be expected by extrapolation of Eq. (2.6.34) from

the fit for $\ell \geq 4$, [10] as shown by the apparent absence of two-point correlations for angles greater than about 60°. The discrepancy with theory is even worse when the integrated Sachs-Wolfe effect is taken into account. This result has since been confirmed by observations with the WMAP satellite, [11] to be discussed in Chapter 7. It is quite possible that this discrepancy is due to a combination of foreground contamination and cosmic variance, [12] which according to Eq. (2.6.9) is 63% for $\ell = 2$.

We will have to wait to discuss calculations of C_ℓ for large ℓ, above the range of validity of Eq. (2.6.34), until Chapter 7, after we have developed the general relativistic theory of cosmological fluctuations in Chapters 5 and 6. In Chapter 7 we will also come back to the observations over the past decade that have refined the COBE measurements and extended them to higher ℓ.

10.3 The Early Universe

We have been exploring the period in the history of the universe when the radiation temperature dropped from a little over 10^4 K down to its present value of 2.725 K . We now want to look back to the era when the temperature was greater than about 10^4 K, well before the energy density in radiation fell below that of baryons and cold dark matter. We will carry the story back to when the temperature was above 10^{10} K, when electron-positron pairs were abundant and neutrinos were in equilibrium with these pairs, and even farther back, as far as our current knowledge of the laws of physics will take us.

[477] There is also an anisotropy produced by time-dependent fluctuations in the gravitational potential of cosmological inhomogeneities (such as concentrations of cold dark matter), known as the Rees-Sciama effect; M. J. Rees and D. W. Sciama, Nature 217, 511 (1968). This is expected to be quite small; see U. Seljak, Astrophys. J. 460, 549 (1996). For possible larger effects due to local structures, see A. Rakić, Syksy Räsänen, and D. J. Schwarz, Mon. Not. Roy. Astron. Soc. 363, L27 (2006) [astro-ph/0601445].

[8] G. F. Smoot et al., Astrophys. J. 396, L1 (1992)

[9] C. L. Bennett et al., Astrophys. J. 464, L1 (1996) [astro-ph/9601067]; A. Kogut et al., ibid, L5 (1996) [astro-ph/9601066]; K. M. Górski et al., ibid, L11 (1996) [astro-ph/9601063]; G. Hinshaw et al., ibid, L17 (1996) [astro-ph/9601088]: E. L. Wright et al., ibid, L21 (1996) [astro-ph/9601059].

[4710] G. Hinshaw et al., Astrophys. J. 464, L25 (1996) [astro-ph/9601061].

[11] D. N. Spergel et al., Astrophys. J. Suppl. 148, 175 (2003).

[12] G. Efstathiou, Mon. Not. Roy. Astron. Soc. 346, L26 (2003) [astro-ph/0306431]; A. Slosar, U. Seljak, and A. Makarov, Phys. Rev. D 69, 123003 (2004) [astro-ph/0403073]; A. Slosar and U. Seljak, Phys. Rev. D 70, 083002 (2004); G. Hinshaw et al., Astrophys. J. Suppl. Ser. 170, 288 (2007).

10.3.1 3.1 Thermal history

We first want to work out the history of the falling temperature of the early universe. In this section we will look back only to a time when the temperature was between 10^4 K and 10^{11} K, which is low enough so that muon-antimuon and hadron-antihadron pairs were no longer being produced in appreciable numbers.

There are two circumstances that greatly simplify this task. The first is that the collision rate of photons with electrons and other charged particles during this era was so much greater than the expansion rate of the universe that the photons and charged particles can be assumed to have been in thermal equilibrium, with a common falling temperature. At sufficiently early times even the neutrinos and perhaps the cold dark matter particles were also in thermal equilibrium with the photons and charged particles; later, when no longer colliding rapidly with other particles, they can be treated separately as free particles. The other circumstance is that the number density of baryons (or more strictly, the number density of baryons minus the number density of antibaryons) is so much less than the number density of photons that we can ignore the chemical potential associated with baryon number. Baryons will be put back into the picture in the following section. Also, because the electron/photon number ratio is so small now, it is reasonable to assume that the universe has always had a very small net lepton number density (the number density of leptons of all sorts minus that of antileptons) per photon. This means that even at temperatures of order 10^{10} K, when electron-positron pairs were abundant and the energy density and pressure were not simply proportional to T^4 and the entropy density was not simply proportional to T^3, the entropy density, energy density, and pressure were functions $s(T), \rho(T)$, and $p(T)$ of the temperature alone.
(The possibility of a non-negligible net lepton number is discussed at the end of this section.)

Before studying the history of the universe during this era we will have to take a brief look at the thermodynamics and statistical mechanics of this sort of matter, in thermal equilibrium with negligible chemical potentials. The condition of thermal equilibrium tells us that the entropy in a co-moving volume is fixed

$$s(T)a^3 = \text{ constant} \tag{3.1.1}$$

The second law of thermodynamics says that any adiabatic change in a system of volume V produces a change in the entropy given by

$$d(s(T)V) = \frac{d(\rho(T)V) + p(T)dV}{T} \tag{3.1.2}$$

Equating the coefficients of dV gives our formula for the entropy density

$$s(T) = \frac{\rho(T) + p(T)}{T} \tag{3.1.3}$$

Also, equating the coefficients of VdT and using Eq. (3.1.3) give the law of conservation of energy:

$$T\frac{dp(T)}{dT} = \rho(T) + p(T) \tag{3.1.4}$$

(For instance, for radiation we have $p(T) = \rho(T)/3$, so Eq. (3.1.4) gives the Stefan-Boltzmann law $\rho = a_{\mathcal{B}}T^4$ with $a_{\mathcal{B}}$ a constant that cannot be determined from thermodynamics alone; Eq. (3.1.3) then gives the entropy density for radiation as $s(T) = 4a_{\mathcal{B}}T^3/3$. This is why in we said in Section 2.2 that the constant $\sigma k_{\mathcal{B}} \equiv 4a_{\mathcal{B}}T^3/3n_B$ may be interpreted as the radiation entropy per baryon.)

With equal numbers of particles and antiparticles, the number density $n(p)dp$ of a species of fermions (such as electrons) or bosons (like photons) of mass m and momentum between p and $p+dp$ is given by the Fermi-Dirac or Bose-Einstein distributions (with zero chemical potential)

$$n(p,T) = \frac{4\pi g p^2}{(2\pi\hbar)^3}\left(\frac{1}{\exp\left(\sqrt{p^2+m^2}/k_{\mathcal{B}}T\right)\pm 1}\right) \tag{3.1.5}$$

where g is the number of spin states of the particle and antiparticle, and the sign is + for fermions and - for bosons. For instance, for photons $g = 2$ (and of course $m = 0$), because photons have two polarization states and they are their own antiparticles, while for electrons and positrons $g = 4$, because they have two spin states and electrons and positrons are distinct particle species. The energy density and pressure of a particle of mass m are given by the integrals [1]

$$\rho(T) = \int_0^\infty n(p,T)dp\sqrt{p^2+m^2} \tag{3.1.6}$$

$$p(T) = \int_0^\infty n(p,T)dp\frac{p^2}{3\sqrt{p^2+m^2}} \tag{3.1.7}$$

The entropy density of this particle is then given by Eq. (3.1.3) as

$$s(T) = \frac{1}{T}\int_0^\infty n(p,T)dp\left[\sqrt{p^2+m^2}+\frac{p^2}{3\sqrt{p^2+m^2}}\right] \tag{3.1.8}$$

In particular, for massless particles Eq. (3.1.6) gives

$$\begin{aligned}\rho(T) &= g\int_0^\infty \frac{4\pi p^3 dp}{(2\pi\hbar)^3}\left(\frac{1}{\exp\left(p/k_{\mathcal{B}}T\right)\pm 1}\right)\\ &= \begin{cases} g a_{\mathcal{B}} T^4/2 & \text{bosons}\\ 7g a_{\mathcal{B}} T^4/16 & \text{fermions}\end{cases}\end{aligned} \tag{3.1.9}$$

and of course $p(T) = \rho(T)/3$ and $s(T) = 4\rho(T)/3T$. In other words, each species and spin of massless fermions makes a contribution to the energy density, pressure, and entropy density that is just the same as for each polarization state of photons, except for an additional factor 7/8.

During a period of thermal equilibrium, the variation with time of the temperature is governed by the equation (3.1.1) of entropy conservation and the Einstein field equation, with curvature neglected,

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G\rho(T)}{3}. \tag{3.1.10}$$

Combining these gives

$$t = -\int \frac{s'(T)dT}{s(T)\sqrt{24\pi G\rho(T)}} + \text{ constant} \tag{3.1.11}$$

(The minus sign is inserted in taking the square root of $\dot{T}^2$, to take account of the fact that the temperature decreases as time passes.) In particular, during any epoch in which the

dominant constituent of the universe is a highly relativistic ideal gas, the entropy and energy densities are

given by

$$s(T) = \frac{2\mathcal{N}a_{\mathcal{B}}T^3}{3} \tag{3.1.12}$$

$$\rho(T) = \frac{\mathcal{N}a_{\mathcal{B}}T^4}{2} \tag{3.1.13}$$

where $\mathcal{N}$ is the number of particle types, counting particles and antiparticles and each spin state separately, and with an extra factor of 7/8 for fermions. Then Eq. (3.1.11) becomes

$$t = \sqrt{\frac{3}{16\pi G\mathcal{N}a_{\mathcal{B}}}}\frac{1}{T^2} + \text{ constant} \tag{3.1.14}$$

With this background, let us now start our history at a time when the temperature was around 10^{11} K, which is in the range $m_\mu \gg k_{\mathcal{B}}T \gg m_e$. Even though it was too cold at this time for reactions like $\upsilon_\mu + e \to \mu + \upsilon_e$ or $\nu_\tau + e \to \tau + \nu_e$, the μ and τ neutrinos and antineutrinos were kept in thermal equilibrium by neutral current reactions, like neutrino-electron scattering or $e^+ + e^- \rightleftharpoons \nu + \bar{\nu}$. Hence the constituents of the universe at this time were photons with two spin states, plus three species of neutrinos and three of antineutrinos, each with one spin state, plus electrons and positrons, each with two spin states, all in equilibrium and all highly relativistic, giving

$$\mathcal{N} = 2 + \frac{7}{8}(6+4) = \frac{43}{4} \tag{3.1.15}$$

so that Eq. (3.1.14) gives, in cgs units:

$$t = \sqrt{\frac{3c^2}{172\pi G a_{\mathcal{B}}}}\frac{1}{T^2} + \text{ constant } = 0.994\,\text{sec}\left[\frac{T}{10^{10}\text{ K}}\right]^{-2} + \text{ constant .} \tag{3.1.16}$$

For instance, with muons ignored and the mass of the electron neglected, it took 0.0098 sec for the temperature to drop from a value 10^{12} K to 10^{11} K, and another 0.98 sec for the temperature to drop to 10^{10} K.

At a temperature of about 10^{10} K neutrinos were just going out of equilibrium and beginning a free expansion. The weak interaction cross section for neutrino-electron scattering is roughly $\sigma_{\text{wk}} \approx (\hbar G_{\text{wk}} k_{\mathcal{B}}T)^2$, where $G_{\text{wk}} \simeq 1.16 \times 10^{-5}\text{GeV}^{-2}$ is the weak interaction coupling constant, and the factor $\hbar^2$ is included to convert a quantity with the units (energy) $^{-2}$ to a quantity with the units (length) ^{2} of a cross section. (Recall that we are using units with $c = 1$.) The number density of electrons at temperatures above 10^{10} K is roughly given by $n_e \approx (k_{\mathcal{B}}T/\hbar)^3$, so the collision rate of a neutrino with electrons or positrons at such temperatures is

$$\Gamma_\upsilon = n_e\sigma_{\text{wk}} \approx G_{\text{wk}}^2\left(k_{\mathcal{B}}T\right)^5/\hbar$$

This may be compared with the expansion rate, which during the radiationdominated era is of the order

$$H \approx \sqrt{G\left(k_{\mathcal{B}}T\right)^4/\hbar^3},$$

[471] Eq. (3.1.6) follows directly from the definition of $n(p,T)$, and Eq. (3.1.7) can then be derived from Eq. (3.1.4). Or both Eqs. (3.1.6) and (3.1.7) can be obtained from Eqs. (B.41) and (B.43).

with the factor $\hbar^{-3}$ included to convert a quantity with the units (energy) ^{4} to a quantity with the units mass/length ^{3} of a mass density. The ratio of the collision rate to the expansion rate is thus

$$\frac{\Gamma_\nu}{H} \approx G_{\rm wk}^2(\hbar/G)^{1/2}\,(k_{\mathcal B}T)^3 \simeq \left(\frac{T}{10^{10}\ {\rm K}}\right)^3$$

Hence neutrinos were scattered rapidly enough to remain in thermal equilibrium at temperatures above 10^{10} K. This is just a little greater than $m_e/k_{\mathcal B}$, so for lower temperatures electrons and positrons rapidly disappeared from equilibrium, the collision rate dropped rapidly below $G_{\rm wk}^2\,(k_{\mathcal B}T)^5\,/\hbar$, and hence the ratio Γ_ν/H dropped rapidly below unity. The neutrinos then began a free expansion, in which (as we saw in Section 2.1) the number density distribution n_ν continued to keep the form (3.1.5), with a temperature $T_\nu \propto 1/a$.

At lower temperatures we must take into consideration the finite mass of the electron, so the temperature T of the electrons, positrons, and photons (which were still in equilibrium with each other) no longer fell as $1/a$. On the other hand, the freely expanding massless neutrinos preserved a FermiDirac momentum distribution, ^{2} with a temperature that continued to drop as $1/a$. We must therefore now distinguish between the photon temperature T, and the neutrino temperature T_ν.

The entropy density of the photons, electrons, and positrons is

$$\begin{aligned} s(T) =& \frac{4a_{\mathcal B}T^3}{3} + \frac{4}{T}\int_0^\infty \frac{4\pi p^2 dp}{(2\pi\hbar)^3}\left(\sqrt{p^2+m_e^2} + \frac{p^2}{3\sqrt{p^2+m_e^2}}\right) \\ & \times \frac{1}{\exp\left(\sqrt{p^2+m_e^2}/k_{\mathcal B}T\right)+1} \\ =& \frac{4a_{\mathcal B}T^3}{3}\mathcal{S}\left(m_e/k_{\mathcal B}T\right) \end{aligned} \tag{3.1.17}$$

where, recalling that $a_{\mathcal B} = \pi^2 k_{\mathcal B}^4/15\hbar^3$,
$\mathcal{S}(x) \equiv 1 + \frac{45}{2\pi^4}\int_0^\infty y^2 dy\left(\sqrt{y^2+x^2} + \frac{y^2}{3\sqrt{y^2+x^2}}\right)\frac{1}{\exp\sqrt{y^2+x^2}+1}$.

The entropy conservation law (3.1.1) gives $a^3T^3\mathcal{S}\left(m_e/k_{\mathcal B}T\right)$ constant, and since $T_\nu \propto 1/a$, this means that T_ν is proportional to $T\mathcal{S}^{1/3}\left(m_e/k_{\mathcal B}T\right)$. The temperatures were equal for $k_{\mathcal B}T \gg m_e$, and

$$\mathcal{S}(0) = 1 + 2\frac{7}{8} = 11/4 \tag{3.1.19}$$

so

$$T_v = (4/11)^{1/3}T\mathcal{S}^{1/3}\left(m_e/k_{\mathcal B}T\right) \tag{3.1.20}$$

The ratio T/T_ν rose from very close to unity for $T > 10^{11}$ K to 1.008 at $T = 10^{10}$ K and to 1.346 at $T = 10^9$ K. To find the asymptotic value of T/T_ν without a computer calculation, we note that $\mathcal{S}(\infty) = 1$, so for $k_{\mathcal B}T \ll m_e$, Eq. (3.1.20) gives

^{472} This is not exact; even at temperatures under 10^{10} K, the neutrino distribution was slightly affected by weak interaction processes, such as $e^- + e^+ \to \upsilon + \bar{\upsilon}$. See A. D. Dolgov, S. H. Hansen, and D. V. Semikoz, Nucl. Phys. B 503, 426 (1997); 543, 269 (1999); G. Mangano, G. Miele, S. Pastor, and M. Peioso, Phys. Lett. B534, 8 (2002). For a review, see A. D. Dolgov, Phys. Rep. 370, 333 (2002). The weak interactions provide some thermal contact between the neutrinos and the plasma, which is being heated by electron-positron annihilation, so the effect is to slightly increase the neutrino energy density, by an amount usually represented as an increase in the effective number of neutrino species, from 3 to 3.04. This effect is neglected in what follows.

$$T/T_v \to (11/4)^{1/3} = 1.401 \tag{3.1.21}$$

In particular, at the present time, when $T = 2.725$ K, the neutrino temperature is 1.945 K . Unfortunately there does not seem to be any way of detecting such a neutrino background.

With three flavors of neutrinos and antineutrinos, the total energy density during this period is

$$\begin{aligned}\rho(T) &= 6 \cdot \frac{7}{8} \cdot \frac{a_{\mathcal{B}} T_v^4}{2} + a_{\mathcal{B}} T^4 + 4 \int_0^\infty \frac{4\pi p^2 dp}{(2\pi\hbar)^3} \frac{\sqrt{p^2 + m_e^2}}{\exp\left(\sqrt{p^2 + m_e^2}/k_{\mathcal{B}} T\right) + 1} \\ &= a_{\mathcal{B}} T^4 \mathcal{E}\left(m_e/k_{\mathcal{B}} T\right)\end{aligned} \tag{3.1.22}$$

where

$$\mathcal{E}(x) = 1 + \frac{21}{8}\left(\frac{4}{11}\right)^{4/3} \mathcal{S}^{4/3}(x) + \frac{30}{\pi^4} \int_0^\infty \frac{y^2 \sqrt{y^2 + x^2} dy}{\exp \sqrt{y^2 + x^2} + 1} \tag{3.1.23}$$

We insert Eqs. (3.1.17) and (3.1.22) in Eq. (3.1.11), and find

$$\begin{aligned}t &= \int \left(\frac{\left(m_e/k_{\mathcal{B}} T\right) \mathcal{S}'\left(m_e/k_{\mathcal{B}} T\right)}{\mathcal{S}\left(m_e/k_{\mathcal{B}} T\right)} - 3 \right) \frac{dT}{T\sqrt{24\pi G a_{\mathcal{B}} T^4 \mathcal{E}\left(m_e/k_{\mathcal{B}} T\right)}} \\ &= t_e \int \left(3 - \frac{x\mathcal{S}'(x)}{\mathcal{S}(x)}\right) \mathcal{E}^{-1/2}(x) x dx\end{aligned} \tag{3.1.24}$$

where $x \equiv m_e/k_{\mathcal{B}} T$, and in cgs units

$$t_e \equiv \left(24\pi G c^6 a_{\mathcal{B}} m_e^4/k_{\mathcal{B}}^4\right)^{-1/2} = 4.3694\text{sec}. \tag{3.1.25}$$

The values of T/T_ν and of the time required for the temperature to fall to T (calculated from Eq. (3.1.11)) are given for various values of T in Table 3.1.

After the era of electron-positron annihilation, the energy density of the universe was dominated for a long while by photons, neutrinos, and antineutrinos, all of them highly relativistic, so during this period we have $s(T) \propto T^3$, and

$$\rho(T) = a_{\mathcal{B}} T^4 + \frac{7}{8} \cdot 3 \cdot a_{\mathcal{B}} T_v^4 = a_{\mathcal{B}} T^4 \left(1 + \frac{7}{8} \cdot 3 \cdot (4/11)^{4/3}\right) = 3.363 a_{\mathcal{B}} T^4/2. \tag{3.1.26}$$

Table 3.1: Ratio of electron-photon temperature T to neutrino temperature T_{ν} and the time t required for the temperature to drop from 10^{11} K to T, for various values of T.

T(K)	T/T_v	t(sec)
10^{11}	1.000	0
6×10^{10}	1.000	0.0177
3×10^{10}	1.001	0.101
2×10^{10}	1.002	0.239
10^{10}	1.008	0.998
6×10^{9}	1.022	2.86
3×10^{9}	1.080	12.66
2×10^{9}	1.159	33.1
10^{9}	1.345	168
3×10^{8}	1.401	1980
10^{8}	1.401	1.78×10^{4}
10^{7}	1.401	1.78×10^{6}
10^{6}	1.401	1.78×10^{8}

That is, during this period the effective number of particle species is $\mathcal{N} = 3.363$. Using Eq. (3.1.14) then gives, in cgs units:

$$t = \sqrt{\frac{3c^2}{3.363 \cdot 16 \cdot \pi a_{\mathcal{B}} G}} \frac{1}{T^2} + \text{ constant } = 1.78\,\text{sec} \left[\frac{T}{10^{10}\text{ K}}\right]^{-2} + \text{ constant .} \tag{3.1.27}$$

For instance, the time required for the universe to cool from a temperature of 10^9 K (where electrons and positrons have mostly annihilated) to a temperature of 10^8 K is 1.76×10^4sec, or 4.9 hours.

According to Eq. (3.1.27), the time required for the temperature to drop to 10^6 K from much higher values is 1.78×10^8sec, or 5.64 years. At lower temperatures we must take into account the energy density of nonrelativistic matter, and Eq. (3.1.27) no longer applies. We saw in Section 2.3 that for $\Omega_M h^2 = 0.15$, it took an additional 360,000 years for the universe to cool to the temperature 3,000 K of last scattering.

So far in this section we have been assuming that neutrinos are massless, and that the net neutrino number of each of the three types (that is, the number of neutrinos minus the number of antineutrinos) is much less than the number of photons. In the general case of an ideal gas of particles of mass m, the number $n(p)dp$ of particles of momentum between p and $p + dp$ is given by the Fermi-Dirac and Bose-Einstein distributions

$$n(p, T, \mu)dp = \frac{4\pi g p^2 dp}{(2\pi\hbar)^3} \left(\frac{1}{\exp\left[\left(\sqrt{p^2 + m^2} - \mu\right)/k_{\mathcal{B}}T\right] \pm 1} \right) \tag{3.1.28}$$

where μ is the chemical potential for the particle in question, a quantity that is conserved in any reaction occurring rapidly in thermal equilibrium, and again g is the number of spin states of the particle and antiparticle, and the sign is + for fermions and - for bosons. This reduces to Eq. (3.1.5) in the case of zero chemical potential, and it yields the number density (2.3.1) for non-relativistic particles with $p \ll m$ and $k_{\mathcal{B}}T \ll m$. During the whole of the era of interest here, electrons and positrons rapidly annihilated into photons, so their chemical potentials were equal and opposite, and since we are assuming charge neutrality and neglecting

the tiny number of baryons per photon, we can conclude that the chemical potentials of the electrons and photons were much less than $k_{\mathcal{B}}T$. At temperatures at which neutrinos and antineutrinos were in thermal equilibrium with electrons, positrons, and photons, reactions like $e^+ + e^- \rightleftharpoons v_i + \bar{v}_i$ were occurring rapidly (where $i = e, \mu, \tau$ label the three types of neutrino), so the chemical
potential μ_i of each type of neutrino was equal and opposite to the chemical potential of the corresponding antineutrino. But if we do not assume zero net neutrino (or lepton) number, then there is no a priori reason why the μ_i had to be less than $k_{\mathcal{B}}T$.

If neutrino masses are less than about 1 eV , then they may be neglected at the temperatures of interest in this section. The observations of oscillations between different flavors of neutrinos from the sun, nuclear reactors, and cosmic rays shows [3] that the two differences in the squares of the masses of the three types of neutrinos of definite mass (which are mixtures of neutrinos of electron, muon, and tau flavor) are $8.0^{+0.4}_{-0.3} \times 10^{-5}\text{eV}^2$ and between 1.9 and 3.0 times 10^{-3}eV^2. Thus the neutrino masses are all much less than 1 eV , unless they are highly degenerate, which there is no reason to expect. If degenerate, then from the absence of anomalies in the low-energy beta decay of tritium, their common mass must be less than 2eV.[3] Whether degenerate or not, it is clear from this that all three neutrino types (if there are only three) have masses very much less than 1 MeV , and were therefore highly relativistic at the time that they went of thermal equilibrium with electrons and positrons, at a temperature of about 10^{10} K. Once out of equilibrium, their momentum simply decayed as $1/a$ (as shown in Section 1.1), so if their chemical potential was negligible their momentum distribution remained the same as that of photons, with a temperature less by a factor $(4/11)^{1/3}$. Thus once $k_{\mathcal{B}}T$ dropped below the smallest neutrino mass, their energy density became just $n_\nu \sum_\nu m_\nu = (3/11) n_\gamma \sum_\nu m_\nu$. (For $k_{\mathcal{B}}T$ much larger than the mass, the integral for the number density of each spin state of fermions is 3/4 the corresponding integral for bosons, and after neutrinos decouple $T_\nu^3 = (4/11)T_\gamma^3$.) With a non-zero chemical potential the energy density is larger. The agreement between theory (with massless neutrinos) and observation for the cosmic microwave background anisotropies discussed in Sections 2.6 and 7.2 and for the large scale structure discussed in Chapter 8 shows that the sum of the three neutrino masses is less than 0.68 eV (95% confidence level), [4] so if they are degenerate the common mass is less than 0.23 eV . This result has been contradicted by the observation of neutrinoless double beta decay in a Heidelberg-Moscow experiment, [5] which suggests a value greater than 1.2 eV for the sum of neutrino masses. [6] There has not yet been an opportunity to confirm the double beta decay results, and for the present it seems reasonable to continue to neglect neutrino masses.

With neutrino masses neglected, the energy density, pressure, and net lepton number density of neutrinos and antineutrinos of type i is

$$\begin{aligned}
\rho_i =& 3p_i \\
=& \frac{4\pi}{(2\pi\hbar)^3} \int p^3 dp \left(\frac{1}{\exp\left[(p-\mu_i)/k_{\mathcal{B}}T\right]+1} + \frac{1}{\exp\left[(p+\mu_i)/k_{\mathcal{B}}T\right]+1} \right) \\
=& 4\pi \frac{(k_{\mathcal{B}}T)^4}{(2\pi\hbar)^3} \mathcal{P}\left(\mu_i/k_{\mathcal{B}}T\right) \qquad (3.1.29) \\
n_i =& \frac{4\pi}{(2\pi\hbar)^3} \int p^2 dp
\end{aligned}$$

[473] W.-M. Yao et al. (Particle Data Group), J. Phys. G. 33, 1 (2006).
[4] D. N. Spergel et al., Astrophys. J. Suppl. Ser. 170, 377 (2007) [astro-ph/0603449].
[5] H. V. Klapdor-Kleingroth, I. V. Krivosheina, A. Dietz and O. Chkvorets, Phys. Lett. B 586, 198 (2004).
[6] A. De La Macorra, A. Melchiorri, P. Serra, and R. Bean, Astropart. Phys. 27, 406 (2007) [astroph/0608351].

$$\times\left(\frac{1}{\exp\left[\left(p-\mu_i\right)/k_{\mathcal{B}}T\right]+1}-\frac{1}{\exp\left[\left(p+\mu_i\right)/k_{\mathcal{B}}T\right]+1}\right)$$
$$=4\pi\frac{\left(k_{\mathcal{B}}T\right)^3}{(2\pi\hbar)^3}\mathcal{M}\left(\mu_i/k_{\mathcal{B}}T\right) \tag{3.1.30}$$

where

$$\mathcal{P}(x)\equiv\int_0^\infty\left[\left(e^{y-x}+1\right)^{-1}+\left(e^{y+x}+1\right)^{-1}\right]y^3dy \tag{3.1.31}$$
$$\mathcal{M}(x)\equiv\int_0^\infty\left[\left(e^{y-x}+1\right)^{-1}-\left(e^{y+x}+1\right)^{-1}\right]y^2dy \tag{3.1.32}$$

As we have seen, at temperatures above 10^{10} K the energy density of the photons and electron-positron pairs is $11a_{\mathcal{B}}T^4/4$, and the pressure is one-third as great, so the total energy density and pressure are given by

$$\rho=3p=T^4\left[\frac{11a_{\mathcal{B}}}{4}+\frac{4\pi k_{\mathcal{B}}^4}{(2\pi\hbar)^3}\sum_{i=e,\mu,\tau}\mathcal{P}\left(\mu_i/k_{\mathcal{B}}T\right)\right]. \tag{3.1.33}$$

The equation (1.5.20) of energy conservation tells us that under these circumstances ρa^4 is constant, while the conservation of each type of neutrino number also tells us that n_ia^3 is constant. Since ρ and n_i depend in different ways on the chemical potentials, this requires that as the universe expands in this era the $\mu_i/k_{\mathcal{B}}T$ remain constant, and also $T\propto 1/a$, just as in the case of zero chemical potential.

As the temperature dropped below 10^{10} K the temperature of the photons and electron-positron pairs no longer varied as $1/a$, but as we have seen the neutrinos and antineutrinos entered on a free expansion. With each neutrino's momentum p varying as $1/a$, the form of the Fermi-Dirac distributions for each type of massless neutrino was preserved, with a temperature $T_\nu\propto 1/a$ and $\mu_i/k_{\mathcal{B}}T$ constant, just as before decoupling. We
conclude that to a good approximation $T_\nu\propto 1/a$ and each $\mu_i\propto 1/a$ throughout the whole era of interest here.

The only effect that a non-zero neutrino chemical potential would have on the calculations of this section is that it would increase the total energy density and hence shorten the time scale. For any non-zero μ the function $\mathcal{P}\left(\mu/k_{\mathcal{B}}T\right)$ is greater than $\mathcal{P}(0)=7\pi^4/60$, so if $\mu_i\neq 0$ then $\rho_i>7\pi^2\left(k_{\mathcal{B}}T\right)^4/120(\hbar)^3=(7/8)a_{\mathcal{B}}T^4$. In particular, if $|\mu_i|\gg k_{\mathcal{B}}T$ then $\mathcal{P}\left(\mu_i/k_{\mathcal{B}}T\right)\simeq\left(\mu_i/k_{\mathcal{B}}T\right)^4/4$, so $\rho_i\simeq\pi\left(\mu_i\right)^4/(2\pi\hbar)^3\gg a_{\mathcal{B}}T^4$, and these neutrinos (or antineutrinos, if $\exp\left(-\mu_i/k_{\mathcal{B}}T\right)$ is large) dominated the energy density of the universe, at least until the cross-over of non-relativistic matter and radiation. Inspection of Eq. (3.1.28) shows that for a chemical potential $\mu\gg k_{\mathcal{B}}T$, the fermion distribution function $n(p,T,\mu)dp$ is equal to $4\pi gp^2dp/(2\pi\hbar)^3$ for particle energies less than μ, and then falls off rapidly for higher energies, indicating that all fermion energy levels up to energy μ are filled, while higher energy levels are empty. This is the case of complete neutrino degeneracy. [7] Experiments on the beta decay of tritium, ${}^3\mathrm{H}\rightarrow{}^3\mathrm{He}+e^-+\bar{\nu}_e$ show that $|\mu_e|$ is less than a few eV , because otherwise for $\mu_e>0$ there would be a rise in the electron spectrum beyond the expected endpoint $m\left({}^3\mathrm{H}\right)-m\left({}^3\mathrm{He}\right)$, due to absorption of degenerate cosmic neutrinos in the reaction $\nu_e+{}^3\mathrm{H}\rightarrow{}^3\mathrm{He}+e^-$, while for $\mu_e<0$ the Pauli exclusion principle would produce a dip in the electron spectrum within a few eV of its expected endpoint, where antineutrino energies are less than a few eV, putting them within the degenerate antineutrino sea. (The absence of a dip within a few eV of the expected electron spectrum endpoint also sets a limit of a few eV on the electron neutrino mass.) This yields an upper bound on the timeindependent quantity $|\mu_e|/k_{\mathrm{B}}T$ of order 10^4, which is very much

weaker than the upper bound that will be provided in the following section by considerations of cosmological nucleosynthesis.

10.3.2 3.2 Cosmological nucleosynthesis

We have worked out the thermal history of the universe from temperatures above 10^{10} K down to the crossover temperature $\approx 10^4$ K, ignoring the presence of a small number of nucleons (and a small excess of electrons over positrons). Now let us consider what happens to the nucleons during this era. [1]

The weak interactions allow neutron-proton conversion through six processes:

$$n + v \rightleftharpoons p + e^-, \quad n + e^+ \rightleftharpoons p + \bar{v}, \quad n \rightleftharpoons p + e^- + \bar{v} \tag{3.2.1}$$

(Here v is v_e; the other neutrino flavors do not contribute to these reactions.) In this range of temperatures $k_{\mathcal{B}}T \ll m_N$, so the nucleons can be treated as essentially at rest. The initial and final lepton energies are therefore simply related, by

$$\begin{array}{ll} E_e - E_v = Q & \text{for } n + v \rightleftharpoons p + e^-, \\ E_v - E_e = Q & \text{for } n + e^+ \rightleftharpoons p + \bar{v} \\ E_v + E_e = Q & \text{for } n \rightleftharpoons p + e^- + \bar{v}, \end{array} \tag{3.2.2}$$

where

$$Q = m_n - m_p = 1.293\text{MeV} \tag{3.2.3}$$

The total rates at which an individual neutron is converted to a proton or a proton to a neutron take the form

$$\lambda(n \to p) = A \int \left(1 - \frac{m_e^2}{(Q+q)^2}\right)^{1/2} \frac{(Q+q)^2 q^2 dq}{(1 + e^{q/k_{\mathcal{B}}T_v})\,(1 + e^{-(Q+q)/k_{\mathcal{B}}T})} \tag{3.2.4}$$

[477] The possibility of cosmic neutrino degeneracy was raised by S. Weinberg, Phys. Rev. 128, 1457 (1962).

[1] I outlined the history of these calculations in The First Three Minutes (Basic Books, 1977, 1988). Briefly, the first calculations of cosmological nucleosynthesis were undertaken by Ralph Alpher, George Gamow, and Robert Herman in the late 1940s; see G&C, Chapter 15, footnotes 51 and 52. In this

[47]work it was assumed that nucleons start as pure neutrons, which then convert to protons by the process of neutron beta decay. It was then pointed out that the conversion of neutrons into protons and vice versa occurs primarily through two-particle collisions, and that the rapid rate of these processes at very early times has the consequence that nucleons start as 50% neutrons and 50% protons, by C. Hayashi, Prog. Theor. Phys. (Japan) 5, 224 (1950). Following this, a modern calculation of the evolution of the neutron/proton ratio was presented by R. A. Alpher, J. W. Follin, Jr., and R. C. Herman, Phys. Rev. 92, 1347 (1953), but the results were not applied to the problem of cosmological nucleosynthesis. Several authors noted that the abundance of helium in the universe is too large to be accounted for by stellar nucleosynthesis; see G. Burbidge, Pub. Astron. Soc. Pacific 70, 83 (1958); F. Hoyle and R. J. Tayler, Nature 203, 1108 (1964). The modern theory of cosmological nucleosynthesis is due to P. J. E. Peebles, Astron. J. 146, 542 (1966). (Related calculations done by Ya. B. Zel'dovich, Adv. Astron. Astrophys. 3, 241 (1965) were not known in the West until much later.) Nucleosynthesis calculations were then extended to more nuclides and reactions by R. V. Wagoner, W. A. Fowler, and F. Hoyle, Astrophys. J. 148, 3 (1967), and many small corrections were included by D. A. Dicus, E. W. Kolb, A. M. Gleeson, E. C. G. Sudarshan, V. L. Teplitz, and M. S. Turner, Phys. Rev. D 26, 2694 (1982). Modern reviews are given by G. Steigman, in Measuring and Modeling the Universe - Carnegie Observatories Astrophysics Series, Volume 2, ed. W. Freedman (Cambridge University Press, Cambridge, UK) [astro-ph/0307244]; in The Local Group as an Astrophysical Laboratory - Proceedings of the May 2003 STScI Symposium [astro-ph/0308511]; in Chemical Abundances and Mixing in Stars in the Milky Way and its Satellites Proceedings of the ESO/Arcetrei Workshop, eds. L. Pasquini and S. Randich (Springer-Verlag) [astroph/0501591]; and Int. J. Mod. Phys. E15, 1 (2006) [astro-ph/0511534]. For discussions emphasizing analytic calculations, see G&C, Section 15.7, and V. Mukhanov, Int. J. Theor. Phys. 43, 669 (2004) [astro-ph/0303073].

$$\lambda(p \to n) = A \int \left(1 - \frac{m_e^2}{(Q+q)^2}\right)^{1/2} \frac{(Q+q)^2 q^2 dq}{(1+e^{-q/k_{\mathcal{B}}T_v})\,(1+e^{(Q+q)/k_{\mathcal{B}}T})} \tag{3.2.5}$$

where [2]

$$A \equiv \frac{G_{\rm wk}^2\,(1+3g_A^2)\cos^2\theta_C}{2\pi^3\hbar} \tag{3.2.6}$$

and the integrals over q run from $-\infty$ to $+\infty$, leaving out a gap from $q = -Q - m_e$ to $q = -Q + m_e$, where the square root would be imaginary. These rates include the effect of the Pauli Principle in the presence of partly filled lepton seas. For instance, the cross section for the process $n + e^+ \to p + \bar{v}$ is $2\pi^2\hbar^3 A E_v^2/v_e$, the number density of positrons of each helicity with momentum between p_e and $p_e + dp_e$ is $4\pi p_e^2 dp_e (2\pi\hbar)^{-3}\left[\exp\left(E_e/k_{\mathcal{B}}T\right)+1\right]^{-1}$, and the fraction of unfilled antineutrino levels with energy E_ν is

$$1 - \left[\exp\left(E_\nu/k_{\mathcal{B}}T_\nu\right)+1\right]^{-1} = \left[\exp\left(-E_\nu/k_{\mathcal{B}}T_\nu\right)+1\right]^{-1}$$

so the total rate per neutron of the process $n + e^+ \to p + \bar{v}$ is

$$\begin{aligned}\lambda\left(n + e^+ \to p + \bar{v}\right) =& A \int_0^\infty E_v^2 p_e^2 dp_e \left[\exp\left(E_e/k_{\mathcal{B}}T\right)+1\right]^{-1} \\ & \times \left[\exp\left(-E_v/k_{\mathcal{B}}T_v\right)+1\right]^{-1}\end{aligned}$$

Changing the variable of integration to $q \equiv -E_v = -Q - E_e$, we see that this partial rate is just the part of the integral (3.2.4) that runs from $q = -\infty$ to $q = -Q - m_e$. Likewise, the part of the integral that runs from $q = -Q + m_e$ to $q = 0$ is supplied by the neutron-decay process $n \to p + e^- + \bar{\nu}$, with $q = -E_\nu$, and the part of the integral that runs from $q = 0$ to $q = +\infty$ arises from the process $n + v \to p + e^-$, with $q = E_v$. Similar remarks apply to the integral (3.2.5).

With the rates (3.2.4)-(3.2.5) known in principle, we can calculate the change in the ratio X_n of neutrons to all nucleons from the differential equation

$$\frac{dX_n}{dt} = -\lambda(n \to p) X_n + \lambda(p \to n)\,(1 - X_n) \tag{3.2.7}$$

As a check, note that for a time-independent temperature T equal to T_ν, the two rates (3.2.4) and (3.2.5) would have the ratio

$$\frac{\lambda(p \to n)}{\lambda(n \to p)} = \exp\left(-Q/k_{\mathcal{B}}T\right) \qquad \text{for } T = T_v \tag{3.2.8}$$

It follows then that in this case, Eq. (3.2.7) would have the time-independent solution expected in thermal equilibrium

$$\frac{X_n}{X_p} = \frac{X_n}{1 - X_n} = \exp\left(-Q/k_{\mathcal{B}}T\right) \tag{3.2.9}$$

It is the inequality of T and T_ν as well as the time-dependence of these temperatures that drives X_n/X_p away from its equilibrium value (3.2.9).

For $k_{\mathcal{B}}T \gg Q$ we can evaluate the integrals (3.2.4)-(3.2.5) by setting $T_\nu = T$ and $Q = m_e = 0$, so in this case

[472] Here $G_{\rm wk} = 1.16637(1) \times 10^{-5}{\rm GeV}^{-2}$ is the weak coupling constant, measured from the rate of the decay process $\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu$; $g_A = 1.257$ is the axial vector coupling of beta decay, measured from the rate of neutron decay; and θ_C is the Cabibbo angle, with $\cos\theta_C = 0.9745$ (6), measured from the rate of O^{14} beta decay and other $0^+ \to 0^+$transitions.

$$\lambda(n \to p) = \lambda(p \to n) = A \int_{-\infty}^{+\infty} \frac{q^4 dq}{(1 + e^{q/k_{\mathcal{B}}T})(1 + e^{-q/k_{\mathcal{B}}T})}$$
$$= \frac{7}{15}\pi^4 A (k_{\mathcal{B}}T)^5 = 0.400\text{sec}^{-1} \left(\frac{T}{10^{10}\text{ K}}\right)^5 \tag{3.2.10}$$

For comparison, the time t for the temperature to drop to T from much higher values is given by Eq. (3.1.16) as $0.99\text{sec}\,(T/10^{10}\text{ K})^{-2}$. Also, $H \simeq 1/2t$. The ratio λ/H is therefore

$$\lambda/H \simeq 0.8 \times \left(\frac{T}{10^{10}\text{ K}}\right)^3 \tag{3.2.11}$$

This ratio is larger than 1 for $T > 1.1 \times 10^{10}$ K. It is true that temperatures near this lower bound are not much larger than Q/k, and T_ν at this epoch is not precisely equal to T, so the rates $\lambda(n \to p)$ and $\lambda(p \to n)$ are not precisely equal, and neither is given precisely by Eq. (3.2.10). Nevertheless Eq. (3.2.10) gives the order of magnitude of these rates in this temperature range, so we can still rely on the conclusion that $\lambda(n \to p)/H$ and $\lambda(p \to n)/H$ are large down to these temperatures. This means that the initial value of X_n at temperatures larger than about 3×10^{10} K is given by the condition that the right-hand side of Eq. (3.2.7) should vanish:

$$X_n \to \frac{\lambda(p \to n)}{\lambda(p \to n) + \lambda(n \to p)} \tag{3.2.12}$$

If X_n were larger or smaller than this, then the right-hand side of Eq. (3.2.7) would be large and respectively negative or positive, so X_n would be rapidly driven to the value (3.2.12). For $T > 10^{10}$ K the neutrino temperature was within 1% of the photon-electron-positron temperature, so at these temperatures the rates have the ratio (3.2.8). Thus at the temperatures $T > 3 \times 10^{10}$ K at which the neutron fraction is given by the equilibrium formula (3.2.12), this formula gives simply

$$X_n = \frac{1}{1 + \exp(Q/k_{\mathcal{B}}T)} \tag{3.2.13}$$

It is crucially important that this ratio is fixed at high temperatures by the condition of zero lepton chemical potential, so that we do not need to make any a priori assumptions about the initial neutron/proton ratio. The results of a numerical integration of Eq. (3.2.7) with initial condition provided by Eq. (3.2.13) are presented in Table 3.2. [3]

We have seen that for $k_{\mathcal{B}}T > m_e$ the ratios $\lambda(p \to n)/H$ and $\lambda(n \to p)/H$ varied roughly as T^3, so there was a rather sharp end to the equilibrium era, in which $\lambda \gg H$, at a temperature between 3×10^{10} K and 10^{10} K. A little later, at a temperature between 10^{10} K and 3×10^9 K the two-body and threebody neutron-proton conversion reaction rates became negligible, due in part to the disappearance of electron-positron pairs. Neutron-proton conversion continued mostly through the process of neutron decay, with a mean lifetime τ_n of 885.7 ± 0.9sec, so the neutron fraction became proportional to $\exp(-t/\tau_n)$. This is confirmed by fitting the numerical results presented in Table 3.2 with an exponential that decays with the observed rate of neutron decay, which gives

$$X_n \to 0.1609 \exp\left(\frac{-t}{885.7\text{sec}}\right) \tag{3.2.14}$$

The conversion of neutrons into protons was eventually stopped by the formation of complex nuclei, in which neutrons are stable. As a guide to this process, note that in thermal (and

chemical) equilibrium the number density of a nuclear species i is given by formulas like Eqs. (2.3.1):

$$n_i = g_i e^{\mu_i/k_{\mathcal{B}}T} \left(\frac{2\pi m_i k_{\mathcal{B}} T}{h^2}\right)^{3/2} \exp\left(-m_i/k_{\mathcal{B}}T\right) \tag{3.2.15}$$

where m_i is the mass of nucleus i and g_i is the number of its spin states. If these nuclei can be built up rapidly out of Z_i protons and $A_i - Z_i$ neutrons,

Table 3.2: Neutron fraction X_{n} as a function of temperature or time (with neglect of the formation of complex nuclei).

T(K)	t(sec)	X_n
10^{12}	0.0001	0.4962
3×10^{11}	0.0011	0.4875
10^{11}	0.0099	0.4626
3×10^{10}	0.1106	0.3798
10^{10}	1.008	0.2386
3×10^{9}	12.67	0.1654
1.3×10^{9}	91.09	0.1458
1.2×10^{9}	110.2	0.1425
1.1×10^{9}	135.1	0.1385
10^{9}	168.1	0.1333
9×10^{8}	212.7	0.1268
8×10^{8}	274.3	0.1182
7×10^{8}	362.6	0.1070
6×10^{8}	496.3	0.0919
3×10^{8}	1980	0.0172
10^{8}	17780	3.07×10^{-10}

then the chemical potential of nuclei of type i is

$$\mu_i = Z_i \mu_p + (A_i - Z_i)\, \mu_n. \tag{3.2.16}$$

We can eliminate the unknown nucleon chemical potentials by forming the quantities

$$\frac{n_i}{n_p^{Z_i} n_n^{A_i - Z_i}} = \frac{g_i}{2^{A_i}} A_i^{3/2} \left(\frac{2\pi m_N k_{\mathcal{B}} T}{h^2}\right)^{3(1-A_i)/2} e^{B_i/k_{\mathcal{B}}T}, \tag{3.2.17}$$

where B_i is the binding energy, defined by

$$m_i = Z_i m_p + (A_i - Z_i)\, m_n - B_i. \tag{3.2.18}$$

(In deriving Eq. (3.2.17), we neglect the binding energy and the neutronproton mass difference outside the exponential.) Eq. (3.2.17) can be expressed in terms of the dimensionless ratios

$$X_i \equiv n_i/n_N, \quad X_p \equiv n_p/n_N, \quad X_n \equiv n_n/n_N \tag{3.2.19}$$

[473] These results are somewhat different from those given in Table 15.5 of G&C, because in 1972 only two types of neutrinos were known, and we are now assuming three types of massless neutrinos, which increases the expansion rate at a given temperature. We are also using the modern value 1.257 for the axial vector coupling constant of beta decay in Eq. (3.2.6), instead of the value 1.18 used in G&C.

where n_N is here the number density of all nucleons. In these terms,

$$X_i = \frac{g_i}{2} X_p^{Z_i} X_n^{A_i - Z_i} A_i^{3/2} \epsilon^{A_i - 1} e^{B_i/k_{\mathcal{B}}T} \tag{3.2.20}$$

where ϵ is the dimensionless quantity

$$\begin{aligned} \epsilon &\equiv \frac{1}{2} n_N h^3 \left(2\pi m_N k_{\mathcal{B}} T\right)^{-3/2} \\ &= 2.96 \times 10^{-11} \left(\frac{a}{10^{-10} a_0}\right)^{-3} \left(\frac{T}{10^{10}\ \mathrm{K}}\right)^{-3/2} \Omega_B h^2 \end{aligned} \tag{3.2.21}$$

in which we have used $n_N = 3\Omega_B H_0^2 \left(a_0/a\right)^3 / 8\pi G m_N$ for the number density of nucleons, bound or free. During the period of interest (after electronpositron annihilation) the temperature T goes as $1/a$, so Eq. (3.2.21) may be written

$$\epsilon = 1.46 \times 10^{-12} \left(\frac{T}{10^{10}\ \mathrm{K}}\right)^{+3/2} \Omega_B h^2 \tag{3.2.22}$$

The coefficient ϵ is very small for temperatures in the range of interest, so in equilibrium a nuclear species i is nearly absent until the temperature drops to the value

$$T_i \simeq \frac{B_i}{k\left(A_i - 1\right) \left|\ln \epsilon\right|} \tag{3.2.23}$$

For $\Omega_B h^2 \simeq 0.02$ this temperature is 0.75×10^9 K for deuterium, 1.4×10^9 K for H^3, 1.3×10^9 K for He^3, and 3.1×10^9 K for He^4, with only a very weak dependence on $\Omega_B h^2$. The binding energy per nucleon for heavier nuclei is similar to that of He^4, so they have similar values of T_i.

If thermal and chemical equilibrium were really maintained during the time that the temperature drops from around 10^{10} K to below 10^9 K, then during this time He^4 and heavier nuclei would appear first, followed by He^3 and H^3 (which would later beta decay to He^3), followed finally by H^2. But this is not what happens. The density at this time is too low for any but two-body reactions to compete with the expansion rate, so nuclei are built up by a chain of two-body processes: first $p + n \to d + \gamma$, then $d + d \to H^3 + p$ and $d + d \to He^3 + n$, and next $d + H^3 \to He^4 + n$ and $d + He^3 \to He^4 + p$, as well as slower processes involving photons.

There is no trouble with the first step. The rate of deuterium production per free neutron is

$$\begin{aligned} \lambda_d &= 4.55 \times 10^{-20}\ \mathrm{cm}^3/\mathrm{sec} \times n_p = 511 \mathrm{sec}^{-1} \left(\frac{a}{10^{-9} a_0}\right)^{-3} X_p \Omega_B h^2 \\ &= 2.52 \times 10^4 \mathrm{sec}^{-1} \left(\frac{T}{10^{10}\ \mathrm{K}}\right)^3 X_p \Omega_B h^2 \end{aligned} \tag{3.2.24}$$

Multiplying by the time (3.1.27), this gives

$$\lambda_d t \simeq 4.5 \times 10^4 \left(\frac{T}{10^{10}\ \mathrm{K}}\right) X_p \Omega_B h^2, \tag{3.2.25}$$

which remains substantially greater than unity until well after the temperature drops below 10^9 K. Therefore the deuterium abundance during the period of interest is given to a good approximation by its equilibrium value, which according to Eq. (3.2.20) is

$$X_d = 3\sqrt{2} X_p X_n \epsilon \exp\left(\frac{B_d}{k_{\mathcal{B}} T}\right). \tag{3.2.26}$$

The trouble is that, because of the small binding energy of the deuteron, the temperature $T_d \approx 0.7 \times 10^9$ K is quite small, so deuterons remained rare until long after He^4 would have been abundant in thermal equilibrium. With deuterons rare, the two-deuteron processes $d + d \to H^3 + p$ and $d + d \to He^3 + n$ had small rates per deuteron, blocking further nucleosynthesis. (The rarity of deuterons had no effect on the rate per deuteron of radiative processes like $p + d \to He^3 + \gamma$ and $n + d \to H^3 + \gamma$, and these reactions are included in modern nucleosynthesis calculations, but they have intrinsically small cross sections.) When finally the temperature dropped below T_d the neutrons that were still extant were very rapidly assembled into the most deeply bound of the light elements, He^4. Further cosmological nucleosynthesis was blocked by the non-existence of stable nuclear species with atomic weight 5 or 8 . (In stars this blockage is overcome [4] by the brief formation of the unstable Be^8 nucleus in collisions of two He^4 nuclei, followed by resonant capture of another He^4 nucleus to form an excited state of C^{12}, but the time available in the early universe was too short for this to have been effective then.) Thus, to a good approximation, the fraction by weight Y of He^4 formed in the early universe is just equal to twice [5] the fraction X_n of all nucleons that are neutrons (because in He^4 each neutron is accompanied with a bound proton) at the time that deuterium becomes abundant enough to allow the build-up of heavier nuclei.

This actually occurs at a temperature somewhat above T_d. The first steps in building up heavier elements from deuterium are the reactions $d + d \to H^3 + p$ and $d + d \to He^3 + n$. These are exothermic processes, so for initial velocity $v \to 0$ these have cross sections proportional to $1/v$. At the temperatures of interest here, we have

$$\left\langle \sigma \left(d + d \to H^3 + p \right) v \right\rangle \simeq 1.8 \times 10^{-17} \text{ cm}^3/\text{sec}$$
$$\left\langle \sigma \left(d + d \to He^3 + n \right) v \right\rangle \simeq 1.6 \times 10^{-17} \text{ cm}^3/\text{sec}$$

so the total rate of these processes per deuteron is [6]

$$\begin{aligned} \Lambda &= \left[\left\langle \sigma \left(d + d \to H^3 + p \right) v \right\rangle + \left\langle \sigma \left(d + d \to He^3 + n \right) v \right\rangle \right] X_d n_N \\ &\simeq 1.9 \times 10^7 \left(T/10^{10} \text{ K} \right)^3 \left(\Omega_B h^2 \right) X_d \text{sec}^{-1} \end{aligned}$$

This may be compared with the expansion rate, which after electron-positron annihilation is given by Eq. (3.1.27) as $H = 1/2t = 0.28 \left(T/10^{10} \text{ K}\right)^2 \text{sec}^{-1}$, so for T in the neighborhood of 10^9 K, we have $\Lambda = H$ at $X_d \simeq 1.2 \times 10^{-7}/\Omega_B h^2$, which is 0.6×10^{-5} for $\Omega_B h^2 = 0.02$. This value is reached in thermal equilibrium at a temperature $\simeq 10^9$ K, with only a weak dependence on $\Omega_B h^2$, so nucleosynthesis began at around 10^9 K, not at 0.75×10^9 K. According to table 3.2, this happened when $t = 168$ seconds, so according to Eq. (3.2.14), the abundance by weight of helium formed at this time was about

$$Y_p \simeq 2 \times 0.1609 \times \exp(-168/885) \simeq 0.27$$

(The conventional subscript p stands for "primordial.") The larger the nucleon density, the higher the temperature at which nucleosynthesis began, and so the less time there was for neutron decay before nucleosynthesis, leading to a higher final He^4 abundance. The results of modern calculations are usually given for different values of η, the ratio of nucleons to photons, which according to Eqs. (2.1.12) and (2.1.13) is related to $\Omega_B h^2$ by

[474] E. E. Salpeter, Astrophys. J. 115, 326 (1952).

[5] In a little more detail, with n_{He} helium nuclei and n_{H} hydrogen nuclei per unit proper volume, the helium mass per unit proper volume in atomic mass units is $4n_{\text{He}}$, while the total mass per unit proper

$$\Omega_B h^2 = 3.65\eta \times 10^7,$$

so for instance $\Omega_B h^2 = 0.02$ corresponds to $\eta = 5.5 \times 10^{-10}$. (The nucleon density at a given temperature can be expressed in terms of η without knowing the present microwave background temperature, which made η a more convenient parameter than $\Omega_B h^2$ before the 1990s, when the present microwave background temperature was not yet accurately known.) Detailed calculations [7] give $Y_p = 0.232$ for $\eta = 2 \times 10^{-10}$ and $Y_p = 0.240$ for $\eta = 4 \times 10^{-10}$.

The primordial helium/hydrogen ratio is inferred from spectroscopic study of HII regions (regions of ionized hydrogen) containing low abundances of "metals" (elements other than hydrogen or helium), especially in blue compact galaxies which have not yet formed many stars. Observations in the 1990s were divided between those giving lower values [8] $Y_p = 0.234 \pm 0.002 \pm 0.005$, and higher values [9] $Y_p = 0.243 \pm 0.003$. Even with this division among the observers, the helium abundance clearly called for a hot universe, with η in the range of 10^{-10} to 5×10^{-10}. More recent observations of HII regions, combined with new atomic data used to interpret the spectroscopic observations, have led to a more precise determination: [10] $Y_p = 0.2477 \pm 0.0029$, corresponding to $\eta = (5.813 \pm 1.81) \times 10^{-10}$. The uncertainty in the value of η given by observations of helium abundance is still quite large, because Y_p is only weakly dependent on the baryon/photon ratio.

Helium abundance is more useful as a test of the expansion rate than of the value of η. For instance, if there were four flavors of massless neutrinos that went out of equilibrium at temperatures between 10^{11} K and 10^{10} K, then at temperatures below 3×10^9 K, the effective number of particle species would have been $2+(7/8)\cdot 8\cdot(4/11)^{4/3} = 3.817$ instead of 3.363 , so the time required to drop from 3×10^9 K to any lower temperature would have been shortened by a factor $\sqrt{3.363/3.817} = 0.94$. Shortening the time to reach a given temperature increases the neutron fraction at that temperature, and hence increases the abundance of helium produced at the temperature $T \simeq 10^9$ K of nucleosynthesis. It is actually the shortening of the time scale at temperatures between 3×10^{10} K and 3×10^9 K, when electronpositron pairs were disappearing, that would have the largest effect on the helium abundance. Detailed calculations show [11] that for each additional

neutrino species this effect increases Y by an amount 0.01276 for $\eta = 10^{-10}$, and by 0.01369 for $\eta = 5 \times 10^{-10}$. The agreement between theory and observation for the helium abundance gave an upper bound of four light neutrino flavors before this was accurately measured to be just three flavors in measurements of the Z^0 decay width, part of which is due to the unobserved processes $Z^0 \to v + \bar{\nu}$. (Strictly speaking, the Z^0 width measures the number of neutrinos with masses less than $m_Z/2 = 45.6$GeV, while the helium abundance measures the number of neutrinos with masses less than about 1 MeV .) Although the number of light neutrino flavors is now definitely known on the basis of Z^0 decay to be no greater than three, there may be other light particles left over from the very early universe that contribute to the expansion rate,

[47] volume is $4n_{\rm He} + n_{\rm H}$, so the fractional abundance by weight of helium is $Y = 4n_{\rm He}/\left(4n_{\rm He} + n_{\rm H}\right)$. But the numbers of protons and neutrons per unit proper volume are $n_p = 2n_{\rm He} + n_{\rm H}$ and $n_n = 2n_{\rm He}$, so the fraction of nucleons that are neutrons is $X_n = 2n_{\rm He}/\left(4n_{\rm He} + n_{\rm H}\right)$. Hence $Y = 2X_n$.

[6] V. Mukhanov, ref. 1.

[477] R. E. Lopez and M. S. Turner, Phys. Rev. D 59, 103502 (1999) [astro-ph/9807279].

[8] K. A. Olive and G. Steigman, Astrophys. J. Suppl. 97, 49 (1995); K. A. Olive, E. Skillman, and G. Steigman, Astrophys. J. 483, 788 (1997).

[9] Y. I. Izotov, T. X. Thuan, and V. A. Lipovetsky, Astrophys. J. 435, 647 (1994); Astrophys. J. Suppl. 108, 1 (1997); Y. I. Izotov and T. X. Thuan, Astrophys. J. 497, 227 (1998); 500, 188 (1998).

[10] A. Peimbert, M. Peimbert, and V. Luridiana, Astrophys. J. 565, 668 (2002); V. Luridiana, A. Peimbert, M. Peimbert, and M. Cerviño, Astrophys. J. 592, 846 (2003); M. Peimbert, V. Luridiana, and A. Peimbert, Astrophys. J. 667, (2007) [astro-ph/0701580].

[11] R. E. Lopez and M. S. Turner, ref. 7.

and for these the helium abundance continues to provide useful upper bounds.

The nuclear reactions that built up helium from free neutrons at $T \approx 10^9$ K were not perfectly efficient, but left over a small residue, [12] like an unburned ash, of the light elements H^2, H^3, He^3, Li^7, and Be^7. The nuclei of H^3 decayed later by β^+decay to He^3, and the nuclei of Be^7 decayed later by electron capture to Li^7, leaving us with H^2, He^3, and Li^7, as well as protons and He^4. The calculated abundances are shown in Figure 3.1.

The higher the baryon density, the more complete will be the incorporation of neutrons into He^4, and hence the smaller the resulting abundance of deuterium. We saw earlier that nucleosynthesis began when $X_d \simeq 1.2 \times 10^{-7}/\Omega_B h^2$, or $X_d \simeq 0.6 \times 10^{-5}$ for $\Omega_B h^2 = 0.02$. The deuterium fraction continued to rise for a while, as the temperature dropped and the exponential in Eq. (3.2.26) increased, but X_d then decreased again as the incorporation of free neutrons into deuterium reduced the factor X_n in Eq. (3.2.26), and deuterium was converted to H^3 and He^3, and thence to He^4. The final results, shown in Figure 3.1, are not very different from the deuterium fraction $X_d \simeq 1.2 \times 10^{-7}/\Omega_B h^2$ that we found at $T = 10^9$ K, and in particular exhibit a strong decline with increasing values of $\Omega_B h^2$.

The measurement of the deuterium abundance is complicated by the fact that deuterium has a small binding energy, and can readily be destroyed in stars. Any measurement of the deuterium abundance therefore gives a lower bound on the primordial deuterium abundance, and hence an upper bound on the baryon density.

In the past, the deuterium/hydrogen ratio was measured in various ways:

1. Interstellar medium. Spectroscopic studies of the interstellar medium gave a deuterium/hydrogen ratio [13] of $\left(1.60 \pm 0.09^{+0.05}_{-0.10}\right) \times 10^{-5}$.

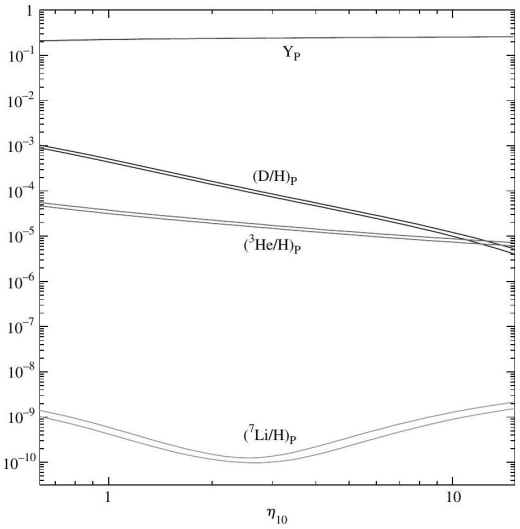


Figure 3.1: Calculated primordial abundances of deuterium, He^{3}, and Li^7 relative to hydrogen, and the fraction Y_P of the primordial mass of the universe in He^4, as functions of $\eta_{10} \equiv 10^{10}\eta$, where η is the ratio of nucleons to photons in the present universe. The widths of the bands indicate the effect of uncertainties in nuclear reaction rates. From G. Steigman, Int. J. Mod. Phys. E15, 1 (2006) [astro-ph/0511534].

2. Solar wind. Deuterium was converted to He^3 in the sun before it went onto the main sequence, so measurements of He^3 in the solar wind and meteorites is believed to give the

[4712] The calculated abundances cited below are given by S. Burles, K. M. Nollett, and M. S. Turner, Astrophys. J. 552, L1 (2001).

[13] J. L. Linsky et al., Astrophys. J. 402, 694 (1993); 451, 335 (1995).

total pre-solar value of He^3 and deuterium. Subtracting the abundance of He^3 obtained in other ways gave [14] a deuterium/hydrogen ratio of $(2.6 \pm 0.6 \pm 1.4) \times 10^{-5}$.

3. Jovian atmosphere. Spectroscopic studies of the Jovian atmosphere gave [15] a deuterium/hydrogen ratio of $(5 \pm 2) \times 10^{-5}$.

More recently, it has been possible to measure the deuterium/hydrogen ratio in very early intergalactic matter, by observing deuterium as well as hydrogen absorption lines in the spectra of quasistellar objects, due to

Table 3.3: Five quasi-stellar objects in whose spectra deuterium and hydrogen absorption lines are observed, together with the redshift of the intergalactic cloud responsible for the absorption, and the ratio of the numbers of deuterium and hydrogen atoms in the clouds inferred from the relative strength of the absorption lines, from ref. 17. (Observational uncertainties represent one standard deviation.)

QSO	z	D/H
PKS 1937-1009	3.572	$3.25 \pm 0.3 \times 10^{-5}$
Q1009+299	2.504	$3.98^{+0.59}_{-0.67} \times 10^{-5}$
HS 0105+1619	2.536	$2.54 \pm 0.23 \times 10^{-5}$
Q1243+3047	2.525675	$2.42^{+0.35}_{-0.25} \times 10^{-5}$
Q2206-199	2.0762	$1.65 \pm 0.35 \times 10^{-5}$

absorption in intervening intergalactic clouds of large redshift. [16] In 2003 the results of several years of observations of deuterium and hydrogen absorption lines in the spectra of four quasi-stellar objects were put together with results for one more QSO, [17] with the results shown in Table 3.3. The best value of the deuterium/hydrogen number ratio was found to be $2.78^{+0.44}_{-0.38} \times 10^{-5}$, from which is inferred a baryon/photon ratio $\eta = 5.9 \pm 0.5 \times 10^{-10}$, corresponding to

$$\Omega_B h^2 = 0.0214 \pm 0.0020.$$

Even for H_0 as small as 50 km/sec/Mpc, it was clear that Ω_B is much less than the fraction that all non-relativistic matter contributes to the critical density, which we have seen had been given as $\Omega_M \simeq 0.2$ by studies of galaxy clusters and as $\Omega_M \simeq 0.3$ by the redshift-distance relation of type Ia supernovae. It is this discrepancy that provided the original evidence for non-baryonic dark matter in the universe.

The discrepancy has become sharper through measurements of anisotropies in the cosmic microwave background. As we will see in Section 7.2, from these measurements it is possible to infer that $\Omega_B h^2 = 0.0223^{+0.0007}_{-0.0009}$ and $\Omega_M h^2 = 0.127^{+0.007}_{-0.013}$, indicating a total mass density that is from 5 to 6 times larger than the density of ordinary baryonic matter. The nature of the missing matter is discussed in Section 3.4.

The primordial abundance of He^3, like that of deuterium, is a monotonically decreasing function of the baryon/photon ratio. On the basis of a long term study of galactic HII regions

[4714] K. A. Olive, lectures given at the Advanced School on Cosmology and Particle Physics, Peniscola, Spain, June 1998, and Theoretical and Observational Cosmology Summer School, Cargese, France, August 1998 [astro-ph/9901231], and references therein.

[15] H. B. Niemann et al. Science 272, 846 (1996).

[4716] For a summary of early results, see K. A. Olive, ref. 13; S. Sarker, talk at the Second International Workshop on Dark Matter in Astro- and Particle Physics, Heidelberg, July 1998 [astro-ph/9903183].

[17] D. Kirkman, D. Tytler, N. Suzuki, J. M. O'Meara, and D. Lubin, Astrophys. J. Suppl. 149, 1 (2003) [astro-ph/0302006], and references cited therein.

and planetary nebula, a 2002 study [18] concluded that the He^3/H ratio in interstellar space is less than $1.1 \pm 0.2 \times 10^{-5}$, which is consistent with the value $1.04 \pm 0.06 \times 10^{-5}$ calculated for $\Omega_B h^2 = 0.0214$. But, unlike deuterium, He^3 is both produced and destroyed in stars, so it is not clear whether the observed interstellar abundance really represents the primordial abundance of He^3. Indeed, it had been thought that low-mass stars would inject a good deal of He^3 into the interstellar medium, in which case the apparent agreement between the amount observed in the interstellar medium and the amount expected from cosmological nucleosynthesis would actually represent a discrepancy. [19] This apparent discrepancy may have been removed by detailed calculations [20] of the movement of He^3 into the interior of these low-mass stars, which indicate that these stars do not in fact emit much He^3.

Some Li^6 is produced cosmologically, but in such small quantities (an abundance of about 10^{-13} to 10^{-14} that of hydrogen) that it has generally not been considered useful as a test of cosmological theories. [21] Much more attention has been given to Li^7. Its abundance has a more complicated dependence on the baryon/photon ratio, because Li^7 was formed in two different ways: directly, by the reactions $H^3 + He^4 \rightarrow Li^7 + \gamma$, and indirectly by $He^3 + He^4 \rightarrow Be^7 + \gamma$, followed much later by $e^- + Be^7 \rightarrow v + Li^7$. As we go to higher baryon densities, the amount of Li^7 produced directly increases at first, but then begins to decrease as Li^7 is destroyed in the reaction $p + Li^7 \rightarrow He^4 + He^4$. Eventually the indirect reaction takes over, and the Li^7 abundance rises again. The minimum Li^7/ hydrogen ratio is calculated to be about 2×10^{-10}, and is reached at a baryon/photon ratio of about 3×10^{-10}, corresponding to $\Omega_B h^2 = 0.01$. The observed Li^7/H ratio was reported [22] in 2000 to be $2.07^{+0.15}_{-0.04} \times 10^{-10}$, close to this minimum. A subsequent study [23] of 63 dwarf stars in the galactic halo gave a Li^7/H ratio of $(2.37 \pm 0.05) \times 10^{-10}$. Either result is less than the value 3×10^{-10} predicted for $\Omega_B h^2 = 0.0214$, but it is plausible that although

Li^7 is produced in stars and in the interaction of cosmic rays with matter, the Li^7 abundance has been depleted by convection in stellar atmospheres. Observation of Li^7 abundances in stars of varying temperature in the globular cluster NGC 6397 gave results in agreement with a theory of convective depletion of Li^7, with an assumed initial Li^7/H ratio equal to the expected value 3×10^{-10} (calculated taking $\Omega_B h^2$ to have the value estimated from deuterium abundance and microwave background anisotropies). [24] With the one possible exception of Li^7, there is now complete agreement between observations of light element abundances and calculations of cosmological nucleosynthesis, adopting the value $\Omega_B h^2$ provided by observations of anisotropies in the cosmic microwave background.

Although the baryon mass density inferred from cosmological nucleosynthesis and the cosmic microwave background is considerably less than the total mass density, it is also considerably greater than the density of baryonic matter observed in stars and luminous interstellar matter. [25] Some of this dark baryonic matter is in intergalactic space, but a fair fraction is believed to be present in galaxies, in the form of brown dwarf stars and clouds of hydrogen molecules. [26] But this is more a problem for the astrophysics of galaxies than for cosmology.

[4718] T. M. Bania, R. T. Rood, and D. S. Balser, Nature 415, 54 (2002).

[19] N. Hata et al., Phys. Rev. Lett. 75, 3977 (1995); K. A. Olive et al., Astrophys. J. 444, 680 (1995).

[20] P. P. Eggleston, D. S. P. Dearborn, and J. C. Lattanzio, Science 314, 1580 (2006).

[21] For a review, see E. Vangioni-Flam, M. Cassé, R. Cayrel J. Audouze, M. Spite, and F. Spite, New Astron. 4, 245 (1999).

[22] T. K. Suzuki, Y. Yoshii, and T. C. Beers, Astrophys. J. 540, 99 (2000) [astro-ph/0003164]. For earlier observations, see M. Spite and F. Spite, Nature 297, 483 (1982); S. G. Ryan, J. E. Norris, and T. C. Beers, Astrophys. J. 523, 654 (1999).

[23] J. Melendez and I. Ramirez, Astrophys. J. 615, L33 (2004).

10.3.3 3.3 Baryonsynthesis and leptonsynthesis

We saw in the previous section that the ratio η of nucleons to photons at the time of nucleosynthesis had the tiny value $\simeq 5 \times 10^{-10}$. At earlier times, when the temperature was above 10^{13} K, nucleons would not yet have formed from their three constituent quarks, and there would have been roughly as many quark-antiquark pairs in thermal equilibrium as photons. But the conservation of baryon number (one-third the number of quarks minus the number of antiquarks) during the annihilation process tells us that before annihilation there must have been a slight excess, roughly of order η per photon, of quarks over antiquarks, so that some quarks would survive to form nucleons when all the antiquarks had annihilated with quarks. There was also a slight excess of electrons over positrons, to maintain the charge neutrality of the universe. It is conceivable that there is a compensating excess of antineutrinos over neutrinos, so that the total lepton number density (the number density of electrons, muons, tauons, and neutrinos, minus the number density of their antiparticles) vanishes, but it seems more natural to assume that before lepton-antilepton annihilation

there was also a slight imbalance of leptons and antileptons, comparable to the excess of quarks over antiquarks.

These tiny imbalances in the numbers of quarks and antiquarks and of leptons and antileptons might be explained if the baryon and lepton number densities were generated by physical processes in a universe that at some early time had equal number of particles and antiparticles of all sorts. We could then hope to calculate η from first principles, and understand why it is so small. In 1967 Sakharov [1] outlined three conditions that must be met for this to be possible:

1. Obviously, in order for an excess of baryons over antibaryons or an inequality of leptons and antileptons to arise in a universe that begins with equal numbers of particles and antiparticles of each type, some physical processes must violate the conservation of baryon number or lepton number.

2. A universe with equal numbers of particles and antiparticles of each type (and each momentum and helicity) is invariant under the symmetry operators C (the exchange of particles with antiparticles) and CP (the exchange of particles with antiparticles, combined with a change of sign of all three-dimensional coordinate vectors), while a state with an excess of baryons over antibaryons or an imbalance of leptons and antileptons is clearly not invariant under either C or CP. Hence to produce such an state, some physical process must violate invariance under both C and CP . It is true that, whether or not C and/or CP are conserved, any relativistic quantum field theory will respect a symmetry [2] under CPT, the simultaneous exchange of particle with antiparticles, combined with a change of sign of all three-dimensional coordinate vectors, combined with a change in the direction of time's flow, but this does not prevent the production of baryon or lepton number, because the time-reversal symmetry T is violated by the expansion of the universe.

3. A little less obviously, in order to produce an excess of baryons over antibaryons or an imbalance of leptons and antileptons out of a state with equal numbers of particles and antiparticles, the universe must at some time depart from a state of thermal (including chemical) equilibrium. This is because in a state of thermal equilibrium, if baryon and/or lepton conservation are not respected, and all conserved quantities like electric charge vanish, then since chemical potentials must be conserved in all reactions, all chemical

[4][7][24] A. J. Korn et al., Nature 442, 657 (2006) [astro-ph/0608201].
[25] M. Fukugita, C. J. Hogan, and P. J. E. Peebles, Astrophys. J. 503, 518 (1998).
[26] See e.g. M. Roncadelli, Recent Research Devel. Astron. & Astrophys. 1, 407 (2003).

potentials must vanish. The CPT symmetry implies that even if C and CP are not conserved, the masses of particles and their

antiparticles are precisely equal, so with vanishing chemical potentials particles and antiparticles will have identical distribution functions, such as (3.1.28) with $\mu = 0$ for ideal gases. Thus, whatever the rates for various processes, no net baryon or lepton number will be produced. [3] An exception to this reasoning is presented at the end of this section.

All three conditions are now known to be satisfied:

(1) There is no direct experimental evidence for the nonconservation of baryon number, but a very weak baryon number nonconservation is expected according to modern views of the standard model of elementary particles. According to these views, the standard model is not a fundamental theory, which might be expected to be contain only interactions whose coupling parameters are either dimensionless or proportional to positive powers of mass, so that all infinities that arise in the standard model can be absorbed into a renormalization of these coupling parameters. Rather, we now think that the standard model is only an effective field theory, valid at energies much less than some fundamental mass scale M, which might be the Planck mass $M_{\mathcal{P}} \equiv G^{-1/2} = 1.22 \times 10^{19}$GeV, or perhaps the energy scale $\approx 10^{15}$ to 10^{16}GeV at which the three independent (suitably normalized) gauge coupling parameters of the standard model become equal. We would expect such an effective field theory to contain every possible interaction allowed by the gauge symmetries of the strong, weak, and electromagnetic interactions, but all but a finite number of these coupling parameters will have the dimensions of negative powers of mass. These "non-renormalizable" couplings are thus suppressed at energy $E \ll M$ by powers of E/M. Now, the gauge symmetries of the standard model do not allow any unsuppressed interactions among quarks and leptons that violate baryon or lepton number, so this picture makes it plausible that baryon and lepton number would be automatically conserved to a good approximation for energies $E \ll M$, even if baryon and lepton conservation are not respected by whatever fundamental theory describes physics at energies of order M. In other words, baryon and lepton number conservation may be mere "accidental" symmetries. In this case, there is no reason to exclude any suppressed interactions that violate conservation of baryon and lepton number. The least suppressed interactions of this sort are an interaction involving two lepton doublets and two scalar doublets, which is suppressed by a factor M^{-1} and violates lepton but not baryon conservation, and an interaction suppressed by a factor M^{-2} involving three quark fields and one lepton field, which violates both lepton and baryon

conservation. [4] (Also, as we shall see below, there is a quantum-mechanical violation of baryon-number and lepton-number conservation in the standard electroweak theory.) There is experimental evidence for the first but not the second interaction. With a coupling parameter of order $1/\,(10^{16}\text{GeV})$, when the neutral scalar fields are replaced by their vacuum expectation values, this interaction provides a neutrino mass of order 10^{-2}eV, in good agreement with the results of neutrino oscillation experiments. (The existence of non-zero neutrino masses means that helicity $+1/2$ neutrinos, which are conventionally assigned lepton number $+1$, can be changed to helicity $-1/2$ neutrinos with lepton number -1 by a Lorentz transformation.) The second interaction would lead to decay processes like $p \to \pi^0 + e^+$, with decay rates of order $m_p^5/M^4\hbar$. Such events have not been seen, but could easily have escaped detection.

(2) The violation of invariance under C was discovered in 1957, while the violation of invariance under CP was discovered in 1964.[5]

(3) The expansion of the universe tends to pull states out of thermal equilibrium, either because the cooling temperature makes reaction rates decrease below the expansion rate, or because

[471] A. D. Sakharov, JETP Lett. 5, 24 (1967).

[2] See QTF, Vol. I, Sec. 5.8.

[473] For a more detailed argument and references to earlier discussions, see S. Weinberg, Phys. Rev. Lett. 42, 850 (1978).

as it cools the universe goes through first-order phase transitions, similar to the condensation of water vapor or the freezing of liquid water. Although this in itself does not violate the conservation of baryon or lepton number, it opens the door for physical processes that do violate these conservation laws, as well as C and CP conservation, to create an imbalance between baryons and antibaryons and/or between leptons and antileptons.

All this just goes to show that it is possible for physical processes to produce a non-zero cosmological baryon and lepton number. It remains to find a specific theory in which the observed ratio of baryons to photons could be produced. [6] There are several theories of this type:

1. Delayed decay of heavy particles [7]

Suppose there is a species of heavy " X " particle, which decays into a pair of different channels, with baryon numbers B_1 and B_2 and lepton numbers L_1 and L_2, and branching ratios r and $1-r$. (For instance, in some grand

unified theories there are "leptoquarks" with mass of order 10^{15}GeV, that decay into either two quarks, with baryon number 2/3 and lepton number zero, or into a lepton and antiquark, with baryon number $-1/3$ and lepton number $+1$.) The antiparticle will then decay into channels with baryon numbers $-B_1$ and $-B_2$ and lepton numbers $-L_1$ and $-L_2$, with branching ratios $\bar{r}$ and $1-\bar{r}$. (Invariance under CPT tells us that any particle has the same total decay rate as its antiparticle, but as long as C and CP conservation are violated, it is possible for particles and antiparticles to have different branching ratios for their different decay channels.) The average total baryon number produced in the decay of one X particle and the decay of one of the corresponding antiparticles is then

$$\Delta B = rB_1 + (1-r)B_2 - \bar{r}B_1 - (1-\bar{r})B_2 = (r-\bar{r})\,(B_1 - B_2)\,, \tag{3.3.1}$$

and likewise for lepton number.
Similarly, there may be some heavy " N " particle that is its own antiparticle, and that decays with branching ratio r into one channel with baryon number B_1 and lepton number L_1, and with branching ratio $1-r$ into the antichannel with baryon number $-B_1$ and lepton number $-L_1$. On the average, each decay produces a baryon number

$$\Delta B = rB_1 - (1-r)B_1 \tag{3.3.2}$$

and likewise for lepton number. If C and CP conservation are violated, then it is possible to have $r \neq 1-r$, so that a net baryon number and lepton number may be produced. For instance, in some grand unified theories there are neutral fermions that decay both into scalar particles and leptons and into their antiparticles, producing a net lepton number if the branching ratios for these channels are unequal.

Such processes do not produce any net baryon or lepton number in equilibrium, because the inverse to the decay processes will destroy precisely as much baryon and lepton number as the decay processes create. The conditions for thermal equilibrium will be violated if H falls below the X-particle decay rate, but as long as $k_{\mathcal{B}}T$ remained above all particle masses,

[474] For a review, with references to the original literature, see QTF, Vol. II, Section 21.5.
[5] See QTF, Vol. I, Section 3.3.
[6] For the earliest attempts in this direction, see M. Yoshimura, Phys. Rev. Lett. 41, 281 (1978); 42, 746(E) (1979); S. Dimopoulos and L. Susskind, Phys. Rev. D 18, 4500 (1979); Phys. Lett. 81B, 416 (1979); A. Yu. Ignatiev, N. V. Krosnikov, V. A. Kuzmin, and A. N. Tavkhelidze, Phys. Lett. 76B, 436 (1978); B. Toussaint, S. B. Treiman, F. Wilczek, and A. Zee, Phys. Rev. D 19, 1036 (1978); J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, Phys. Lett. 80B, 360 (1979); 82B, 464(E) (1979).
[7] S. Weinberg, Phys. Rev. Lett. 42, 850 (1978).

whatever the rates of various processes, the expansion preserved the equilibrium form for all particle distributions, with a redshifted temperature $T \propto 1/a$. Specifically, if $k_{\mathcal{B}}T$ was still above all particle masses when $H \equiv \dot{a}/a$ fell below the decay rate Γ_X, the inverse decay process would re-create as many X particles as had decayed. On the other hand, if at the time that $H \approx \Gamma_X, k_{\mathcal{B}}T$ was less than the mass m_X of the X particles, then the Boltzmann factor $e^{-m_X/k_{\mathcal{B}}T}$ would have blocked the inverse decay, and the X particles and antiparticles would have disappeared, yielding a net baryon number $(r-\bar{r})(B_1 - B_2)$ for each original heavy particle-antiparticle pair.

The condition for this to work is then that $k_{\mathcal{B}}T \leq m_X$ at the temperature when $H = \Gamma_X$. We can estimate that (here taking $\hbar = c = 1$)

$$H = \sqrt{\frac{8\pi G a_{\mathcal{B}} T^4 (\mathcal{N}/2)}{3}} = 1.66\,(k_{\mathcal{B}}T)^2\,\mathcal{N}^{1/2}/m_{\mathcal{P}} \tag{3.3.3}$$

where $m_{\mathcal{P}} \equiv G^{-1/2} = 1.22 \times 10^{19}$GeV, and $\mathcal{N}$ is the total number of helicity states of all elementary particles and antiparticles, with an extra factor 7/8 for fermions. (We assume here that the main contribution to the energy density of the universe at these temperatures comes from the large number $\mathcal{N}$ of highly relativistic particle types, rather than from the X particles themselves.) The decay rate of the X particle will be m_X times some dimensionless parameter α_X, which characterizes the strength of the interactions responsible for the decay and the number of decay channels. Hence decays start to be significant at a temperature T_X, given by

$$k_{\mathcal{B}}T_X \approx \sqrt{\alpha_X m_X m_{\mathcal{P}}/\mathcal{N}^{1/2}} \tag{3.3.4}$$

For this to be less than m_X, we must have

$$m_X \geq \mathcal{N}^{-1/2}\alpha_X m_{\mathcal{P}} \tag{3.3.5}$$

This is a fairly severe lower bound on m_X. For instance, if the X particles decay through ordinary electroweak interactions, then $\alpha_X \approx 10^{-2}$, so if $\mathcal{N} \approx 100$ we must have m_X greater than about 10^{16}GeV. From the point of view of theories that unify the strong and electroweak interactions, this is not an unreasonable value for the mass.

Assuming that this condition is satisfied, we can easily use Eq. (3.3.1) to make an estimate of η. At temperatures far above the heavy particle mass the number density of pairs of X particles and antiparticles is of the same order as the number density of photons. The entropy density (using energy units for temperature, with $k_{\mathcal{B}} = 1$) at this time is of the order of $\mathcal{N}$ times the number density of photons, so the ratio of the number densities of X and $\bar{X}$ pairs to the entropy density is of the order of $1/\mathcal{N}$. With Eq. (3.3.4) satisfied, after the disappearance of these pairs the ratio of the baryon number to the entropy densities will be of order $(r-\bar{r})(B_1 - B_2)/\mathcal{N}$. The entropy density varies as a^{-3}, and provided that baryon number is subsequently conserved so does the baryon number density, and so the baryon number to entropy ratio will remain unchanged. The present entropy density is of the order of the photon number density, and at present the only baryons are nucleons, with no antinucleons, so we expect a nucleon to photon ratio

$$\eta \approx (r-\bar{r})(B_1 - B_2)/\mathcal{N} \tag{3.3.6}$$

Typically $B_1 - B_2$ is of order unity (in the leptoquark example mentioned above, it is equal to unity) but $r-\bar{r}$ is generally very small, both because CP conservation is weakly violated, and because the CPT theorem tells us that $r = \bar{r}$ in the lowest order of perturbation theory. The precise value of $r - \bar{r}$ is very model-dependent, but values of η of the desired order, 5×10^{-10},

appear quite natural. [8] This idea runs into difficulty in inflationary theories, which as we will see in Chapter 10 generally require that the temperature was never high enough to produce particles with masses satisfying Eq. (3.3.5).

2. Nonperturbative electroweak baryon and lepton number nonconservation

In the standard model of weak, electromagnetic, and strong interactions baryon and lepton number are automatically conserved to all orders of perturbation theory, but not when certain non-perturbative effects are taken into account. [9] This produces reactions that violate baryon and lepton number conservation, but such reactions are suppressed at low temperatures by a factor $\exp(-8\pi^2/g^2) \simeq 10^{-162}$, where g is the $SU(2)$ electroweak coupling constant. This tiny exponential is actually a barrier penetration factor, which accompanies the quantum mechanical tunneling transition through the barrier between topologically different configurations of the gauge fields. At high temperatures, above about 300 GeV , thermal fluctuations allow passage over this barrier, and the exponential suppression disappears. [10] Nevertheless, by themselves these reactions do not produce an appreciable net baryon or lepton number, both because they take place at a time of nearly perfect thermal equilibrium, [11] and because they are suppressed by small parameters associated with the need to violate CP conservation as well as the conservation of baryon and lepton number. [12]

3. Leptogenesis [13]

Although the non-perturbative effects of electroweak interactions described in the previous paragraph do not by themselves provide a way of accounting

for the observed baryon/photon ratio, they can convert a cosmological lepton number density into a baryon number density, or vice versa. The only truly conserved quantum numbers in the $SU(3) \times SU(2) \times U(1)$ standard model are those associated with gauge symmetries - the electroweak isospin generator T_3, the electroweak hypercharge Y (defined so that the electric charge in units of e is $T_3 - Y$), and a pair of generators of the $SU(3)$ gauge group of quantum chromodynamics - together with $B-L$, the total baryon number minus the total lepton number, which is conserved because of a cancelation between Feynman diagrams containing loops of quarks or leptons. This creates a further problem for the proposal that the observed baryon number density of the universe is created in the decay of a leptoquark into both two quarks and into a lepton and antiquark. Although these channels have different values for B and L, they both have the same value (equal to 2/3) for $B - L$. Hence if the universe starts with equal numbers of particles and antiparticles of all types, then even if leptoquark decay produces equal non-zero baryon and lepton number densities, all truly conserved quantities will remain zero, so when thermal equilibrium is established at lower temperatures it will be with zero values

[478] D. V. Nanopoulos and S. Weinberg, Phys. Rev. D 20, 2484 (1979).

[9] G. 't Hooft, Phys. Rev. Lett. 37, 8 (1976). Also see QTF, Vol. II, Section 23.5.

[10] V. A. Kuzmin, V. A. Rubakov, and M. E. Shaposhnikov, Phys. Lett. 155B, 36 (1985). The transition between field configurations is dominated by intermediate field configurations known as sphalerons; see N. S. Manton, Phys. Rev. D 28, 2019 (1983); F. R. Klinkhammer and N. S. Manton, Phys. Rev. D 30, 2212 (1984); R. F. Dashen, B. Hasslacher, and A. Neveu, Phys. Rev. D 10, 4138 (1974).

[11] The absence of a first-order phase transition in the electroweak theory is shown by K. Kajantie, M. Laine, K. Rummukainen, and M. Shaposhnikov, Nucl. Phys. B 466, 189 (1996); K. Rummukainen, M. Tsypin, K. Kajantie, and M. Shaposhnikov, Nucl. Phys. B 532 (1998); F. Csikor, Z. Fodor, and J. Heitger, Phys. Rev. Lett. 82, 21 (1999); and earlier references cited therein.

[12] M. E. Shaposhnikov, JETP Lett. 44, 465 (1986); Nucl. Phys. B 287, 757 (1987).

[13] M. Fukugita and T. Yanagida, Phys. Lett. B 174, 45 (1986). For a review, see W. Buchmüller, R. D. Peccei, and T. Yanagida, Ann. Rev. Nucl. Part. Sci. 55, 311 (2005) [hep-ph/0502169].

for all chemical potentials. (This is demonstrated below.) Such a state has equal numbers of particles and antiparticles, and hence zero densities of B and L as well as $B-L$.

On the other hand, if some heavy particle (such as the N particle mentioned above) in the early universe decays in such a way as to produce a non-zero density of $B-L$ this will persist through the period of thermal equilibrium, though the relative densities of B and L may change. In general, whatever mixture of baryon and lepton number is produced when the heavy particle decay, and even if only lepton number is produced, we would expect the densities of baryon and lepton number to be comparable in a subsequent period of thermal equilibrium.

This can be made quantitative. [14] Suppose in thermal equilibrium there are a set of conserved quantum numbers Q_a, such as T_3, Y, and $B-L$. Suppose also that there are several species i of particles in equilibrium, such as quarks, leptons, etc., each carrying a value q_{ai} for the quantum number Q_a. The chemical potentials μ_i for these particles must be conserved for all reactions in thermal equilibrium, which requires that they be linear combinations of the conserved quantum numbers:

$$\mu_i = \sum_a q_{ai}\mu_a \tag{3.3.7}$$

with coefficients μ_a that can be regarded as chemical potentials for the different conserved quantities. The densities of the different particle species

can then be expressed as functions of the μ_a and the temperature, and these relations can be used to calculate the densities n_a of the different conserved quantum numbers as functions of the μ_a and the temperature. But there are just as many n_a as there are μ_a, so these relations can be inverted to give the μ_a in terms of the n_a and the temperature, from which we can calculate the density of anything else as functions of the n_a and the temperature, including the density of a nonconserved quantity like baryon or lepton number that is not among the Q_a.

In the case that interests us here, at temperatures above about 10^{16} K, all particles of the Standard Model are highly relativistic, so that their masses can be neglected. The number density of particle species i is then

$$\begin{aligned} n_i &= \frac{g_i}{(2\pi\hbar)^3}\int \frac{d^3p}{e^{(p-\mu_i)/k_{\mathcal{B}}T}\mp 1} \\ &= 4\pi g_i\left(\frac{k_{\mathcal{B}}T}{2\pi\hbar}\right)^3\int_0^\infty \frac{x^2dx}{e^{x-\mu_i/k_{\mathcal{B}}T}\mp 1} \end{aligned} \tag{3.3.8}$$

where g_i is the number of helicity (and other sources of multiplicity) states for each species, and the $\mp$ sign is - for bosons and + for fermions. The antiparticle density $\bar{n}_i$ will be given by the same formula, but with μ_i replaced with $-\mu_i$, so the difference is

$$n_i - \bar{n}_i = 8\pi g_i\left(\frac{k_{\mathcal{B}}T}{2\pi\hbar}\right)^3 \sinh\left(\frac{\mu_i}{k_{\mathcal{B}}T}\right)\int_0^\infty \frac{x^2e^x dx}{e^{2x}\mp 2e^x\cosh\left(\mu_i/k_{\mathcal{B}}T\right)+1} \tag{3.3.9}$$

In the situation that concerns us here, the imbalance between particles and antiparticles of all sorts is small, so $|\mu_i| \ll 1$ for all particle species. In this case,

$$n_i - \bar{n}_i = 8\pi g_i\left(\frac{k_{\mathcal{B}}T}{2\pi\hbar}\right)^3 \frac{\mu_i}{k_{\mathcal{B}}T}\int_0^\infty \frac{x^2e^x dx}{\left(e^x\mp 1\right)^2} \tag{3.3.10}$$

The integral over x has the value $\pi^2/3$ for bosons and $\pi^2/6$ for fermions, so we can write this as

[4714] J. A. Harvey and M. S. Turner, Phys. Rev. D 42, 3344 (1990).

$$n_i - \bar{n}_i = f(T)\tilde{g}_i\mu_i \tag{3.3.11}$$

where

$$f(T) \equiv \frac{4\pi^3}{3}\frac{(k_{\mathcal{B}}T)^2}{(2\pi\hbar)^3} \tag{3.3.12}$$

and $\tilde{g}_i$ is the number of spin states, but with an extra factor of 2 for bosons. Using Eq. (3.3.7), Eq. (3.3.11) becomes

$$n_i - \bar{n}_i = f(T)\tilde{g}_i \sum_a q_{ai}\mu_a \tag{3.3.13}$$

In particular, the density of the conserved quantum number Q_a is

$$n_a = \sum_i q_{ai}\,(n_i - \bar{n}_i) = f(T)\sum_b M_{ab}\mu_b, \tag{3.3.14}$$

where M is the matrix

$$M_{ab} \equiv \sum_i \tilde{g}_i q_{ai} q_{bi} \tag{3.3.15}$$

This matrix is positive-definite (in the sense that, for any set of real numbers ξ_a, we have $\sum_{ab} M_{ab}\xi_a\xi_b > 0$, unless all ξ_a vanish), and therefore it has an inverse M^{-1}. We can thus invert the relation (3.3.14), and find $\mu_a = \sum_b M^{-1}_{ab} n_b/f(T)$. Using this in Eq. (3.3.13) gives

$$n_i - \bar{n}_i = \sum_{ab} \tilde{g}_i q_{ai} M^{-1}_{ab} n_b \tag{3.3.16}$$

for any particle species i. Note in particular that if the densities n_b of all the conserved quantum numbers vanish, then there is an equal number of particles and antiparticles of every kind, as mentioned above.

In order to deal with the case where some conserved quantities such as $B-L$ have non-zero densities, we need to calculate the matrix M_{ab}. The particles of the Standard Model are listed in Table 3.4. For N_g generations of quarks and leptons and N_d scalar doublets, the independent elements of the matrix M_{ab} are

$$M_{B-LB-L} = \frac{13N_g}{3}, \quad M_{B-LY} = -\frac{8N_g}{3}, \quad M_{YY} = \frac{10N_g}{3} + N_d \tag{3.3.17}$$

We don't need any of the matrix elements involving T_3, because the sum of the T_3 values vanishes for all the particles with any given values of $B-L$ and Y, so that $M_{B-LT_3} = M_{YT_3} = 0$. That is, the matrix M_{ab} is blockdiagonal, with a 2×2 block having a and b running over $B-L$ and Y. This has the consequence that without bothering to calculate $M_{T_3T_3}$, we can calculate that

$$M^{-1}_{B-LB-L} = \frac{10N_g}{3D} + \frac{N_d}{D}, \quad M^{-1}_{B-LY} = \frac{8N_g}{3D}, \quad M^{-1}_{YY} = \frac{13N_g}{3D} \tag{3.3.18}$$

where D is the determinant

$$D = \frac{22N_g^2}{3} + \frac{13N_gN_d}{3} \tag{3.3.19}$$

(Similarly, because the sum of the color quantum numbers vanishes for all the particles with any given values of $B-L$ and Y, we do not need to take

Table 3.4: Particles of the Standard Model, together with the number $\tilde{g}$ of their helicity and color states (with an extra factor 2 for bosons), and the values of their baryon number, lepton number, and gauge quantum numbers. Only one "generation" of quarks and leptons and only one doublet of scalar fields are shown. The subscripts L and R denote the helicity states of quarks u and d and leptons v and e. Antiparticles are not shown separately, and the photon and Z^0 are not shown because they are their own antiparticles, and so do not contribute to the densities of any quantum numbers. Color quantum numbers are not shown, for reasons given in the text.

Particle	$\tilde{g}$	B	L	T_3	Y
u_L	3	1/3	0	1/2	-1/6
d_L	3	1/3	0	-1/2	-1/6
u_R	3	1/3	0	0	-2/3
d_R	3	1/3	0	0	1/3
v_L	1	0	1	1/2	1/2
e_L	1	0	1	-1/2	1/2
e_R	1	0	1	0	1
W^+	4	0	0	1	0
φ^+	2	0	0	1/2	-1/2
φ^0	2	0	0	-1/2	-1/2
gluons	4	0	0	0	0

the color quantum numbers into account here.) Thus if $B-L$ is the only conserved quantum number with a non-vanishing number density, then Eqs. (3.3.16) and (3.3.18) tell us that the baryon number density in thermal equilibrium is

$$\begin{aligned} n_B &\equiv \sum_i B_i \left(n_i - \tilde{n}_i\right) = \sum_i \tilde{g}_i B_i \left((B-L)_i M^{-1}_{B-LB-L} + Y_i M^{-1}_{YB-L}\right) n_{B-L} \\ &= \left(\frac{4}{3} M^{-1}_{B-LB-L} - \frac{2}{3} M^{-1}_{YB-L}\right) N_g n_{B-L} \\ &= \left(\frac{8N_g + 4N_d}{22N_g + 13N_d}\right) n_{B-L} \end{aligned} \tag{3.3.20}$$

For instance, in the minimal experimentally allowed case, with $N_g = 3$ and $N_d = 1$, this gives $n_B = (28/79)n_{B-L}$. In any case, n_B turns out to be of
the same order of magnitude as n_{B-L}, as anticipated above. The reason for the smallness of n_B/n_γ in this scenario would be traced to the smallness of CP violation in the out-of-equilibrium heavy particle decay that produces a non-vanishing density of $B - L$.

4. Affleck-Dine mechanism [15]

It is possible for baryon number nonconservation to occur in the nonequilibrium dynamics of a scalar field that carries a non-zero baryon number. Both baryon-number conservation and CP-invariance need to be violated in the Lagrangian of the scalar field. Such theories find their motivation in supersymmetry, which lies outside the scope of this book.

5. Equilibrium baryon synthesis

The violation of CPT by the expansion of the universe means that the part of the Hamiltonian that is odd in C and CP can have a non-zero expectation value ΔE_i in the state of a single

baryon of type i, and of course opposite expectation value $-\Delta E_i$ in the state of the corresponding antibaryon. Then even the universe starts in a state of zero baryon number, if it enters a state of thermal equilibrium in which baryon-number nonconserving processes occur rapidly, although there will be no chemical potential associated with baryon number, the difference in energy of baryons and antibaryons will lead to a net baryon number. When baryon-number nonconserving processes become ineffective the resulting baryon number density will survive, simply decreasing as a^{-3}. As long as baryon-conserving collisions remain sufficiently rapid, the one-particle distribution will have the form appropriate for thermal equilibrium, but now with a non-vanishing baryonic chemical potential.

As a general class of theories of this sort, suppose that in the expanding universe, there is a term in the Lagrangian density of the form

$$\Delta\mathcal{L}(x) = -\sqrt{-\operatorname{Det} g}V_\mu(x)J_B^\mu(x), \tag{3.3.21}$$

where $V_\mu(x)$ is a classical vector field and $J_B^\mu(x)$ is the current associated with baryon number (for which $J_B^0(x)$ is the baryon density, which is odd under C and CP). Two proposals of this sort have been made: the vector field could be $V_\mu = M^{-1}\partial_\mu\varphi$, where φ is some scalar field [16] and M is some large mass; or the vector field could be $V_\mu = M^{-2}\partial_\mu R$, where R is the curvature scalar [17] and again M is some large mass. In any case, the isotropy and homogeneity of the Robertson-Walker metric requires that

V_i vanishes, while V_0 is a function only of time. This interaction then shifts the energy of the state of a single particle of type i with baryon number b_i by an amount $\Delta E_i = V_0(t)b_i$. In thermal (including chemical) equilibrium, the baryon number density will be

$$n_B(t) = \sum_i b_i g_i \int \frac{4\pi p^2 dp}{(2\pi\hbar)^3}\left[e^{\left(\sqrt{p^2+m_i^2}+b_iV_0(t)\right)/k_{\mathcal{B}}T(t)} \mp 1\right]^{-1} \tag{3.3.22}$$

where the sum over i runs over all particle (and antiparticle) types; g_i and b_i are the number of spin states and the baryon number of a particle of type i; and the upper and lower signs again apply to bosons and fermions, respectively. If baryon non-conserving collisions shut off suddenly at time t_1, then subsequently the baryon number density will be

$$n_B(t) = \left(\frac{a\left(t_1\right)}{a(t)}\right)^3 \sum_i b_i g_i \int \frac{4\pi p^2 dp}{(2\pi\hbar)^3}\left[e^{\left(\sqrt{p^2+m_i^2}+b_iV_0(t_1)\right)/k_{\mathcal{B}}T(t_1)} \mp 1\right]^{-1} \tag{3.3.23}$$

For $V_0 = 0$ the cancelation between baryons and antibaryons of course makes this vanish. Since the baryon/entropy ratio is known to be small, we expect $V_0\left(t_1\right)$ to be small; to first order in $V_0\left(t_1\right)$, the baryon number density for $t > t_1$ will be

$$\begin{aligned} n_B(t) = &\left(\frac{a\left(t_1\right)}{a(t)}\right)^3 \frac{V_0\left(t_1\right)}{k_{\mathcal{B}}T\left(t_1\right)} \sum_i b_i^2 g_i \\ &\times \int \frac{4\pi p^2 dp}{(2\pi\hbar)^3} e^{\sqrt{p^2+m_i^2}/k_{\mathcal{B}}T(t_1)}\left[e^{\sqrt{p^2+m_i^2}/k_{\mathcal{B}}T(t_1)} \mp 1\right]^{-2} \end{aligned} \tag{3.3.24}$$

[4715] L. Affleck and M. Dine, Nucl. Phys. B 249, 361 (1985); M. Dine, L. Randall, and S. Thomas, Phys. Rev. Lett. 75, 398 (1995); Nucl. Phys. B 458, 291 (1996).

[16] A. G. Cohen and D. B. Kaplan, Phys. Lett. B 199, 251 (1987). The production of baryon number is suppressed when the scalar field oscillates rapidly; see A. Dolgov, K. Freese, R. Rangarajan, and M. Srednicki, Phys. Rev. D 56, 6155 (1997) [hep-ph/9610405].

[17] H. Davoudiasl, R. Kitano, G. D. Kribs, H. Murayama, and P. J. Steinhardt, Phys. Rev. Lett. 93, 201301 (2004).

A similar mechanism could also be responsible for lepton synthesis.
The crucial confirmation of any theory of baryon synthesis would be a successful prediction of the present baryon/photon ratio. So far, none of the proposals discussed here are anywhere near this goal.

10.3.4 3.4 Cold dark matter

We saw in Section 3.2 that considerations of cosmological nucleosynthesis lead to the conclusion that most of the mass in the universe is not in the form of ordinary baryonic matter, i.e. atomic nuclei and electrons. We will see in Chapter 7 that this conclusion is powerfully reinforced by observations of anisotropies in the cosmic microwave background. So we face the question, if the particles that make up most of the mass of the universe are not baryons, then what are they?

We know that this matter is dark, in the sense that it does not interact significantly with radiation, both because we don't see it, and also because it has not lost its kinetic energy sufficiently to relax into the disks of galaxies, as has baryonic matter. This means in particular that these particles must be electrically neutral. [1] Detailed studies of the dynamics of galaxy clusters indicate that the dark matter particles must also be cold, in the sense that their velocities are highly non-relativistic. [2]

The study of a double galaxy cluster 1E0657-558 (the "bullet cluster," with $z = 0.296$) has provided vivid direct evidence of the existence of dark matter, which does not have non-gravitational interactions with itself or with ordinary baryonic matter. [3] The galaxies in this cluster are mostly grouped into two distinct subclusters, while hot gas (observed through its emission of X-rays) is concentrated between these subclusters. The interpretation is that two clusters of galaxies have collided; the galaxies which have little chance of close encounters have mostly continued on their original paths, while the two clouds of hot gas that previously accompanied them have collided and remained closer to the center of the double cluster. The total matter density in 1E0657-558 is mapped out through its effect in gravitationally deflecting light from more distant galaxies along the same line of sight. (Gravitational lensing is discussed in Chapter 9.) In this way, it is found that most of the matter in 1E0657-558 is not associated with the hot gas, but like the galaxies forms two subclusters that have evidently passed through each other without appreciable interaction. The ratio of the mass in hot gas to the mass in all matter is estimated to be about 1/6, in line with the value of Ω_B/Ω_M previously inferred from measurements of deuterium abundance and luminosity distance as a function of redshift, or from anisotropies in the cosmic microwave background.

Elementary particle theory offers several candidates for the particles making up the cold dark matter.

A. Weakly interacting massive particles (WIMPs)

Massive particles may survive to the present if they carry some sort of conserved additive or multiplicative quantum number. If there is a non-zero

[471] There are particularly strong limitations on the number density of any sort of charged stable particles that might be left over from the big bang, which are set by mass spectroscopy, the analysis of the charged particles contained in samples of matter according to their ratio of mass to charge. The number of electrically charged exotic particles with masses in the range of 6 GeV to 330 GeV has been found to be less than 10^{-21} of the number of nucleons, by P. F. Smith and J. R. J. Bennett, Nucl. Phys. B 149, 525 (1979).
[2] P. J. E. Peebles, Astrophys. J. 263, L1 (1983); G. R. Blumenthal, S. M. Faber, J. R. Primack, and M. J. Rees, Nature 311, 517 (1984).
[3] D. Clowe et al., Astrophys. J. 648, L109 (2006) [astro-ph/0608407].

chemical potential associated with this quantum number, then of course some particles (or antiparticles) must be left over after all the antiparticles (or particles) have annihilated. But even if there are no non-zero chemical potentials for these particles, so that the initial number densities of particles and antiparticles are equal, if they can only annihilate with their antiparticles then once their number density becomes sufficiently low the collision rate eventually becomes too small to reduce the density further. [4] We will call these particles L-particles (for "left-over"), to distinguish them from the other particles into which they may annihilate, which we will assume are all approximately in thermal and chemical equilibrium during the period of annihilation. The annihilation rate per particle of the L particles and antiparticles is $n\langle\sigma v\rangle$, where n is their number density, and $\langle\sigma v\rangle$ is the average value of the product of annihilation cross section and relative velocity. The rate of decrease in the number of L particles in a co-moving volume a^3 is then $na^3 \times n\langle\sigma v\rangle$. There is also an n-independent rate of creation of these particle-antiparticle pairs from the thermal background. Since this must balance the annihilation rate when everything is in equilibrium, in general the creation rate per volume a^3 must equal $n_{\rm eq}^2\, a^3\langle\sigma v\rangle$, where $n_{\rm eq}$ is the number density of L particles and of antiparticles in equilibrium. The number na^3 of L particles and of antiparticles in a co-moving volume a^3 is therefore governed by a Boltzmann equation

$$\frac{d\,(na^3)}{dt} = -\left(n^2 - n_{\rm eq}^2\right) a^3\langle\sigma v\rangle. \tag{3.4.1}$$

For very high temperatures with $k_{\mathcal{B}}T \gg m_L$ the equilibrium density $n_{\rm eq}$ varies as T^3, and T varies as $1/a$, so Eq. (3.4.1) has a solution $n = n_{\rm eq}$. Eventually, with the decrease in temperature below the L-particle mass, the equilibrium density drops so low that the creation term in Eq. (3.4.1) becomes negligible, and we have

$$\frac{d\,(na^3)}{dt} = -n^2a^3\langle\sigma v\rangle. \tag{3.4.2}$$

The solution of Eq. (3.4.2) is

$$\frac{1}{n(t)a^3(t)} = \text{ constant } + \int \frac{\langle\sigma v\rangle dt}{a^3(t)}$$

or, in other words,

$$n(t)a^3(t) = \frac{n\,(t_1)\,a^3\,(t_1)}{1 + n\,(t_1)\,a^3\,(t_1)\int_{t_1}^{t}\langle\sigma v\rangle a^{-3}\,(t')\,dt'}$$

where t_1 is any convenient time chosen late enough so that the creation term in Eq. (3.4.1) may be neglected for $t' > t_1$.

The important point here is that the integral in the denominator converges for $t \to \infty$. The denominator $a^3(t)$ increases like $t^{3/2}$ when the energy density is dominated by relativistic particles, and even faster later, when it is dominated by non-relativistic particles and/or vacuum energy. Also, if annihilation is possible from states of zero orbital angular momentum then σv approaches a constant for low energies, so its thermal average $\langle\sigma v\rangle$ approaches a constant for low temperatures, and hence for late times. The contribution of states of higher orbital angular momentum decreases with decreasing temperature, so if s wave annihilation is forbidden by selection rules the integral converges even faster. Because the integral converges, the particle number in a co-moving volume a^3 approaches a finite limit:

[474] B. W. Lee and S. Weinberg, Phys. Rev. Lett. 39, 165 (1977); D. D. Dicus, E. W. Kolb, and V. L. Teplitz, Phys. Rev. Lett. 39, 168 (1977); E. W. Kolb and K. A. Olive, Phys. Rev. D 33, 1202 (1986).

$$n(t)a^3(t) \to \frac{n\left(t_1\right)a^3\left(t_1\right)}{1+n\left(t_1\right)a^3\left(t_1\right)\int_{t_1}^{\infty}\langle\sigma v\rangle a^{-3}\left(t'\right)dt'} \tag{3.4.3}$$

Let us assume that the annihilation of L particles and antiparticles took place during a time when the density of the universe was dominated by relativistic particles, so that $a \propto 1/T$, and the time is given by Eq. (3.1.14):

$$dt = -2\sqrt{\frac{3}{16\pi G\mathcal{N}a_{\mathcal{B}}}}\frac{dT}{T^3}$$

or, using the formula $a_{\mathcal{B}} = \pi^2 k_{\mathcal{B}}^4/15$ with $\hbar = c = 1$,

$$dt = -\sqrt{\frac{45}{4\pi^3 G\mathcal{N}}}m_L^{-2}\frac{dx}{x^3} \tag{3.4.4}$$

where $x \equiv k_{\mathcal{B}}T/m_L$. Eq. (3.4.1) therefore takes the form

$$\frac{du(x)}{dx} = B\left[u^2(x) - u_{\rm eq}^2(x)\right] \tag{3.4.5}$$

where $u(x)$ is the dimensionless quantity of interest

$$u \equiv n/\left(k_{\mathcal{B}}T\right)^3 \tag{3.4.6}$$

$u_{\rm eq}\ (k_{\mathcal{B}}T/m_L)$ is its equilibrium value, and B is the dimensionless parameter

$$B = \sqrt{\frac{45}{4\pi^3 G\mathcal{N}}}m_L\langle\sigma v\rangle \tag{3.4.7}$$

We see that the left-over value of $n/\left(k_{\mathcal{B}}T\right)^3$ depends only on B, and on the spin of the L particles, which we need to know to give a formula for $u_{\rm eq}\left(x\right)$:

$$u_{\rm eq}(x) = \frac{2s_L+1}{(2\pi)^3}\int_0^{\infty}\frac{4\pi y^2 dy}{\exp\sqrt{x^{-2}+y^2}\pm 1} \tag{3.4.8}$$

where $y \equiv p/k_{\mathcal{B}}T$, and as usual the sign is $+1$ for fermions and -1 for bosons.

For instance, for $k_{\mathcal{B}}T \ll m_L$ the heavy particles are non-relativistic, and the mean value of the low-energy annihilation cross section times velocity is a constant,

$$\langle\sigma v\rangle = G_{\rm wk}^2 m_L^2\mathcal{F}/2\pi \tag{3.4.9}$$

where $G_{\rm wk} = 1.1664\times 10^{-5}{\rm GeV}^{-2}$ is the weak coupling constant, and $\mathcal{F}$ is a fudge factor to take account of the number of annihilation channels and the details of the interaction responsible for the annihilation. This gives

$$B = 1.59\times 10^8\left(m_L[{\rm GeV}]\right)^3\mathcal{F}\mathcal{N}^{-1/2} \tag{3.4.10}$$

The solution of Eq. (3.4.5) for various values of B and $s_L = 1/2$ is shown in Figure 3.2. We see that $u(x)$ drops steeply from a constant B-independent value for $x > 1$ to a constant B-dependent value for $x < 0.01$.

For m_L in the GeV range B is quite large, and for such values of B (and $s_L = 1/2$) the asymptotic value of u is well approximated by [4]

$$u(0) \simeq 6.1B^{-0.95} \tag{3.4.11}$$

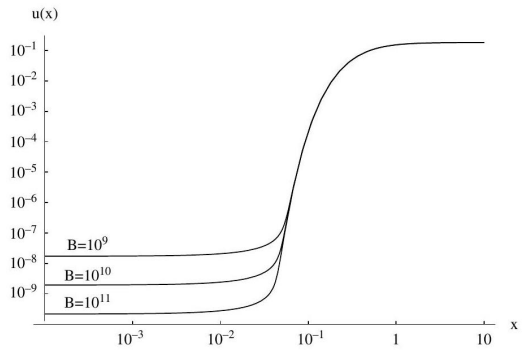


Figure 3.2: The function $u(x) \equiv n /\left(k_{\mathcal{B}} T\right)^{3}$ vs $x \equiv k_{\mathcal{B}}T/m_L$, for a dark matter particle of spin 1/2 and mass m_L, with various values of the parameter B defined by Eq. (3.4.7).

(A value $u(0) \propto B^{-1}$ is what we would expect if the second term in the denominator in Eq. (3.4.3) were much greater than unity.) The present mass density of these heavy particles and antiparticles is then (recalling that electron-positron annihilation increases the photon temperature by a factor $(11/4)^{1/3}$)

$$\begin{aligned}\rho_L &= 2m_L u(0)(4/11)\left(k_{\mathcal{B}}T_{\gamma 0}\right)^3 \\ &= \left(2.15\times 10^{-28}\ \mathrm{g/cm^3}\right)\left(m_L(\mathrm{GeV})\right)^{-1.85}\left(\mathcal{F}/\sqrt{\mathcal{N}}\right)^{-0.95}\end{aligned} \tag{3.4.12}$$

or in other words

$$\Omega_L \equiv 8\pi G\rho_L/3H_0^2 = 11.5h^{-2}\left(m_L(\mathrm{GeV})\right)^{-1.85}\left(\mathcal{F}/\sqrt{\mathcal{N}}\right)^{-0.95} \tag{3.4.13}$$

Note that Ω_L is a decreasing function of m_L, because heavy L particles annihilate more effectively than light ones. If we assume that the left-over heavy particles make up most of the cosmic mass density, then $\Omega_M \simeq \Omega_L$, and so

$$m_L(\mathcal{F}/\sqrt{\mathcal{N}})^{0.51} \simeq 3.7\left(\Omega_M h^2\right)^{-.54}\,\mathrm{GeV} \tag{3.4.14}$$

which is 10 GeV for $\Omega_M h^2 = 0.15$. Otherwise this provides a lower bound on the mass of these particles.

Incidentally, similar arguments apply to the annihilation of nucleons and antinucleons, if the universe has zero net baryon number. The difference here of course is that instead of $\langle\sigma v\rangle$ being of order $G_F^2 m_L^2$, it is roughly of order m_π^{-2}. We can thus estimate the density of left-over baryon-antibaryon pairs by taking $m_L = m_N, \mathcal{N} \approx 10$, and replacing $\mathcal{F}$ in Eq. (3.4.13) with $2\pi m_\pi^{-2}/G_F^2 m_N^2 = 2.7\times 10^{12}$, so that Eq. (3.4.13) would give a baryonantibaryon density parameter $\Omega_{B+\bar{B}}h^2 \approx 6\times 10^{-11}$. This is much less than the present observed density parameter of baryons, ruling out the possibility that the baryons around us are just those that missed being annihilated in an initially baryon-antibaryon symmetric universe, and then somehow became segregated from the antibaryons.

Returning to the cold dark matter, originally it was thought that the L particles might be heavy neutrinos. They could not be any of the three known neutrino types, which as discussed at the end of Section 3.1 have masses at most of the order of 1 eV , but there could be a fourth generation of very heavy leptons that have negligible mixing with the known leptons. The mass of a new heavy neutrino would have to be greater than $m_Z/2 = 45$ GeV , to block the decay of the Z^0 into a heavy neutrino and antineutrino, which if it occurred would destroy the present excellent agreement between theory and experiment for the total decay rate of the Z^0 particle.

If the
L particles were heavy neutrinos, then the fudge factor $\mathcal{F}$ in Eq. (3.4.9) would be of the order of the number of types of particles into which the L particles might annihilate. If we assume very roughly that $\mathcal{F} \approx \mathcal{N} \approx 100$, then Eq. (3.4.13) gives $\Omega_L h^2 \leq 10^{-3}$. So it does not seem that the cold dark matter could consist of a new heavy neutrino.

The most plausible candidate for the L particle is one of the new particles required by supersymmetry. [5] In many supersymmetric theories there is a multiplicatively conserved quantum number R, which takes the value $+1$ for all the known particles of the Standard Model, and -1 for their supersymmetric partners. (Multiplicative conservation means that the product of the Rs for all the particles in the final state of any reaction is the same as for the initial state.) Among other things, this conservation law tells us that the lightest particle with $R = -1$ (which is often called the LSP, for "lightest supersymmetric particle") is stable, although two of these particles could annihilate into ordinary particles with $R = +1$. To judge which is the lightest particle with $R = -1$, it is necessary to distinguish between two possible pictures of supersymmetry breaking. [6] In both pictures supersymmetry is spontaneously broken by non-perturbative effects in some hidden sector of particles with a large typical mass M_S, which interact through some strong force that is not felt by the known particles of the Standard Model or their superpartners. The supersymmetry breaking in the hidden sector also gives the gravitino (the superpartner of the graviton with spin 3/2) a mass $m_g \approx \sqrt{G} M_S^2$.

In one picture of supersymmetry breaking, the breakdown of supersymmetry is communicated to the particles of the Standard Model and their superpartners by the electroweak and ordinary strong forces of the Standard Model. In this case M_S would have to be of the order of 100 GeV to 100 TeV , and the gravitino mass would be at most of order 1 eV . The lightest particle with $R = -1$ would be the gravitino, which would be too light to furnish the cold dark matter. Even if gravitinos were once in equilibrium with other particles, and did not annihilate or decay, by the same arguments as in Section 3.1 their number density now would be less than the number density of photons by a factor of order $2/\mathcal{N}$, where $\mathcal{N}$ is the effective number of relativistic particle states at the time that gravitinos went out of thermal equilibrium. Photons with typical energy $k_{\mathcal{B}} T_{\gamma 0}$ have an energy density parameter given by Eq. (2.1.8) as $\Omega_\gamma h^2 = 2.47 \times 10^{-5}$, so in order for gravitinos to furnish cold dark matter with $\Omega_M h^2 \simeq 0.15$, the mass of the gravitino would have to be of order

$0.15 \mathcal{N} k_{\mathcal{B}} T_\gamma 0 / 2 \Omega_\gamma h^2$, which for $\mathcal{N} \approx 100$ is roughly 70eV.[7] Thus in this picture of supersymmetry breaking, gravitinos are not likely to be the cold dark matter, though this is not absolutely ruled out. The next-to-lightest particles with $R = -1$ would all presumably decay over billions of years into gravitinos and ordinary particles, so they could not furnish the cold dark matter either. (The rate of decay into relativistic gravitinos with helicity $\pm 3/2$ would be suppressed by a factor G, but particles with $R = -1$ could decay into gravitinos with helicity $\pm 1/2$ with the decay rate suppressed only by a factor M_S^{-2} and by some powers of the gauge couplings of the Standard Model.)

In the other picture, supersymmetry breaking is mediated by the gravitational field and its superpartners, and because these interactions are very weak M_S must be correspondingly large, of order 10^{11} to 10^{13}GeV. The gravitino mass would be of the same order of magnitude as the masses of the superpartners of the known particles of the Standard Model, so it might or might not be the lightest particle with $R = -1$.

If the lightest particle with $R = -1$ is not a gravitino, then it could be either a sneutrino, the spin 0 superpartner of the neutrino, or a neutralino, the spin 1/2 superpartner of some mixture of the neutral gauge and scalar bosons of the Standard Model. [8] Whichever of these is the

[475] This and other possibilities are discussed in a comprehensive review by G. Bertone, D. Hooper, and J. Silk, Phys. Rep. 405, 279 (2005) [hep-ph/0404175].
[6] See QTF, Vol. III, Sec. 28.3.

lightest, the decay process will involve the exchange of a particle with $R = -1$, having a mass $\tilde{M}$ that is expected to be of the order of a TeV or so. Hence its annihilation amplitude (aside from factors of 2 and π) will be of order $g^2/\tilde{M}^2$, where g is a typical electroweak coupling. This is smaller than the weak coupling constant $G_{\rm wk}$ by a factor of order $m_W^2/\tilde{M}^2$, so the fudge factor $\mathcal{F}$ in the annihilation rate constant (3.4.9) will be of order $m_W^4\mathcal{N}_A/\tilde{M}^4 \approx 10^{-4}\mathcal{N}_A$, where $\mathcal{N}_A$ is the number of annihilation channels. Taking $\mathcal{N}_A/\sqrt{\mathcal{N}} \approx 1$, Eq. (3.4.14) tells us that to furnish cold dark matter with $\Omega_M h^2 \simeq 0.15$, the lightest particle with $R = -1$ (if not the gravitino) would have to have a mass of order 1 TeV . This is similar to estimates of the masses of the superpartners of the particles of the Standard Model in typical supersymmetric models, a circumstance that greatly encourages the hope that the particles of dark matter will be found to be created in experiments at high energy accelerators.

Cosmological considerations rule out the possibility that the lightest particle with $R = -1$ in this picture of supersymmetry breaking is the gravitino. Unlike the superpartners of the particles of the Standard Model, the

gravitino has a two-body annihilation amplitude proportional to two factors [9] of the gravitational coupling constant $\sqrt{G}$, so the annihilation rate is proportional to G^2, and is therefore much too small for annihilation to play any significant role in reducing the number density of gravitinos. If, because of particle annihilations after gravitino decoupling, the number density of gravitinos were now, say, 1% of the number density of photons, then, as we have seen, in order for their mass density now not to exceed cosmological bounds their mass would have to be less than roughly 100 eV . This is quite inconsistent with the gravitino masses expected in theories of gravity mediated supersymmetry breaking. But if the gravitinos are not the lightest particles with $R = -1$ then they can decay, reducing their present mass density to acceptable values even if their masses are quite high. [10] The coupling of the gravitino to other fields is proportional to $\sqrt{G}$, so on dimensional grounds the decay rate Γ_g of a gravitino at rest is roughly of the order of Gm_g^3. This is to be compared with the rate of expansion of the universe, which at temperature T is of order $\sqrt{G\left(k_{\mathcal{B}}T\right)^4}$. (We are here ignoring factors of order 10-100, including those involving non-gravitational coupling constants and the number of particle species.) When the cosmic temperature drops to the value $k_{\mathcal{B}}T \approx m_g$ at which gravitinos become non-relativistic, the ratio of their decay rate to the expansion rate is of order $\sqrt{G}m_g = m_g/m_{\rm Planck} \ll 1$, so gravitino decay becomes significant only after this time, when the gravitinos are highly non-relativistic. As we have seen, their number density will be of order $\left(k_{\mathcal{B}}T\right)^3$, so their energy density will then be of order $m_g\left(k_{\mathcal{B}}T\right)^3$, which is greater than the energy density of order $\left(k_{\mathcal{B}}T\right)^4$ of the photons and other relativistic particles in thermal equilibrium at temperature T, and therefore makes the dominant contribution to the cosmic gravitational field that governs the rate of expansion of the universe. The expansion rate under these conditions is therefore of order $\sqrt{Gm_g\left(k_{\mathcal{B}}T\right)^3}$, and gravitino decay becomes significant when this equals the gravitino decay rate of order Gm_g^3, and therefore at a temperature

$$k_{\mathcal{B}}T_g \approx G^{1/3}m_g^{5/3}$$

After they decay, their energy must go into the energy of photons and other relativistic particles, so the temperature T_g' after decay is related to the temperature T_g calculated above by the energy conservation condition

[477] H. Pagels and J. R. Primack, Phys. Rev. Lett. 48, 223 (1982) first remarked that in order for the gravitino not to give $\Omega_M \gg 1$, its mass would have to be less than about 1 keV .

[8] S. Weinberg, Phys. Rev. Lett. 50, 387 (1983).

[479] This does not include a factor of $\sqrt{G}$ in the gravitino mass, because this factor is multiplied by the scale of supersymmetry breaking, which for gravitationally mediated supersymmetry breaking is very large, giving a gravitino mass that is comparable to the mass of other supersymmetric particles.

[10] S. Weinberg, Phys. Rev. Lett. 48, 223 (1982).

$m_g T_g^3 \approx T_g'^4$, and hence

$$k_{\mathcal{B}} T_g' \approx G^{1/4} m_g^{3/2}$$

In particular, since $k_{\mathcal{B}} T_g \ll m_g$, we have $T_g' \gg T_g$. If T_g were less than the temperature $T_n \simeq 0.1$MeV at which cosmological nucleosynthesis can occur, then gravitinos would still be abundant before nucleosynthesis, giving a higher energy density and hence a faster expansion, so that there would be less time for free neutrons to decay before being incorporated into complex nuclei, and hence more helium would be produced when nucleosynthesis occurs. Also, the ratio of the photon and baryon densities would have been subsequently increased by gravitino decay, so this ratio at the time of nucleosynthesis would have been considerably less than is usually estimated from the present cosmic microwave background temperature, and so nuclear reactions would have incorporated neutrons more completely into helium, and less deuterium would be left today. The present agreement between theory and observation for the cosmic helium and deuterium abundances would thus be destroyed. This problem is avoided if $T_g > 0.1$MeV, but it can also be avoided under the much weaker condition that $T_g' > 0.4$MeV, because then after the gravitinos decay the temperature would have been high enough to break up the excess helium and give cosmological nucleosynthesis a fresh start as the universe recools. This condition requires that $m_g > 10$TeV. This limit on m_g corresponds to a supersymmetry breaking scale $M_S > 10^{11}$GeV for $m_g \approx \sqrt{G} M_S^2$.

It may be possible to detect cosmic WIMPs through observation of the recoil of atomic nuclei from which they scatter elastically, [11] as for instance the scattering of heavy neutrinos through the neutral current weak interaction. [12] This is being pursued by a number of collaborations: DAMA, [13] CRESST, [14] EDELWEISS, [15] UK Dark Matter, [16] CDMS, [17] and WARP. [18] Assuming that WIMPs are more or less at rest in the halo of our galaxy, the motion of the solar system through the halo produces a WIMP "wind" with

a speed v_w of about 220 km/s. A nucleus of mass Am_N when struck with a WIMP of mass $m_w \gg Am_N$ traveling at this velocity will recoil with a typical velocity of order v_w, and hence a kinetic energy of order $Am_N v_w^2/2 \approx 100A$ eV . These recoils can be detected by observing the ionization of atoms struck by the recoiling nucleus, or by detecting light emitted by these atoms, or by detecting vibrations in the crystal lattice of the detector. The mass density ρ_h in the halo of our galaxy near earth is estimated from observations of stellar motions to be about 0.3GeV/cm^3, giving a number density for WIMP mass m_w of ρ_h/m_w, from which the elastic scattering rate can be calculated for any assumed values of the WIMP mass and the scattering cross section. Failing to observe nuclear recoil events then excludes some region of the m_w-cross-section plane.

The greatest problem in these experiments is distinguishing true WIMP events from background, caused by natural radioactivity and cosmic rays. The best hope for distinguishing events from background is to exploit the motion of the earth around the sun. [19] This orbital motion adds about 15 km/s to the speed of the WIMP wind in summer, and subtracts an

[4][7][11] For reviews of current experiments, see Y. Ramachers, Nucl. Phys. B. Proc. Suppl. [astro-ph/ 0211500]; G. Chardin, in Cryogenic Particle Detection, ed. C. Ens (Springer, Heidelberg, 2005) [astroph/0411503]; R. J. Gaitskell, Ann. Rev. Nucl. Part. Sci. 54, 315 (2004); J. Ellis, K. Olive, Y. Santoso, and V. C. Spanos, Phys. Rev. D 71, 095007 (2005) [hep-ph/0502001]; K. Freese, Nucl. Instrum. Meth. A559, 337 (2006) [astro-ph/0508279]; L. Baudis, Int. J. Mod. Phys. A21, 1925 (2006) [astro-ph/0511805].

[12] A. Drukier and L. Stodolsky, Phys. Rev. D 30, 2295 (1985); M. Goodman and E. Witten, Phys. Rev. D 31, 3059 (1985).

[13] R. Bernabei et al., Phys. Lett. B480, 23 (2000).

[14] G. Angloher et al., Astropart. Phys. 23, 325 (2005) [astro-ph/0408006].

[15] V. Sanglard et al., Phys. Rev. D 71, 122002 (2005) [astro-ph/0503265].

[16] G. J. Alner et al., Astropart. Phys. 23, 444 (2005).

[17] D. S. Akerib et al., Phys. Rev. Lett. 93, 211301 (2004) [astro-ph/0405033], and astro-ph/0507190.

[18] P. Benetti et al., astro-ph/0701286.

equal amount in winter (though this depends on the halo model), so one may expect a 7% seasonal modulation of true WIMP events. The DAMA collaboration reported just such a modulation, but almost all of the region in the m_w-cross-section plane that would account for this observation was subsequently apparently excluded (using a different detection scheme) by the CDMS collaboration. So far, experiments set an upper bound on the effective cross section of about 10^{-42} to 10^{-43} cm^2 for $m_w > 50$GeV, and much larger for $m_w < 50$GeV, the precise bound depending on assumptions about the distribution of WIMPs in the galactic halo. [20] (For comparison, the effective cross section for the low energy scattering of the neutralinos of supersymmetric theories on nucleons is expected to be less than about 10^{-41} cm^2.)

There is also a possibility of an indirect detection of WIMPs, through observation of gamma rays or other particles produced when pairs of WIMPs annihilate in regions of high WIMP concentration. [21] Gamma rays do not penetrate the earth's atmosphere, so they have to be detected either through the observation of showers of charged particles produced in the atmosphere, as for instance by the Cerenkov radiation associated with these particles, or by gamma ray telescopes carried by balloons or by satellites in orbit above the earth's atmosphere. If WIMP-WIMP annihilation produces

just a pair of gamma rays, then each will carry a unique energy equal to the WIMP mass, giving a very clear WIMP signal. So far, there has been no sign of such a monochromatic gamma ray. Of course the annihilation may produce other particles along with gamma rays, or particles whose decay then produces gamma rays, but in either case this would yield a continuum of gamma ray energies, which would be much harder to identify as coming from WIMP annihilation. The Cangaroo-II [22] and HESS [23] atmospheric Cerenkov detector collaborations have both reported continuum sources of gamma rays coming from near the center of our galaxy, but with very different spectra, which can be interpreted as coming from annihilation of WIMPs with a mass of about 1 TeV or 19 TeV , respectively.

For decades there has been evidence of monochromatic gamma rays coming from the galactic center, but at the energy of 511 keV expected from electron-positron annihilation, rather than WIMP-WIMP annihilation. In 1970 a balloon-borne pair of gamma ray detectors found evidence of a gamma ray line around 500 keV coming more-or-less from the direction of the galactic center. [24] The evidence for this has since become much stronger through observations made by the INTEGRAL (International GammaRay Astrophysics Laboratory) satellite, [25] which found a gamma ray line coming from the galactic center with an energy within about 3 keV of 511 keV . It is possible that this is due to the decay of a relatively light WIMP into electron-positron pairs, with the positrons then losing energy by ionization, after which they annihilate with ambient electrons. [26]

There is also a report of an excess of positrons in cosmic rays found by the HEAT balloon experiment, [27] which might or might not come from WIMP annihilation. Very recently it has been suggested that an excess of microwave emission from the direction of the center of our galaxy observed by the WMAP satellite may come from synchrotron emission by relativistic electrons and positrons produced in WIMP annihilation. [28]

It is too early to reach any definite conclusions from any of these experiments about WIMPs as candidates for the particles of dark matter.

[4719] A. K. Drukier, K. Freese, and D. N. Spergel, Phys. Rev. D 33, 3495 (1986); K. Freese, J. A. Frieman, and A. Gould, Phys. Rev. D 37, 3388 (1988).

[20] A. Bottino, F. Donato, N. Fornengo, and S. Scopel, Phys. Rev. D 72, 083521 (2005) [hep-ph/ 0508270].

[21] For a review, see P. Gondolo, hep-ph/0501134.

[4722] K. Tsuchiya et al., Astrophys. J. 606, L115 (2004) [astro-ph/0403592].

[23] D. Horns, Phys. Lett. B607, 225 (2005). But see F. Aharonian,Phys. Rev. Lett. 97, 221102 (2006)

[24] W. N. Johnson, III, F. R. Harnden, Jr., and R. C. Haymes, Astrophys. J. 172, L1 (1972.)

[25] P. Jean et al., Astron. Astrophys. 407, 55 (2003) [astro-ph/0309484]; J. Knödlseder et al., Astron. Astrophys. 411, 457 (2003) [astro-ph/0309442]; E. Churazov et al., Mon. Not. Roy. Astron. Soc. 357, 1377 (2005) [astro-

B. Axions and axinos

Axions [29] are light neutral spinless particles that are made necessary [30] by the spontaneous breakdown of a symmetry that first appeared in a model [31] that was proposed to explain why non-perturbative effects of the strong interactions do not violate CP invariance. For our present purposes, all we need to know is that the dominant part of the effective action (with $\hbar = c = 1$) that describes the axion field φ takes the form (B.44) with potential $V(\varphi) = m_a^2\varphi^2/2$:

$$I[\varphi] = \int d^4x \sqrt{-\det g}\left(-\frac{1}{2}g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi - \frac{1}{2}m_a^2\varphi^2\right) \tag{3.4.15}$$

The axion mass m_a is a complicated function of temperature, but for temperatures well below 10^{12} K it takes a well-known constant value, related to the energy scale M at which the Peccei-Quinn symmetry is broken by

$$m_{a0} = \frac{F_\pi m_\pi}{M}\frac{\sqrt{m_d m_u}}{m_d + m_u} \simeq \frac{13\text{MeV}}{M[\text{GeV}]} \tag{3.4.16}$$

(Here $F_\pi = 184$MeV is the pion decay amplitude, and m_d and m_u are the down and up quark masses appearing in the Lagrangian of quantum chromodynamics, for which $m_d/m_u \simeq 1.85$.) In the original Peccei-Quinn model, [31] M was of the order of the electroweak symmetry breaking scale, $M \approx 100$GeV, but it was soon realized that this is experimentally ruled out. Axion fields interact with ordinary matter through factors $\partial_\mu\varphi/M$, so with $M \geq 100$GeV axions interact so weakly that they emerge without attenuation from reactor cores or stellar interiors. The production rate of axions is proportional to $1/M^2$, so the failure to observe effects of axion emission from stars or nuclear reactors sets a lower bound on M, and hence an upper bound on m_a. In particular, limits on the rate of cooling of red giant stars by axion emission give [32] $M > 10^7$GeV, while observations of the supernova SN1987A indicate [33] that $M > 10^{10}$GeV. A generalization of the Peccei-Quinn model was then proposed, [34] in which M is an arbitrary parameter. As we shall see, cosmological considerations provide an upper bound [35] on M, which together with the above lower bounds leaves a narrow window of allowed values.

The cosmological axion field is supposed to be spatially homogeneous. [36] In a Robertson-Walker metric, the energy density and pressure are given by Eq. (B.66) and (B.67) as

$$\rho_a = \frac{1}{2}\dot{\varphi}^2 + \frac{1}{2}m_a^2\varphi^2, \quad p_a = \frac{1}{2}\dot{\varphi}^2 - \frac{1}{2}m_a^2\varphi^2, \tag{3.4.17}$$

so the equation (1.1.32) of energy conservation (or the Euler-Lagrange equation derived directly from the action (3.4.15)) gives the field equation

ph/0411351]; J. Knödlseder et al., Astron. Astrophys. 441, 513 (2005) [astroph/0506026]; G. Weidenspointer et al., Astron. Astrophys. 450, 1013 (2006) [astro-ph/0601673].

[26] C. Boehm, D. Hooper, J. Silk, M. Casse, and J. Paul, Phys. Rev. Lett. 92, 1301 (2004) [astro-ph/ 0309686].

[27] S. W. Barwick et al., Phys. Rev. Lett. 75, 390 (1995); Astrophys. J. 482, L191 (1997).

[28] D. Hooper, D. P. Finkbeiner, and G. Dobler, 0705.3655.

[4729] See QTF, Vol II, Sec. 23.6. For a review, see P. Sikivie, in Axions - Lecture Notes on Physics, ed. M. Kuster (Springer-Verlag, to be published) [astro-ph/0610440].

[30] S. Weinberg, Phys. Rev. Lett. 40, 223 (1978); F. Wilczek, Phys. Rev. Lett. 40, 279 (1978).

[31] R. D. Peccei and H. Quinn, Phys. Rev. Lett. 38, 1440 (1977); Phys. Rev. D 16, 1791 (1977).

[32] D. A. Dicus, E. W. Kolb, V. I. Teplitz, and R. V. Wagoner,Phys. Rev. D 18, 1829 (1978); Phys. Rev. D 22, 839 (1980).

[33] For reviews, see M. S. Turner, Phys. Rep. 197, 67 (1990); G.G. Raffelt, Phys. Rep. 198. 1 (1990); P. Sikivie, ref. 26.

[34] J. E. Kim, Phys. Rev. Lett. 43, 103 (1979); M. Dine, W. Fischler, and M. Srednicki, Phys. Lett. 104B, 199 (1981); M. B. Wise, H. Georgi, and S. L. Glashow, Phys. Rev. Lett. 47, 402 (1981).

$$\ddot{\varphi} + 3H(t)\dot{\varphi} + m_a^2(t)\varphi = 0 \tag{3.4.18}$$

At early times $H(t) \gg m_a(t)$, so we can ignore $m_a(t)$, and Eq. (3.4.17) has solutions $\varphi =$ constant, and $\varphi \propto 1/a^3$. Rejecting the singular solution, we see that at early times $\varphi(t)$ is frozen at a value φ_0, which (absent fine tuning) would be expected to be of order M. Later, when $H(t)$ dropped below $m_a(t), \varphi(t)$ began a rapid oscillation, so that in this case Eq. (3.4.18) can be solved using the WKB approximation, which gives

$$\varphi(t) \to \varphi_1 \left(\frac{a\,(t_1)}{a(t)}\right)^{3/2} \cos\left(\int_0^t m_a(t)dt + \alpha\right) \tag{3.4.19}$$

where t_1 is the time at which $H\,(t_1) = m_{a0}, \varphi_1$ is a constant of order $\varphi_0 \approx M$, and α is a phase that cannot be determined without a more detailed study, but which fortunately we do not need to know. (For instance, if $m_a(t)$ has the constant value m_{a0} and the universe is radiation-dominated in the era of interest, then $\varphi_1 = 2^{3/2}\pi^{-1/2}\Gamma(5/4)\varphi_0 = 1.446\varphi_0$, and $\alpha = -3\pi/8$.) Since $a(t)$ for $m_a \gg H$ was varying much more slowly than the phase of the cosine, the energy density at late times is given by Eq. (3.4.17) as

$$\rho_a(t) \to \frac{1}{2}m_a^2\varphi_1^2 \left(\frac{a\,(t_1)}{a(t)}\right)^3 \tag{3.4.20}$$

In order to project this forward to the present, we note that if the universe was radiation-dominated at time t_1, with $\mathcal{N}_1$ the effective number of types of particles with masses much less than $k_{\mathcal{B}}T\,(t_1)$ (counting each spin state

of particles and antiparticles separately, and including an extra factor 7/8 for fermions), then the photon temperature at present is

$$T_{\gamma 0} = \mathcal{N}_1^{1/3} T\,(t_1)\, a\,(t_1)\,/a\,(t_0)$$

the factor $\mathcal{N}_1^{1/3}$ being inserted to take account of the heating of photons by the annihilation of particles and antiparticles between times t_1 and t_0. The temperature $T\,(t_1)$ may be determined by noting that the expansion rate at time t_1 is

$$m_a \equiv H\,(t_1) = \sqrt{\frac{8\pi G\mathcal{N}_1 a_{\mathcal{B}} T^4/2}{3}} = \sqrt{\frac{4\pi^3 G\mathcal{N}_1}{45}}\,(k_{\mathcal{B}}T_1)^2\,.$$

Using these results in Eq. (3.4.20) then gives the present axion energy density

$$\rho_a\,(t_0) \simeq \frac{1}{2}m_a^{1/2}\mathcal{N}_1^{-1/4}\varphi_0^2 \left(\frac{4\pi^3 G}{45}\right)^{3/4} (k_{\mathcal{B}}T_{\gamma_0})^3 \tag{3.4.21}$$

We expect φ_0 to be of the same order of magnitude as the symmetry breaking scale M, so using Eq. (3.4.16) and ignoring all factors of order unity,

$$\rho_a \approx \frac{F_\pi^2 m_\pi^2 G^{3/4}\,(k_{\mathcal{B}}T_{\gamma 0})^3}{m_a^{3/2}} \tag{3.4.22}$$

Equivalently, the axion density provides a fraction Ω_a of the critical density given by

[4735] J. Preskill, M. B. Wise, and F. Wilczek, Phys. Lett. 120B, 127 (1983); L. F. Abbott and P. Sikivie, Phys. Lett. 120B, 127 (1983); M. Dine and W. Fischler, Phys. Lett. 120B, 137 (1983). For reviews, see J. E. Kim, Phys. Rep. 150, 1 (1987); M. S. Turner, Phys. Rep. 197, 68 (1990).

[36] Components with non-zero wave number are doubtless present, but their energy density decays more rapidly than the energy density of the spatially homogeneous coherent field $\varphi(t)$.

$$\Omega_a h^2 \approx \left(m_a/10^{-5}\text{eV}\right)^{-3/2} \tag{3.4.23}$$

Because the axion field is spatially homogeneous, for $m_a \gg H$ its energy takes the form of massive particles that are essentially at rest. If axions furnish the whole of the cold dark matter, then $m_a \approx 10^{-5}$eV, corresponding to $M \approx 10^{12}$GeV. Otherwise, these numbers provide a lower bound on m_a and an upper bound on M.

Axions are much too weakly interacting to be detected in the sort of nuclear recoils looked for in searches for WIMPs. One possibility is to observe the conversion of cosmic axions into photons in intense magnetic fields. [37] This approach has already been used to put a limit on the parameters of axions that would be produced by the sun. [38] The axion field $\varphi(x)$ would be expected to have an interaction with the electromagnetic field of

the form $g_{a\gamma}\varphi\mathbf{E}\cdot\mathbf{B}$, with $g_{a\gamma}$ of order $\alpha/2\pi M$, so that photons in the sun could convert into axions in the presence of the strong electric fields around atomic nuclei, and these solar axions would then convert back into photons in intense laboratory magnetic fields. No such photons were seen, indicating that for axions with mass less than 0.02eV, $|g_{a\gamma}| < 1.16 \times 10^{-10}\text{GeV}^{-1}$, a much more restrictive limit than provided by earlier experiments of this sort. [39] In a different sort of search, [40] the Axion Dark Matter Experiment, a microwave cavity was used to search for axions in our galactic halo, and put upper limits on the axion density in the narrow mass range (1.98 to 2.17) $\times 10^{-6}$eV. None of these experiments are in conflict with axion models of dark matter, but a plausible improvement in the sensitivity of this sort of experiment may rule out these models, or perhaps find axions.

In supersymmetric theories, the axion would be partnered with a spin one-half particle, the axino, which would probably be the lightest particle with $R = -1$, and hence stable. Axinos could be produced non-thermally, [41] through the decay of other particles with $R = -1$, or thermally. [42] It appears that the axino provides another plausible candidate for the particle of cold dark matter.

10.4 Inflation

We can have some confidence in the story of the evolution of the universe from the time of electron-positron annihilation to the present, as told in the previous three chapters. About earlier times, so far we can only speculate. In the past quarter century these speculations have centered on the idea that before the period of radiation domination, during which the RobertsonWalker scale factor $a(t)$ was growing as $\sqrt{t}$, there was an earlier period of inflation, when the energy density of the universe was dominated by a slowly varying vacuum energy, and $a(t)$ grew more-or-less exponentially. The possibility of an early exponential expansion had been noticed by several authors, [1] but at first it attracted little attention. It was Alan Guth [2] who incited interest in the possibility of inflation by noting what it was good for.

Guth noticed that, in a model of grand unification he was considering (with Henry Tye), scalar fields could get caught in a local minimum of the potential, which in his work corresponded to a state with an unbroken grand unified symmetry. The energy of empty space would then have remained constant for a while as the universe expanded, which would produce

[4737] P. Sikivie, Phys. Rev. Lett. 51, 1415 (1983); Phys. Rev. D 32, 2988 (1985).

[38] K. Zioutas et al. (CERN Axion Solar Telescope collaboration), Phys. Rev. Lett. 94, 121301 (2005) [hep-ex/0411033].

[4739] This and earlier experiments are reviewed by G. G. Raffelt, contribution to XI International Workshop on Neutrino Telescopes, hep-ph/0504152.

[40] L. D. Duffy et al., Phys. Rev. D 74, 012006 (2006) [astro-ph/0603108], and earlier references cited therein.

[41] L. Covi, J. E. Kim, and L. Roszkowski, Phys. Rev. Lett. 82, 4180 (1999).

[42] L. Covi, H. B. Kim, J. E. Kim, and L. Roszkowski, J. High Energy Phys. 0105, 033 (2001) [hepph/0101009]; A. Brandenburg and F. D. Steffen, J. Cosm. & Astropart. Phys. 0408. 008 (2004) [hep-ph/0405158].

a constant rate of expansion, meaning that $a(t)$ would have grown exponentially. Eventually this inflation would be stopped by quantum-mechanical barrier penetration, after which the scalar field would start rolling down the potential toward a global minimum, corresponding to the present universe. In itself this would have been a result of no great immediate importance. But then it occurred to Guth that the existence of an era of inflation would solve one of the outstanding problems of cosmology, mentioned here in Section 1.5. It is known as the "flatness problem:" Why was the curvature of space was so small in the early universe? Guth soon also discovered that inflation would solve other cosmological puzzles, some of which he had not even realized were puzzles. These problems along with the flatness problem will be discussed in Section 4.1.

As Guth and others soon realized, his version of inflation had a fatal problem, to be described in Section 4.2. Guth's "old inflation" was soon replaced with a "new inflation" model, due to Andrei Linde [3] and Andreas Albrecht and Paul Steinhardt. [4] The essential element introduced by

theories of new inflation was a nearly exponential expansion during the slow roll of one or more scalar fields down a potential hill, which is the main subject of Section 4.2. This provided a basis for "chaotic inflation" and "eternal inflation" and other variants, some of which are briefly described in Section 4.3.

So far, the details of inflation are unknown, and the whole idea of inflation remains a speculation, though one that is increasingly plausible. Aside from the classic problems that inflation solved at the beginning, it has had one significant experimental success: a prediction of some of the properties of the fluctuations in the cosmic microwave background and large scale structure. We will come to this in Chapter 10, after we take up the evolution of fluctuations in the early universe in Chapters 5 and 6 and the observation of these fluctuations in Chapters 7 and 8 and Section 9.5.

10.4.1 4.1 Three puzzles

In this section we will outline three classic cosmological problems, and work out the extent of the inflation required to solve each of them. For this purpose we will here simply assume that the universe went through an early period of exponential expansion, without worrying yet about how this came about.

A. Flatness

As we saw in Section 1.6 and 1.8, the observed Type Ia supernova redshiftdistance relation and measurements of the ages of the oldest stars are consistent with a vanishing spatial curvature parameter Ω_K, though a nonvanishing curvature can be accommodated by changing Ω_M. Including data from the cosmic microwave background temperature fluctuations, discussed in Section 7.2, favors $\Omega_K = 0$. Although there is still room for a small non-zero Ω_K, it seems quite safe to conclude from these observations that $|\Omega_K| < 1$. But Ω_K is just the present value of the dimensionless time dependent curvature parameter $-K/a^2H^2 = -K/\dot{a}^2$, with K constant. From the time the temperature dropped to about 10^4 K until near the present, $a(t)$ has been increasing as $t^{2/3}$, so $|K|/\dot{a}^2$ has also been increasing as $t^{2/3} \propto T^{-1}$. Thus, if $|\Omega_K| < 1$, then at 10^4 K the curvature parameter $|K|/\dot{a}^2$ could not have been greater than about 10^{-4}. Earlier,

[471] A. A. Starobinsky, JETP Lett. 30, 682 (1979); Phys. Lett. B 91, 99 (1980); D. Kazanas, Astrophys. J. 241, L59 (1980); K. Sato, Mon. Not. Roy. Astron. Soc. 195, 467 (1981).

[2] A. Guth, Phys. Rev. D 23, 347 (1981). Guth tells the story of this work in The Inflationary Universe: The Quest for a New Theory of Cosmic Origins (Helix Books/Addison Wesley, 1997).

[3] A. D. Linde, Phys. Lett. B 108, 389 (1982); 114, 431 (1982); Phys. Rev. Lett. 48, 335 (1982).

[4] A. Albrecht and P. Steinhardt, Phys. Rev. Lett. 48, 1220 (1982).

$a(t)$ was increasing as $t^{1/2}$, so $|K|/\dot{a}^2$ was increasing as $t \propto T^{-2}$. In order for $|K|/\dot{a}^2$ at 10^4 K to be no greater than about 10^{-4}, it is necessary that $|K|/\dot{a}^2$ was at most about 10^{-16} at the temperature $T \approx 10^{10}$ K of electron-positron annihilation (roughly, the beginning of the period of neutron-proton conversion that results in the observed helium abundance), and even smaller at earlier times. [5] This is not a paradox-there is no reason why the curvature should not have been very small-but it is the sort of thing physicists would like to explain if we can.

What Guth realized was that during inflation $\dot{a}/a$ would have been roughly constant, so $|K|/a^2H^2$ would have been decreasing more or less like a^{-2}. So to understand why space was so flat at the beginning of the present big bang it is not necessary to make any arbitrary assumptions; if the radiation-dominated big bang was preceded by a sufficient period of inflation, it would necessarily have started with negligible curvature.

To put this quantitatively, suppose the universe began with a period of inflation during which $a(t)$ increased by some large factor $e^{\mathcal{N}}$, followed by a period of radiation dominance lasting until the time of radiation-matter equality, followed in turn by a period of matter dominance and then a period dominated by vacuum energy. If $|K|/a^2H^2$ had a value of order unity at the beginning of inflation, then at the time t_I of the end of inflation $|K|/a^2H^2$ would have had a value $|K|/a_I^2H_I^2$ of order $e^{-2\mathcal{N}}$ (where a_I and H_I are the Robertson-Walker scale factor and expansion rate at this time), and today we will have

$$|\Omega_K| = \frac{|K|}{a_0^2H_0^2} = e^{-2\mathcal{N}}\left(\frac{a_IH_I}{a_0H_0}\right)^2, \tag{4.1.1}$$

Thus the flatness problem is avoided if the expansion during inflation has the lower bound

$$e^{\mathcal{N}} > \frac{a_IH_I}{a_0H_0}. \tag{4.1.2}$$

To evaluate this we will make the somewhat risky assumption that not much happens to the cosmic scale factor and expansion rate from the end of inflation to the beginning of the radiation-dominated era, so that

$$a_IH_I \simeq a_1H_1, \tag{4.1.3}$$

the subscript 1 denoting the beginning of the radiation-dominated era. We can express the expansion a_0/a_1 of the universe since the start of the radiation-dominated era in terms of ratios of expansion rates by noting that over the whole of the radiation and matter-dominated era, the expansion

rate was

$$H = \frac{H_{\rm EQ}}{\sqrt{2}}\sqrt{\left(\frac{a_{\rm EQ}}{a}\right)^3 + \left(\frac{a_{\rm EQ}}{a}\right)^4},$$

where $a_{\rm EQ} = a_0\Omega_R/\Omega_M$ and $H_{\rm EQ} = \sqrt{2\Omega_M}H_0\,(a_0/a_{\rm EQ})^{3/2}$ are the scale factor and expansion rate at matter-radiation equality. Setting $a = a_1 \ll a_{\rm EQ}$ gives

$$H_1 = \frac{H_{\rm EQ}}{\sqrt{2}}\left(\frac{a_{\rm EQ}}{a_1}\right)^2. \tag{4.1.4}$$

Using this relation to eliminate a_1, we can put the bound (4.1.2) in the more useful form

$$e^{\mathcal{N}} > \left(\frac{\Omega_M a_{\rm EQ}}{a_0}\right)^{1/4}\sqrt{\frac{H_1}{H_0}} = \Omega_R^{1/4}\sqrt{\frac{H_1}{H_0}} = \left(\Omega_R\frac{\rho_1}{\rho_{0,\,\rm crit}}\right)^{1/4} = \frac{[\rho_1]^{1/4}}{0.037h{\rm eV}}, \tag{4.1.5}$$

[475] R. H. Dicke and P. J. E. Peebles, in General Relativity - An Einstein Centenary Survey, eds. S. Hawking and W. Israel (Cambridge University Press, 1979).

where ρ_1 is the energy density at the beginning of the radiation-dominated era, and $\rho_{0,\,\mathrm{crit}} = [3.00 \times 10^{-3}\mathrm{eV}]^4\, h^2$ is the critical density (1.5.28).

To go further, we need some idea of the energy density at the end of inflation. The success of the theory of cosmological nucleosynthesis shows that ρ_1 cannot be less than the energy density at the time of the beginning of neutron-proton conversion, roughly $[\,1\ \mathrm{MeV}]^4$, in which case Eq. (4.1.5) with $h = 0.7$ requires that the universe expanded during inflation by at least a factor 4×10^7, or $17e$-foldings. At the other extreme, we would not expect ρ_1 to be greater than the Planck energy density $G^{-2} = [1.22 \times 10^{19}\mathrm{GeV}]^4$, in which case Eq. (4.1.5) with $h = 0.7$ would require that the expansion during inflation was at least by a factor 5×10^{29}, or $68e$-foldings. We will see some evidence in Section 10.3 that ρ_1 is of order $[2 \times 10^{16}\mathrm{GeV}]^4$, in which case $e^{\mathcal{N}}$ for $h = 0.7$ would have to be at least 8×10^{26}, so that $\mathcal{N} > 62$.

This is the least convincing of the arguments for inflation, because the small value of $|K|/\dot{a}^2$ in the past could be explained by the assumption (one that was often made before anyone heard of inflation) that space is precisely flat, so that $K = 0$ now and always. On the other hand, as we will discuss in Section 4.3, inflation opens up the interesting possibility that the universe in the large is not at all homogeneous and isotropic, and that its apparent flatness of the cosmic metric is just the result of inflation.

It would be quite a coincidence if inflation lasted for precisely the right number of e-foldings so that $|K|/\dot{a}^2$ would have decreased during inflation from an initial value of order unity just enough so that its subsequent increase during the radiation and matter-dominated eras would have
brought it back to a value of order unity now. It seems more likely that the present value $|\Omega_K|$ would be either much larger or much smaller than unity, and since observations tell us that it is not much larger than unity, inflationary theories suggest that $|\Omega_K| \ll 1$. (But this sort of reasoning would also suggest that the vacuum energy is much less than the present matter density, which we now know is not the case.)

B. Horizons

From the beginning the observed high degree of isotropy of the cosmic microwave radiation background posed a problem. Recall that the horizon size in a matter- or radiation-dominated universe is of order t, which, because $a(t)$ has increased as $t^{2/3}$ since the time of last scattering, was of order $d_H \approx H_0^{-1}\,(1+z_L)^{-3/2}$ at the time of last scattering. (See Eq. (2.6.32).) Also, according to Eq. (2.6.29), the angular diameter distance d_A to the surface of last scattering is of order $H_0^{-1}\,(1+z_L)^{-1}$, so the horizon at the time of last scattering now subtends an angle of order $d_H/d_A \approx (1+z_L)^{-1/2}$ radians, which for $z_L \simeq 1100$ is about 1.6°. Therefore in a matter- or radiation-dominated universe no physical influence could have smoothed out initial inhomogeneities and brought points at a redshift z_L that are separated by more than a few degrees to the same temperature, in contradiction with the nearly perfect isotropy of the microwave background at large angular scales observed ever since the background radiation was discovered. Inflation provides an explanation: during the inflationary era the part of the universe that we can observe would have occupied a tiny space, and there would have been plenty of time for everything in this space to be homogenized.

To work out what this means for the expansion during inflation, first recall that as discussed in Section 1.13, the proper horizon size at the time t_L of last scattering is

$$d_H\,(t_L) \equiv a\,(t_L) \int_{t_*}^{t_L} \frac{dt}{a(t)} \tag{4.1.6}$$

with t_* (possibly equal to $-\infty$) the beginning of the era of inflation. We have seen that the contribution to the integral from the radiation and matterdominated eras is much too small

to account for the isotropy of the microwave radiation background, so we will assume that the integral is dominated by an era of inflation. For definiteness, we assume that during inflation $a(t)$ increased exponentially at a rate H_I, so that

$$a(t) = a(t_*) \exp(H_I(t - t_*)) = a_I \exp(-H_I(t_I - t))$$

where t_I is again the time of the end of inflation, and $a_I = a(t_I)$. With $\mathcal{N} \equiv H_I(t_I - t_*)$ the number of e-foldings of expansion during inflation, Eq. (4.1.6) gives

$$d_H(t_L) = \frac{a(t_L)}{a_I H_I}\left[e^{\mathcal{N}} - 1\right] \tag{4.1.7}$$

In order to have any hope of solving the horizon problem, we must have $e^{\mathcal{N}} \gg 1$, so we can drop the term -1 in square brackets in Eq. (4.1.7).

To account for the observed high degree of isotropy of the cosmic microwave background at large angular scales we need $d_H(t_L) > d_A(t_L)$, where $d_A(t_L)$ is the angular-diameter distance of the surface of last scattering. According to Eq. (2.6.29),

$$d_A(t_L) \approx \frac{a(t_L)}{H_0 a_0} \tag{4.1.8}$$

The condition $d_H(t_L) > d_A(t_L)$ for the isotropy of the cosmic microwave background is then

$$e^{\mathcal{N}} > \frac{a_I H_I}{a_0 H_0}. \tag{4.1.9}$$

This is the same as the condition (4.1.2) for the solution of the flatness problem. If we again make the assumption that not much happens between the end of inflation and the beginning of the radiation-dominated era, and use Eq. (4.1.3), then to solve the horizon problem we again need $\mathcal{N} > 17$ if $\rho_1 \simeq [1\text{MeV}]^4$, $\mathcal{N} > 62$ if $\rho_1 \simeq [2 \times 10^{16}\text{GeV}]^4$, and $\mathcal{N} > 68$ if $\rho_1 \simeq G^{-2}$. We will see in Chapter 10 that whether 17 or 62 or 68e-foldings are needed to solve the horizon problem, it is only that number of e-foldings before the end of inflation that can be explored through observations of nonuniformities in the present universe.

It should be noted that the time t_L of last scattering does not enter in the bound (4.1.9) on $\mathcal{N}$, so this is also the condition that the whole sky at any redshift $z < z_{\text{EQ}}$ was within the horizon at the time that light observed now with that redshift left its source. Indeed, Eq. (4.1.9) is also the condition that the horizon size at the present should be greater than the size of the observable universe, which is roughly $1/H_0$.

C. Monopoles

In grand unified theories local symmetry under some simple symmetry group is spontaneously broken at an energy $M \approx 10^{16}$GeV to the gauge symmetry of the Standard Model under the group $SU(3) \times SU(2) \times U(1)$.

In all such cases, the scalar fields that break the symmetry can be left in twisted configurations that carry non-zero magnetic charge and that cannot be smoothed out through any continuous processes. [6] This poses a problem for some cosmological models. [7] The scalar fields before this phase transition would have necessarily been uncorrelated at distances larger than the horizon distance, the farthest distance that light could have traveled since the initial singularity. At an early time t in the standard big bang theory the horizon distance was of order $t \approx \left(G(k_{\mathcal{B}}T)^4\right)^{-1/2}$ (where $G \simeq (10^{19}\text{GeV})^{-2}$ is Newton's constant), so the number density of monopoles produced at the time that the temperature drops to $M/k_{\mathcal{B}}$ would have been of

order $t^{-3} \approx (GM^4)^{3/2}$, which is smaller than the photon density $\approx M^3$ at $T \approx M/k_{\mathcal{B}}$ by a factor of order $(GM^2)^{3/2}$. For $M \approx 10^{16}$GeV this factor is of order 10^{-9}. If monopoles did not find each other to annihilate, then this ratio would remain roughly constant to the present, but with at least 10^9 microwave background photons per nucleon today, this would give at least one monopole per nucleon, in gross disagreement with what is observed.

This potential paradox was one of the factors leading to interest in inflationary cosmological models. In such models, a period of exponential expansion that occurred before the monopoles were produced would have greatly extended the horizon, and an exponential expansion that occurred after the production of monopoles (but before photons were created in a period of reheating) would have greatly reduced the monopole to photon ratio. To be specific, the search for monopoles in iron ore, seawater, etc. shows that there are fewer than 10^{-6} per gram, or about 10^{-30} monopoles per nucleon, and hence fewer than about 10^{-39} monopoles per photon. [8] (With this abundance, even if the monopole mass were as large as 10^{19}GeV, they would make a negligible contribution to the cosmic mass density.) In order for inflation to have reduced the monopole/photon ratio by a factor 10^{-30}, it must have increased the horizon size (at some time before the reheating that creates photons) by a factor 10^{10}. That is, the horizon size $e^{\mathcal{N}}/H_1$ after inflation must be greater than the previous estimate $(GM^4)^{-1/2}$ by at least a factor 10^{10}. For $H_1 \approx (GM^4)^{1/2}$ this requires the number $\mathcal{N}$ of e-foldings to be greater than $\ln 10^{10} = 23$. Of course, another possible solution of the monopole problem is that inflation ends at a temperature below the grand unification scale M, so that there never was a time when the grand unification group was unbroken. An even simpler possibility, which does not rely on inflation, is that there may be no simple gauge group that is

spontaneously broken to the gauge group $SU(3) \times SU(2) \times U(1)$ of the Standard Model.

The most serious of the above three problems is the horizon problem. As we have seen, there are possible solutions of the flatness and monopole problems that do not rely on inflation. Also, any number of e-foldings of inflation that solves the horizon problem automatically solves not only the flatness problem, but also the monopole problem. If the radiationdominated era begins with an energy density $\rho_1 > [10^{15}\text{GeV}]^4$ then for inflation to solve the flatness and horizon problems we need at least 59 e-foldings of inflation, which is more than enough to avoid the monopole problem, while if $\rho_1 < [10^{15}\text{GeV}]^4$ GeV then in the usual picture of grand unification there would be no monopoles at all.

10.4.2 4.2 Slow-roll inflation

In Guth's original work, inflation was conceived to be due to a delayed first-order phase transition, in which a scalar field was initially trapped in a local minimum of some potential, and then leaked through the potential barrier and rolled toward a true minimum of the potential. It was soon realized [1] that this idea does not work, because of what has come to be called the graceful exit problem. The transition from the supercooled initial "false vacuum" phase to the lower energy "true vacuum" phase could not have occurred everywhere simultaneously, but here and there in small bubbles of true vacuum, which rapidly expanded into the background of false vacuum, in which the scalar field would have been still trapped in its local minimum, [2] like water droplets forming in supercooled water vapor. The trouble is that the latent heat released in the phase transition would have wound up in the bubble walls, leaving the interiors of the bubbles essentially empty, so that the only places where there would be energy that could grow into the present contents of the universe would be highly inhomogeneous and anisotropic.

[476] For a discussion, see QTF, Vol. II, Sec. 23.3.

[7] Ya. B. Zel'dovich and M. Yu. Khlopov, Phys. Lett. B 79, 239 (1978); J. Preskill, Phys. Rev. Lett. 43, 1365 (1979). For a review, see J. Preskill, Annual Rev. Nucl. Part. Science 34, 461 (1984).

[8] For a review, see Particle Data Group, Phys Lett. B 582, 1001 (2004).

At first Guth thought the bubbles in inflationary cosmologies would have merged, leading to our present more-or-less homogeneous universe, but this could not have happened; because the background false-vacuum space continued to inflate, the bubble walls would have moved too fast away from each other ever to have coalesced.

Guth's version of inflation was soon supplanted by a version due to Linde and to Albrecht and Steinhardt, known as 'new inflation. [3] Originally new inflation was formulated in a particular model of the breakdown of a grand unified symmetry, using a symmetry-breaking mechanism introduced by Coleman and E. Weinberg. [4] With this mechanism the zero-temperature potential for a scalar field φ is artificially adjusted to have zero second derivative at $\varphi = 0$. One-loop radiative corrections then give a potential equal to a known positive factor times $\varphi^4 \ln(\varphi/M)$, where M is a free constant; changing the value of M amounts to changing the φ^4 coupling constant. This potential has an unstable stationary point at $\varphi = 0$ and a minumum at $\varphi_0 = Me^{-1/4}$. At finite temperature T there is also a quadratic term in the potential, proportional to $T^2\varphi^2$, which makes the stationary point at $\varphi = 0$ into a local minimum. Again the phase transition occurs by forming bubbles, but for low temperature the potential barrier is very small, and so the scalar field in the interior of the bubble starts with φ nearly zero. The field then rolls slowly down the potential, in the manner discussed in Section 1.12, while the universe (including the bubble) undergoes an exponential expansion. Eventually the field energy is converted into ordinary particles, filling the bubble. Our observable universe is supposed to occupy a small part of one such bubble.

The consequences of the new inflationary theories turned out to depend on the slow roll of the scalar field after bubble formation, rather than the process of bubble formation itself. Indeed, the important aspects of inflation do not really require any assumptions about grand unification or the Coleman-E. Weinberg mechanism. All we need to assume is that there is a scalar field φ, known as the inflaton, which at some early time takes a value at which the potential $V(\varphi)$ is large but quite flat. The scalar field "rolls" very slowly at first down this potential, so that the Hubble constant decreases only slowly, and the universe experiences a more-or-less exponential inflation before the field changes very much.

To put this quantitatively, recall that the energy density (B.66) and pressure (B.67) of a spatially homogeneous scalar field $\varphi(t)$ with potential $V(\varphi)$ in a Robertson-Walker spacetime take the form (with $\hbar = c = 1$)

$$\rho = \frac{1}{2}\dot{\varphi}^2 + V(\varphi), \quad p = \frac{1}{2}\dot{\varphi}^2 - V(\varphi),$$

so the energy conservation equation $\dot{\rho} = -3H(\rho + p)$ takes the form:

$$\ddot{\varphi} + 3H\dot{\varphi} + V'(\varphi) = 0, \tag{4.2.1}$$

where $H \equiv \dot{a}/a$ is the time dependent expansion rate, which during the period of scalar field energy dominance is given by

$$H = \sqrt{\frac{8\pi G\rho}{3}} = \sqrt{\frac{8\pi G}{3}\left(\frac{1}{2}\dot{\varphi}^2 + V(\varphi)\right)} \tag{4.2.2}$$

[471] S. W. Hawking, I. G. Moss, and J. M. Stewart, Phys. Rev. D 26, 2681 (1982); A. H. Guth and E. J. Weinberg, Nucl. Phys. B 212, 321 (1983).

[2] For a description of this process and references to the original literature on bubble formation in quantum field theory, see QTF II, Section 23.8.

[473] A. D. Linde, Phys. Lett. B 108, 389 (1982); 114, 431 (1982); Phys. Rev. Lett. 48, 335 (1982); A. Albrecht and P. Steinhardt, Phys. Rev. Lett. 48, 1220 (1982).

[4] S. Coleman and E. Weinberg, Phys. Rev. D 7, 1888 (1973).

From Eqs. (4.2.1) and (4.2.2) we can derive an extremely useful formula for $\dot{H}$. By taking the time derivative of the square of Eq. (4.2.2) and then using Eq. (4.2.1), we have

$$2H\dot{H} = \frac{8\pi G}{3}\left(\dot{\varphi}\ddot{\varphi} + V'(\varphi)\dot{\varphi}\right) = -8\pi G H\dot{\varphi}^2,$$

and therefore

$$\dot{H} = -4\pi G\dot{\varphi}^2 \tag{4.2.3}$$

Now, in order to have a nearly exponential expansion, the fractional change $|\dot{H}/H|(1/H)$ in H during an expansion time $1/H$ must be much less than unity. That is, we must have

$$|\dot{H}| \ll H^2. \tag{4.2.4}$$

With Eqs. (4.2.3) and (4.2.2), this requires that

$$\dot{\varphi}^2 \ll |V(\varphi)|. \tag{4.2.5}$$

This has the consequence that $p \simeq -\rho$, and also

$$H \simeq \sqrt{\frac{8\pi G V(\varphi)}{3}} \tag{4.2.6}$$

Usually it is also assumed that the fractional change $|\ddot{\varphi}/\dot{\varphi}|(1/H)$ in $\dot{\varphi}$ during an expansion time $1/H$ is much less than unity. That is,

$$|\ddot{\varphi}| \ll H|\dot{\varphi}|. \tag{4.2.7}$$

This has the consequence that we may drop the inertial term $\ddot{\varphi}$ in Eq. (4.2.1), which then becomes

$$\dot{\varphi} = -\frac{V'(\varphi)}{3H} = -\frac{V'(\varphi)}{\sqrt{24\pi G V(\varphi)}}. \tag{4.2.8}$$

The fractional change of the expansion rate H in an expansion time $1/H$ will then be

$$\frac{|\dot{H}|}{H^2} = \frac{1}{2}\sqrt{\frac{3}{8\pi G}}\left|\frac{V'(\varphi)\dot{\varphi}}{V^{3/2}(\varphi)}\right| = \frac{1}{16\pi G}\left(\frac{V'(\varphi)}{V(\varphi)}\right)^2, \tag{4.2.9}$$

so the exponential expansion of the universe will last for many e-foldings if

$$\left|\frac{V'(\varphi)}{V(\varphi)}\right| \ll \sqrt{16\pi G}. \tag{4.2.10}$$

According to Eq. (4.2.8), the condition on the potential for the inequality (4.2.5) to be satisfied is that

$$\left|\frac{V'(\varphi)}{V(\varphi)}\right| \ll \sqrt{24\pi G}, \tag{4.2.11}$$

which is guaranteed by the inequality (4.2.10). Also, Eq. (4.2.8) gives

$$\ddot{\varphi} = -\frac{V''(\varphi)\dot{\varphi}}{3H} + \frac{V'(\varphi)\dot{H}}{3H^2} = \frac{V''(\varphi)V'(\varphi)}{9H^2} - \frac{V'^3}{48\pi G V^2}. \tag{4.2.12}$$

The inequality (4.2.10) ensures that the absolute value of the last term on the right-hand side is much less than $|V'(\varphi)|$, so the condition for $|\ddot{\varphi}|$ to be much less than $|V'(\varphi)|$ is that $|V''(\varphi)| \ll 9H^2$, or, in other words,

$$\left|\frac{V''(\varphi)}{V(\varphi)}\right| \ll 24\pi G. \tag{4.2.13}$$

Eqs. (4.2.10) and (4.2.13) are the two "flatness" conditions needed to insure the slow roll of both φ and $\dot{\varphi}$. It is possible in principle that the second flatness condition (4.2.13) may not be satisfied for a potential that does satisfy the first flatness condition (4.2.10), but this is unusual, and in particular is not possible for the simple potentials discussed below.

Under these conditions the expansion is generally not strictly exponential, but it can easily be exponentially large. Suppose that during some time interval the field $\varphi(t)$ shifts from an initial value φ_1 to a final value φ_2, with $0 < V(\varphi_2) < V(\varphi_1)$, with both inequalities (4.2.10) and (4.2.13) assumed valid over this range of φ. The Robertson-Walker scale factor will increase during this period by a factor

$$\begin{aligned}\frac{a(t_2)}{a(t_1)} &= \exp\left[\int_{t_1}^{t_2} H dt\right] = \exp\left[\int_{\varphi_1}^{\varphi_2} \frac{H d\varphi}{\dot{\varphi}}\right] \\ &\simeq \exp\left[-\int_{\varphi_1}^{\varphi_2}\left(\frac{8\pi G V(\varphi)}{V'(\varphi)}\right) d\varphi\right]\end{aligned} \tag{4.2.14}$$

In this range the potential is positive and decreases as $\varphi(t)$ runs from φ_1 to φ_2, so the argument of the exponential in Eq. (4.2.14) is positive. Condition (4.2.10) tells us that this argument is much greater than $\sqrt{4\pi G}\,|\varphi_1 - \varphi_2|$, so this flatness condition guarantees that we get a large number of e-foldings
in any time interval in which φ changes by an amount at least as large as $1/\sqrt{4\pi G} = 3.4 \times 10^{18}$GeV.

It is important to note that such large values of the scalar field do not necessarily rule out the classical treatment of gravitation on which we have been relying. The condition that allows us to neglect quantum gravitational effects is that the energy density should be much less than the Planck energy density:

$$|V(\varphi)| \ll (4\pi G)^{-2} \tag{4.2.15}$$

This condition can be satisfied, even if φ is comparable to the Planck mass $\approx G^{-1/2}$, by supposing that $V(\varphi)$ is proportional to a sufficiently small coupling constant. Neither the flatness conditions (4.2.10), (4.2.13) nor the growth (4.2.14) of $a(t)$ for a given change in $\varphi(t)$ depend on the value of such a coupling constant.

Depending on the potential shape, the flatness conditions (4.2.10) and (4.2.13) may provide either conditions on the initial value of the scalar field or on the parameters of the potential itself. As an example of the first sort, consider the power-law potential

$$V(\varphi) = g\varphi^{\alpha} \tag{4.2.16}$$

with g and α arbitrary real parameters, except that we assume that $g > 0$ and take $|\alpha|$ larger but not orders of magnitude larger than unity. The flatness conditions (4.2.10) and (4.2.13) are then both satisfied for $|\varphi| \gg 1/\sqrt{4\pi G}$, irrespective of the value of the coupling constant g. The number of e foldings of expansion for a scalar field that starts at a value φ_1 and ends at a much smaller value is given by Eq. (4.2.14) as $8\pi G\varphi_1^2/\alpha$, so for instance for $\alpha = 4$ we get the $62e$-foldings needed to avoid the horizon problem for inflation ending at a temperature $2 \times 10^{16}\text{GeV}/k_{\mathcal{B}}$ if $|\varphi_1| > \sqrt{31/\pi G}$. On the other hand, for this potential the condition (4.2.15) for the neglect of quantum gravitational effects does put an upper bound on $|g|$. For instance, in the case $\alpha = 4$, the potential (4.2.16) will satisfy condition (4.2.15) if $g \ll (\sqrt{4\pi G}|\varphi|)^{-4}$ so, with $|\varphi|$ just large enough to get $62e$ foldings of exponential expansion, to avoid quantum

gravity corrections we need $g \ll 2 \times 10^{-5}$. The need for a very small coupling can be avoided in theories with more than one scalar field, such as "hybrid inflation" theories, [5] in which the effective self-coupling of one scalar field is very small because the other scalar field has a very small mass and hence a small expectation value.

There is a well-known example of a rather different sort for which Eqs. (4.2.1) and (4.2.2) can be solved exactly, the exponential potential: [6]

$$V(\varphi) = g \exp(-\lambda\varphi), \tag{4.2.17}$$

with g and λ arbitrary constants. It is easy to verify that Eqs. (4.2.1) and (4.2.2) are satisfied by [7]

$$\varphi(t) = \frac{1}{\lambda} \ln\left(\frac{8\pi G g \epsilon^2 t^2}{3-\epsilon}\right), \tag{4.2.18}$$

and

$$a \propto t^{1/\epsilon}, \quad H = 1/\epsilon t, \tag{4.2.19}$$

where ϵ is the positive dimensionless quantity

$$\epsilon \equiv \frac{\lambda^2}{16\pi G}. \tag{4.2.20}$$

The flatness conditions (4.2.10) and (4.2.13) here respectively read $\epsilon \ll 1$ and $\epsilon \ll 3/2$, so both are satisfied if and only if

$$\epsilon \ll 1, \tag{4.2.21}$$

with no constraint on the values of either g or φ. This exact solution is useful as a check of approximate calculations for more general potentials. For instance, Eq. (4.2.18) may be compared with the solution of the approximate equation (4.2.8), which for the exponential potential is

$$\varphi(t) = \frac{1}{\lambda} \ln\left(\frac{8\pi G g \epsilon^2 t^2}{3}\right), \tag{4.2.22}$$

The difference between this and the exact solution (4.2.18) is evidently negligible for $\epsilon \ll 1$. Likewise, the increase in the Robertson-Walker scale factor during a time interval from t_1 to t_2 in which the field drops from φ_1 to φ_2 is

$$\frac{a(t_2)}{a(t_1)} = \left(\frac{t_2}{t_1}\right)^{1/\epsilon} = e^{(\varphi_2-\varphi_1)\lambda/2\epsilon}, \tag{4.2.23}$$

which is exactly the same as the "slow-roll" result (4.2.14) for the exponential potential.

It is assumed in these theories that when $V(\varphi)$ dropped sufficiently far, the inequalities (4.2.10) and (4.2.13) were in general no longer be satisfied, and φ began a damped oscillation around the minimum of $V(\varphi)$, which is at the present value φ_0. Eventually φ would have approached close enough to φ_0 so that we can approximate the potential as a quadratic,

[475] A. D. Linde, Phys. Lett. B 259, 38 (1991); Phys. Rev. D 49, 748 (1994) [astro-ph/9307002].

[476] L. F. Abbott and M. B. Wise, Nucl. Phys. B 244, 541 (1984); F. Lucchin and S. Matarrese, Phys. Rev. D 32, 1316 (1985); Phys. Lett. B 164, 282 (1985); D. H. Lyth and E. D. Stewart, Phys. Lett. B 274, 168 (1992).

[7] This is an exact solution for all values of t, but it is not the most general solution. Since Eqs. (4.2.1) is a second-order differential equation, it has a two-parameter set of solutions. The particular solution (4.2.18) is an attractor, which the general solutions approach for large t.

$$V(\varphi) = \frac{1}{2} m^2 (\varphi - \varphi_0)^2 \tag{4.2.24}$$

(In order to account for the small present value of the vacuum energy, it is necessary to assume that, for reasons that remain entirely mysterious, the minimum value of the potential is very close to zero.) This is just like the field theory of spinless particles with mass m and negligible velocity. In order to have ended inflation, there must also be some coupling of the inflaton scalar field to other fields, including the fields of ordinary matter and radiation, so that the energy density of the inflaton field decreased as

$$\rho_\varphi(t) = \rho_\varphi(t_I) \left(\frac{a(t_I)}{a(t)} \right)^3 e^{-\Gamma(t-t_I)} \tag{4.2.25}$$

where Γ is the rate of decay of the φ quanta into other particles, and t_I is taken at the beginning of the inflaton oscillation and decay. This is known as the period of reheating. [8] It is this period in which the entropy observed in the present universe is supposed to be generated. [9]

The energy density ρ_M of the particles into which φ decayed satisfies a conservation equation like Eq. (1.1.32), but corrected to take account of the flow of energy from the inflaton:

$$\dot{\rho}_M + 3H(\rho_M + p_M) = \Gamma \rho_\varphi \tag{4.2.26}$$

For definiteness, we will assume that the decay products of the inflaton are highly relativistic, so that $p_M = \rho_M/3$. Then the solution of Eq. (4.2.26) is

$$\rho_M(t) = \frac{\rho_\varphi(t_I)\,\Gamma a^3(t_I)}{a^4(t)} \int_{t_I}^{t} a(t')\, e^{-\Gamma(t'-t_I)} dt' \tag{4.2.27}$$

In contemporary models of inflation this is the source of all the matter and radiation in the present universe. (In using this relation it is important

to keep in mind the possibility of parameteric amplification of the density $\rho_\varphi(t_I)$. That is, scalar field expectation values can increase the effective mass of the inflaton field φ, and thereby increase its energy density. [10])

The matter energy density (4.2.27) starts equal to zero at $t = t_I$, then rises at first, and finally falls as the density is attenuated by the expansion of the universe. It is of some interest to find the value of $\rho_M(t)$ at its maximum, because this tells us the maximum temperature ever reached, which controls the kinds of relics - cold dark matter, baryons, monopoles, axions - left over from the early universe. This maximum density can be easily calculated in two extreme cases, for $\Gamma \gg H(t_I)$ and $\Gamma \ll H(t_I)$.

For $\Gamma \gg H(t_I)$, we can express Eq. (4.2.27) as a power series in $H(t_I)/\Gamma$ by repeated integration by parts

$$\rho_M(t) = \rho_\varphi(t_I) \left(\frac{a(t_I)}{a(t)} \right)^4 \left(1 + \frac{H(t_I)}{\Gamma} + \ldots \right) \tag{4.2.28}$$

We see that in this case ρ_M jumped up almost immediately to the value $\rho_\varphi(t_I)$, and then decreased with the usual a^{-4} factor, so in this case all the energy of the inflaton field at the end of inflation went into ordinary matter and radiation. This was the assumption made in deriving lower bounds on the number of e-foldings of inflation in the previous section.

[478] The term reheating is a historical relic of early theories of inflation in which it was assumed that the zero-temperature slow roll of the inflaton field followed an earlier period of high temperature.

[9] A. D. Dolgov and A. D. Linde, Phys. Lett. 116B, 329 (1982); L. F. Abbott, E. Farhi, and M. B. Wise, Phys. Lett. 117B, 29 (1982).

For $\Gamma \ll H(t_I)$, the maximum value of $\rho_M(t)$ was reached at a time when the exponential $e^{-\Gamma(t'-t_I)}$ in Eq. (4.2.27) had not yet begun to decay. (This will be checked below.) Setting this factor equal to unity, we have

$$\rho_M(t) \simeq \frac{\rho_\varphi(t_I)\,\Gamma a^3(t_I)}{a^4(t)} \int_{t_I}^{t} a(t')\,dt' \tag{4.2.29}$$

At this time the energy density of the universe was still dominated by the inflaton, so $a(t) = a(t_I)(t/t_I)^{2/3}$, and Eq. (4.2.29) becomes

$$\rho_M(t) \simeq \frac{3}{5}\Gamma t_I \rho_\varphi(t_I) \left(\frac{t_I}{t}\right)^{8/3} \left(\left(\frac{t}{t_I}\right)^{5/3} - 1\right). \tag{4.2.30}$$

This reached a maximum at $t = (8/3)^{3/5} t_I$, which incidentally confirms that, under the assumption that $\Gamma \ll H(t_I) = 2/3t_I$, the argument of the exponential in Eq. (4.2.27) at this maximum was still negligible. At this maximum, the matter density is

$$\rho_{M,\max} = (3/8)^{8/5}\Gamma t_I \rho_\varphi(t_I) = 0.139\,(\Gamma/H(t_I))\,\rho_\varphi(t_I). \tag{4.2.31}$$

In this case the maximum energy density at the beginning of the radiationdominated era would have been much less than the energy density in the inflaton field at the end of inflation.

10.4.3 4.3 Chaotic inflation, eternal inflation

It was soon realized that some "new inflation" models actually entail an endless production of inflating bubbles. [1] This has come to be called "eternal inflation." [2] To take one example, if the inflaton scalar field at a given point in space once had a value of unstable equilibrium (like at the top of a potential hill) then the probability that the inflaton field was still at this value after a time t decreased as $\exp(-\gamma t)$. However, the volume in which the scalar field had this value was meanwhile increasing as $\exp(+3Ht)$, so as long as $3H > \gamma$ the volume of space that still undergoes inflation eternally increases exponentially.

We have been assuming that the scalar field is initially independent of position, aside from small perturbations, about which more in Chapter 10. Soon after the introduction of new inflation, the possibilities of inflationary theory were greatly expanded and improved when Linde [3] proposed the theory of "chaotic inflation," in which initially one or more scalar fields varied in a random way with position. Here and there one would have found patches of space in which an inflaton field took a nearly uniform value at which the potential satisfied the slow-roll conditions (4.2.10) and (4.2.13), as for instance a value substantially greater than the Planck mass for a power-law potential. Inflation will then have occured in such a patch, provided the patch was initially sufficiently large.

It is necessary to require that the uniform patch be sufficiently large, because the scalar field Lagrangian density given by Eq. (B.63) contains a term involving spatial derivatives, which for a non-uniform scalar field contributes a term $-a^{-2}\nabla^2\varphi$ on the left-hand side of Eq. (4.2.1). In order for this term not to interfere with the slow-roll analysis of the previous section, we need the scale L of proper distances over which φ takes a roughly constant initial value to be greater than $|\varphi/V'(\varphi)|^{1/2}$. For instance, for the

[4][7][1][0] L. Kofman, A. D. Linde, and A. A. Starobinsky, Phys. Rev. Lett. 73, 3195 (1994); Phys. Rev. D 56, 3258 (1997).

[471] P. J. Steinhardt, in The Very Early Universe - Proceedings of the Nuffield Workshop, 1982, eds. G. W. Gibbon S. W. Hawking, and S. T. C. Siklos (Cambridge University Press, 1983): 251; A. Vilenkin, Phys. Rev. D 27, 2848 (1983).

[2] For reviews of this and other variants of inflation, see A. Guth, talk given at the Pritzker Symposium on

potential $V(\varphi) = g\varphi^4$, this condition gives

$$L \gg \frac{1}{2\sqrt{g}|\varphi|} = \left(\frac{1}{g(4\pi G)^2\varphi^4}\right)^{1/2} (\sqrt{\pi G}\varphi)\sqrt{4\pi G} \tag{4.3.1}$$

The classical field theory condition (4.2.15) makes the first factor much larger than unity, while the slow-roll condition (4.2.10) makes the second factor also much larger than unity, so the patch size must be very much greater than the third factor, which is essentially the Planck length. Such relatively large uniform patches may be quite rare, but that is no argument against this hypothesis, because life can only arise in big bangs that stem from such patches.

A sufficiently large patch will inflate to an enormous size, which to observers deep inside seems highly homogeneous and isotropic. In this way, chaotic inflation solves a puzzle that had not generally been realized to be a puzzle, even when the first inflation theories were being developed. It explains not just why the Robertson-Walker metric in which we find ourselves was remarkably flat in the past; it also explains why we find ourselves in a Robertson-Walker metric at all. Unfortunately, it is hard to see how we will ever observe any part of the universe beyond our inflated patch. The validity of the idea of chaotic inflation will probably have to come from progress in fundamental physics, which may verify the existence of a suitable inflaton field, rather than from astronomical observation.

Even when the scalar field in a patch of the space was large enough to start a slow roll inflation, quantum fluctuations in smaller regions within that patch would subsequently have driven the inflaton field to even higher values, so that these regions will begin an earlier stage of inflation. [4] In this way, chaotic inflation turns out also to be eternal.

10.5 General Theory of Small Fluctuations

In most of the work described in the previous chapters, the universe has been treated as isotropic and homogeneous, with a gravitational field described by the Robertson-Walker metric. This is of course just an approximation, which ignores many of the most interesting things in the universe: galaxy clusters, galaxies, stars, us. We now turn to an analysis of these departures from homogeneity and isotropy.

In this chapter we will lay a foundation for this analysis, by deriving the general relativistic equations that govern small fluctuations, and drawing general conclusions about their implications. Chapter 6 will apply this formalism to the evolution of structure, from the radiation-dominated era to near the present. In Chapters 7 and 8 we apply the results obtained in Chapter 6 to the observed fluctuations in the cosmic microwave background and to the growth of structure. Chapter 9 deals with gravitational lensing, which may in the long run provide the best tool for analyzing the large scale structure of dark matter. These chapters are kept at a general level, independent of detailed assumptions about an inflationary era before the radiation-dominated era. Chapter 10 will then explore the implications of inflationary theories for the calculations of Chapters 6 through 9. Some readers may prefer to skip on immediately to Chapter 6, using the present chapter as a source of useful formulas, while others will do better to read these chapters in order.

---

the Status of Inflationary Cosmology, January 1999, astro-ph/0002188; A. Linde, J. Phys. Conf. Ser. 24, 151 (2005) [hep-th/0503195].

[3] A.D. Linde, Phys. Lett. 129B, 177 (1983).

[474] A. D. Linde, Mod. Phys. Lett. A1, 81 (1986); Phys. Lett. B 175, 395 (1986); A. S. Goncharev, A. D. Linde, and V. F. Mukhanov, Int. J. Mod. Phys. A 2, 561 (1987); A. D. Linde, D. A. Linde, and A. Mezhlumian, Phys. Rev. D 49, 1783 (1994) [gr-qc/9306035].

10.5.1 5.1 Field equations

As an essential feature of the analysis presented here, we assume that during most of the history of the universe all departures from homogeneity and isotropy have been small, so that they can be treated as first-order perturbations. [1] Because the observable universe is nearly homogeneous, and its spatial curvature either vanishes or is negligible until very near the present, we will take the unperturbed metric to have the Robertson-Walker form (1.1.11), with curvature constant $K = 0$. (Effects of a possible finite curvature at times close to the present will be included in Chapters 7 through 9

where they are relevant.) The total perturbed metric is then

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}, \tag{5.1.1}$$

where $\bar{g}_{\mu\nu}$ is the unperturbed $K = 0$ Robertson-Walker metric

$$\bar{g}_{00} = -1, \quad \bar{g}_{i0} = \bar{g}_{0i} = 0, \quad \bar{g}_{ij} = a^2(t)\delta_{ij}, \tag{5.1.2}$$

and $h_{\mu\nu} = h_{\nu\mu}$ is a small perturbation. (Here and from now on, a bar over any quantity denotes its unperturbed value.) The perturbation to the inverse of a general matrix M is $\delta M^{-1} = -M^{-1}(\delta M)M^{-1}$, so the inverse metric is perturbed by

$$h^{\mu\nu} \equiv g^{\mu\nu} - \bar{g}^{\mu\nu} = -\bar{g}^{\mu\rho}\bar{g}^{\nu\sigma}h_{\rho\sigma}, \tag{5.1.3}$$

with components

$$h^{ij} = -a^{-4}h_{ij}, \quad h^{i0} = a^{-2}h_{i0}, \quad h^{00} = -h_{00} \tag{5.1.4}$$

Note the - sign in the last expression in Eq. (5.1.3); in our notation, the perturbation $\delta g^{\mu\nu}$ to $g^{\mu\nu}$ is not given by simply using the unperturbed metric to raise the indices on $\delta g_{\mu\nu}$.

The metric perturbation produces a perturbation to the affine connection

$$\delta\Gamma^{\mu}_{\nu\lambda} = \frac{1}{2}\bar{g}^{\mu\rho}\left[-2h_{\rho\sigma}\bar{\Gamma}^{\sigma}_{\nu\lambda} + \partial_\lambda h_{\rho\nu} + \partial_\nu h_{\rho\lambda} - \partial_\rho h_{\lambda\nu}\right]. \tag{5.1.5}$$

For $K = 0$, the only non-vanishing components of the unperturbed affine connection are given by Eqs. (1.1.17)-(1.1.19) as

$$\bar{\Gamma}^{i}_{j0} = \bar{\Gamma}^{i}_{0j} = \frac{\dot{a}}{a}\delta_{ij}, \quad \bar{\Gamma}^{0}_{ij} = a\dot{a}\delta_{ij}$$

Thus Eq. (5.1.5) gives the components of the perturbed affine connection as

[471] The study of first-order cosmological fluctuations was initiated by E. Lifshitz, J. Phys. U.S.S.R. 10, 116 (1946). Classical second-order corrections were worked out by K. Tomita, Prog. Theor. Phys. 37, 831 (1967); 45, 1747 (1970); 47, 416 (1971). For recent work, see K. Tomita, Phys. Rev. D 71, 3504 (2005); N. Bartolo, S. Matarrese, and A. Riotto, J. Cosm. & Astropart. Phys. 0606, 024 (2006) [astro-ph/0604416].

$$\delta\Gamma^i_{jk} = \frac{1}{2a^2}\left(-2a\dot{a}h_{i0}\delta_{jk} + \partial_k h_{ij} + \partial_j h_{ik} - \partial_i h_{jk}\right) \tag{5.1.6}$$

$$\delta\Gamma^i_{j0} = \frac{1}{2a^2}\left(-\frac{2\dot{a}}{a}h_{ij} + \dot{h}_{ij} + \partial_j h_{i0} - \partial_i h_{j0}\right) \tag{5.1.7}$$

$$\delta\Gamma^0_{ij} = \frac{1}{2}\left(2a\dot{a}\delta_{ij}h_{00} - \partial_j h_{i0} - \partial_i h_{j0} + \dot{h}_{ij}\right) \tag{5.1.8}$$

$$\delta\Gamma^i_{00} = \frac{1}{2a^2}\left(2\dot{h}_{i0} - \partial_i h_{00}\right) \tag{5.1.9}$$

$$\delta\Gamma^0_{i0} = \frac{\dot{a}}{a}h_{i0} - \frac{1}{2}\partial_i h_{00} \tag{5.1.10}$$

$$\delta\Gamma^0_{00} = -\frac{1}{2}\dot{h}_{00} \tag{5.1.11}$$

In particular, we will need

$$\delta\Gamma^\lambda_{\lambda\mu} = \partial_\mu\left[\frac{1}{2a^2}h_{ii} - \frac{1}{2}h_{00}\right].$$

To write the Einstein equations, we need the perturbation to the Ricci tensor

$$\begin{aligned}\delta R_{\mu\kappa} =& \frac{\partial\delta\Gamma^\lambda_{\mu\lambda}}{\partial x^\kappa} - \frac{\partial\delta\Gamma^\lambda_{\mu\kappa}}{\partial x^\lambda} \\ &+ \delta\Gamma^\eta_{\mu\nu}\bar{\Gamma}^v_{\kappa\eta} + \delta\Gamma^v_{\kappa\eta}\bar{\Gamma}^\eta_{\mu\nu} - \delta\Gamma^\eta_{\mu\kappa}\bar{\Gamma}^v_{\nu\eta} - \delta\Gamma^v_{\nu\eta}\bar{\Gamma}^\eta_{\mu\kappa}\end{aligned} \tag{5.1.12}$$

with components

$$\begin{aligned}\delta R_{jk} =& -\frac{1}{2}\partial_j\partial_k h_{00} - \left(2\dot{a}^2 + a\ddot{a}\right)\delta_{jk}h_{00} - \frac{1}{2}a\dot{a}\delta_{jk}\dot{h}_{00} \\ &+ \frac{1}{2a^2}\left(\nabla^2 h_{jk} - \partial_i\partial_j h_{ik} - \partial_i\partial_k h_{ij} + \partial_j\partial_k h_{ii}\right) \\ &- \frac{1}{2}\ddot{h}_{jk} + \frac{\dot{a}}{2a}\left(\dot{h}_{jk} - \delta_{jk}\dot{h}_{ii}\right) + \frac{\dot{a}^2}{a^2}\left(-2h_{jk} + \delta_{jk}h_{ii}\right) + \frac{\dot{a}}{a}\delta_{jk}\partial_i h_{i0} \\ &+ \frac{1}{2}\left(\partial_j\dot{h}_{k0} + \partial_k\dot{h}_{j0}\right) + \frac{\dot{a}}{2a}\left(\partial_j h_{k0} + \partial_k h_{j0}\right)\end{aligned} \tag{5.1.13}$$

$$\begin{aligned}\delta R_{0j} =& \delta R_{j0} = \frac{\dot{a}}{a}\partial_j h_{00} + \frac{1}{2a^2}\left(\nabla^2 h_{j0} - \partial_j\partial_i h_{i0}\right) - \left(\frac{\ddot{a}}{a} + \frac{2\dot{a}^2}{a^2}\right)h_{j0} \\ &+ \frac{1}{2}\frac{\partial}{\partial t}\left[\frac{1}{a^2}\left(\partial_j h_{kk} - \partial_k h_{kj}\right)\right]\end{aligned} \tag{5.1.14}$$

$$\begin{aligned}\delta R_{00} =& \frac{1}{2a^2}\nabla^2 h_{00} + \frac{3\dot{a}}{2a}\dot{h}_{00} - \frac{1}{a^2}\partial_i\dot{h}_{i0} \\ &+ \frac{1}{2a^2}\left[\ddot{h}_{ii} - \frac{2\dot{a}}{a}\dot{h}_{ii} + 2\left(\frac{\dot{a}^2}{a^2} - \frac{\ddot{a}}{a}\right)h_{ii}\right]\end{aligned} \tag{5.1.15}$$

In general, we can put the Einstein field equations (B.71) in the form

$$R_{\mu\nu} = -8\pi G S_{\mu\nu}, \tag{5.1.16}$$

where

$$S_{\mu\nu} = T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}g^{\rho\sigma}T_{\rho\sigma}. \tag{5.1.17}$$

(A cosmological constant can be accommodated by including a term in $T_{\mu\nu}$ proportional to $g_{\mu\nu}$, with a constant coefficient.) The perturbation to the
energy-momentum tensor and metric produces a perturbation to the source tensor $S_{\mu\nu}$:

$$\delta S_{\mu\nu} = \delta T_{\mu\nu} - \frac{1}{2}\bar{g}_{\mu\nu}\delta T^{\lambda}_{\lambda} - \frac{1}{2}h_{\mu\nu}\bar{T}^{\lambda}_{\lambda}. \tag{5.1.18}$$

We are not assuming that the contents of the universe form a perfect fluid, but the rotational and translational invariance of the unperturbed energymomentum tensor $\bar{T}^{\mu\nu}$ require that it takes the perfect fluid form (B.42):

$$\bar{T}_{\mu\nu} = \bar{p}\bar{g}_{\mu\nu} + (\bar{p}+\bar{\rho})\bar{u}_{\mu}\bar{u}_{\nu} \tag{5.1.19}$$

where $\bar{\rho}(t), \bar{p}(t)$, and $\bar{u}^{\mu}$ are the unperturbed energy density, pressure, and velocity four-vector, respectively, with $\bar{u}^0 = 1$ and $\bar{u}^i = 0$. Also, we use the unperturbed Einstein equations (1.5.17) and (1.5.18) to write $\bar{\rho}$ and $\bar{p}$ in terms of the Robertson-Walker scale factor and its derivatives

$$\bar{\rho} = \frac{3}{8\pi G}\left(\frac{\dot{a}^2}{a^2}\right), \quad \bar{p} = -\frac{1}{8\pi G}\left(\frac{2\ddot{a}}{a} + \frac{\dot{a}^2}{a^2}\right). \tag{5.1.20}$$

It follows in particular that the unperturbed energy-momentum tensor has the trace

$$\bar{T}^{\lambda}_{\lambda} = 3\bar{p} - \bar{\rho} = -\frac{3}{4\pi G}\left(\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2}\right).$$

Thus Eq. (5.1.18) gives

$$\delta S_{jk} = \delta T_{jk} - \frac{a^2}{2}\delta_{jk}\delta T^{\lambda}_{\lambda} + \frac{3}{8\pi G}\left(\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2}\right)h_{jk}, \tag{5.1.21}$$

$$\delta S_{j0} = \delta T_{j0} + \frac{3}{8\pi G}\left(\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2}\right)h_{j0} \tag{5.1.22}$$

$$\delta S_{00} = \delta T_{00} + \frac{1}{2}\delta T^{\lambda}_{\lambda} + \frac{3}{8\pi G}\left(\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2}\right)h_{00}. \tag{5.1.23}$$

The Einstein equations (5.1.16) thus take the form

$$\begin{aligned}
-8\pi G\left(\delta T_{jk} - \frac{a^2}{2}\delta_{jk}\delta T^{\lambda}_{\lambda}\right) &= -\frac{1}{2}\partial_j\partial_k h_{00} - \left(2\dot{a}^2 + a\ddot{a}\right)\delta_{jk}h_{00} \\
&\quad -\frac{1}{2}a\dot{a}\delta_{jk}\dot{h}_{00} + \frac{1}{2a^2}\left(\nabla^2 h_{jk} - \partial_i\partial_j h_{ik} - \partial_i\partial_k h_{ij} + \partial_j\partial_k h_{ii}\right) \\
-\frac{1}{2}\ddot{h}_{jk} + \frac{\dot{a}}{2a}\left(\dot{h}_{jk} - \delta_{jk}\dot{h}_{ii}\right) + \left(\frac{\dot{a}^2}{a^2} + \frac{3\ddot{a}}{a}\right)h_{jk} + \left(\frac{\dot{a}^2}{a^2}\right)\delta_{jk}h_{ii} \\
+\frac{\dot{a}}{a}\delta_{jk}\partial_i h_{i0} + \frac{1}{2}\left(\partial_j\dot{h}_{k0} + \partial_k\dot{h}_{j0}\right) + \frac{\dot{a}}{2a}\left(\partial_j h_{k0} + \partial_k h_{j0}\right) &\qquad (5.1.24) \\
-8\pi G\delta T_{j0} = \frac{\dot{a}}{a}\partial_j h_{00} + \frac{1}{2a^2}\left(\nabla^2 h_{j0} - \partial_j\partial_i h_{i0}\right) + \left(\frac{\dot{a}^2}{a^2} + \frac{2\ddot{a}}{a}\right)h_{j0} \\
+\frac{1}{2}\frac{\partial}{\partial t}\left[\frac{1}{a^2}\left(\partial_j h_{kk} - \partial_k h_{kj}\right)\right] &\qquad (5.1.25) \\
-8\pi G\left(\delta T_{00} + \frac{1}{2}\delta T^{\lambda}_{\lambda}\right) = \frac{1}{2a^2}\nabla^2 h_{00} + \frac{3\dot{a}}{2a}\dot{h}_{00} - \frac{1}{a^2}\partial_i\dot{h}_{i0} \\
+\frac{1}{2a^2}\left[\ddot{h}_{ii} - \frac{2\dot{a}}{a}\dot{h}_{ii} + 2\left(\frac{\dot{a}^2}{a^2} - \frac{\ddot{a}}{a}\right)h_{ii}\right] + 3\left(\frac{\dot{a}^2}{a^2} + \frac{\ddot{a}}{a}\right)h_{00}. &\qquad (5.1.26)
\end{aligned}$$

The components of the energy momentum tensor are subject to the conservation condition that $T^{\mu}{}_{\nu;\mu} = 0$, which to first order in perturbations gives

$$\partial_\mu \delta T^\mu_\nu + \bar{\Gamma}^\mu_{\mu\lambda}\delta T^\lambda_\nu - \bar{\Gamma}^\lambda_{\mu\nu}\delta T^\mu_\lambda + \delta\Gamma^\mu_{\mu\lambda}\bar{T}^\lambda_\nu - \delta\Gamma^\lambda_{\mu\nu}\bar{T}^\mu_\lambda = 0 \tag{5.1.27}$$

in which the perturbations to the energy-momentum tensor $\delta T^{\mu}{}_{\nu}$ with mixed indices can be calculated from

$$\delta T^\mu_\nu = \bar{g}^{\mu\lambda}\left[\delta T_{\lambda\nu} - h_{\lambda\kappa}\bar{T}^\kappa_\nu\right] \tag{5.1.28}$$

Setting v equal to a spatial coordinate index j gives the equation of momentum conservation

$$\partial_0 \delta T^0_j + \partial_i \delta T^i_j + \frac{2\dot{a}}{a}\delta T^0_j - a\dot{a}\delta T^j_0 - (\bar{\rho}+\bar{p})\left(\frac{1}{2}\partial_j h_{00} - \frac{\dot{a}}{a}h_{j0}\right) = 0, \tag{5.1.29}$$

while setting v equal to the time coordinate index 0 gives the equation of energy conservation

$$\partial_0 \delta T^0_0 + \partial_i \delta T^i_0 + \frac{3\dot{a}}{a}\delta T^0_0 - \frac{\dot{a}}{a}\delta T^i_i - \left(\frac{\bar{\rho}+\bar{p}}{2a^2}\right)\left(-\frac{2\dot{a}}{a}h_{ii} + \dot{h}_{ii}\right) = 0. \tag{5.1.30}$$

As remarked in Appendix B, these conservation equations are not independent conditions, but may be derived from the Einstein field equations. However, it is often convenient to use either or both in place of one or two of the field equations. Also, in the frequently encountered case where the constituents of the universe are non-interacting fluids (such as one fluid consisting of cold dark matter and another consisting of ordinary matter
and radiation) these conservation equations are satisfied separately by each fluid, information that could not be derived from the field equations.

The results obtained so far are repulsively complicated. Fortunately, the spatial isotropy and homogeneity of the unperturbed metric and energymomentum tensor allow us to simplify these results by decomposing the perturbations into scalars, divergenceless vectors, and divergenceless traceless symmetric tensors, which are not coupled to each other by the field equations or conservation equations. The perturbation to the metric can always be put in the form [2]

$$h_{00} = -E \tag{5.1.31}$$

$$h_{i0} = a\left[\frac{\partial F}{\partial x^i} + G_i\right] \tag{5.1.32}$$

$$h_{ij} = a^2\left[A\delta_{ij} + \frac{\partial^2 B}{\partial x^i \partial x^j} + \frac{\partial C_i}{\partial x^j} + \frac{\partial C_j}{\partial x^i} + D_{ij}\right] \tag{5.1.33}$$

where the perturbations $A, B, C_i, D_{ij} = D_{ji}, E, F$, and G_i are functions of $\mathbf{x}$ and t, satisfying the conditions

$$\frac{\partial C_i}{\partial x^i} = \frac{\partial G_i}{\partial x^i} = 0, \quad \frac{\partial D_{ij}}{\partial x^i} = 0, \quad D_{ii} = 0 \tag{5.1.34}$$

To carry out a similar decomposition of the energy-momentum tensor, we first note that for a perfect fluid we would have

$$T_{\mu\nu} = p g_{\mu\nu} + (\rho + p)u_\mu u_\nu \tag{5.1.35}$$

with

$$g^{\mu\nu}u_\mu u_\nu = -1 \tag{5.1.36}$$

Recalling that $\bar{u}_i = 0$ and $\bar{u}_0 = -1$, we find that the normalization condition Eq. (5.1.36) gives

$$\delta u^0 = \delta u_0 = h_{00}/2 \tag{5.1.37}$$

while δu_i is an independent dynamical variable. (Note that $\delta u^\mu \equiv \delta\left(g^{\mu\nu}u_\nu\right)$ is not given by $\bar{g}^{\mu\nu}\delta u_\nu$.) Then the first-order perturbation to the energymomentum tensor for a perfect fluid is

$$\delta T_{ij} = \bar{p}h_{ij} + a^2\delta_{ij}\delta p, \quad \delta T_{i0} = \bar{p}h_{i0} - (\bar{\rho}+\bar{p})\delta u_i, \quad \delta T_{00} = -\bar{\rho}h_{00} + \delta\rho \tag{5.1.38}$$

More generally, we can always put the perturbed energy-momentum tensor in a form like that of the perturbed metric. In general, we define $\delta\rho$ just as for a perfect fluid, as the difference between δT_{00} and $-\bar{\rho}h_{00}$, but $\delta\rho$ is not necessarily given by varying the temperature and chemical potentials in the formula for ρ that applies in thermal equilibrium. Also, in general we define the velocity perturbation δu_i times $\bar{\rho}+\bar{p}$ as for a perfect fluid, as the difference between $-\delta T_{i0}$ and $\bar{p}h_{i0}$, and we decompose δu_i into the gradient of a scalar velocity potential δu and a divergenceless vector δu_i^V. Finally, we define $a^2\delta p$ as the coefficient of δ_{ij} in the difference between δT_{ij} and $\bar{p}\delta_{ij}$, again without assuming that δp is given by varying the temperature and chemical potentials in the formula for p that applies in thermal equilibrium. The other terms in δT_{ij}, denoted $\partial_i\partial_j\pi^S, \partial_i\pi_j^V + \partial_j\pi_i^V$, and π_{ij}^T, represent dissipative corrections to the inertia tensor. That is, we write

$$\delta T_{ij} = \bar{p}h_{ij} + a^2\left[\delta_{ij}\delta p + \partial_i\partial_j\pi^S + \partial_i\pi_j^V + \partial_j\pi_i^V + \pi_{ij}^T\right] \tag{5.1.39}$$

$$\delta T_{i0} = \bar{p}h_{i0} - (\bar{\rho}+\bar{p})\left(\partial_i\delta u + \delta u_i^V\right) \tag{5.1.40}$$

$$\delta T_{00} = -\bar{\rho}h_{00} + \delta\rho \tag{5.1.41}$$

where π_i^V, π_{ij}^T, and δu_i^V satisfy conditions analogous to the conditions (5.1.34) satisfied by C_i, D_{ij}, and G_i :

$$\partial_i\pi_i^V = \partial_i\delta u_i^V = 0, \quad \partial_i\pi_{ij}^T = 0, \quad \pi_{ii}^T = 0 \tag{5.1.42}$$

To repeat, these formulas can be taken as a definition of the quantities $\delta\rho$, δp, and $\delta u_i \equiv \partial_i\delta u + \delta u_i^V$, as well as of the anisotropic inertia terms π^S, π^V, and π^T, which characterize departures from the perfect fluid form of the energy-momentum tensor. The perturbed mixed components (5.1.28) of the energy-momentum tensor, which are needed in the conservation laws, now take the form

$$\begin{aligned}
\delta T_j^i &= \delta_{ij}\delta p + \partial_i\partial_j\pi^S + \partial_i\pi_j^V + \partial_j\pi_i^V + \pi_{ij}^T \\
\delta T_0^i &= a^{-2}(\bar{\rho}+\bar{p})\left(a\partial_i F + aG_i - \partial_i\delta u - \delta u_i^V\right) \\
\delta T_i^0 &= (\bar{\rho}+\bar{p})\left(\partial_i\delta u + \delta u_i^V\right), \quad \delta T_0^0 = -\delta\rho, \\
\delta T_\lambda^\lambda &= 3\delta p - \delta\rho + \nabla^2\pi^S.
\end{aligned} \tag{5.1.43}$$

With these decompositions, and again using Eqs. (5.1.20), the Einstein field equations (5.1.24)-(5.1.26) and conservation equations (5.1.29) and (5.1.30) fall into three classes of coupled equations:

[472] To see this, we define F, A, and B, as the solutions of $\nabla^2 F = a^{-1}\partial_i h_{i0}, \nabla^2 A + \nabla^4 B = a^{-2}\partial_i\partial_j h_{jk}$, and $3A + \nabla^2 B = a^{-2}h_{ii}$, then define C_i as the solution of $\nabla^2 C_i = a^{-2}\partial_j h_{jk} - \partial_i\left[A + \nabla^2 B\right]$, and then use Eqs. (5.1.32) and (5.1.33) to define G_i and D_{jk}.

Scalar (compressional) modes

These are the most complicated, involving the eight scalars $E, F, A, B, \delta\rho,\ \delta p, \pi^S$, and δu. The part of Eq. (5.1.24) proportional to δ_{jk} gives

$$\begin{aligned}-4\pi G a^2\left[\delta\rho-\delta p-\nabla^2\pi^S\right] =&\frac{1}{2}a\dot{a}\dot{E}+\left(2\dot{a}^2+a\ddot{a}\right)E+\frac{1}{2}\nabla^2 A-\frac{1}{2}a^2\ddot{A}\\ &-3a\dot{a}\dot{A}-\frac{1}{2}a\dot{a}\nabla^2\dot{B}+\dot{a}\nabla^2 F\end{aligned} \tag{5.1.44}$$

The part of Eq. (5.1.24) of the form $\partial_j\partial_k S$ (where S is any scalar) gives

$$\partial_j\partial_k\left[16\pi G a^2\pi^S+E+A-a^2\ddot{B}-3a\dot{a}\dot{B}+2a\dot{F}+4\dot{a}F\right]=0 \tag{5.1.45}$$

The part of Eq. (5.1.25) of the form $\partial_j S$ (where S is again any scalar) gives

$$8\pi G a(\bar{\rho}+\bar{p})\partial_j\delta u=-\dot{a}\partial_j E+a\partial_j\dot{A} \tag{5.1.46}$$

Eq. (5.1.26) gives

$$\begin{aligned}-4\pi G\left(\delta\rho+3\delta p+\nabla^2\pi^S\right)=&-\frac{1}{2a^2}\nabla^2 E-\frac{3\dot{a}}{2a}\dot{E}-\frac{1}{a}\nabla^2\dot{F}-\frac{\dot{a}}{a^2}\nabla^2 F\\ &+\frac{3}{2}\ddot{A}+\frac{3\dot{a}}{a}\dot{A}-\frac{3\ddot{a}}{a}E+\frac{1}{2}\nabla^2\ddot{B}+\frac{\dot{a}}{a}\nabla^2\dot{B}\end{aligned} \tag{5.1.47}$$

The part of the momentum conservation condition (5.1.29) that is a derivative ∂_j is

$$\partial_j\left[\delta p+\nabla^2\pi^S+\partial_0[(\bar{\rho}+\bar{p})\delta u]+\frac{3\dot{a}}{a}(\bar{\rho}+\bar{p})\delta u+\frac{1}{2}(\bar{\rho}+\bar{p})E\right]=0 \tag{5.1.48}$$

and the energy-conservation condition (5.1.30) is

$$\begin{aligned}&\delta\dot{\rho}+\frac{3\dot{a}}{a}(\delta\rho+\delta p)+\nabla^2\left[-a^{-1}(\bar{\rho}+\bar{p})F+a^{-2}(\bar{\rho}+\bar{p})\delta u+\frac{\dot{a}}{a}\pi^S\right]\\ &+\frac{1}{2}(\bar{\rho}+\bar{p})\partial_0\left[3A+\nabla^2 B\right]=0.\end{aligned} \tag{5.1.49}$$

In Eqs. (5.1.48) and (5.1.49), $\delta\rho, \delta p$, and π^S are elements of the perturbation to the total energy-momentum tensor, but the same equations apply to each constituent of the universe that does not exchange energy and momentum with other constituents.

Vector (vortical) modes

These involve the four divergenceless vectors $G_i, C_i, \delta u_i^V$, and π_i^V. The part of Eq. (5.1.24) of the form $\partial_k V_j$ (where V_j is any vector satisfying $\partial_j V_j = 0$) gives

$$\partial_k\left[16\pi G a^2\pi_j^V-a^2\ddot{C}_j-3a\dot{a}\dot{C}_j+a\dot{G}_j+2\dot{a}G_j\right]=0, \tag{5.1.50}$$

while the part of Eq. (5.1.25) of the form V_j (where V_j is again any vector satisfying $\partial_j V_j = 0$) gives

$$8\pi G(\bar{\rho}+\bar{p})a\delta u_j^V=\frac{1}{2}\nabla^2 G_j-\frac{a}{2}\nabla^2\dot{C}_j. \tag{5.1.51}$$

The part of the momentum conservation equation (5.1.29) that takes the form of a divergenceless vector is

$$\nabla^2\pi_j^V + \partial_0\left[(\bar\rho+\bar p)\delta u_j^V\right] + \frac{3\dot a}{a}(\bar\rho+\bar p)\delta u_j^V = 0, \tag{5.1.52}$$

In particular, for a perfect fluid $\pi_i^V = 0$, and Eq. (5.1.52) tells us that $(\ \bar\rho+\bar p)\delta u_j^V$ decays as $1/a^3$. In this case, both Eqs. (5.1.50) and (5.1.51) imply that the quantity $G_j - a\dot C_j$ (which we will see in Section 5.3 is the only physically relevant combination of metric components for vector perturbations) decays as $1/a^2$. Because they decay, vector modes have not played a large role in cosmology.

Tensor (radiative) modes

These involve only the two traceless divergenceless symmetric tensors D_{ij} and π_{ij}^T. There is only one field equation here: the part of Eq. (5.1.24) of the form of a divergenceless traceless tensor is the wave equation for gravitational radiation

$$-16\pi G a^2\pi_{ij}^T = \nabla^2 D_{ij} - a^2\ddot D_{ij} - 3a\dot a\dot D_{ij}. \tag{5.1.53}$$

The above equations for scalar, vector, and tensor perturbations do not form a complete set. This is in part because we still have the freedom to make changes in the coordinate system, of the same order as the physical perturbations. In Section 5.3 we will see how to remove this freedom by a choice of "gauge."

But even after the gauge has been fixed, the equations for the scalar modes will still not form a complete set, unless the pressure p and anisotropic inertia π^S can be expressed as functions of the energy density ρ. This is the
case, for instance, for a constituent of the universe such as cold dark matter, whose particles have negligible velocities, and do not interact with other constituents. For such constituents, we can simply set p and π^S equal to zero. Things are only a little more complicated for constituents of the universe whose particles' velocities cannot be neglected, but which experience collisions that are sufficiently rapid to maintain local thermal equilibrium. In such cases, π^S can be neglected, and if the particles are highly relativistic p is simply $\rho/3$. Even where the particles are only moderately relativistic, the pressure in thermal equilibrium can usually be expressed as a function of ρ and one or more number densities n that satisfy the condition that the current nu^μ is conserved

$$(nu^\mu)_{;\mu} = 0. \tag{5.1.54}$$

This condition tells us that the unperturbed number density satisfies $\bar n \propto a^{-3}$, while the perturbation satisfies

$$\frac{\partial}{\partial t}\left(\frac{\delta n}{\bar n}\right) + \frac{1}{a^2}\nabla^2\delta u + \frac{1}{2}\left(3\dot A + \nabla^2\dot B\right) - \frac{1}{a}\nabla^2 F = 0. \tag{5.1.55}$$

With p given as a function of ρ and n, after gauge fixing the field equations and conservation equations (5.1.48), (5.1.49), and (5.1.55) form a complete set of equations for the scalar modes.

Similarly, even after gauge fixing, the equations for vector and tensor modes do not form a complete set unless we have formulas for π_i^V and π_{ij}^T, respectively. This is no problem for perfect fluids, for which $\pi_i^V = \pi_{ij}^T = 0$.

In the general case local thermal equilibrium is not maintained, and we must calculate $\delta\rho, \delta p, \pi^S, \pi_i^V$ and π_{ij}^T by following changes in the distribution of individual particle positions and momenta, which are governed by Boltzmann equations. The Boltzmann equations for photons and neutrinos are derived in Appendix H, and used in Chapters 6 and 7.

10.5.2 5.2 Fourier decomposition and stochastic initial conditions

In order to simplify our work we want to make full use of the symmetries of the problem. We have already used the rotational symmetry of the field and conservation equations in sorting out perturbations into scalar, vector, and tensor modes, and we will apply rotational symmetry again later in this section. The equations also have a symmetry under translations in space, which can best be exploited by working with the Fourier components of the perturbations. As long as we treat perturbations as infinitesimal, there is no coupling between the Fourier components of different wave number.

Further, only the initial conditions but not the equations themselves depend on the direction of the co-moving wave number $\mathbf{q}$.

Let us first see how this works out for the scalar modes. Because the unperturbed metric, energy density, etc. are independent of position, the general solution for the corresponding perturbed quantities may be written as a superposition of plane wave solutions, whose spatial variation is given by factors $\exp(i\mathbf{q}\cdot\mathbf{x})$, where $\mathbf{q}$ is a real wave vector, over which we must integrate. The coefficient of $\exp(i\mathbf{q}\cdot\mathbf{x})$ in each of the scalar perturbations $A(\mathbf{x},t), B(\mathbf{x},t)$, etc. is a solution of coupled ordinary differential equations with time as the independent variable, obtained by replacing $\partial/\partial x^j$ with iq_j everywhere in the equations presented in Section 5.1. The differential equations obtained in this way depend on $q \equiv |\mathbf{q}|$, but not on the direction of $\mathbf{q}$, so the solutions can be written as superpositions of independent q-dependent normal modes, each characterized by a set of perturbations $A_{nq}(t), B_{nq}(t)$, etc., with an overall normalization factor $\alpha_n(\mathbf{q})$ carrying a discrete index n that labels the various independent solutions. These normalization factors depend on the initial conditions, which of course are not rotationally invariant - if they were then there would be no galaxies or stars - so the $\alpha_n(\mathbf{q})$ depend on the direction of $\mathbf{q}$, but the solutions $A_{nq}(t), B_{nq}(t)$ can be chosen to depend only on $q \equiv |\mathbf{q}|$. That is, we write

$$A(\mathbf{x},t) = \sum_n \int d^3q \alpha_n(\mathbf{q}) A_{nq}(t) e^{i\mathbf{q}\cdot\mathbf{x}}$$
$$B(\mathbf{x},t) = \sum_n \int d^3q \alpha_n(\mathbf{q}) B_{nq}(t) e^{i\mathbf{q}\cdot\mathbf{x}} \tag{5.2.1}$$

and likewise (with the same $\alpha_n(\mathbf{q})$) for E and F as well as for $\delta\rho, \delta p, \delta u, \pi^S$ and any other rotational scalars.

Now, as discussed in Section 2.6, we expect the scalar variables $A(\mathbf{x},t)$, $B(\mathbf{x},t)$, etc. to be stochastic variables, characterized by averages of their products. The solutions $\{A_{nq}(t), B_{nq}(t), \ldots\}$ are ordinary fixed functions, not stochastic variables, so the stochastic nature of the scalar variables arises from the stochastic character of the initial conditions, embodied in the factors $\alpha_n(\mathbf{q})$. Under the assumption that the scalar variables are governed by Gaussian distributions, of the sort discussed in Appendix E, all averages of scalar quantities can be expressed in terms of bilinear averages $\langle A(\mathbf{x},t)A(\mathbf{y},t)\rangle, \langle A(\mathbf{x},t)B(\mathbf{y},t)\rangle$, etc. Let us consider the average of the product of any two real scalar quantities $X(\mathbf{x},t)$ and $Y(\mathbf{y},t)$. It turns out to be very convenient to use Eq. (5.2.1) for X and its complex conjugate for Y, so that

$$\langle X(\mathbf{x},t)Y(\mathbf{y},t)\rangle = \sum_{nm} \int d^3q \int d^3q' X_{nq}(t) Y^*_{mq}(t) \left\langle \alpha_n(\mathbf{q}) \alpha^*_m(\mathbf{q}') \right\rangle e^{i\mathbf{q}\cdot\mathbf{x}} e^{-i\mathbf{q}'\cdot\mathbf{y}}$$

We assume that although the initial conditions are not translationally invariant, they are governed by a translationally invariant probability distribution function, so that $\langle X(\mathbf{x},t)Y(\mathbf{y},t)\rangle$ should be a function only of $\mathbf{x}-\mathbf{y}$. This immediately tells us that $\langle \alpha_n(\mathbf{q})\alpha^*_m(\mathbf{q}')\rangle$ must be proportional to a delta function, $\delta^3(\mathbf{q}-\mathbf{q}')$. Furthermore, although the initial conditions are not

rotationally invariant, we also assume that they are governed by a rotationally invariant probability distribution, so that the function of $\mathbf{q}$ multiplying this delta function can only depend on the magnitude $q \equiv |\mathbf{q}|$ of the wave vector, not on its direction. That is, we can write

$$\langle \alpha_n(\mathbf{q}) \alpha_m^*(\mathbf{q}') \rangle = \mathcal{P}_{nm}(q) \delta^3(\mathbf{q} - \mathbf{q}'), \tag{5.2.2}$$

and so

$$\langle X(\mathbf{x}, t) Y(\mathbf{y}, t) \rangle = \sum_{nm} \int d^3q X_{nq}(t) Y_{mq}^*(t) \mathcal{P}_{nm}(q) \exp(i\mathbf{q} \cdot (\mathbf{x} - \mathbf{y}) \tag{5.2.3}$$

The task of the theory of cosmological perturbations is twofold: to find the solutions $A_{nq}(t), B_{nq}(t)$, etc. of the differential equations under suitable assumptions about the constituents of the universe, and to calculate the spectral functions $\mathcal{P}_{nm}(q)$ in a theory of the origin of fluctuations in the very early universe.

Inspection of Eq. (5.2.2) shows immediately that $\mathcal{P}_{mn}(q)$ is a Hermitian matrix

$$\mathcal{P}_{nm}^*(q) = \mathcal{P}_{mn}(q) \tag{5.2.4}$$

and it is positive, in the sense that

$$\sum_{nm} \mathcal{P}_{nm}(q) \xi_n \xi_m^* > 0 \tag{5.2.5}$$

for any set of complex numbers ξ_n (or functions of q) that are not all zero.
In general, for an arbitrary choice of independent solutions, there is no reason why $\mathcal{P}_{nm}(q)$ should also be diagonal, so in general there will be interference between the different modes. However, it is sometimes convenient to choose the solutions so that $\mathcal{P}_{nm}(q)$ is simply equal to δ_{nm}. To see that this can always be done, we recall a theorem of matrix algebra, which says that, because it is positive and Hermitian, $\mathcal{P}_{nm}(q)$ can be put in the form

$$\mathcal{P}_{nm}(q) = \sum_r Z_{nr}(q) Z_{mr}^*(q) \tag{5.2.6}$$

for some square matrix $Z_{nr}(q)$, with r running over as many values as n. (That is, in matrix notation, $\mathcal{P} = ZZ^\dagger$.) We can then redefine the
independent solutions, defining

$$\tilde{A}_{rq}(t) \equiv \sum_n Z_{nr}(q) A_{nq}(t), \quad \tilde{B}_{rq}(t) \equiv \sum_n Z_{nr}(q) B_{nq}(t), \text{ etc. },$$

with a corresponding redefinition of the normalization factors

$$\alpha_n(q) \equiv \sum_r Z_{nr}(q) \tilde{\alpha}_r(q)$$

chosen so that

$$A(\mathbf{x}, t) = \sum_r \int d^3q \tilde{\alpha}_r(\mathbf{q}) \tilde{A}_{rq}(t) e^{i\mathbf{q}\cdot\mathbf{x}}$$

and likewise (with the same $\tilde{\alpha}_r(\mathbf{q})$) for B, E, and F, as well as for $\delta\rho, \delta p, \delta u$, etc. The advantage of this is that now the relevant bilinear averages are

$$\langle \tilde{\alpha}_r(\mathbf{q}) \tilde{\alpha}_s^*(\mathbf{q}') \rangle = \delta_{rs} \delta^3(\mathbf{q} - \mathbf{q}') \tag{5.2.7}$$

This is the result that was to be proved. With the solutions $\tilde{A}_{rq}(t)$, etc. chosen so that the $\tilde{\alpha}_r(\mathbf{q})$ satisfy Eq. (5.2.7), the different modes are uncorrelated. That is, any binary average is a sum over the individual modes. For instance,

$$\langle A(\mathbf{x},t)A(\mathbf{y},t)\rangle = \sum_r \int d^3q e^{i\mathbf{q}\cdot(\mathbf{x}-\mathbf{y})} \left|\tilde{A}_{rq}(t)\right|^2 \tag{5.2.8}$$

In this way, the problem of calculating the spectral function $\mathcal{P}_{nm}(q)$ is traded for the problem of finding the correct linear combination of solutions for which Eq. (5.2.7) applies.

Incidentally, there is no problem in calculating the averages $\langle\alpha_n\alpha_m\rangle$ if we know the averages $\langle\alpha_n\alpha_m^*\rangle$, because the α_n satisfy a reality condition. To derive this condition, note that the differential equations for $A_{nq}(t)$, $B_{nq}(t)$, etc. are real, so the set of complex conjugates $A_{nq}^*(t), B_{nq}^*(t)$, etc. of solution n is also a solution, and can therefore be expressed as a set of linear combinations $\sum_m c_{nm}(q)A_{mq}(t), \sum_m c_{nm}(q)B_{mq}(t)$, etc. (Of course, in the case of a real solution, $c_{nm}(q) = \delta_{nm}$.) The functions $A(\mathbf{x},t), B(\mathbf{x},t)$ are real, so by taking the complex conjugate of Eqs. (5.2.1) and replacing the integration variable $\boldsymbol{q}$ with $-\boldsymbol{q}$ we see that

$$\sum_n \alpha_n^*(\mathbf{q})c_{nm}(q) = \alpha_m(-\mathbf{q}) \tag{5.2.9}$$

In particular, it follows that

$$\langle\alpha_n(\mathbf{q})\alpha_m(\mathbf{q}')\rangle = \sum_l \mathcal{P}_{nl}(q)c_{lm}(q)\delta^3(\mathbf{q}+\mathbf{q}')$$

Now let us consider the tensor modes. These are completely characterized by the traceless divergenceless symmetric tensor $D_{ij}(\mathbf{x},t)$. (The stress tensor $\pi_{ij}^T(\mathbf{x},t)$ is not an independent dynamical quantity but, as we will see in Section 6.6, it is given by a linear functional of $\dot{D}_{ij}$-specifically, the solution of a linear integral equation with $\dot{D}_{ij}$ as the inhomogeneous term.) If we write D_{ij} as a Fourier integral

$$D_{ij}(\mathbf{x},t) = \int d^3q e^{i\mathbf{q}\cdot\mathbf{x}}\mathcal{D}_{ij}(\mathbf{q},t), \tag{5.2.10}$$

then the Fourier transform $\mathcal{D}$ must satisfy the conditions

$$\mathcal{D}_{ij} = \mathcal{D}_{ji}, \quad \mathcal{D}_{ii} = 0, \quad q_i\mathcal{D}_{ij} = 0. \tag{5.2.11}$$

For a given wave vector $\mathbf{q}$, there are just two independent matrices satisfying these conditions. For instance, for $\mathbf{q}$ in the three-direction, Eq. (5.2.11) requires that

$$\mathcal{D}_{11} = -\mathcal{D}_{22}, \quad \mathcal{D}_{12} = \mathcal{D}_{21}, \quad \mathcal{D}_{i3} = \mathcal{D}_{3i} = 0, \tag{5.2.12}$$

so all $\mathcal{D}_{ij}$ can be expressed in terms of the two independent components $\mathcal{D}_{11}$ and $\mathcal{D}_{12}$. (These components are frequently denoted $h^+ = \mathcal{D}_{11}$ and $h^\times = \mathcal{D}_{12}$.) It is convenient to classify the possible $\mathcal{D}_{ij}$ by their transformation properties under a rotation by an angle θ around the three-axis. [1] It is easily seen that under such a rotation,

$$\begin{aligned}
\mathcal{D}_{11} &\to \cos^2\theta\mathcal{D}_{11} + \cos\theta\sin\theta\mathcal{D}_{12} + \sin\theta\cos\theta\mathcal{D}_{21} + \sin^2\theta\mathcal{D}_{22} \\
&= \cos 2\theta\mathcal{D}_{11} + \sin 2\theta\mathcal{D}_{12} \\
\mathcal{D}_{12} &\to \cos^2\theta\mathcal{D}_{12} - \cos\theta\sin\theta\mathcal{D}_{11} + \sin\theta\cos\theta\mathcal{D}_{22} - \sin^2\theta\mathcal{D}_{21} \\
&= -\sin 2\theta\mathcal{D}_{11} + \cos 2\theta\mathcal{D}_{12},
\end{aligned}$$

or more succinctly

$$\mathcal{D}_{11} \mp i\mathcal{D}_{12} \to e^{\pm 2i\theta}\left[\mathcal{D}_{11} \mp i\mathcal{D}_{12}\right] \tag{5.2.13}$$

For this reason, the linear combinations $\mathcal{D}_{11} \mp i\mathcal{D}_{12}$ are said to have helicity ± 2. (A wave of helicity λ consists of quanta with angular momentum in the direction of motion equal to $\hbar\lambda$.) We will write $\mathcal{D}_{ij}(\mathbf{q}, t)$ as a sum over helicities:

$$\mathcal{D}_{ij}(\mathbf{q}, t) = \sum_{\lambda=\pm 2} e_{ij}(\hat{q}, \lambda)\mathcal{D}(\mathbf{q}, \lambda, t), \tag{5.2.14}$$

where for $\mathbf{q}$ in the three-direction

$$\begin{aligned} e_{11}(\hat{z}, \pm 2) = -e_{22}(\hat{z}, \pm 2) = \mp ie_{12}(\hat{z}, \pm 2) = \mp ie_{21}(\hat{z}, \pm 2) = \frac{1}{\sqrt{2}} \\ e_{i3} = e_{3i} = 0 \end{aligned} \tag{5.2.15}$$

while for $\hat{q}$ in any other direction $e_{ij}(\hat{q}, \pm 2)$ is defined by applying on each of the indices i and j a standard rotation, that takes the three-direction into the direction of $\hat{q}$.

For $\mathbf{q}$ in the three-direction, the combination $\mathcal{D}_{11}(\mathbf{q}, t) \mp i\mathcal{D}_{12}(\mathbf{q}, t)$ is proportional to $\mathcal{D}(\mathbf{q}, \pm 2, t)$, so according to the transformation rule (5.2.13), a rotation by an angle θ around the direction of $\mathbf{q}$ changes $\mathcal{D}(\mathbf{q}, \pm 2, t)$ by a factor $\exp(\pm 2i\theta)$. There is nothing special about the three-direction, and the same is true for $\mathbf{q}$ in any direction.

The quantities $\mathcal{D}(\mathbf{q}, \lambda, t)$ in Eq. (5.2.14) satisfy the λ-independent second-order field equation (5.1.53), with the Laplacian ∇^2 replaced with $-q^2 \equiv -\mathbf{q}^2$:

$$\ddot{\mathcal{D}}(\mathbf{q}, \lambda, t) + 3\frac{\dot{a}}{a}\dot{\mathcal{D}}(\mathbf{q}, \lambda, t) + \frac{q^2}{a^2}\mathcal{D}(\mathbf{q}, \lambda, t) = 16\pi G\pi^T(\mathbf{q}, \lambda, t) \tag{5.2.16}$$

where $\pi^T(\mathbf{q}, \lambda, t)$ is the Fourier transform of the tensor part of the anisotropic inertia tensor:

$$\pi^T_{ij}(\mathbf{x}, t) = \sum_{\lambda=\pm 2}\int d^3q e^{i\mathbf{q}\cdot\mathbf{x}} e_{ij}(\hat{q}, \lambda)\pi^T(\mathbf{q}, \lambda, t) \tag{5.2.17}$$

As already mentioned, $\pi^T(\mathbf{q}, \lambda, t)$ is a helicity-independent linear functional of $\dot{\mathcal{D}}(\mathbf{q}, \lambda, t)$. Therefore, just as in the absence of anisotropic inertia, there are two independent solutions (distinguished by a label N) that, aside from normalization factors β, are independent of λ and of the direction of $\mathbf{q}$. Thus we can write

$$\mathcal{D}(\mathbf{q}, \lambda, t) = \sum_N \beta_N(\mathbf{q}, \lambda)\mathcal{D}_{Nq}(t) \tag{5.2.18}$$

and

$$\pi^T(\mathbf{q}, \lambda, t) = \sum_N \beta_N(\mathbf{q}, \lambda)\pi^T_{Nq}(t) \tag{5.2.19}$$

the sum over N running over the labels of the two independent solutions $\mathcal{D}_{Nq}(t)$ of the field equation

$$\ddot{\mathcal{D}}_{Nq}(t) + 3\frac{\dot{a}}{a}\dot{\mathcal{D}}_{Nq}(t) + \frac{q^2}{a^2}\mathcal{D}_{Nq}(t) = 16\pi G\pi^T_{Nq}(t) \tag{5.2.20}$$

Putting together Eqs. (5.2.10), (5.2.14), and (5.2.18), we have

[471] See, for instance, G&C, Sec. 10.2.

$$D_{ij}(\mathbf{x},t) = \sum_{N,\lambda} \int d^3q e^{i\mathbf{q}\cdot\mathbf{x}} e_{ij}(\hat{q},\lambda)\beta_N(\mathbf{q},\lambda)\mathcal{D}_{Nq}(t) \tag{5.2.21}$$

The normalization factors $\beta_N(\mathbf{q},\lambda)$ like the $\alpha_n(\mathbf{q})$ are stochastic variables. The translational invariance of the probability distribution governing these factors tells us that $\langle \beta_N(\mathbf{q},\lambda)\beta^*_{N'}(\mathbf{q}',\lambda')\rangle$ is proportional to $\delta^3(\mathbf{q}-\mathbf{q}')$. Under a rotation by an angle θ around the direction of $\mathbf{q}$, the product $\beta_N(\mathbf{q},\lambda)\beta^*_{N'}(\mathbf{q},\lambda')$ is changed by a factor $\exp(i\theta(\lambda-\lambda'))$, so the rotational invariance of the probability distribution requires further that $\langle \beta_N(\mathbf{q},\lambda)\beta^*_{N'}(\mathbf{q}',\lambda')\rangle$ is also proportional to $\delta_{\lambda\lambda'}$. For the most part, we will also assume that the probability distribution is invariant under the space-inversion operator P . This operator reverses the direction of momentum but not of angular momentum, so it changes the sign of the helicity, and therefore with this assumption, aside from the factor $\delta_{\lambda\lambda'}$, the mean value $\langle \beta_N(\mathbf{q},\lambda)\beta^*_{N'}(\mathbf{q}',\lambda')\rangle$ is independent of helicity. [2] We have then

$$\langle \beta_N(\mathbf{q},\lambda)\beta^*_{N'}(\mathbf{q}',\lambda')\rangle = \mathcal{P}^{\rm grav}_{NN'}(q)\delta_{\lambda\lambda'}\delta^3(\mathbf{q}-\mathbf{q}') \tag{5.2.22}$$

Rotational and translational invariance also tell us that there is no correlation between the normalization factors for scalar and tensor modes:

$$\langle \beta_N(\mathbf{q},\lambda)\alpha^*_n(\mathbf{q}')\rangle = 0 \tag{5.2.23}$$

Just as in the case of scalar modes, it is always possible to choose the two tensor modes so that $\mathcal{P}^{\rm grav}_{NN'}(q)$ is just $\delta_{NN'}$, but we need to know the probability distribution governing initial conditions in order to decide which linear combination of modes have this property.

The average of a product of two tensor perturbations is given by Eqs. (5.2.21) and (5.2.22) as

$$\begin{aligned}\langle D_{ij}(\mathbf{x},t)D_{kl}(\mathbf{y},t)\rangle =& \sum_{NN'}\int d^3q \mathcal{P}^{\rm grav}_{NN'}(q)e^{i\mathbf{q}\cdot(\mathbf{x}-\mathbf{y})} \\ & \times \Pi_{ij,k\ell}(\hat{q})\mathcal{D}_{Nq}(t)\mathcal{D}^*_{N'q}(t)\end{aligned} \tag{5.2.24}$$

where [3]

$$\begin{aligned}\Pi_{ij,k\ell}(\hat{q}) \equiv& \sum_\lambda e_{ij}(\hat{q},\lambda)e^*_{k\ell}(\hat{q},\lambda) \\ =& \delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk} - \delta_{ij}\delta_{k\ell} + \delta_{ij}\hat{q}_k\hat{q}_\ell + \delta_{k\ell}\hat{q}_i\hat{q}_j \\ & - \delta_{ik}\hat{q}_j\hat{q}_\ell - \delta_{i\ell}\hat{q}_j\hat{q}_k - \delta_{jk}\hat{q}_i\hat{q}_\ell - \delta_{j\ell}\hat{q}_i\hat{q}_k + \hat{q}_i\hat{q}_j\hat{q}_k\hat{q}_\ell\end{aligned} \tag{5.2.25}$$

Formulas like Eqs. (5.2.24) and (5.2.3) will be applied to the cosmic microwave background in Section 7.4.

We will see in Section 5.4 that the anisotropic inertia π_{Nq} becomes negligible during the era when the physical wave number q/a is much less than the expansion rate $\dot{a}/a$. In this case, one tensor mode becomes dominant. In the absence of anisotropic inertia, the gravitational wave equation (5.2.20) in the limit $q/a \ll H$ becomes

$$\ddot{\mathcal{D}}_{Nq}(t) + 3\frac{\dot{a}}{a}\dot{\mathcal{D}}_{Nq}(t) = 0$$

[472] The same conclusion follows if we assume that the probability distribution is invariant under CP, where C is the charge conjugation operator. Neither P nor CP is an exact symmetry of nature, but CP-invariance is a better approximation than P -invariance. The product CPT (where T is the reversal of the direction of time) is an exact symmetry in any quantum field theory, but T -invariance is broken by the expansion of the universe.

This has two obvious solutions:

$$\mathcal{D}_{1q}(t) = 1, \quad \mathcal{D}_{2q}(t) = \int_t^\infty \frac{dt'}{a^3(t')}$$

(The integral in the second solution converges, because $a(t) \propto t^{1/2}$ in the radiation-dominated era, and grows even faster in the matter-dominated and vacuum-dominated eras.) Since $\mathcal{D}_{2q}(t) \to 0$ at late times, for generic initial conditions the gravitational waves will eventually be dominated by the first solution. Thus in this case, to evaluate bilinear averages like (5.2.24), we only need to know $\mathcal{P}_{11}^{\rm grav}(q)$. Alternatively, we could take $\mathcal{D}_{1q}(t)$ to be a constant $\mathcal{D}_q^o$ (the o superscript denoting "outside the horizon"), chosen so that $\mathcal{P}_{11}^{\rm grav}(q) = 1$. In Section 5.4 we will see that similar remarks usually apply to the scalar modes.

10.5.3 5.3 Choosing a gauge

The equations derived in Section 5.1 have two unsatisfactory features. First, even with the simplifications introduced by decomposing the equations into

scalar, vector, and tensor modes, the equations for the scalar modes are still fearsomely complicated. Second, among the solutions of these equations are unphysical scalar and vector modes, corresponding to a mere coordinate transformation of the unperturbed Robertson-Walker metric and energy-momentum tensor. We can eliminate the second problem and ameliorate the first by fixing the coordinate system, adopting suitable conditions on the full perturbed metric and/or energy-momentum tensor. We will deal here with the coordinate-dependent perturbations $A(\mathbf{x}, t), B(\mathbf{x}, t)$, etc., but all of the results of this section could be applied just as well to the Fourier components $A_{nq}(t), B_{nq}(t)$, etc., by simply replacing each Laplacian with $-q^2$.

Consider a spacetime coordinate transformation

$$x^\mu \to x'^\mu = x^\mu + \epsilon^\mu(x) \tag{5.3.1}$$

with $\epsilon^\mu(x)$ small in the same sense that $h_{\mu\nu}, \delta\rho$, and other perturbations are small. Under this transformation, the metric tensor will be transformed to

$$g'_{\mu\nu}(x') = g_{\lambda\kappa}(x)\frac{\partial x^\lambda}{\partial x'^\mu}\frac{\partial x^\kappa}{\partial x'^\nu} \tag{5.3.2}$$

Instead of working with such transformations, which affect the coordinates and unperturbed fields as well as the perturbations to the fields, it is more convenient to work with so-called gauge transformations, which affect only the field perturbations. For this purpose, after making the coordinate transformation (5.3.1), we relabel coordinates by dropping the prime on the coordinate argument, and we attribute the whole change in $g_{\mu\nu}(x)$ to a change in the perturbation $h_{\mu\nu}(x)$. The field equations should thus be invariant under the gauge transformation $h_{\mu\nu}(x) \to h_{\mu\nu}(x) + \Delta h_{\mu\nu}(x)$, where

$$\Delta h_{\mu\nu}(x) \equiv g'_{\mu\nu}(x) - g_{\mu\nu}(x) \tag{5.3.3}$$

with the unperturbed Robertson-Walker metric $\bar{g}_{\mu\nu}(x)$ left unchanged, and corresponding gauge transformations of other perturbations. To first order in $\epsilon(x)$ and $h_{\mu\nu}(x)$, Eq. (5.3.3) is

[473] To obtain the final formula for $\Pi_{ij,k\ell}(\hat{q})$, one can use the conditions that $\Pi_{ij,k\ell}(\hat{q})$ is a tensor function of $\hat{q}$, symmetric in i and j and in k and ℓ, that $\Pi_{ij,k\ell}(\hat{q}) = \Pi^*_{k\ell,ij}(\hat{q})$, and that $\hat{q}_i\Pi_{ij,k\ell}(\hat{q}) = 0$, to show that $\Pi_{ij,k\ell}(\hat{q})$ is proportional to the quantity on the right-hand side of Eq. (5.2.25). The constant of proportionality can then be found by using Eq. (5.2.15), letting $\hat{q}$ be in the three-direction and setting i, j, k, and ℓ all equal to 1 .

$$\Delta h_{\mu\nu}(x) = g'_{\mu\nu}(x') - \frac{\partial g_{\mu\nu}(x)}{\partial x^\lambda}\epsilon^\lambda(x) - g_{\mu\nu}(x)$$
$$= -\bar{g}_{\lambda\mu}(x)\frac{\partial \epsilon^\lambda(x)}{\partial x^\nu} - \bar{g}_{\lambda\nu}(x)\frac{\partial \epsilon^\lambda(x)}{\partial x^\mu} - \frac{\partial \bar{g}_{\mu\nu}(x)}{\partial x^\lambda}\epsilon^\lambda(x), \tag{5.3.4}$$

or in more detail

$$\Delta h_{ij} = -\frac{\partial \epsilon_i}{\partial x^j} - \frac{\partial \epsilon_j}{\partial x^i} + 2a\dot{a}\delta_{ij}\epsilon_0 \tag{5.3.5}$$
$$\Delta h_{i0} = -\frac{\partial \epsilon_i}{\partial t} - \frac{\partial \epsilon_0}{\partial x^i} + 2\frac{\dot{a}}{a}\epsilon_i \tag{5.3.6}$$
$$\Delta h_{00} = -2\frac{\partial \epsilon_0}{\partial t} \tag{5.3.7}$$

with all quantities evaluated at the same spacetime coordinate point, and indices now raised and lowered with the Robertson-Walker metric, so that $\epsilon_0 = -\epsilon^0$ and $\epsilon_i = a^2\epsilon^i$. The field equations will be invariant only if the same gauge transformation is applied to all tensors, and in particular to the energy-momentum tensor, so that we must transform $\delta T_{\mu\nu}(x) \to \delta T_{\mu\nu}(x) + \Delta\delta T_{\mu\nu}(x)$, where $\Delta\delta T_{\mu\nu}$ is given by a formula [1] analogous to Eq. (5.3.4):

$$\Delta\delta T_{\mu\nu} = -\bar{T}_{\lambda\mu}(x)\frac{\partial \epsilon^\lambda(x)}{\partial x^\nu} - \bar{T}_{\lambda\nu}(x)\frac{\partial \epsilon^\lambda(x)}{\partial x^\mu} - \frac{\partial \bar{T}_{\mu\nu}(x)}{\partial x^\lambda}\epsilon^\lambda(x) \tag{5.3.8}$$

or in more detail

$$\Delta\delta T_{ij} = -\bar{p}\left(\frac{\partial \epsilon_i}{\partial x^j} + \frac{\partial \epsilon_j}{\partial x^i}\right) + \frac{\partial}{\partial t}\left(a^2\bar{p}\right)\delta_{ij}\epsilon_0 \tag{5.3.9}$$
$$\Delta\delta T_{i0} = -\bar{p}\frac{\partial \epsilon_i}{\partial t} + \bar{\rho}\frac{\partial \epsilon_0}{\partial x^i} + 2\bar{p}\frac{\dot{a}}{a}\epsilon_i \tag{5.3.10}$$
$$\Delta\delta T_{00} = 2\bar{\rho}\frac{\partial \epsilon_0}{\partial t} + \dot{\bar{\rho}}\epsilon_0 \tag{5.3.11}$$

Note that we use δ to signify a perturbation, while Δ here denotes the change in a perturbation associated with a gauge transformation.

In order to write these gauge transformations in terms of the scalar, vector, and tensor components introduced in Section 5.1, it is necessary to decompose the spatial part of ϵ^μ into the gradient of a spatial scalar plus a divergenceless vector:

$$\epsilon_i = \partial_i\epsilon^S + \epsilon_i^V, \quad \partial_i\epsilon_i^V = 0 \tag{5.3.12}$$

Then the transformations (5.3.5)-(5.3.7) and (5.3.9)-(5.3.11) give the gauge transformations of the metric perturbation components defined by

Eqs. (5.1.31)-(5.1.33):

$$\Delta A = \frac{2\dot{a}}{a}\epsilon_0, \quad \Delta B = -\frac{2}{a^2}\epsilon^S,$$
$$\Delta C_i = -\frac{1}{a^2}\epsilon_i^V, \quad \Delta D_{ij} = 0, \quad \Delta E = 2\dot{\epsilon}_0 \tag{5.3.13}$$
$$\Delta F = \frac{1}{a}\left(-\epsilon_0 - \dot{\epsilon}^S + \frac{2\dot{a}}{a}\epsilon^S\right), \quad \Delta G_i = \frac{1}{a}\left(-\dot{\epsilon}_i^V + \frac{2\dot{a}}{a}\epsilon_i^V\right)$$

[471] The right-hand sides of Eqs. (5.3.4) and (5.3.8) are known as the Lie derivatives of the metric and energy-momentum tensor, respectively. For a discussion of Lie derivatives, see G&C, Secs. 10.9 and 12.3.

and of the perturbations (5.1.39)-(5.1.41) to the pressure, energy density, and velocity potential

$$\Delta\delta p = \dot{\bar{p}}\epsilon_0, \quad \Delta\delta\rho = \dot{\bar{\rho}}\epsilon_0, \quad \Delta\delta u = -\epsilon_0 \tag{5.3.14}$$

The other ingredients of the energy-momentum tensor are gauge invariant:

$$\Delta\pi^S = \Delta\pi_i^V = \Delta\pi_{ij}^T = \Delta\delta u_i^V = 0 \tag{5.3.15}$$

Note in particular that the conditions $\pi^S = \pi_i^V = \pi_{ij}^T = 0$ for a perfect fluid and the condition $\delta u_i^V = 0$ for potential (i.e., irrotational) flow are gauge invariant.

For the field equations to be gauge-invariant, similar transformations must of course be made on any other ingredients in these equations. For instance, any four-scalar $s(x)$ for which $s'(x') = s(x)$ under arbitrary fourdimensional coordinate transformations would undergo the change $\Delta\delta s(x) \equiv s'(x) - s(x) = s'(x) - s'(x')$, which to first order in perturbations is

$$\Delta\delta s(x) = s(x) - s(x') = -\frac{\partial\bar{s}(t)}{\partial x^\mu}\epsilon^\mu(x) = \dot{\bar{s}}(t)\epsilon_0 \tag{5.3.16}$$

This applies for instance to the number density n or a scalar field φ. For a perfect fluid both p and ρ are defined as scalars, and the gauge transformations in Eq. (5.3.14) of δp and $\delta\rho$ are other special cases of Eq. (5.3.16). Likewise, for a perfect fluid the gauge transformation in Eq. (5.3.14) of δu can be derived from the vector transformation law of u_μ. Because the gauge transformation properties of $\delta\rho, \delta p, \delta u$, etc. do not depend on the conservation laws, Eqs. (5.3.14) and (5.3.15) apply to each individual constituent of the universe in any case in which the energy-momentum tensor is a sum of terms for different constituents of the universe, even if these individual terms are not separately conserved.

We can eliminate the gauge degrees of freedom either by working only with gauge-invariant quantities, [2] or by choosing a gauge. The tensor

quantities π_{ij}^T and D_{ij} appearing in Eq. (5.1.53) are already gauge invariant, and no gauge-fixing is necessary or possible. For the vector quantities π_i^V, δu_i^V, C_i and G_i, we can write Eqs. (5.1.50)-(5.1.52) in terms of the gaugeinvariant quantities π_i^V, δu_i^V and $\tilde{G}_i \equiv G_i - a\dot{C}_i$, or we can fix a gauge for these quantities by choosing ϵ_i^V so that either C_i or G_i vanishes. (Note that $\tilde{G}_i$ is the vector field that we showed in Section 5.1 to decay as $1/a^2$ in the absence of anisotropic inertia.) For the scalar perturbations it is somewhat more convenient to fix a gauge. There are several frequently considered possibilities.

A. Newtonian gauge

Here we choose ϵ^S so that $B = 0$, and then choose ϵ_0 so that $F = 0$. Both choices are unique, so that after choosing Newtonian gauge, there is no remaining freedom to make gauge transformations. It is conventional to write E and A in this gauge as

$$E \equiv 2\Phi, \quad A \equiv -2\Psi, \tag{5.3.17}$$

so that (now considering only scalar perturbations) the perturbed metric has components

$$g_{00} = -1 - 2\Phi, \quad g_{0i} = 0, \quad g_{ij} = a^2\delta_{ij}[1 - 2\Psi]. \tag{5.3.18}$$

The gravitational field equations (5.1.44)-(5.1.47) then take the form

[472] J. M. Bardeen, Phys. Rev. D 22, 1882 (1980).

$$-4\pi G a^2\left[\delta\rho - \delta p - \nabla^2\pi^S\right] = a\dot{a}\dot{\Phi} + \left(4\dot{a}^2 + 2a\ddot{a}\right)\Phi - \nabla^2\Psi + a^2\ddot{\Psi} + 6a\dot{a}\dot{\Psi} \tag{5.3.19}$$

$$-8\pi G a^2\partial_i\partial_j\pi^S = \partial_i\partial_j[\Phi - \Psi] \tag{5.3.20}$$

$$4\pi G a(\bar{\rho} + \bar{p})\partial_i\delta u = -\dot{a}\partial_i\Phi - a\partial_i\dot{\Psi} \tag{5.3.21}$$

$$4\pi G\left(\delta\rho + 3\delta p + \nabla^2\pi^S\right) = +\frac{1}{a^2}\nabla^2\Phi + \frac{3\dot{a}}{a}\dot{\Phi} + 3\ddot{\Psi} + \frac{6\dot{a}}{a}\dot{\Psi} + \frac{6\ddot{a}}{a}\Phi \tag{5.3.22}$$

and the equations (5.1.48) and (5.1.49) of momentum and energy conservation become (aside from modes of zero wave number)

$$\delta p + \nabla^2\pi^S + \partial_0[(\bar{\rho} + \bar{p})\delta u] + \frac{3\dot{a}}{a}(\bar{\rho} + \bar{p})\delta u + (\bar{\rho} + \bar{p})\Phi = 0 \tag{5.3.23}$$

$$\delta\dot{\rho} + \frac{3\dot{a}}{a}(\delta\rho + \delta p) + \nabla^2\left[a^{-2}(\bar{\rho} + \bar{p})\delta u + \frac{\dot{a}}{a}\pi^S\right] - 3(\bar{\rho} + \bar{p})\dot{\Psi} = 0 \tag{5.3.24}$$

In particular, Eq. (5.3.20) shows that Φ and Ψ are not physically independent fields; they differ only by a term arising from the anisotropic part of the stress tensor, and in particular they are equal for a perfect fluid, for which $\pi^S = 0$. The perturbation to the number density of a species of particle whose total number is conserved will satisfy relation (5.1.55), which in Newtonian gauge reads

$$\frac{\partial}{\partial t}\left(\frac{\delta n}{\bar{n}}\right) + \frac{1}{a^2}\nabla^2\delta u - 3\dot{\Psi} = 0 \tag{5.3.25}$$

Given an equation of state for p as a function of ρ (or, if p depends also on other quantities like n, then given also field equations for those quantities) and given also a formula for π^S as a linear combination of the other perturbations (such as for instance the formula $\pi^S = 0$ for a perfect fluid) we can regard Eqs. (5.3.21), (5.3.23), and (5.3.24) (and, where needed, Eq. (5.3.25)) as equations of motion for $\Psi, \delta u$, and $\delta\rho$, respectively, with Φ given in terms of Ψ by Eq. (5.3.20). The remaining equations provide a constraint on the solution of this coupled system of equations. By subtracting $3/a^2$ times Eq. (5.3.19) from Eq. (5.3.22) and then using Eqs. (5.3.20) and (5.3.21) to eliminate π^S and Φ, we find that [2]

$$a^3\delta\rho - 3Ha^3(\bar{\rho} + \bar{p})\delta u - \left(\frac{a}{4\pi G}\right)\nabla^2\Psi = 0 \tag{5.3.26}$$

This is a constraint rather than an equation of motion, because the equations of motion (5.3.21), (5.3.23), and (5.3.24) imply that the left-hand side of Eq. (5.3.26) is time-independent, so that Eq. (5.3.26) only has to be imposed as an initial condition.

B. Synchronous gauge

Here we choose ϵ_0 so that $E = 0$, and then choose ϵ^S so that again $F = 0$. Considering only scalar perturbations, the complete perturbed metric is then

$$g_{00} = -1, \quad g_{0i} = 0, \quad g_{ij} = a^2\left[(1 + A)\delta_{ij} + \frac{\partial^2 B}{\partial x^i\partial x^j}\right] \tag{5.3.27}$$

In this gauge, the Einstein field equations (5.1.44)-(5.1.47) take the form

$$-4\pi G a^2\left[\delta\rho-\delta p-\nabla^2\pi^S\right]=\frac{1}{2}\nabla^2 A-\frac{1}{2}a^2\ddot{A}-3a\dot{a}\dot{A}-\frac{1}{2}a\dot{a}\nabla^2\dot{B} \tag{5.3.28}$$

$$-16\pi G a^2\pi^S=A-a^2\ddot{B}-3a\dot{a}\dot{B} \tag{5.3.29}$$

$$8\pi G a(\bar{\rho}+\bar{p})\delta u=a\dot{A} \tag{5.3.30}$$

$$-4\pi G\left(\delta\rho+3\delta p+\nabla^2\pi^S\right)=\frac{3}{2}\ddot{A}+\frac{3\dot{a}}{a}\dot{A}+\frac{1}{2}\nabla^2\ddot{B}+\frac{\dot{a}}{a}\nabla^2\dot{B} \tag{5.3.31}$$

and the equations (5.1.48) and (5.1.49) of momentum and energy conservation read

$$\delta p+\nabla^2\pi^S+\partial_0[(\bar{\rho}+\bar{p})\delta u]+\frac{3\dot{a}}{a}(\bar{\rho}+\bar{p})\delta u=0, \tag{5.3.32}$$

$$\delta\dot{\rho}+\frac{3\dot{a}}{a}(\delta\rho+\delta p)+\nabla^2\left[a^{-2}(\bar{\rho}+\bar{p})\delta u+\frac{\dot{a}}{a}\pi^S\right]+\frac{1}{2}(\bar{\rho}+\bar{p})\partial_0\left[3A+\nabla^2 B\right]=0. \tag{5.3.33}$$

Note that in this gauge the equation of momentum conservation, which furnishes the equation of motion (the Navier-Stokes equation) for an imperfect fluid, does not depend at all on the perturbed metric, while the equation of energy conservation may be written as

$$\delta\dot{\rho}+\frac{3\dot{a}}{a}(\delta\rho+\delta p)+\nabla^2\left[a^{-2}(\bar{\rho}+\bar{p})\delta u+\frac{\dot{a}}{a}\pi^S\right]+(\bar{\rho}+\bar{p})\psi=0 \tag{5.3.34}$$

where

$$\psi\equiv\frac{1}{2}\left[3\dot{A}+\nabla^2\dot{B}\right]=\frac{\partial}{\partial t}\left(\frac{h_{ii}}{2a^2}\right). \tag{5.3.35}$$

We need A and B separately to calculate the motion of individual particles, but the effect of gravitation on a perfect or imperfect fluid is entirely governed by the quantity ψ. Inspection of the field equation (5.3.31) shows that it provides a differential equation for just this combination of scalar fields:

$$-4\pi G a^2\left(\delta\rho+3\delta p+\nabla^2\pi^S\right)=\frac{\partial}{\partial t}\left(a^2\psi\right) \tag{5.3.36}$$

Also, in synchronous gauge the equation (5.1.55) for particle conservation takes the form

$$\frac{\partial}{\partial t}\left(\frac{\delta n}{\bar{n}}\right)+a^{-2}\nabla^2\delta u+\psi=0. \tag{5.3.37}$$

Given an equation of state for p as a function of ρ (and perhaps n) and a formula expressing π^S as a linear combination of the other scalar perturbations we can use Eqs. (5.3.32), (5.3.34), (5.3.36) (and perhaps

Eq. (5.3.37)) to find solutions for the three independent perturbations δu, $\delta\rho$, and ψ, respectively. The left-over equations (5.3.28)-(5.3.30) are not needed, for a reason given in Section 5.1: the full set of equations (5.3.28)(5.3.33) are not independent, because the equations of energy and momentum conservation can be derived from the Einstein field equations.

If we need to know A and B separately we can find them from ψ and $\delta\rho$. By adding 3 times Eq. (5.3.28), plus 1/2 the Laplacian of Eq. (5.3.29), plus a^2 times Eq. (5.3.31), we obtain the simple relation

$$\nabla^2 A = -8\pi G a^2 \delta\rho + 2Ha^2\psi, \tag{5.3.38}$$

where as usual $H \equiv \dot{a}/a$. After A is found in this way, we can find B from A and ψ by solving Eq. (5.3.35).

Synchronous gauge was widely used in early calculations of the evolution of perturbations in cosmology, starting with the ground-breaking work of Lifshitz in 1946.[3] In the 1980s synchronous gauge became unpopular, because of a feature emphasized by Bardeen: [2] even after we impose the conditions $E = F = 0$, we are left with a residual gauge invariance. We can see from Eq. (5.3.13) that E and F are not affected by a gauge transformation with

$$\epsilon_0(\mathbf{x}, t) = -\tau(\mathbf{x}), \quad \epsilon^S(\mathbf{x}, t) = a^2(t)\tau(\mathbf{x}) \int a^{-2}(t)dt \tag{5.3.39}$$

where $\tau(\mathbf{x})$ is an arbitrary function of $\mathbf{x}$, but not of t. But under this transformation A and B do change

$$\Delta A = -\frac{2\dot{a}\tau}{a}, \quad \Delta B = -2\tau \int a^{-2}(t)dt \tag{5.3.40}$$

In particular, the combination (5.3.35) undergoes the change

$$\Delta\psi = -3\tau \frac{d}{dt}\left(\frac{\dot{a}}{a}\right) - a^{-2}\nabla^2\tau. \tag{5.3.41}$$

Also, the changes in the perturbations to the energy density, pressure, and velocity potential are given by Eqs. (5.3.14) and (5.3.39) as

$$\Delta\delta p = -\dot{\bar{p}}\tau, \quad \Delta\delta\rho = -\dot{\bar{\rho}}\tau, \quad \Delta\delta u = \tau, \tag{5.3.42}$$

while π^S is invariant. The same transformation rules apply for any one of the individual constituents of the universe. Any scalar perturbation δs such as the number density perturbation δn or a scalar field perturbation

$\delta\varphi$ undergoes a change like that of the pressure and density perturbations:

$$\Delta\delta s = -\dot{\bar{S}}\tau \tag{5.3.43}$$

The reader can check that all of the equations (5.3.28)-(5.3.34) and (5.3.37) are invariant under these residual gauge transformations. This being the case, for any solution $\psi, \delta p, \delta\rho, \delta u, \delta n$, etc. of these equations there will be another solution $\psi + \Delta\psi, \delta p + \Delta\delta p, \delta\rho + \Delta\delta\rho, \delta u + \Delta\delta u, n + \Delta\delta n$, etc., and since the field equations are linear, this means that $\Delta\psi, \Delta\delta p, \Delta\delta\rho, \Delta\delta u, \Delta\delta n$, etc., is also a solution. (For this solution there is no scalar anisotropic inertia, because π^S is gauge invariant.) This is a nuisance, because in finding solutions of the field equations we keep having to check that our solution represents a physical disturbance, not a mere change of gauge.

However, this is not a problem if we can remove the residual gauge symmetry in any natural way. This is the case if the universe contains a fluid (such as cold dark matter) whose individual particles are moving at speeds much less than that of light. In this case, the space-space components T_D^{ij} of the energy-momentum tensor for that fluid are negligible, so we can take $\bar{p}_D = 0$, $\bar{\rho}_D \propto a^{-3}$, and $\delta p_D = \pi_D^S = 0$ in the equation (5.3.32) of momentum conservation for this fluid, which tells us then that δu_D is time-independent. According to Eq. (5.3.42) (which

[473] E. Lifshitz, J. Phys. USSR 10, 116 (1946). Also see G&C, Sec. 15.10.

applies separately to each constituent of the universe) a time-independent velocity potential δu_D can always be removed in synchronous gauge by a residual gauge transformation, with $\tau = -\delta u_D$. There is then no longer any ambiguity in the choice of gauge. These features make synchronous gauge particularly convenient in dealing with the later stages of cosmological evolution, when cold dark matter plays a prominent role.

C. Newtonian/synchronous conversion

We will find it convenient to do calculations using Newtonian gauge in some eras, and synchronous gauge in others. To connect results for different eras, we need to be able to convert them from one gauge to another. [4]

Suppose first that we begin in Newtonian gauge, and make an infinitesimal coordinate transformation $x^\mu \to x^\mu + \epsilon^\mu$, with $\epsilon_i = \partial_i \epsilon^S$ so as not to induce vector perturbations. According to Eqs. (5.3.13) and (5.3.17), in order to give E the synchronous gauge value $E = 0$, we need

$$\dot{\epsilon}_0 = -\Phi \tag{5.3.44}$$

Then according to Eq. (5.3.13), to keep $F = 0$ we need

$$\frac{\partial}{\partial t}\left(\frac{\epsilon^S}{a^2}\right) = -\frac{\epsilon_0}{a^2}. \tag{5.3.45}$$

The A and B components of the spatial metric in synchronous gauge are given by Eqs. (5.3.13) and (5.3.17):

$$A = -2\Psi + 2H\epsilon_0, \quad B = -\frac{2}{a^2}\epsilon^S, \tag{5.3.46}$$

where $H \equiv \dot{a}/a$. In particular, the field ψ is

$$\psi = -3\dot{\Psi} + 3\frac{\partial}{\partial t}(H\epsilon_0) + \frac{\nabla^2 \epsilon_0}{a^2}. \tag{5.3.47}$$

Also, Eq. (5.3.14) allows us to calculate the synchronous gauge pressure perturbation δp^s, energy density perturbation $\delta\rho^s$, and velocity potential δu^s from the corresponding quantities $\delta p, \delta\rho$, and δu in Newtonian gauge:

$$\delta p^s = \delta p + \epsilon_0 \dot{\bar{p}}, \quad \delta\rho^s = \delta\rho + \epsilon_0 \dot{\bar{\rho}}, \quad \delta u^s = \delta u - \epsilon_0. \tag{5.3.48}$$

Given Φ we can calculate ϵ_0 from Eq. (5.3.44), and then given Ψ we can obtain ψ from Eq. (5.3.47) and the synchronous gauge pressure, energy density, and velocity potential perturbations from Eq. (5.3.48). The quantity ϵ_0 is determined by Eq. (5.3.44) only up to a time-independent function of position, so the values of the synchronous gauge quantities A, B, ψ, $\tilde{p}, \tilde{\rho}$, and $\delta\tilde{u}$ are only determined up to a residual gauge transformation (5.3.40)-(5.3.42).

Next suppose that we begin in synchronous gauge, with metric fields A and B, and want to convert to Newtonian gauge. According to Eq. (5.3.13), to make g_{ij} proportional to δ_{ij} we need to take

$$\epsilon^S = a^2 B/2. \tag{5.3.49}$$

Then to keep $g_{i0} = 0$, Eq. (5.3.13) tells us that we must take

$$\epsilon_0 = -a^2 \dot{B}/2 \tag{5.3.50}$$

[474] See, for instance, C.-P. Ma and E. Bertschinger, Astrophys. J. 455, 7 (1995).

Using Eq. (5.3.13) again together with the definitions (5.3.18), we have then

$$\Phi = \dot{\epsilon}_0 = -\frac{1}{2}\frac{\partial}{\partial t}\left(a^2\dot{B}\right) \tag{5.3.51}$$

$$\Psi = -\frac{1}{2}\left(A + \frac{2\dot{a}}{a}\epsilon_0\right) = -\frac{1}{2}A + \frac{a\dot{a}}{2}\dot{B}. \tag{5.3.52}$$

In contrast to the previous case, Eqs. (5.3.51) and (5.3.52) give Φ and Ψ uniquely. Not only that - it is easy to see that the results for Φ and Ψ are unaffected if the A and B with which we start are subjected to the residual gauge transformation (5.3.40).

D. Other gauges

In choosing a gauge, it is not necessary to impose conditions only on the scalar fields appearing in the metric tensor. [5] Instead, some of the gauge conditions can impose constraints on the scalars appearing in the energymomentum tensor. For instance, in co-moving gauge we choose ϵ_0 so that $\delta u = 0$ (which for scalar perturbations makes the velocity perturbation δu^i vanish). Where the only "matter" is a single scalar field, as in popular theories of inflation, this means that the time coordinate is defined so that at any given time the scalar field equals its unperturbed value, with all perturbations relegated to components of the metric. [6] In the constant density gauge we choose ϵ_0 so that $\delta\rho = 0$. In either case, after fixing ϵ_0 we can make F vanish with a suitable choice of ϵ^S, so that the scalar perturbations still have $g_{i0} = 0$. Note that although this procedure fixes ϵ_0, it only fixes ϵ^S up to terms of the form $a^2(t)\tau(\mathbf{x})$, so these gauges share the drawback of synchronous gauge, of leaving a residual gauge symmetry.

10.5.4 5.4 Conservation outside the horizon

The perturbations that concern us are believed to have originated in quantum fluctuations during an era of inflation in the very early universe, discussed in Chapter 10. That early time and the much later time when these perturbations are observed are separated by a time interval in which the equations governing perturbations are not well known. For instance, at the end of the era of inflation there was a time of so-called reheating, during which the energy of the vacuum was transferred to ordinary matter and radiation, but we have no idea what particles were first created during reheating or how the energy transfer took place. Later, there was presumably a time when some particles effectively stopped interacting with the rest of matter and radiation, and became what is now observed as cold dark matter, but we can only guess when this was and how it happened.

The only reason that we are able to use inflationary theories to make any predictions at all about observable perturbations despite these uncertainties is that the wavelengths of the perturbations that concern us were outside the horizon during a period that extends from well before the end of inflation until relatively near the present (and hence includes the times of reheating and cold dark matter decoupling), in the sense that the physical wave number q/a during this period was much less than the expansion rate $H \equiv \dot{a}/a$. During inflation q/a decreases with time more-or-less exponentially while H is roughly constant, so all perturbations originally had $q/a \gg H$ but except for very short wavelengths eventually have $q/a \ll H$. Then during the radiation and matter-dominated eras a increases like $t^{1/2}$ or $t^{2/3}$ while H falls like

[475] A great variety of different gauges are described by H. Kodama and M. Sasaki, Prog. Theor. Phys. Suppl. 78, 1 (1984).

[6] For instance, this gauge was used in calculations of non-Gaussian corrections to cosmological correlations by J. Maldacena, J. High Energy Phys. 0305, 013 (2003) [astro-ph/0210603].

$1/t$, so except for the longest wavelengths the perturbations that had $q/a \ll H$ at the end of inflation eventually again have $q/a \gg H$. During the intervening period when perturbations were outside the horizon, the scalar and tensor fluctuations were subject to certain conservation laws, that allow us to connect the distant past to the relatively recent past.

As we shall see below, for scalar modes it is the quantity defined in Newtonian gauge by [1]

$$\mathcal{R}_q \equiv -\Psi_q + H\delta u_q \tag{5.4.1}$$

that in certain circumstances is conserved outside the horizon. (Recall that δu is the velocity potential for the total energy-momentum tensor.) Equivalently, there is another quantity [2]

$$\zeta_q \equiv -\Psi_q + \frac{\delta\rho_q}{3(\bar{\rho}+\bar{p})} \tag{5.4.2}$$

that according to the constraint (5.3.26) is related to $\mathcal{R}_q$ by

$$\zeta_q = \mathcal{R}_q - \frac{q^2\Psi_q}{12\pi G(\bar{\rho}+\bar{p})a^2} \tag{5.4.3}$$

The difference is of relative order $(q/aH)^2$, so for $q/a \ll H$ the quantity ζ_q is conserved outside the horizon if $\mathcal{R}_q$ is.

The importance of $\mathcal{R}_q$ in the work of ref. 1 is that (as will be shown below) it is conserved outside the horizon in inflation driven by a single scalar

field. But this is not enough for our purposes. After inflation, the universe becomes filled with a number of types of matter and radiation, and $\mathcal{R}_q$ is not necessarily conserved outside the horizon under these circumstances. What we need is a theorem that says that, whatever the constituents of the universe, there is always a solution of the field equations for which $\mathcal{R}_q$ and ζ_q are conserved outside the horizon. Such a solution is called "adiabatic," for reasons explained later. If cosmological fluctuations are described by such a solution during inflation, then they will continue to be described by such a solution and $\mathcal{R}_q$ and ζ_q will remain constant as long as the perturbation is outside the horizon, because this is a solution under all circumstances.

This result (and a corresponding result for tensor modes) is contained in the following theorem: [3] Whatever the contents of the universe, there are two independent adiabatic physical scalar solutions of the Newtonian gauge field equations for which the quantity $\mathcal{R}_q$ is time-independent in the limit $q/a \ll H$, and there is one tensor mode for which the tensor amplitude $\mathcal{D}_q$ is time-independent in the limit $q/a \ll H$. In this limit one of the scalar modes has $\mathcal{R}_q \neq 0$; the scalar metric components are

$$\Phi_q(t) = \Psi_q(t) = \mathcal{R}_q\left[-1 + \frac{H(t)}{a(t)}\int_{\mathcal{T}}^{t} a\left(t'\right)dt'\right] \tag{5.4.4}$$

the perturbation to any four-scalar $s(x)$ (such as the energy density, pressure, inflaton field, etc.) is given by

$$\delta s_q(t) = -\frac{\mathcal{R}_q\dot{\bar{s}}(t)}{a(t)}\int_{\mathcal{T}}^{t} a\left(t'\right)dt' \tag{5.4.5}$$

and the perturbation to the velocity potential is

[471] The constancy of $\mathcal{R}$ was noted in various special cases by J. M. Bardeen, Phys. Rev. D 22, 1882 (1980); D. H. Lyth, Phys. Rev. D 31, 1792 (1985). For reviews, see J. Bardeen, in Cosmology and Particle Physics, eds. Li-zhi Fang and A. Zee (Gordon & Breach, New York, 1988); A. R. Liddle and D. H. Lyth, Cosmological Inflation and Large Scale Structure (Cambridge University Press, Cambridge, UK, 2000).

[2] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, Phys. Rev. D 28, 679 (1983). This quantity was re-introduced by D. Wands, K. A. Malik, D. H. Lyth, and A. R. Liddle, Phys. Rev. D 62, 043527 (2000).

$$\delta u_q(t) = \frac{\mathcal{R}_q}{a(t)} \int_{\mathcal{T}}^{t} a\,(t')\,dt' \tag{5.4.6}$$

where $\mathcal{T}$ is an arbitrary initial time, the same in all integrals. The other scalar mode has $\mathcal{R}_q = 0$, and

$$\Phi_q(t) = \Psi_q(t) = \frac{\mathcal{C}_q H(t)}{a(t)}, \quad \delta s_q(t) = -\frac{\mathcal{C}_q \dot{\bar{s}}(t)}{a(t)}, \quad \delta u_q(t) = \frac{\mathcal{C}_q}{a(t)} \tag{5.4.7}$$

where $\mathcal{C}_q$ is time-independent. For $q/a \ll H$, the anisotropic inertia components π_q^S, π_{iq}^V, and π_{ijq}^T vanish in both adiabatic scalar modes and in the conserved tensor mode, even when some mean free times are comparable

with the Hubble time. These are physical solutions for scalar as well as tensor modes, because the choice of Newtonian gauge leaves no remaining gauge freedom.

The proof of the theorem is based on the observation that in the special case of a spatially homogeneous universe, the coordinate space Newtonian gauge field equations and dynamical equations for matter and radiation (as well as the condition of spatial homogeneity) are invariant under coordinate transformations that are not symmetries of the unperturbed metric. [4] In Newtonian gauge, general first-order spatially homogeneous scalar and tensor perturbations to the metric take the form

$$h_{00} = -2\Phi(t), \quad h_{i0} = 0, \quad h_{ij} = -2\delta_{ij}a^2(t)\Psi(t) + a^2(t)D_{ij}(t)$$

with D_{ij} subject to the condition that $D_{ii} = 0$. Spatial homogeneity also requires that all pressures, densities, velocity potentials, etc. are functions only of time. The Newtonian gauge field equations for these spatially homogeneous perturbations are necessarily invariant under those gauge transformations of the form (5.3.4)-(5.3.7), (5.3.14)-(5.3.16) that preserve the conditions for Newtonian gauge and spatial homogeneity. Eq. (5.3.7) shows that in order for h_{00} to remain spatially homogeneous, ϵ_0 must be of the form

$$\epsilon_0(\mathbf{x}, t) = \epsilon(t) + \chi(\mathbf{x}),$$

so that

$$\Delta\Phi = \dot{\epsilon}$$

Eq. (5.3.6) then shows that in order for the h_{i0} to remain equal to zero, ϵ_i must have the form

$$\epsilon_i(\mathbf{x}, t) = a^2(t)f_i(\mathbf{x}) - a^2(t)\frac{\partial\chi(\mathbf{x})}{\partial x^i}\int \frac{dt}{a^2(t)}$$

Eq. (5.3.5) then shows that

$$\Delta h_{ij} = -a^2\left(\frac{\partial f_i}{\partial x^j} + \frac{\partial f_j}{\partial x^i}\right) + 2\delta_{ij}a\dot{a}[\epsilon + \chi] - 2\frac{\partial^2\chi}{\partial x^i \partial x^j}\int \frac{dt}{a^2}$$

In order not to introduce any $\mathbf{x}$-dependence in h_{ij}, we must take χ constant, in which case by shifting it into ϵ it can be taken to be zero, and we must also take f_i to have the form

[473] S. Weinberg, Phys. Rev. D 67, 123504 (2003) [astro-ph/0302326]. Also see the appendix of S. Weinberg, Phys. Rev. D 69, 023503 (2004) [astro-ph/0306304].

$f_i(\mathbf{x}) = \omega_{ij}x^j$, with ω_{ij} a constant matrix. (An $\mathbf{x}$-independent term in f_i would have no effect on the metric or anything

else, and so can be ignored here.) Then

$$\Delta h_{ij} = -a^2\left[\omega_{ij} + \omega_{ji}\right] + 2\delta_{ij}a\dot{a}\epsilon.$$

We compare this with Eq. (5.1.33) (with the Newtonian gauge conditions $B = C_i = 0$ and $A = -2\Psi$), which gives

$$\Delta h_{ij} = -2a^2\delta_{ij}\Delta\Psi + a^2\Delta D_{ij}$$

Matching the terms in Δh_{ij} that are either proportional to δ_{ij} or traceless gives

$$\Delta\Psi = \frac{1}{3}\omega_{ii} - H\epsilon,$$
$$\Delta D_{ij} = -\omega_{ij} - \omega_{ji} + \frac{2}{3}\delta_{ij}\omega_{kk}.$$

(Note that an antisymmetric term in ω_{ij} would have no effect, because the unperturbed metric is invariant under three-dimensional rotations.) The corresponding gauge transformations of the quantities appearing in the energy-momentum tensor and of general scalars are given by Eqs. (5.3.14)(5.3.16). Since $\{h_{\mu\nu}, T_{\mu\nu}\}$ and $\{h_{\mu\nu} + \Delta h_{\mu\nu}, T_{\mu\nu} + \Delta T_{\mu\nu}\}$ are both solutions of the field equations and conservation equations, their difference must also be such a solution. We conclude that there is always a spatially homogeneous solution of the Newtonian gauge field and conservation equations, with scalar perturbations

$$\Psi = H\epsilon - \omega_{ii}/3, \quad \Phi = -\dot{\epsilon} \tag{5.4.8}$$
$$\delta p = -\dot{\bar{p}}\epsilon, \quad \delta\rho = -\dot{\bar{\rho}}\epsilon, \quad \delta u = \epsilon, \quad \pi^S = 0 \tag{5.4.9}$$

and more generally for any four-scalar s

$$\delta s = -\dot{\bar{S}}\epsilon. \tag{5.4.10}$$

(The reader can check that when we drop all spatial gradients, then the perturbations (5.4.8)-(5.4.10) satisfy the Newtonian gauge field equations (5.3.19)-(5.3.22), the conservation laws (5.3.23)-(5.3.24) and (5.3.25), and the constraint (5.3.26).) There is also a spatially homogeneous solution with a tensor perturbation

$$D_{ij} \propto \omega_{ij} - \frac{1}{3}\delta_{ij}\omega_{kk}, \quad \pi^T_{ij} = 0. \tag{5.4.11}$$

(This is obviously a solution of Eq. (5.1.53), but it includes the not so obvious information that the equations that determine π^T_{ij} necessarily give $\pi^T_{ij} = 0$ for D_{ij} constant, even if some particle collision rates become comparable

with H.) Equivalently, Eqs. (5.4.8)-(5.4.11) are solutions for the Fourier transforms of the perturbations with zero wave number.

So far, ϵ is an arbitrary function of time, and ω_{ij} is an unrelated arbitrary constant 4×4 matrix. But these are just gauge modes for zero wave number. On the other hand, if they can be extended to non-zero wave number they become physical modes, since the choice of

[474] In this respect, the theorem proved here is similar to the Goldstone theorem of quantum field theory; see J. Goldstone, Nuovo Cimento 9, 154 (1961); J. Goldstone, A. Salam, and S. Weinberg, Phys. Rev. 127, 965 (1962). The modes for which $\mathcal{R}_q$ or $\mathcal{D}_q$ are constant outside the horizon take the place here of the Goldstone bosons that become free particles for long wavelength.

Newtonian gauge leaves no residual gauge symmetries except for zero wave number. For the tensor modes there is no problem; in this case there are no field equations that disappear for zero wave number, so the solution with D_{ij} timeindependent automatically has an extension to a physical mode for non-zero wave number. But matters are more complicated for the scalar modes.

For the scalar modes the field equation (5.3.20) disappears in the limit of zero wave number, so to get a physical mode we must impose on the perturbations the condition

$$\Phi = \Psi \tag{5.4.12}$$

(The condition $\delta u = \epsilon$ that is required for Eq. (5.3.21) to be satisfied for $q \neq 0$ is already satisfied, according to Eq. (5.4.9).) Inserting Eq. (5.4.8) in (5.4.12) gives a differential equation for ϵ :

$$\dot{\epsilon} = -H\epsilon + \omega_{kk}/3 \tag{5.4.13}$$

Also, Eq. (5.4.8) for Ψ and Eq. (5.4.9) for δu give the quantity $\mathcal{R}$ defined by Eq. (5.4.1) the time-independent value

$$\mathcal{R} = \omega_{kk}/3 \tag{5.4.14}$$

There is a general solution of Eq. (5.4.13) for $\epsilon(t)$ with $\mathcal{R} \neq 0$:

$$\epsilon(t) = \frac{\mathcal{R}}{a(t)} \int_{\mathcal{T}}^{t} a\left(t'\right) dt' \tag{5.4.15}$$

with $\mathcal{T}$ arbitrary. Using Eq. (5.4.15) in Eq. (5.4.8) gives the explicit solution (5.4.4) for large wavelengths

$$\Psi = \Phi = \mathcal{R}\left[-1 + \frac{H(t)}{a(t)} \int_{\mathcal{T}}^{t} a\left(t'\right) dt'\right] \tag{5.4.16}$$

Eq. (5.4.9) gives

$$\frac{\delta p}{\dot{\bar{p}}} = \frac{\delta \rho}{\dot{\bar{\rho}}} = -\delta u = -\frac{\mathcal{R}}{a(t)} \int_{\mathcal{T}}^{t} a\left(t'\right) dt' \tag{5.4.17}$$

and more generally for any four-scalar Eq. (5.4.10) gives

$$\frac{\delta s}{\dot{\bar{s}}} = -\frac{\mathcal{R}}{a(t)} \int_{\mathcal{T}}^{t} a\left(t'\right) dt' \tag{5.4.18}$$

Because for any $\mathcal{T}$, Eq. (5.4.15) is a solution of Eq. (5.4.13) with the same value of $\mathcal{R}$, the difference of two of these solutions with different values of $\mathcal{T}$ is also a solution, but with $\mathcal{R} = 0$. For this solution

$$\epsilon(t) = \frac{\mathcal{C}}{a(t)} \tag{5.4.19}$$

with $\mathcal{C}$ another constant, and

$$\Psi = \Phi = \frac{\mathcal{C}H(t)}{a(t)} \tag{5.4.20}$$

$$\frac{\delta \rho}{\dot{\bar{\rho}}} = \frac{\delta p}{\dot{\bar{p}}} = -\delta u = \frac{\delta s}{\dot{\bar{s}}} = -\frac{\mathcal{C}}{a(t)} \tag{5.4.21}$$

and likewise for individual constituents of the universe. (Since $a(t)$ increases and $H(t)$ decreases with time, this is a decaying mode, which is usually assumed to play no significant role at late times.) Note in particular that these scalar modes have equal values for $\delta\rho_\alpha/\dot{\bar{\rho}}_\alpha$ for all individual constituents α of the universe, whether or not energy is separately conserved for these constituents. For this reason, such perturbations are called adiabatic. [5] For the same reason, any other solutions are called entropic. (Sometimes such other solutions are called zero-curvature modes, but this is misleading, because setting $\Phi = \Psi = 0$ does not usually give a solution at all.)

We have shown that for sufficiently small wave number (in practice, this means $q/a \ll H$) there are always two adiabatic physical solutions for scalar perturbations that take the form (5.4.16)-(5.4.18) and (5.4.20)-(5.4.21), with $\mathcal{R}$ and $\mathcal{C}$ arbitrary constants, and a physical solution for tensor perturbations that takes the form (5.4.11) with ω_{ij} an arbitrary constant matrix. Since the equations we have solved are homogeneous, it follows that there are also solutions of the same form for which $\mathcal{R}, \mathcal{C}$, and ω_{ij} are arbitrary timeindependent functions of q. This concludes the proof. [6]

This theorem shows that if the perturbations produced during inflation are actually in the modes found above at the end of inflation, then they stay in these modes as long as the perturbations remain outside the horizon, and in particular $\mathcal{R}_q$ and $\mathcal{D}_q$ remain constant, since these modes are solutions of the equations whatever the constituents of the universe may become. In particular, the wavelengths that will interest us are far outside the horizon during the era of reheating that is supposed to follow inflation, so if the scalar perturbations are adiabatic at the end of inflation, then reheating cannot generate entropic perturbations. [7]

The question that is left unanswered by this theorem is whether the scalar perturbations produced during inflation are actually in these modes at the end of inflation. This is often a matter of counting. We know that there are always two independent adiabatic solutions of the differential equations governing the scalar fluctuations, so if these equations have no more than two independent solutions, then any perturbations must be adiabatic. As we will see in Chapter 10, this counting shows that all solutions are adiabatic for inflation with a single inflaton field.

But it must not be thought that if observation of the cosmic microwave background reveals purely adiabatic perturbations at the time of last scattering, then the perturbations must have been adiabatic at the end of inflation. In a state of complete local thermal equilibrium in which all conserved quantum numbers vanish (such as is usually assumed to have existed at some early time in theories of cosmological baryonsynthesis or leptonsynthesis, like those discussed in Section 3.3) there are only two scalar degrees of freedom, the temperature and the gravitational potential $\Phi = \Psi$. They are governed by coupled first-order ordinary differential equations, so there are just two independent solutions, which must be adiabatic since there are always at least two adiabatic solutions. Thus whatever happens during inflation, if the universe subsequently spends sufficient time in a state of local thermal equilibrium with no non-zero conserved quantities, then the perturbations become adiabatic, and they remain adiabatic subsequently, even when the conditions of local thermal equilibrium are no longer satisfied. [8]

[475] For instance, if several constituents (such as an electron-positron plasma, and photons) each have a density and pressure that depends (even when perturbed) only on the temperature, then $\delta\rho_\alpha(T) = \rho'_\alpha(T)\delta T$ and $d\bar{\rho}_\alpha(T)/dt = \rho'_\alpha(T)\dot{\bar{T}}$, so $\delta\rho_\alpha/\dot{\bar{\rho}}_\alpha = \delta T/\dot{\bar{T}}$, and likewise for pressure and any other scalars.

[6] The existence of solutions with $\delta s_q/\dot{\bar{s}}$ equal for all four-scalars s such as energy densities, pressures, etc. (but not the detailed solutions (5.4.4)-(5.4.7)) seems to have been generally accepted for a long time. An intuitive "separate universe" argument for the existence of solutions for which $\delta\rho_{\alpha q}/\dot{\bar{\rho}}_\alpha$ are equal for all constituents α of the universe has been given by D. H. Lyth and D. Wands, Phys. Rev. D 68, 103516 (2003); also see D. Wands, K. A. Malik, D. H. Lyth, and A. R. Liddle, Phys. Rev. D 62, 043627 (2000); A. R. Liddle and D. H. Lyth, Cosmological Inflation and Large-Scale Structure (Cambridge University Press, 2000). This reasoning has been extended beyond perturbation theory by D. H. Lyth, K. A. Malik, and M. Sasaki, J. Cosm. & Astropart. Phys. 0505, 004 (2005) [astro-ph/0411220]. But this

The use of conservation laws to connect different cosmological eras is not limited to Newtonian gauge. Indeed, both ζ_q and $\mathcal{R}_q$ can be put in

a gauge invariant form. It is only necessary to remark that according to Eqs. (5.3.13) and (5.3.14), the quantities $A/2 - H\delta\rho/\dot{\bar{\rho}}$ and $A/2 + H\delta u$ are gauge invariant, and that they reduce in Newtonian gauge to the quantities (5.4.2) and (5.4.1), respectively, so in any gauge

$$\zeta_q = A_q/2 - H\delta\rho_q/\dot{\bar{\rho}}, \quad \mathcal{R}_q = A_q/2 + H\delta u_q, \tag{5.4.22}$$

provided of course that $A, \delta\rho$, and δu are all calculated in the same gauge.
For instance, in synchronous gauge Eq. (5.3.38) gives

$$q^2 A_q = 8\pi G a^2 \delta\rho_q - 2Ha^2\psi_q, \tag{5.4.23}$$

so in this gauge ζ_q and $\mathcal{R}_q$ can be expressed in terms of the convenient gravitational variable ψ_q by

$$\begin{aligned} q^2\zeta_q &= -a^2 H\psi_q + 4\pi G a^2\delta\rho_q - q^2 H\delta\rho_q/\dot{\bar{\rho}} \\ q^2\mathcal{R}_q &= -a^2 H\psi_q + 4\pi G a^2\delta\rho_q + q^2 H\delta u_q \end{aligned} \tag{5.4.24}$$

of course with the understanding that $\delta\rho_q$ and δu_q are here calculated in synchronous gauge. This result can also be derived directly from the Newtonian gauge formulas (5.4.1) and (5.4.2) by using the rules for transforming from Newtonian to synchronous gauge given in the previous section.

There is a convenient general formula for the rate of change of ζ_q and $\mathcal{R}_q$ that holds for $q/a \ll H$ whether or not the cosmological fluctuations are in an adiabatic mode. To derive this formula, we use the Newtonian gauge energy-conservation law (5.3.24), which for $q/a \ll H$ gives

$$\delta\dot{\rho}_q + 3H\left(\delta\rho_q + \delta p_q\right) = 3(\bar{\rho} + \bar{p})\dot{\Psi}_q, \tag{5.4.25}$$

and the corresponding unperturbed conservation law

$$\dot{\bar{\rho}} = -3H(\bar{\rho} + \bar{p}). \tag{5.4.26}$$

It is then a matter of simple algebra to calculate that for $q/a \ll H$

$$\dot{\zeta}_q = \frac{\dot{\bar{\rho}}\delta p_q - \dot{\bar{p}}\delta\rho_q}{3(\bar{\rho} + \bar{p})^2}. \tag{5.4.27}$$

For $q/a \ll H$, the same formula then also gives $\dot{\mathcal{R}}_q$. It should be noted that according to Eq. (5.3.14) the quantity on the right-hand side is gaugeinvariant, so we can check whether or not ζ_q and $\mathcal{R}_q$ are time-independent by evaluating $\dot{\bar{\rho}}\delta p_q - \dot{\bar{p}}\delta\rho_q$ in any gauge.

In particular, if for arbitrary perturbations $\bar{p} + \delta p$ is a function $F(\bar{\rho} + \delta\rho)$ of the perturbed energy density alone, then $\dot{\bar{p}} = F'(\bar{\rho})\dot{\bar{\rho}}$ and $\delta p_q = F'(\bar{\rho})\delta\rho_q$, so the terms in the numerator of

[47] sort of argument only shows that there is a solution satisfying this condition for zero wave number. As we have seen, there are indeed many such solutions for zero wave number, most of which have no physical significance because they cannot be extended to finite wave number. The proof presented here shows that the requirement that the solution can be extended to finite wave number yields just two solutions, described by Eqs. (5.4.4)-(5.4.7). It is this requirement that makes it necessary for the infinitesimal redefinition of the time coordinate, used in the "separate universe" argument to generate the solutions for zero wave number, to be accompanied with an infinitesimal rescaling of the space coordinate.

[7] S. Weinberg, Phys. Rev. D 70, 043541 (2004) (astro-ph/0401313).

[8] S. Weinberg, Phys. Rev. D 70, 083522 (2004) (astro-ph/0405397).

the right-hand side of Eq. (5.4.27) cancel,
and so ζ_q and $\mathcal{R}_q$ are conserved outside the horizon, even if the fluctuations are not in an adiabatic mode. [9] This is the case for instance if all abundant particles are highly relativistic, in which case the total perturbed pressure is 1/3 the total perturbed energy density, or if they are highly non-relativistic, in which case the total perturbed pressure is negligible. On the other hand, the theorem proved earlier in this section tells us that ζ_q and $\mathcal{R}_q$ are conserved outside the horizon in the adiabatic modes, even in cases (such as inflation with a single scalar field) for which the perturbed pressure is not given as a function only of the perturbed energy density.

The proof of Eq. (5.4.27) depended only on the energy conservation equations (5.4.25) and (5.4.26), not on the gravitational field equations. Therefore, if there is a constituent α of the universe that is energetically isolated, in the sense of not exchanging energy with the rest of the matter and radiation of the universe, then for $q/a \ll H$ the quantity

$$\zeta_{\alpha q} \equiv -\Psi_q + \delta\rho_{\alpha q}/3\left(\bar{\rho}_\alpha + \bar{p}_\alpha\right) \tag{5.4.28}$$

calculated using the energy density ρ_α and pressure p_α of the isolated constituent, satisfies the gauge-invariant relation [10]

$$\dot{\zeta}_{\alpha q} = \frac{\dot{\bar{\rho}}_\alpha \delta p_{\alpha q} - \dot{\bar{p}}_\alpha \delta\rho_{\alpha q}}{3\left(\bar{\rho}_\alpha + \bar{p}_\alpha\right)^2} \tag{5.4.29}$$

In particular, $\zeta_{\alpha q}$ (and hence also $\mathcal{R}_{\alpha q}$) is conserved outside the horizon if the perturbed pressure of the isolated constituent is a function only of its perturbed energy density. [11] (As we will see in Section 6.1, this is actually a good approximation for each of the individual constituents of the universe at times after the temperature dropped below about 10^{10} K, when neutrinos and antineutrinos no longer had significant interactions with matter and radiation.) In this case, Eq. (5.4.28) can be regarded as a convenient formula for the fractional density fluctuation outside the horizon in Newtonian gauge

$$\frac{\delta\rho_{\alpha q}}{3\left(\bar{\rho}_\alpha + \bar{p}_\alpha\right)} = \zeta_{\alpha q} + \Psi_q \tag{5.4.30}$$

with $\zeta_{\alpha q}$ time-independent.

There is a corresponding result in synchronous gauge: For long wavelengths the energy conservation equation (5.3.34) for an energetically isolated component of the universe takes the form

$$\delta\dot{\rho}^s_{\alpha q} + 3H\left(\delta\rho^s_{\alpha q} + \delta p^s_{\alpha q}\right) + \left(\bar{\rho}_\alpha + \bar{p}_\alpha\right)\psi_q = 0, \tag{5.4.31}$$

the superscript s denoting synchronous gauge. If also p_α is either negligible or a function only of ρ_α, then this can be written as

$$\frac{d}{dt}\left(\frac{\delta\rho^S_{\alpha q}}{\bar{\rho}_\alpha + \bar{p}_\alpha}\right) = -\psi_q,$$

[479] If the total pressure is a function only of the total energy density, then the three first-order differential equations (5.3.21), (5.3.23) and (5.3.24) form a closed set of equations for $\Psi = \Phi, \delta u$, and $\delta\rho$, so with the constraint (5.3.26) there must be just two independent solutions for these quantities, which therefore must take the adiabatic form (5.4.4) and (5.4.6), or (5.4.7). But the complete solution is not necessarily adiabatic, because other four-scalars (such as the energy densities or pressures of individual components of the universe) may not be given by Eq. (5.4.9).

[10] S. Bashinsky and U. Seljak, Phys. Rev. D 69, 083002 (2004) [astro-ph/0310198].

[11] D. Wands et al., ref. 2.

so that

$$\frac{\delta\rho^s_{\alpha q}}{\bar\rho_\alpha + \bar p_\alpha} = -\int \psi_q dt + c_{\alpha q} \tag{5.4.32}$$

The integration constants $c_{\alpha q}$ depend on what we take as the lower limit on the integral of ψ_q, but their differences have an absolute significance, and are simply related to the differences of the $\zeta_{\alpha q}$. The gauge transformation equation (5.3.14) and the equation $\dot{\bar\rho}_\alpha = -3H\left(\bar\rho_\alpha + \bar p_\alpha\right)$ for separate energy conservation shows that for any two components of the universe whose energy is separately conserved, we have

$$\frac{\delta\rho^S_{\alpha q}}{\bar\rho_\alpha + \bar p_\alpha} - \frac{\delta\rho^S_{\beta q}}{\bar\rho_\beta + \bar p_\beta} = \frac{\delta\rho_{\alpha q}}{\bar\rho_\alpha + \bar p_\alpha} - \frac{\delta\rho_{\beta q}}{\bar\rho_\beta + \bar p_\beta} \tag{5.4.33}$$

Comparing Eqs. (5.4.30) and (5.4.33), we see that

$$c_{\alpha q} - c_{\beta q} = 3\left(\zeta_{\alpha q} - \zeta_{\beta q}\right). \tag{5.4.34}$$

In the special case in which the perturbations are adiabatic, according to Eq. (5.4.9) the quantities $\delta\rho_{\alpha q}/\left(\bar\rho_\alpha + \bar p_\alpha\right) = -3H\delta\rho_{\alpha q}/\dot{\bar\rho}_\alpha$ are all equal, and hence the $\zeta_{\alpha q}$ are all equal, and in fact equal to ζ_q. According to Eq. (5.4.34), the $c_{\alpha q}$ will then also be all equal. The cosmic microwave background can be used to measure the differences of the $\zeta_{\alpha q}$ or $c_{\alpha q}$ before perturbations re-enter the horizon, and hence to decide whether the cosmological fluctuations are truly adiabatic. So far, as we will see in Section 7.2, it seems they are.

10.6 Evolution of Cosmological Fluctuations

We will now apply the formalism developed in the previous chapter to work out the evolution of cosmological fluctuations, from a temperature $\approx 10^9$ K when electron positron annihilation is substantially complete and neutrinos have decoupled from matter and radiation, down to the relatively recent time when the matter fluctuations become too large to be treated as firstorder perturbations. Our results will be applied to analyze the observed fluctuations in the cosmic microwave background in the next chapter, and large scale structure in Chapter 8.

For reasons discussed in Section 5.4, the connection between the evolution of fluctuations in this era and what happened at earlier times will appear in a few parameters including $\mathcal{R}_q$ (or ζ_q), which are conserved during the many e-foldings of expansion when the perturbations to the various constituents of the universe were still outside the horizon - that is, when the physical wave number q/a was much less than the expansion rate $\dot a/a$. The values of these parameters when outside the horizon thus characterize the strength of the various perturbations. In Chapter 10 we will see what can be understood about these parameters from a study of the much earlier era of inflation, and in this way work out what large scale structure and fluctuations in the cosmic microwave background can tell us about the era of inflation.

Section 6.1 presents the equations governing scalar perturbations. These equations take a simple hydrodynamic form for cold dark matter and the baryonic plasma, but for calculations of high accuracy it is necessary to use the Boltzmann equations of kinetic theory to follow the detailed distribution of photons and neutrinos in phase space. [1] These equations are too complicated for an analytic treatment; that is a task for comprehensive computer programs such as CMBfast [2] and CAMB. [3] Unfortunately such computer programs do not lend themselves to an exposition aimed at an

[471] P. J. E. Peebles and J. T. Yu, Astrophys. J. 162, 815 (1970); R. A. Sunyaev and Ya. B. Zel'dovich, Astrophys. Space Sci. 7, 3 (1970).

[2] The original code for the Boltzmann equations for photons and neutrinos and the other dynamical and

understanding of the physical phenomena involved. Therefore in subsequent sections we will present hydrodynamic calculations that are simple enough to be done analytically, aside from a few numerical integrations, and yet realistic enough so that the results are a good approximation to the more accurate results of computer programs.

The general equations and initial conditions for our analytic treatment of scalar modes are given in Section 6.2. Our analytic treatment of scalar modes then divides into the study of two wavelength regimes: wavelengths long enough to have come within the horizon during the matter-dominated era, to be considered in Section 6.3, and wavelengths short enough to have come within the horizon during the radiation-dominated era, considered in Section 6.4. Section 6.5 will show how to interpolate between the solutions found for these long and short wavelengths. Section 6.6 treats the evolution of tensor perturbations.

In Chapter 7 the results of this chapter for both scalar and tensor modes are applied to the anisotropies and polarization of the cosmic microwave background. Chapter 8 takes the treatment of matter perturbations beyond the time of last scattering, with results applied to observations of the cosmic distribution of matter.

10.6.1 6.1 Scalar perturbations - kinetic theory

It seems highly likely that from the beginning of the period of interest here, from just after $e^+ - e^-$annihilation at a temperature $T \approx 10^9$ K, down to the time of last scattering when $T \simeq 3,000$ K, the universe consisted of just four components: photons, cold dark matter, neutrinos, and a baryonic plasma consisting of free electrons, ions, and neutral atoms. In this section we will consider the perturbations in scalar modes to each of these four constituents in turn, adopting for this purpose the synchronous gauge described in Section 5.3. Each perturbed quantity $X(\mathbf{x}, t)$ (such as $\delta\rho, \delta p,\ A, B,$ etc.) is written as a Fourier integral and a sum over modes, as in Eq. (5.2.1):

$$X(\mathbf{x}, t) = \sum_n \int d^3q \alpha_n(\mathbf{q}) X_{nq}(t) e^{i\mathbf{q}\cdot\mathbf{x}} \tag{6.1.1}$$

where $\alpha_n(\mathbf{q})$ is the stochastic parameter for the n-th mode. In particular, the metric perturbation in synchronous gauge is given by $\delta g_{00} = 0, \delta g_{i0} = 0$, and

$$\delta g_{ij}(\mathbf{x}, t) = a^2(t) \sum_n \int d^3q \alpha_n(\mathbf{q}) \left[A_{nq}(t)\delta_{ij} - q_i q_j B_{nq}(t)\right] e^{i\mathbf{q}\cdot\mathbf{x}} \tag{6.1.2}$$

(Note that q is the co-moving wave number, related to the physical wave number k by $k = q/a$. It is common to define a so that $a = 1$ at the present time, so that the co-moving wave number equals the present value of the physical wave number, in which case the co-moving wave number is often denoted k. We will instead leave the normalization of a arbitrary, and reserve the symbol k for the physical wave number q/a.) In this section we will consider any one mode, dropping the label n; the equations we find will have a number of solutions, which define the various modes.

gravitational equations on which CMBfast is based was written by C.-P. Ma and E. Bertschinger, Astrophys. J. 455, 7 (1995) [astro-ph/9506072]. An important element discussed below was added by U. Seljak and M. A. Zaldarriaga, Astrophys. J. 469, 437 (1996).[astro-ph/9603033]. Also see M. Zaldarriaga, U. Seljak, and E. Bertschinger, Astrophys. J. 494, 491 (1998); M. Zaldarriaga and U. Seljak, Astrophys. J. 129, 431 (2000). The program is available on the website www. cmbfast. org.

[3] This program is based on CMBfast. It was written by A. Lewis and A. Challinor, and is available at camb.info/.

Cold dark matter

The individual cold dark matter particles are assumed to move too slowly for them to produce an appreciable pressure or anisotropic inertia, and, as shown in Section 5.3, the absence of pressure or anisotropic inertia allows us to adopt a particular synchronous gauge in which the cold dark matter fluid velocity u^i_D vanishes. That is, the coordinate mesh is tied to the dark matter particles in such a way that they remain at rest despite fluctuations in the gravitational field in which they move. Cold dark matter is therefore characterized solely by a total density $\bar{\rho}_D(t) + \delta\rho_D(\mathbf{x}, t)$, with the unperturbed density $\bar{\rho}_D(t)$ simply decreasing as $a^{-3}(t)$, and the Fourier transform of the perturbation $\delta\rho_D(\mathbf{x}, t)$ governed by the equation (5.3.34) of energy conservation with zero pressure and velocity

$$\delta\dot{\rho}_{Dq} + 3H\delta\rho_{Dq} = -\bar{\rho}_D\psi_q \tag{6.1.3}$$

(Recall that $\psi_q \equiv \left(3\dot{A}_q - q^2\dot{B}_q\right)/2$.)

Baryonic plasma

The Coulomb interactions of electrons and atomic nuclei are sufficiently strong so that they act together as a single perfect fluid. In the era of interest both electrons and nuclei are highly non-relativistic, so the baryonic plasma has negligible pressure and anisotropic inertia, and therefore the unperturbed density $\bar{\rho}_B(t)$ goes as $a^{-3}(t)$, and the Fourier transform of the density perturbation $\delta\rho_B(\mathbf{x}, t)$ is governed by the energy-conservation equation (5.3.34), but now with a non-zero velocity potential $\delta u_B(\mathbf{x}, t)$:

$$\delta\dot{\rho}_{Bq} + 3H\delta\rho_{Bq} - \left(q^2/a^2\right)\bar{\rho}_B\delta u_{Bq} = -\bar{\rho}_B\psi_q. \tag{6.1.4}$$

On the other hand, Thomson scattering allows the baryonic plasma to exchange momentum with photons, so it is the combination of photons and plasma that satisfies the equation (5.3.32) of momentum conservation, which here reads

$$\delta p_{\gamma q} - q^2\pi^S_{\gamma q} + [\partial_0 + 3H]\left[\bar{\rho}_B\delta u_{Bq} + \frac{4}{3}\bar{\rho}_\gamma\delta u_{\gamma q}\right] = 0 \tag{6.1.5}$$

where δp_γ and δu_γ are the pressure perturbation and velocity potential of the photons. We can regard this as the equation of motion of the baryonic plasma, with the photonic quantities $\pi^S_{\gamma q}, p_{\gamma q}, \delta u_{\gamma q}$ calculated as described in the next subsection.

Photons

Well before the time of recombination the density of free electrons was high enough so that photons could be described hydrodynamically: Thomson scattering gave the photons a total momentum locked to that of the baryonic plasma, so that $\delta u_{\gamma q} = \delta u_{Bq}$, and a momentum distribution that was isotropic in the co-moving frame, so that $\delta p_{\gamma q} = \delta\rho_{\gamma q}/3$ and $\pi^S_{\gamma q} = 0$. But for a highly accurate treatment of photons around the time of recombination it is necessary to treat them kinetically, studying the distribution of photons in momentum space, and taking account of photon polarization. As discussed in Appendix H, this distribution is an Hermitian number density matrix $n^{ij}(\mathbf{x}, \mathbf{p}, t)$, defined so that the number of photons in a space volume $\prod_i dx^i$ with momenta in a momentum-space volume $\prod_i dp_i$, weighted with the probability of their having polarization e^i, is $e_i e^*_j n^{ij}(\mathbf{x}, \mathbf{p}, t)\prod_k dx^k dp_k$, and $p_i n^{ij} = 0$. (The polarization vectors satisfy the normalization condition $g_{ij}e^i e^{j*} = 1$ and the transversality condition $p_i e^i = 0$,

and $e_i \equiv g_{ik}e^k$. For further discussion of polarization, see Appendix G.) For small perturbations, this distribution is conveniently written in the form

$$n^{ij}(\mathbf{x},\mathbf{p},t) = \frac{1}{2}\bar{n}_\gamma\left(a(t)p^0(\mathbf{x},\mathbf{p},t)\right)\left[g^{ij}(\mathbf{x},t) - \frac{g^{ik}(\mathbf{x},t)g^{jl}(\mathbf{x},t)p_k p_l}{\left[p^0(\mathbf{x},\mathbf{p},t)\right]^2}\right] + \delta n^{ij}(\mathbf{x},\mathbf{p},t) \tag{6.1.6}$$

where

$$p^0(\mathbf{x},\mathbf{p},t) \equiv \sqrt{g^{ij}(\mathbf{x},t)p_i p_j} \tag{6.1.7}$$

Here $\bar{n}_\gamma(p)$ is the equilibrium phase space number density

$$\bar{n}_\gamma(p) \equiv \frac{1}{(2\pi)^3}\left[\exp\left(p/k_{\mathcal{B}}a(t)\bar{T}(t)\right) - 1\right]^{-1} \tag{6.1.8}$$

with $\bar{T}(t)$ the unperturbed temperature of the baryonic plasma, and δn^{ij} is a small intrinsic perturbation. The first term in Eq. (6.1.6) is just the distribution matrix for unpolarized photons in thermal equilibrium at temperature $\bar{T}(t)$, written in a general spatial coordinate system, so δn includes the dynamical rather than the purely geometric effect of metric perturbations on the photon distribution. (Note that the factor $a(t)$ in the argument of $\bar{n}_\gamma$ in

Eq. (6.1.6) is canceled by the factor $a(t)$ multiplying $\bar{T}(t)$ in Eq. (6.1.8); this factor is introduced because in the era of interest $\bar{T}(t) \propto a^{-1}(t)$, so that as we have defined it, $\bar{n}_\gamma$ is a time-independent function of its argument.)

Since the sum over polarizations of $e_i e_j^*$ is g_{ij}, the phase space density of photons is $g_{ij}n^{ij}$. Hence the total energy-momentum tensor of the photons is

$$T^\mu_{\gamma\nu} = \frac{1}{\sqrt{\text{Det}\, g}}\int\left[\prod_{k=1}^{3}dp_k\right] g_{ij}n^{ij}\frac{p^\mu p_v}{p^0} \tag{6.1.9}$$

The first term in Eq. (6.1.6) contains first-order perturbations arising from the metric perturbations in $p^i = g^{ij}p_j$ and $p^0 = -p_0 = \sqrt{g^{ij}p_i p_j}$, in the metric determinant Detg, and in the factor g_{ij} in $g_{ij}n^{ij}$. It is straightforward though tedious to show directly that all these contributions to $\delta T^\mu_{\gamma\,\nu}$ cancel, but this can be seen more easily by noting that Eq. (6.1.9) shows that the contribution of the first term in Eq. (6.1.6) to $\delta T^i_{\gamma j}(\mathbf{x},t)$, $\delta T^0_{\gamma j}(\mathbf{x},t)$, $\delta T^j_{\gamma 0}(\mathbf{x},t)$, and $\delta T^0_{\gamma 0}(\mathbf{x},t)$ are local functions of the three-metric $g_{ij}(\mathbf{x},t)$ that transform under general spatial coordinate transformations as a mixed three-tensor, a covariant three-vector, a contravariant three-vector, and a three-scalar, respectively. But there are no non-trivial local functions of the three-metric that transform in this way under spatial coordinate transformations! This leaves only the contribution from the trace of δn^{ij}_γ. That is, the total first-order perturbations to the mixed components of the energy-momentum tensor are

$$\delta T^i_{\gamma j}(\mathbf{x},t) = \frac{1}{a^4(t)}\int\left(\prod_{k=1}^{3}dp_k\right)a^2(t)\delta n^{kk}_\gamma(\mathbf{x},\mathbf{p},t)\frac{p_i p_j}{\sqrt{p_k p_k}} \tag{6.1.10}$$

$$\delta T^0_{\gamma j}(\mathbf{x},t) = \frac{1}{a^3(t)}\int\left(\prod_{k=1}^{3}dp_k\right)a^2(t)\delta n^{kk}_\gamma(\mathbf{x},\mathbf{p},t)p_j \tag{6.1.11}$$

$$\delta T^0_{\gamma 0}(\mathbf{x},t) = -\frac{1}{a^4(t)}\int\left(\prod_{k=1}^{3}dp_k\right)a^2(t)\delta n^{kk}_\gamma(\mathbf{x},\mathbf{p},t)\sqrt{p_k p_k} \tag{6.1.12}$$

Evidently, all we need to calculate these perturbations is the integral of $\delta n_\gamma^{kk}(\mathbf{x},\mathbf{p},t)$ over photon energy, weighted with a single factor of energy. We therefore introduce a dimensionless intensity matrix $J_{ij}(\mathbf{x},\hat{p},t)$, defined by

$$a^4(t)\bar{\rho}_\gamma(t)J_{ij}(\mathbf{x},\hat{p},t) \equiv a^2(t)\int_0^\infty \delta n_\gamma^{ij}(\mathbf{x},p\hat{p},t)4\pi p^3 dp \tag{6.1.13}$$

where $\bar{\rho}_\gamma(t) \equiv a^{-4}(t)\int 4\pi p^3 \bar{n}_\gamma(p)dp$ is the unperturbed photon energy density. The components (6.1.10)-(6.1.12) then become

$$\delta T_{\gamma j}^i(\mathbf{x},t) = \bar{\rho}_\gamma(t)\int \frac{d^2\hat{p}}{4\pi} J_{kk}(\mathbf{x},\hat{p},t)\hat{p}_i\hat{p}_j \tag{6.1.14}$$

$$\delta T_{\gamma j}^0(\mathbf{x},t) = a(t)\bar{\rho}_\gamma(t)\int \frac{d^2\hat{p}}{4\pi} J_{kk}(\mathbf{x},\hat{p},t)p_j \tag{6.1.15}$$

$$\delta T_{\gamma 0}^0(\mathbf{x},t) = -\bar{\rho}_\gamma(t)\int \frac{d^2\hat{p}}{4\pi} J_{kk}(\mathbf{x},\hat{p},t) \tag{6.1.16}$$

As shown in Eq. (H.37) of Appendix H, the perturbation δn^{ij} is governed by the Boltzmann equation

$$\begin{aligned}
&\frac{\partial \delta n^{ij}(\mathbf{x},\mathbf{p},t)}{\partial t} + \frac{\hat{p}_k}{a(t)}\frac{\partial \delta n^{ij}(\mathbf{x},\mathbf{p},t)}{\partial x^k} + \frac{2\dot{a}(t)}{a(t)}\delta n^{ij}(\mathbf{x},\mathbf{p},t) \\
&\quad -\frac{1}{4a^2(t)}p\bar{n}'_\gamma(p)\hat{p}_k\hat{p}_l\frac{\partial}{\partial t}\left(a^{-2}(t)\delta g_{kl}(\mathbf{x},t)\right)(\delta_{ij}-\hat{p}_i\hat{p}_j) \\
&\quad = -\omega_c(t)\delta n^{ij}(\mathbf{x},\mathbf{p},t) + \frac{3\omega_c(t)}{8\pi}\int d^2\hat{p}_1 \\
&\times\left[\delta n^{ij}(\mathbf{x},p\hat{p}_1,t) - \hat{p}_i\hat{p}_k\delta n^{kj}(\mathbf{x},p\hat{p}_1,t) - \hat{p}_j\hat{p}_k\delta n^{ik}(\mathbf{x},p\hat{p}_1,t)\right. \\
&\qquad \left. +\hat{p}_i\hat{p}_j\hat{p}_k\hat{p}_l\delta n^{kl}(\mathbf{x},p\hat{p}_1,t)\right] \\
&\quad -\frac{\omega_c}{2a^3}p_k\delta u_{Bk}\bar{n}'_\gamma(p)(\delta_{ij}-\hat{p}_i\hat{p}_j)
\end{aligned} \tag{6.1.17}$$

where now $p \equiv \sqrt{p_i p_i}, \hat{p} \equiv \mathbf{p}/p; \omega_c(t)$ is the collision rate of a photon with electrons in the baryonic plasma; and δu_{Bk} is the peculiar velocity of the baryonic plasma. We can derive a Boltzmann equation for the dimensionless intensity matrix by multiplying Eq. (6.1.17) with $4\pi p^3$ and integrating over $p \equiv \sqrt{p_i p_i}$, using

$$4\pi\int_0^\infty p^4\bar{n}'_\gamma(p)dp = -16\pi\int_0^\infty p^3\bar{n}_\gamma(p)dp = -4a^4(t)\rho_\gamma(t)$$

Writing the intensity matrix and plasma velocity as Fourier transforms

$$J_{ij}(\mathbf{x},\hat{p},t) = \int d^3q J_{ij}(\mathbf{q},\hat{p},t)e^{i\mathbf{q}\cdot\mathbf{x}}, \quad \delta u_{Bk}(\mathbf{x},t) = \int d^3q\,\delta u_{Bi}(\mathbf{q},t)e^{i\mathbf{q}\cdot\mathbf{x}} \tag{6.1.18}$$

10.6.2 6.1 Scalar perturbations - kinetic theory

we find

$$\begin{aligned}
&\frac{\partial J_{ij}(\mathbf{q},\hat{p},t)}{\partial t}+i\frac{\hat{p}\cdot\mathbf{q}}{a(t)}J_{ij}(\mathbf{q},\hat{p},t)\\
&\quad+\alpha(\mathbf{q})\left[\dot{A}_q(t)-(\mathbf{q}\cdot\hat{p})^2\dot{B}_q(t)\right](\delta_{ij}-\hat{p}_i\hat{p}_j)\\
&=-\omega_c(t)J_{ij}(\mathbf{q},\hat{p},t)+\frac{3\omega_c(t)}{8\pi}\int d^2\hat{p}_1\\
&\quad\times\left[J_{ij}\left(\mathbf{q},\hat{p}_1,t\right)-\hat{p}_i\hat{p}_kJ_{kj}\left(\mathbf{q},\hat{p}_1,t\right)-\hat{p}_j\hat{p}_kJ_{ik}\left(\mathbf{q},\hat{p}_1,t\right)\right.\\
&\quad\left.+\hat{p}_i\hat{p}_j\hat{p}_k\hat{p}_lJ_{kl}\left(\mathbf{q},\hat{p}_1,t\right)\right]\\
&\quad+\frac{2\omega_c(t)}{a(t)}\left[\delta_{ij}-\hat{p}_i\hat{p}_j\right]\hat{p}_k\delta u_{Bk}(\mathbf{q},t)
\end{aligned}\tag{6.1.19}$$

in which we have used Eqs. (5.3.27) and (6.1.2) for the metric perturbation.
The intensity matrix and plasma velocity are proportional to the stochastic parameter $\alpha(\mathbf{q})$ for whatever mode is under consideration, which contains all information about initial conditions. Apart from this factor, there are no preferred directions in the problem, so the coefficient of $\alpha(\mathbf{q})$ in the intensity matrix can be decomposed into a sum of terms proportional to the two symmetric three-dimensional tensors $\delta_{ij}-\hat{p}_i\hat{p}_j$ and $(\hat{q}_i-(\hat{q}\cdot\hat{p})\hat{p}_i)(\hat{q}_j-(\hat{q}\cdot\hat{p})\hat{p}_j)$ that vanish when contracted with $\hat{p}_i$ or $\hat{p}_j$, with coefficients that depend on the directions $\hat{q}$ and $\hat{p}$ only through the scalar product $\hat{q}\cdot\hat{p}$. This decomposition is conventionally written as

$$\begin{aligned}
J_{ij}(\mathbf{q},\hat{p},t)=\alpha(\mathbf{q})\left\{\frac{1}{2}\left(\Delta_T^{(S)}(q,\hat{q}\cdot\hat{p},t)-\Delta_P^{(S)}(q,\hat{q}\cdot\hat{p},t)\right)(\delta_{ij}-\hat{p}_i\hat{p}_j)\right.\\
\left.+\Delta_P^{(S)}(q,\hat{q}\cdot\hat{p},t)\right)\left[\frac{(\hat{q}_i-(\hat{q}\cdot\hat{p})\hat{p}_i)(\hat{q}_j-(\hat{q}\cdot\hat{p})\hat{p}_j)}{1-(\hat{p}\cdot\hat{q})^2}\right]\Bigg\}.
\end{aligned}\tag{6.1.20}$$

(The subscripts T and P stand for "temperature" and "polarization." Note that the trace J_{ii}, which is all that appears in the energy-momentum tensor, is proportional solely to $\Delta_T^{(S)}$, but we need to keep track of $\Delta_P^{(S)}$ because it is linked to $\Delta_T^{(S)}$ through the dynamical equations.) Similarly, the integral over $\hat{p}$ appearing on the right-hand side of Eq. (6.1.19) may be expressed in terms of a pair of "source functions" $\Phi(q,t)$ and $\Pi(q,t)$ as

$$\int\frac{d^2\hat{p}}{4\pi}J_{ij}(\mathbf{q},\hat{p},t)=\alpha(\mathbf{q})\left[\delta_{ij}\Phi(q,t)+\frac{1}{2}\hat{q}_i\hat{q}_j\Pi(q,t)\right],\tag{6.1.21}$$

and as usual we write

$$\delta u_{Bi}(\mathbf{q},t)=i\alpha(\mathbf{q})q_i\delta u_{Bq}(t)\tag{6.1.22}$$

Inserting Eqs. (6.1.20)-(6.1.22) in Eq.(6.1.19) yields the coupled Boltzmann equations for $\Delta_T^{(S)}$ and $\Delta_P^{(S)}$:

$$\begin{aligned}
&\dot{\Delta}_P^{(S)}(q,\mu,t)+i\left(\frac{q\mu}{a(t)}\right)\Delta_P^{(S)}(q,\mu,t)=-\omega_c(t)\Delta_P^{(S)}(q,\mu,t)\\
&\quad+\frac{3}{4}\omega_c(t)\left(1-\mu^2\right)\Pi(q,t)
\end{aligned}\tag{6.1.23}$$

$$\begin{aligned}
&\dot{\Delta}_T^{(S)}(q,\mu,t)+i\left(\frac{q\mu}{a(t)}\right)\Delta_T^{(S)}(q,\mu,t)=-\omega_c(t)\Delta_T^{(S)}(q,\mu,t)\\
&\quad-2\dot{A}_q(t)+2q^2\mu^2\dot{B}_q(t)\\
&\quad+3\omega_c(t)\Phi(q,t)+\frac{3}{4}\omega_c(t)\left(1-\mu^2\right)\Pi(q,t)+4iq\mu\omega_c(t)\delta u_{Bq}(t)
\end{aligned}\tag{6.1.24}$$

with Φ and Π defined by Eq. (6.1.21).
The usual approach [4] to the solution of these Boltzmann equations is through an expansion of $\Delta_T^{(S)}$ and $\Delta_P^{(S)}$ in partial wave amplitudes:

$$\Delta_T^{(S)}(q,\mu,t) = \sum_{\ell=0}^{\infty} i^{-\ell}(2\ell+1)P_\ell(\mu)\Delta_{T,\ell}^{(S)}(q,t) \tag{6.1.25}$$

$$\Delta_P^{(S)}(q,\mu,t) = \sum_{\ell=0}^{\infty} i^{-\ell}(2\ell+1)P_\ell(\mu)\Delta_{P,\ell}^{(S)}(q,t) \tag{6.1.26}$$

To derive the Boltzmann equations for the partial wave amplitudes, we use the recursion and normalization relations for Legendre polynomials
$(2\ell+1)\mu P_\ell(\mu) = (\ell+1)P_{\ell+1}(\mu) + \ell P_{\ell-1}(\mu)$, $\quad \int_{-1}^{+1} P_\ell^2(\mu)d\mu = \frac{2}{2\ell+1}$.
Multiplying Eqs. (6.1.23) and (6.1.24) by $P_\ell(\mu)$ and integrating over μ then gives

$$\begin{aligned}
&\dot{\Delta}_{P,\ell}^{(S)} + \frac{q}{a(2\ell+1)}\left[(\ell+1)\Delta_{P,\ell+1}^{(S)} - \ell\Delta_{P,\ell-1}^{(S)}\right] \\
&\quad = -\omega_c\Delta_{P,\ell}^{(S)} + \frac{1}{2}\omega_c\Pi\left(\delta_{\ell 0} + \frac{\delta_{\ell 2}}{5}\right),
\end{aligned} \tag{6.1.27}$$

$$\begin{aligned}
&\dot{\Delta}_{T,\ell}^{(S)} + \frac{q}{a(2\ell+1)}\left[(\ell+1)\Delta_{T,\ell+1}^{(S)} - \ell\Delta_{T,\ell-1}^{(S)}\right] \\
&\quad = -2\dot{A}_q\delta_{\ell 0} + 2q^2\dot{B}_q\left(\frac{\delta_{\ell 0}}{3} - \frac{2\delta_{\ell 2}}{15}\right) \\
&\quad - \omega_c\Delta_{T,\ell}^{(S)} + \omega_c\left(3\Phi + \frac{1}{2}\Pi\right)\delta_{\ell 0} + \frac{1}{10}\omega_c\Pi\delta_{\ell 2} \\
&\quad - \frac{4}{3}q\omega_c\delta u_{Bq}\delta_{\ell 1}
\end{aligned} \tag{6.1.28}$$

We can express the source functions in terms of $\Delta_{T,\ell}^{(S)}$ and $\Delta_{P,\ell}^{(S)}$ by inserting Eq.(6.1.20) in Eq. (6.1.21), and using the integral formula, that for any function $f(\hat{q}\cdot\hat{p})$,

$$\frac{1}{4\pi}\int d^2\hat{p} f(\hat{q}\cdot\hat{p})\hat{p}_i\hat{p}_j = \mathcal{A}\delta_{ij} + \mathcal{B}\hat{q}_i\hat{q}_j$$

where

$$\mathcal{A} = \frac{1}{4}\int_{-1}^{1} d\mu f(\mu)\left(1-\mu^2\right) = \frac{1}{6}\int_{-1}^{1} d\mu f(\mu)\left(P_0(\mu) - P_2(\mu)\right)$$
$$\mathcal{B} = \frac{1}{4}\int_{-1}^{1} d\mu f(\mu)\left(3\mu^2-1\right) = \frac{1}{2}\int_{-1}^{1} d\mu f(\mu)P_2(\mu)$$

(The general form of the integrals is dictated by rotational invariance, while the formulas for the coefficients are found by contracting the integral with δ_{ij} and with $\hat{q}_i\hat{q}_j$, and then solving the resulting pair of linear equations for $\mathcal{A}$ and $\mathcal{B}$.) Using this, we can easily evaluate the two terms in the integral (6.1.21) (here dropping the arguments q and t):

[474] M. L. Wilson and J. Silk, Astrophys. J. 243, 14 (1981); J. R. Bond and G. Efstathiou, Astrophys. J. 285, L45 (1984); R. Crittenden, J. R. Bond, R. L. Davis, G. Efstathiou, and P. Steinhardt, Phys. Rev. Lett. 71, 324 (1993); C. -P. Ma and E. Bertschinger, Astrophys. J. 455, 7 (1995).

$$\frac{1}{8\pi}\int d^2\hat{p}\left(\Delta_T^{(S)}(\hat{q}\cdot\hat{p})-\Delta_P^{(S)}(\hat{q}\cdot\hat{p})\right)(\delta_{ij}-\hat{p}_i\hat{p}_j)$$
$$=\int_{-1}^{+1}d\mu\left(\Delta_T^{(S)}(\mu)-\Delta_P^{(S)}(\mu)\right)\left[\left(\frac{1}{6}P_0(\mu)+\frac{1}{12}P_2(\mu)\right)\delta_{ij}-\frac{1}{4}P_2(\mu)\hat{q}_i\hat{q}_j\right]$$
$$\frac{1}{4\pi}\int d^2\hat{p}\Delta_P^{(S)}(\hat{q}\cdot\hat{p})\frac{(\hat{q}_i-(\hat{q}\cdot\hat{p})\hat{p}_i)\,(\hat{q}_j-(\hat{q}\cdot\hat{p})\hat{p}_j)}{1-(\hat{q}\cdot\hat{p})^2}$$
$$=\int_{-1}^{+1}d\mu\Delta_P^{(S)}(\mu)\left[\left(\frac{1}{12}P_0(\mu)+\frac{1}{6}P_2(\mu)\right)\delta_{ij}+\left(\frac{1}{4}P_0(\mu)-\frac{1}{2}P_2(\mu)\right)\hat{q}_i\hat{q}_j\right]$$

In this way, and using the partial wave expansions (6.1.25) and (6.1.26), we find

$$\Phi=\frac{1}{6}\left[2\Delta_{T,0}^{(S)}-\Delta_{P,0}^{(S)}-\Delta_{T,2}^{(S)}-\Delta_{P,2}^{(S)}\right] \tag{6.1.29}$$
$$\Pi=\Delta_{P,0}^{(S)}+\Delta_{T,2}^{(S)}+\Delta_{P,2}^{(S)} \tag{6.1.30}$$

In the same way, using Eq. (6.1.20) in Eqs. (6.1.14)-(6.1.16) gives

$$\delta T^i_{\gamma j}(\mathbf{x},t)=\bar{\rho}_\gamma(t)\int d^3q e^{i\mathbf{q}\cdot\mathbf{x}}\alpha(\mathbf{q})\int_{-1}^{1}d\mu\Delta_T^{(S)}(q,\mu,t)$$
$$\times\left[\frac{1}{6}\left(P_0(\mu)-P_2(\mu)\right)\delta_{ij}+\frac{1}{2}P_2(\mu)\hat{q}_i\hat{q}_j\right]$$
$$\delta T^0_{\gamma j}(\mathbf{x},t)=a(t)\bar{\rho}_\gamma(t)\int d^3q e^{i\mathbf{q}\cdot\mathbf{x}}\alpha(\mathbf{q})\frac{\hat{q}_i}{2}\int_{-1}^{1}d\mu\Delta_T^{(S)}(q,\mu,t)P_1(\mu)$$
$$\delta T^0_{\gamma 0}(\mathbf{x},t)=-\frac{1}{2}\bar{\rho}_\gamma(t)\int d^3q e^{i\mathbf{q}\cdot\mathbf{x}}\alpha(\mathbf{q})\int_{-1}^{1}d\mu\Delta_T^{(S)}(q,\mu,t)P_0(\mu)$$

in which we have used the formula

$$\frac{1}{4\pi}\int d^2\hat{p}f(\hat{q}\cdot\hat{p})\hat{p}_i=\frac{\hat{q}_i}{2}\int_{-1}^{1}d\mu f(\mu)P_1(\mu)$$

Comparing this with the first three of Eqs. (5.1.43), and again using the partial wave expansions (6.1.25) and (6.1.26), we find

$$\delta p_{\gamma q}=\frac{\bar{\rho}_\gamma}{3}\left(\Delta_{T,0}^{(S)}+\Delta_{T,2}^{(S)}\right) \tag{6.1.31}$$
$$q^2\pi^S_{\gamma q}=\bar{\rho}_\gamma\Delta_{T,2}^{(S)} \tag{6.1.32}$$
$$\delta\rho_{\gamma q}=\bar{\rho}_\gamma\Delta_{T,0}^{(S)} \tag{6.1.33}$$
$$q\delta u_{\gamma q}=-\frac{3}{4}\Delta_{T,1}^{(S)} \tag{6.1.34}$$

As a check of Eqs. (6.1.31)-(6.1.33), note that when used in the last of Eqs. (5.1.43) these results give $\delta T^\lambda_\gamma\lambda=0$, a necessary consequence of the masslessness of the photon for any distribution of photon momenta.

Eqs. (6.1.31)-(6.1.34) show that to work out the cosmological evolution of the gravitational field and its effect on other perturbations in scalar modes, all we need to know about photons is $\Delta_{T,\ell}^{(S)}$ for $\ell\leq 2$. But of course the evolution of these three amplitudes is coupled by the Boltzmann equations (6.1.27)-(6.1.30) to both $\Delta_{T,\ell}^{(S)}$ and $\Delta_{P,\ell}^{(S)}$ for all higher ℓ. In computer

programs like CMBfast and CAMB, the partial wave expansion is cut off at a sufficiently high value of ℓ; in the latest version of CMBfast, the maximum value of ℓ is taken as $\ell_{\max} = 12$, in which case the computer has to solve $2(\ell_{\max}+1) = 26$ coupled ordinary differential equations for each value of q, not counting the other equations that describe the evolution of the baryonic plasma, cold dark matter, neutrinos, and the gravitational field.

As we will see in the next chapter, the interpretation of observations of the cosmic microwave radiation background requires calculation of $\Delta^{(S)}_{T,\ell}$ and $\Delta^{(S)}_{P,\ell}$ for ℓ ranging to values well over 1,000 . Originally this was done by a direct use of the coupled Boltzmann equations (6.1.27) and (6.1.28), [4] but this required hours or even days of computer time for each theoretical model. A great improvement was introduced with the suggestion to use instead a formal solution of the Boltzmann equation (6.1.19), in the form of a "line-of-sight" integral, [5] which in matrix notation takes the form

$$\begin{aligned} J_{ij}(\mathbf{q},\hat{p},t) = \alpha(\mathbf{q})\int_{t_1}^{t} dt' \exp\left(-i\mathbf{q}\cdot\hat{p}\int_{t'}^{t}\frac{dt''}{a(t'')} - \int_{t'}^{t} dt''\omega_c(t'')\right) \\ \times\Big[-(\delta_{ij}-\hat{p}_i\hat{p}_j)\left(\dot{A}_q(t') - (\hat{p}\cdot\mathbf{q})^2\dot{B}_q(t')\right) \\ + \frac{3\omega_c(t')}{2}(\delta_{ij}-\hat{p}_i\hat{p}_j)\,\Phi(q,t') \\ + \frac{3\omega_c(t')}{4}\Pi(q,t')\,(\hat{q}_i-\hat{p}_i(\hat{q}\cdot\hat{p}))\,(\hat{q}_j-\hat{p}_j(\hat{q}\cdot\hat{p})) \\ + \frac{2\omega_c(t')}{a(t')}\left[\delta_{ij}-\hat{p}_i\hat{p}_j\right]\hat{p}_k\delta u_k(\mathbf{q},t')\Big] \\ + J_{ij}(\mathbf{q},\hat{p},t_1)\exp\left(-i\mathbf{q}\cdot\hat{p}\int_{t_1}^{t}\frac{dt'}{a(t')} - \int_{t_1}^{t} dt'\omega_c(t')\right) \end{aligned} \tag{6.1.35}$$

where t_1 is any arbitrary initial time. If we choose t_1 to be sufficiently early, before recombination, so that $\omega_c(t_1) \gg H(t_1)$, and take t at any time

after recombination, the final term in Eq. (6.1.35) may be neglected, and we have

$$\begin{aligned} J_{ij}(\mathbf{q},\hat{p},t) = \alpha(\mathbf{q})\int_{t_1}^{t} dt' \exp\left(-i\mathbf{q}\cdot\hat{p}\int_{t'}^{t}\frac{dt''}{a(t'')} - \int_{t'}^{t} dt''\omega_c(t'')\right) \\ \times\Big[-(\delta_{ij}-\hat{p}_i\hat{p}_j)\left(\dot{A}_q(t') - (\hat{p}\cdot\mathbf{q})^2\dot{B}_q(t')\right) \\ + \frac{3\omega_c(t')}{2}(\delta_{ij}-\hat{p}_i\hat{p}_j)\,\Phi(q,t') \\ + \frac{3\omega_c(t')}{4}\Pi(q,t')\,(\hat{q}_i-\hat{p}_i(\hat{q}\cdot\hat{p})\,(\hat{q}_j-\hat{p}_j(\hat{q}\cdot\hat{p}) \\ + \frac{2\omega_c(t')}{a(t')}\left[\delta_{ij}-\hat{p}_i\hat{p}_j\right]\hat{p}_k\delta u_k(\mathbf{q},t')\Big] \end{aligned} \tag{6.1.36}$$

In terms of the temperature and polarization amplitudes defined by the decomposition (6.1.20), the line-of-sight solution reads

[475] U. Seljak and M. A. Zaldarriaga, Astrophys. J. 469, 437 (1996). [astro-ph/9603033].

$$\Delta_T^{(S)}(q,\mu,t) = \Delta_P^{(S)}(q,\mu,t) + 2\int_{t_1}^{t} dt' \exp\left[-iq\mu\int_{t'}^{t}\frac{dt''}{a\,(t'')} - \int_{t'}^{t}\omega_c\,(t'')\,dt''\right]$$
$$\times\left[-\dot{A}_q\,(t') + \mu^2 q^2 \dot{B}_q\,(t') + \frac{3}{2}\omega_c\,(t')\,\Phi\,(q,t') + \frac{2i\mu q\omega_c\,(t')}{a\,(t')}\delta u_q\,(t')\right] \tag{6.1.37}$$

$$\Delta_P^{(S)}(q,\mu,t) = \frac{3}{4}\left(1-\mu^2\right)\int_{t_1}^{t} dt' \exp\left[-iq\mu\int_{t'}^{t}\frac{dt''}{a\,(t'')} - \int_{t'}^{t}\omega_c\,(t'')\,dt''\right]$$
$$\times\,\omega_c\,(t')\,\Pi\,(q,t') \tag{6.1.38}$$

where δu_q is the scalar velocity potential, defined by $\delta u_k(\mathbf{q},t) = iq_k u_q(t)$. Once Φ and Π have been calculated from Eqs. (6.1.29) and (6.1.30), Eqs. (6.1.37) and (6.1.38) can be used to calculate $\Delta_{T,\ell}^{(S)}$ and $\Delta_{P,\ell}^{(S)}$ for arbitrarily high values of ℓ. (It is also possible to use Eq. (6.1.36) as a substitute for the partial wave expansion in calculating the source terms Φ and Π. Integrating Eq. (6.1.36) over $\hat{p}$ yields integral equations: [6] expressions for $\Phi(q,t)$ and $\Pi(q,t)$ as integrals from t_1 to t in which the integrand is a linear function of $\Phi\,(q.t')$ and $\Pi\,(q,t')$ for $t' < t$. This approach will be applied to tensor modes in Section 6.6.)

Neutrinos

The number density $n_\nu(\mathbf{x},\mathbf{p},t)$ of each species of massless neutrinos (or antineutrinos) in phase space can be conveniently expressed in terms of an intrinsic perturbation $\delta n_\nu(\mathbf{x},\mathbf{p},t)$ by a formula like Eq. (6.1.6):

$$n_v(\mathbf{x},\mathbf{p},t) = \bar{n}_v\left(a(t)p^0(\mathbf{x},\mathbf{p},t)\right) + \delta n_v(\mathbf{x},\mathbf{p},t) \tag{6.1.39}$$

where $\bar{n}_\nu$ is the equilibrium phase space density of each neutrino species

$$\bar{n}_v(p) \equiv \frac{1}{(2\pi)^3}\left[\exp\left(p/k_{\mathcal{B}}a(t)\bar{T}(t)\right)+1\right]^{-1} \tag{6.1.40}$$

As shown in Appendix H, the perturbation $\delta n_\nu(\mathbf{x},\mathbf{p},t)$ satisfies the same Boltzmann equation as the photon phase space density $a^2(t)n_\gamma^{kk}(\mathbf{x},\mathbf{p},t)$, except that for $T \ll 10^{10}$ K, the terms proportional to the collision frequency are absent:

$$0 = \frac{\partial\delta n_v(\mathbf{x},\mathbf{p},t)}{\partial t} + \frac{\hat{p}_k}{a(t)}\frac{\partial\delta n_v(\mathbf{x},\mathbf{p},t)}{\partial x^k}$$
$$-\frac{1}{2}p\bar{n}_v'(p)\hat{p}_k\hat{p}_l\frac{\partial}{\partial t}\left(a^{-2}(t)\delta g_{kl}(\mathbf{x},t)\right) \tag{6.1.41}$$

The contribution of each species of neutrino to the perturbations to the energy-momentum tensor is given by formulas (6.1.10)-(6.1.12), except that $\delta n_v(\mathbf{x},\mathbf{p},t)$ appears instead of $a^2(t)\delta n_\gamma^{kk}(\mathbf{x},\mathbf{p},t)$. Once again, all we need for this purpose is a dimensionless direction-dependent intensity, defined by a formula like Eq. (6.1.13):

$$a^4(t)\bar{\rho}_v(t)J(\mathbf{x},\hat{p},t) \equiv N_v\int_0^\infty \delta n_v(\mathbf{x},\mathbf{p},t)4\pi p^3 dp \tag{6.1.42}$$

where N_v is the number of species of neutrino, counting antineutrinos separately, and $\bar{\rho}_\nu \equiv N_\nu a^{-4}\int 4\pi p^3\bar{n}_\nu(p)dp$. Then the total neutrino and antineutrino contribution to the energy-momentum tensor is

[476] S. Weinberg, Phys. Rev. D 74, 063517 (2006) [astro-ph/0607076]; D. Baskaran, L. P. Grishchuk, and A. G. Polnarev, Phys. Rev. D 74, 063517 [gr-qc/0605100].

$$\delta T^i_{vj}(\mathbf{x},t) = \bar{\rho}_v(t)\int \frac{d^2\hat{p}}{4\pi} J(\mathbf{x},\hat{p},t)\hat{p}_i\hat{p}_j \tag{6.1.43}$$

$$\delta T^0_{vj}(\mathbf{x},t) = a(t)\bar{\rho}_v(t)\int \frac{d^2\hat{p}}{4\pi} J(\mathbf{x},\hat{p},t)p_j \tag{6.1.44}$$

$$\delta T^0_{v0}(\mathbf{x},t) = -\bar{\rho}_v(t)\int \frac{d^2\hat{p}}{4\pi} J(\mathbf{x},\hat{p},t) \tag{6.1.45}$$

just as in Eqs. (6.1.14)-(6.1.16). Rotational and translational invariance allow us to express $J(\mathbf{x},\hat{p},t)$ as a Fourier integral of the form

$$J(\mathbf{x},\hat{p},t) = \int \alpha(\mathbf{q})\Delta^{(S)}_v(q,\hat{q}\cdot\hat{p},t)e^{i\mathbf{q}\cdot\mathbf{x}}d^3q \tag{6.1.46}$$

To derive a Boltzmann equation for $\Delta^{(S)}_\nu$, we multiply Eq. (6.1.41) with $4\pi|\mathbf{p}|^3$ and integrate over $|\mathbf{p}|$, and find

$$\frac{\partial\Delta^{(S)}_v(q,\mu,t)}{\partial t} + i\frac{q\mu}{a(t)}\Delta^{(S)}_v(q,\mu,t) = -2\dot{A}_q(t) + 2q^2\mu^2\dot{B}_q(t) \tag{6.1.47}$$

In computer programs like CMBfast, the Boltzmann equation for neutrinos as well as for photons is solved by a partial wave expansion. One writes

$$\Delta^{(S)}_v(q,\mu,t) = \sum_{\ell=0}^{\infty} i^{-\ell}(2\ell+1)P_\ell(\mu)\Delta^{(S)}_{v,\ell}(q,t) \tag{6.1.48}$$

Inserting this in Eq. (6.1.46) and then in Eqs. (6.1.43)-(6.1.45) and then comparing the results with the first three of Eqs. (5.1.43) gives the perturbed pressure, scalar anisotropic inertia, perturbed energy density, and velocity potential of the neutrinos

$$\delta p_{vq}(t) = \frac{\bar{\rho}_v(t)}{3}\left(\Delta^{(S)}_{v,0}(q,t) + \Delta^{(S)}_{v,2}(q,t)\right) \tag{6.1.49}$$

$$q^2\pi^S_{vq}(t) = \bar{\rho}_v(t)\Delta^{(S)}_{v,2}(q,t) \tag{6.1.50}$$

$$\delta\rho_v q(t) = \bar{\rho}_v(t)\Delta^{(S)}_{v,0}(q,t) \tag{6.1.51}$$

$$q\delta u_v q = -\frac{3}{4}\Delta^{(S)}_{v,1}(q,t) \tag{6.1.52}$$

To derive the Boltzmann equations for the partial wave amplitudes, we multiply Eq. (6.1.47) with $P_\ell(\mu)$ and integrate over μ :

$$\begin{aligned}\dot{\Delta}^{(S)}_{v,\ell}(q,t) + \frac{q}{a(2\ell+1)}\left[(\ell+1)\Delta^{(S)}_{v,\ell+1}(q,t) - \ell\Delta^{(S)}_{v,\ell-1}(q,t)\right] \\ = -2\dot{A}_q(t)\delta_{\ell 0} + q^2\dot{B}_q(t)\left(\frac{2\delta_{\ell 0}}{3} - \frac{4\delta_{\ell 2}}{15}\right)\end{aligned} \tag{6.1.53}$$

In the current version of CMBfast, this equation is cut off at a maximum value of ℓ equal to 25 . Instead of relying on a truncated partial wave expansion, it is possible to write a solution (here not merely a formal solution) of Eq. (6.1.47) as another line-of-sight integral:

$$\Delta_v^{(S)}(q,\mu,t) = -2\int_{t_1}^{t} \exp\left(-iq\mu\int_{t'}^{t}\frac{dt''}{a(t'')}\right)\left(\dot{A}_q(t') - \mu^2 q^2 \dot{B}_q(t')\right) + \Delta_v^{(S)}(q,\mu,t_1)\exp\left(-iq\mu\int_{t_1}^{t}\frac{dt'}{a(t')}\right) \tag{6.1.54}$$

where t_1 is any arbitrary initial time. If t_1 is taken at some time after the neutrinos went out of thermal equilibrium with the baryonic plasma, but early enough so that gravitational field perturbations have not yet had a chance to distort the neutrino distribution, then $\Delta_\nu^{(S)}(q,\mu,t_1)$ arises only from the temperature perturbation at time t_1 :

$$\Delta_v^{(S)}(q,\mu,t_1) = 4\left[\frac{\delta T_q(t_1)}{\bar{T}(t_1)} + i\frac{\mu q \delta u_q(t_1)}{a(t_1)}\right] \tag{6.1.55}$$

in which the second term in square brackets arises from the Doppler effect due to the streaming of the electron-positron-photon plasma.

The integrals over direction in Eqs. (6.1.43)-(6.1.45) can be done analytically; for this purpose we need the formulas

$$\int\frac{d^2\hat{p}}{4\pi}\hat{p}_i\hat{p}_j\hat{p}_k\hat{p}_l e^{-i\hat{p}\cdot\mathbf{v}} = (\delta_{ij}\delta_{kl}+\delta_{ik}\delta_{jl}+\delta_{il}\delta_{kj})\,j_2(v)/v^2 - (\delta_{ij}\hat{v}_k\hat{v}_l+\delta_{ik}\hat{v}_j\hat{v}_l+\delta_{il}\hat{v}_k\hat{v}_j+\delta_{jk}\hat{v}_i\hat{v}_l+\delta_{jl}\hat{v}_k\hat{v}_i+\delta_{kl}\hat{v}_i\hat{v}_j)\,j_3(v)/v + \hat{v}_i\hat{v}_j\hat{v}_k\hat{v}_l j_4(v)$$

$$\int\frac{d^2\hat{p}}{4\pi}\hat{p}_i\hat{p}_j\hat{p}_k e^{-i\hat{p}\cdot\mathbf{v}} = -i\,(\delta_{ij}\hat{v}_k+\delta_{jk}\hat{v}_i+\delta_{ki}\hat{v}_j)\,j_2(v)/v + i\hat{v}_i\hat{v}_j\hat{v}_k j_3(v)$$

$$\int\frac{d^2\hat{p}}{4\pi}\hat{p}_i\hat{p}_j e^{-i\hat{p}\cdot\mathbf{v}} = \delta_{ij}j_1(v)/v - \hat{v}_i\hat{v}_j j_2(v) \tag{6.1.56}$$

$$\int\frac{d^2\hat{p}}{4\pi}\hat{p}_i e^{-i\hat{p}\cdot\mathbf{v}} = -i\hat{v}_i j_1(v)$$

$$\int\frac{d^2\hat{p}}{4\pi}e^{-i\hat{p}\cdot\mathbf{v}} = j_0(v), \tag{6.1.57}$$

here $v\equiv|\mathbf{v}|$, and $j_\ell(v)$ is the usual spherical Bessel function. Again (6.1.58)

omparing the results with the first three of Eqs. (5.1.43), we obtain explicit (6.1.59)

ormulas for the perturbed pressure, scalar anisotropic inertia, perturbed (6.1.60)

energy density, and velocity potential of the neutrinos: [7]

$$\delta p_{\nu q}(t) = -2\bar{\rho}_\nu(t)\int_{t_1}^{t}dt' \times\left[\dot{A}_q(t')\,K_1\left(q\int_{t'}^{t}\frac{dt''}{a(t'')}\right) - q^2\dot{B}_q(t')\,K_2\left(q\int_{t'}^{t}\frac{dt''}{a(t'')}\right)\right] \tag{6.1.61}$$

$$q^2\pi_{\nu q}^S(t) = -2\bar{\rho}_\nu(t)\int_{t_1}^{t}dt' \times\left[\dot{A}_q(t')\,j_2\left(q\int_{t'}^{t}\frac{dt''}{a(t'')}\right) - q^2\dot{B}_q(t')\,K_3\left(q\int_{t'}^{t}\frac{dt''}{a(t'')}\right)\right] \tag{6.1.62}$$

$$q\delta u_\nu q(t) = \frac{3a(t)}{2}\int_{t_1}^{t}dt'$$

$$\times \left[\dot{A}_q(t') j_1\left(q \int_{t'}^{t} \frac{dt''}{a(t'')} \right) - q^2 \dot{B}_q(t') K_4\left(q \int_{t'}^{t} \frac{dt''}{a(t'')} \right) \right] \tag{6.1.63}$$

$$\delta\rho_\nu q(t) = -2\bar{\rho}_\nu(t) \int_{t_1}^{t} dt' \times \left[\dot{A}_q(t') j_0\left(q \int_{t'}^{t} \frac{dt''}{a(t'')} \right) - q^2 \dot{B}_q(t') K_5\left(q \int_{t'}^{t} \frac{dt''}{a(t'')} \right) \right] \tag{6.1.64}$$

where

$$K_1(v) \equiv j_1(v)/v \tag{6.1.65}$$
$$K_2(v) \equiv j_2(v)/v^2 - j_3(v)/v \tag{6.1.66}$$
$$K_3(v) \equiv -2j_2(v)/v^2 + 5j_3(v)/v - j_4(v) \tag{6.1.67}$$
$$K_4(v) \equiv 3j_2(v)/v - j_3(v) \tag{6.1.68}$$
$$K_5(v) \equiv j_1(v)/v - j_2(v) \tag{6.1.69}$$

Using these formulas, one no longer needs the truncated partial wave expansion for neutrinos.

Gravitation

It only remains to give the equations of motion for the scalar gravitational field components $\dot{A}$ and $\dot{B}$. (Note that A and B themselves are nowhere needed.) It is convenient to take one of these as Eq. (5.3.31), written as

$$\frac{\partial}{\partial t}\left[a^2\psi_q\right] = -4\pi G a^2 \times \left(\delta\rho_{Dq} + \delta\rho_{Bq} + \delta\rho_{\gamma q} + \delta\rho_{\nu q} + 3\delta p_{\gamma q} + 3\delta p_{\nu q} - q^2\pi^S_{\gamma q} - q^2\pi^S_{\nu q}\right) \tag{6.1.70}$$

where

$$\psi_q \equiv \frac{1}{2}\left(3\dot{A}_q - q^2\dot{B}_q\right) \tag{6.1.71}$$

The other can be taken from Eq. (5.3.30)

$$\dot{A}_q = 8\pi G\left[\frac{4}{3}\bar{\rho}_\gamma \delta u_{\gamma q} + \frac{4}{3}\bar{\rho}_\nu \delta u_{\nu q} + \bar{\rho}_B \delta u_{Bq}\right]. \tag{6.1.72}$$

(Recall that we have adopted the particular synchronous gauge in which $\delta u_{Dq} = 0$.) After solving the first-order differential equation (6.1.70) for ψ_q and using Eq. (6.1.72) to find $\dot{A}_q$, the other component is trivially given by the definition Eq. (6.1.71) as

$$q^2\dot{B}_q = 3\dot{A}_q - 2\psi_q. \tag{6.1.73}$$

With this, as long as we truncate the partial wave expansions used for photons and neutrinos, we have a closed system of ordinary differential equations for all the perturbations, which can be straightforwardly solved by computer for any given initial conditions.

To find initial conditions, we note that at a sufficiently early time t_1, well before the era of recombination, (say, for $\bar{T}(t_1) > 10^5$ K), the collision rate of photons with the baryonic plasma

[477] These formulas are given also for massive neutrinos by S. Weinberg, ref. 6.

is so great that photons are in thermal and kinetic equilibrium with the plasma. Under these conditions, the photon distribution δn^{ij} arises only from a perturbation to the temperature in the first term of Eq. (H.31), including the Doppler shift due to the photon streaming velocity, which in equilibrium is the same as the baryonic plasma velocity $\delta \mathbf{u}_B$:
$\delta n_\gamma^{ij}(\mathbf{x}, \mathbf{p}, t_1) = -\frac{1}{2} a^{-2}(t_1) [\delta_{ij} - \hat{p}_i \hat{p}_j] \bar{n}'_\gamma(p) p \left[\frac{\delta T(\mathbf{x}, t_1)}{\bar{T}(t_1)} + \frac{\hat{p}_k \delta u_{Bk}(\mathbf{x}, t_1)}{a(t_1)} \right]$.
(The factor $1/a(t_1)$ in the Doppler term in Eq. (6.1.74) was explained in connection with Eq. (H.13).) Note that for ω_c very large, the coefficients of ω_c on the right-hand side of equation (6.1.17) must cancel, which gives an
initial condition consistent with Eq. (6.1.74). Multiplying with $4\pi p^3$ and integrating over p, we find a corresponding condition on J_{ij} :

$$J_{ij}(\mathbf{x}, \hat{p}, t_1) = 2 [\delta_{ij} - \hat{p}_i \hat{p}_j] \left[\frac{\delta T(\mathbf{x}, t_1)}{\bar{T}(t_1)} + \frac{\hat{p}_k \delta u_k(\mathbf{x}, t_1)}{a(t_1)} \right] \tag{6.1.75}$$

Note that $J_{ij}(\mathbf{x}, \hat{p}, t_1)$ receives contributions only from scalar and vector perturbations, not from tensor perturbations.

Similar remarks apply to neutrinos, except that we must go back to an earlier time, when the temperature was a little below 10^{10} K, so that neutrinos were already traveling freely, but not enough time had elapsed for the gravitational field perturbation to have altered the equilibrium form of the neutrino phase space distribution.

This still leaves us with the necessity of stipulating initial values for A_q, $B_q, \delta_{Dq}, \delta_{Bq}, \delta u_{Bq}$, and δT_q. For this, we must go back to a time early enough so that the wave numbers of interest were outside the horizon, in the sense that $q/a \ll H$. In the following section the needed initial values will be worked out for the dominant adiabatic mode, with a normalization expressed in terms of the quantity $\mathcal{R}_q$, given in synchronous gauge outside the horizon by Eq. (5.4.24).

10.6.3 6.2 Scalar perturbations - the hydrodynamic limit

The system of equations described in the previous section is much too complicated to allow an analytic solution. Fortunately, until near the time of recombination the rate of collisions of photons with free electrons was so great that photons were in local thermal equilibrium with the baryonic plasma, and so photons at these times can be treated hydrodynamically, like the plasma and cold dark matter. This approach loses its validity around the time of recombination, but a fair degree of accuracy will be preserved in Section 6.4 by taking into account the damping caused by the growing mean free times in this era. After the time of recombination photons traveled more or less freely, and their path can be followed by solving their equation of motion. Neutrinos are more of a problem, but at very early times perturbations were outside the horizon, so at these times $q^2 \pi_{\nu q}^S$ and $q \delta u_{\nu q}$ were negligible and Eqs. (6.1.49)-(6.1.51) show that $\delta \rho_{\nu q} = 3 \delta p_{\nu q}$, just as if the neutrinos were in local thermal equilibrium, while at late times the universe became matter dominated and neutrinos made only a small contribution to the cosmic gravitational field.

With these justifications, in order to allow an analytic treatment, in this and the next three sections we will adopt a hydrodynamic approach. [1] To be specific, for the most part we will neglect anisotropic inertia, take $\delta u_{\gamma q}$ equal to δu_{Bq}, and take $p_{\gamma q} = \rho_{\gamma q}/3$ and $p_{\nu q} = \rho_{\nu q}/3$. This necessarily entails the loss of some accuracy, but our aim in this chapter (and in the next two chapters) is not to calculate the course of cosmic evolution and its observational consequences with the high level of accuracy that would optimize the extraction of cosmological parameters from the latest data on the cosmic microwave background and large scale structure. Rather, we wish here to elucidate the physics of cosmic evolution, and clarify the dependence of observables on cosmological assumptions. Fortunately, the results we obtain from this analytic treatment

will turn out in Chapter 7 to yield predictions for anisotropies in the cosmic microwave background that are quite similar to those obtained by comprehensive computer programs, using the full Boltzmann equations described in the previous section. We are thereby reassured that the hydrodynamic approach captures the essence of what is going on in the early universe.

Under the above assumptions, the Fourier transforms with co-moving wave number q of the synchronous gauge perturbations are governed by the gravitational field equation (5.3.36):

$$\frac{d}{dt}\left(a^2\psi_q\right) = -4\pi G a^2\left(\delta\rho_{Dq} + \delta\rho_{Bq} + 2\delta\rho_{\gamma q} + 2\delta\rho_{\nu q}\right) \tag{6.2.1}$$

the equations (5.3.34) of energy conservation for each of the four fluids

$$\delta\dot{\rho}_{\gamma q} + 4H\delta\rho_{\gamma q} - \left(4q^2/3a^2\right)\bar{\rho}_\gamma\delta u_{\gamma q} = -(4/3)\bar{\rho}_\gamma\psi_q \tag{6.2.2}$$

$$\delta\dot{\rho}_{Dq} + 3H\delta\rho_{Dq} = -\bar{\rho}_D\psi_q \tag{6.2.3}$$

$$\delta\dot{\rho}_{Bq} + 3H\delta\rho_{Bq} - \left(q^2/a^2\right)\bar{\rho}_B\delta u_{\gamma q} = -\bar{\rho}_B\psi_q \tag{6.2.4}$$

$$\delta\dot{\rho}_{\nu q} + 4H\delta\rho_{\nu q} - \left(4q^2/3a^2\right)\bar{\rho}_\nu\delta u_{\nu q} = -(4/3)\bar{\rho}_\nu\psi_q \tag{6.2.5}$$

and the equations (5.3.32) of momentum conservation for the photonbaryon plasma and the neutrinos:

$$\frac{d}{dt}\left(\left(\frac{4}{3}\bar{\rho}_\gamma + \bar{\rho}_B\right)\delta u_{\gamma q}\right) + 3H\left(\left(\frac{4}{3}\bar{\rho}_\gamma + \bar{\rho}_B\right)\delta u_{\gamma q}\right) = -(1/3)\delta\rho_{\gamma q} \tag{6.2.6}$$

$$\frac{d}{dt}\left(\bar{\rho}_\nu\delta u_{\nu q}\right) + 3H\bar{\rho}_\nu\delta u_{\nu q} = -(1/4)\delta\rho_{\nu q} \tag{6.2.7}$$

It is very convenient to rewrite these equations in terms of the dimensionless fractional perturbations [2]

$$\delta_{\alpha q} \equiv \frac{\delta\rho_{\alpha q}}{\bar{\rho}_\alpha + \bar{p}_\alpha} \tag{6.2.8}$$

where α runs over γ, D, B, and ν. Taking into account that $a^4\bar{\rho}_\gamma, a^3\bar{\rho}_D$, $a^3\bar{\rho}_B$, and $a^4\bar{\rho}_\nu$ are all time-independent, Eqs. (6.2.1)-(6.2.7) now read

$$\frac{d}{dt}\left(a^2\psi_q\right) = -4\pi G a^2\left(\bar{\rho}_D\delta_{Dq} + \bar{\rho}_B\delta_{Bq} + \frac{8}{3}\bar{\rho}_\gamma\delta_{\gamma q} + \frac{8}{3}\bar{\rho}_\nu\delta_{\nu q}\right) \tag{6.2.9}$$

$$\dot{\delta}_{\gamma q} - \left(q^2/a^2\right)\delta u_{\gamma q} = -\psi_q \tag{6.2.10}$$

$$\dot{\delta}_{Dq} = -\psi_q \tag{6.2.11}$$

$$\dot{\delta}_{Bq} - \left(q^2/a^2\right)\delta u_{\gamma q} = -\psi_q \tag{6.2.12}$$

$$\dot{\delta}_{\nu q} - \left(q^2/a^2\right)\delta u_{\nu q} = -\psi_q \tag{6.2.13}$$

$$\frac{d}{dt}\left(\frac{(1+R)\delta u_{\gamma q}}{a}\right) = -\frac{1}{3a}\delta_{\gamma q} \tag{6.2.14}$$

$$\frac{d}{dt}\left(\frac{\delta u_{\nu q}}{a}\right) = -\frac{1}{3a}\delta_{\nu q} \tag{6.2.15}$$

[47]1 Other analytic or semi-analytic treatments of the evolution of fluctuations have been given by W.Hu and N. Sugiyama, Astrophys. J. 444, 489 (1995); 471, 542 (1996); V. Mukhanov, Int. J. Theor. Phys. 43, 623 (2004) [astro-ph/0303072]. The treatment given here is in my opinion more transparent though somewhat less accurate than that of Hu and Sugiyama, and (because we allow ourselves the use of a computer to do numerical integrals) it is more accurate than that of Mukhanov.

where $R \equiv 3\bar{\rho}_B/4\bar{\rho}_\gamma$. Eqs. (6.2.1)-(6.2.7) or (6.2.9)-(6.2.15) form a closed system of seven first-order differential equations for ψ_q, the four density perturbations, and the plasma and neutrino velocity potentials, so there must be seven independent solutions.

Before trying to find solutions valid up to the time of recombination, we must first consider the initial conditions to be imposed. These initial conditions will distinguish the different independent solutions. At sufficiently early times the universe was in a radiation-dominated era, when $\bar{\rho}_M \ll \bar{\rho}_R$, where

$$\bar{\rho}_M \equiv \bar{\rho}_D + \bar{\rho}_B, \quad \bar{\rho}_R \equiv \bar{\rho}_\gamma + \bar{\rho}_v \tag{6.2.16}$$

so that to a good approximation $a \propto \sqrt{t}$ and $8\pi G\bar{\rho}_R/3 = 1/4t^2$, while $R \ll 1$. (This fixes our definition of the zero of time.) If we take $a \propto \sqrt{t}$ and $R \ll 1$, Eqs. (6.2.9)-(6.2.15) become

$$\frac{d}{dt}\left(t\psi_q\right) = -4\pi G t\left(\bar{\rho}_D\delta_{Dq} + \bar{\rho}_B\delta_{Bq} + \frac{8}{3}\bar{\rho}_\gamma\delta_{\gamma q} + \frac{8}{3}\bar{\rho}_\nu\delta_{\nu q}\right) \tag{6.2.17}$$

$$\dot{\delta}_{\gamma q} = \dot{\delta}_{Bq} = -\psi_q + \left(q^2/a^2\right)\delta u_{\gamma q} \tag{6.2.18}$$

$$\dot{\delta}_{Dq} = -\psi_q \tag{6.2.19}$$

$$\dot{\delta}_{vq} = -\psi_q + \left(q^2/a^2\right)\delta u_{vq} \tag{6.2.20}$$

$$\frac{d}{dt}\left(\frac{\delta u_{\gamma q}}{\sqrt{t}}\right) = -\frac{1}{3\sqrt{t}}\delta_{\gamma q} \tag{6.2.21}$$

$$\frac{d}{dt}\left(\frac{\delta u_{vq}}{\sqrt{t}}\right) = -\frac{1}{3\sqrt{t}}\delta_{vq} \tag{6.2.22}$$

At very early times the perturbation was outside the horizon, in this sense that $q/aH \ll 1$, but we have not yet dropped the terms in Eqs. (6.2.18) and (6.2.20) proportional to q^2, because in some modes there are cancelations in the calculation of the conserved quantity $\mathcal{R}_q$ outside the horizon that require us to take such terms into account. Also, we have not dropped the terms on the right-hand side of Eq. (6.2.17) proportional to $\bar{\rho}_D$ or $\bar{\rho}_B$, because even though we are now assuming that $\bar{\rho}_D$ and $\bar{\rho}_B$ are much less than $\bar{\rho}_\gamma$ and $\bar{\rho}_\nu$, we want to leave open the possibility of modes in which the fractional fluctuations in the dark matter and/or baryon density are much larger than the fractional fluctuations in the photon and neutrino densities.

Mode 1

This is the dominant adiabatic mode - adiabatic, in the sense that all the $\delta_{\alpha q}$ become equal at very early times. (As discussed in Section 5.4, only these modes are present in inflationary theories with a single scalar field, or if the universe was ever earlier in a state of complete local thermal equilibrium with no non-zero conserved quantities.) Inspection of Eqs. (6.2.17)-(6.2.22) shows that, if we make the ansatz,

$$\delta_{\gamma q} = \delta_{Bq} = \delta_{Dq} = \delta_{\nu q} \equiv \delta_q, \quad \delta u_{\gamma q} = \delta u_{\nu q} \equiv \delta u_q \tag{6.2.23}$$

and if we now drop the baryon and dark matter-dominated terms in Eq. (6.2.17), and consider times early enough so that we can drop the q^2/a^2 terms in Eqs. (6.2.18) and (6.2.20), then Eqs. (6.2.18)-(6.2.22) and (6.2.17) become

[472] Note that this differs from a commonly used convention, according to which $\delta_{\alpha q}$ would be defined as $\delta\rho_{\alpha q}/\bar{\rho}_\alpha$.

$$\dot{\delta}_q = -\psi_q, \tag{6.2.24}$$

$$\frac{d}{dt}\left(\frac{\delta u_q}{\sqrt{t}}\right) = -\frac{1}{3\sqrt{t}}\delta_q, \tag{6.2.25}$$

and

$$\frac{d}{dt}\left(t\psi_q\right) = -\frac{1}{t}\delta_q \tag{6.2.26}$$

Combining Eqs. (6.2.24) and (6.2.26) gives a second-order differential equation for δ_q :

$$\frac{d}{dt}\left(t\frac{d}{dt}\delta_q\right) - \frac{1}{t}\delta_q = 0$$

This has two solutions, with $\delta_q \propto t$ and $\delta_q \propto 1/t$, and for each solution Eqs. (6.2.24) and (6.2.25) give solutions for ψ_q and δu_q. The growing solution has $\delta_q \propto t$ and $\delta u_q \propto t^2$, and gives our first adiabatic mode:

$$\delta_{\gamma q} = \delta_{Bq} = \delta_{\nu q} = \delta_{Dq} = \frac{q^2 t^2 \mathcal{R}_q^o}{a^2}, \tag{6.2.27}$$

$$\psi_q = -\frac{t q^2 \mathcal{R}_q^o}{a^2}, \tag{6.2.28}$$

$$\delta u_{\gamma q} = \delta u_{\nu q} = -\frac{2t^3 q^2 \mathcal{R}_q^o}{9a^2}, \tag{6.2.29}$$

We have normalized this mode so that the quantity given by Eq. (5.4.24) as

$$q^2\mathcal{R}_q \equiv -a^2 H\psi_q + 4\pi G a^2 \delta\rho_q + q^2 H\delta u_q \tag{6.2.30}$$

takes the time-independent value $q^2\mathcal{R}_q^o$ for $q/a \ll H$ (the superscript o standing for "outside the horizon").

Mode 2

The solution of Eqs. (6.2.24)-(6.2.26) which goes as $\delta_q \propto 1/t$ for $t \to 0$ gives us our second adiabatic solution:

$$\delta_{\gamma q} = \delta_{Bq} = \delta_{\nu q} = \delta_{Dq} = \epsilon_q/t, \quad \psi_q = \epsilon_q/t^2, \quad \delta u_{\gamma q} = \delta u_{\nu q} = \frac{2\epsilon_q}{3} \tag{6.2.31}$$

with ϵ_q an arbitrary time-independent function of q. The calculation of $\mathcal{R}_q$ for this solution has the problem that the first two terms in Eq. (6.2.30) for $q^2\mathcal{R}_q$ cancel to zeroth order in q^2/a^2H^2, leaving us with an unknown residue in $q^2\mathcal{R}_q$ of order q^2/a^2H^2, and hence an unknown term in $\mathcal{R}_q$ of zeroth order in q^2/a^2H^2. Fortunately, in this mode we can find a solution to Eqs. (6.2.17)-(6.2.22) that is valid to all orders in q/aH, as long as
$\rho_M \ll \rho_R$:

$$\delta_{\gamma q} = \delta_{Bq} = \delta_{\nu q} = \epsilon_q/t, \quad \delta_{Dq} = \frac{\epsilon_q}{t}\left[1 - \frac{q^2}{3H^2a^2}\ln\left(\frac{q}{Ha}\right)\right] \tag{6.2.32}$$

$$\psi_q = \frac{\epsilon_q}{t^2}\left[1 + \frac{2q^2t^2}{3a^2}\right] \tag{6.2.33}$$

$$\delta u_{\gamma q} = \delta u_{\nu q} = \frac{2\epsilon_q}{3} \tag{6.2.34}$$

Using this in Eq. (6.2.30) shows that this mode has $\mathcal{R}_q = 0$ to all orders in q/aH as long as $\bar{\rho}_M \ll \bar{\rho}_R$.

The other five modes are non-adiabatic, in the sense that some of the $\delta_{\alpha q}$ are unequal even for $q/a \ll H$. One particularly simple mode can serve as an illustration:

Mode 3

$$\begin{gathered} \delta_{Dq} = \frac{\epsilon_q \bar{\rho}_B}{\bar{\rho}_B + \bar{\rho}_D}, \quad \delta_{Bq} = -\frac{\epsilon_q \bar{\rho}_D}{\bar{\rho}_B + \bar{\rho}_D} \\ \psi_q = 0, \quad \delta_{\gamma q} = \delta_{\nu q} = 0, \quad \delta u_{\gamma q} = \delta u_{\nu q} = 0 \end{gathered} \tag{6.2.36}$$

again with ϵ_q time-independent but otherwise arbitrary. This just amounts to a perturbation in the time-independent ratio of the densities of baryons and dark matter, and is an exact solution for all times. It is an isocurvature mode, in both the sense that $\mathcal{R}_q = 0$, and also that $\psi_q = 0$.

As already indicated, these results apply only at times early enough so that $\bar{\rho}_M \ll \bar{\rho}_R$ (and, for mode 1, $q/a \ll H$). As an aid to extending these early-time solutions to later times, note that for all times before recombination, the difference of Eqs. (6.2.10) and (6.2.12) gives

$$\frac{d}{dt}\left(\delta_{Bq} - \delta_{\gamma q}\right) = 0 \tag{6.2.37}$$

We see that any solution that satisfies the adiabatic condition, that $\delta_{Bq} = \delta_{\gamma q}$ at early times, when the perturbation is far outside the horizon, will continue thereafter to have

$$\delta_{Bq} = \delta_{\gamma q} \tag{6.2.38}$$

Eqs. (6.2.9)-(6.2.15) are then reduced to

$$\frac{d}{dt}\left(a^2 \psi_q\right) = -4\pi G a^2 \left[\bar{\rho}_D \delta_{Dq} + \left(\bar{\rho}_B + \frac{8}{3}\bar{\rho}_\gamma\right)\delta_{\gamma q} + \frac{8}{3}\bar{\rho}_\nu \delta_{\nu q}\right] \tag{6.2.39}$$

$$\dot{\delta}_{\gamma q} - \left(q^2/a^2\right)\delta u_{\gamma q} = -\psi_q \tag{6.2.40}$$

$$\dot{\delta}_{Dq} = -\psi_q \tag{6.2.41}$$

$$\dot{\delta}_{\nu q} - \left(q^2/a^2\right)\delta u_{\nu q} = -\psi_q \tag{6.2.42}$$

$$\frac{d}{dt}\left(\frac{(1+R)\delta u_{\gamma q}}{a}\right) = -\frac{1}{3a}\delta_{\gamma q} \tag{6.2.43}$$

$$\frac{d}{dt}\left(\frac{\delta u_{\nu q}}{a}\right) = -\frac{1}{3a}\delta_{\nu q} \tag{6.2.44}$$

where again $R \equiv 3\bar{\rho}_B/4\bar{\rho}_R$. These equations apply to modes 1 and 2 at all times before recombination. We will be chiefly interested in mode 1 , since the perturbations of mode 2 decay by a factor $1/t^2$ relative to those of mode 1 during the part of the radiation dominated era when the perturbation is still outside the horizon.

It is not possible to find an analytic solution of even the reduced set of equations (6.2.39)-(6.2.44) that would be valid for all times and wave numbers. They can be treated analytically, however, in two wavelength regimes: long wavelengths, for which $q \ll q_{\rm EQ}$, and short wavelengths, for which $q \gg q_{\rm EQ}$, where $q_{\rm EQ}$ is the wave number for which $q/a = H$ at matter-radiation equality. Recall that, once inflation is over, q/a decreases more slowly than H, so that for long wavelengths, the wave number is so small that $\bar{\rho}_M$ becomes equal to $\bar{\rho}_R$ when q/a is still much less than H, while for short wavelengths, the wave number is so large that q/a becomes equal to H when $\bar{\rho}_M$ is still much less than $\bar{\rho}_R$. For long wavelengths we will be able

to find analytic solutions of Eqs. (6.2.39)-(6.2.44) in both the early era, when the perturbation is outside the horizon, and in the later era, when the expansion is dominated by non-relativistic matter, and patch them together in the era of overlap, when the perturbation is still outside the horizon and the universe is already matter dominated. Conversely, for short wavelengths we will be able to find analytic solutions of these equations in the early era when the universe is radiation dominated and in the later era when the perturbation is deep inside the horizon, and patch them together in the era of overlap, when the universe is still radiation dominated and the perturbation is already deep inside the horizon. The two cases of long and short wavelength are considered in Sections 6.3 and 6.4, respectively.

To calculate the critical wave number $q_{\rm EQ}$ for which $q/a = H$ at matter-radiation equality, we recall that the redshift of matter-radiation equality is given by Eq. (2.2.5) as $1 + z_{\rm EQ} = \Omega_M/\Omega_R = \Omega_M h^2/4.15 \times 10^{-5}$.

Eq. (2.2.9) gives the Hubble rate during the radiation dominated era as $H = 2.1 \times 10^{-20}(1+z)^2 s^{-1}$, but at radiation-matter equality the contribution of matter to the total energy density makes H larger by a factor $\sqrt{2}$, so for fluctuations that just enter the horizon at matter-radiation equality the physical wave number and Hubble rate at that time are given by Eq. (2.2.9) as $q_{\rm EQ}/a_{\rm EQ} = H_{\rm EQ} = \sqrt{2} \times 2.1 \times 10^{-20}\,(1 + z_{\rm EQ})^2\, s^{-1} = 1.72 \times 10^{-11}\,(\Omega_M h^2)^2$ s^{-1}.
This corresponds to a critical physical wavelength at present given by

$$\lambda_0 \equiv \frac{2\pi}{q_{\rm EQ}/a_0} = \frac{2\pi\,(1 + z_{\rm EQ})}{q_{\rm EQ}/a_{\rm EQ}} = 85\left(\Omega_M h^2\right)^{-1} {\rm Mpc} \tag{6.2.45}$$

For comparison, the size of the local supercluster, estimated from the distance between our galaxy and the Virgo cluster, is about 15 Mpc . Perturbations that are now observed to extend over distances that are larger or smaller than $\lambda_{\rm EQ}$ came within the horizon before or after matter-radiation equality, respectively. Using the present mass density ρ_{M0} given by Eq. (1.5.28) and (1.5.39), the average mass now contained within a sphere of diameter λ_0 is

$$\frac{\pi}{6}\rho_{M0}\lambda_0^3 = 0.9 \times 10^{17}\left(\Omega_M h^2\right)^{-2} M_\odot \tag{6.2.46}$$

This may be compared with the mass of a large galaxy, about $10^{12} M_\odot$. Thus any perturbation that is relevant to the formation of galaxies or even clusters of galaxies would have been well within the horizon at the time of radiation-matter equality.

We can also identify a corresponding critical multipole order $\ell_{\rm EQ}$ of anisotropies in the cosmic microwave background. As remarked in Section 2.6, the integral over wave numbers for the multipole coefficient C_ℓ is dominated by co-moving wave numbers of order ℓ/r_L, where r_L is the radial coordinate of the surface of last scattering, and hence the integral for C_ℓ is dominated by wave numbers of order $q_{\rm EQ}$ that just come into the horizon at matter-radiation equality if ℓ is of order $\ell_{\rm EQ} = q_{\rm EQ} r_L$. This can be written

$$\ell_{\rm EQ} = \left(\frac{q_{\rm EQ}}{a_{\rm EQ}}\right)\left(\frac{a_{\rm EQ}}{a_0}\right)\left(\frac{a_0}{a_L}\right) a_L r_L = \frac{H_{\rm EQ}\,(1 + z_L)}{(1 + z_{\rm EQ})} a_L r_L$$

where a_L is the Robertson-Walker scale factor at last scattering. We recall that $H_{\rm EQ} = \sqrt{2\Omega_M}\,(1 + z_{\rm EQ})^{3/2}\, H_0$, and $a_L r_L = d_A$, the angular diameter distance of the surface of last scattering. Also, $1 + z_{\rm EQ} = \Omega_M/\Omega_R$. Hence

$$\ell_{\rm EQ} = \Omega_M \sqrt{2/\Omega_R} H_0 d_A\,(1 + z_L) \tag{6.2.47}$$

For instance, if we take sample parameters $\Omega_M = 0.26$ and $\Omega_\Lambda = 0.74$, then as noted in Section 2.6 $d_A = 3.38 H_0^{-1}\,(1 + z_L)^{-1}$. Taking $\Omega_R = 8.01 \times 10^{-5}$ (corresponding to $T_0 = 2.725$ K and $H_0 = 72$ kmsec^{-1}Mpc^{-1}), Eq. (6.2.47) gives a critical multipole order $\ell_{\rm EQ} = 140$.

Multipole coefficients for larger values of ℓ arise only from perturbations that entered the horizon during the radiation dominated era.

In what follows, for both scalar and tensor modes,we will find it convenient to introduce a dimensionless rescaled wave number,

$$\kappa \equiv \frac{\sqrt{2}q}{q_{\rm Eq}} = \frac{(q/a_0)\sqrt{\Omega_R}}{H_0\Omega_M} = \frac{q/a_0}{0.052\Omega_M h^2 {\rm Mpc}^{-1}}. \tag{6.2.48}$$

In the cases of long and short wavelength, we have $\kappa \ll 1$ and $\kappa \gg 1$, respectively.

The calculations of the next three sections are necessarily complicated and perhaps tedious. As a guide, it may help to say that the results at which we are aiming are Eqs. (6.5.15) and (6.5.16) for the perturbations to the dark matter density and gravitational field in the whole of the matter-dominated era, and Eqs. (6.5.17) and (6.5.18) for the perturbations to the photon and baryonic plasma density and velocity potential in the matter-dominated era, up to the time of the decoupling of matter and radiation.

10.6.4 6.3 Scalar perturbations - long wavelengths

We first consider perturbations with wavelengths that are long enough so that they are still outside the horizon at the time of radiation-matter equality. As discussed at the end of the previous section, such perturbations are responsible for multipole moments of the cosmic microwave background anisotropies with $\ell < 140$. Because $q\bar\rho_R/aH\bar\rho_M$ is constant during the radiation dominated era, when $a \propto \sqrt{t}$ and $H \propto a^{-2}$, and we are here assuming it is much less than one when $\bar\rho_M = \bar\rho_R$, it follows that for these wavelengths we have

$$\frac{q}{aH} \ll \frac{\bar\rho_M}{\bar\rho_R} \tag{6.3.1}$$

throughout the radiation-dominated era. (Recall that $\bar\rho_M \equiv \bar\rho_D + \bar\rho_B$ and $\bar\rho_R \equiv \bar\rho_\gamma + \bar\rho_\nu$.) On the other hand, during the matter-dominated era when $a \propto t^{2/3}$ and $H \propto a^{-3/2}$, it is $q^2\bar\rho_R/a^2H^2\bar\rho_M$ that remains constant, and since this quantity is assumed to be much less than one when $\bar\rho_M = \bar\rho_R$, for these wavelengths we have

$$\frac{q^2}{a^2H^2} \ll \frac{\bar\rho_M}{\bar\rho_R} \tag{6.3.2}$$

throughout the matter-dominated era.

We cannot give a single analytic formula for these perturbations during the whole era from just after electron-positron annihilation until near the present, but fortunately we can find analytic solutions in two eras: first, the era when the perturbations are outside the horizon, and, second, the era when the energy density of the universe is dominated by non-relativistic matter. (The anisotropic inertia due to neutrinos can be neglected in both eras, because it is negligible outside the horizon, and irrelevant when the energy density of neutrinos is much less than that of matter.) For the long wavelengths considered in this section, that are still outside the horizon at the time that the matter density becomes equal to that of radiation, these eras overlap. This allows us to take the initial condition in the second era from the results for the first era in the period in which they overlap.

A. Outside the horizon

The perturbations of greatest interest, corresponding to adiabatic modes, are governed by Eqs. (6.2.39)-(6.2.44), with the fractional perturbations $\delta_{\alpha q}$ defined by Eq. (6.2.8), and with $\delta_{Bq} = \delta_{\gamma q}$. For $q/a \ll H$, this gives

$$\frac{d}{dt}\left(a^2\psi_q\right) = -4\pi G a^2\left[\bar{\rho}_D\delta_{Dq} + \left(\bar{\rho}_B + \frac{8}{3}\bar{\rho}_\gamma\right)\delta_{\gamma q} + \frac{8}{3}\bar{\rho}_\nu\delta_{\nu q}\right] \tag{6.3.3}$$

$$\dot{\delta}_{\gamma q} = \dot{\delta}_{\nu q} = \dot{\delta}_{Dq} = -\psi_q \tag{6.3.4}$$

$$\frac{d}{dt}\left(\frac{(1+R)\delta u_{\gamma q}}{a}\right) = -\frac{1}{3a}\delta_{\gamma q} \tag{6.3.5}$$

$$\frac{d}{dt}\left(\frac{\delta u_{\nu q}}{a}\right) = -\frac{1}{3a}\delta_{\nu q} \tag{6.3.6}$$

These can be solved analytically, at least for some of the modes described in the previous section. Most importantly, for the adiabatic solutions with all $\delta_{\alpha q}$ equal outside the horizon, Eqs. (6.3.3) and (6.3.4) become

$$\frac{d}{dt}\left(a^2\psi_q\right) = -4\pi G a^2\left[\bar{\rho}_M + \frac{8}{3}\bar{\rho}_R\right]\delta_q \tag{6.3.7}$$

$$\dot{\delta}_q = -\psi_q \tag{6.3.8}$$

where

$$\delta_q \equiv \delta_{\gamma q} = \delta_{\nu q} = \delta_{Bq} = \delta_{Dq} \tag{6.3.9}$$

and we recall that $\bar{\rho}_M \equiv \bar{\rho}_B + \bar{\rho}_D$ and $\bar{\rho}_R \equiv \bar{\rho}_\gamma + \bar{\rho}_\nu$. Inserting Eq. (6.3.8) in (6.3.7) then gives a second-order differential equation for δ_q :

$$\frac{d}{dt}\left(a^2\frac{d}{dt}\delta_q\right) = 4\pi G a^2\left[\bar{\rho}_M + \frac{8}{3}\bar{\rho}_R\right]\delta_q \tag{6.3.10}$$

To solve Eq. (6.3.10), it is very convenient to replace the dependent variable t with $y \equiv a/a_{\rm EQ} = \bar{\rho}_M/\bar{\rho}_R$, where $a_{\rm EQ}$ is the Robertson-Walker scale factor at matter-radiation equality. Then $\bar{\rho}_M = \rho_{\rm EQ}/y^3$ and $\bar{\rho}_R = \rho_{\rm EQ}/y^4$, where $\rho_{\rm EQ}$ is the common density of matter and radiation when they are equal. Using the Friedmann formula for the expansion rate, we have then

$$\frac{d}{dt} = \frac{H_{\rm EQ}}{\sqrt{2}}\frac{\sqrt{1+y}}{y}\frac{d}{dy} \tag{6.3.11}$$

and Eq. (6.3.10) becomes

$$y\sqrt{1+y}\frac{d}{dy}\left(y\sqrt{1+y}\frac{d}{dy}\delta_q\right) - \frac{3}{2}\left(y+\frac{8}{3}\right)\delta_q = 0 \tag{6.3.12}$$

This has two independent solutions,

$$\delta^{(a)} = y^{-2}\left(16 + 8y - 2y^2 + y^3\right), \quad \delta^{(b)} = y^{-2}\sqrt{1+y}$$

We are looking for a solution that, as found for Mode 1 in the previous section, vanishes like $t \propto y^2$ for $t \to 0$, so we must take our solution to be proportional to $\delta^{(a)} - 16\delta^{(b)}$, which for $y \to 0$ approaches $5y^2/8$. Adjusting the normalization of this solution to match Eq. (6.2.27), we have then

$$\delta_q = \frac{4q^2\mathcal{R}_q^o}{5H_{\rm EQ}^2 a_{\rm EQ}^2 y^2}\left(16+8y-2y^2+y^3-16\sqrt{1+y}\right), \tag{6.3.13}$$

$$\delta u_{\gamma q} = -\frac{\sqrt{2}y}{3H_{\rm EQ}(1+R)}\int_0^y \frac{dy'}{\sqrt{1+y'}}\delta_q\left(y'\right), \tag{6.3.14}$$

$$\delta u_{\nu q} = -\frac{\sqrt{2}y}{3H_{\rm EQ}}\int_0^y \frac{dy'}{\sqrt{1+y'}}\delta_q\left(y'\right), \tag{6.3.15}$$

$$\psi_q = \frac{\sqrt{2}q^2\mathcal{R}_q^o}{5H_{\rm EQ}a_{\rm EQ}^2 y^4}\left(2\sqrt{1+y}\left(32+8y-y^3\right)-64-48y\right). \tag{6.3.16}$$

With this normalization factor, $\mathcal{R}_q$ takes the time-independent value $\mathcal{R}_q^o$ outside the horizon.

B. The matter-dominated era

We can also solve Eqs. (6.2.39)-(6.2.44) analytically in the matter-dominated era, when $\bar\rho_M \gg \bar\rho_R$, whether or not the perturbation is outside the horizon. For simplicity, we will also assume that $\bar\rho_B \ll \bar\rho_D$. (Their ratio is actually
about 0.2.) However, since $\bar\rho_\gamma$ is also much less than $\bar\rho_D$ in the matterdominated era, we will not assume that $\bar\rho_B$ is negligible compared with $\bar\rho_\gamma$ in this era. Recall that in the matter-dominated era ${}^1a \propto t^{2/3}$ and $8\pi G\bar\rho_M/3 = H^2 = 4/9t^2$. If we keep only the term in Eq. (6.2.39) proportional to $\bar\rho_D$, but now make no assumption about the relative magnitude of q/a and H, then Eqs. (6.2.39)-(6.2.44) become

$$\frac{d}{dt}\left(t^{4/3}\psi_q\right) = -\frac{2}{3}t^{-2/3}\delta_{Dq} \tag{6.3.17}$$

$$\dot\delta_{\gamma q} - \left(q^2/a^2\right)\delta u_{\gamma q} = -\psi_q \tag{6.3.18}$$

$$\dot\delta_{\nu q} - \left(q^2/a^2\right)\delta u_{\nu q} = -\psi_q \tag{6.3.19}$$

$$\dot\delta_{Dq} = -\psi_q \tag{6.3.20}$$

$$\frac{d}{dt}\left(t^{-2/3}(1+R)\delta u_{\gamma q}\right) = -\frac{1}{3}t^{-2/3}\delta_{\gamma q} \tag{6.3.21}$$

$$\frac{d}{dt}\left(t^{-2/3}\delta u_{\nu q}\right) = -\frac{1}{3}t^{-2/3}\delta_{\nu q} \tag{6.3.22}$$

where once again $R \equiv 3\bar\rho_B/4\bar\rho_R \propto a$.
There are two solutions to Eqs. (6.3.17) and (6.3.20), one with $\psi_q \propto t^{-1/3}$ and $\delta_{Dq} \propto t^{2/3}$, the other with $\psi_q \propto t^{-2}$ and $\delta_{Dq} \propto t^{-1}$. To evaluate the coefficients of these solutions in mode 1 , we must compare these results with those given by Eqs. (6.3.13) and (6.3.16) in the era where both sets of results apply, the era (which exists because of our assumption of long wavelength) when both $q/a \ll H$ and $\bar\rho_R \ll \bar\rho_M$, i. e., $y \gg 1$. Eqs. (6.3.13) and (6.3.16) give in mode 1

$$\delta_{Dq} = \frac{4q^2 y\mathcal{R}_q^o}{5H_{\rm EQ}^2 a_{\rm EQ}^2} = \frac{9q^2t^2\mathcal{R}_q^o}{10a^2} \tag{6.3.23}$$

$$\psi_q = \frac{-2\sqrt{2}q^2\mathcal{R}_q^o}{5H_{\rm EQ}a_{\rm EQ}^2 y^{1/2}} = -\frac{3q^2t\mathcal{R}_q^o}{5a^2} \tag{6.3.24}$$

(The final expression in Eqs. (6.3.23)-(6.3.24) is derived using the result that for $\bar\rho_R \ll \bar\rho_D$, the Hubble rate is $H = 2/3t = H_{\rm EQ}/\sqrt{2}y^{3/2}$, and eliminating $H_{\rm EQ}$.) For $a \propto t^{2/3}$, these match

the solution with $\delta_{Dq} \propto t^{2/3}$ and $\psi_q \propto t^{-1/3}$, while it is easy to see that Mode 2 gives the other solution in the matter-dominated era, the one with $\psi_q \propto t^{-2}$ and $\delta_{Dq} \propto t^{-1}$. We conclude that if Mode 1 is dominant outside the horizon, then the dark

matter density perturbation and gravitational field perturbation are given in the matter-dominated era by Eqs. (6.3.23) and (6.3.24).

This is a good place to pause, and make contact with the more elementary treatment of Section 2.6. Under the assumption that the fluctuation in the Newtonian gravitational potential $\phi(\mathbf{x}, t)$ at around the time of last scattering is dominated by fluctuations in the dark matter density, its Fourier transform is given by using Poisson's equation, with the Laplacian replaced with $-q^2/a^2$:

$$\delta\phi(\mathbf{q}, t) = -4\pi G \left(a^2(t)/q^2\right) \delta\rho_D(\mathbf{q}, t) = -4\pi G \left(a^2(t)/q^2\right) \bar{\rho}_D(t)\delta_{Dq}(t)$$

Using the Mode 1 solution Eq. (6.3.23) and the Friedmann equation $H^2 = (2/3t)^2 = 8\pi G\bar{\rho}_D/3$, we see that $\delta\phi(\mathbf{q}, t)$ takes the time-independent value

$$\delta\phi(\mathbf{q}) = -\frac{3}{5}\alpha(\mathbf{q})\mathcal{R}_q^o$$

Then $\langle\delta\phi(\mathbf{q})\delta\phi\left(\mathbf{q}'\right)\rangle = \mathcal{P}_\phi(q)\delta^3\left(\mathbf{q}+\mathbf{q}'\right)$, with the correlation function $\mathcal{P}_\phi$ introduced in Section 2.6 equal to

$$\mathcal{P}_\phi(q) = \frac{9}{25}\left|\mathcal{R}_q^o\right|^2.$$

The behavior $\mathcal{P}_\phi(q) = N_\phi^2 q^{-3}$ that was found in Section 2.6 to yield a temperature multipole coefficient $C_\ell = 8\pi N_\phi^2/9\ell(\ell+1)$ thus corresponds to the assumption that $\left|\mathcal{R}_q^o\right|^2 = N^2 q^{-3}$, with $N_\phi^2 = 9N^2/25$.

Even if we assume that the gravitational field is dominated by cold dark matter, we must still consider the perturbations to the photon density and the photon-baryon velocity as preparation for calculating the contribution of intrinsic temperature fluctuations and Doppler effect to the cosmic microwave background anisotropies in the next chapter. Equations (6.3.18) and (6.3.21) are a pair of coupled inhomogeneous first-order differential equations for the two unknowns $\delta_{\gamma q}$ and $\delta u_{\gamma q}$, with a forcing term proportional to ψ_q. Surprisingly, with ψ_q given by Eq. (6.3.24), there is a simple exact solution of these equations:

$$\delta_{\gamma q}^{(1)} = \frac{3q^2t^2(1+3R)\mathcal{R}_q^o}{5a^2\left(t^2q^2/a^2+2R\right)}, \quad \delta u_{\gamma q}^{(1)} = -\frac{3t^3q^2\mathcal{R}_q^o}{5a^2\left(t^2q^2/a^2+2R\right)} \tag{6.3.25}$$

(We will not be concerned with the remaining equations, (6.3.19) and (6.3.22), which are needed only to calculate the neutrino perturbations.)

To the particular solution (6.3.25), we must add a suitable solution of the homogeneous version of Eqs. (6.3.18) and (6.3.21),

$$\dot{\delta}_{\gamma q}^{(2)} = \left(q^2/a^2\right)\delta u_{\gamma q}^{(2)}, \quad \frac{d}{dt}\left(t^{-2/3}(1+R)\delta u_{\gamma q}^{(2)}\right) = -\frac{1}{3}t^{-2/3}\delta_{\gamma q}^{(2)}, \tag{6.3.26}$$

or, eliminating the velocity potential,

$$\frac{d}{dt}\left(t^{-2/3}(1+R)a^2\frac{d}{dt}\delta_{\gamma q}^{(2)}\right) + \frac{q^2}{3}t^{-2/3}\delta_{\gamma q}^{(2)} = 0 \tag{6.3.27}$$

[471] The zero of time is chosen in this subsection so that in the matter-dominated era $a \propto t^{2/3}$. In other words, t could here be defined as $2/3H$.

with coefficients chosen so that $\delta^{(1)}_{\gamma q} + \delta^{(2)}_{\gamma q}$ matches the solution found earlier outside the horizon, when $q^2t^2/a^2 \ll 1$.

Using the fact that in the matter dominated era $R \propto a \propto t^{2/3}$, we can find a general solution of Eq. (6.3.27) as a linear combination of the functions

$$F\left(\tfrac{1}{4} - \tfrac{1}{4}\sqrt{1-16\eta}, \tfrac{1}{4} + \tfrac{1}{4}\sqrt{1-16\eta}, \tfrac{1}{2}, -R\right),$$
$$\sqrt{R}F\left(\tfrac{3}{4} - \tfrac{1}{4}\sqrt{1-16\eta}, \tfrac{3}{4} + \tfrac{1}{4}\sqrt{1-16\eta}, \tfrac{3}{2}, -R\right),$$

where F is the Gauss hypergeometric function (also known as ${}_2F_1$), and η is the quantity

$$\eta \equiv \frac{3q^2t^2}{4a^2R}$$

which is time-independent during the matter-dominated era. Unfortunately, this does not provide much insight into the behavior of the solutions. Instead, at this point we will make the further assumption that the wavelength is only moderately long, in the sense that $\eta \gg 1$. That is, although q is small enough so that $q/aH \ll 1$ at matter-radiation equality, we assume that it is sufficiently large so that throughout the matter-dominated era we have

$$R \ll t^2q^2/a^2 \ll \bar{\rho}_M/\bar{\rho}_R \tag{6.3.28}$$

Each term in this inequality is proportional to $t^{2/3}$ during the matterdominated era, so if Eq. (6.3.28) holds at any time during this era, then it holds throughout it. This assumption will allow us to find solutions of Eq. (6.3.27) in terms of elementary functions. In the opposite case of extremely long wavelengths, for which $t^2q^2/a^2 \ll R$ throughout the matterdominated era, we can solve the general solution of Eqs. (6.3.18) and (6.3.21)
as a power series in the quantity t^2q^2/a^2R. We are more interested here in the case of moderately long wavelengths, because as we will see in Section 6.5, it is this case that can be connected to the case of short wavelengths by a smooth extrapolation.

For perturbations satisfying the inequality (6.3.28), when the perturbation is outside the horizon and for some time after it re-enters the horizon, we will have $R \ll 1$. During the period when $R \ll 1$, the homogeneous equations (6.3.26) have the exact solution:

$$\delta^{(2)}_{\gamma q} = c_q \cos(\sqrt{3}qt/a) + d_q \sin(\sqrt{3}qt/a) \tag{6.3.29}$$
$$\delta u^{(2)}_{\gamma q} = \frac{a}{\sqrt{3}q}\left[-c_q \sin(\sqrt{3}qt/a) + d_q \cos(\sqrt{3}qt/a)\right] \tag{6.3.30}$$

with c_q and d_q constant. To this, we must add the inhomogeneous solution (6.3.25), in the limit $R \ll q^2t^2/a^2$:

$$\delta^{(1)}_{\gamma q} = \frac{3(1+3R)\mathcal{R}^o_q}{5}, \quad \delta u^{(1)}_{\gamma q} = -\frac{3t\mathcal{R}^o_q}{5}. \tag{6.3.31}$$

We can evaluate the constants c_q and d_q by requiring that for $qt/a \ll 1$ (which according to Eq. (6.3.28) also implies that $R \ll 1$), the total photon density perturbation $\delta^{(1)}_{\gamma q} + \delta^{(2)}_{\gamma q}$ must approach $\delta_{Dq} \to 9q^2t\mathcal{R}^o_q/10a^2$. This gives

$$c_q = -\frac{3\mathcal{R}^o_q}{5}, \quad d_q = 0 \tag{6.3.32}$$

so that, as long as $R \ll 1$,

$$\delta^{(2)}_{\gamma q} = -\frac{3\mathcal{R}^o_q}{5}\cos(\sqrt{3}qt/a), \tag{6.3.33}$$

Eventually R becomes non-negligible, but under the assumption (6.3.28), by then qt/a will be much larger than one, and we can solve the homogeneous equations (6.3.26) using the WKB approximation. Inspection of Eq. (6.3.27) suggests that for $qt/a \gg 1$ the density fluctuation will oscillate rapidly, with phase

$$\varphi \equiv \int_0^t \frac{q dt}{a\sqrt{3(1+R)}} = \frac{\sqrt{3}qt}{a\sqrt{R}} \ln(\sqrt{R} + \sqrt{1+R}) \tag{6.3.34}$$

Using φ as the independent variable instead of t, and recalling that during the matter-dominated era $a \propto t^{2/3}$, Eq. (6.3.27) becomes

$$\frac{d^2\delta_{\gamma q}}{d\varphi^2} + \frac{1}{2}\left(\frac{d\ln(1+R)}{d\varphi}\right)\frac{d\delta_{\gamma q}}{d\varphi} + \delta_{\gamma q} = 0.$$

We try for a solution of the form $\mathcal{A}e^{\pm i\varphi}$, with $\mathcal{A}$ varying slowly with φ, so that we can neglect the second derivative of $\mathcal{A}$ in the first term on the left and neglect the first derivative of $\mathcal{A}$ in the second term on the left, which is already small because of the factor $d\ln(1+R)/d\varphi$. This gives $dA/d\varphi \simeq -(A/4)d\ln(1+R)/d\varphi$, and hence the general WKB solutions of Eq. (6.3.27) are

$$\delta_{\gamma q}^{(2\pm)} \propto (1+R)^{-1/4}\exp(\pm i\varphi).$$

Clearly, the linear combination of these solutions that merge smoothly with the results for $R \ll 1$ is obtained by replacing the argument of the cosine in Eq. (6.3.33) with φ, and multiplying with $(1+R)^{-1/4}$. Adding the inhomogeneous term (6.3.25), the total photon and baryon fractional density perturbations for moderately long wavelengths in the matter-dominated era are

$$\delta_{\gamma q} = \delta_{Bq} = \frac{3\mathcal{R}_q^o}{5}\left[1 + 3R - (1+R)^{-1/4}\cos\varphi\right] \tag{6.3.35}$$

We can then use Eqs. (6.3.18) and (6.3.24) to calculate the velocity potential

$$\delta u_{\gamma q} = \frac{3t\mathcal{R}_q^o}{5}\left[-1 + \frac{a}{\sqrt{3}qt(1+R)^{3/4}}\sin\varphi\right] \tag{6.3.36}$$

(Here we neglect a term in square brackets of order Ra^2/t^2q^2.) As a check, note that early in the matter-dominated era, when $qt/a \ll 1$, we have $R \ll 1$, so $\varphi \to \sqrt{3}qt/a \to 0$, and hence Eq. (6.3.35) gives $\delta_{\gamma q} = \delta_{Bq} \to 9q^2t^2\mathcal{R}_q^o/10a^2$, in agreement with Eq. (6.3.23) and the condition that for adiabatic modes all $\delta_{\alpha q}$ are equal outside the horizon. This condition is satisfied by Eq. (6.3.35) even if the inequality (6.3.28) is not satisfied, as long as both $R \ll 1$ and $qt/a \ll 1$.

The results (6.3.35) and (6.3.36) for $\delta_{\gamma q} = \delta_{Bq}$ and $\delta u_{\gamma q}$ in the case of moderately long wavelengths all apply only up to the time of last scattering. On the other hand, to the extent that the energy density after last scattering is dominated by dark matter, δ_{Dq} and ψ_q are unaffected by the decoupling of radiation from the baryonic plasma, and continue to be given by Eq. (6.3.23) and (6.3.24) until either vacuum energy and possibly spatial curvature become significant or the perturbations become too strong to be treated as first-order perturbations.

10.6.5 6.4 Scalar perturbations - short wavelengths

We next consider adiabatic perturbations with wavelengths that are short enough so that they are already well within the horizon at the time of
radiation-matter equality. As discussed at the end of Section 6.2, such perturbations are responsible for multipole moments of the cosmic microwave background anisotropies with $\ell \gg 140$,

and also for the onset of gravitational condensations that lead to the formation of structures on the scale of galaxies or clusters of galaxies. Following the same reasoning as at the beginning of the previous section, for these wavelengths in the radiationdominated era

$$\frac{q}{aH} \gg \frac{\bar{\rho}_M}{\bar{\rho}_R}, \tag{6.4.1}$$

and in the matter-dominated era

$$\frac{q^2}{a^2H^2} \gg \frac{\bar{\rho}_M}{\bar{\rho}_R}, \tag{6.4.2}$$

where as before $\bar{\rho}_M \equiv \bar{\rho}_D + \bar{\rho}_B$ and $\bar{\rho}_R \equiv \bar{\rho}_\gamma + \bar{\rho}_\nu$.
As shown in Section 6.2, because we are considering adiabatic perturbations, the fractional perturbations $\delta_{\alpha q} \equiv \delta\rho_{\alpha q}/\left(\bar{\rho}_\alpha + \bar{p}_\alpha\right)$ are subject to the condition $\delta_{Bq} = \delta_{\gamma q}$. Consequently the perturbations are governed by Eqs. (6.2.39)-(6.2.44). Again, we cannot give a single analytic solution of these equations during the whole era from just after electron-positron annihilation until near the present, but fortunately we can find analytic solutions in two overlapping eras: first, the era when the energy density of the universe is dominated by radiation (photons and neutrinos), and second, the era when perturbations are well within the horizon.

A. The radiation-dominated era

We can solve Eqs. (6.2.39)-(6.2.44) analytically in the radiation-dominated era, when $\bar{\rho}_R \gg \bar{\rho}_M$. We will assume tentatively that at this time the photon and neutrino density fluctuations on the right-hand side of the gravitational field equation (6.2.39) dominate over the dark matter density fluctuations, an assumption we will check later in this section. In this era $a \propto t^{1/2}$, so Eqs. (6.2.39)-(6.2.44) take the form

$$\frac{d}{dt}\left(t\psi_q\right) = -\frac{32\pi G t}{3}\left(\bar{\rho}_\gamma\delta_{\gamma q} + \bar{\rho}_\nu\delta_{\nu q}\right) \tag{6.4.3}$$

$$\dot{\delta}_{\gamma q} - \left(q^2/a^2\right)\delta u_{\gamma q} = -\psi_q \tag{6.4.4}$$

$$\dot{\delta}_{Dq} = -\psi_q \tag{6.4.5}$$

$$\frac{d}{dt}\left(t^{-1/2}\delta u_{\gamma q}\right) = -\frac{1}{3}t^{-1/2}\delta_{\gamma q} \tag{6.4.6}$$

$$\dot{\delta}_{vq} - \left(q^2/a^2\right)\delta u_{vq} = -\psi_q \tag{6.4.7}$$

$$\frac{d}{dt}\left(t^{-1/2}\delta u_{vq}\right) = -\frac{1}{3}t^{-1/2}\delta_{vq} \tag{6.4.8}$$

We are interested in adiabatic solutions for which all $\delta_{\alpha q}$ and $\delta u_{\alpha q}$ become equal at early times, so since the differential equations here are the same for photons and neutrinos, these adiabatic solutions have

$$\delta_{\gamma q} = \delta_{\nu q}, \quad \delta u_{\gamma q} = \delta u_{\nu q}$$

For these modes, Eq. (6.4.3) now simplifies to

$$\frac{d}{dt}\left(t\psi_q\right) = -\frac{32\pi G\bar{\rho}_R t}{3}\delta_{\gamma q} = -\frac{1}{t}\delta_{\gamma q}$$

Also, assuming that the cosmological perturbations are in the growing adiabatic mode, Mode 1, they satisfy the initial conditions (6.2.27)-(6.2.30): For $q/aH \ll 1$

$$\delta_{\gamma q} = \delta_{Bq} = \delta_{\nu q} \to \frac{q^2 t^2 \mathcal{R}_q^o}{a^2}, \quad \delta_{Dq} \to \frac{q^2 t^2 \mathcal{R}_q^o}{a^2} \tag{6.4.9}$$

$$\psi_q \to -\frac{t q^2 \mathcal{R}_q^o}{a^2}, \quad \delta u_{\gamma q} = \delta u_{\nu q} \to -\frac{2 t^3 q^2 \mathcal{R}_q^o}{9 a^2} \tag{6.4.10}$$

where $\mathcal{R}_q^o$ is the value of $\mathcal{R}_q$ outside the horizon. The reader can check that the solution of Eqs. (6.4.3)-(6.4.8) satisfying these initial conditions is [1]

$$\delta_{\gamma q} = \delta_{Bq} = \delta_{\nu q} = 3\mathcal{R}_q^o \left(\frac{2}{\Theta} \sin\Theta - \left(1 - \frac{2}{\Theta^2}\right) \cos\Theta - \frac{2}{\Theta^2} \right) \tag{6.4.11}$$

$$\psi_q = \frac{3\mathcal{R}_q^o}{t} \left(\frac{2}{\Theta} \sin\Theta + \frac{2}{\Theta^2} \cos\Theta - \frac{2}{\Theta^2} - 1 \right) \tag{6.4.12}$$

$$\delta_{Dq} = -6\mathcal{R}_q^o \int_0^\Theta \left(\frac{2}{\vartheta^3} \sin\vartheta + \frac{2}{\vartheta^4} \cos\vartheta - \frac{2}{\vartheta^4} - \frac{1}{\vartheta^2} \right) \vartheta d\vartheta \tag{6.4.13}$$

$$\delta u_{\gamma q} = \delta u_{\nu q} = 4t\mathcal{R}_q^0 \left(\frac{\sin\Theta}{2\Theta} - \frac{1 - \cos\Theta}{\Theta^2} \right) \tag{6.4.14}$$

where

$$\Theta \equiv \frac{2qt}{\sqrt{3}a} \tag{6.4.15}$$

Note that the fractional perturbations $\delta_{\alpha q}$ are all of the same order of magnitude for moderate values of Θ. This justifies the neglect of the matter terms in Eq. (6.2.39) when $\bar{\rho}_R$ is much greater than $\bar{\rho}_D$ and q/aH is not very much larger than unity. The condition for continuing to neglect the matter terms in Eq. (6.2.39) for perturbations deep inside the horizon, when Θ becomes large, will be discussed below.

B. Deep inside the horizon

We can also find a solution of Eqs. (6.2.39)-(6.2.44) when the wavelength is well within the horizon, in the sense that $q/a \gg H$, whether or not $\bar{\rho}_D$ is negligible compared with $\bar{\rho}_R$. For $q/a \gg H$ we can distinguish two different kinds of solutions: "slow modes," for which time derivatives yield factors of order H, and "fast modes," for which time derivatives acting on the perturbations yield factors of order of q/a, as well as other terms with factors of H instead. Eqs. (6.2.39)-(6.2.44) are a sixth-order system of differential equations, so they have six independent solutions. We are going to identify four independent fast solutions and two independent slow solutions, so we can be sure that there are no solutions other than what we have called fast and slow modes.

[471] By replacing t with Θ as the dependent variable, Eqs. (6.4.3), (6.4.4), and (6.4.6) are put in the form of a parameter-free third-order system of differential equations for $\delta_{\gamma q} = \delta_{Bq} = \delta_{\nu q}, \delta u_{\gamma q}/t = \delta u_{\nu q}/t$ and $t\psi_q$. After finding the solution that matches the initial conditions (6.4.9), (6.4.10), we can solve Eq. (6.4.5) for δ_{Dq} by an integration. Aside from normalization, this solution for the various fluctuations is equivalent to that given for the Newtonian potential in a different gauge in Eq. (48) of S. Bashinsky & E. Bertschinger Phys. Rev. D 65, 123008 (2002).

1. Fast modes

Up to now we have ignored the neutrino anisotropic inertia, but for the rapidly oscillating fast modes we must take into account its effect of the long neutrino mean free path in damping the neutrino density and velocity perturbations. In considering the fast modes deep inside the horizon, we shall simply assume that this damping allows us to ignore δ_{vq} in Eq. (6.2.39). Turning to the other perturbations in a fast mode with fractional rates of change of order q/a, Eq. (6.2.41) shows that δ_{Dq} is of order $\psi_q/(q/a)$, so the dark-matter term on the right-hand side of Eq. (6.2.39) is of order

$$4\pi G a^2 \overline{\rho_D}\psi_q/(q/a) \leq 3H^2 a^2 \psi_q/2(q/a)$$

while the left-hand side of Eq. (6.2.39) is of order $(q/a)a^2\psi_q$, and the dark matter term is therefore less than the left-hand side of Eq. (6.2.39) by a factor less than of order $\leq H^2/(q/a)^2$, which deep in the horizon is much less than unity. Dropping the dark matter term in Eq. (6.2.39), we see that $\delta_{\gamma q}$ must be at least of order $(q/a)\psi_q/H^2$. Eq. (6.2.43) then shows that $\delta u_{\gamma q}$ is of order $\delta_{\gamma q}/(q/a)$, and hence at least of order ψ_q/H^2. Both terms
on the left-hand side of Eqs. (6.2.42) are then larger than ψ_q by factors of order $q^2/a^2H^2 \gg 1$, so we can drop ψ_q on the right-hand side. (That is, because the wavelength is short, pressure gradients exert a much larger force on the baryon-photon plasma than gravitation.) Hence if we neglect all terms in Eqs. (6.2.40), (6.2.43), (6.2.39), and (6.2.41) that are suppressed by factors $H^2/(q/a)^2$ (but not terms arising from derivatives of a that are only suppressed by factors $H/(q/a)$), these equations become

$$\dot{\delta}_{\gamma q} = \left(q^2/a^2\right)\delta u_{\gamma q} \tag{6.4.16}$$

$$\frac{d}{dt}\left(\frac{(1+R)\delta u_{\gamma q}}{a}\right) = -\frac{1}{3a}\delta_{\gamma q} \tag{6.4.17}$$

$$\frac{d}{dt}\left(a^2\psi_q\right) = -\frac{16\pi G a^2}{3}\bar{\rho}_\gamma(R+2)\delta_{\gamma q} \tag{6.4.18}$$

$$\dot{\delta}_{Dq} = -\psi_q \tag{6.4.19}$$

where, as before, $R \equiv 3\bar{\rho}_B/4\bar{\rho}_\gamma$. Equations (6.4.16) and (6.4.17) have two independent solutions for $\delta_{\gamma q}$ and $\delta u_{\gamma q}$. Given these solutions, and looking only for fast modes, Eq. (6.4.18) then has a unique solution for ψ_q, and Eq. (6.4.19) then has a unique solution for δ_{Dq}. (Possible constant terms that might be included in the solutions for $a^2\psi_q$ or δ_{Dq} would contribute to the slow modes, not the fast modes.) Together with the two strongly damped solutions of Eqs. (6.2.42) and (6.2.44) for the neutrino perturbations, there are four independent fast modes, as promised.

By eliminating $\delta u_{\gamma q}$ from Eqs. (6.4.16) and (6.4.17), we obtain a secondorder differential equation for $\delta_{\gamma q}$ alone

$$\frac{d}{dt}\left(a(1+R)\frac{d}{dt}\delta_{\gamma q}\right) + \frac{q^2}{3a}\delta_{\gamma q} = 0 \tag{6.4.20}$$

If a and R were constant, this would be just the wave equation for a sound wave, with physical wave number q/a and velocity ${}^2v_s = 1/\sqrt{3(1+R)}$. With a and R varying at a relatively slow fractional rate H, Eq. (6.4.20) can be solved for $q/a \gg H$ by the WKB approximation. For this purpose, we introduce a new independent variable, the phase $\varphi \equiv q\int_0^t dt/a\sqrt{3(1+R)}$, and rewrite Eq. (6.4.20) as

$$\frac{d^2\delta_{\gamma q}}{d\varphi^2} + \frac{1}{2}\frac{d\ln(1+R)}{d\varphi}\frac{d\delta_{\gamma q}}{d\varphi} + \delta_{\gamma q} = 0$$

Writing $\delta_{\gamma q} = \mathcal{A}\exp(\pm i\varphi)$, and neglecting $d^2\mathcal{A}/d\varphi^2$ in the first term on the left and neglecting $d\mathcal{A}/d\varphi$ in the second term (which is already small because of the factor $d\ln(1+R)/d\varphi$), we find $d\mathcal{A}/d\varphi = -(\mathcal{A}/4)d\ln(1+R)/d\varphi$, so the WKB solutions are [3]

$$\delta^{\pm}_{\gamma q} = (1+R)^{-1/4}\exp\left[\pm iq\int_0^t \frac{dt}{a\sqrt{3(1+R)}}\right] \tag{6.4.21}$$

There is a further complication that must be taken into account for fast modes, though it is not important for slow modes. The amplitude of a sound wave whose physical wave number k is larger than the inverse mean path of the particles in a relativistic medium is damped by viscosity and heat conduction, at a rate given in general by [4]

$$\Gamma = \frac{k^2}{2(\rho+p)}\left\{\zeta + \frac{4}{3}\eta + \chi\left(\frac{\partial\rho}{\partial T}\right)_n^{-1} \right.$$
$$\left.\left[\rho + p - 2T\left(\frac{\partial p}{\partial T}\right)_n + v_s^2 T\left(\frac{\partial\rho}{\partial T}\right)_n - \frac{n}{v_s^2}\left(\frac{\partial p}{\partial n}\right)_T\right]\right\}, \tag{6.4.22}$$

where η, χ, and ζ are the coefficients of shear viscosity, heat conduction, and bulk viscosity, respectively, defined in Appendix B; v_s is the sound speed; n is any number density on which the fluid properties may depend; and subscripts indicate the quantities held constant in taking partial derivatives. For the baryon-photon fluid, for which n is the baryon number density, we have $\rho = \rho_B + \rho_\gamma, p = \rho_\gamma/3$, with $\rho_B \propto n$ and $\rho_\gamma \propto T^4$. Also, as we have seen $v_s^2 = 1/\sqrt{3(1+R)}$. Hence, (setting $k = q/a$) the damping rate becomes

$$\Gamma = \frac{3q^2}{8a^2\bar{\rho}_\gamma(1+R)}\left\{\zeta + \frac{4}{3}\eta + \frac{\chi T R^2}{3(1+R)}\right\}. \tag{6.4.23}$$

The viscosity and heat conduction coefficients for photons interacting with a non-relativistic plasma with mean free time $t_\gamma = 1/\sigma_{\mathcal{T}} n_e$ are

$$\eta = \frac{16}{45}\bar{\rho}_\gamma t_\gamma, \quad \chi T = \frac{4}{3}\bar{\rho}_\gamma t_\gamma, \quad \zeta = 0 \tag{6.4.24}$$

so in this case the damping rate is [5]

$$\Gamma = \frac{q^2 t_\gamma}{6a^2(1+R)}\left\{\frac{16}{15} + \frac{R^2}{1+R}\right\}. \tag{6.4.25}$$

The effect is to replace Eq. (6.4.21) for the fast mode amplitudes with

$$\delta^{\pm}_{\gamma q} = (1+R)^{-1/4}\exp\left[\pm iq\int_0^t \frac{dt}{a\sqrt{3(1+R)}} - \int_0^t \Gamma dt\right] \tag{6.4.26}$$

This damping of the fast modes is known as Silk damping. [6] With this result, $\delta u^{\pm}_{\gamma q}$for these fast modes can be obtained from Eq. (6.4.16), $\psi^{\pm}_q$can be obtained from Eq. (6.4.18), and then $\delta^{\pm}_{Dq}$can be obtained from Eq. (6.4.19).

[472] Note that the condition of constant entropy gives $d\rho_B/\bar{\rho}_B = d\rho_\gamma/(\bar{\rho}_\gamma + \bar{p}_\gamma) = (3/4)d\rho_\gamma/\bar{\rho}_\gamma$, so $v_s^2 = dp/d\rho = d\rho_\gamma/3(d\rho_\gamma + d\rho_B) = 1/3(1+R)$.

[473] P. J. E. Peebles and J. T. Yu, Astrophys. J. 162, 815 (1970). R. A. Sunyaev and Ya. B. Zel'dovich, Astrophys. & Space Sci. 7, 3 (1970).

[4] S. Weinberg, Astrophys. J. 168, 175 (1971).

2. Slow modes

For solutions whose fractional rate of change is of order $H = O(1/t)$, we can run through the same sort of counting of powers of q/aH as for fast modes, but with very different results. From Eqs. (6.2.41), (6.2.43), and (6.2.44), we see that δ_{Dq} is of order ψ_q/H while $\delta_{\gamma q}$ and $\delta_{\nu q}$ are of order $H\delta u_{\gamma q}$ and $H\delta u_{\nu q}$, respectively. The terms $\dot{\delta}_{\gamma q}$ and $\dot{\delta}_{\nu q}$ on the left-hand side of Eqs. (6.2.40) and (6.2.42) are then of order $H^2\delta u_{\gamma q}$ and $H^2\delta u_{\nu q}$, and hence are less than the terms $(q^2/a^2)\,\delta u_{\gamma q}$ and $(q^2/a^2)\,\delta u_{\nu q}$ by factors of order H^2a^2/q^2, and may be dropped, giving instead

$$\left(q^2/a^2\right)\delta u_{\gamma q} = \left(q^2/a^2\right)\delta u_{\nu q} = \psi_q, \tag{6.4.27}$$

so Eqs. (6.2.43) and (6.2.44) show that $\delta_{\gamma q}$ and $\delta_{\nu q}$ are of order $(a^2H/q^2)\,\psi_q$.
The ratios of the photon and neutrino terms on the right-hand side of Eq. (6.2.39) to the dark matter term are then of order

$$\frac{\text{photons \& neutrinos}}{\text{dark matter}} = O\left[\left(\frac{\bar{\rho}_B + 8\bar{\rho}_R/3}{\bar{\rho}_D}\right)\left(\frac{a^2H^2}{q^2}\right)\right],$$

where as before $\bar{\rho}_R \equiv \bar{\rho}_\gamma + \bar{\rho}_\nu$. The term in the numerator proportional to $\bar{\rho}_B$ contributes much less than unity to this ratio, because $\bar{\rho}_B/\bar{\rho}_D \approx 1/5$ and we are now assuming that $aH/q \ll 1$. According to Eq. (6.4.2), the contribution of the term proportional to $\bar{\rho}_R$ is much less than unity throughout the matter-dominated era. It actually begins to be much less than unity during the radiation-dominated era, when the ratio of radiation density to dark matter density falls below the critical value $(\bar{\rho}_R/\bar{\rho}_D)_{\rm crit}$, given by

$$\left(\frac{\bar{\rho}_R}{\bar{\rho}_D}\right)_{\rm crit} = \left(\frac{3}{8}\right)^{1/3}\left(\frac{q}{aH}\frac{\bar{\rho}_R}{\bar{\rho}_D}\right)^{2/3}. \tag{6.4.28}$$

According to Eqs. (6.4.1), the right-hand side is constant and much greater than unity throughout the radiation-dominated era for the wavelengths considered in this section, so $\bar{\rho}_R/\bar{\rho}_D$ will fall below this critical value well before radiation-matter equality. (That is, strong pressure forces keep perturbations to the baryon-photon plasma density small enough so that their effect on the gravitational field is negligible once $\bar{\rho}_\gamma/\bar{\rho}_B$ falls below the critical value (6.4.28), even though the unperturbed radiation density is at first still larger than the dark matter density.) From then on the photon and neutrino terms may be neglected in Eq. (6.2.39) for the slow modes, yielding

$$\frac{d}{dt}\left(a^2\psi_q\right) = -4\pi G\bar{\rho}_D a^2\delta_{Dq} \tag{6.4.29}$$

The remaining equations, (6.2.41), (6.2.43) and (6.2.44) are unchanged:

[475] This damping rate was first calculated by N. Kaiser, Mon. Not. Roy. Astron. Soc. 202, 1169 (1983), and is derived here in Appendix H. The formulas for the shear viscosity and heat conduction coefficients are obtained by comparing formulas (6.4.23) and (6.4.25) for the acoustic damping rate, taking into account that the bulk viscosity vanishes because energy and momentum are transported by relativistic particles; see ref. 4. The damping rate calculated by ref. 4 had given the correct values for χ and ζ, but it gave a value for η that was 3/4 the correct value of Kaiser, quoted here in Eq. (6.4.24). This was because its results were based on calculations of L. H. Thomas, Quart. J. Math. (Oxford) 1, 239 (1930), that had assumed isotropic scattering and ignored photon polarization. (The same value for η had been given earlier by C. Misner, Astrophys. J. 151, 431 (1968).) Kaiser's results are calculated using the correct differential cross section for Thomson scattering and take photon polarization into account, and therefore supersede the value for η quoted in ref. 4 and Chapter 15 of G&C.

[6] J. Silk, Nature 215, 1155 (1972).

$$\dot{\delta}_{Dq} = -\psi_q \tag{6.4.30}$$

$$\frac{d}{dt}\left(\frac{(1+R)\delta u_{\gamma q}}{a}\right) = -\frac{1}{3a}\delta_{\gamma q} \tag{6.4.31}$$

$$\frac{d}{dt}\left(\frac{\delta u_{\nu q}}{a}\right) = -\frac{1}{3a}\delta_{\nu q} \tag{6.4.32}$$

Using Eq. (6.4.30) in Eq. (6.4.29) yields a second-order differential equation for δ_{Dq} :[7]

$$\frac{d}{dt}\left(a^2\frac{d\delta_{Dq}}{dt}\right) = 4\pi G a^2 \bar{\rho}_D \delta_{Dq}. \tag{6.4.33}$$

It is convenient once again to convert the independent variable from t to $y \equiv a/a_{\rm EQ} = \bar{\rho}_M/\bar{\rho}_R$, using the Friedmann equation

$$\frac{\dot{y}^2}{y^2} = \frac{8\pi G}{3}\left(\bar{\rho}_M + \bar{\rho}_R\right) = \frac{8\pi G \rho_{\rm EQ}}{3}\left(y^{-3} + y^{-4}\right) \tag{6.4.34}$$

with $\rho_{\rm EQ}$ the values of $\bar{\rho}_M$ and $\bar{\rho}_R$ when they are equal. Then (6.4.33) becomes what is sometimes known as the Mészáros equation:

$$y(1+y)\frac{d^2\delta_D}{dy^2} + \left(1+\frac{3y}{2}\right)\frac{d\delta_D}{dy} - \frac{3}{2}(1-\beta)\delta_D = 0 \tag{6.4.35}$$

where $\beta \equiv \bar{\rho}_B/\bar{\rho}_M = \Omega_B/\Omega_M$. The independent solutions of Eq. (6.4.35) for $\beta = 0$ were given by Mészáros, [7] and by Groth and Peebles [8]

$$\delta_{Dq}^{(1)} = 1 + \frac{3y}{2}, \quad \delta_{Dq}^{(2)} = \left(1+\frac{3y}{2}\right)\ln\left(\frac{\sqrt{1+y}+1}{\sqrt{1+y}-1}\right) - 3\sqrt{1+y}. \tag{6.4.36}$$

Subsequently Hu and Sugiyama [9] gave two independent solutions for general β :

$$\delta_{Dq} \propto (1+y)^{-\alpha_\pm} F\left(\alpha_\pm, \alpha_\pm + \frac{1}{2}, 2\alpha_\pm + \frac{1}{2}; \frac{1}{1+y}\right)$$

where F is the Gauss hypergeometric function and

$$\alpha_\pm = \frac{1 \pm \sqrt{1+24\beta}}{4}$$

In order to obtain our final results in an analytic form, we will continue to drop corrections proportional to $\beta \equiv \bar{\rho}_B/\bar{\rho}_M \approx 1/6$, while keeping those proportional to $R \equiv 3\bar{\rho}_B/4\bar{\rho}_\gamma$, so we shall use the slow solutions (6.4.36) for $\beta = 0$. (Corrections for the finite value of β are discussed in the following section.) From these two solutions, we can find unique corresponding slow solutions for $\psi_q, \delta u_{\gamma q}, \delta u_{\nu q}, \delta_{\gamma q}$ and $\delta_{\nu q}$ by successive use of Eqs. (6.4.30), (6.4.27), (6.4.31) and (6.4.32). (The neutrino perturbations are of no known observational interest.) We have thus found two slow modes, giving six in all.

[477] Eq. (6.4.33) was first derived by P. Mészáros, Astron. Astrophys. 37, 225 (1974), who simply ignored fluctuations in the radiation energy density. The argument given here for the neglect of perturbations in the radiation density in Eq. (6.4.33) was given by S. Weinberg, Astrophys. J. 581, 810 (2002). It applies only to the slow mode part of the solution; in the fast mode it is the perturbations in the dark matter density that become negligible for small wavelength. This paper also gives comments on other attempts to justify the neglect of perturbations in the radiation density in Eq. (6.4.33).

C. Matching

Fortunately, for small wavelength there is an overlap in the two eras in which we have found solutions for δ_D, etc., satisfying both conditions $q/a \gg \dot{a}/a$

and $\bar{\rho}_M \ll \bar{\rho}_R$. In this period the solution (6.4.11)-(6.4.14) found for the radiation-dominated era can be decomposed into a fast and a slow mode. Since the variable Θ defined by Eq. (6.4.15) is here much larger than unity, we have

$$\delta^{\rm fast}_{\gamma q} = \delta^{\rm fast}_{Bq} = \delta^{\rm fast}_{\nu q} = 3\mathcal{R}^o_q \left(\frac{2}{\Theta} \sin\Theta - \left(1 - \frac{2}{\Theta^2}\right) \cos\Theta \right),$$
$$\rightarrow -3\mathcal{R}^o_q \cos\Theta \tag{6.4.37}$$

$$\psi^{\rm fast}_q = \frac{3\mathcal{R}^o_q}{t} \left(\frac{2}{\Theta} \sin\Theta + \frac{2}{\Theta^2} \cos\Theta \right) \rightarrow \frac{6\mathcal{R}^o_q}{t\Theta} \sin\Theta, \tag{6.4.38}$$

$$\delta^{\rm fast}_{Dq} = -6\mathcal{R}^o_q \int_{\Theta_1}^{\Theta} \left(\frac{2}{\vartheta^3} \sin\vartheta + \frac{2}{\vartheta^4} \cos\vartheta \right) \vartheta d\vartheta \rightarrow \frac{12\mathcal{R}^o_q}{\Theta^2} \cos\Theta, \tag{6.4.39}$$

$$\delta u^{\rm fast}_{\gamma q} = \delta u^{\rm fast}_{\nu q} = 4t\mathcal{R}^0_q \left(\frac{\sin\Theta}{2\Theta} + \frac{\cos\Theta}{\Theta^2} \right) \rightarrow \frac{2t\mathcal{R}^0_q}{\Theta} \sin\Theta, \tag{6.4.40}$$

and

$$\delta^{\rm slow}_{\gamma q} = \delta^{\rm slow}_{Bq} = \delta^{\rm slow}_{\nu q} = -\frac{6\mathcal{R}^o_q}{\Theta^2} \tag{6.4.41}$$

$$\psi^{\rm slow}_q = -\frac{3\mathcal{R}^o_q}{t} \left(\frac{2}{\Theta^2} + 1 \right) \rightarrow -\frac{3\mathcal{R}^o_q}{t} \tag{6.4.42}$$

$$\delta^{\rm slow}_{Dq} = -6\mathcal{R}^o_q \int_0^{\Theta_1} \left(\frac{2}{\vartheta^3} \sin\vartheta + \frac{2}{\vartheta^4} \cos\vartheta - \frac{2}{\vartheta^4} - \frac{1}{\vartheta^2} \right) \vartheta d\vartheta$$
$$+ 6\mathcal{R}^o_q \left[-\frac{1}{\Theta^2} + \frac{1}{\Theta_1^2} + \ln\left(\Theta/\Theta_1\right) \right]$$
$$\rightarrow 6\mathcal{R}^o_q \left(-\frac{1}{2} + \gamma + \ln\Theta \right) \tag{6.4.43}$$

$$\delta u^{\rm slow}_{\gamma q} = \delta u^{\rm slow}_{\nu q} = -\frac{4t\mathcal{R}^0_q}{\Theta^2} \tag{6.4.44}$$

where again $\Theta \equiv 2qt/\sqrt{3}a$, Θ_1 is any constant in the range $1 \ll \Theta_1 \ll \Theta$, and $\gamma = 0.5772\ldots$ is the Euler constant. [10]

Matching the solutions in the radiation-dominated era and deep within the horizon is straightforward for the fast modes. In the radiation-dominated era, when $R \ll 1$, damping is negligible (because the mean free time is very short), and $a \propto \sqrt{t}$, the argument of the cosine in Eq. (6.4.37) may be expressed as the integral

[478] E. J. Groth and P. J. E. Peebles, Astron. Astrophys. 41, 143 (1975).

[9] W. Hu and N. Sugiyama, Astrophys. J. 471, 542 (1996).

[4710] To evaluate the asymptotic limit of the integral in Eq. (6.4.43), we can rewrite this integral as the sum of three terms, each of which converges at $\vartheta = 0$:

$$\int_0^{\Theta_1} \left(\frac{2}{\vartheta^3} \sin\vartheta + \frac{2}{\vartheta^4} \cos\vartheta - \frac{2}{\vartheta^4} - \frac{1}{\vartheta^2} \right) \vartheta d\vartheta = \int_0^{\Theta_1} \left(\frac{1}{\vartheta^2} \sin\vartheta + \frac{2}{\vartheta^3} \cos\vartheta - \frac{2}{\vartheta^3} \right) d\vartheta$$
$$+ \int_0^{\Theta_1} \left(\frac{\sin\vartheta}{\vartheta} - \frac{1}{1+\vartheta} \right) \frac{d\vartheta}{\vartheta} + \int_0^{\Theta_1} \left(\frac{1}{1+\vartheta} - 1 \right) \frac{d\vartheta}{\vartheta}$$

$$q\int_0^t \frac{dt}{a\sqrt{3(1+R)}} \to \frac{2qt}{a\sqrt{3}} = \Theta$$

Hence the linear combination of the two fast solutions (6.4.26) that fits smoothly with the result (6.4.37) is:

$$\delta_{\gamma q}^{\rm fast} = -\frac{3\mathcal{R}_q^o}{(1+R)^{1/4}} e^{-\int_0^t \Gamma dt} \cos\left(\int_0^t \frac{qdt}{a\sqrt{3(1+R)}}\right) \tag{6.4.45}$$

By a successive use of Eqs. (6.4.16), (6.4.18), and (6.4.19) (and ignoring the time dependence of all factors except the rapidly oscillating sines and cosines), we then also find that for $q/a \gg H$:

$$\delta u_{\gamma q}^{\rm fast} = \frac{a\sqrt{3}\mathcal{R}_q^o}{q(1+R)^{3/4}} e^{-\int_0^t \Gamma dt} \sin\left(\int_0^t \frac{qdt}{a\sqrt{3(1+R)}}\right). \tag{6.4.46}$$

$$\psi_q^{\rm fast} = 16\sqrt{3}\pi G\bar\rho_\gamma(2+R)(1+R)^{1/4}(a/q)\mathcal{R}_q^o e^{-\int_0^t \Gamma dt} \sin\left(\int_0^t \frac{qdt}{a\sqrt{3(1+R)}}\right),$$

$$\delta_{Dq}^{\rm fast} = 48\pi G\bar\rho_\gamma(2+R)(1+R)^{3/4}(a/q)^2\mathcal{R}_q^o e^{-\int_0^t \Gamma dt} \cos\left(\int_0^t \frac{qdt}{a\sqrt{3(1+R)}}\right).$$

We will use Eqs. (6.4.45)-(6.4.48) in dealing with baryon acoustic oscillations in Section 8.1.

The reader can easily check that in the overlap era, when the universe is radiation dominated, the perturbation is deep inside the horizon, and damping is negligible, Eqs. (6.4.45) and (6.4.46) give the same results for $\delta_{\gamma q}^{\rm fast}$ and $\delta u_{\gamma q}^{\rm fast}$ as Eqs. (6.4.37) and (6.4.40). On the other hand, Eqs. (6.4.47) and (6.4.48) give results for $\psi_q^{\rm fast}$ and $\delta_{Dq}^{\rm fast}$ that differ from

Eqs. (6.4.38) and (6.4.39) by a factor $\bar\rho_\gamma/\bar\rho_R$. This is because in deriving Eqs. (6.4.38) and (6.4.39) we treated the neutrinos as a perfect fluid throughout the radiation-dominated era, although this is valid only outside the horizon when anisotropic inertia is negligible, while in deriving Eqs. (6.4.47) and (6.4.48) we assumed that deep inside the horizon neutrino density fluctuations are so damped by anisotropic inertia that they can be neglected as a source of gravitational field perturbations. This discrepancy is a small price to pay for the simplicity gained by these approximations, especially since it will turn out that the slow contributions to ψ_q and δ_{Dq} are much larger than the fast contributions.

Next let us consider the slow modes. Here there is a complication in matching solutions in the radiation-dominated era and deep inside the horizon. In deriving Eqs. (6.4.41)-(6.4.43) we have assumed that in the radiation dominated era we can neglect perturbations in the baryon and dark matter densities as a source of the gravitational field perturbations, but in exploring the solutions deep inside the horizon we found that this assumption is violated for the slow modes once the radiation/dark matter density ratio drops below the limit (6.4.28), even if this ratio is still much larger than unity. Indeed, we can see from Eqs. (6.4.41) and (6.4.43) that for large Θ, $\delta_{Dq}^{\rm slow}/\delta_{\gamma q}^{\rm slow} \to -\Theta^2\ln\Theta$, so the assumption under which Eqs. (6.4.41)(6.4.44) were derived breaks down once $\Theta^2\ln\Theta$ becomes comparable to $\bar\rho_R/\bar\rho_D$, which is close to when $\bar\rho_\gamma/\bar\rho_D$ falls below the critical value (6.4.28). Therefore we have to interpolate between the results (6.4.41)-(6.4.44), which are valid early in the radiation-dominated era, when $q^2t^2/a^2 \ll \bar\rho_R/\bar\rho_D$,

[47] For large Θ_1, these three integrals approach the values $-1/2, 1-\gamma$, and $-\ln\Theta_1$, respectively, giving a total of $1/2-\gamma-\ln\Theta_1$, which when combined with the term on the second line of Eq. (6.4.43) yields the quoted result.

and the results later in the radiation-dominated era and in the matter-dominated era, when the Mészáros equation (6.4.35) applies.

This is easiest for the dark matter density perturbation, because its timedependence turns out to have the same form deep in the horizon in the radiation-dominated era both before and after $\bar{\rho}_R/\bar{\rho}_D$ falls below the critical value (6.4.28). Before this time in the radiation-dominated era $\delta_{Dq}^{\rm slow}$ is given by Eq. (6.4.43). After this time, the gravitational field perturbation becomes dominated by dark matter, and the dark matter density perturbation becomes a linear combination of the solutions (6.4.36) of the Mészáros equation, which in the radiation-dominated era when $y \ll 1$ become

$$\delta_{Dq}^{(1)} \to 1, \quad \delta_{Dq}^{(2)} \to -\ln(y/4) - 3 \tag{6.4.49}$$

The linear combination of these two solutions that fits smoothly with Eq. (6.4.43) is then

$$\delta_{Dq}^{\rm slow} = 6\mathcal{R}_q^0 \left\{ \left[-\frac{7}{2} + \gamma + \ln\left(\frac{4q\sqrt{2}}{\sqrt{3}H_{\rm EQ}a_{\rm EQ}} \right) \right] \delta_{Dq}^{(1)} - \delta_{Dq}^{(2)} \right\} \tag{6.4.50}$$

where $H_{\rm EQ}$ and $a_{\rm EQ}$ are the expansion rate and Robertson-Walker scale factor at matter-radiation equality. (Here we use $t = 1/2H = y^2/\sqrt{2}H_{\rm EQ}$.) The slow part of the gravitational field, radiation velocity potential, and radiation density perturbations are given by a successive use of Eqs. (6.4.30), (6.4.27), and (6.4.31) as

$$\psi_q^{\rm slow} = -\dot{\delta}_{Dq}^{\rm slow} \tag{6.4.51}$$

$$\delta u_{\gamma q}^{\rm slow} = -\left(a^2/q^2\right) \dot{\delta}_{Dq}^{\rm slow} \tag{6.4.52}$$

$$\delta_{\gamma q}^{\rm slow} = \frac{3a}{q^2}\frac{d}{dt}\left(a(1+R)\frac{d}{dt}\delta_{Dq}^{\rm slow} \right). \tag{6.4.53}$$

In particular, in the matter-dominated era we have $y \gg 1, \delta_{Dq}^{(1)} \to 3y/2$, and $\delta_{Dq}^{(2)} \to 4/15y^{3/2}$, so Eqs. (6.4.50)-(6.4.53) become

$$\delta_{Dq}^{\rm slow} \to \frac{9\mathcal{R}_q^0 a}{a_{\rm EQ}} \left[-\frac{7}{2} + \gamma + \ln\left(\frac{4\kappa}{\sqrt{3}} \right) \right], \tag{6.4.54}$$

$$\psi_q^{\rm slow} \to -\frac{6\mathcal{R}_q^0 a}{a_{\rm EQ}t} \left[-\frac{7}{2} + \gamma + \ln\left(\frac{4\kappa}{\sqrt{3}} \right) \right]. \tag{6.4.55}$$

$$\delta u_{\gamma q}^{\rm slow} \to -\frac{6\mathcal{R}_q^0 a^3}{a_{\rm EQ}tq^2} \left[-\frac{7}{2} + \gamma + \ln\left(\frac{4\kappa}{\sqrt{3}} \right) \right], \tag{6.4.56}$$

$$\delta_{\gamma q}^{\rm slow} \to \frac{6\mathcal{R}_q^o a^3(1+3R)}{a_{\rm EQ}t^2q^2} \left[-\frac{7}{2} + \gamma + \ln\left(\frac{4\kappa}{\sqrt{3}} \right) \right]. \tag{6.4.57}$$

We have here again introduced a dimensionless rescaled wave number

$$\kappa \equiv \frac{q\sqrt{2}}{a_{\rm EQ}H_{\rm EQ}} = \frac{(q/a_0)\sqrt{\Omega_R}}{H_0\Omega_M} = \frac{19.3\,(q/a_0)\left[{\rm Mpc}^{-1}\right]}{\Omega_M h^2}, \tag{6.4.58}$$

in which we have used the relations $H_{\rm EQ} = \sqrt{2}\left(H_0\sqrt{\Omega_M}\right)(a_0/a_{\rm EQ})^{3/2}$ and $a_0/a_{\rm EQ} = \Omega_M/\Omega_R$. In terms of the wave number $q_{\rm EQ}$ introduced in Section 6.2, for which the perturbation just comes into the horizon at matter-radiation equality, we have $\kappa = \sqrt{2}q/q_{\rm EQ}$, so the assumption of short wavelength made in this section essentially amounts to the condition that $\kappa \gg 1$.

The full solution up to the time of recombination is given by adding the contributions of the fast mode and slow mode:

$$\begin{aligned}\delta_{Dq} \to & \frac{9\mathcal{R}_q^0 a}{a_{\rm EQ}}\left[-\frac{7}{2}+\gamma+\ln\left(\frac{4\kappa}{\sqrt{3}}\right)\right] \\ & + 48\pi G\bar{\rho}_\gamma(2+R)(1+R)^{3/4}(a/q)^2\mathcal{R}_q^o \\ & \times e^{-\int_0^t \Gamma dt}\cos\left(\int_0^t \frac{qdt}{a\sqrt{3(1+R)}}\right) \end{aligned} \tag{6.4.59}$$

$$\begin{aligned}\psi_q \to & -\frac{6\mathcal{R}_q^0 a}{a_{\rm EQ}t}\left[-\frac{7}{2}+\gamma+\ln\left(\frac{4\kappa}{\sqrt{3}}\right)\right] \\ & + 16\sqrt{3}\pi G\bar{\rho}_\gamma(2+R)(1+R)^{1/4}(a/q)\mathcal{R}_q^o \\ & \times e^{-\int_0^t \Gamma dt}\sin\left(\int_0^t \frac{qdt}{a\sqrt{3(1+R)}}\right) \end{aligned} \tag{6.4.60}$$

$$\begin{aligned}\delta u_{\gamma q} \to & -\frac{6\mathcal{R}_q^0 a^3}{a_{\rm EQ}tq^2}\left[-\frac{7}{2}+\gamma+\ln\left(\frac{4\kappa}{\sqrt{3}}\right)\right] \\ & + \frac{a\sqrt{3}\mathcal{R}_q^o}{q(1+R)^{3/4}}e^{-\int_0^t \Gamma dt}\sin\left(\int_0^t \frac{qdt}{a\sqrt{3(1+R)}}\right) \end{aligned} \tag{6.4.61}$$

$$\begin{aligned}\delta_{Bq}=\delta_{\gamma q} \to & \frac{6\mathcal{R}_q^o a^3(1+3R)}{a_{\rm EQ}t^2q^2}\left[-\frac{7}{2}+\gamma+\ln\left(\frac{4\kappa}{\sqrt{3}}\right)\right] \\ & - \frac{3\mathcal{R}_q^o}{(1+R)^{1/4}}e^{-\int_0^t \Gamma dt}\cos\left(\int_0^t \frac{qdt}{a\sqrt{3(1+R)}}\right) \end{aligned} \tag{6.4.62}$$

with Γ and κ given by Eqs. (6.4.25) and (6.4.58).
Each of the different perturbations (6.4.59)-(6.4.62) is dominated by either its fast or slow term. First, let us consider δ_{Dq} and ψ_q. Comparison of Eqs. (6.4.48) and (6.4.47) with Eqs. (6.4.54) and (6.4.55) shows that the ratios of the fast and slow contributions to δ_{Dq} and ψ_q are of order

$$\frac{\delta_{Dq}^{\rm fast}}{\delta_{Dq}^{\rm slow}} \sim \frac{a^2H^2}{q^2}\left(\frac{\bar{\rho}_R}{\bar{\rho}_M}\right)^2, \quad \frac{\psi_q^{\rm fast}}{\psi_q^{\rm slow}} \sim \frac{aH}{q}\left(\frac{\bar{\rho}_R}{\bar{\rho}_M}\right)^2. \tag{6.4.63}$$

We are assuming in this section that the wavelengths are short enough so that perturbations enter the horizon during the radiation-dominated era, so these ratios are much less than one at matter-radiation equality. Subsequently the ratio of fast to slow contributions to δ_{Dq} decreases as $1/a^3$, while for ψ_q the ratio of fast to slow contributions decreases as $1/a^{5/2}$. Hence it is a good approximation to take $\delta_{Dq}=\delta_{Dq}^{\rm slow}$ and $\psi_q=\psi_q^{\rm slow}$ throughout the matter-dominated era. Nevertheless, because the fast terms in Eqs. (6.4.59) and (6.4.60) have an oscillatory dependence on q, we will need to take them into account when we consider baryon acoustic oscillations in Section 8.1.

We also need $\delta_{\gamma q}$ and $\delta u_{\gamma q}$ in calculations of the cosmic microwave background anisotropies. Comparison of Eqs. (6.4.45) and (6.4.46) with Eqs. (6.4.57) and (6.4.56) shows that the ratios of the fast and slow contributions to $\delta_{\gamma q}$ and $u_{\gamma q}$ are (apart from the damping of the fast terms) of order

$$\frac{\delta_{\gamma q}^{\rm fast}}{\delta_{\gamma q}^{\rm slow}} \sim \frac{q^2}{a^2H^2}\frac{\bar{\rho}_R}{\bar{\rho}_M}, \quad \frac{\delta u_{\gamma q}^{\rm fast}}{\delta u_{\gamma q}^{\rm slow}} \sim \frac{q}{aH}\frac{\bar{\rho}_R}{\bar{\rho}_M}. \tag{6.4.64}$$

These ratios are much larger than unity at horizon entry, and remain so until matter-radiation equality. After that the ratio of fast to slow contributions to $\delta_{\gamma q}$ remains of the same order of magnitude, while for $\delta u_{\gamma q}$ the ratio of fast to slow terms decreases like $1/\sqrt{a}$. Hence, once the wavelength enters the horizon, the slow contribution to $\delta_{\gamma q}$ is nominally smaller than the fast term, while the slow contribution to $\delta u_{\gamma q}$ remains nominally smaller than the fast term until late in the matter-dominated era. Nevertheless, we will keep the slow as well as fast terms here, because they are not affected by the damping that suppresses the fast terms. Also, even though it is relatively small, the slow term in δ_{Rq} will be found in Section 7.2 to produce a characteristic effect in the plot of the cosmic microwave background multipole coefficients C_ℓ vs. ℓ that would not be present with the fast term alone.

10.6.6 6.5 Scalar perturbations - interpolation & transfer functions

In the previous two sections we found analytic results for wavelengths that are long enough to enter the horizon well after matter-radiation equality, or short enough to enter the horizon well before matter-radiation equality. Unfortunately, this leaves out wavelengths that enter the horizon around the time of matter-radiation equality. It is wavelengths of this magnitude that make the dominant contributions to the first acoustic peak at around $\ell = 200$ in the multipole coefficients of the cosmic microwave background anisotropies. In this section we will consider how to construct formulas for the fluctuations that interpolate between the results of Sections 6.3 and 6.4, concentrating on results in the matter-dominated era, which are of the greatest observational interest.

We can get a good clue to this interpolation by first considering the limit of negligible baryon mass density and negligible damping, for which the form of the solution in the matter-dominated era can be found exactly, with no limitations on wavelength. In this limit, and leaving aside the neutrinos, the equations for the perturbations in the matter-dominated era are given for
all wavelengths by Eqs. (6.3.17), (6.3.18), (6.3.20) and (6.3.21), with $R = 0$:

$$\frac{d}{dt}\left(t^{4/3}\psi_q\right) = -\frac{2}{3}t^{-2/3}\delta_{Dq} \tag{6.5.1}$$

$$\dot{\delta}_{\gamma q} - \left(q^2/a^2\right)\delta u_{\gamma q} = -\psi_q \tag{6.5.2}$$

$$\dot{\delta}_{Dq} = -\psi_q \tag{6.5.3}$$

$$\frac{d}{dt}\left(t^{-2/3}\delta u_{\gamma q}\right) = -\frac{1}{3}t^{-2/3}\delta_{\gamma q} \tag{6.5.4}$$

with $a \propto t^{2/3}$. These have four independent exact solutions, three of which are quite simple:

Solution 1:

$$\delta_{Dq} = \frac{3q^2t^2}{2a^2}, \psi_q = -\frac{q^2t}{a^2}$$
$$\delta_{\gamma q} = 1, \delta u_{\gamma q} = -t$$

Solution 2:

$$\delta_{Dq} = \psi_q = 0$$
$$\delta_{\gamma q} = -\cos\left(q\int_0^t \frac{dt}{\sqrt{3}a}\right), \quad \delta u_{\gamma q} = \frac{a}{\sqrt{3}q}\sin\left(q\int_0^t \frac{dt}{\sqrt{3}a}\right).$$

Solution 3:

$$\delta_{Dq} = \psi_q = 0$$
$$\delta_{\gamma q} = \sin\left(q\int_0^t \frac{dt}{\sqrt{3}a}\right), \quad \delta u_{\gamma q} = \frac{a}{\sqrt{3}q}\cos\left(q\int_0^t \frac{dt}{\sqrt{3}a}\right).$$

The fourth solution is more complicated; it decays, with $\delta_{Dq} \propto 1/t$ and $\psi_q \propto 1/t^2$. As we will see, the initial conditions are satisfied by a linear combination of the first three solutions, so the fourth solution will not concern us here.

Without any loss of generality, we can write the linear combination of solutions 1, 2, and 3 that fits the solution at earlier time in a form that simplifies this fit, as [1]

$$\delta_{Dq} = \frac{9q^2t^2\mathcal{R}_q^o\mathcal{T}(\kappa)}{10a^2} \tag{6.5.5}$$
$$\psi_q = -\frac{3q^2t\mathcal{R}_q^o\mathcal{T}(\kappa)}{5a^2} \tag{6.5.6}$$

$$\delta_{\gamma q} = \delta_{\nu q} = \frac{3\mathcal{R}_q^o}{5}\left[\mathcal{T}(\kappa) - \mathcal{S}(\kappa)\cos\left(q\int_0^t \frac{dt}{\sqrt{3}a} + \Delta(\kappa)\right)\right] \tag{6.5.7}$$
$$\delta u_{\gamma q} = \delta u_{\nu q} = \frac{3t\mathcal{R}_q^o}{5}\left[-\mathcal{T}(\kappa) + \mathcal{S}(\kappa)\frac{a}{\sqrt{3}qt}\sin\left(q\int_0^t \frac{dt}{\sqrt{3}a} + \Delta(\kappa)\right)\right] \tag{6.5.8}$$

where $\mathcal{S}(\kappa), \mathcal{T}(\kappa)$, and $\Delta(\kappa)$ are time-independent dimensionless functions of the dimensionless rescaled wave number introduced in Sections 6.2 and 6.4 :

$$\kappa \equiv \frac{q\sqrt{2}}{a_{\rm EQ}H_{\rm EQ}} = \frac{(q/a_0)\sqrt{\Omega_R}}{H_0\Omega_M} = \frac{19.3\,(q/a_0)\left[{\rm Mpc}^{-1}\right]}{\Omega_M h^2} \tag{6.5.9}$$

in which $a_{\rm EQ}$ and $H_{\rm EQ}$ are respectively the Robertson-Walker scale factor and expansion rate at matter-radiation equality. These are known as transfer functions. (These functions can only depend on κ because they must be independent of the normalization of the spatial coordinates and are dimensionless.) The division we made in the previous two sections between long and short wavelengths can be expressed in terms of the parameter κ. In the matter-dominated era we have $t = 2/3H = (t/3H_{\rm EQ})\,(a/a_{\rm EQ})^{3/2}$ and $\bar\rho_R/\bar\rho_M = a_{\rm EQ}/a$, so

$$\frac{t^2q^2}{a^2}\frac{\bar\rho_R}{\bar\rho_M} = \frac{4\kappa^2}{9}$$

Hence t^2q^2/a^2 is much less or much greater than $\bar\rho_M/\bar\rho_R$, according as the parameter κ is much less or much greater than one.

We choose to write the linear combination of solutions 1,2 , and 3 in the form (6.5.5)-(6.5.8) because it leads to simple values for the coefficients $\mathcal{S}(\kappa), \mathcal{T}(\kappa)$, and $\Delta(\kappa)$ for $\kappa \ll 1$. For $\bar\rho_B = 0$, the results (6.3.23), (6.3.24), (6.3.35), (6.3.36) obtained earlier for moderately long wavelengths satisfying $t^2q^2/a^2 \ll \bar\rho_M/\bar\rho_R$, or in other words in the limit $\kappa \ll 1$, are consistent with Eqs. (6.5.5)-(6.5.8), and tell us that in this limit [2]

[471] From now on, we define the zero of time so that $a \propto t^{1/2}$ in the radiation-dominated era. This is different from the definition of the zero of time in Solutions 1 and 2, where time is defined so that $a \propto t^{2/3}$ in the matter-dominated era. However, the difference this makes in the integral $\int_0^t dt/a$ is just a constant, which can be absorbed into $\Delta(\kappa)$.

$$\mathcal{T}(\kappa) \to 1, \quad \mathcal{S}(\kappa) \to 1, \quad \Delta(\kappa) \to 2\kappa/\sqrt{3}. \tag{6.5.10}$$

(The case of extremely long wavelengths does not arise here, because it requires that $t^2q^2/a^2 \ll R$, and for the present we are taking $R = 0$.)

Similarly, for $\bar{\rho}_B = 0$ and $\Gamma = 0$, the results (6.4.59)-(6.4.62) obtained earlier for short wavelengths satisfying $t^2q^2/a^2 \gg \bar{\rho}_M/\bar{\rho}_R$, or in other words

in the limit $\kappa \gg 1$, are consistent with Eqs. (6.5.5)-(6.5.8), and tell us that in this limit [3]

$$\mathcal{T}(\kappa) \to \frac{45}{2\kappa^2}\left[-\frac{7}{2} + \gamma + \ln\left(\frac{4\kappa}{\sqrt{3}}\right)\right], \quad \mathcal{S}(\kappa) \to 5, \quad \Delta(\kappa) \to 0 \tag{6.5.11}$$

As we shall see in Chapter 8, the decrease of $\mathcal{T}(\kappa)$ for large κ is very important in determining the intensity of fluctuations in the dark matter density at various wavelengths. (We can understand this decrease qualitatively by following the history of the dark matter density fluctuations. According to Eq. (6.2.27), for all wavelengths δ_{Dq} grows like t at early times, when the universe is radiation dominated and the wavelength is outside the horizon. For $\kappa \ll 1$ this growth continues until the universe becomes matter dominated, after which Eqs. (6.3.13) and (6.3.23) tell us that δ_{Dq} grows like $t^{2/3}$ both before and after the wavelength enters the horizon. In contrast, for $\kappa \gg 1$ the growth of δ_{Dq} like t continues only until the wavelength enters the horizon, which in this case is during the radiation-dominated era, after which according to Eq. (6.4.50) δ_{Dq} grows only logarithmically with time until the universe becomes matter dominated, after which according to Eq. (6.4.54) it grows as $t^{2/3}$, just as in the case $\kappa \ll 1$. The end of the period of growth proportional to t is at the time when $q/a \approx 1/t$, which since $a \propto \sqrt{t}$ is at $t \propto 1/q^2$, so the growth until matter-radiation equality is proportional to $\ln q/q^2$, accounting for the asymptotic behavior of $\mathcal{T}(\kappa)$. That is, the smallness of $\mathcal{T}(\kappa)$ for $\kappa \gg 1$ reflects not a decay of δ_{Dq}, but rather the failure of δ_{Dq} to grow appreciably during the interval from horizon-entry to matter-radiation equality.)

For values of κ of order unity, we have to find some way of interpolating between these two limiting cases. This can be done almost "by hand"; it turns out that almost any smooth interpolation between the limits $\kappa \ll 1$ and $\kappa \gg 1$ gives reasonable results for the cosmic microwave background anisotropies. For better precision, we need to solve the full coupled equations (6.2.9)-(6.2.15) (still with $\bar{\rho}_B = 0$) numerically for general wavelengths and general values of $y \equiv \bar{\rho}_M/\bar{\rho}_R$, imposing the initial conditions found in Section 6.1 for $y \ll 1$, and then comparing this solution for $y \gg 1$ with Eqs. (6.5.5)-(6.5.8). These equations can be put in a dimensionless form by using y as the independent variable, and writing

$$\delta_{Dq} = \kappa^2\mathcal{R}_q^0 d(y)/4, \quad \delta_{\gamma q} = \delta_{\nu q} = \kappa^2\mathcal{R}_q^0 r(y)/4$$
$$\psi_q = \left(\kappa^2 H_{\rm EQ}/4\sqrt{2}\right)\mathcal{R}_q^0 f(y), \quad \delta u_{\gamma q} = \delta u_{\nu q} = \left(\kappa^2\sqrt{2}/4H_{\rm EQ}\right)\mathcal{R}_q^0 g(y)$$

10.6.7 6.5 Scalar perturbations - interpolation & transfer functions

Eqs. (6.2.9)-(6.2.15) then become

[472] The limit $2\kappa/\sqrt{3}$ for $\Delta(\kappa)$ is the difference for $y \equiv \bar{\rho}_M/\bar{\rho}_R \gg 1$ between the phase $\sqrt{3}qt/a = 2\kappa\sqrt{y}/\sqrt{3}$ in Eq. (6.3.34) (for $R \ll 1$), and the phase $q\int_0^t dt/\sqrt{3}a = 2\kappa[\sqrt{1+y}-1]/\sqrt{3}$ in Eqs. (6.5.7) and (6.5.8).

[473] This result for $T(\kappa)$ was given by S. Weinberg, Astrophys. J. 581, 810 (2002) [astro-ph/0207375]. The fact that the Mészáros equation (6.4.35) implies a fall-off of the transfer function for large wave number k like $\ln k/k^2$ had been pointed out by W. Hu and N. Sugiyama, Astrophys. J. 444, 489 (1995).

$$\sqrt{1+y}\frac{d}{dy}\left(y^2 f(y)\right) = -\frac{3d(y)}{2} - \frac{4r(y)}{y}$$

$$\sqrt{1+y}\frac{d}{dy}d(y) = -yf(y)$$

$$\sqrt{1+y}\frac{d}{dy}r(y) - \frac{\kappa^2}{y}g(y) = -yf(y)$$

$$\sqrt{1+y}\frac{d}{dy}\left(\frac{g(y)}{y}\right) = -\frac{1}{3}r(y)$$

In this notation, the initial conditions (6.2.27)-(6.2.29) read

$$d(y) \to r(y) \to y^2,$$

$$f(y) \to -2, \quad g(y) \to -\frac{y^4}{9}.$$

When evaluated for $y \gg 1$, the numerical solutions of these equations [4] match Eqs. (6.5.5)-(6.5.8), with transfer functions given in Table 6.1. Inspection of this table shows that these numerical results agree with the analytic results (6.5.10) and (6.5.11) in the limits $\kappa \ll 1$ and $\kappa \gg 1$, respectively, although κ must be quite large before the asymptotic results (6.5.11) are reached.

As we will see in the following two chapters, the microwave background anisotropies and the correlation function of large scale structure are given by integrals involving one or more of these transfer functions, so it will be useful to give approximate analytic formulas for these functions. Dicus [5] has found "fitting formulas," which to a good approximation agree with the asymptotic formulas (6.5.10) and (6.5.11) for $\kappa \ll 1$ and $\kappa \gg 1$, and (except for $\Delta(\kappa)$ at large κ) generally match the numerical results of Table 6.1 at intermediate values of κ to better than 2% :

$$\mathcal{T}(\kappa) \simeq \frac{\ln\left[1+(0.124\kappa)^2\right]}{(0.124\kappa)^2}\left[\frac{1+(1.257\kappa)^2+(0.4452\kappa)^4+(0.2197\kappa)^6}{1+(1.606\kappa)^2+(0.8568\kappa)^4+(0.3927\kappa)^6}\right]^{1/2}_{(6.5.12)} \tag{6.5.12}$$

$$\mathcal{S}(\kappa) \simeq \left[\frac{1+(1.209\kappa)^2+(0.5116\kappa)^4+5^{1/2}(0.1657\kappa)^6}{1+(0.9459\kappa)^2+(0.4249\kappa)^4+(0.1657\kappa)^6}\right]^2 \tag{6.5.13}$$

[474] I thank D. Dicus for this numerical calculation.

[5] D. Dicus, private communication.

Table 6.1: The Scalar Transfer Functions

κ	$\mathcal{S}$	$\mathcal{T}$	Δ	κ	$\mathcal{S}$	$\mathcal{T}$	Δ
0.1	1.0167	0.9948	0.1207	10	3.9895	0.1608	0.3270
0.2	1.0551	0.9780	0.2240	11	4.0546	0.1440	0.3147
0.3	1.1147	0.9569	0.3156	12	4.1172	0.1298	0.2962
0.4	1.1891	0.9339	0.3852	13	4.1841	0.1178	0.2850
0.5	1.2680	0.9101	0.4423	14	4.2175	0.1075	0.2747
0.6	1.3529	0.8860	0.4800	15	4.2676	0.0985	0.2604
0.7	1.4388	0.8620	0.5148	16	4.3135	0.0907	0.2541
0.8	1.5195	0.8384	0.5336	17	4.3336	0.0838	0.2438
0.9	1.6081	0.8154	0.5531	18	4.3796	0.0777	0.2339
1.	1.6801	0.7930	0.5637	19	4.4043	0.0723	0.2296
1.2	1.8330	0.7502	0.5784	20	4.4233	0.0675	0.2195
1.4	1.9777	0.7104	0.5854	25	4.5271	0.0496	0.1920
1.6	2.1126	0.6734	0.5842	30	4.6051	0.0383	0.1713
1.8	2.2354	0.6391	0.5782	35	4.6650	0.0305	0.1542
2.	2.3451	0.6074	0.5700	40	4.7087	0.0249	0.1396
2.5	2.5895	0.5378	0.5537	45	4.7389	0.0209	0.1276
3.	2.7839	0.4798	0.5334	50	4.7605	0.0177	0.1182
3.5	2.9473	0.4311	0.5094	55	4.7794	0.0153	0.1111
4.	3.0970	0.3898	0.4854	60	4.7992	0.0134	0.1053
4.5	3.2346	0.3545	0.4659	65	4.8192	0.0118	0.0997
5.	3.3506	0.3241	0.4509	70	4.8365	0.0105	0.0940
5.5	3.4114	0.2976	0.4367	75	4.8487	0.0094	0.0885
6.	3.5181	0.2726	0.4203	80	4.8563	0.0084	0.0838
6.5	3.5953	0.2531	0.4029	85	4.8622	0.0077	0.0803
7.	3.6754	0.2361	0.3884	90	4.8695	0.0070	0.0776
7.5	3.7473	0.2200	0.3782	95	4.8792	0.0064	0.0751
8.	3.8015	0.2056	0.3695	100	4.8895	0.0059	0.0722
8.5	3.8432	0.1927	0.3590				
9.	3.8865	0.1810	0.3465				
9.5	3.9380	0.1704	0.3350				

$$\Delta(\kappa) \simeq \left[\frac{(0.1585\kappa)^2 + (0.9702\kappa)^4 + (0.2460\kappa)^6}{1 + (1.180\kappa)^2 + (1.540\kappa)^4 + (0.9230\kappa)^6 + (0.4197\kappa)^8}\right]^{1/4}. \tag{6.5.14}$$

These transfer functions are shown in Figure 6.1. A fitting formula for $\mathcal{T}(\kappa)$ that includes the effects of neutrino anisotropic inertia and fits the

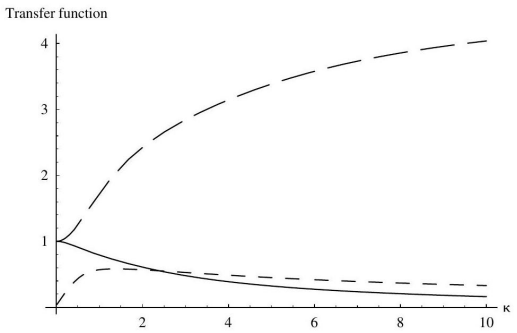


Figure 6.1: The transfer functions $\mathcal{T}(\kappa)$ (solid curve), $\mathcal{S}(\kappa)$ (long dashes), and $\Delta(\kappa)$ (short dashes), as functions of the rescaled wave number κ.

CAMB numerical results is given by Eisenstein and Hu. [6] Because of the neglect of neutrino anisotropic inertia, the results for $\mathcal{T}(\kappa)$ of Eq. (6.5.12) and Table 6.1 are about 4% too low.

Now we have to consider how to take damping and the non-zero ratio of baryon to photon density into account. Eqs. (6.3.23), (6.3.24) and (6.4.59), (6.4.60) show that in the limits of either short or long wavelength, the leading terms in δ_{Dq} and ψ_q are unaffected by either damping or baryons, so in leading order we can simply use Eqs. (6.5.5) and (6.5.6):

$$\delta_{Dq} = \frac{9q^2t^2\mathcal{R}_q^o\mathcal{T}(\kappa)}{10a^2} \tag{6.5.15}$$

$$\psi_q = -\frac{3q^2t\mathcal{R}_q^o\mathcal{T}(\kappa)}{5a^2} \tag{6.5.16}$$

Damping affects the short-wavelength results (6.4.61) and (6.4.62) for $\delta u_{\gamma q}$ and $\delta_{\gamma q}$ by multiplying the sines and cosines with a factor $\exp\left(-\int_0^t \Gamma dt\right)$. This factor is absent in the long-wavelength results (6.3.36) and (6.3.35) for these perturbations, but this factor is essentially unity anyway for long wavelengths, because $\Gamma \propto q^2$. Hence for all wavelengths we can take damping into account by simply multiplying the sines and cosines in Eqs. (6.5.7) and (6.5.8) by $\exp\left(-\int_0^t \Gamma dt\right)$. Finally, in both the short-wavelength results (6.4.62) and (6.4.61) and the long-wavelength results (6.3.35) and (6.3.36),

the effect of the non-zero baryon density is to multiply the non-oscillatory term in $\delta_{\gamma q}$ with $1+3R$, multiply the cosine in $\delta_{\gamma q}$ with $(1+R)^{-1/4}$, and multiply the sine in $\delta u_{\gamma q}$ with $(1+R)^{-3/4}$, so it is highly plausible that a finite ratio of baryon density to photon density can be taken into account at all wavelengths by making the same alterations in Eqs. (6.5.7) and (6.5.8):

$$
\begin{aligned}
\delta_{\gamma q}=\delta_{B q} &= \frac{3 \mathcal{R}_{q}^{o}}{5}\left[\mathcal{T}(\kappa)(1+3 R)\right. \\
&\left.(1+R)^{-1 / 4} e^{-\int_{0}^{t} \Gamma d t} \mathcal{S}(\kappa) \cos \left(\int_{0}^{t} \frac{q d t}{a \sqrt{3(1+R)}}+\Delta(\kappa)\right)\right] \tag{6.5.17}
\end{aligned}
$$

[476] D. J. Eisenstein and W. Hu, Astrophys. J. 496, 605 (1998).

$$\delta u_{\gamma q} = \delta_{Bq} = \frac{3\mathcal{R}_q^o}{5}[-t\mathcal{T}(\kappa) + \frac{a}{\sqrt{3}q(1+R)^{3/4}} e^{-\int_0^t \Gamma dt} \mathcal{S}(\kappa) \sin\left(\int_0^t \frac{qdt}{a\sqrt{3(1+R)}} + \Delta(\kappa)\right)] \tag{6.5.18}$$

To repeat, these results agree with the results of Section 6.4 for short wavelength, they agree with the results of Section 6.3 for long wavelength (for which Γ is negligible), and they agree with the results found in this section for all wavelengths when damping and the baryon density are neglected.

We must say a little more about the function $\mathcal{T}(\kappa)$, which is commonly known as the transfer function. It is conventional to write this transfer function as a function of a variable Q :

$$Q \equiv \frac{q}{a_0} \times \frac{1\text{Mpc}}{\Omega_M h^2}, \tag{6.5.19}$$

which according to Eq. (6.5.9) is the same as $\kappa/19.3$. In these terms, Eqs. (6.5.10) and (6.5.11) for $\mathcal{T}(\kappa)$ now read

$$\mathcal{T}(\kappa) \to \begin{cases} 1 & Q \to 0 \\ \frac{\ln(2.40Q)}{(4.07Q)^2} & Q \to \infty \end{cases} \tag{6.5.20}$$

A numerical solution of the equations for growth of dark matter density fluctuations has been fit for large Q with the formula [7]

$$\mathcal{T}_{\text{BBKS}}(\kappa) \simeq \frac{\ln(1+2.34Q)}{2.34Q} \times \left[1 + 3.89Q + (16.1Q)^2 + (5.46Q)^3 + (6.71Q)^4\right]^{-1/4} \tag{6.5.21}$$

This goes to $\ln(2.34Q)/(3.96Q)^2$ for large Q, in excellent agreement with the analytic result (6.5.20). However, it should be noted that although the BBKS transfer function (6.5.21) is correctly normalized at $Q = 0$, it does not give a good account of the behavior of the transfer function for $Q \ll 1$. The power series expansion of Eq. (6.5.21) contains odd as well as even terms in Q, which is not consistent with the requirement that it must be an analytic function of the three-vector $\mathbf{q}$. The Dicus fitting formula (6.5.12) was constructed to be analytic in the three-vector $\mathbf{q}$ at $\mathbf{q} = 0$, and in fact fits the numerical results of Table 6.1 better than the BBKS transfer function (6.5.21). In our analysis of cosmic microwave background anisotropies, we will use the fitting formula (6.5.12) instead of Eq. (6.5.21).

All this has been for $\beta \equiv \bar{\rho}_B/\bar{\rho}_M \ll 1$, though we have now taken into account a non-negligible ratio $R = 3\bar{\rho}_B/4\bar{\rho}_\gamma$. The corrections to the transfer function for finite values of β are roughly 10%; they have been calculated analytically [8] for $\kappa \gg 1$ to lowest order in β, and numerically [9] for general κ and for selected values of β to all orders in β. In using the transfer function in calculations of large scale structure, it is common to use a simple modification of the fitting formula. Peacock and Dodds [10] proposed in effect that the same fitting formula could be used for finite baryon density, but with $\kappa = \left(k\sqrt{\Omega_R}/H_0\Omega_M\right)\exp\left(\Omega_B\right)$. This worked well for the limited range of cosmological parameters studied, but it is physically impossible for the transfer function to have this sort of dependence on Ω_B. There is no way that the physical processes during the radiation-dominated era that are responsible for the transfer function to know anything about the time at which we happen to measure cosmological parameters like $\bar{\rho}_M$.

[477] J. M. Bardeen, J. R. Bond, N. Kaiser, & A. S. Szalay, Astrophys. J. 304, 15 (1986).

Aside from κ, the transfer function can only depend on quantities such as $\beta \equiv \bar{\rho}_B/\bar{\rho}_M = \Omega_B/\Omega_M$, which is constant, or $\Omega_M h^2$, which (for a known present radiation temperature) tells us the matter density at any given temperature. Indeed, Sugiyama [11] pointed out that the correction factor $\exp(\Omega_B)$ actually works well only for values of Ω_M close to unity. As an alternative that would apply for smaller values of Ω_M, he proposed a correction factor $\exp(\Omega_B + \Omega_B/\Omega_M)$, which is physically impossible for the same reason as $\exp(\Omega_B)$. Another difficulty with all these suggestions is that the baryon correction must disappear for small wave number, because in this case pressure forces are negligible, and baryons behave just like cold dark matter. A baryon correction that satisfies all these physical criteria has been proposed by Eisenstein and Hu: [6] The transfer function is evaluated

with κ taken as

$$\kappa = \left(\frac{k\sqrt{\Omega_R}}{H_0\Omega_M}\right)\left[\alpha + \frac{1-\alpha}{1+(0.43ks)^4}\right]^{-1}, \tag{6.5.22}$$

where

$$\alpha = 1 - 0.328\ln\left(431\Omega_M h^2\right)\beta + 0.38\ln\left(22.3\Omega_M h^2\right)\beta^2,$$

and s is the acoustic horizon at the time $t = t_L$ of last scattering, projected (as is k) to the present:

$$s \equiv a(t_0)\int_0^{t_L}\frac{v_s dt}{a(t)} = (1+z_L)\,d_H$$

with d_H given by Eq. (2.6.32).

10.6.8 6.6 Tensor perturbations

We next turn to the tensor modes. These are considerably simpler to study than the scalar modes, so what took five sections to analyze for the scalar modes will be treated here in just one section. As in the case of scalar modes, we begin by setting down the full set of equations used in computer programs like CMBfast, and then move on to approximations.

A. Cold dark matter and baryonic plasma

As already mentioned, the particles of both the cold dark matter and baryonic plasma move too slowly to contribute any anisotropic inertia. In tensor modes there are no perturbations to densities or streaming velocities, so there are no perturbations to either the cold dark matter or baryonic plasma that need to be followed here.

B. Gravitation

According to Section 5.1, in tensor modes the gravitational perturbation takes the form:

$$\delta g_{ij}(\mathbf{x},t) = a^2(t)D_{ij}(\mathbf{x},t) \tag{6.6.1}$$

with $D_{ij}(\mathbf{x},t)$ satisfying the wave equation (5.1.53):

$$\ddot{D}_{ij} + 3H\dot{D}_{ij} - a^{-2}\nabla^2 D_{ij} = 16\pi G\pi^T_{ij}, \tag{6.6.2}$$

and the trace and transversality conditions

[478] S. Weinberg, ref. 2.
[9] J. A. Holtzman, Astrophys. J. Suppl. 71, 1 (1989).
[10] J. A. Peacock and S.J. Dodds, Mon. Not. Roy. Astron. Soc. 267, 1020 (1994).
[11] N. Sugiyama, Astrophys. J. 100, 281 (1995).

$$D_{ii} = 0, \quad \partial_i D_{ij} = 0 \tag{6.6.3}$$

We will see that the anisotropic inertia tensor π_{ij}^T is a linear functional of $\dot{D}_{ij}$, so Eq. (6.6.2) has two independent solutions. During the period when the perturbation is outside the horizon, the anisotropic inertia tensor π_{ij}^T and the $a^{-2}\nabla^2 D_{ij}$ term are negligible, and Eq. (6.6.2) becomes

$$\ddot{D}_{ij} + 3H\dot{D}_{ij} = 0.$$

Hence outside the horizon one of the solutions is constant, while the other decays as $\int a^{-3}dt$, which in the radiation dominated era goes as $t^{-1/2}$. For all interesting wave numbers the perturbation remains outside the horizon during many e-foldings of cosmic expansion, so the decaying mode becomes negligible, and we can consider only the other mode. The metric perturbation and anisotropic inertia can therefore be put in the form

$$D_{ij}(\mathbf{x}, t) = \sum_{\lambda=\pm 2} \int d^3q e^{i\mathbf{q}\cdot\mathbf{x}} \beta(\mathbf{q}, \lambda) e_{ij}(\hat{q}, \lambda) \mathcal{D}_q(t), \tag{6.6.4}$$

$$\pi_{ij}^T(\mathbf{x}, t) = \sum_{\lambda=\pm 2} \int d^3q e^{i\mathbf{q}\cdot\mathbf{x}} \beta(\mathbf{q}, \lambda) e_{ij}(\hat{q}, \lambda) \pi_q^T(t). \tag{6.6.5}$$

where $\beta(\mathbf{q}, \lambda)$ is a stochastic parameter for the single non-decaying mode with wave number $\mathbf{q}$ and helicity λ; $e_{ij}(\hat{q}, \lambda)$ is the corresponding polarization tensor, defined in Section 5.2, with $e_{ii} = q_i e_{ij} = 0$; and $\mathcal{D}_q(t)$ is the solution of the wave equation

$$\ddot{\mathcal{D}}_q + 3H\dot{\mathcal{D}}_q + a^{-2}q^2\mathcal{D}_q = 16\pi G\pi_q^T. \tag{6.6.6}$$

We will return later to the solution of this equation.

C. Photons

The Boltzmann equation for the photon density matrix perturbation $\delta n^{ij}(\mathbf{x}, t)$ in the tensor mode is given by Eq. (H.35) as

$$\begin{aligned}
&\frac{\partial \delta n_\gamma^{ij}(\mathbf{x}, \mathbf{p}, t)}{\partial t} + \frac{\hat{p}_k}{a(t)} \frac{\partial \delta n_\gamma^{ij}(\mathbf{x}, \mathbf{p}, t)}{\partial x^k} + \frac{2\dot{a}(t)}{a(t)} \delta n_\gamma^{ij}(\mathbf{x}, \mathbf{p}, t) \\
&- \frac{1}{4a^2(t)} p\bar{n}_\gamma'(p) \hat{p}_k \hat{p}_l \dot{D}_{kl}(\mathbf{x}, t) \left(\delta_{ij} - \hat{p}_i \hat{p}_j\right) \\
&= -\omega_c(t) \delta n_\gamma^{ij}(\mathbf{x}, \mathbf{p}, t) + \frac{3\omega_c(t)}{8\pi} \int d^2\hat{p}_1 \\
&\times \left[\delta n_\gamma^{ij}\left(\mathbf{x}, p\hat{p}_1, t\right) - \hat{p}_i \hat{p}_k \delta n_\gamma^{kj}\left(\mathbf{x}, p\hat{p}_1, t\right) - \hat{p}_j \hat{p}_k \delta n_\gamma^{ik}\left(\mathbf{x}, p\hat{p}_1, t\right)\right. \\
&\left. + \hat{p}_i \hat{p}_j \hat{p}_k \hat{p}_l \delta n_\gamma^{kl}\left(\mathbf{x}, p\hat{p}_1, t\right)\right]
\end{aligned} \tag{6.6.7}$$

in which we have used Eq. (6.6.1) for the metric perturbation. As in the case of scalar modes, instead of $\delta n_\gamma^{ij}(\mathbf{x}, \mathbf{p}, t)$, we will concentrate on the fractional intensity matrix defined by the analog of Eq. (6.1.13):

$$a^4(t)\bar{\rho}_\gamma(t) J_{ij}(\mathbf{x}, \hat{p}, t) \equiv a^2(t) \int_0^\infty \delta n_\gamma^{ij}(\mathbf{x}, p\hat{p}, t) 4\pi p^3 dp \tag{6.6.8}$$

We seek a solution in the form

$$J_{ij}(\mathbf{x},\hat{p},t)=\sum_{\lambda=\pm 2}\int d^3q e^{i\mathbf{q}\cdot\mathbf{x}}\beta(\mathbf{q},\lambda)J_{ij}(\mathbf{q},\hat{p},t,\lambda) \tag{6.6.9}$$

The Boltzmann equation (6.6.7) then takes the form

$$\begin{aligned}\frac{\partial J_{ij}(\mathbf{q},\hat{p},t,\lambda)}{\partial t}+i\frac{\mathbf{q}\cdot\hat{p}}{a(t)}J_{ij}(\mathbf{q},\hat{p},t,\lambda)\\ +\hat{p}_k\hat{p}_l e_{kl}(\hat{q},\lambda)\dot{\mathcal{D}}_q(t)\left(\delta_{ij}-\hat{p}_i\hat{p}_j\right)\\ =-\omega_c(t)J_{ij}(\mathbf{q},\hat{p},t,\lambda)+\frac{3\omega_c(t)}{2}\\ \times\left[\mathcal{J}_{ij}(\mathbf{q},t,\lambda)-\hat{p}_i\hat{p}_k\mathcal{J}_{kj}(\mathbf{q},t,\lambda)-\hat{p}_j\hat{p}_k\mathcal{J}_{ik}(\mathbf{q},t,\lambda)\right.\\ \left.+\hat{p}_i\hat{p}_j\hat{p}_k\hat{p}_l\mathcal{J}_{kl}(\mathbf{q},t,\lambda)\right]\end{aligned} \tag{6.6.10}$$

where

$$\mathcal{J}_{ij}(\mathbf{q},t,\lambda)\equiv\int\frac{d^2\hat{p}}{4\pi}J_{ij}(\mathbf{q},\hat{p},t,\lambda) \tag{6.6.11}$$

Furthermore, because $J_{ij}(\mathbf{q},\hat{p},t,\lambda)$ for a given helicity λ must be a linear combination of the polarization tensor components $e_{kl}(\hat{q},\lambda)$ with the same λ, while $\hat{q}_k e_{kl}(\hat{q},\lambda)$ and $e_{kk}(\hat{q},\lambda)$ both vanish, the only possible form of $\mathcal{J}_{ij}(\mathbf{q},t,\lambda)$ allowed by rotational invariance is just $e_{ij}(\hat{q},\lambda)$ times some function of $q\equiv|\mathbf{q}|$ and t. This relation is conventionally written

$$\mathcal{J}_{ij}(\mathbf{q},t,\lambda)=-\frac{2}{3}e_{ij}(\hat{q},\lambda)\Psi(q,t) \tag{6.6.12}$$

Rotational invariance allows the intensity matrix perturbation to be written in the form

$$\begin{aligned}J_{ij}(\mathbf{q},\hat{p},t,\lambda)=&\frac{1}{2}\left(\delta_{ij}-\hat{p}_i\hat{p}_j\right)\hat{p}_k\hat{p}_l e_{kl}(\hat{q},\lambda)\\ &\times\left(\Delta_T^{(T)}(q,\hat{p}\cdot\hat{q},t)+\Delta_P^{(T)}(q,\hat{p}\cdot\hat{q},t)\right)\\ &+\left(e_{ij}(\hat{q},\lambda)-\hat{p}_i\hat{p}_k e_{kj}(\hat{q},\lambda)-\hat{p}_j\hat{p}_k e_{ik}(\hat{q},\lambda)+\hat{p}_i\hat{p}_j\hat{p}_k\hat{p}_l e_{kl}(\hat{q},\lambda)\right)\\ &\times\Delta_P^{(T)}(q,\hat{p}\cdot\hat{q},t)\end{aligned} \tag{6.6.13}$$

(Here the superscript T stands for "tensor," while the subscript T stands for "temperature." The coefficients are chosen so that J_{ii} is proportional to Δ_T, and the polarization is proportional to Δ_P.) A third term proportional to $\left(\hat{q}_i-\hat{p}_i(\hat{p}\cdot\hat{q})\right)\left(\hat{q}_j-\hat{p}_j(\hat{p}\cdot\hat{q})\right)\hat{p}_k\hat{p}_l e_{kl}$ would be allowed by symmetry principles, but is not generated in the Boltzmann equation by Thomson scattering. Using Eq. (6.6.13) in Eq. (6.6.10) yields separate Boltzmann equations for $\Delta_T^{(T)}$ and $\Delta_P^{(T)}$:

$$\begin{aligned}&\frac{\partial}{\partial t}\Delta_T^{(T)}(q,\mu,t)+ia^{-1}(t)q\mu\Delta_T^{(T)}(q,\mu,t)\\ &\quad=-2\dot{\mathcal{D}}_q(t)-\omega_c(t)\Delta_T^{(T)}(q,\mu,t)+\omega_c(t)\Psi(q,t)\end{aligned} \tag{6.6.14}$$

$$\begin{aligned}&\frac{\partial}{\partial t}\Delta_P^{(T)}(q,\mu,t)+ia^{-1}(t)q\mu\Delta_P^{(T)}(q,\mu,t)\\ &\quad=-\omega_c(t)\Delta_P^{(T)}(q,\mu,t)-\omega_c(t)\Psi(q,t)\end{aligned} \tag{6.6.15}$$

The functions $\Delta_T^{(T)}$ and $\Delta_P^{(T)}$ may be expanded in Legendre polynomials

$$\Delta_T^{(T)}(q,\hat{p}\cdot\hat{q},t)=\sum_{\ell=0}^{\infty}i^{-\ell}(2\ell+1)P_\ell(\hat{q}\cdot\hat{p})\Delta_{T,\ell}^{(T)}(q,t) \tag{6.6.16}$$

$$\Delta_P^{(T)}(q,\hat{p}\cdot\hat{q},t)=\sum_{\ell=0}^{\infty}i^{-\ell}(2\ell+1)P_\ell(\hat{q}\cdot\hat{p})\Delta_{P,\ell}^{(T)}(q,t) \tag{6.6.17}$$

Using the familiar recursion relation

$$zP_\ell(z)=\frac{\ell+1}{2\ell+1}P_{\ell+1}(z)+\frac{\ell}{2\ell+1}P_{\ell-1}(z)$$

we find that the Boltzmann equations (6.6.14) and (6.6.15) now read

$$\begin{aligned}&\dot{\Delta}_{T,\ell}^{(T)}+\frac{q}{a(2\ell+1)}\left((\ell+1)\Delta_{T,\ell+1}^{(T)}-\ell\Delta_{T,\ell-1}^{(T)}\right)\\&\quad=\left(-2\dot{\mathcal{D}}_q+\omega_c\Psi\right)\delta_{\ell,0}-\omega_c\Delta_{T,\ell}^{(T)}\end{aligned} \tag{6.6.18}$$

$$\begin{aligned}&\dot{\Delta}_{P,\ell}^{(T)}+\frac{q}{a(2\ell+1)}\left((\ell+1)\Delta_{P,\ell+1}^{(T)}-\ell\Delta_{P,\ell-1}^{(T)}\right)\\&\quad=-\omega_c\Psi\delta_{\ell,0}-\omega_c\Delta_{P,\ell}^{(T)}\end{aligned} \tag{6.6.19}$$

To calculate the source term $\Psi(q,t)$ in terms of partial waves, one first integrates Eq. (6.6.13) over $\hat{p}$, using the formulas

$$\int d^2\hat{p}f(\hat{p}\cdot\hat{q})\hat{p}_i\hat{p}_k e_{jk}(\hat{q})=\frac{1}{2}e_{ij}(\hat{q})\int d^2\hat{p}f(\hat{p}\cdot\hat{q})\left(1-(\hat{p}\cdot\hat{q})^2\right)$$

$$\int d^2\hat{p}f(\hat{p}\cdot\hat{q})\hat{p}_i\hat{p}_j\hat{p}_k\hat{p}_l e_{kl}(\hat{q})=\frac{1}{4}e_{ij}(\hat{q})\int d^2\hat{p}f(\hat{p}\cdot\hat{q})\left(1-(\hat{p}\cdot\hat{q})^2\right)^2$$

where f is any function of $\hat{p}\cdot\hat{q}$ and e_{ij} is any symmetric traceless matrix function of $\hat{q}$ with $\hat{q}_i e_{ij}=0$ and $e_{ii}=0$. The integral gives

$$\begin{aligned}\Psi(q,t)=-\frac{3}{2}\int\frac{d^2\hat{p}}{4\pi}\Bigg[&-\frac{1}{8}\left(1-(\hat{p}\cdot\hat{q})^2\right)^2\Delta_T^{(T)}(q,\hat{p}\cdot\hat{q},t)\\&+\left((\hat{p}\cdot\hat{q})^2+\frac{1}{8}\left(1-(\hat{p}\cdot\hat{q})^2\right)^2\right)\Delta_P^{(T)}(q,\hat{p}\cdot\hat{q},t)\Bigg].\end{aligned} \tag{6.6.20}$$

Inserting the partial wave expansions (6.6.16) and (6.6.17) then gives: [1]

$$\begin{aligned}\Psi(q,t)=&\frac{1}{10}\Delta_{T,0}^{(T)}(q,t)+\frac{1}{7}\Delta_{T,2}^{(T)}(q,t)+\frac{3}{70}\Delta_{T,4}^{(T)}(q,t)-\frac{3}{5}\Delta_{P,0}^{(T)}(q,t)\\&+\frac{6}{7}\Delta_{P,2}^{(T)}(q,t)-\frac{3}{70}\Delta_{P,4}^{(T)}(q,t)\end{aligned} \tag{6.6.21}$$

Eqs. (6.6.18) and (6.6.19) thus form a closed system of coupled differential equations for the partial wave amplitudes produced by a given gravitational field perturbation $\mathcal{D}_q(t)$. Of course, their solution requires a truncation of the partial wave expansion at some maximum ℓ.

The solution is of interest in itself, because we can measure anisotropies in the cosmic microwave background. It will be applied to these anisotropies in Chapter 7. It is also needed in calculations of the tensor anisotropic inertia. In tensor modes, the only non-vanishing contribution of photons to the energy-momentum tensor is to the space-space component (6.1.10)

$$\begin{aligned}\delta T^i_{\gamma j}(\mathbf{x},t) =& a^{-4}(t)\int d^3p a^2(t)\delta n^{kk}(\mathbf{x},\mathbf{p},t)p\hat{p}_i\hat{p}_j\\ =&\bar{\rho}_\gamma(t)\sum_\lambda\int d^3q\beta(\mathbf{q},\lambda)e^{i\mathbf{q}\cdot\mathbf{x}}\int\frac{d^2\hat{p}}{4\pi}J_{kk}(\mathbf{q},\hat{p},t,\lambda)\hat{p}_i\hat{p}_j\\ =&\bar{\rho}_\gamma(t)\sum_\lambda\int d^3q\beta(\mathbf{q},\lambda)e^{i\mathbf{q}\cdot\mathbf{x}}e_{kl}(\hat{q},\lambda)\\ &\times\int\frac{d^2\hat{p}}{4\pi}\Delta^{(T)}_T(q,\hat{p}\cdot\hat{q},t)\hat{p}_i\hat{p}_j\hat{p}_k\hat{p}_l\\ =&\bar{\rho}_\gamma(t)\sum_\lambda\int d^3q\beta(\mathbf{q},\lambda)e^{i\mathbf{q}\cdot\mathbf{x}}e_{ij}(\hat{q},\lambda)\\ &\times\frac{1}{4}\int\frac{d^2\hat{p}}{4\pi}\Delta^{(T)}_T(q,\hat{p}\cdot\hat{q},t)\left(1-(\hat{p}\cdot\hat{q})^2\right)^2\end{aligned}\tag{6.6.22}$$

For tensor modes there is no pressure perturbation, so according to Eq. (5.1.43), this is the same as the anisotropic inertia tensor $\pi^T_{\gamma ij}(\mathbf{x},t)$. Comparing Eq. (6.6.22) with Eq. (6.6.5) thus gives

$$\begin{aligned}\pi^T_{\gamma q}(t) &= \frac{\bar{\rho}_\gamma(t)}{4}\int\frac{d^2\hat{p}}{4\pi}\Delta^{(T)}_T(q,\hat{p}\cdot\hat{q},t)\left(1-(\hat{p}\cdot\hat{q})^2\right)^2\\ &= 2\bar{\rho}_\gamma(t)\left[\frac{1}{15}\Delta^{(T)}_{T,0}(q,t)+\frac{2}{21}\Delta^{(T)}_{T,2}(q,t)+\frac{1}{35}\Delta^{(T)}_{T,4}(q,t)\right].\end{aligned}\tag{6.6.23}$$

Experience shows that to accurately calculate the partial wave amplitudes up to $\ell=4$, which appear in Eqs. (6.6.21) and (6.6.23), one needs to solve the Boltzmann equations for the partial wave amplitudes up to larger values of ℓ, up to $\ell=10$. Once Ψ is calculated in this way, we can calculate $J_{ij}(\mathbf{q},\hat{p},t,\lambda)$ for very much higher values of ℓ by using the "line of sight" solution of Eq. (6.6.10):

$$\begin{aligned}J_{ij}(\mathbf{q},\hat{p},t,\lambda) = &\int_{t_1}^t dt'\exp\left(-i\mathbf{q}\cdot\hat{p}\int_{t'}^t\frac{dt''}{a\left(t''\right)}-\int_{t'}^t dt''\omega_c\left(t''\right)\right)\\ &\times\left[-\hat{p}_k\hat{p}_l\left(\delta_{ij}-\hat{p}_i\hat{p}_j\right)e_{kl}(\hat{q},\lambda)\dot{\mathcal{D}}_q\left(t'\right)-\omega_c\left(t'\right)\Psi\left(q,t'\right)\right.\\ &\left.\left(e_{ij}(\hat{q},\lambda)-\hat{p}_i\hat{p}_k e_{kj}(\hat{q},\lambda)-\hat{p}_j\hat{p}_k e_{ik}(\hat{q},\lambda)+\hat{p}_i\hat{p}_j\hat{p}_k\hat{p}_l e_{kl}(\hat{q},\lambda)\right)\right]\end{aligned}\tag{6.6.24}$$

where t_1 is any time that is early enough before recombination so that $\omega_c\left(t_1\right)\gg H\left(t_1\right)$, which allows us to drop a term proportional to $J_{ij}\left(\mathbf{q},\hat{p},t_1,\lambda\right)$. In terms of the temperature and polarization amplitudes defined by the decomposition (6.6.13), the line-of-sight integrals read

$$\begin{aligned}\Delta^{(T)}_T(q,\mu,t) =& -\Delta^{(T)}_P(q,\mu,t)-2\int_{t_1}^t dt'\exp\left[-iq\mu\int_{t'}^t\frac{dt''}{a\left(t''\right)}\right.\\ &\left.-\int_{t'}^t\omega_c\left(t''\right)dt''\right]\times\mathcal{D}_q\left(t'\right)\end{aligned}\tag{6.6.25}$$

$$\begin{aligned}\Delta^{(T)}_P(q,\mu,t) =& -\int_{t_1}^t dt'\exp\left[-iq\mu\int_{t'}^t\frac{dt''}{a\left(t''\right)}-\int_{t'}^t\omega_c\left(t''\right)dt''\right]\\ &\times\omega_c\left(t'\right)\Psi\left(q,t'\right)\end{aligned}\tag{6.6.26}$$

[471] R. Crittenden, J. R. Bond, R. L. Davis, G. Efstathiou, and P. J. Steinhardt, Phys. Rev. Lett. 71, 324 (1993) [astro-ph/9303014].

This line of sight integral also provides an alternative to the use of the truncated partial wave expansion. [2] We can derive an integral equation for

$\Psi(q,t)$ by simply analytically integrating Eq. (6.6.24) over $\hat{p}$. Equating the coefficients of e_{ij} on both sides gives the integral equation

$$\Psi(q,t) = \frac{3}{2}\int_{t_1}^{t} dt' \exp\left[-\int_{t'}^{t} \omega_c\left(t''\right) dt''\right] \times \left[-2\dot{\mathcal{D}}_q\left(t'\right) K\left(q\int_{t'}^{t}\frac{dt''}{a\left(t''\right)}\right) + \omega_c\left(t'\right) F\left(q\int_{t'}^{t}\frac{dt''}{a\left(t''\right)}\right)\Psi\left(q,t'\right)\right] \tag{6.6.27}$$

where $K(v)$ and $F(v)$ are the functions

$$K(v) \equiv j_2(v)/v^2, \quad F(v) \equiv j_0(v) - 2j_1(v)/v + 2j_2(v)/v^2. \tag{6.6.28}$$

Eq. (6.6.27) can be solved efficiently either by iteration, or by numerical recipes appropriate for integral equations of the Volterra type. Once $\Psi(q,t)$ is calculated in this way, the complete photon intensity matrix can be obtained by a numerical integration in Eq. (6.6.24).

Neutrinos

The Boltzmann equation for the perturbation $\delta n_\nu(\mathbf{x},\mathbf{p},t)$ to the neutrino phase space density is given by Eq. (H.14) as

$$\frac{\partial \delta n_v(\mathbf{x},\mathbf{p},t)}{\partial t} + \frac{\hat{p}_i}{a(t)}\frac{\partial \delta n_v(\mathbf{x},\mathbf{p},t)}{\partial x^i} = \frac{p\bar{n}'(p)}{2}\hat{p}_i\hat{p}_j\dot{D}_{ij}(\mathbf{x},t), \tag{6.6.29}$$

in which we have used Eq. (6.6.1) for the metric perturbation. As in the case of scalar modes, instead of δn_ν, we find it more convenient to deal with a dimensionless intensity perturbation J, defined by the analog of Eq. (6.1.42):

$$a^4(t)\bar{\rho}_v(t)J(\mathbf{x},\hat{p},t) \equiv N_v\int_0^\infty \delta n_v(\mathbf{x},\mathbf{p},t)4\pi p^3 dp \tag{6.6.30}$$

where N_ν is the number of species of neutrino, counting antineutrinos separately, and $\bar{\rho}_\nu \equiv N_\nu a^{-4}\int 4\pi p^3 \bar{n}_\nu(p)$. This satisfies a Boltzmann equation

$$\frac{\partial J(\mathbf{x},\hat{p},t)}{\partial t} + \frac{\hat{p}_i}{a(t)}\frac{\partial J(\mathbf{x},\hat{p},t)}{\partial x^i} = -2\hat{p}_i\hat{p}_j\dot{D}_{ij}(\mathbf{x},t), \tag{6.6.31}$$

with D_{ij} given by Eq. (6.6.4). We will be able to find a solution in the form

$$J(\mathbf{x},\hat{p},t) = \sum_{\lambda=\pm 2}\int d^3q e^{i\mathbf{q}\cdot\mathbf{x}}\beta(\mathbf{q},\lambda)e_{ij}(\hat{q},\lambda)\hat{p}_i\hat{p}_j\Delta_v^{(T)}(q,\hat{p}\cdot\hat{q},t) \tag{6.6.32}$$

Then Eq. (6.6.29) becomes an equation for $\Delta_\nu^{(T)}$:

$$\dot{\Delta}_v^{(T)}(q,\mu,t) + \frac{iq\mu}{a(t)}\Delta_v^{(T)}(q,\mu,t) = -2\dot{\mathcal{D}}_q(t) \tag{6.6.33}$$

This can be solved by a partial wave expansion

$$\Delta_v^{(T)}(q,\mu,t) = \sum_\ell i^{-\ell}(2\ell+1)\Delta_{v,\ell}^{(T)}(q,t) \tag{6.6.34}$$

[472] S. Weinberg, Phys. Rev. D 74, 063517 (2006) [astro-ph/0607076]; D. Baskaran, L. P. Grishchuk, and A. G. Polnarev, Phys. Rev. D 74, 083008 (2006) [gr-qc/0605100].

with $\Delta^{(T)}_{\nu,\ell}(q,t)$ satisfying

$$\dot{\Delta}^{(T)}_{v,\ell} + \frac{q}{a(2\ell+1)}\left((\ell+1)\Delta^{(T)}_{v,\ell+1} - \ell\Delta^{(T)}_{T,\ell-1}\right) = -2\dot{\mathcal{D}}_q(t)\delta_{\ell 0}. \tag{6.6.35}$$

This of course needs to be truncated at some more-or-less arbitrary maximum value of ℓ. But instead we can find a direct solution of Eq. (6.6.33), as a line of sight integral

$$\Delta^{(T)}_v(q,\mu,t) = -2\int_{t_1}^{t} dt' \exp\left(-iq\mu\int_{t'}^{t}\frac{dt''}{a\,(t'')}\right)\dot{\mathcal{D}}_q\,(t') \tag{6.6.36}$$

We take t_1 soon enough after the decoupling of neutrinos at $T \approx 10^{10}$ K so that at this time the distribution of neutrinos is still essentially that of local thermal equilibrium. In this case the perturbation to this distribution arises only from the perturbations δT_ν and $\delta\mathbf{u}_\nu$ to the neutrino temperature and streaming velocity, which do not have tensor components, so we do not need to include an initial value term $\Delta^{(T)}_v\,(q,\mu,t_1)$ on the right-hand side of Eq. (6.6.36).

It will be a long while before anyone measures the angular distribution of cosmic neutrinos, so the only use to be made of calculations of δn_ν is in calculating components of the energy momentum tensor. In the tensor mode, the only non-vanishing component is $\delta T^i_{\nu j}$, given by the same formula (6.6.22) as for photons, except that $\Delta^{(T)}_\nu$ appears instead of $\Delta^{(T)}_T$:

$$\begin{aligned}
\delta T^i_{vj}(\mathbf{x},t) =& a^{-4}(t)\int d^3p\,\delta n_v(\mathbf{x},\mathbf{p},t)p\hat{p}_i\hat{p}_j\\
=&\bar{\rho}_v(t)\sum_\lambda\int d^3q\,\beta(\mathbf{q},\lambda)e^{i\mathbf{q}\cdot\mathbf{x}}\int\frac{d^2\hat{p}}{4\pi}J(\mathbf{q},\hat{p},t,\lambda)\hat{p}_i\hat{p}_j\\
=&\bar{\rho}_v(t)\sum_\lambda\int d^3q\,\beta(\mathbf{q},\lambda)e^{i\mathbf{q}\cdot\mathbf{x}}e_{kl}(\hat{q},\lambda)\\
&\times\int\frac{d^2\hat{p}}{4\pi}\Delta^{(T)}_v(q,\hat{p}\cdot\hat{q},t)\hat{p}_i\hat{p}_i\hat{p}_j\hat{p}_k\\
\\
=&\bar{\rho}_v(t)\sum_\lambda\int d^3q\,\beta(\mathbf{q},\lambda)e^{i\mathbf{q}\cdot\mathbf{x}}e_{ij}(\hat{q},\lambda)\\
&\times\frac{1}{4}\int\frac{d^2\hat{p}}{4\pi}\Delta^{(T)}_v(q,\hat{p}\cdot\hat{q},t)\left(1-(\hat{p}\cdot\hat{q})^2\right)^2
\end{aligned} \tag{6.6.37}$$

This is the neutrino contribution to the anisotropic inertia tensor π^T_{ij}. Comparing this with Eq. (6.6.5) then gives

$$\pi^T_{vq}(t) = \frac{\bar{\rho}_v(t)}{4}\int\frac{d^2\hat{p}}{4\pi}\Delta^{(T)}_v(q,\hat{p}\cdot\hat{q},t)\left(1-(\hat{p}\cdot\hat{q})^2\right)^2. \tag{6.6.38}$$

As for photons, this can be evaluated using the partial wave expansion:

$$\pi^T_{vq}(t) = 2\bar{\rho}_v(t)\left[\frac{1}{15}\Delta^{(T)}_{v,0}(q,t) + \frac{2}{21}\Delta^{(T)}_{v,2}(q,t) + \frac{1}{35}\Delta^{(T)}_{v,4}(q,t)\right]. \tag{6.6.39}$$

But for neutrinos it is easier to use the explicit solution (6.6.36). Using Eqs. (6.6.36) and (6.1.55) in the third expression of Eq. (6.6.35) and comparing with Eq. (6.6.5) gives [3]

$$\pi^T_{vq}(t) = -4\bar{\rho}_\gamma(t)\int_{t_1}^{t}dt'K\left(q\int_{t'}^{t}\frac{dt''}{a\,(t'')}\right)\dot{\mathcal{D}}_q\,(t') \tag{6.6.40}$$

where

$$K(v) \equiv j_2(v)/v^2 = -\frac{\sin v}{v^3} - \frac{3\cos v}{v^4} + \frac{3\sin v}{v^5}. \tag{6.6.41}$$

We will use this below in calculating the decay of gravitational waves that exit from the horizon in the radiation-dominated era.

We have found a complete set of differential equations for the tensor perturbations. Now we shall turn to the calculation of $\mathcal{D}_q(t)$, which provides the essential input in the calculation of tensor anisotropies in the cosmic microwave background. We begin by neglecting the tensor anisotropic inertia π_{ij}^T, returning to a consideration of its effects at the end of this section. The anisotropic inertia is negligible during most of the history of the universe, when the cosmic energy density is dominated by one or more perfect fluids, and it is never very large. With this approximation, the field equation (6.6.6) governing the Fourier components of the tensor component D_{ij} of the metric perturbation is simply

$$\ddot{\mathcal{D}}_q(t) + 3\frac{\dot{a}}{a}\dot{\mathcal{D}}_q(t) + \frac{q^2}{a^2}\mathcal{D}_q(t) = 0. \tag{6.6.42}$$

As already mentioned earlier in this section, at very early times when $q/a \ll \dot{a}/a$ one solution is constant while the other decays, so as long as this era lasts sufficiently long we can neglect the decaying solution, and take our initial condition that $\mathcal{D}_q(t)$ goes to a constant $\mathcal{D}_q^o$ at early times.

To treat the evolution of $\mathcal{D}_q(t)$ at later times, it is convenient once again to change the independent variable from t to $y \equiv a/a_{\rm EQ} = \bar{\rho}_M/\bar{\rho}_R$, where $a_{\rm EQ}$ is the value of the Robertson-Walker scale factor at matter-radiation equality. Assuming that the energy density of the universe is governed by radiation and non-relativistic matter, [4] we can put Equation (6.3.11) in the form

$$\frac{H_{\rm EQ}dt}{\sqrt{2}} = \frac{ydy}{\sqrt{1+y}} \tag{6.6.43}$$

where $H_{\rm EQ}$ is the expansion rate at matter-radiation equality. Then Eq. (6.6.42) becomes

$$(1+y)\frac{d^2\mathcal{D}_q}{dy^2} + \left(\frac{2(1+y)}{y} + \frac{1}{2}\right)\frac{d\mathcal{D}_q}{dy} + \kappa^2\mathcal{D}_q = 0 \tag{6.6.44}$$

where κ is the dimensionless rescaled wave number (6.5.9):

$$\kappa \equiv \frac{\sqrt{2}q}{a_{\rm EQ}H_{\rm EQ}} = \frac{(q/a_0)\sqrt{\Omega_R}}{H_0\Omega_M} = \frac{19.3\,(q/a_0)\left[{\rm Mpc}^{-1}\right]}{\Omega_M h^2}. \tag{6.6.45}$$

We also have the initial condition, that $\mathcal{D}_q \to \mathcal{D}_q^o$ for $y \ll 1$. As in the case of scalar modes, we can find analytic solutions in two extreme cases, for $\kappa \ll 1$ and $\kappa \gg 1$.

Consider first the case $\kappa \gg 1$. In this limit we already have $q/a \gg \dot{a}/a$ at matter-radiation equality, so horizon entry occurs early in the radiationdominated era. Hence for $\kappa \gg 1$ there are two overlapping eras; an era when the universe is radiation dominated extending to early times when the perturbation is outside the horizon, and an era when the perturbation is well inside the horizon extending to the present, in both of which we will be able to find analytic solutions. Because these eras overlap, we will be able to match the analytic solutions in the era of overlap, and in this way relate the gravitational wave amplitude at late times to the initial condition when the perturbation is outside the horizon.

First, in the era when the universe is radiation dominated $y \ll 1$, and Eq. (6.6.44) becomes

[473] S. Weinberg, Phys. Rev. D 69, 023503 (2004) [astro-ph/0306304].

$$\frac{d^2\mathcal{D}_q}{dy^2} + \left(\frac{2}{y}\right)\frac{d\mathcal{D}_q}{dy} + \kappa^2\mathcal{D}_q = 0 \tag{6.6.46}$$

The solution which approaches a constant $\mathcal{D}_q^o$ for $y \to 0$ is

$$D_q = \frac{\sin(\kappa y)}{\kappa y}\mathcal{D}_q^o \tag{6.6.47}$$

Next, consider the case of perturbations deep inside the horizon, that is, for which $q/a \gg H$. In this case we can find solutions using the WKB approximation for a completely arbitrary dependence of $a(t)$ on time, so that our results will apply even in the presence of vacuum energy, whether or not it is constant. For this purpose, we must put Eq. (6.6.42) into a standard form with no first derivative term [5] by introducing a new independent variable,

$$x \equiv \int a^{-3}(t)dt$$

so that Eq. (6.6.42) becomes

$$\frac{d^2\mathcal{D}_q}{dx^2} + q^2a^4\mathcal{D}_q = 0 \tag{6.6.48}$$

We write $\mathcal{D}_q = A\exp\left(\pm iq\int a^2dx\right)$, and keep only terms in Eq. (6.6.48) of order q^2 and q. This gives $dA/dx = -\left(A/2a^2\right)da^2/dx$, so $A \propto 1/a$, and the WKB solutions for $\mathcal{D}_q$ are $a^{-1}\exp\left(\pm iq\int a^2dx\right) = a^{-1}\exp\left(\pm iq\int a^{-1}dt\right)$. The factor $1/a$ gives a gravitational wave energy density that decreases as a^{-4}, a factor of a^{-1} representing the redshift of individual gravitons, and a factor a^{-3} arising from the dilution of gravitons as the universe expands. [6]

To find the correct linear combination of the two WKB solutions, we have to match them to the solution (6.6.47) in the radiation-dominated era, by considering these solutions in the intermediate range where both $q/a \gg H$ and $y \ll 1$. In the radiation-dominated era vacuum energy is presumably negligible, so here $q/aH = \kappa a_{\rm EQ}H_{\rm EQ}/\sqrt{2}aH = \kappa y$, and for $\kappa \gg 1$ there does exist a range of y for which $\kappa y \gg 1$ even though $y \ll 1$. In this range of y, we have

$$q\int\frac{dt}{a} = q\int\frac{da}{Ha^2} = \kappa\int\frac{yda}{a} = \kappa y$$

so (recalling that $y \propto a$) the linear combination of $a^{-1}\exp\left(+iq\int a^{-1}dt\ \right)$ and $a^{-1}\exp\left(-iq\int a^{-1}dt\right)$ that matches Eq. (6.6.47) where both are valid is

$$\mathcal{D}_q = \frac{1}{\kappa y}\sin\left(q\int_0^t\frac{dt'}{a\left(t'\right)}\right)\mathcal{D}_q^o. \tag{6.6.49}$$

To repeat, this is the solution deep in the horizon, whatever the contents of the universe. For times near the present, this is

$$\mathcal{D}_q = \frac{\sqrt{\Omega_R H_0^2}}{k}\sin\left(\eta + k\left(t - t_0\right)\right)\mathcal{D}_q^o, \tag{6.6.50}$$

[474] The evolution of the tensor amplitude under more general assumptions is considered by L . A. Boyle and P. J. Steinhardt, astro-ph/0512014.

[475] The standard way of eliminating the first derivative term in Eq. (6.6.42) is to change the dependent instead of the independent variable. For instance, see V. S. Mukhanov, H. A. Feldman, and R. H. Brandenberger, Phys. Rep. 215, 203 (1992); K. Ng and A. Speliotopoulos, Phys. Rev. D 52, 2112 (1995); and more recently, J. R. Pritchard and M. Kamionkowski, Ann. Phys. 318, 2 (2005). The results obtained in this section are much simpler than with this standard method, because by changing the independent rather than the dependent variable we encounter no turning point in the WKB solution.

[6] See G&C, Sec. 15.10.

where $k \equiv q/a_0$, and $\eta \equiv k a_0 \int_0^{t_0} dt/a(t)$.
It is possible that cosmological gravitational waves might be detected directly. [7] Tensor modes detectable in this way would certainly have $\kappa \gg 1$, the case we have been considering, so that they would enter the horizon much earlier than the time of recombination, and hence could provide an opportunity for a direct observation of the universe at very early times, and even for the exploration of physics at higher energies than can be reached by conventional particle accelerators. The existing ground-based Laser Interferometer Gravitational-Wave Observatory (LIGO) operates at about 100 Hz ; perturbations with this frequency re-entered the horizon when the cosmic temperature was about 10^8GeV, but LIGO does not have the sensitivity required to detect cosmological gravitational waves. But cosmological gravitational waves might be detected by space-borne laser interferometers. For instance, two detectors of cosmological gravitational waves have been under consideration: the Big Bang Observer in the U.S., and the Decihertz Laser Interferometer Gravitational Wave Observatory in Japan. Both operate at frequencies around 0.01 to 0.1 Hz , and according to Eq. (6.6.43), a gravitational wave with a frequency of $qc/2\pi a_0 = 10^{-2}$ Hz would have $\kappa = 1.3 \times 10^{15}/\Omega_M h^2 \gg 1$. The wavelengths to which such detectors would be sensitive would short enough to have come into the horizon when the cosmic temperature was 10^4 to 10^5GeV. In this case, the changes in the time dependence of $a(t)$ associated with changes in the equation of state of matter at various annihilation thresholds produce distinctive features in the spectrum of the tensor modes. [8] For reasonable assumptions about the pri-

mordial tensor spectrum, it seems likely [9] that a tensor perturbation strong enough to be detected directly at short wavelengths would be detected also indirectly at longer wavelengths through its effect on the polarization of the cosmic microwave background, to be discussed in Section 7.4.

In calculating the effect of cosmological gravitational waves on the cosmic microwave background, we need also to consider values of κ of order unity or less. Let us therefore turn to the other case that can be treated analytically, the case $\kappa \ll 1$. Here we still have $q/a \ll \dot{a}/a$ at matter-radiation equality, so the perturbation remains outside the horizon and $\mathcal{D}_q(t)$ remains equal to the constant value $\mathcal{D}_q^0$ until well into the matter-dominated era, when $y \gg 1$. In the limit $y \gg 1$, and whatever the value of κ, Eq. (6.6.42) becomes

$$y\frac{d^2\mathcal{D}_q}{dy^2} + \frac{5}{2}\frac{d\mathcal{D}_q}{dy} + \kappa^2\mathcal{D}_q = 0 \tag{6.6.51}$$

This has two independent solutions, which can be written as functions of $\kappa\sqrt{y}$:

$$\begin{aligned} \frac{2j_1(2\kappa\sqrt{y})}{\kappa\sqrt{y}} &= \frac{-\cos(2\kappa\sqrt{y})}{\kappa^2 y} + \frac{\sin(2\kappa\sqrt{y})}{2\kappa^3 y^{3/2}} \\ -\frac{2n_1(2\kappa\sqrt{y})}{\kappa\sqrt{y}} &= \frac{\sin(2\kappa\sqrt{y})}{\kappa^2 y} + \frac{\cos(2\kappa\sqrt{y})}{2\kappa^3 y^{3/2}} \end{aligned} \tag{6.6.52}$$

In the matter-dominated era, $q/aH = \kappa a_{\rm EQ} H_{\rm EQ}/\sqrt{2}aH = \kappa\sqrt{y}$, so we must impose the condition that the solution should approach the constant value $\mathcal{D}_q^o$ for $\kappa\sqrt{y} \ll 1$. In order to satisfy this condition, we must exclude the second solution, which becomes singular for

[477] For recent studies, see N. Seto, S. Kawamura, and T. Nakamura, Phys. Rev. Lett. 87, 221103 (2001) [astro-ph/0108011]; A. Buonanno, gr-qc/0303085; A. Cooray, astro-ph/0503118; T. L. Smith, M. Kamionkowski, and A. Cooray, Phys. Rev. D 73, 023504 (2006) [astro-ph/0506422]; G. Efstathiou and S. Chongchitnan, Prog. Theor. Phys. Suppl. 163, 204 (2006) [astro-ph/0603118]; B. Friedman, A. Cooray, and A. Melchiorri, Phys. Rev. D 74, 123509 (2006) [astro-ph/0610220]. Also see NASA web page universe.nasa.gov/program/bbo.html.
[8] Y. Watanabe and E. Komatsu, Phys. Rev. D 73, 123515 (2006) [astro-ph/0604176].

$\kappa\sqrt{y} \to 0$. In this limit the first solution approaches the constant 4/3, so the correct solution when $\kappa \ll 1$ is

$$\mathcal{D}_q \to \frac{3\mathcal{D}_q^o}{4}\left(\frac{-\cos(2\kappa\sqrt{y})}{\kappa^2 y} + \frac{\sin(2\kappa\sqrt{y})}{2\kappa^3 y^{3/2}}\right). \tag{6.6.53}$$

(Note that although Eq. (6.6.51) applies only for $y \gg 1$, Eq. (6.6.53) is valid for $\kappa \ll 1$ and any y, because it correctly gives $\mathcal{D}_q(t) = \mathcal{D}_q^o$ as long as $\kappa\sqrt{y} \ll 1$, which for $\kappa \ll 1$ applies until $y \gg 1$.)

The interpolation between the cases $\kappa \gg 1$ and $\kappa \ll 1$ is simplest in the matter-dominated era, when $y \gg 1$. This is only a fair approximation at the time of the decoupling of matter and radiation, when $y \approx 3$, but it

becomes a good approximation thereafter. In this case the tensor modes are governed by Eq. (6.6.51), and the general solution is a linear combination of the two solutions (6.6.52), which for any κ can conveniently be written

$$\mathcal{D}_q = \frac{3\mathcal{D}_q^o\mathcal{U}(\kappa)}{4\kappa^2}\left(\frac{-\cos(2\kappa\sqrt{y}+\Xi(\kappa))}{y} + \frac{\sin(2\kappa\sqrt{y}+\Xi(\kappa))}{2\kappa y^{3/2}}\right), \tag{6.6.54}$$

where $\mathcal{U}$ and Ξ are real dimensionless functions of κ. From Eq. (6.6.53) we see that for $\kappa \ll 1$,

$$\mathcal{U}(\kappa) \to 1, \quad \Xi(\kappa)/\kappa^3 \to 0 \tag{6.6.55}$$

which is why we chose to write the physical solution in the form (6.6.54). Also, for $y \gg 1$ Eq. (6.6.49) becomes

$$\mathcal{D}_q = \frac{\sin(2\kappa(\sqrt{1+y}-1))}{\kappa y}\mathcal{D}_q^o \tag{6.6.56}$$

and therefore for $\kappa \gg 1$,

$$\mathcal{U}(\kappa) \to \frac{4\kappa}{3}, \quad \Xi(\kappa) \to \frac{\pi}{2} - 2\kappa \tag{6.6.57}$$

The values of $\mathcal{U}(\kappa)$ and $\Xi(\kappa)$ for general κ must be found by a computer calculation of the solution of Eq. (6.6.44). The results [10] are presented in Table 6.2. It can be seen that the analytic asymptotic limits (6.6.55) and (6.6.57) agree quite well with the computer calculation for $\kappa \ll 1$ and $\kappa \gg 1$, respectively.

Now let us consider the effect of anisotropic inertia. [11] The anisotropic inertia tensor is the sum of the contributions from photons and neutrinos, but photons are have a short mean free time before the era of recombination, and make only a small contribution to the total energy density afterwards, so their contribution to the anisotropic inertia is small. This leaves neutrinos (including antineutrinos), which have been travelling essentially without collisions [12] since the temperature dropped below

[479] Smith, Kamionkowski, and Cooray, ref. 7.

[4710] D. Dicus, private communication.

[11] S. Weinberg, ref. 3. For results of an earlier computer calculation, see J. R. Bond, in Cosmology and Large Scale Structure, Les Houches Session LX, eds. R. Schaeffer, J. Silk, and J. Zinn-Justin (Elsevier, Amsterdam, 1996).

[12] For very short wavelengths that entered the horizon before the time of electron-positron annihilation, it is necessary to take into account collisions of neutrinos with each other and with electrons and positrons. This is considered by M. Lattanzi and G. Montani, Mod. Phys. Lett. A 20, 2607 (2005) [astro-ph/0508364].

Table 6.2: The Tensor Transfer Functions

κ	$\mathcal{U}$	Ξ	κ	$\mathcal{U}$	Ξ
0.001	1.0000	-2.991×10^{-11}	1.8	2.990	-2.418
0.002	1.0000	-9.557×10^{-10}	2.0	3.250	-2.781
0.003	1.0000	-7.239×10^{-9}	2.5	3.906	-3.706
0.004	1.0000	-9.236×10^{-8}	3.0	4.565	-4.646
0.005	1.0001	-9.236×10^{-8}	3.5	5.222	-5.597
0.006	1.0001	-2.286×10^{-7}	4.0	5.886	-6.555
0.007	1.0002	-4.908×10^{-7}	4.5	6.551	-7.518
0.008	1.0002	-9.498×10^{-7}	5.0	7.216	-8.484
0.009	1.0003	-1.697×10^{-6}	5.5	7.878	-9.453
0.01	1.0004	-2.847×10^{-6}	6.0	8.541	-10.425
			6.5	9.208	-11.398
0.02	1.0016	-7.848×10^{-5}	7.0	9.875	-12.372
0.03	1.0037	-4.655×10^{-4}	7.5	10.538	-13.347
0.04	1.0065	-1.397×10^{-3}	8.0	11.201	-14.324
0.05	1.0095	-2.809×10^{-3}	8.5	11.869	-15.302
0.06	1.0128	-4.397×10^{-3}	9.0	12.538	-16.279
0.07	1.0165	-6.023×10^{-3}	10.0	13.863	-18.237
0.08	1.0210	-7.929×10^{-3}	11.0	15.202	-20.196
0.09	1.0260	-1.044×10^{-2}	12.0	16.526	-22.157
0.1	1.0310	-1.352×10^{-2}	13.0	17.867	-24.119
0.2	1.0960	-5.802×10^{-2}	14.0	19.190	-26.081
0.3	1.1800	-0.1293	15.0	20.532	-28.044
0.4	1.2766	-0.2223	16.0	21.854	-30.001
0.5	1.3816	-0.3327	17.0	23.197	-31.973
0.6	1.4926	-0.4568	18.0	24.52	-33.94
0.7	1.6079	-0.5919	19.0	25.86	-35.90
0.8	1.7265	-0.7357	20.0	27.18	-37.87
0.9	1.848	-0.8866	21.0	28.53	-39.83
1.0	1.970	-1.043	22.0	29.85	-41.80
1.1	2.095	-1.205	23.0	31.19	-43.77
1.2	2.220	-1.370	24.0	32.51	-45.73
1.4	2.474	-1.171	25.0	33.86	-47.70
1.6	2.731	-2.061			

about 10^{10} K, and which make up a good fraction of the energy density of the universe until cold dark matter becomes important, at a temperature about 10^4 K. The tensor part of the anisotropic inertia tensor is given by Eq. (6.6.40), so the gravitational wave equation (6.6.6) now becomes an
integro-differential equation [13]
$\ddot{\mathcal{D}}_q(t) + 3\frac{\dot{a}}{a}\dot{\mathcal{D}}_q(t) + \frac{q^2}{a^2}\mathcal{D}_q(t) = -64\pi G\bar{\rho}_\nu(t)\int_0^t K\left(q\int_{t'}^t \frac{dt''}{a(t'')}\right)\dot{\mathcal{D}}_q(t')\,dt'$,
with $K(v)$ given by Eq. (6.6.41). Note that despite the presence of anisotropic inertia, for $q/a \ll H$ this has a solution with $\mathcal{D}_q(t)$ time-independent, in accordance with the general theorem of Section 5.4.

This wave equation becomes particularly simple for wavelengths short enough to enter the horizon during the radiation-dominated era (though well after $e^+ - e^-$annihilation), that is, for

$\kappa \gg 1$. We will define the zero of time so that in this era $a \propto \sqrt{t}$. It is convenient now to write $\mathcal{D}_q$ as a function of the variable

$$u \equiv q \int_0^t \frac{dt'}{a\,(t')} = \frac{2qt}{a(t)}, \tag{6.6.59}$$

instead of t. Using the Friedmann equation $8\pi G\bar{\rho}/3 = H^2 = 1/4t^2$, the gravitational wave equation (6.6.58) in the radiation-dominated era becomes

$$\frac{d^2}{du^2}\mathcal{D}_q(u) + \frac{2}{u}\frac{d}{du}\mathcal{D}_q(u) + \mathcal{D}_q(u) = -\frac{24 f_v}{u^2}\int_0^u K\,(u-u')\,\frac{d\mathcal{D}_q\,(u')}{du'}du' \tag{6.6.60}$$

where

$$f_v \equiv \frac{\bar{\rho}_v}{\bar{\rho}_v + \bar{\rho}_\gamma} = \frac{3\cdot(7/8)\cdot(4/11)^{4/3}}{1+3\cdot(7/8)\cdot(4/11)^{4/3}} = 0.4052. \tag{6.6.61}$$

Late in the radiation-dominated era, the factor $1/u^2$ makes the right-hand side negligible, so $\mathcal{D}_q(u)$ approaches a solution of the homogeneous equation. In general, one might expect a linear combination of $\sin u/u$ and $\cos u/u$, but in fact no $\cos u/u$ term appears in the solution. [14] A numerical solution of Eq. (6.6.60) shows that if $\mathcal{D}_q(u)$ takes the value $\mathcal{D}_q^o$ for $u \ll 1$ then for $u \gg 1$ (but still in the radiation-dominated era)

$$\mathcal{D}_q(u) \to \mathcal{D}_q^o \alpha \frac{\sin(u)}{u}, \tag{6.6.62}$$

where $\alpha = 0.8026$. This then serves instead of Eq. (6.6.47) as an initial condition for the subsequent evolution of the gravitational wave amplitude, so that later, during the matter-dominated era, the effect of damping on the amplitude (neglecting δ) is simply to multiply the result given in Eq. (6.6.54) by a factor α. [15]

The damping of gravitational waves of longer wavelength is considerably more complicated. Because of a shift in the phase of the oscillation, the effect of anisotropic inertia on the amplitude of gravitational waves at the time of decoupling is a sensitive function of wave number, and for some wave numbers can even be an enhancement instead of a damping, but typically the amplitude is damped for $\kappa = O(1)$ by roughly 5%.

Because the damping effect is small anyway for $\kappa \ll 1$, it will be an adequate approximation for all wavelengths to take the gravitational wave amplitude in the matter dominated era to be given by multiplying the result (6.6.54) with a factor $\alpha(\kappa)$:

$$\mathcal{D}_q = \frac{3\mathcal{D}_q^o \mathcal{U}(\kappa)\alpha(\kappa)}{4\kappa^2}\left(\frac{-\cos(2\kappa\sqrt{y}+\Xi(\kappa))}{y} + \frac{\sin(2\kappa\sqrt{y}+\Xi(\kappa))}{2\kappa y^{3/2}}\right) \tag{6.6.63}$$

with $\alpha(\kappa)$ some function of κ that rises smoothly from $\alpha(\kappa) = 0.8026$ for $\kappa \gg 1$ to $\alpha(\kappa) \simeq 1$ for $\kappa \ll 1$. For instance, we can take $\alpha(\kappa) \simeq (1 + .8026\kappa)/(1+\kappa)$. All observable effects of cosmological gravitational waves will be reduced by this factor $\alpha(\kappa)$.

[4713] The lower bound on this integral should in principle be taken as the time of neutrino decoupling, at a temperature of about 10^{10} K, but it is a good approximation to take it at a time $t = 0$, defined by writing the scale factor during the radiation-dominated era as $a \propto t^{1/2}$.

[14] D. Dicus and W. Repko, Phys. Rev. 72, 088302(2005) [astro-ph/0509096] have shown analytically that no $\cos u/u$ term appears in this solution, and have given an analytic solution as a rapidly convergent sum of spherical Bessel functions. The absence of a $\cos u/u$ term was shown very generally on causality grounds by S. Bashinsky, astro-ph/0505502.

[4715] The effect of possible neutrino masses and/or chemical potentials is considered by K. Ichiki, M. Yamaguchi, and J. Yokayama, Pub. Astron. Soc. Pacific 119, 30 (2007) [hep-ph/0611121].

10.7 Anisotropies in the Microwave Sky

We will now return to the theory of anisotropies in the cosmic microwave background, introduced in Section 2.6. In Section 7.1 we derive general formulas for the observed temperature fluctuation. Then in Sections 7.2 and 7.3 we combine these results with the analysis of cosmic evolution in Chapter 6 and introduce a series of approximations that simplify the evaluation of the multipole coefficients for scalar and tensor modes, respectively. Section 7.4 deals with the polarization of the microwave background.

10.7.1 7.1 General formulas for the temperature fluctuation

In this section we will derive general formulas for the contribution of scalar and tensor modes to the observed temperature fluctuation. We begin by carrying the solution of the Boltzmann equations in synchronous gauge given in Section 6.1 forward to the present. When implemented with computer programs such as CMBfast or CAMB, the approach provides numerical results of great accuracy, but neither the derivation of the formulas for temperature fluctuations nor the computer programs are physically very transparent. We will then show how these results can be simplified by making the approximation of a sharp transition from thermal equilibrium to complete transparency at a moment t_L of last scattering. This ignores the scattering of photons by matter that becomes reionized at a redshift of order 10. For temperature correlations, the corrections due to reionization are very simple for multipole orders ℓ greater than about 20, and will be included in Section 7.2. In Section 7.2 we will also partly make up for the approximation of a sharp drop from opacity to transparency by including effects of viscous damping during this transition, and by including effects of averaging over t_L. (At the end of this section we will show how the same simplified results can also be obtained in a more general gauge by following photon trajectories from t_L to the present, with no need to use the Boltzmann equation.) The results of the sudden-decoupling approximation obtained here will be used together with other approximations in Section 7.2 to derive analytic expressions for the temperature multipole coefficients, that require computer calculations only to carry out a single numerical integration.

Because the proper energy density of black body radiation is proportional to the fourth power of the temperature, the fractional perturbation in temperature of radiation coming from a direction $\hat{n}$ is one-fourth the
fractional perturbation in the proper energy density of photons travelling in direction $\hat{p} = -\hat{n}$ at our position $\mathbf{x} = 0$ and time $t = t_0$. Eqs. (6.1.13) or (6.6.8) thus give

$$\frac{\Delta T(\hat{n})}{T_0} = \frac{1}{4} J_{ii}\left(\mathbf{x} = 0, -\hat{n}, t_0\right), \tag{7.1.1}$$

where $J_{ij}(\mathbf{x}, \hat{p}, t)$ is the fractional density matrix, defined by Eqs. (6.1.13) for scalar modes and (6.6.8) for tensor modes. Using the decompositions (6.1.18), (6.1.20) for scalar modes or (6.6.9), (6.6.13) for tensor modes, the scalar and tensor contributions to the temperature fluctuation are

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)^{(S)} = \frac{1}{4}\int d^3q \alpha(\mathbf{q}) \Delta_T^{(S)}\left(q, -\hat{q}\cdot\hat{n}, t_0\right), \tag{7.1.2}$$

and

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)^{(T)} = \frac{1}{4}\sum_{\lambda=\pm 2}\int d^3q \beta(\mathbf{q},\lambda)\hat{n}_k\hat{n}_l e_{kl}(\hat{q},\lambda)\Delta_T^{(T)}\left(q, -\hat{q}\cdot\hat{n}, t_0\right), \tag{7.1.3}$$

where $\alpha(\mathbf{q})$ and $\beta(\mathbf{q},\lambda)$ are the stochastic parameters for whatever modes are assumed to dominate the scalar and tensor perturbations, respectively; $e_{kl}(\hat{q},\lambda)$ is the polarization

tensor defined in Section 6.6 for a gravitational wave with wave number $\mathbf{q}$ and helicity λ; and $\Delta_T^{(S)}(q,\mu,t)$ and $\Delta_T^{(T)}(q,\mu,t)$ are amplitudes appearing in the decompositions (6.1.20) and (6.6.13) of the scalar and tensor contributions to the fractional density matrix J_{ij}.

In order to display the angular dependence of the temperature shift, we use the line-of-sight integrals (6.1.37) and (6.1.38) for scalar modes and (6.6.25) and (6.6.26) for tensor modes, which give, for the scalar modes

$$\begin{aligned}\left(\frac{\Delta T(\hat{n})}{T_0}\right)^{(S)} =& \frac{1}{4}\int d^3q\alpha(\mathbf{q}) \\ & \times \int_{t_1}^{t_0} dt \exp\left[i\mathbf{q}\cdot\hat{n}\int_t^{t_0}\frac{dt'}{a\left(t'\right)} - \int_t^{t_0} dt'\omega_c\left(t'\right)\right] \\ & \times \Big[-2\dot{A}_q(t) + 2(\mathbf{q}\cdot\hat{n})^2\dot{B}_q(t) + 3\omega_c(t)\Phi(q,t) \\ & - 4i(\mathbf{q}\cdot\hat{n})\omega_c(t)\delta u_{Bq}(t)/a(t) \\ & + \frac{3}{4}\left(1-(\hat{q}\cdot\hat{n})^2\right)\omega_c(t)\Pi(q,t)\Big] \end{aligned} \tag{7.1.4}$$

and for the tensor modes

$$\begin{aligned}\left(\frac{\Delta T(\hat{n})}{T_0}\right)^{(T)} =& \frac{1}{4}\sum_{\lambda=\pm 2}\int d^3q\beta(\mathbf{q},\lambda)\hat{n}_k\hat{n}_l e_{kl}(\hat{q},\lambda) \\ & \times \int_{t_1}^{t_0} dt \exp\left[i\mathbf{q}\cdot\hat{n}\int_t^{t_0}\frac{dt'}{a\left(t'\right)} - \int_t^{t_0} dt'\omega_c\left(t'\right)\right] \\ & \times \Big[-2\dot{\mathcal{D}}_q(t) + \omega_c(t)\Psi(q,t)\Big] \end{aligned} \tag{7.1.5}$$

As a reminder: t_1 is any time sufficiently early before recombination so that any photon present then would have been scattered many times before the present; $\omega_c(t)$ is the photon collision rate at time t; $A_q(t)$ and $B_q(t)$ are the scalar fields in the perturbation to the metric in synchronous gauge, defined in Section 5.2; $\Phi(q,t)$ and $\Pi(q,t)$ are the scalar source functions, defined by Eq. (6.1.21); $\delta u_{Bq}(t)$ is the scalar velocity potential of the baryonic plasma; $\mathcal{D}_q(t)$ is the gravitational wave amplitude, defined in Section 5.2; and $\Psi(q,t)$ is the tensor source function, defined by Eqs. (6.6.11) and (6.6.12).

Eq. (7.1.4) does not give the expression for the scalar temperature fluctuation in its most convenient form. When we pass to the limit of a sharp moment of last scattering, its terms will not correspond to the decomposition of the fluctuation into Sachs-Wolfe, Doppler, intrinsic, and integrated Sachs-Wolfe terms, discussed in Section 2.6. The individual terms in Eq. (7.1.4) are not even invariant under the limited class of gauge transformations (5.3.39)-(5.3.42) that preserve the conditions for synchronous gauge. We will therefore rewrite Eq. (7.1.4) by using the identity

$$\begin{aligned}&\exp\left(i\mathbf{q}\cdot\hat{n}\int_t^{t_0}\frac{dt'}{a\left(t'\right)}\right)(\mathbf{q}\cdot\hat{n})^2\dot{B}_q(t) = \\ &-\exp\left(i\mathbf{q}\cdot\hat{n}\int_t^{t_0}\frac{dt'}{a\left(t'\right)}\right)\frac{d}{dt}\left(a^2(t)\ddot{B}_q(t) + a(t)\dot{a}(t)\dot{B}(t)\right) \\ &+\frac{d}{dt}\left\{\exp\left(i\mathbf{q}\cdot\hat{n}\int_t^{t_0}\frac{dt'}{a\left(t'\right)}\right)\right. \\ &\left.\times\left[a^2(t)\ddot{B}_q(t) + a(t)\dot{a}(t)\dot{B}_q(t) + ia(t)\mathbf{q}\cdot\hat{n}\dot{B}_q(t)\right]\right\} \end{aligned} \tag{7.1.6}$$

Using this in Eq. (7.1.4) and integrating by parts, yields our final formula for the scalar temperature fluctuation

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)^{(S)} = \left(\frac{\Delta T(\hat{n})}{T_0}\right)^{(S)}_{\rm early} + \left(\frac{\Delta T(\hat{n})}{T_0}\right)^{(S)}_{\rm ISW}, \tag{7.1.7}$$

where the first, "early," term receives contributions only from times with an appreciable free electron density, before recombination and after reionization,

$$\begin{aligned}\left(\frac{\Delta T(\hat{n})}{T_0}\right)^{(S)}_{\rm early} &= \int d^3q\alpha(\mathbf{q}) \\ &\times \int_{t_1}^{t_0} dt \exp\left[i\mathbf{q}\cdot\hat{n}\int_t^{t_0}\frac{dt'}{a\,(t')} - \int_t^{t_0} dt'\omega_c\,(t')\right] \\ &\times \omega_c(t)\left[\frac{3}{4}\Phi(q,t) + \frac{3}{16}\left(1-(\hat{q}\cdot\hat{n})^2\right)\Pi(q,t)\right. \\ &- \frac{1}{2}a^2(t)\ddot{B}_q(t) - \frac{1}{2}a(t)\dot{a}(t)\dot{B}_q(t) \\ &\left.-i(\mathbf{q}\cdot\hat{n})\left(\delta u_{Bq}(t)/a(t) + a(t)\dot{B}_q(t)/2\right)\right]\end{aligned} \tag{7.1.8}$$

and the second, "integrated Sachs-Wolfe term," receives contributions from the whole period from t_1 to the present,

$$\begin{aligned}\left(\frac{\Delta T(\hat{n})}{T_0}\right)^{(S)}_{\rm ISW} &= -\frac{1}{2}\int d^3q\alpha(\mathbf{q})\int_{t_1}^{t_0} dt \\ &\times \exp\left[i\mathbf{q}\cdot\hat{n}\int_t^{t_0}\frac{dt'}{a\,(t')} - \int_t^{t_0} dt'\omega_c\,(t')\right] \\ &\times \frac{d}{dt}\left[A_q(t) + a^2(t)\ddot{B}_q(t) + a(t)\dot{a}(t)\dot{B}_q(t)\right]\end{aligned} \tag{7.1.9}$$

(Here we are ignoring a surface term in the integration by parts at $t = t_0$, because it is linear in $\hat{n}$, and therefore just contributes to the $\ell = 0$ and $\ell = 1$ partial waves. This "late" term is calculated by a different method later in this section, and given in Eq.(7.1.38).)

The integrated Sachs-Wolfe term (7.1.9) represents the effect of changing gravitational fields during the passage of the microwave photons from last scattering to the present. As already noted in Section 2.6, the ISW term would vanish if the gravitational field from last scattering to the present (or more precisely, at times t when the transparency $\exp\left(-\int_t^{t_0}\omega_c\,(t')\,dt'\right)$ was non-negligible) arose solely from the density of cold matter. (This feature provides another reason, apart from gauge invariance, for the rearrangement of terms that led to Eq. (7.1.7).) Under this approximation, Eqs. (6.5.15) and (6.5.16) show that $\delta_{Dq} \propto t^{2/3}$ and $\psi_q \propto t^{-1/3}$. Eq. (5.3.38) gives $q^2A_q = 8\pi a^2\bar{\rho}_D\delta_{Dq} - 2Ha^2\psi_q$, both terms of which are constant in the matter-dominated era, so $\dot{A}_q$ does not contribute to Eq. (7.1.9). With $\dot{A}_q$ negligible, $\psi_q = -q^2\dot{B}_q$, so $\dot{B}_q \propto t^{-1/3}$, and therefore both terms $a^2\ddot{B}_q$ and $a\dot{a}\dot{B}_q$ in Eq. (7.1.9) are time-independent, and therefore do not contribute to the ISW effect either. For this reason the ISW term (7.1.9) is relatively small. The early-time ISW effect depends sensitively on the ratio of matter to radiation at last scattering, which is an aid in using observations of temperature fluctuations to measure $\Omega_M h^2$. The late-time ISW effect receives its main contribution from times near the present, when the dark matter density falls below the vacuum energy density. Anisotropies subtend larger angles when viewed nearby than from great distances, so the late-time ISW term in the temperature

anisotropy contributes to the temperature multipole coefficients C_ℓ only for relatively small ℓ, say $\ell < 20$. It is the integrated Sachs-Wolfe effect that causes the predicted values of $\ell(\ell+1)C_\ell$ to rise as ℓ drops from around 20 to smaller values. This effect has been difficult to see in present data on the cosmic microwave background temperature fluctuations. [1]

It is the "early" term (7.1.8) that makes the largest contribution to the scalar temperature fluctuation for $\ell > 20$. The terms in Eq. (7.1.8) proportional to $\mathbf{q} \cdot \hat{n}$ represent the Doppler effect, [2] while the other terms give the combined effect of gravitational time dilation and intrinsic temperature fluctuations.

Let's pause to check invariance under the limited set of gauge transformations that preserve the conditions for synchronous gauge. These transformations induce the changes $\Delta \dot{B}_q = -2\tau/a^2$ and $\Delta \delta u_{Bq} = \tau$, where τ is an arbitrary function of $\mathbf{x}$, so the Doppler term proportional to $i\mathbf{q} \cdot \hat{n}$ in

Eq. (7.1.8) is separately gauge invariant. Also, ${}^3\Delta\Phi = -(4/3)\dot{\bar{T}}\tau/\bar{T} = (4/3)\dot{a}\tau/a$ and $\Delta\Pi = 0$, so the other terms in Eq. (7.1.8) are also gauge invariant; and $\Delta A = -2\dot{a}\tau/a$, so also Eq. (7.1.9) is invariant under this limited set of gauge transformations.

Our results so far are exact, aside from the use of first-order perturbation theory and the assumption that photons interact only through purely elastic Thomson scattering. (In Eqs. (7.1.2) and (7.1.3) we are assuming that the scalar and tensor fluctuations are each dominated by a single mode, but it would be trivial to introduce a sum over modes in these expressions.) These results are equivalent to those given by Seljak and Zaldarriaga, [4] which are used in computer programs like CMBfast and CAMB. At the cost of only a small loss of numerical accuracy, they can be greatly simplified if we now make the approximation of a sharp transition from thermal equilibrium to perfect transparency at a definite time t_L.

The integrand of the "early" contribution (7.1.8) to the scalar temperature perturbation contains a factor $P(t)$ equal to the probability distribution of the last photon scattering

$$P(t) = \omega_c(t) \exp\left(- \int_t^{t_0} dt' \omega_c\left(t'\right) \right) \tag{7.1.10}$$

Assuming that $\omega_c(t)$ drops sharply at time t_L from a value much greater than the expansion rate to zero, the function $P(t)$ is non-zero only in a narrow interval around t_L. But $P(t)$ is a normalized probability distribution:

[471] The ISW effect can be detected through its correlation with inhomogeneities in the distribution of matter (which are also linear in $\alpha(\mathbf{q})$), as suggested by R. G. Crittenden and N. Turok, Phys. Rev. Lett. 76, 575 (1996) [astro-ph/9510072]. This effect has been seen in the correlation of data from the WMAP satellite (discussed in the next section) with various surveys, by P. Fosalba and E. Gaztañaga, Mon. Not. Roy. Astron. Soc. 350, L37 (2004) [astro-ph/0305468]; P. Fosalba, E. Gaztañaga, and F. Castander, Astrophys. J. 597, L89 (2003) [astro-ph/0307249]; N. Ashfordi, Y-S. Loh, and M. A. Strauss, astro-ph/0308260; S. P. Boughn and R. G. Crittenden, New Astron. Rev. 49, 75 (2005) [astroph/0404470]; N. Padmanabhan, C. M. Hirata, U. Seljak, D. Schlegel, J. Brinkmann and D. P. Schneider, Phys. Rev. D 72, 043525 (2005) [astro-ph/0410360]. More recent cross-correlations of anisotropies seen by the three-year Wilkinson Microwave Anisotropy Probe (discussed in the next section) with galaxies in the Sloan Digital Sky Survey and with radio galaxy data from the NRAO VLA Sky Survey give $0.7 \leq \Omega_V \leq 0.82$ and $0.3 \leq \Omega_V \leq 0.8$ at a 95% confidence level, respectively; see A. Cabré et al., astro-ph/0603690; D. Pietrobon, A. Balbi, and D. Marinucci, Phys. Rev. D 74, 043524 (2006) [astroph/0606475]. Unfortunately, cosmic variance limits the accuracy with which this approach can be used to study the time dependence of the vacuum energy.

[2] In a gauge in which cold dark matter remains at rest, the effect of gravitational perturbations on the cold dark matter particles is canceled by the definition of surfaces of equal time, so in this gauge the velocity perturbations of the baryonic plasma arise solely from pressure forces, not from gravitational forces. This is why the gravitational term proportional to $\dot{B}_q$ appears accompanying the plasma velocity potential in Eq. (7.1.8); it represents the velocity that in a different gauge would be given to the photonbaryon plasma by gravitational forces.

$$\int_{t_1}^{t_0} P(t)dt = 1 \tag{7.1.11}$$

so the integral over t in Eq. (7.1.8) can be evaluated by dropping the factor $P(t)$ and setting $t = t_L$:

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)^{(S)}_{\text{early}} \simeq \int d^3q\alpha(\mathbf{q}) \exp\left[i\mathbf{q}\cdot\hat{n}\int_{t_L}^{t_0}\frac{dt'}{a\,(t')}\right]$$
$$\times\left[\frac{3}{4}\Phi\,(q,t_L) + \frac{3}{16}(1-(\hat{q}\cdot\hat{n}))^2\Pi\,(q,t_L)\right.$$

7.1 General formulas for the temperature fluctuation

$$-\frac{1}{2}a^2(t)\ddot{B}_q\,(t_L) - \frac{1}{2}a\,(t_L)\,\dot{a}\,(t_L)\,\dot{B}_q\,(t_L)$$
$$\left.-i(\mathbf{q}\cdot\hat{n})\left(\delta u_q\,(t_L)\,/a\,(t_L) + a\,(t_L)\,\dot{B}_q\,(t_L)\,/2\right)\right]. \tag{7.1.12}$$

Under the same assumption of a rapid drop in $\omega_c(t)$ from a large value for $t < t_L$ to a negligible value for $t > t_L$, the factor $\exp\left(-\int_t^{t_0} dt'\omega_c\,(t')\right)$ in Eq. (7.1.9) rises sharply from zero for $t < t_L$ to unity for $t > t_L$, so Eq (7.1.9) becomes

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)^{(S)}_{\text{ISW}} \simeq -\frac{1}{2}\int d^3q\alpha(\mathbf{q})\int_{t_L}^{t_0} dt\exp\left[i\mathbf{q}\cdot\hat{n}\int_t^{t_0}\frac{dt'}{a\,(t')}\right]$$
$$\times\frac{d}{dt}\left[A_q(t) + a^2(t)\ddot{B}_q(t) + a(t)\dot{a}(t)\dot{B}_q(t)\right] \tag{7.1.13}$$

The same approximation applied to the tensor contribution (7.1.5) gives

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)^{(T)} = \frac{1}{4}\sum_{\lambda\simeq\pm2}\int d^3q\beta(\mathbf{q},\lambda)\hat{n}_k\hat{n}_l e_{kl}(\hat{q},\lambda)$$
$$\times\left[-2\int_{t_L}^{t_0} dt\dot{\mathcal{D}}_q(t)\exp\left(i\mathbf{q}\cdot\hat{n}\int_t^{t_0}\frac{dt'}{a\,(t')}\right)\right.$$
$$\left.+\Psi\,(q,t_L)\exp\left(i\mathbf{q}\cdot\hat{n}\int_{t_L}^{t_0}\frac{dt'}{a\,(t')}\right)\right] \tag{7.1.14}$$

Further, in local thermal equilibrium photons are unpolarized and have an isotropic momentum distribution, so $\Delta^{(S)}_{P,\ell}$, $\Delta^{(T)}_{P,\ell}$, and $\Delta^{(T)}_{T,\ell}$ vanish for all ℓ, and $\Delta^{(S)}_{T,\ell}$ vanishes for all ℓ except $\ell = 0$, so that the formulas (6.1.29), (6.1.30), and (6.6.21) for the source functions give $\Pi = \Psi = 0$

[473] This follows from the rule (5.3.42), that the change $\Delta\delta s$ in the perturbation δs to a scalar s with unperturbed value $\bar{s}$ is $\Delta s = -\dot{\bar{s}}\tau$. In using this rule, we note that the unperturbed photon distribution is isotropic and unpolarized, so that, of the terms in Eqs. (6.1.29) and (6.1.30) for Φ and Π, the only one that has an unperturbed value is $\Delta^{(S)}_{T,0}/3$. For the purposes of assessing the gauge transformation property of this term, we must define the unperturbed value of $\Delta^{(S)}_{T,0}$ so that its perturbation is the fractional photon density fluctuation $4\delta T/\bar{T}$, and so its unperturbed value is $4\ln\bar{T}$.

[4] U. Seljak and M. Zaldarriaga, Astrophys. J. 459, 437 (1996) [astro-ph/9603033].

and $\Phi = \Delta_{T,0}^{(S)}/3 = 4\delta T/3\bar{T}$ in local thermal equilibrium. [5] The assumption of a sharp drop from a very high to a very low photon collision frequency then

allows Eqs. (7.1.12) and (7.1.14) to be further simplified to

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)_{\text{early}}^{(S)} \simeq \int d^3q\alpha(\mathbf{q}) \exp\left[i\mathbf{q}\cdot\hat{n}\int_{t_L}^{t_0}\frac{dt'}{a\left(t'\right)}\right] \times \left[F(q) + i(\hat{q}\cdot\hat{n})G(q)\right] \tag{7.1.15}$$

where $F(q)$ and $G(q)$ are the form factors

$$F(q) = \frac{\delta T_q\left(t_L\right)}{\bar{T}\left(t_L\right)} - \frac{1}{2}a^2(t)\ddot{B}_q\left(t_L\right) - \frac{1}{2}a\left(t_L\right)\dot{a}\left(t_L\right)\dot{B}_q\left(t_L\right) \tag{7.1.16}$$

$$G(q) = -q\left(\delta u_{\gamma q}\left(t_L\right)/a\left(t_L\right) + a\left(t_L\right)\dot{B}_q\left(t_L\right)/2\right) \tag{7.1.17}$$

and

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)^{(T)} = -\frac{1}{2}\sum_{\lambda\simeq\pm 2}\int d^3q\beta(\mathbf{q},\lambda)\hat{n}_k\hat{n}_l e_{kl}(\hat{q},\lambda) \times \int_{t_L}^{t_0} dt\dot{\mathcal{D}}_q(t)\exp\left(i\mathbf{q}\cdot\hat{n}\int_t^{t_0}\frac{dt'}{a\left(t'\right)}\right) \tag{7.1.18}$$

We will now make the same approximation, of a sudden drop in opacity at t_L, and use it to derive formulas for the scalar and tensor temperature fluctuations by following photon trajectories, without needing to use the Boltzmann equation formalism of Section 6.1. Because it is easy, we will carry out this derivation in a more general class of gauges. After this derivation, we will check that it yields the results (7.1.15) and (7.1.13) for scalar fluctuations in synchronous gauge, and (7.1.18) for tensor fluctuations. The reader who is comfortable with the derivation of these formulas using the Boltzmann equation in synchronous gauge may want to skip the rest of this section.

We start with some general remarks, that apply equally to scalar and tensor perturbations, and that for scalar perturbations apply in any gauge in which $g_{i0} = 0$, including both Newtonian and synchronous gauge. Continuing to neglect a possible unperturbed spatial curvature, we write the perturbed metric in any gauge with $g_{i0} = 0$ in the form

$$g_{00} = -1 - E(\mathbf{x},t), \quad g_{i0} = 0, \quad g_{ij} = a^2(t)\delta_{ij} + h_{ij}(\mathbf{x},t). \tag{7.1.19}$$

A light ray travelling toward the center of the Robertson-Walker coordinate system from the direction $\hat{n}$ will have a co-moving radial coordinate r related to t by

$$0 = g_{\mu\nu}dx^\mu dx^\nu = -(1+E(r\hat{n},t))dt^2 + \left(a^2(t) + h_{rr}(r\hat{n},t)\right)dr^2 \tag{7.1.20}$$

[475] This can be seen formally by taking the limit $\omega_c \to \infty$ in the Boltzmann equations (6.1.27), (6.1.28), (6.6.18), and (6.6.19). This gives

$$\Delta_{P,\ell}^{(S)} = \frac{1}{2}\Pi\left(\delta_{\ell 0} + \delta_{\ell 2}\right), \quad \Delta_{T,\ell}^{(S)} = \left(3\Phi + \frac{1}{2}\Pi\right)\delta_{\ell 0} + \frac{1}{2}\Pi\delta_{\ell 2}$$
$$\Delta_{T,\ell}^{(T)} = \delta_{\ell 0}\Psi, \quad \Delta_{P,\ell}^{(T)} = -\delta_{\ell 0}\Psi$$

The formulas (6.1.29), (6.1.30), and (6.6.21) for the source functions then read $\Phi = \Phi - \Pi/12, \Pi = 3\Pi/5$, and $\Psi = 7\Psi/10$, which require that $\Pi = \Psi = 0$ and $\Phi = \Delta_{T,0}^{(S)}/3$.

or in other words

$$\frac{dr}{dt} = -\left(\frac{a^2 + h_{rr}}{1+E}\right)^{-1/2} \simeq -\frac{1}{a} + \frac{h_{rr}}{2a^3} - \frac{E}{2a}. \tag{7.1.21}$$

(To first order in perturbations we don't have to worry about a deflection of the ray from the radial direction, because $\bar{g}_{r\theta} = \bar{g}_{r\phi} = 0$, so that any deflection would have to be of first order, and its effect in the term $h_{ij}dx^i dx^j$ would therefore be of second order.)

We now make the approximation that the transition of cosmic matter from opacity to transparency occurred suddenly at a time t_L of last scattering, at a red shift $1 + z_L \simeq 1090$. With this approximation, the relevant first-order solution of Eq. (7.1.21) is

$$r(t) = s(t) + \int_{t_L}^{t} \frac{dt'}{a\,(t')} N\left(s\,(t')\,\hat{n}, t'\right) \tag{7.1.22}$$

where

$$N(\mathbf{x}, t) \equiv \frac{1}{2}\left[\frac{h_{rr}(\mathbf{x}, t)}{a^2(t)} - E(\mathbf{x}, t)\right], \tag{7.1.23}$$

and $s(t)$ is the zeroth order solution for the radial coordinate which has the value r_L at $t = t_L$:

$$s(t) = r_L - \int_{t_L}^{t} \frac{dt'}{a\,(t')} = \int_{t}^{t_0} \frac{dt'}{a\,(t')} \tag{7.1.24}$$

In particular, if the ray reaches $r = 0$ at a time t_0, then Eq. (7.1.22) gives

$$0 = s\,(t_0) + \int_{t_L}^{t_0} \frac{dt}{a(t)} N(s(t)\hat{n}, t) = r_L + \int_{t_L}^{t_0} \frac{dt}{a(t)} (N(s(t)\hat{n}, t) - 1). \tag{7.1.25}$$

A time interval δt_L between the departure of successive light wave crests at the time t_L of last scattering produces a time interval δt_0 between arrival of successive crests at t_0 given by the variation of Eq. (7.1.25):

$$\begin{aligned} 0 = \quad & \frac{\delta t_L}{a\,(t_L)}\Big[1 - N\,(r_L\hat{n}, t_L) + \int_{t_L}^{t_0} \frac{dt}{a(t)} \left(\frac{\partial N(r\hat{n}, t)}{\partial r}\right)_{r=s(t)}\Big] \\ & + \delta t_L \delta u_\gamma^r\,(r_L\hat{n}, t_L) + \frac{\delta t_0}{a\,(t_0)}\left[-1 + N\,(0, t_0)\right]. \end{aligned} \tag{7.1.26}$$

(The term on the right-hand side involving the radial velocity δu_γ^r of the photon gas or photon-electron-nucleon fluid arises from the change with
time of the radial coordinate r_L of the light source in Eq. (7.1.25). We don't consider the variation of the argument $s(t)\hat{n}$ in N, because to zeroth order r_L and t_L are related in such a way that $s(t) = 0$ for all r_L, so its variation with r_L is of first order, and the effect of this variation on N would be of second order.) The total rate of change of the quantity $N(s(t)\hat{n}, t)$ in Eq. (7.1.25) is

$$\frac{d}{dt} N(s(t)\hat{n}, t) = \left(\frac{\partial}{\partial t} N(r\hat{n}, t)\right)_{r=s(t)} - \frac{1}{a(t)} \left(\frac{\partial N(r\hat{n}, t)}{\partial r}\right)_{r=s(t)}$$

so Eq. (7.1.26) may be written

$$0 = \frac{\delta t_L}{a(t_L)}\left[1 - N(0,t_0) + \int_{t_L}^{t_0} dt \left\{\frac{\partial}{\partial t} N(r\hat{n},t)\right\}_{r=s(t)}\right] + \delta t_L \delta u_\gamma^r(r_L\hat{n},t_L) + \frac{\delta t_0}{a(t_0)}\left[-1 + N(0,t_0)\right] \tag{7.1.27}$$

This gives the ratio of the coordinate time interval between crests when emitted and received, but what we want is the ratio of the proper time intervals

$$\delta\tau_L = \sqrt{1 + E(r_L,t_L)}\delta t_L, \quad \delta\tau_0 = \sqrt{1 + E(0,t_0)}\delta t_0 \tag{7.1.28}$$

which to first order gives the ratio of the received and emitted frequencies as

$$\frac{v_0}{v_L} = \frac{\delta\tau_L}{\delta\tau_0} = \frac{a(t_L)}{a(t_0)}\left[1 + \frac{1}{2}\left(E(r_L\hat{n},t_L) - E(0,t_0)\right) - \int_{t_L}^{t_0}\left\{\frac{\partial}{\partial t} N(r\hat{n},t)\right\}_{r=s(t)} dt - a(t_L)\,\delta u_\gamma^r(r_L\hat{n},t_L)\right] \tag{7.1.29}$$

The temperature observed at the present time t_0 coming from direction $\hat{n}$ is then

$$T(\hat{n}) = (v_0/v_L)\left(\bar{T}(t_L) + \delta T(r_L\hat{n},t_L)\right)$$

where now we have added a term δT to take account of the intrinsic temperature fluctuation at time t_L. Likewise, in the absence of perturbations the temperature observed in all directions would be

$$T_0 = \left(a(t_L)/a(t_0)\right)\bar{T}(t_L)$$

so the fractional shift in the radiation temperature observed coming from direction $\hat{n}$ from its unperturbed value is [6]

$$\frac{\Delta T(\hat{n})}{T_0} \equiv \frac{T(\hat{n}) - T_0}{T_0} = \frac{v_0}{a(t_L)\,v_L/a(t_0)} - 1 + \frac{\delta T(r_L\hat{n},t_L)}{\bar{T}(t_L)} = \frac{1}{2}\left(E(r_L\hat{n},t_L) - E(0,t_0)\right) - \int_{t_L}^{t_0} dt\left\{\frac{\partial}{\partial t} N(r\hat{n},t)\right\}_{r=s(t)} - a(t_L)\,\delta u_\gamma^r(r_L\hat{n},t_L) + \frac{\delta T(r_L\hat{n},t_L)}{\bar{T}(t_L)} \tag{7.1.30}$$

Because tensor and scalar perturbations are uncorrelated, we will treat their contribution to C_ℓ independently.

For scalar perturbations in any gauge with $h_{i0} = 0$, the metric perturbation is given by Eqs. (5.1.31)-(5.1.33) as

$$h_{00} = -E, \quad h_{ij} = a^2\left[A\delta_{ij} + \frac{\partial^2 B}{\partial x^i \partial x^j}\right] \tag{7.1.31}$$

Also for scalar perturbations the radial photon fluid velocity is given in terms of a velocity potential δu_γ as

$$\delta u_\gamma^r = \bar{g}^{r\mu}\frac{\partial \delta u_\gamma}{\partial x^\mu} = \frac{1}{a^2}\frac{\partial \delta u_\gamma}{\partial r} \tag{7.1.32}$$

Thus Eq. (7.1.30) gives the scalar contribution to the temperature fluctuation

$$\begin{aligned}\left(\frac{\Delta T(\hat{n})}{T_0}\right)^S &= \frac{1}{2}\left(E\left(r_L\hat{n}, t_L\right) - E\left(0, t_0\right)\right) - \int_{t_L}^{t_0} dt \left\{\frac{\partial}{\partial t}N(r\hat{n}, t)\right\}_{r=s(t)} \\ &- \frac{1}{a\left(t_L\right)}\left(\frac{\partial \delta u_\gamma\left(r\hat{n}, t_L\right)}{\partial r}\right)_{r=r_L} + \frac{\delta T\left(r_L\hat{n}, t_L\right)}{\bar{T}\left(t_L\right)}\end{aligned} \tag{7.1.33}$$

where now

$$N = \frac{1}{2}\left[A + \frac{\partial^2 B}{\partial r^2} - E\right]. \tag{7.1.34}$$

Eq. (7.1.33) is not in the form that is most useful for our purposes. The total temperature fluctuation given by Eq. (7.1.33) is invariant with respect to the limited class of gauge transformations that leave $g_{i0} = 0$, but this is not true of its individual terms, including even the integral of $(\partial N/\partial t)_{r=s(t)}$.

It will be much more convenient to rewrite Eq. (7.1.33) in a way that leaves the integral term separately gauge invariant.

For this purpose, we make use of an identity corresponding to Eq. (7.1.6):

$$\begin{aligned}&\left\{\frac{\partial^2 \dot{B}(r\hat{n}, t)}{\partial r^2}\right\}_{r=s(t)} \\ &= -\frac{d}{dt}\left[\left\{a^2(t)\ddot{B}(r\hat{n}, t) + a(t)\dot{a}(t)\dot{B}(r\hat{n}, t) + a(t)\frac{\partial \dot{B}(r\hat{n}, t)}{\partial r}\right\}_{r=s(t)}\right] \\ &\quad + \left\{\frac{\partial}{\partial t}\left(a^2(t)\ddot{B}(r\hat{n}, t) + a(t)\dot{a}(t)\dot{B}(r\hat{n}, t)\right)\right\}_{r=s(t)}.\end{aligned}$$

Together with Eq. (7.1.34), this gives the integrand in Eq. (7.1.33) as

$$\begin{aligned}\left\{\frac{\partial}{\partial t}N(r\hat{n}, t)\right\}_{r=s(t)} &= -\frac{1}{2}\frac{d}{dt}\left[\left\{a^2(t)\ddot{B}(r\hat{n}, t) + a(t)\dot{a}(t)\dot{B}(r\hat{n}, t) + a(t)\frac{\partial \dot{B}(r\hat{n}, t)}{\partial r}\right\}_{r=s(t)}\right] \\ &\quad + \frac{1}{2}\left\{\frac{\partial}{\partial t}\left(a^2(t)\ddot{B}(r\hat{n}, t) + a(t)\dot{a}(t)\dot{B}(r\hat{n}, t)\right.\right. \\ &\quad \left.\left. + A(r\hat{n}, t) - E(r\hat{n}, t)\right)\right\}_{r=s(t)}\end{aligned}$$

The scalar fractional temperature fluctuation (7.1.33) may therefore be written

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)^S = \left(\frac{\Delta T(\hat{n})}{T_0}\right)^S_{\text{early}} + \left(\frac{\Delta T(\hat{n})}{T_0}\right)^S_{\text{late}} + \left(\frac{\Delta T(\hat{n})}{T_0}\right)^S_{\text{ISW}}, \tag{7.1.36}$$

where

[476] This result is essentially that first found by R. K. Sachs and A. M. Wolfe, Astrophys. J. 147, 73 (1967).

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)^S_{\text{early}} = -\frac{1}{2}a^2(t_L)\,\ddot{B}(r_L\hat{n},t_L) - \frac{1}{2}a(t_L)\,\dot{a}(t_L)\,\dot{B}(r_L\hat{n},t_L) + \frac{1}{2}E(r_L\hat{n},t_L) + \frac{\delta T(r_L\hat{n},t_L)}{\bar{T}(t_L)} - a(t_L)\left[\frac{\partial}{\partial r}\left(\frac{1}{2}\dot{B}(r\hat{n},t_L) + \frac{1}{a^2(t_L)}\delta u_\gamma(r\hat{n},t_L)\right)\right]_{r=r_L}, \tag{7.1.37}$$

$$\left(\frac{\Delta T(\hat{n})}{T}\right)^S_{\text{late}} = \frac{1}{2}a^2(t_0)\,\ddot{B}(0,t_0) + \frac{1}{2}a(t_0)\,\dot{a}(t_0)\,\dot{B}(0,t_0) - \frac{1}{2}E(0,t_0) + a(t_0)\left[\frac{\partial}{\partial r}\left(\frac{1}{2}\dot{B}(r\hat{n},t_0) + \frac{\delta u_\gamma(r\hat{n},t_0)}{a^2(t_0)}\right)\right]_{r=0}, \tag{7.1.38}$$

$$\left(\frac{\Delta T(\hat{n})}{T}\right)^S_{\text{ISW}} = -\frac{1}{2}\int_{t_L}^{t_0} dt\left[\frac{\partial}{\partial t}\left(a^2(t)\ddot{B}(r\hat{n},t) + a(t)\dot{a}(t)\dot{B}(r\hat{n},t) + A(r\hat{n},t) - E(r\hat{n},t)\right)\right]_{r=s(t)} \tag{7.1.39}$$

The "late" term (7.1.38) (which was ignored earlier in this section) is the sum of a direction-independent term and a term proportional to $\hat{n}$, which has been added to represent the anisotropy due to the local cosmic gravitational field and velocity. It only affects terms in the multipole expansion of the temperature correlation function with $\ell = 0$ and $\ell = 1$, so it can be ignored if from now on we consider only multipole orders $\ell \geq 2$.

Now let us check the separate gauge invariance of the ISW, Doppler, and the remaining combined terms, at least for the limited class of gauge transformations that preserve the condition $g_{i0} = 0$, or in the notation of Eq. (5.1.32), $F = 0$. Eq. (5.3.13) tells us that these gauge transformations have

$$\epsilon_0 = -a^2\frac{\partial}{\partial t}\left(\frac{\epsilon^S}{a^2}\right),$$

and so $\Delta\dot{B} = 2\epsilon_0/a^2$. According to Eqs. (5.3.13) and (5.3.16), a gauge transformation therefore shifts the terms $-a^2\ddot{B}/2, -a\dot{a}\dot{B}/2, E/2$, and $\delta T/\bar{T}$ in the combined gravitational and intrinsic temperature fluctuations by the amounts $-a^2\partial\left(\epsilon_0/a^2\right)/\partial t, -\dot{a}\epsilon_0/a, \dot{\epsilon}_0$, and $\epsilon_0\dot{\bar{T}}/\bar{T} = -\epsilon_0\dot{a}/a$, giving no net change; it shifts the terms $a\dot{B}/2$ and $\delta u_\gamma/a$ in the Doppler contribution by ϵ_0/a and $-\epsilon_0/a$, giving no net change; and it shifts the four terms $a^2\ddot{B}, 2a\dot{a}\dot{B}, A$, and $-E$ in the integrand of the integrated Sachs-Wolfe term by the amounts $2\dot{\epsilon}_0 - 4\epsilon_0\dot{a}/a, 2\epsilon_0\dot{a}/a, 2\epsilon_0\dot{a}/a$, and $-2\dot{\epsilon}_0$, respectively, giving no net change. Thus the integrated Sachs-Wolfe term, the Doppler term, and the combined gravitational and intrinsic temperature terms are separately gauge invariant. In particular, both Newtonian and synchronous gauge will give the same results for each of these three contributions.

Let us now assume that from last scattering to the present the scalar contributions to the fluctuations are dominated by a single mode, so that
any perturbation $X(\mathbf{x},t)$ (such as $B, E, \delta u_\gamma$, or δT) can be written as

$$X(\mathbf{x},t) = \int d^3q\,\alpha(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{x}}X_q(t) \tag{7.1.40}$$

with $\alpha(\mathbf{q})$ a stochastic variable (the same for all X), normalized so that

$$\langle\alpha(\mathbf{q})\alpha^*(\mathbf{q}')\rangle = \delta^3(\mathbf{q}-\mathbf{q}'). \tag{7.1.41}$$

Then Eqs. (7.1.37) and (7.1.39) give

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)^S_{\text{early}} = \int d^3q\alpha(\mathbf{q})e^{i\mathbf{q}\cdot\hat{n}r(t_L)}(F(q) + i\hat{q}\cdot\hat{n}G(q)) \tag{7.1.42}$$

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)^S_{\text{ISW}} = -\frac{1}{2}\int_{t_0}^{t_L} dt \int d^3q\alpha(\mathbf{q})e^{i\mathbf{q}\cdot\hat{n}s(t)}\frac{d}{dt}\Big(a^2(t)\ddot{B}_q(t) + a(t)\dot{a}(t)\dot{B}_q(t) + A_q(t) - E_q(t)\Big) \tag{7.1.43}$$

where

$$F(q) = -\frac{1}{2}a^2(t_L)\,\ddot{B}_q(t_L) - \frac{1}{2}a(t_L)\,\dot{a}(t_L)\,\dot{B}_q(t_L) + \frac{1}{2}E_q(t_L) + \frac{\delta T_q(t_L)}{\bar{T}(t_L)} \tag{7.1.44}$$

$$G(q) = -q\left(\frac{1}{2}a(t_L)\,\dot{B}_q(t_L) + \frac{1}{a(t_L)}\delta u_{\gamma q}(t_L)\right) \tag{7.1.45}$$

As we have seen, the form factors $F(q)$ and $G(q)$ and the integrand of the ISW term are separately invariant under gauge transformations that leave g_{i0} equal to zero.

In synchronous gauge $E_q = 0$, so Eqs. (7.1.42)-(7.1.45) are the same as the previously derived results (7.1.15), (7.1.13), (7.1.17), and (7.1.18). In Newtonian gauge $B = 0$ and $E = 2\Phi$, so the form factors are

$$F(q) = \Phi_q(t_L) + \frac{\delta T_q(t_L)}{\bar{T}(t_L)} \tag{7.1.46}$$

$$G(q) = -\frac{1}{a(t_L)}\delta u_{\gamma q}(t_L) \tag{7.1.47}$$

(Of course, δT_q and $\delta u_{\gamma q}$ are different in Newtonian gauge from what they are in synchronous gauge.) Also, the integrated Sachs-Wolfe term is

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)_{\text{ISW}} = -2\int_{t_L}^{t_0} dt \int d^3q\alpha(\mathbf{q})e^{i\mathbf{q}\cdot\hat{n}s(t)}\dot{\Phi}_q(t) \tag{7.1.48}$$

It is left as an exercise for the reader to show that Φ_q is constant during the matter dominated era. Therefore, as we saw earlier in synchronous
gauge, the integrated Sachs-Wolfe effect receives contributions only from departures from strict matter dominance.

Finally, we consider the tensor contribution to temperature fluctuations. In the approximation of a sudden transition from thermal equilibrium to transparency at time t_L, the only contribution of tensor perturbations to the observed temperature anisotropy comes from the term $h_{rr}/2a^2$ in the definition (7.1.23) of N, which according to Eq. (5.1.33) contains a term $D_{rr}/2$. Using this in Eq. (7.1.29) gives the tensor contribution to the temperature fluctuation

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)^T = -\frac{\hat{n}_i\hat{n}_j}{2}\int_{t_L}^{t_0}\left\{\frac{\partial}{\partial t}D_{ij}(r\hat{n},t)\right\}_{r=s(t)} dt \tag{7.1.49}$$

The gravitational wave amplitude $D_{ij}(\mathbf{x},t)$ can be expressed as a Fourier integral (5.2.21)

$$D_{ij}(\mathbf{x},t) = \sum_{\lambda=\pm 2}\int d^3q e^{i\mathbf{q}\cdot\mathbf{x}}e_{ij}(\hat{q},\lambda)\beta(\mathbf{q},\lambda)\mathcal{D}_q(t) \tag{7.1.50}$$

where $\mathcal{D}_q(t)$ is the dominant solution of the wave equation (5.2.16):

$$\ddot{\mathcal{D}}_q(t) + 3\frac{\dot{a}}{a}\dot{\mathcal{D}}_q(t) + \frac{q^2}{a^2}\mathcal{D}_q(t) = 16\pi G\pi_q^T(t), \tag{7.1.51}$$

$e_{ij}(\lambda, \hat{q})$ are polarization tensors defined in Section 5.2, and $\beta(\mathbf{q}, \lambda)$ is a stochastic variable chosen to satisfy Eq. (5.2.22). The tensor mode contribution to the temperature fluctuation is then

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)^T = -\frac{1}{2}\sum_{\lambda=\pm 2}\int d^3q\hat{n}_i\hat{n}_j e_{ij}(\lambda, q)\beta(\mathbf{q}, \lambda)\int_{t_L}^{t_0} dt e^{is(t)\mathbf{q}\cdot\hat{n}}\dot{\mathcal{D}}_q(t) \tag{7.1.52}$$

This is the same as our previously derived result (7.1.18). We will return to the tensor term in C_ℓ in Section 7.3, after we have studied the scalar term in the next section.

10.7.2 7.2 Temperature multipole coefficients: Scalar modes

We will now apply the results of the previous section to the calculation of the contribution $C^S_{TT,\ell}$ of scalar modes to the multipole coefficients of temperature-temperature angular correlations:

$$C_{TT,\ell} = \frac{1}{4\pi}\int d^2\hat{n}\int d^2\hat{n}' P_\ell\left(\hat{n}\cdot\hat{n}'\right)\left\langle\Delta T(\hat{n})\Delta T\left(\hat{n}'\right)\right\rangle \tag{7.2.1}$$

where $\Delta T(\hat{n})$ is the stochastic variable giving the departure from the mean of the temperature observed in the direction $\hat{n}$, and $\langle\ldots\rangle$ denotes an average over the position of the observer, or equivalently, over the sequence of accidents that led to the particular pattern of temperature fluctuations we observe. (We are including a label TT to distinguish this multipole coefficient from those in temperature-polarization or polarization-polarization correlations, which are the subject of Section 7.4.) Of course this is not what is observed; the observed quantity is

$$C^{\rm obs}_{TT,\ell} = \frac{1}{4\pi}\int d^2\hat{n}\int d^2\hat{n}' P_\ell\left(\hat{n}\cdot\hat{n}'\right)\Delta T(\hat{n})\Delta T\left(\hat{n}'\right)$$

but as shown in Section 2.6, the cosmic variance, the mean square fractional difference between this and Eq. (7.2.1), is $2/2\ell+1$, and therefore may be neglected for $\ell \gg 1$. In this section we will consider only the contribution $C^S_{TT,\ell}$ of scalar modes to $C_{TT,\ell}$; as we saw in Section 5.2, tensor and scalar modes do not interfere, so we can take up the contribution of tensor modes separately in the following section.

First let's apply the results that were obtained in the previous section by using the kinetic theory approach described in Section 6.1. The use of the Boltzmann equation yields formulas (7.1.7)-(7.1.9) for the temperature fluctuation. To calculate the coefficients in a partial wave expansion of the temperature fluctuation, we use the familiar expansion (2.6.16) of a plane wave in Legendre polynomials, together with the addition theorem for spherical harmonics:

$$e^{i\hat{q}\cdot\hat{n}\rho} = 4\pi\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} i^\ell j_\ell(\rho)Y_\ell^m(\hat{n})Y_\ell^{m*}(\hat{q}). \tag{7.2.2}$$

Using this in Eqs. (7.1.7)-(7.1.9), and replacing factors of $i\hat{q}\cdot\hat{n}$ with derivatives of the spherical Bessel function j_ℓ, the scalar contribution to the temperature fluctuation observed in a direction $\hat{n}$ is given by

$$(\Delta T(\hat{n}))^{(S)} = \sum_{\ell m} a^S_{T,\ell m}Y_\ell^m(\hat{n}), \tag{7.2.3}$$

where

$$a^S_{T,\ell m} = 4\pi i^\ell T_0 \int d^3q \alpha(\mathbf{q}) Y^{m*}_\ell(\hat{q}) \int_{t_1}^{t_0} dt \\ \times \left[j_\ell(qr(t)) F(q,t) + j'_\ell(qr(t)) G(q,t) + j''_\ell(qr(t)) H(q,t) \right]. \tag{7.2.4}$$

Here $\alpha(\mathbf{q})$ is the stochastic parameter for the dominant scalar mode, normalized to satisfy Eq. (5.2.7):

$$\langle \alpha(\mathbf{q}) \alpha^* (\mathbf{q}') \rangle = \delta^3 (\mathbf{q} - \mathbf{q}') \tag{7.2.5}$$

t_1 is any time sufficiently early before recombination so that a photon present then would have scattered many times before the present; $r(t)$ is the radial coordinate of a point from which light emitted at time t would reach us at the present time t_0;

$$r(t) \equiv \int_t^{t_0} \frac{dt'}{a(t')} \tag{7.2.6}$$

and $F(q,t), G(q,t)$, and $H(q,t)$ are time-dependent form-factors, given by

$$\begin{aligned} F(q,t) =& \exp\left[-\int_t^{t_0} \omega_c(t') dt' \right] \\ & \times \left\{ \omega_c(t) \left[\frac{3}{4}\Phi(q,t) + \frac{3}{16}\Pi(q,t) \right.\right. \\ & \left. -\frac{1}{2}a^2(t)\ddot{B}_q(t) - \frac{1}{2}a(t)\dot{a}(t)\dot{B}_q(t) \right] \\ & \left. -\frac{1}{2}\frac{d}{dt}\left[A_q(t) + a^2(t)\ddot{B}_q(t) + a(t)\dot{a}(t)\dot{B}_q(t) \right] \right\} \end{aligned} \tag{7.2.7}$$

$$\begin{aligned} G(q,t) =& -q\omega_c(t) \exp\left[-\int_t^{t_0} \omega_c(t') dt' \right] \\ & \times \left[\delta u_{Bq}(t)/a(t) + a(t)\dot{B}_q(t)/2 \right] \end{aligned} \tag{7.2.8}$$

$$H(q,t) = \frac{3}{16}\omega_c(t) \exp\left[-\int_t^{t_0} \omega_c(t') dt' \right] \Pi(q,t) \tag{7.2.9}$$

As a reminder: $\omega_c(t)$ is the photon collision frequency at time t; $A_q(t)$ and $B_q(t)$ are the scalar fields in the perturbation to the metric in synchronous gauge, defined in Section 5.2; $\Phi(q,t)$ and $\Pi(q,t)$ are the scalar source functions, defined by Eq. (6.1.21); and $\delta u_{Bq}(t)$ is the scalar velocity potential of the baryonic plasma. A subscript T has been appended to the $a_{\ell m}$ introduced in Section 2.6, to indicate that these are partial wave coefficients in the temperature rather than the polarization, and a superscript S has been included to distinguish scalar from tensor contributions.

Together with the orthonormality property (2.6.19) of Legendre polynomials, Eqs. (7.2.1) and (7.2.3)-(7.2.5) give the scalar multipole coefficients [1]

$$C_{TT,\ell}^S = 16\pi^2 T_0^2 \int_0^\infty q^2 dq \times \mid \int_{t_1}^{t_0} dt \, [j_\ell(qr(t))F(q,t) + j'_\ell(qr(t))G(q,t) + j''_\ell(qr(t))H(q,t)] \mid^2 \quad (7.2.10)$$

This formula gives results of high accuracy, but the computer calculations used to calculate the multipole coefficients $C_{TT,\ell}^S$ in this way are not particularly revealing. Instead, we will apply a series of approximations that lead to a simple analytic formula for the $C_{TT,\ell}^S$.

First, we will neglect the integrated Sachs-Wolfe effect, given by the last term in Eq. (7.2.7). This effect is important only for relatively small values of ℓ, where cosmic variance intrudes on measurements of $C_{TT,\ell}^S$.

Next, we assume a sudden transition from perfect opacity to perfect transparency at a definite time t_L,[2] and a single dominant mode of perturbation. The fractional temperature fluctuation then takes the form (7.1.15):

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right) = \int d^3q \alpha(\mathbf{q}) e^{i\mathbf{q}\cdot\hat{n}r_L}(F(q) + i\hat{q}\cdot\hat{n}G(q)) \quad (7.2.11)$$

which was also derived at the end of Section 7.1 by following photon trajectories after the time of last scattering. (Here $r_L = r\,(t_L)$ is the co-moving radius of the surface of last scattering.) For the present we will not use formulas (7.1.16) and (7.1.17) for the form factors $F(q)$ and $G(q)$, but will proceed for general form factors, returning later to the specific form factors (7.1.16) and (7.1.17).

Using Eq. (7.2.2) in Eq. (7.2.11), and replacing $i\hat{q}\cdot\hat{n}$ in the Doppler term with $\partial/\partial\,(qr_L)$, we again have the partial-wave expansion (7.2.3), but this

time with

$$a_{T,\ell m}^S = 4\pi i^\ell T_0 \int d^3q \alpha(\mathbf{q}) Y_\ell^{m*}(\hat{q}) \left[j_\ell\,(qr_L)\,F(q) + j'_\ell\,(qr_L)\,G(q)\right]. \quad (7.2.12)$$

Inserting this in Eq. (7.2.1) and using Eq. (7.2.5) gives the multipole coefficients

$$C_{TT,\ell}^S = \frac{1}{2\ell+1}\sum_{m=-\ell}^{\ell} |a_{\ell m}|^2 = 16\pi^2 T_0^2 \int_0^\infty q^2 dq \times \left[j_\ell\,(qr_L)\,F(q) + j'_\ell\,(qr_L)\,G(q)\right]^2 \quad (7.2.13)$$

This is a standard result, but it does not provide transparent information about the dependence of $C_{TT,\ell}^S$ on ℓ. For this purpose, we now make a further approximation: we specialize to the most interesting case of large ℓ, where cosmic variance can be neglected. In this case, we can use an approximate formula for the spherical Bessel functions: [3]

[471] The derivatives of the spherical Bessel functions in Eq. (7.2.10) can be expressed as time derivatives, and then integrating by parts this can be written at the integral of a single form factor times $j_\ell(qr(t))$. The result is equivalent to Eq. (16) of M. Zaldarriaga and U. Seljak, Phys. Rev. D 55, 1830 (1997) [astro-ph/9609170], except for a difference of normalization: the source functions Π and Φ used here are 4 times those of Zaldarriaga and Seljak.

[2] To correct for whatever inaccuracy is introduced by this approximation, we will later include the effect of damping of acoustic oscillations before t_L, and average the temperature fluctuation over the time of last scattering.

$$j_\ell(\rho) \to \begin{cases} \cos b \cos[v(\tan b - b) - \pi/4]/v\sqrt{\sin b} & \rho > v \\ 0 & \rho < v \end{cases} \tag{7.2.14}$$

where $v \equiv \ell + 1/2$, and $\cos b \equiv v/\rho$, with $0 \le b \le \pi/2$. This approximation is valid for $|v^2 - \rho^2| \gg v^{4/3}$. Hence for $\ell \gg 1$, this formula can be used over most of the ranges of integration in Eq. (7.2.13). Furthermore, for $\rho > v \gg 1$ the phase $v(\tan b - b)$ in Eq. (7.2.14) is a very rapidly increasing function of ρ, so the derivative acting on the spherical Bessel function in Eq. (7.2.13) can be taken to act chiefly on this phase:

$$j'_\ell(\rho) \to \begin{cases} -\cos b\sqrt{\sin b} \sin[v(\tan b - b) - \pi/4]/v & \rho > v \\ 0 & \rho < v \end{cases} \tag{7.2.15}$$

(Letting the derivative act on the factor $1/\sqrt{\sin b}$ in Eq. (7.2.14) introduces an apparent divergence in the integral at $\rho = \nu$, but this divergence is spurious; for ρ very close to v the approximation (7.2.14) breaks down, and

there is no singularity.) Using these limits in Eq. (7.2.13) and changing the variable of integration from q to $b = \cos^{-1}(v/qr_L)$ gives

$$\begin{aligned} C^S_{TT,\ell} = & \frac{16\pi^2 v}{r_L^3} \int_0^{\pi/2} \frac{db}{\cos^2 b} \\ & \times \left[F\left(\frac{v}{r_L \cos b}\right) \cos[v(\tan b - b) - \pi/4] \right. \\ & \left. - \sin b G\left(\frac{v}{r_L \cos b}\right) \sin[v(\tan b - b) - \pi/4] \right]^2 \end{aligned} \tag{7.2.16}$$

For $v \gg 1$ the functions $\cos[v(\tan b - b) - \pi/4]$ and $\sin[v(\tan b - b) - \pi/4]$ oscillate very rapidly, so $\cos^2[\nu(\tan b - b) - \pi/4]$ and $\sin^2[\nu(\tan b - b) - \pi/4]$ average to 1/2, while $\cos[\nu(\tan b - b) - \pi/4] \sin[\nu(\tan b - b) - \pi/4]$ averages to zero. Dropping the distinction between ℓ and $v = \ell + 1/2$, and changing the variable of integration again, from b to $\beta = 1/\cos b$, Eq. (7.2.16) then becomes [4]

$$\begin{aligned} \ell(\ell+1) C^S_{TT,\ell} = & \frac{8\pi^2 \ell^3}{r_L^3} \int_1^\infty \frac{\beta d\beta}{\sqrt{\beta^2 - 1}} \\ & \times \left[F^2\left(\frac{\ell\beta}{r_L}\right) + \frac{\beta^2 - 1}{\beta^2} G^2\left(\frac{\ell\beta}{r_L}\right) \right]. \end{aligned} \tag{7.2.17}$$

We will see that the form factors $F(q)$ and $G(q)$ fall off rapidly for large q, in part because $\left|\mathcal{R}^o_q\right|^2$ decreases more or less like q^{-3}, so the integral over β converges at $\beta = \infty$, and in fact is dominated by small values of β. The integral of the F^2 term thus receives its greatest contribution from $\beta \approx 1$, or in other words, for $q \approx \ell/r_L$, or $q/a_L \approx \ell/d_A$, where $d_A = r_L a_L$ is the angular diameter distance of the surface of last scattering. On the other hand, the factor $\beta^2 - 1$ multiplying G^2 in Eq. (7.2.17) kills the contribution of β values very close to unity, so the Doppler term proportional to G^2 makes a relatively small contribution to the multipole coefficients.

Even without a detailed calculation of the form factors $F(q)$ and $G(q)$, we know that they depend on the baryon and total matter densities at last scattering, which for a given present

[473] See, e.g., I. S. Gradshteyn & I. M. Ryzhik, Table of Integrals, Series, and Products, translated, corrected and enlarged by A. Jeffrey (Academic Press, New York, 1980): formula 8.453.1.

microwave background temperature can be expressed in terms of $\Omega_B h^2$ and $\Omega_M h^2$, but since spatial curvature and dark energy are (presumably) negligible at last scattering, the form factors cannot depend on Ω_K or Ω_Λ or H_0. Thus there is a high

degree of degeneracy here; all dependence on Ω_K or Ω_Λ or H_0 can only enter in the single parameter r_L, the Robertson-Walker radius of the surface of last scattering. If we assume as commonly done that $\Omega_K = 0$, so that $\Omega_\Lambda \simeq 1 - \Omega_M$, then temperature anisotropies can tell us the values of H_0 as well as Ω_M and Ω_B, but they cannot tell us that $\Omega_K = 0$ unless we have information about H_0 from other source. Likewise, they cannot distinguish quintessence from a constant vacuum energy. Furthermore, r_L and ℓ appear in $\ell(\ell+1)C^S_{TT,\ell}$ only in the ratio ℓ/r_L, so the values of Ω_K, Ω_Λ, and H_0 can only effect the scale of the ℓ-dependence of $\ell(\ell+1)C^S_{TT,\ell}$. For instance, the values of these parameters can affect the positions of the peaks in $\ell(\ell+1)C^S_{TT,\ell}$ by a common factor, but cannot affect their heights.

We next make the approximation that the gravitational field perturbations at last scattering are dominated by perturbations in the dark matter density. We already remarked in connection with the ISW effect in the previous section that in this case $\dot{A}_q$ vanishes (because each term on the right-hand side of Eq. (5.3.38) is time-independent). The field ψ_q whose evolution we followed in Chapter 6 is defined in general as $\psi_q \equiv \left(3\dot{A}_q - q^2\dot{B}_q\right)/2$, so here $\dot{B}_q = -2\psi_q/q^2$, and since Eq. (6.5.16) gives $\psi_q \propto t^{-1/3}$, we have also $\ddot{B}_q = 2\psi_q/3tq^2$. Also, $a \propto t^{2/3}$, so Eqs. (7.1.16) and (7.1.17) give the form factors as

$$F(q) = \frac{1}{3}\delta_{\gamma q}(t_L) + \frac{a^2(t_L)\,\psi_q(t_L)}{3q^2 t_L} \tag{7.2.18}$$

$$G(q) = -q\delta u_{\gamma q}(t_L)/a(t_L) + a(t_L)\,\psi_q(t_L)/q \tag{7.2.19}$$

in which we have used $\delta T_q/\bar{T} = \delta\rho_{\gamma q}/4\bar{\rho}_\gamma = \delta_{\gamma q}/3$.
As our next approximation, we shall use the results (6.5.16), (6.5.17), and (6.5.18) of our analysis of cosmic evolution in Chapter 6, in which we neglected the baryon/dark matter density ratio:

$$\psi_q(t_L) = -\frac{3q^2 t_L \mathcal{R}^o_q \mathcal{T}(\kappa)}{5a_L^2} \tag{7.2.20}$$

$$\begin{aligned}\delta_{\gamma q}(t_L) =& \frac{3\mathcal{R}^o_q}{5}\left[\mathcal{T}(\kappa)(1+3R_L)\right.\\ &-(1+R_L)^{-1/4}\,e^{-\int_0^t L dt}\mathcal{S}(\kappa)\\ &\left.\times\cos\left(\int_0^{t_L}\frac{q dt}{a(t)\sqrt{3(1+R(t))}}+\Delta(\kappa)\right)\right]\end{aligned} \tag{7.2.21}$$

$$\begin{aligned}\delta u_{\gamma q}(t_L) =& \frac{3\mathcal{R}^o_q}{5}\left[-t_L\mathcal{T}(\kappa)\right.\\ &+\frac{a_L}{\sqrt{3}q(1+R_L)^{3/4}}e^{-\int_0^t L^2\Gamma dt}\mathcal{S}(\kappa)\\ &\left.\times\sin\left(\int_0^{t_L}\frac{q dt}{a(t)\sqrt{3(1+R(t))}}+\Delta(\kappa)\right)\right]\end{aligned} \tag{7.2.22}$$

[474] A more rigorous but rather more complicated derivation was given by S. Weinberg, Phys. Rev. D 64, 123512 (2001) [astro-ph/0103281]. The contribution of $F(q)$ had earlier been calculated by J. R. Bond, "Theory and Observations of the Cosmic Background Radiation," in Cosmology and Large Scale Structure, eds. R. Schaeffer, J. Silk, M. Spiro and J. Zinn-Justin (Elsevier, 1996), Section 5.1.3.

Here $R(t) \equiv 3\bar{\rho}_B(t)/4\bar{\rho}_\gamma(t), R_L \equiv R(t_L)$, and $a_L \equiv a(t_L); \mathcal{T}(\kappa), \mathcal{S}(\kappa)$, and $\Delta(\kappa)$ are the transfer functions defined and calculated in Section 6.5; $\kappa \equiv \sqrt{2}q/q_{\rm EQ}$, where $q_{\rm EQ}$ is the wave number that comes into the horizon at matter-radiation equality; and $\Gamma(t)$ is the acoustic damping rate (6.4.25). Using Eqs. (7.2.20)-(7.2.22) in Eqs. (7.2.18) and (7.2.19) gives the form factors as

$$\begin{aligned} F(q) =& \frac{\mathcal{R}_q^o}{5}\left[3\mathcal{T}(\kappa)R_L\right. \\ & - (1+R_L)^{-1/4}\, e^{-\int_0^{t_L}\Gamma dt}\mathcal{S}(\kappa) \\ & \left.\times\cos\left(\int_0^{t_L}\frac{qdt}{a(t)\sqrt{3(1+R(t))}} + \Delta(\kappa)\right)\right] \end{aligned} \tag{7.2.23}$$

$$\begin{aligned} G(q) =& -\frac{\sqrt{3}\mathcal{R}_q^o}{5\,(1+R_L)^{3/4}} e^{-\int_0^{t_L}\Gamma dt}\mathcal{S}(\kappa) \\ & \times\sin\left(\int_0^{t_L}\frac{qdt}{a(t)\sqrt{3(1+R(t))}} + \Delta(\kappa)\right) \end{aligned} \tag{7.2.24}$$

Note that the "slow" terms have canceled in $G(q)$, and would have canceled in $F(q)$ if it were not for a finite baryon/photon density ratio at last scattering.

Now we must take up a complication that arises only for the "fast" part of the form factors in the case of short wavelengths. We have been assuming that the opacity of the universe drops to zero instantaneously at a time t_L of last scattering, but of course the drop takes place during some finite interval of time, over which the form factors must be averaged. This makes little difference for the contribution of the slow modes, but for large wave numbers the cosines and sines in Eqs. (7.2.23) and (7.2.24) are rapidly oscillating functions of t_L, so the fast terms can be significantly reduced by this averaging. This is similar to what in other contexts is called Landau damping, except that usually Landau damping arises from a spread in the frequency of an oscillation, while here it is produced by a spread in the moment at which the oscillating amplitude is observed.

To continue with our analytic treatment, since the probability of last scattering is a sharply peaked function of time, we can approximate it as a Gaussian: the probability that last scattering occurs between t and $t+dt$ will be taken in the form

$$P(t)dt = \frac{\exp\left[-(t-t_L)^2/2\sigma_t^2\right]dt}{\sigma_t\sqrt{2\pi}} \tag{7.2.25}$$

In Eqs. (7.2.23) and (7.2.24), we must make the replacements

$$\left\{\begin{array}{c}\cos\left(\int_0^{t_L}\omega dt+\Delta\right)\\ \sin\left(\int_0^{t_L}\omega dt+\Delta\right)\end{array}\right\} \to \int_{-\infty}^{+\infty} P(t)dt \left\{\begin{array}{c}\cos\left(\int_0^{t}\omega dt+\Delta\right)\\ \sin\left(\int_0^{t}\omega dt+\Delta\right)\end{array}\right\} \tag{7.2.26}$$

where $\omega = q/a\sqrt{3(1+R)}$. For a sharply peaked distribution function $P(t)$, we can do these integrals by expanding the arguments of sines and cosines to first order in $t - t_L$:

$$\int_0^t \omega dt \simeq \int_0^{t_L}\omega dt + \omega_L\,(t-t_L)$$

The integrals (7.2.26) are now easily done

$$\int_{-\infty}^{+\infty} P(t)dt \left\{ \begin{array}{c} \cos\left(\int_0^t \omega dt + \Delta\right) \\ \sin\left(\int_0^t \omega dt + \Delta\right) \end{array} \right\}$$
$$\simeq \exp\left(-\omega_L^2\sigma_t^2/2\right) \left\{ \begin{array}{c} \cos\left(\int_0^{t_L} \omega dt + \Delta\right) \\ \sin\left(\int_0^{t_L} \omega dt + \Delta\right) \end{array} \right\} \tag{7.2.27}$$

Thus, the whole effect of this averaging is to introduce an additional damping factor $\exp\left(-\omega_L^2\sigma_t^2/2\right)$ in the fast part of the form factors. Both Γ and ω_L^2 are proportional to q^2, so we may write

$$\int_0^{t_L} \Gamma dt + \omega_L^2\sigma_t^2/2 = q^2 d_D^2/a_L^2 \tag{7.2.28}$$

where d_D is a damping length, given by Eqs. (6.4.25) and (7.2.27) as

$$d_D^2 = d_{\rm Silk}^2 + d_{\rm Landau}^2 \tag{7.2.29}$$
$$d_{\rm Silk}^2 = a_L^2 \int_0^{t_L} \frac{t_\gamma}{6a^2(1+R)} \left\{ \frac{16}{15} + \frac{R^2}{(1+R)} \right\} dt \tag{7.2.30}$$
$$d_{\rm Landau}^2 = \frac{\sigma_t^2}{6\left(1+R_L\right)} \tag{7.2.31}$$

with $t_\gamma = 1/\omega_c$ the photon mean free time, and $R \equiv 3\bar{\rho}_B/4\bar{\rho}_\gamma$.
To evaluate the Silk damping term, we recall that $R \propto a$, so

$$t_\gamma = \frac{1}{n_e \sigma_{\mathcal{T}} c} = \frac{R^3}{n_{B0} R_0^3 (1-Y) X \sigma_{\mathcal{T}} c} \tag{7.2.32}$$

where $R_0 = 3\Omega_B/4\Omega_\gamma$ is the present value of $R, Y \simeq 0.24$ is the fraction of nucleons in the form of un-ionized helium around the time of last scattering, $n_{B0} = 3H_0^2\Omega_B/8\pi G m_N$ is the present number density of baryons, and $X(R)$ is the fractional ionization, calculated in Section 2.3. Also,

$$dt = \frac{dR}{RH_0\sqrt{\Omega_M\left(R_0/R\right)^3 + \Omega_R\left(R_0/R\right)^4}} = \frac{RdR}{H_0\sqrt{\Omega_M}R_0^{3/2}\sqrt{R_{\rm EQ}+R}} \tag{7.2.33}$$

where $R_{\rm EQ} \equiv \Omega_R R_0/\Omega_M = 3\Omega_R\Omega_B/4\Omega_M\Omega_\gamma$ is the value of R at matterradiation equality. Putting this all together, the Silk damping length is given by

$$d_{\rm Silk}^2 = \frac{R_L^2}{6(1-Y)n_{B0}\sigma_{\mathcal{T}} c H_0\sqrt{\Omega_M}R_0^{9/2}} \times \int_0^{R_L} \frac{R^2 dR}{X(R)(1+R)\sqrt{R_{\rm EQ}+R}} \left\{ \frac{16}{15} + \frac{R^2}{(1+R)} \right\} \tag{7.2.34}$$

Also, the standard deviation σ_t in the time of last scattering is related to the standard deviation σ in the temperature of last scattering, calculated in Section 2.3, by $\sigma_t = 3t_L\sigma/2T_L$, so

$$d_{\rm Landau}^2 = \frac{3\sigma^2 t_L^2}{8T_L^2\left(1+R_L\right)} \tag{7.2.35}$$

The form factors (7.2.23) and (7.2.24) may now be written as explicit functions of wave number

$$F(q) = \frac{\mathcal{R}_q^o}{5}\left[3T\left(qd_T/a_L\right)R_L \right. \\ \left. -\left(1+R_L\right)^{-1/4}\mathcal{S}\left(qd_T/a_L\right)e^{-q^2d_D^2/a_L^2}\cos\left(qd_H/a_L+\Delta\left(qd_T/a_L\right)\right)\right] \tag{7.2.36}$$

$$G(q) = -\frac{\sqrt{3}\mathcal{R}_q^o}{5\left(1+R_L\right)^{3/4}}e^{-q^2d_D^2/a_L^2}\mathcal{S}\left(qd_T/a_L\right)\sin\left(qd_H/a_L+\Delta\left(qd_T/a_L\right)\right) \tag{7.2.37}$$

where d_T is a length defined by writing the argument of the transfer functions as $\kappa = qd_T/a_L$, so that Eq. (6.5.9) gives

$$d_T = \frac{\sqrt{\Omega_R}}{\left(1+z_L\right)H_0\Omega_M} = \frac{0.0177}{\Omega_M h^2}\text{Mpc} \tag{7.2.38}$$

Also, d_H is the acoustic horizon distance at last scattering, given by

$$d_H \equiv a_L\int_0^{t_L}\frac{dt}{a\sqrt{3(1+R)}} \\ = \frac{2}{H_0\sqrt{3R_L\Omega_M}\left(1+z_L\right)^{3/2}}\ln\left(\frac{\sqrt{1+R_L}+\sqrt{R_{\rm EQ}+R_L}}{1+\sqrt{R_{\rm EQ}}}\right) \tag{7.2.39}$$

where again $R_{\rm EQ} = 3\Omega_R\Omega_B/4\Omega_M\Omega_\gamma$ and $R_L = 3\Omega_B/4\Omega_\gamma\left(1+z_L\right)$.
Before using these form factors in Eq. (7.2.17), there is one more complication that needs to be mentioned. At a redshift $z_{\rm reion}$ of order 10 , the neutral hydrogen left over from the time of recombination becomes reionized by ultraviolet light from the first generation of massive stars. The photons of the cosmic microwave background have a small but nonnegligible probability $1-\exp\left(-\tau_{\rm reion}\right)$ (where $\tau_{\rm reion}$ is the optical depth of the reionized plasma) of being scattered by the electrons set free by this reionization. The temperature anisotropy ΔT is then the sum of two terms. One arises from photons that are not scattered by the reionized hydrogen, and is just equal to the anisotropy we have calculated times the probability $\exp\left(-\tau_{\rm reion}\right)$ of no scattering. The other term arises from scattered photons, and since this scattering occurs at redshifts much less than $z_L \simeq 1,090$, we see the anisotropies at a smaller distance, and hence at much lower values of ℓ. Thus the effect on $C_{TT,\ell}^S$ of scattering by the reionized plasma is simply to multiply $C_{TT,\ell}^S$ by a factor $\exp\left(-2\tau_{\rm reion}\right)$, except for very small values of ℓ, where in any case cosmic variance interferes with the interpretation of observations. This means that observations of temperature anisotropies alone cannot effectively disentangle the reionization probability from the over-all scale of the function $\mathcal{R}_q^0$ that characterizes primordial fluctuations; they can only tell us the value of $\left|\mathcal{R}_q^o\right|^2\exp\left(-2\tau_{\rm reion}\right)$. The measurement of the polarization of microwave photons (discussed in Section 7.4) produced by scattering after reionization suggests that $\exp\left(-2\tau_{\rm reion}\right)\approx 0.8$.

It is conventional to parameterize the quantity $\mathcal{R}_q^o$ in a form equivalent to

$$\left|\mathcal{R}_q^o\right|^2 = N^2q^{-3}\left(\frac{q/a_0}{k_{\mathcal{R}}}\right)^{n_S-1} \tag{7.2.40}$$

with n_S perhaps varying with wave number. This parameterization is convenient, because then

$$\left(\frac{\ell}{r_L}\right)^3 \left|\mathcal{R}^o_{\beta\ell/r_L}\right|^2 = \frac{N^2}{\beta^3}\left(\frac{\ell\beta}{k_{\mathcal{R}} d_A\,(1+z_L)}\right)^{n_S-1}$$

and it will turn out that n_S is not very different from 1 . It is only $N^2 k_{\mathcal{R}}^{1-n_S}$ that enters in the normalization of $\left|\mathcal{R}^o_q\right|^2$, so the choice of $k_{\mathcal{R}}$ is arbitrary; it is conventional to take it as $k_{\mathcal{R}} = 0.05\text{Mpc}^{-1}$.

We conclude from all this that for reasonably large values of ℓ (say, $\ell > 20$), where we can ignore cosmic variance and the integrated Sachs-Wolfe effect, use the large ℓ approximations that led to Eq. (7.2.17), and treat the effect of reionization as a simple factor $\exp(-2\tau_{\text{reion}})$, the quantity usually quoted as giving the scalar contribution to the multipole coefficients is

$$\begin{aligned}\frac{\ell(\ell+1)C^S_{TT,\ell}}{2\pi} &= \frac{4\pi T_0^2 N^2 e^{-2\tau_{\text{reion}}}}{25}\int_1^\infty d\beta \left(\frac{\beta\ell}{\ell_{\mathcal{R}}}\right)^{n_S-1} \\ &\times\left\{\frac{1}{\beta^2\sqrt{\beta^2-1}}\left[3\mathcal{T}\left(\beta\ell/\ell_T\right)R_L\right.\right. \\ &\left.-(1+R_L)^{-1/4}\,\mathcal{S}\left(\beta\ell/\ell_T\right)e^{-\beta^2\ell^2/\ell_D^2}\cos\left(\beta\ell/\ell_H+\Delta\left(\beta\ell/\ell_T\right)\right)\right]^2 \\ &\left.+\frac{3\sqrt{\beta^2-1}}{\beta^4\,(1+R_L)^{3/2}}e^{-2\beta^2\ell^2/\ell_D^2}\mathcal{S}^2\left(\beta\ell/\ell_T\right)\sin^2\left(\beta\ell/\ell_H+\Delta\left(\beta\ell/\ell_T\right)\right)\right\} \end{aligned} \tag{7.2.41}$$

where

$$\ell_D \equiv d_A/d_D, \quad \ell_T \equiv d_A/d_T, \quad \ell_H \equiv d_A/d_H, \quad \ell_{\mathcal{R}} = (1+z_L)\,k_{\mathcal{R}} d_A \tag{7.2.42}$$

and again d_A is the angular diameter distance of the surface of last scattering:
$d_A \equiv r_L a_L = \frac{1}{\Omega_K^{1/2} H_0(1+z_L)}\sinh\left[\Omega_K^{1/2}\int_{1/(1+z_L)}^1 \frac{dx}{\sqrt{\Omega_\Lambda x^4+\Omega_K x^2+\Omega_M x}}\right]$,
with Ω_R neglected, and $\Omega_K = 1-\Omega_\Lambda-\Omega_M$. If we assume as discussed earlier that the integral over β is dominated by values $\beta \approx 1$, neglect the Doppler term, and for the moment neglect the term proportional to the transfer function $\mathcal{T}$, then Eq. (7.2.41) shows that $C^S_{TT,\ell}$ has peaks at $\chi(\ell) =$
$\pi, 2\pi, 3\pi$, etc., where $\chi(\ell) = \ell/\ell_H + \Delta\left(\ell/\ell_T\right)$ is the phase of the cosine in Eq. (7.2.41) for $\beta = 1$. The presence of the positive term $3\mathcal{T}R_L$ arising from the inertia of the baryonic plasma enhances and slightly shifts the peaks for $\chi = \pi, 3\pi$, etc., where the cosine is negative, and reduces the peaks for $\chi = 2\pi$, 4π, etc., where the cosine is positive. For large ℓ the Silk and Landau damping factor $\exp(-\beta^2\ell^2/d_D^2)$ damps out all these peaks. As we will soon see, the results of numerical calculations exhibit this expected pattern, of a sequence of decreasing peaks, with odd peaks somewhat enhanced over the even peaks.

We can now read off the dependence of the quantities appearing in $C^S_{TT,\ell}$ on various cosmological parameters. Taking as fixed the well established value of the present microwave temperature, (which yields values for $\Omega_\gamma h^2$ and $\Omega_R h^2 = \Omega_\gamma h^2 + \Omega_v h^2$), and also fixing the values of t_L, z_L, and σ, which are only weakly dependent on other cosmological parameters, we see that

- $R_L \propto \Omega_B h^2$.
- The integral in Eq. (7.2.34) for d^2_{Silk} is a complicated but not very sensitive function of $\Omega_B h^2$ and $\Omega_M h^2$. Aside from the integral, d^2_{Silk} is proportional to $\left(\Omega_B h^2\right)^{-7/2}\left(\Omega_M h^2\right)^{-1/2}$.
- d^2_{Landau} depends on $\Omega_B h^2$ through a factor $(1+R_L)^{-1}$.

- $d_T \propto (\Omega_M h^2)^{-1}$
- Aside from a slowly varying logarithm, $d_H \propto (\Omega_B h^2)^{-1/2} (\Omega_M h^2)^{-1/2}$.
- Only d_A depends on $H_0, \Omega_\Lambda, \Omega_M, \Omega_B$, or Ω_K apart from a dependence on $\Omega_B h^2$ and $\Omega_M h^2$. (For any observationally allowed values of Ω_Λ or Ω_K, the effects of a constant vacuum energy or of spatial curvature would be quite negligible at times before recombination, so their effect on $C^S_{TT,\ell}$ is limited to their influence on the propagation of light after recombination, that is, on d_A.)

To see how well the various approximations we have made work in practice, we shall calculate $C^S_{TT,\ell}$ for a realistic set of values for cosmological parameters taken from a fit [5] to data on the microwave background from the CBI, ACBAR, and first-year WMAP observations (about which more below). These are the same parameters that have been used in a full-scale computer calculation whose results are readily available, [6] so that we will

conveniently be able to compare Eq. (7.2.41) with an accurate numerical calculation. The cosmological parameters of this set are

$$\Omega_M h^2 = 0.13299, \quad \Omega_B h^2 = 0.02238, \quad h = 0.71992, \quad \Omega_\Lambda = 1 - \Omega_M. \tag{7.2.44}$$

and [7]

$$\begin{aligned} n_S =& 0.95820, \quad k_{\mathcal{R}} = 0.05\text{Mpc}^{-1}, \quad N^2 = 1.736 \times 10^{-10} \\ & e^{-2\tau_{\text{reion}}} = 0.80209 \end{aligned} \tag{7.2.45}$$

We take $T_0 = 2.725$ K, which yields $\Omega_\gamma h^2 = 2.47 \times 10^{-5}$, and we assume three flavors of massless neutrinos, which gives $\Omega_R h^2 = 1.6813 \Omega_\gamma h^2 = 4.15 \times 10^{-5}$. We will also adopt the parameters describing recombination calculated in Section 2.3 (which are not very sensitive to other cosmological parameters) for $\Omega_B h^2 = 0.02$ and $\Omega_M h^2 = 0.15$; in particular,

$$1 + z_L = 1,090, \quad \sigma = 262 \text{ K}, \quad t_L = 370,000\text{yrs}. \tag{7.2.46}$$

From the values (7.2.44) of $\Omega_M h^2$ and $\Omega_B h^2$, we find

$$R_0 = 679.6, \quad R_L = 0.6234, \quad R_{\text{EQ}} = 0.2121. \tag{7.2.47}$$

Then Eqs. (7.2.38), (7.2.39), (7.2.43) give

$$d_T = 0.1331\text{Mpc}, \quad d_H = 0.1351\text{Mpc}, \quad d_A = 12.99\text{Mpc}, \tag{7.2.48}$$

while the damping lengths are given by Eqs. (7.2.34), (7.2.35) and (7.2.29):
$d_{\text{Silk}} = 0.006555\text{Mpc}$, $d_{\text{Landau}} = 0.004809\text{Mpc}$, $d_D = 0.008130\text{Mpc}$.

Finally, the parameters (7.2.42) appearing in Eq. (7.2.41) are

$$\ell_T = 97.60, \quad \ell_H = 96.15, \quad \ell_D = 1598, \quad \ell_{\mathcal{R}} = 708 \tag{7.2.50}$$

while the factor multiplying the integral is

$$\frac{4\pi T_0^2 N^2 e^{-2\tau_{\text{reion}}}}{25} = 519.7\mu \text{ K}^2. \tag{7.2.51}$$

[475] D.N. Spergel et al., Astrophys. J. Suppl. 148, 175 (2003) [astro-ph/0302209].
[6] http://lambda.gsfc.nasa.gov/data/map/powspec/wmap_lcdm_pl_model_yr1_ v1.txt

Using these values, the result of using Eq. (7.2.41) with the parameter set of reference 6 is

$$\frac{\ell(\ell+1)C^S_{TT,\ell}}{2\pi} = 519.7\mu K^2 \int_1^\infty d\beta \left(\frac{\beta\ell}{708}\right)^{-0.0418}$$
$$\times \left\{ \frac{1}{\beta^2\sqrt{\beta^2-1}}[3(0.6234)\mathcal{T}(\beta\ell/97.6)\right.$$
$$-(1.6234)^{-1/4}\mathcal{S}(\beta\ell/97.6)e^{-(\beta\ell/1598)^2}\cos(\beta\ell/96.15+\Delta(\beta\ell/97.6))\Big]^2$$
$$+\frac{3\sqrt{\beta^2-1}}{\beta^4(1.6234)^{3/2}}e^{-2(\beta\ell/1598)^2}\mathcal{S}^2(\beta\ell/97.6)$$
$$\left.\times\sin^2(\beta\ell/96.15+\Delta(\beta\ell/97.6))\right\} \qquad (7.2.52)$$

The integral over β converges very rapidly, and can be done with a cut-off at $\beta = 5$; raising the cut-off to $\beta = 50$ has a negligible effect.

The results are shown in Figure 7.1, in comparison with the more accurate Boltzmann hierarchy calculation of reference 6, based on Eq. (7.2.10). Evidently the hydrodynamic calculation does quite well; like the computer calculation of reference 6, it shows a high first peak, followed by two nearly equal lower peaks, followed by a decaying tail punctuated by successively lower peaks. To give a quantitative comparison, Tables 7.1 and 7.2 show the

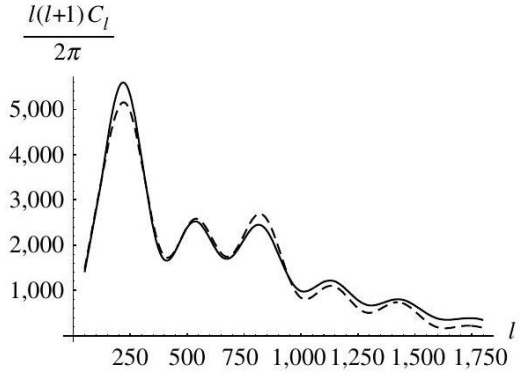


Figure 7.1: The scalar multipole coefficient $\ell(\ell+1) C_{\ell}^{S} / 2 \pi$ in square microKelvin, vs. ℓ, for the cosmological parameters given in Eqs. (7.2.44) and (7.2.45). The hydrodynamic approximation (7.2.41) is indicated by the dashed curve, while for comparison the solid curve gives the more accurate large scale computer calculation of reference 6, based on Eq. (7.2.10).

Table 7.1: Comparison of the results for the values of ℓ at the first five peaks in $\ell(\ell+1) C_{TT,\ell}^{S} / 2 \pi$, as given by Eq. (7.2.52) and as given by reference 6 .

Eq. (7.2.52):	220	541	821	1134	1425
reference 6:	219	536	813	1127	1425

[477] The parameter N^2 is related to what is often given as an "amplitude" A by $4\pi N^2 \equiv 20,000\pi A/9T_{\gamma 0}(\mu K)^2 = 2.95 \times 10^{-9}A$. For instance, see L. Verde et al., Astrophys. J. Suppl. 148, 195 (2003) [astro-ph/0302218]. It should be noted that, because of different conventions used in writing Fourier integrals, the quantity $\mathcal{R}_k$ in this paper is equal to what we have defined as $\mathcal{R}^o_{a_0 k}$ times a factor $(2\pi)^{3/2}$. Reference 6 uses $A = 0.73935$, corresponding to the value of N^2 given in Eq. (7.2.45).

Table 7.2: Comparison of the results for the values of $\ell(\ell+1) C_{T T, \ell}^{S} / 2 \pi$ (in μK^2) at the peaks listed in Table 7.1, as given by Eq. (7.2.52) and as given by reference 6.

Eq. (7.2.52):	5155	2694	2783	1126	746
Reference 6:	5591	2525	2451	1221	806

peak positions and heights calculated using Eq. (7.2.52) and those given by the results of reference 6. In all cases, the peak heights given by Eq. (7.2.52) agree with the more accurate computer results to within 10%, while the results of Eq. (7.2.52) for the peak positions are almost embarrassingly good, in no case being off by more than 1%.

Among other things, the general success of the calculations of this section shows that the evolution of cosmological perturbations is primarily hydrodynamic, in the sense that it can be well described by the equations of hydrodynamics without the full apparatus of coupled Boltzmann equations used in computer calculations like those of reference 6. The Boltzmann equation is implicit in our calculations, because in calculating the Silk damping rate we have used standard values for the shear viscosity and heat conductivity that were obtained by using the Boltzmann equation for photons in an ionized gas, but evidently not much is lost by not solving the Boltzmann equation over again in a cosmological context.

To the extent that the general formula (7.2.41) has been validated by this comparison, we can use it to see what can be learned from measurements of $C_{TT,\ell}$:

- Eq. (7.2.41) shows that the shape of the function $\ell(\ell+1)C^S_{TT,\ell}$, as for instance the set of ratios of the peak positions or of the peak heights, depends on only four quantities: d_D, d_H, d_T, and n_S. Also, with the present radiation temperature and the number of massless neutrinos fixed, the three lengths d_D, d_H, and d_T depend only on the two cosmological parameters $\Omega_B h^2$ and $\Omega_M h^2$, which can therefore be found from the shape of the function $\ell(\ell+1)C^S_{TT,\ell}$.

- Measurement of the scale of this function's ℓ-dependence (as for instance, the ℓ value of any one peak) depends on these two parameters, but also on d_A, which is a function not only of $\Omega_B h^2$ and $\Omega_M h^2$ but also of h and Ω_Λ. It is therefore impossible, without taking the integrated Sachs-Wolfe effect into account, to use measurements of $C^S_{TT,\ell}$ alone to make separate determinations of both h and Ω_Λ. It is often simply assumed that the universe is spatially flat, in which case to a good approximation $\Omega_\Lambda = 1 - \Omega_M$, and then measurements of $C^S_{TT,\ell}$ can be used to determine h as well as $\Omega_B h^2$ and $\Omega_M h^2$.

- Measurement of the magnitude of the function $\ell(\ell+1)C_{TT,\ell}$ for $\ell \gg 1$ (as for instance measurement of the height of any one peak) only tells us the quantity $N^2 \exp(-2\tau_{\rm reion})$, so as already remarked we cannot use it to make a separate determination of N^2 or $\tau_{\rm reion}$. This ambiguity is resolved by measurements of the polarization of the cosmic microwave background, discussed in Section 7.4.

The original discovery of microwave background anisotropy was made by the COBE collaboration, and is discussed in Section 2.6. This only gave information about the anisotropy for relatively small ℓ, well below the position of the first acoustic peak at $\ell \approx 200$. This discovery was followed by a series of balloon-borne and ground-based observations, [8] which gave definite evidence for the first acoustic peak, and some data on higher peaks, extending in the case of the CBI collaboration up to values of ℓ beyond the position of the fifth acoustic peak, at $\ell \approx 1400$. The accuracy of these measurements up to about $\ell \approx 600$ was then greatly improved by observations made by a remarkable satellite mission, known as the Wilkinson Microwave Anisotropy Probe, or WMAP. [9] The WMAP satellite was launched on June 30, 2001, made loops around

the moon to pick up kinetic energy from the Moon's motion, and finally reached an orbit about the equilibrium point known as L2. This point orbits the Sun at the speed needed to keep it about 1.5×10^6 km from Earth, on the other side of the earth from the sun, a location chosen to isolate the instrument from microwave radiation from the sun, earth, or moon. The satellite carries two back-to-back 1.4×1.6 meter

microwave receivers, cooled by thermal radiators to about 90 K , and measures the polarization of the microwave background as well as differences in its temperature over the whole sky.

After a year of observation, $C_{TT,\ell}$ had been measured out to $\ell \simeq 600$ with small errors arising mostly from cosmic variance and foreground emission. [10] These measurements were fit to the results of the " Λ CDM" model we have been studying, with zero curvature; the contents of the universe supposed to consist of photons, three flavors of massless neutrinos, baryonic matter, cold dark matter, and a constant vacuum energy; and a primordial spectrum of purely adiabatic fluctuations given by Eq. (7.2.40), with n_S constant. The values of cosmological parameters derived from this fit were: [11]

- $\Omega_B h^2 = 0.024 \pm 0.001$
- $\Omega_M h^2 = 0.14 \pm 0.02$
- $h = 0.72 \pm 0.05$
- $|N|^2 = (2.1 \pm 0.2) \times 10^{-10}$
- $n_S = 0.99 \pm 0.04$
- $\tau_{\rm reion} = 0.166^{+0.076}_{-0.071}$
  (Measurements of polarization were used here chiefly to measure the optical depth $\tau_{\rm reion}$ of the reionized plasma, [12] which is needed to obtain $|N|^2$ from the value of $|N|^2 \exp(-2\tau_{\rm reion})$ given by the measured values of $C^S_{TT,\ell}$.)

In March 2006 the WMAP group announced the results of the second and third years of observation. [13] The results are shown together with a best fit of $\ell(\ell+1)C^S_{TT,\ell}/2\pi$ to the Λ CDM model (with zero tensor anisotropies) in Figure 7.2. This fit gave the parameters: [14]

- $\Omega_B h^2 = 0.02229 \pm 0.00073$
- $\Omega_M h^2 = 0.1277^{+0.0080}_{-0.0079}$
- $h = 0.732^{+0.031}_{-0.032}$
- $|N|^2 = (1.93 \pm 0.12) \times 10^{-10}$

[478] The collaborations are ARCHEOPS: A. Benoit et al., Astron. Astrophys. 399, L19, L25 (2003) [astro-ph/0210305, 0210306]; CDMP & MAT/TOCO: A. Miller et al., Astrophys. J. Suppl. 140, 115 (2002) [astro-ph/0108030]; BOOMERANG: J. E. Ruhl et al., Astrophys. J. 599, 786 (2003) [astroph/0212229]; MAXIMA: A. T. Lee et al. Astrophys. J. 561, L1 (2001) ; DASI: N. W. Halverson et al., Astrophys. J. 568, 38 (2002) [astro-ph/0104489]; CBI: T. J. Pearson et al., Astrophys. J. 591, 556 (2003) [astro-ph/0205388]; ACBAR: C. L. Kuo et al., Astrophys. J. 600, 32 (2004) [astro-ph/0212289]; and VSA: C. Dickinson et al., Mon. Not. Roy. Astron. Soc. 353, 732 (2004) [astro-ph/0402498].

[9] C. L. Bennett et al., Astrophys. J. Suppl. 148, 1 (2003) [astro-ph/0302207]. Aspects of this mission are treated in detail in other articles in the same volume of Astrophys. J. Suppl.

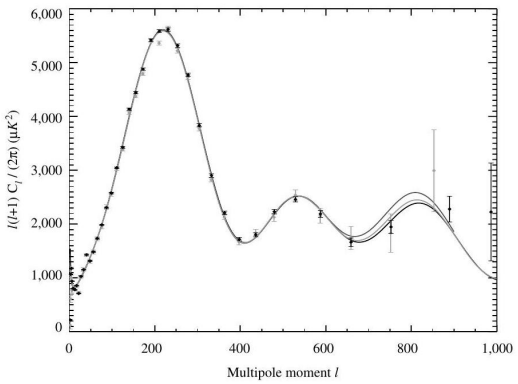


Figure 7.2: Comparison of theory and WMAP observations for the multipole coefficient $\ell(\ell+1) C_{\ell} / 2 \pi$ in square microKelvin vs. ℓ, from D. N. Spergel et al., astro-ph/0603449. The solid data points are the three-year WMAP data, and the light gray data points are the first year WMAP data. The bottom curve is the best fit to the three-year WMAP data. The top curve is the best fit to the first-year WMAP data, and the middle curve is the best fit to the first-year WMAP data combined with data from CBI and ACBAR.

- $n_S = 0.958 \pm 0.016$
- $\tau_{\rm reion} = 0.089 \pm 0.030$

The new results are consistent with those found before, but evidently there has been a significant improvement in the precision with which these parameters are known. This increased precision has now revealed the important information that n_s is very likely somewhat less than unity, as expected on the basis of typical inflationary theories, to be discussed in Chapter 10. It is also important that the optical depth of the reionized plasma found by polarization measurements is considerably less than previously found, suggesting a more plausible time of reionization. [15] Because the magnitude of $C^S_{TT,\ell}$ is proportional to $|N|^2 \exp(-2\tau_{\rm reion})$, the reduction in $\tau_{\rm reion}$ has led to a corresponding reduction in the estimated value of $|N|^2$; there has been very little change in the reported value of $|N|^2 \exp(-2\tau_{\rm reion})$.

It is truly satisfying to see that the values of $h, \Omega_M h^2$, and $\Omega_B h^2$ obtained from the cosmic microwave background anisotropies confirm the values obtained by the very different methods discussed in Chapters 1 and 3. The values of $|N|^2$ and n_s are also in good agreement with values inferred from observations of large scale structure, discussed in the next chapter. The general agreement of theory and observation, both for the microwave background anisotropies alone and for the microwave background anisotropies in conjunction with other observations, goes far to confirm the general assumptions of the cosmological model, including cold dark matter and dark energy, that we have been using.

[4710] G. Hinshaw et al., Astrophys. J. 148, 135 (2003) [astro-ph/0302217].

[11] D.N. Spergel et al., ref. 5. Errors given here represent a 68% confidence range.

[12] A. Kogut et al., Astrophys. J. 148, 161 (2003) [astro-ph/0302213].

[13] Temperature and polarization results are given by G. Hinshaw et al., Astrophys. J. Suppl. Ser. 170, 288 (2007) [astro-ph/0603451] and L. Page et al., Astrophys. J. Suppl. Ser. 170, 335 (2007) [astroph/0603450], respectively.

[14] D. N. Spergel et al., Astrophys. J. Suppl. Ser. 170, 377 (2007) [astro-ph/0603449].

[4715] For the implications of the three-year WMAP measurement of the plasma optical depth for theories of early star formation, see M. A. Alvarez, P. R. Shapiro, K. Ahn, and I. T. Iliev, Astrophys. J. 644, L101 (2006) [astro-ph/0604447]; Z. Haiman and G. L. Bryan, Astrophys. J. 650, 7 (2006) [astro-ph/0603541]; T. R. Choudhury and A. Ferrara, Mon. Not. Roy. Astron. Soc. 371, L55 (2006) [astro-ph/0603617]; M. Shull and A. Venkatesan, astro-ph/0702323.

10.7.3 7.3 Temperature multipole coefficients: Tensor modes

We next consider the contribution of cosmological gravitational radiation to the temperature multipole coefficients. [1] According to Eq. (7.1.5), in the case of a single dominant tensor mode the tensor contribution to the fractional temperature fluctuation is

$$\left(\frac{\Delta T(\hat{n})}{T_0}\right)^{(T)} = -\frac{1}{2}\sum_{\lambda=\pm 2}\int d^3q\beta(\mathbf{q},\lambda)\hat{n}_k\hat{n}_l e_{kl}(\hat{q},\lambda) \times \int_{t_1}^{t_0} dt \exp\left[i\mathbf{q}\cdot\hat{n}\int_t^{t_0}\frac{dt'}{a\,(t')}\right] d(q,t) \tag{7.3.1}$$

where for brevity we have now introduced the quantity

$$d(q,t) \equiv \exp\left[-\int_t^{t_0} dt'\omega_c\,(t')\right]\left(\dot{\mathcal{D}}_q(t) - \frac{1}{2}\omega_c(t)\Psi(q,t)\right) \tag{7.3.2}$$

As a reminder: $\beta(\mathbf{q},\lambda)$ is the stochastic parameter for the mode that does not decay outside the horizon, which is assumed to dominate the tensor perturbation; $e_{kl}(\hat{q},\lambda)$ is the polarization tensor defined in Section 5.2 for a gravitational wave with wave number $\boldsymbol{q}$ and helicity λ; t_1 is any time sufficiently early before recombination so that any photon present then would have been scattered many times before the present; $\omega_c(t)$ is the photon collision rate at time t; $\mathcal{D}_q(t)$ is the gravitational wave amplitude, defined in Section 5.2; and $\Psi(q,t)$ is the tensor source function, defined by Eqs. (6.6.11) and (6.6.12).

Again using Eq. (7.2.2) to expand the exponential in spherical harmonics, the tensor temperature fluctuation has the expansion

$$(\Delta T(\hat{n}))^{(T)} = \sum_{\ell m} a^T_{T,\ell m} Y^m_\ell(\hat{n}) \tag{7.3.3}$$

where

$$a^T_{T,\ell m} = T_0\sum_{\lambda=\pm 2}\int d^3q\beta(\mathbf{q},\lambda) f_{\ell m}(\mathbf{q},\lambda) \tag{7.3.4}$$

$$f_{\ell m}(\mathbf{q},\lambda) \equiv -2\pi\int d^2\hat{n} Y^{m*}_\ell(\hat{n})\hat{n}_k\hat{n}_l e_{kl}(\hat{q},\lambda)\sum_{LM} i^L Y^M_L(\hat{n}) Y^{M*}_L(\hat{q}) \times \int_{t_1}^{t_0} dt j_L(qr(t)) d(q,t) \tag{7.3.5}$$

and $r(t)$ is again the radial coordinate of a point from which light emitted at time t would reach us at the present:

$$r(t) \equiv \int_t^{t_0}\frac{dt'}{a\,(t')} \tag{7.3.6}$$

It will be very convenient first to calculate $f_{\ell m}(\mathbf{q},\lambda)$ for $\mathbf{q}$ in the threedirection $\hat{z}$. In this case we have $Y^M_L(\hat{q}) = \delta_{M0}\sqrt{2L+1/4\pi}$. Also, using Eq. (5.2.15), for $\hat{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ we have

[471] For an early calculation of the first few multipole moments, see V. A. Rubakov, M. V. Sazhin, and A. V. Veryaskin, Phys. Lett. 115B, 189 (1982). A general formula was given by R. Fabbri and M. D. Pollock, Phys. Lett. 125B, 445 (1983).

$$\hat{n}_i\hat{n}_j e_{ij}(\hat{z},\pm 2) = \frac{1}{\sqrt{2}}\sin^2\theta e^{\pm 2i\phi} = 4\sqrt{\frac{\pi}{15}}Y_2^{\pm 2}(\hat{n}) \tag{7.3.7}$$

The integral over $\hat{n}$ is given by a special case of the general formula [2]

$$\int d^2\hat{n}Y_L^M(\hat{n})Y_\Lambda^\mu(\hat{n})Y_\ell^{m*}(\hat{n}) = \sqrt{\frac{(2\Lambda+1)(2\ell+1)}{4\pi(2L+1)}}$$
$$\times C_{\ell\Lambda}(L,M;m,-\mu)C_{\ell\Lambda}(L,0;0,0) \tag{7.3.8}$$

where $C_{\ell\Lambda}(L,M;m,\mu)$ is the usual Clebsch-Gordan coefficient for combining angular momentum quantum numbers ℓ, m and Λ,μ to form

angular momentum quantum numbers L, M. Hence

$$f_{\ell m}(q\hat{z},\pm 2) = -2\sqrt{\frac{\pi(2\ell+1)}{3}}\sum_L i^L C_{\ell 2}(L,0;\pm 2,\mp 2)C_{\ell 2}(L,0;0,0)$$
$$\times \delta_{m,\pm 2}\int_{t_1}^{t_0} dt j_L(qr(t))d(q,t) \tag{7.3.9}$$

In our case we have $\Lambda = 2, \mu = \pm 2$, and $M = 0$, so the relevant ClebschGordan coefficients are

$$C_{\ell 2}(\ell+2,0;0,0) = \sqrt{\frac{3(\ell+2)(\ell+1)}{2(2\ell+1)(2\ell+3)}}$$
$$C_{\ell 2}(\ell+2,0;\pm 2,\mp 2) = \frac{1}{2}\sqrt{\frac{(\ell-1)\ell}{(2\ell+1)(2\ell+3)}}$$
$$C_{\ell 2}(\ell,0;0,0) = -\sqrt{\frac{\ell(\ell+1)}{(2\ell-1)(2\ell+3)}}$$
$$C_{\ell 2}(\ell,0;\pm 2,\mp 2) = \sqrt{\frac{3(\ell-1)(\ell+2)}{2(2\ell-1)(2\ell+3)}}$$
$$C_{\ell 2}(\ell-2,0;0,0) = \sqrt{\frac{3\ell(\ell-1)}{2(2\ell-1)(2\ell+1)}}$$
$$C_{\ell 2}(\ell-2,0;\pm 2,\mp 2) = \frac{1}{2}\sqrt{\frac{(\ell+1)(\ell+2)}{(2\ell-1)(2\ell+1)}}$$

while $C_{\ell 2}(\ell\pm 1,0;0,0) = 0$. Putting this together shows that for $\mathbf{q}$ in the three-direction, the non-vanishing values of the quantity (7.3.5) are [3]

$$f_{\ell m}(q\hat{z},\pm 2) = i^\ell\sqrt{\frac{\pi(2\ell+1)(\ell+2)!}{2(\ell-2)!}}\delta_{m,\pm 2}\int dt d(q,t)$$
$$\times\left[\frac{j_{\ell+2}(qr(t))}{(2\ell+1)(2\ell+3)} + \frac{2j_\ell(qr(t))}{(2\ell-1)(2\ell+3)} + \frac{j_{\ell-2}(qr(t))}{(2\ell-1)(2\ell+1)}\right] \tag{7.3.10}$$

[472] See, e.g., J. M. Blatt and V. F. Weisskopf, Theoretical Nuclear Physics (John Wiley & Sons, New York, 1952): Appendix A, Eq. (5.11).

This can be greatly simplified by iterating the familiar recursion relation

$$j_\ell(\rho)/\rho = \left[j_{\ell-1}(\rho) + j_{\ell+1}(\rho)\right]/(2\ell+1),$$

which gives

$$\frac{j_\ell(\rho)}{\rho^2} = \frac{j_{\ell-2}(\rho)}{(2\ell+1)(2\ell-1)} + \frac{2j_\ell(\rho)}{(2\ell+3)(2\ell-1)} + \frac{j_{\ell+2}(\rho)}{(2\ell+1)(2\ell+3)} \tag{7.3.11}$$

Using this in Eq. (7.3.10), we have

$$f_{\ell m}(q\hat{z},\pm 2) = i^\ell \sqrt{\frac{\pi(2\ell+1)(\ell+2)!}{2(\ell-2)!}}\delta_{m,\pm 2}\int_{t_1}^{t_0} dtd(q,t)\frac{j_\ell(qr(t))}{q^2r^2(t)} \tag{7.3.12}$$

The amplitude (7.3.5) can now be found for a general direction of $\mathbf{q}$ by performing a standard rotation $S(\hat{q})$ that takes the three-axis into the direction $\hat{q}$. (An explicit formula for $S(\hat{q})$ will be given in the next section; it is not needed here.) This gives

$$f_{\ell m}(\mathbf{q},\lambda) = \sum_{\pm} D^{(\ell)}_{m,m'}(S(\hat{q}))f_{\ell m'}(q\hat{z},\lambda) \tag{7.3.13}$$

The coefficients (7.3.4) are then

$$\begin{aligned} a^T_{T,\ell m} =& T_0 i^\ell \sqrt{\frac{\pi(2\ell+1)(\ell+2)!}{2(\ell-2)!}} \sum_{\pm}\int d^3q\beta(\mathbf{q},\pm 2)D^{(\ell)}_{m,\pm 2}(S(\hat{q})) \\ &\times \int_{t_1}^{t_0} dtd(q,t)\frac{j_\ell(qr(t))}{q^2r^2(t)} \end{aligned} \tag{7.3.14}$$

We can now easily calculate the multipole coefficients of the temperature correlation function, defined by

$$\left\langle a^T_{T,\ell m} a^{T*}_{T,\ell' m'}\right\rangle = \delta_{\ell\ell'}\delta_{mm'}C^T_{TT,\ell} \tag{7.3.15}$$

For this purpose, we recall that the stochastic parameter $\beta(\mathbf{q},\lambda)$ is normalized so that

$$\langle\beta(\mathbf{q},\lambda)\beta^*(\mathbf{q},\lambda)\rangle = \delta_{\lambda\lambda'}\delta^3\left(\mathbf{q}-\mathbf{q}'\right) \tag{7.3.16}$$

In the calculation of $C^T_{TT,\ell}$ we encounter an angular integral

$$\int d^2\hat{q}D^{(\ell)}_{m,\pm 2}(S(\hat{q}))D^{(\ell')}_{m',\pm 2}(S(\hat{q}))^*$$

This matrix can readily be seen to commute with all rotation matrices, so that it must be proportional to $\delta_{mm'}\delta_{\ell\ell'}$ with an m-independent coefficient. To calculate this coefficient, we can set $m = m'$ and $\ell = \ell'$ and sum m from $-\ell$ to $+\ell$; using the unitarity of the matrix $D^{(\ell)}_{m,n}(S(\hat{q}))$, this must equal $\int d^2\hat{q} = 4\pi$, so

$$\int d^2\hat{q}D^{(\ell)}_{m,\pm 2}(S(\hat{q}))D^{(\ell')}_{m',\pm 2}(S(\hat{q}))^* = \frac{4\pi}{2\ell+1}\delta_{mm'}\delta_{\ell\ell'} \tag{7.3.17}$$

Using this and Eqs. (7.3.14) and (7.3.16) in Eq. (7.3.15), we see that the sum over helicities in Eq. (7.3.14) now just yields a factor 2, so [4]

[473] This is essentially the result originally obtained by L.F. Abbott and M. B. Wise, Nucl. Phys. B244, 541 (1984), and A. A. Starobinsky, Sov. Astron. Lett. 11(3), 133 (1985), generalized to an arbitrary gravitational wave amplitude $D_q(t)$ and including the correction proportional to Ψ in Eq. (7.3.2).

$$C_{TT,\ell}^{T} = \frac{4\pi^2(\ell+2)!T_0^2}{(\ell-2)!}\int_0^\infty \frac{dq}{q^2}\left|\int_{t_1}^{t_0} dt d(q,t)\frac{j_\ell(qr(t))}{r^2(t)}\right|^2 \tag{7.3.18}$$

(This formula can be obtained more easily by a direct calculation of the temperature correlation function, without calculating the $a_{T,\ell m}^T$, but we will need the $a_{T,\ell m}^T$ in the next section to find the correlation between polarization and temperature anisotropies.)

It remains to calculate $\mathcal{D}_q(t)$ and the source function $\Psi(q,t)$, As we saw in Section 6.6, if the gravitational wave amplitude $\mathcal{D}_q(t)$ is written as a function of $\kappa \equiv q\sqrt{2}/H_{\rm EQ}a_{\rm EQ}$ and $y \equiv a(t)/a_{\rm EQ}$ instead of q and t (the subscript EQ referring to the time of radiation-matter equality), then aside from an over-all factor $\mathcal{D}_q^o$ (equal to the value of $\mathcal{D}_q(t)$ outside the horizon), the amplitude $\mathcal{D}_q(t)$ is independent of any other cosmological parameters. We write this as

$$\mathcal{D}_q(t) = \mathcal{D}_q^o D(\kappa, y) \tag{7.3.19}$$

For a spectrum of gravitational waves that is scale invariant outside the horizon, the quantity $q^3\left|\mathcal{D}_q^o\right|^2$ is a constant. To take account of more general possibilities, it is conventional to write it as proportional to a power n_T of q, or equivalently,

$$q^3\left|\mathcal{D}_q^o\right|^2 = N_T^2\left(\frac{q/a_0}{k_{\mathcal{D}}}\right)^{n_T}, \tag{7.3.20}$$

where $k_{\mathcal{D}}$ is an arbitrary wave number, often taken as $k_{\mathcal{D}} = 0.002{\rm Mpc}^{-1}$, and N_T is a constant analogous to the constant N that describes the strength

of the adiabatic scalar mode. The conventional ratio r of tensor and scalar modes is $r \equiv 4\left|N_T\right|^2/|N|^2$. Also, until vacuum energy becomes important, we have $dt = \sqrt{2}ydy/H_{\rm EQ}\sqrt{1+y}$, so we can write

$$q\int\frac{dt}{a(t)} = \kappa\int\frac{dy}{\sqrt{1+y}} = 2\kappa\sqrt{1+y}$$

The wave equation (6.6.58) then gives

$$\begin{aligned}(1+y)\frac{\partial^2 D(\kappa,y)}{\partial y^2} + \left(\frac{1}{2}+\frac{2(1+y)}{y}\right)\frac{\partial D(\kappa,y)}{\partial y} + \kappa^2 D(\kappa,y)\\ = -\frac{24f_v}{y^2}\int_0^y K\left(2\kappa\left[\sqrt{1+y}-\sqrt{1+y'}\right]\right)\frac{\partial D\left(\kappa,y'\right)}{\partial y'}dy'\end{aligned} \tag{7.3.21}$$

where $f_v \equiv \bar{\rho}_v/\bar{\rho}_R = 0.4052$ and $K(v) \equiv j_2(v)/v^2$, with the initial condition $D(\kappa,0) = 1$. Having found $D(\kappa,y)$ in this way, the source function $\Psi(q,t)$ is calculated in computer programs such as CMBfast and CAMB by using Eq. (6.6.21), with the amplitudes $\Delta_{T,\ell}^{(T)}$ and $\Delta_{P,\ell}^{(T)}$ found by a numerical solution of the coupled Boltzmann equations truncated at some maximum value of ℓ. There is an easier alternative procedure, based on the integral equation (6.6.27): [5]

$$\begin{aligned}\Psi(q,t) = \frac{3}{2}\int_{t_1}^{t} dt' \exp\left[-\int_{t'}^{t}\omega_c\left(t''\right)dt''\right]\\ \times\left[-2\dot{\mathcal{D}}_q\left(t'\right)K\left(q\int_{t'}^{t}\frac{dt''}{a\left(t''\right)}\right)+\omega_c\left(t'\right)F\left(q\int_{t'}^{t}\frac{dt''}{a\left(t''\right)}\right)\Psi\left(q,t'\right)\right]\end{aligned} \tag{7.3.22}$$

[474] This result is derived in a different way by M. Zaldarriaga and U. Seljak, Phys. Rev. D 55, 1830 (1997) [astro-ph/9609170]. Their result is the same as Eq. (7.3.18), if we take their undefined gravitational wave amplitude h to be $\mathcal{D}_q(t)/2$, with their stochastic parameter normalized so that $P_h = 1$.

where $F(v) \equiv j_0(v) - 2j_1(v)/v + 2j_2(v)/v^2$. We can also put the differential optical depth in the form

$$d\tau(y) \equiv \omega_c dt = n_e \sigma_{\mathcal{T}} c dt = \frac{A}{y^2\sqrt{1+y}} X(y) dy$$

where $X(y)$ is the fractional hydrogen ionization calculated in Section 2.3, which depends on $\Omega_B h^2$ and $\Omega_M h^2$ as well as y, and A is the dimensionless constant

$$A \equiv 0.76 \frac{3\Omega_B \Omega_M H_0 \sigma_{\mathcal{T}} c}{8\pi G m_p \Omega_R^{3/2}} = 1.9646 \times 10^5 \left(\Omega_M h^2\right) \left(\Omega_B h^2\right)$$

(It is assumed that the baryons are 76% hydrogen, with the rest un-ionized in the era of interest.) It follows that the source function takes the form

$$\Psi(q,t) = \mathcal{D}_q^o \mathcal{S}(\kappa, y) \tag{7.3.23}$$

where $\mathcal{S}$ depends only on $\Omega_M h^2$ and $\Omega_B h^2$ as well as κ and y, and satisfies the integral equation

$$\begin{aligned}\mathcal{S}(\kappa,y) = \frac{3}{2}\int_0^y dy' \exp\left[-\int_{y'}^y \frac{d\tau\,(y'')}{dy''} dy''\right] \\ \times \left[-2\frac{dD\,(\kappa,y')}{dy'} K\left(2\kappa\left[\sqrt{1+y}-\sqrt{1+y'}\right]\right)\right. \\ \left. + \frac{d\tau\,(y')}{dy'} F\left(2\kappa\left[\sqrt{1+y}-\sqrt{1+y'}\right]\right)\mathcal{S}\,(\kappa,y')\right].\end{aligned} \tag{7.3.24}$$

The multipole coefficient calculated from Eqs. (7.3.18)-(7.3.24) is shown by the solid curve in Figure 7.3. The dashed line in Figure 7.3 shows the result of simplifying the expression for the multipole coefficients in the case $\ell \gg 1$, using the approximate formula (7.2.14) for the spherical Bessel

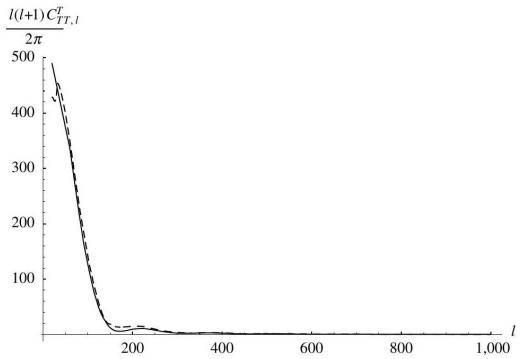


Figure 7.3: The tensor temperature-temperature multipole coefficient $\ell(\ell+1) C_{T T, \ell}^{T} / 2 \pi$ in square microKelvin, vs. ℓ, from ref. 6 . The solid line is the result of using the essentially exact formula (7.3.18); the dashed line gives the result of the large- ℓ approximation (7.3.25). The cosmological parameters used in these calculations are described in the text.

function in Eq. (7.3.18), which gives [6]

[475] S. Weinberg, Phys. Rev. D 74, 063517 (2006) [astro-ph/0607076]; D. Baskaran, L. P. Grishchuk, and A. G. Polnarev, Phys. Rev. D 74, 083005 (2006) [gr-qc/0605100].

$$\ell(\ell+1)C^T_{TT,\ell} \to 4\pi^2 T_0^2 \int_0^\infty q^2 dq \times \left| \int_{r(t)>\ell/q} dt d(q,t) \left\{ \frac{\cos^3 b}{\sqrt{\sin b}} \cos[\ell(\tan b - b) - \pi/4] \right\}_{\cos b = \ell/qr(t)} \right|^2 \tag{7.3.25}$$

The source function $\Psi(q,t)$ is calculated in both cases using the integral equation (7.3.22), instead of using the equivalent formula (6.6.21) together with the truncated Boltzmann hierarchy (6.6.18), (6.6.19). Both of these curves are calculated with a cosmological model that in most respects is consistent with current observations: We assume $\Omega_K = 0$, and constant vacuum energy, with $\Omega_B = 0.0432$, $\Omega_M = 0.257$, $\Omega_\Lambda = 0.743, h = 0.72$, and $T_0 = 2.725$ K. In calculating the photon collision frequency, we use the Recfast recombination code, [7] with helium abundance $Y = 0.24$. The parameters in Eq. (7.3.20) are taken as $N_T^2 = 4.68 \times 10^{-11}$ (corresponding approximately to the WMAP3 value of the scalar amplitude N quoted in the previous section, and $r = 1$) and $n_T = 0$. Reionization is ignored in this calculation. To take into account a different value of r or a finite optical depth τ of the reionized plasma, for $\ell > 10$ it is only necessary to multiply the multipole coefficients shown in Figure 7.3 by $r\exp(-2\tau)$. (In the approximation of a sharp transition at time t_L from thermal equilibrium, in which $\omega_c(t) \gg H(t)$ and $\Psi(q,t) = 0$, to perfect transparency, in which $\omega_c(t) = 0$, Eq. (7.3.25) becomes

$$\ell(\ell+1)C^T_{TT,\ell} \to 4\pi^2 T_0^2 \int_0^\infty q^2 dq \times \left| \int_{r_L>r(t)>\ell/q} dt \dot{\mathcal{D}}_q(t) \left\{ \frac{\cos^3 b}{\sqrt{\sin b}} \cos[\ell(\tan b - b) - \pi/4] \right\}_{\cos b = \ell/qr(t)} \right|^2 \tag{7.3.26}$$

Because of the spread of values of $\tan b - b$, we cannot here make the sort of further simplification that we made for scalar temperature fluctuations in going from Eq. (7.2.16) to Eq. (7.2.17).)

Figure 7.3 shows that $C^T_{TT,\ell}$ has an ℓ-dependence very different from that of the scalar contribution calculated in the previous section, so the agreement between existing observations and theory including only scalar perturbations sets an upper bound on the strength of the tensor perturbations. Under the assumption (as suggested by theories of inflation discussed in Chapter 10) that $\mathcal{D}^o_q$ and $\mathcal{R}^0_q$ have about the same q-dependence, the WMAP collaboration has concluded [8] after three-years of operation that $r < 0.55$, at a 95% confidence level, where $r \equiv 4\left|\mathcal{D}^0_q\right|^2 / \left|\mathcal{R}^0_q\right|^2$. This limit arises chiefly from temperature rather than polarization measurements, because of the much greater signal to noise ratio of the temperature measurements; the limit on r from polarization measurements alone is $r < 2.2$.

Strictly speaking, the limit $r < 0.55$ does rely on polarization measurements, which are used to determine the optical thickness of the reionized plasma. This is needed to determine the value of the slope parameter n_S for scalar modes, which in turn is needed in subtracting the scalar contribution to $C_{TT,\ell}$ from the observed values in order to set a limit on the tensor contribution. But cosmic variance sets a limit to the accuracy with which this subtraction can be made, and in the long run the best upper limits or the actual detection of tensor modes will come directly from polarization measurements, to which we now turn.

[476] R. Flauger and S. Weinberg, Phys. Rev. D 75, 123505 (2007) [astro-ph/0703179].

[7] S. Seager, D. D. Sasselov, and D. Scott, Astrophys. J. 523, L1 (1999) [astro-ph/9909275]; Astrophys. J. Suppl. 128, 407 (2000) [astro-ph/9912182].

10.7.4 7.4 Polarization

Observations of the cosmic microwave background reveal not only its intensity in various directions, but also its polarization. The microwave background is expected to be polarized because of its scattering by free electrons, [1] such as were present around the time of recombination, or during the later period of reionization due to ultraviolet light from the first generation of stars. Polarization measurements have become of importance in learning when reionization began, and in disentangling the effects of reionization from the primordial intensity of fluctuations, and they may become even more important in future, in revealing the effects of gravitational waves

produced during inflation.

In Section 6.1 we introduced a dimensionless photon intensity perturbation matrix $J_{ij}(\mathbf{x}, \hat{p}, t)$, defined by Eq. (6.1.13). It is Hermitian, transverse in the sense that

$$\hat{p}^i J_{ij}(\mathbf{x}, \hat{p}, t) = \hat{p}^j J_{ij}(\mathbf{x}, \hat{p}, t) = 0, \tag{7.4.1}$$

and, as already noted in Section 7.1, its trace at $\mathbf{x} = 0$ and $t = t_0$ is related to the fractional photon temperature fluctuation seen from earth at present in any direction $\hat{n}$ by

$$\frac{\Delta T(\hat{n})}{T_0} = \frac{1}{4} J_{ii}\left(0, -\hat{n}, t_0\right). \tag{7.4.2}$$

For $\hat{n} = -\hat{p}$ in the three-direction $\hat{z}$, any matrix with these properties can be put in the form

$$J_{ij}\left(0, -\hat{z}, t_0\right) = \frac{2}{T_0} \begin{pmatrix} \Delta T(\hat{z}) + Q(\hat{z}) & U(\hat{z}) - iV(\hat{z}) & 0 \\ U(\hat{z}) + iV(\hat{z}) & \Delta T(\hat{z}) - Q(\hat{z}) & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{7.4.3}$$

where Q, U, and V are three real functions of direction with the dimensions of temperature, known as Stokes parameters. [2] In this case the Stokes parameters can be expressed in terms of the matrix J_{ij}, as

$$Q(\hat{z}) \pm iU(\hat{z}) = \frac{T_0}{2} e_{\pm i}(\hat{z}) e_{\pm j}(\hat{z}) J_{ij}\left(0, -\hat{z}, t_0\right) \tag{7.4.4}$$

$$V(\hat{z}) = \frac{T_0}{4} e_{-i}(\hat{z}) e_{+j}(\hat{z}) \left(J_{ij}\left(0, -\hat{z}, t_0\right) - J_{ji}\left(0, -\hat{z}, t_0\right)\right) \tag{7.4.5}$$

where $e_{\pm}(\hat{z}) = (1, \pm i, 0)/\sqrt{2}$ are the polarization vectors for a photon coming from the three-direction. Accordingly, for a photon coming from an arbitrary direction $\hat{n}$, we define the Stokes parameters as

$$Q(\hat{n}) \pm iU(\hat{n}) \equiv \frac{T_0}{2} e_{\pm i}(\hat{n}) e_{\pm j}(\hat{n}) J_{ij}\left(0, -\hat{n}, t_0\right) \tag{7.4.6}$$

$$V(\hat{n}) \equiv \frac{T_0}{4} e_{-i}(\hat{n}) e_{+j}(\hat{n}) \left(J_{ij}\left(0, -\hat{n}, t_0\right) - J_{ji}\left(0, -\hat{n}, t_0\right)\right) \tag{7.4.7}$$

[478] L. Page et al., Astrophys. J. Suppl. Ser. 170, 335 (2007) [astro-ph/0603450].

[1] M. J. Rees, Astrophys. J. 153, L1 (1968). Polarization correlations at small angular separation were considered in early papers on the polarization of the cosmic microwave background; see e.g. A. G. Polnarev, Sov. Astron. 29, 607 (1985); R. Crittenden, R. L. Davis, and P. J. Steinhardt, Astrophys. J. Lett. 417, L13 (1993); D. Coulson, R. Crittenden, and N. Turok, Phys. Rev. Lett. 73, 2390 (1994); A. Kosowsky, Ann. Phys. (N.Y.) 246, 49 (1996). Then all-sky analyses were given by U. Seljak and M. Zaldarriaga, Phys. Rev. Lett. 78, 2054 (1997) [astro-ph/9609169]; M. Zaldarriaga and U. Seljak, Phys. Rev. D 55, 1830 (1997) [astroph/9609170], and by M. Kamionkowski, A. Kosowsky, and A. Stebbins, Phys. Rev. Lett. 78, 2058 (1997) [astro-ph/9609132]; Phys. Rev. D. 55, 7368 (1997) [astro-ph/9611125].

where $e_\pm(\hat{n})$ are the polarization vectors for a photon coming from the $\hat{n}$-direction. Writing $\hat{n}$ in terms of polar and azimuthal angles θ and ϕ as

$$\hat{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \tag{7.4.8}$$

we can take

$$\mathbf{e}_\pm(\hat{n}) = (\hat{\theta} \pm i\hat{\phi})/\sqrt{2} \tag{7.4.9}$$

where $\hat{\theta}$ and $\hat{\phi}$ are orthogonal unit vectors in the plane perpendicular to $\hat{n}$, in the directions of increasing θ and ϕ, respectively:

$$\hat{\theta} \equiv (\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta), \quad \hat{\phi} \equiv (-\sin\phi, \cos\phi, 0) \tag{7.4.10}$$

It is the Stokes parameters that are measured in observations of the microwave background. The scattering of light by non-relativistic electrons does not produce circular polarization, and therefore we expect that all microwave background photons are linearly polarized, so that J_{ij} is real, and therefore $V = 0$. In this case the fractional intensity perturbation $e^i e^j J_{ij}$ reaches a maximum value $2\left(\Delta T + \sqrt{U^2+Q^2}\right)/T_0$ for a real polarization vector $\mathbf{e}$ in a direction $\hat{\theta}\cos\xi + \hat{\phi}\sin\xi$, where $\tan 2\xi = U/Q$. For a polarization vector in the orthogonal direction, $e^i e^j J_{ij}$ takes its minimum value $2\left(\Delta T - \sqrt{U^2+Q^2}\right)/T_0$, giving a total fractional intensity perturbation $4\Delta T/T_0$, as is necessary since the photon energy density goes as T^4.

Now we have to face a complication: As we have defined them, the Stokes parameters are not rotational scalars. That is, under an arbitrary rotation $x_i \to x'_i = R_{ij}x_j$ of the three-dimensional coordinate system, we do not have $Q'(\hat{n}') = Q(\hat{n})$ or $U'(\hat{n}') = U(\hat{n})$. This is because the polarization vectors are not really three-vectors. We note that

$$\mathbf{e}_{\pm i}(\hat{n}) = S_{ij}(\hat{n})\mathbf{e}_{\pm j}(\hat{z}) \tag{7.4.11}$$

where $S_{ij}(\hat{n})$ is a standard rotation that takes the three-axis into the direction $\hat{n}$. For $\hat{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$:

$$S_{ij}(\hat{n}) \equiv \begin{pmatrix} \cos\theta\cos\phi & -\sin\phi & \sin\theta\cos\phi \\ \cos\theta\sin\phi & \cos\phi & \sin\theta\sin\phi \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \tag{7.4.12}$$

and again $\mathbf{e}_\pm(\hat{z}) = (1, \pm i, 0)/\sqrt{2}$. For an arbitrary rotation R, we can write

$$\mathbf{e}_{\pm i}(R\hat{n}) \equiv S_{ij}(R\hat{n})\mathbf{e}_{\pm j}(\hat{z}) = [RS(\hat{n})]_{ik}\left[S^{-1}(\hat{n})R^{-1}S(R\hat{n})\right]_{kj}\mathbf{e}_{\pm j}(\hat{z})$$

Now, $S(R\hat{n})$ takes the three-axis into the direction $R\hat{n}$, R^{-1} takes this into the direction $\hat{n}$, and then $S^{-1}(\hat{n})$ takes this back to the direction of the threeaxis, so $S^{-1}(\hat{n})R^{-1}S(R\hat{n})$ leaves the three-direction invariant, and must
therefore be a rotation by some angle $\psi(R,\hat{n})$ around the three-direction. Acting on $\mathbf{e}_\pm(\hat{z})$, this yields a factor $e^{\pm i\psi}$, and so

$$\mathbf{e}_{\pm i}(R\hat{n}) = e^{\pm i\psi(R,\hat{n})}R_{ij}\mathbf{e}_{\pm j}(\hat{n}). \tag{7.4.13}$$

The matrix J_{ij} is an ordinary three-tensor, in the sense that an arbitrary rotation R takes J_{ij} into J'_{ij}, with

[472] The Stokes parameters are sometimes defined with extra constant factors to give them the dimensions of intensity rather than temperature.

$$J'_{ij}(R\hat{n}) = R_{ik}R_{il}J_{kl}(\hat{n})$$

Because $R^TR = 1$, a rotation R subjects the Stokes parameters (7.4.6) to the transformation $Q \to Q', U \to U'$, with

$$Q'(\hat{n}') \pm iU'(\hat{n}') = e^{\pm 2i\psi(R,\hat{n})}[Q(\hat{n}) \pm iU(\hat{n})]. \tag{7.4.14}$$

For this reason, if we were to expand the Stokes parameters in a series of ordinary spherical harmonics, as we do for scalars like the temperature fluctuation, then the expansion coefficients would not transform under rotations according to the usual representations of the rotation group. Instead, we expand the Stokes parameters $Q(\hat{n})$ and $U(\hat{n})$ seen in a direction $\hat{n}$ in a series of functions ${}^3\mathcal{Y}^m_\ell(\hat{n})$ with the same dependence on the polarization vectors as the Stokes parameters themselves:

$$\begin{aligned} Q(\hat{n}) + iU(\hat{n}) &= \textstyle\sum_{\ell=2}^{\infty}\sum_{m=-\ell}^{\ell} a_{P,\ell m}\mathcal{Y}^m_\ell(\hat{n}), \\ \mathcal{Y}^m_\ell(\hat{n}) &\equiv 2\sqrt{\tfrac{(\ell-2)!}{(\ell+2)!}}e_{+i}(\hat{n})e_{+j}(\hat{n})\tilde{\nabla}_i\tilde{\nabla}_j Y^m_\ell(\hat{n}), \end{aligned} \tag{7.4.16}$$

where $\tilde{\nabla}$ is the angular part of the gradient operator:

$$\tilde{\nabla} \equiv \hat{\theta}\frac{\partial}{\partial\theta} + \frac{\hat{\phi}}{\sin\theta}\frac{\partial}{\partial\phi} \tag{7.4.17}$$

The subscript " P " (for "polarization") on the coefficients $a_{P,\ell m}$ is introduced here to distinguish them from the coefficients entering in the expansion (2.6.1) of the temperature fluctuation, which in this chapter are denoted $a_{T,\ell m}$. Under a rotation R, the ordinary spherical harmonics transform as

$$Y^m_\ell(R\hat{n}) = \sum_{m'} D^{(\ell)}_{m'm}\left(R^{-1}\right) Y^{m'}_\ell(\hat{n}), \tag{7.4.18}$$

where $D^{(\ell)}(R)$ are the unitary irreducible matrices [4] of rank $2\ell+1$ that form a representation of the rotation group, in the sense that $D^{(\ell)}(R_1)\,D^{(\ell)}(R_2) = D^{(\ell)}(R_1R_2)$. It follows then from Eqs. (7.4.13), (7.4.16) and (7.4.18) that

$$\mathcal{Y}^m_\ell(R\hat{n}) = e^{2i\psi(R,\hat{n})}\sum_{m'} D^{(\ell)}_{m'm}\left(R^{-1}\right)\mathcal{Y}^{m'}_\ell(\hat{n}) \tag{7.4.19}$$

and hence also

$$\mathcal{Y}^{m*}_\ell(R\hat{n}) = e^{-2i\psi(R,\hat{n})}\sum_{m'} D^{(\ell)}_{mm'}(R)\mathcal{Y}^{m'}_\ell(\hat{n})^* \tag{7.4.20}$$

For this reason, $\mathcal{Y}^m_\ell$ and its complex conjugate are known as spherical harmonics of spin 2 and -2 , respectively. Just as for ordinary spherical harmonics, it follows from these transformation properties that $\int d\Omega\mathcal{Y}^{m*}_\ell\mathcal{Y}^{m'}_{\ell'}$ vanishes except for $\ell = \ell'$ and $m = m'$, and in that case is independent of m. The factor $2\sqrt{(\ell-2)!/(\ell+2)!}$ is inserted in Eq. (7.4.16) to make these functions satisfy the orthonormality condition

$$\int d\Omega\mathcal{Y}^{m*}_\ell\mathcal{Y}^{m'}_{\ell'} = \delta_{\ell\ell'}\delta_{mm'} \tag{7.4.21}$$

[473] These are a special case of functions introduced in a study of gravitational radiation by E. T. Newman and R. Penrose, J. Math. Phys. 7, 863 (1966). In their notation, which is used in some recent papers on microwave background polarization, the function $\mathcal{Y}^m_\ell$ is denoted ${}_2Y_{\ell,m}$.

just like the ordinary spherical harmonics Y_ℓ^m. Direct evaluation of Eq. (7.4.16) gives [5]

$$\begin{aligned}
\mathcal{Y}_\ell^m(\hat{n}) =& \sqrt{\frac{(\ell-2)!}{(\ell+2)!}}\left[\left(\frac{\partial}{\partial\theta}+i\csc\theta\frac{\partial}{\partial\phi}\right)^2-\cot\theta\left(\frac{\partial}{\partial\theta}+i\csc\theta\frac{\partial}{\partial\phi}\right)\right]Y_\ell^m(\hat{n})\\
=& e^{im\phi}\sqrt{\frac{(\ell-2)!}{(\ell+2)!}\frac{(\ell-|m|)!}{(\ell+|m|)!}\frac{2\ell+1}{4\pi}}\Big[-\ell(\ell+1)P_\ell^{|m|}(\mu)\\
&+\frac{2(m+\mu)(m-\ell\mu)}{1-\mu^2}P_\ell^{|m|}(\mu)\\
&+\frac{2(\ell+|m|)(m+\mu)}{1-\mu^2}P_{\ell-1}^{|m|}(\mu)\Big)\Big]\\
=& e^{im\phi}\left[\frac{(\ell+m)!(\ell-m)!}{(\ell+2)!(\ell-2)!}\frac{2\ell+1}{4\pi}\right]^{1/2}\sin^{2\ell}(\theta/2)(-1)^{(m-|m|)/2}\\
&\times\sum_r\binom{\ell-2}{r}\binom{\ell+2}{r+2-m}\\
&\times(-1)^{\ell-r+m}\cot^{2r+2-m}(\theta/2)
\end{aligned}\tag{7.4.22}$$

where $\mu\equiv\cos\theta$ and $P_\ell^m(\mu)$ are the usual associated Legendre functions. (Here we will only need the first expression in Eq. (7.4.22).) Unlike the ordinary spherical harmonics, the $\mathcal{Y}_\ell^m$ do not satisfy any simple reality condition. Rather, the complex conjugate of Eq. (7.4.15) gives

$$Q(\hat{n})-iU(\hat{n})=\sum_{\ell=2}^{\infty}\sum_{m=-\ell}^{\ell}a_{P,\ell m}^*\mathcal{Y}_\ell^m(\hat{n})^*\tag{7.4.23}$$

where

$$\mathcal{Y}_\ell^m(\hat{n})^*=2\sqrt{\frac{(\ell-2)!}{(\ell+2)!}}e_{-i}(\hat{n})e_{-j}(\hat{n})\tilde{\nabla}_i\tilde{\nabla}_jY_\ell^{-m}(\hat{n}),\tag{7.4.24}$$

in which we have used the reality property of the spherical harmonics employed here, $Y_\ell^{m*}=Y_\ell^{-m}$.

Because $\mathcal{Y}_\ell^m$ does not satisfy any simple reality condition, neither does $a_{P,\ell m}$. Instead, we can define coefficients

$$a_{E,\ell m}\equiv-\left(a_{P,\ell m}+a_{P,\ell-m}^*\right)/2,\quad a_{B,\ell m}\equiv i\left(a_{P,\ell m}-a_{P,\ell-m}^*\right)/2.\tag{7.4.25}$$

This is a useful decomposition, because of the properties of the coefficients under space inversion. If we reverse all three coordinate axes, then $\theta\to\pi-\theta$, while $\phi\to\phi\pm\pi$, so

[474] See, e.g., A. R. Edmonds, Angular Momentum in Quantum Mechanics, (Princeton University Press, Princeton, 1957): Chapter 4; M. E. Rose, Elementary Theory of Angular Momentum (John Wiley & Sons, New York, 1957): Chapter IV; L. D. Landau and E. M. Lifshitz, Quantum Mechanics - Non Relativistic Theory, 3rd edn. (Pergamon Press, Oxford, 1977): Section 58; Wu-Ki Tung, Group Theory in Physics (World Scientific, Singapore, 1985): Sections 7.3 and 8.1. Note that the rotation matrices we use are appropriate for the representation furnished by the spherical harmonics we use, with phases such that $Y_\ell^m(\hat{n})^*=Y_\ell^{-m}(\hat{n})$. The corresponding angular momentum matrices $\mathbf{J}^{(\ell)}$ are not the usual ones, for which the elements of $J_1^{(\ell)}\pm iJ_2^{(\ell)}$ are real and positive. If we changed the phase of the spherical harmonics so that $Y_\ell^m(\hat{n})^*=(-1)^mY_\ell^{-m}(\hat{n})$, then the rotation matrices would be generated by conventional angular momentum matrices, with the elements of $J_1^{(\ell)}\pm iJ_2^{(\ell)}$ real and positive, but then we would have to introduce phases into the definition (7.4.25) of the E and B-type partial wave amplitudes, different from those used by Zaldarriaga and Seljak in ref. 1.

[5] The final expression is taken from Zaldarriaga and Seljak, ref. 1, while the first expression, which is derived here from our definition (7.4.16), is their definition of the spin +2 weighted spherical harmonic.

$\hat{\theta} \to \hat{\theta}$ and $\hat{\phi} \to -\hat{\phi}$. It follows then that the polarization vectors (7.4.9) of definite helicity are interchanged under space inversion; that is, $\mathbf{e}_{\pm}(-\hat{n}) = \mathbf{e}_{\mp}(\hat{n})$, and therefore according to Eq. (7.4.6), under space inversion $Q(\hat{n}) \to Q(-\hat{n})$ while $U(\hat{n}) \to -U(-\hat{n})$ and $V(\hat{n}) \to -V(-\hat{n})$. (The reader will later be able to check these spaceinversion properties of Q and U, by noting that these are the changes that are produced if replace the stochastic parameters $\alpha(\mathbf{q})$ and $\beta(\mathbf{q}, \lambda)$ in Eqs. (7.4.31) and (7.4.40) with their space-inversion transforms $\alpha(-\mathbf{q})$ and $\beta(-\mathbf{q}, -\lambda)$, respectively.) Thus by applying a space inversion to Eq. (7.4.15), we find that space inversion takes the partial wave amplitudes $a_{P,\ell m}$ into
$a'_{P,\ell m}$, where

$$\begin{aligned}
&Q(-\hat{n}) - iU(-\hat{n}) \\
&\quad = 2\sum_{\ell=2}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} a'_{P,\ell m}\sqrt{\frac{(\ell-2)!}{(\ell+2)!}}e_{+i}(\hat{n})e_{+j}(\hat{n})\tilde{\nabla}_i\tilde{\nabla}_j Y_\ell^m(\hat{n}) \\
&\quad = 2(-1)^\ell\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} a'_{P,\ell m}\sqrt{\frac{(\ell-2)!}{(\ell+2)!}}e_{-i}(-\hat{n})e_{-j}(-\hat{n})\tilde{\nabla}_i\tilde{\nabla}_j Y_\ell^m(-\hat{n}) \\
&\quad = (-1)^\ell\sum_{\ell=2}^{\infty}\sum_{m=-\ell}^{\ell} a'_{P,\ell m}\mathcal{Y}_\ell^{-m}(-\hat{n})^*
\end{aligned}$$

the factor $(-1)^\ell$ coming from the reflection property $Y_\ell^m(-\hat{n}) \to (-1)^\ell Y_\ell^m(\hat{n})$. Comparing this with Eq. (7.4.23), we see that space inversion changes $a_{P,\ell m}$ into $a'_{P,\ell m} = (-1)^\ell a^*_{P,\ell -m}$, so it changes $a_{E,\ell m}$ (and also the corresponding temperature multipole coefficient $a_{T,\ell m}$) by a sign$(-1)^\ell$, while it changes $a_{B,\ell m}$ by a sign $-(-1)^\ell$.

These coefficients are stochastic variables, governed by a probability distribution that is presumably invariant under space inversion, so there can be no bilinear correlation between polarization fluctuations of B and E type, or between temperature fluctuations and polarization fluctuations of B type, though there can be correlations between temperature fluctuations and polarization fluctuations of E type. Taking into account also the rotational invariance of the probability distribution, the only non-vanishing bilinear averages are of the form [6]

$$\left\langle a^*_{T,\ell m} a_{T,\ell' m'}\right\rangle = C_{TT,\ell}\delta_{\ell,\ell'}\delta_{m,m'} \tag{7.4.26}$$
$$\left\langle a^*_{T,\ell m} a_{E,\ell' m'}\right\rangle = C_{TE,\ell}\delta_{\ell,\ell'}\delta_{m,m'} \tag{7.4.27}$$
$$\left\langle a^*_{E,\ell m} a_{E,\ell' m'}\right\rangle = C_{EE,\ell}\delta_{\ell,\ell'}\delta_{m,m'} \tag{7.4.28}$$
$$\left\langle a^*_{B,\ell m} a_{B,\ell' m'}\right\rangle = C_{BB,\ell}\delta_{\ell,\ell'}\delta_{m,m'} \tag{7.4.29}$$

All $C_{XY,\ell}$ coefficients have the same dimensions, of square temperature. By their definition, the a_T, a_E, and a_B all satisfy the reality conditions

$$a^*_{T,\ell m} = a_{T,\ell -m}, a^*_{E,\ell m} = a_{E,\ell -m}, a^*_{B,\ell m} = a_{B,\ell -m}, \tag{7.4.30}$$

and hence the coefficients $C_{XY,\ell}$ are all real, while the $C_{XX,\ell}$ are all also positive. The results of measurements of correlations between temperature and/or polarization fluctuations in different directions are generally reported in terms of the $C_{XY,\ell}$.

[476] Here $a_{T,\ell m}$ is the coefficient $a_{\ell m}$ in Eq. (2.6.1). The multipole coefficients $C_{XY,\ell}$ defined here are the same as those defined by Zaldarriaga and Seljak, ref. 1. They are related to coefficients defined by Kamionkowski, Kosowsky, and Stebbins, ref. 1, by $C_\ell^G = C_{\ell,EE}I^2/2T_0^2, C_\ell^C = C_{\ell,BB}I^2/2T_0^2$, $C_\ell^{TG} = -C_{TE,\ell}I/\sqrt{2}T_0^3$, where I is the radiation intensity. (The superscripts G and C stand for "gradient" and "curl", because the pattern of polarization vectors of type E or B resemble the pattern of velocity in potential or solenoidal flow, respectively.)

Another simplification provided by the decomposition of the Stokes parameters into E and B terms is that, as we will see, scalar perturbations can only contribute to the E terms. This result is important, because it means that any sign of a primordial B-type polarization [7] will be clear evidence for cosmological gravitational waves, of the sort that we shall see in Chapter 10 are expected to be produced during inflation.

We now turn to the calculation of the multipole coefficients (7.4.27)(7.4.29). The hydrodynamic treatment that worked reasonably well for temperature correlation functions is not well suited to the treatment of polarization, so we will rely on the more accurate kinetic theory outlined in Section 6.1. Here we must distinguish between scalar and tensor modes. As already mentioned, these do not interfere, so the multipole coefficients can all be written as a sum of a scalar and a tensor term, denoted by superscripts S and T :

$$C_{XY,\ell} = C^S_{XY,\ell} + C^T_{XY,\ell}. \tag{7.4.31}$$

A. Scalar modes

For scalar modes, the matrix J_{ij} in Eq. (7.4.6) is given by the Fourier integral (6.1.18) and the line-of-sight integral (6.1.36):

$$\begin{aligned} J_{ij}\left(0,-\hat{n},t_0\right) = & \int d^3q\alpha(\mathbf{q}) \int_{t_1}^{t_0} dt \exp\left(-i\mathbf{q}\cdot\hat{p}\int_t^{t_0}\frac{dt'}{a\left(t'\right)} - \int_t^{t_0} dt'\omega_c\left(t'\right)\right) \\ & \times \Big[-\left(\delta_{ij}-\hat{n}_i\hat{n}_j\right)\left(\dot{A}_q\left(t'\right) - (\hat{n}\cdot\mathbf{q})^2\dot{B}_q\left(\prime\right)\right) \\ & + \frac{3\omega_c(t)}{2}\left(\delta_{ij}-\hat{n}_i\hat{n}_j\right)\Phi(q,t) \\ & + \frac{3\omega_c(t)}{4}\Pi(q,t)\left(\hat{q}_i-\hat{n}_i(\hat{q}\cdot\hat{n})\left(\hat{q}_j-\hat{n}_j(\hat{q}\cdot\hat{n})\right)\right. \\ & \left. + \frac{2\omega_c(t)}{a(t)}\left[\delta_{ij}-\hat{n}_i\hat{n}_j\right]\hat{n}_k\delta u_k(\mathbf{q},t)\Big]\right. . \end{aligned}$$

Since $\hat{n}_i e_{+i}(\hat{n}) = 0$ and $e_{+i}(\hat{n})e_{+i}(\hat{n}) = 0$, the only term in the integrand that contributes to the Stokes parameters is the one proportional to $\hat{q}_i\hat{q}_j$:

$$\begin{aligned} Q^S(\hat{n}) + iU^S(\hat{n}) = & \frac{3T_0}{8}e_{+i}(\hat{n})e_{+j}(\hat{n})\int d^3q\hat{q}_i\hat{q}_j\alpha(\mathbf{q}) \\ & \times\int_{t_1}^{t_0} dtP(t)\Pi(q,t)\exp\left(i\mathbf{q}\cdot\hat{n}\int_t^{t_0}\frac{dt'}{a\left(t'\right)}\right) \end{aligned} \tag{7.4.32}$$

where $P(t)dt$ is the probability that the last scattering occurs between t and $t+dt$:

$$P(t) = \omega_c(t)\exp\left(-\int_t^{t_0}\omega_c\left(t'\right)dt'\right) \tag{7.4.33}$$

As a reminder: $\alpha(\mathbf{q})$ is the stochastic parameter for whatever mode (presumably the non-decaying adiabatic mode) is assumed to dominate the scalar perturbations; $\omega_c(t)$ is the photon collision rate; t_1 is any time taken early enough so that a photon present at that time would suffer many collisions before the present; and $\Pi(q,t)$ is a source function, given by Eq. (6.1.30)

[477] Weak lensing by foreground objects converts the E-type polarization produced by scalar perturbations into a B-type polarization that may be large enough to interfere with measurements of the primordial B-type polarization; M. Zaldarriaga and U. Seljak, Phys. Rev. D 58, 023003 (1998)[astroph/9803150]. For a review, see A. Lewis and A. Challinor, Phys. Rep. 429, 1 (2006) [astro-ph/0601594].

in terms of the partial wave amplitudes of the temperature and polarization perturbation amplitudes:

$$\Pi(q,t) = \Delta^{(S)}_{T,2}(q,t) + \Delta^{(S)}_{P,0}(q,t) + \Delta^{(S)}_{P,2}(q,t). \tag{7.4.34}$$

To use Eq. (7.4.31) to calculate the EE and TE multipole coefficients, we recall the familiar formula

$$\exp(i\hat{q}\cdot\mathbf{v}) = 4\pi\sum_{\ell,m} i^\ell j_\ell(v) Y^m_\ell(\hat{v}) Y^{m*}_\ell(\hat{q}),$$

where here $\mathbf{v} = \hat{n}q\int_t^{t_0} dt'/a\,(t')$. Acting on this, we can replace $\hat{q}_i$ with $-i\partial/\partial v_i$. Since $\hat{v} = \hat{n}$, we can write $\partial/\partial v_i = \hat{n}_i\partial/\partial v + \tilde{\nabla}_i/v$, where $\tilde{\nabla}$ is the angular gradient operator (7.4.17) acting on $\hat{n}$. Since $\hat{n}_i e_{+i} = 0$ and $e_{+i}e_{+i} = 0$, Eq. (7.4.31) then reads

$$\begin{aligned} Q^S(\hat{n}) + iU^S(\hat{n}) = &-\frac{3\pi T_0}{2} e_{+i}(\hat{n}) e_{+j}(\hat{n}) \tilde{\nabla}_i \tilde{\nabla}_j \sum_{\ell m} i^\ell Y^m_\ell(\hat{n}) \int d^3q \alpha(\mathbf{q}) \\ &\times Y^{m*}_\ell(\hat{q}) \int_{t_1}^{t_0} dt P(t) \Pi(q,t) \frac{j_\ell(qr(t))}{q^2 r(t)^2} \end{aligned} \tag{7.4.35}$$

where here $r(t)$ is the radial coordinate of a point from which light emitted at time t would just reach us at present

$$r(t) = \int_t^{t_0} \frac{dt'}{a\,(t')} \tag{7.4.36}$$

(Note that $\tilde{\nabla}_i\hat{n}_j$ doesn't contribute here, because it equals $\delta_{ij} - \hat{n}_i\hat{n}_j$, which vanishes when contracted with $e_{+i}e_{+j}$.) Comparing Eq. (7.4.34) with Eqs. (7.4.15) and (7.4.16), we see that the scalar contribution to $a_{P,\ell m}$ is

$$a^S_{P,\ell m} = -\frac{3\pi T_0 i^\ell}{4}\sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \int d^3q \alpha(\mathbf{q}) Y^{m*}_\ell(\hat{q}) \int_{t_1}^{t_0} dt \tag{7.4.37}$$

$$\times P(t)\Pi(q,t)\frac{j_\ell(qr(t))}{q^2 r^2(t)} \tag{7.4.38}$$

To check the reality properties of $a_{P,\ell m}$, we note that Eqs. (6.1.18) and (6.1.21) require that $\alpha^*(\mathbf{q})\Pi^*(q,t) = \alpha(-\mathbf{q})\Pi(q,t)$, while $Y^m_\ell(\hat{q}) = (-1)^\ell Y^{-m*}_\ell(-\hat{q})$, and of course $i^{\ell*} = (-1)^\ell i^\ell$, so $a^{S*}_{P,\ell m} = a^S_{P,\ell -m}$. Inspection of Eq. (7.4.25) then shows that, as promised, the scalar modes contribute only an E-type polarization

$$a^S_{E,\ell m} = -a^S_{P,\ell m}, \quad a^S_{B,\ell m} = 0. \tag{7.4.39}$$

We can now give formulas for the scalar contribution to the EE and TE multipole coefficients defined by Eqs. (7.4.28) and (7.4.27). Recalling the normalization condition (7.2.5) of the scalar stochastic parameter $\alpha(\mathbf{q})$, we see from Eqs. (7.4.36) and (7.4.37) that [8]

$$C^S_{EE,\ell} = \frac{9\pi^2 T_0^2}{16}\frac{(\ell+2)!}{(\ell-2)!}\int_0^\infty q^2 dq \left|\int_{t_1}^{t_0} dt P(t)\Pi(q,t)\frac{j_\ell(qr(t))}{q^2 r^2(t)}\right|^2 \tag{7.4.40}$$

Combining Eq. (7.4.36) with the general formula (7.2.4) for the scalar contribution to the temperature partial wave amplitudes, we find in the same way that

$$C^S_{TE,\ell} = -3\pi^2 T_0^2 \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \int_0^\infty q^2 dq \times \int_{t_1}^{t_0} dt P(t) \Pi(q,t) \frac{j_\ell(qr(t))}{q^2 r^2(t)} \times \int_{t_1}^{t_0} [j_\ell(qr(t))F(q,t) + j'_\ell(qr(t))G(q,t) + j''_\ell(qr(t))H(q,t)] \quad (7.4.41)$$

7 Anisotropies in the Microwave Sky

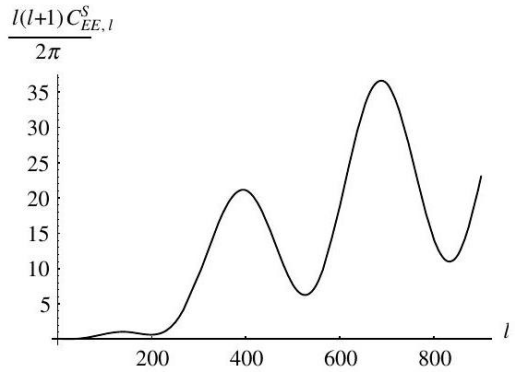


Figure 7.4: The multipole coefficient $\ell(\ell+1)C^S_{EE,\ell}/2\pi$ in square microKelvin, vs. ℓ, for the cosmological parameters given in Eqs. (7.2.44) and (7.2.45).

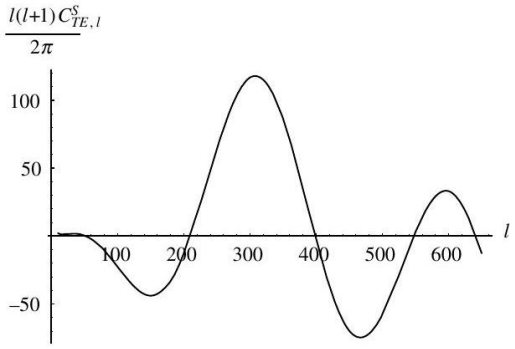


Figure 7.5: The multipole coefficient $\ell(\ell+1) C_{T E, \ell}^{S} / 2 \pi$ in square microKelvin, vs. ℓ, for the cosmological parameters given in Eqs. (7.2.44) and (7.2.45).

where $F(q,t), G(q,t)$, and $H(q,t)$ are the quantities (7.2.7)-(7.2.9). The values of $C^S_{EE,\ell}$ and $C^S_{TE,\ell}$, calculated [9] for the cosmological parameters (7.2.44)-(7.2.45), are shown in Figures 7.4 and 7.5. Comparing these figures with each other and with Figure (7.3), we see that $C^S_{EE,\ell} \ll C^S_{TE,\ell} \ll C^S_{TT,\ell}$. The microwave background polarization is small, because the universe goes swiftly at the time of recombination from a state of nearly perfect thermal equilibrium to one of nearly perfect transparency, and photons are not polarized in thermal equilibrium.

Eqs. (7.4.38) and (7.4.39) must be used for maximum numerical accuracy, but we can get a good idea of the ℓ dependence of the multipole coefficients for $\ell \gg 1$ by using much simpler

[478] This is the same as Eq. (17) of Zaldarriaga and Seljak, ref. 1, except that Π here is 4 times their Π.

approximate versions. The function $P(t)$ in Eq. (7.4.38) is sharply peaked at a time t_L of last scattering, and it has

$\int dt P(t) = 1$, so to a fair approximation we can do the integral over time by simply setting $t = t_L$ and dropping the factor $P(t)$. Also, by using the same approximations that led us from Eq. (7.2.13) to Eq. (7.2.17), we find that for $\ell \gg 1$

$$C^S_{EE,\ell} \approx \frac{9\pi^2\ell^3}{32 r_L^3} \int_1^\infty \frac{d\beta}{\beta\sqrt{\beta^2-1}} \Pi^2\left(\frac{\ell\beta}{r_L}, t_L\right).$$

This does not give a reliable result for the magnitude of $C^S_{EE,\ell}$, because the function $\Pi(q,t)$ varies rapidly around the time of recombination, but since this function falls off rapidly with q, so that the integral is dominated by $\beta \approx 1$, this approximation does suggest that the peaks and valleys in $C^S_{EE,\ell}$ arise chiefly from a factor $\Pi^2\left(\ell/r_L, t_L\right)$. On the other hand, as remarked in Section 7.2, the peaks and valleys in $C^S_{TT,\ell}$ arise chiefly from a factor $F^2\left(\ell/r_L\right)$, and by the same reasoning, the peaks and valleys in $C^S_{TE,\ell}$ arise chiefly from a factor $F\left(\ell/r_L\right)\Pi\left(\ell/r_L, t_L\right)$.

To get an idea of the q-dependence of $\Pi\left(q, t_L\right)$, we note first that, to the extent that the photon polarization arises solely from the last scattering, $\Pi(q,t)$ would be given by just the first term $\Delta^{(S)}_{T,2}(q,t)$ in Eq. (7.4.33). This is because the differential cross section for the scattering of an unpolarized photon of initial momentum $\mathbf{p}$ with $|\mathbf{p}| \ll m_e$ by a non-relativistic electron to yield a photon with real polarization vector $\mathbf{e}$ is [10]

$$\frac{d\sigma}{d\Omega} = \frac{e^4}{32\pi^2 m_e^2}\left(1 - (\hat{p}\cdot\mathbf{e})^2\right).$$

If the distribution of initial momenta $\mathbf{p}$ were spherically symmetric then the average of this cross section over initial directions would be independent of $\mathbf{e}$, so the final photon would be unpolarized. But for a direction-dependent phase space density $n_{\gamma,L}(\mathbf{p})$ of photons with momentum $\mathbf{p}$ at the position and time of last scattering, the intensity of photons after last scattering coming from direction $\hat{n}$ and having polarization vector $\mathbf{e}$ (with $\mathbf{e}^2 = 1$ and e $\cdot\hat{n} = 0$) is proportional to

$$\int d^3p p p n_{\gamma,L}(\mathbf{p})\left(1 - (\hat{p}\cdot\mathbf{e})^2\right).$$

The part of J_{ij} that contributes to the Stokes parameters is thus proportional to the traceless part of $\int d^3p p p n_{\gamma,L}(\mathbf{p})\hat{p}_i\hat{p}_j$, which in turn is proportional to the $\ell = 2$ part of $\Delta^{(S)}_T$ evaluated at the time and position of last scattering.

The additional terms $\Delta^{(S)}_{P,0}(q,t) + \Delta^{(S)}_{P,2}(q,t)$ in $\Pi(q,t)$ take account of whatever polarization the photon may already have acquired by the time of last scattering, and can be expected to be relatively small since photons must be unpolarized under conditions of rapid photon scattering. But then, according to Eq. (6.1.32),

$$\Pi\left(q, t_L\right) \simeq \Delta^{(S)}_{T,2}(q,t) \simeq q^2\pi^S_q\left(t_L\right)/\bar{\rho}_\gamma\left(t_L\right).$$

Recall that, according to the first line of Eq. (5.1.43), $\partial_i\partial_j\pi^S(\mathbf{x},t)$ is the term in the scalar part of $\delta T^i{}_j$ that is not proportional to δ_{ij}, while Eq. (B.50) tells us that for short mean free times, this term equals [11]

$$\partial_i\partial_j\pi^S = a^{-2}\eta_\gamma\partial_i\partial_j\delta\tilde{u}$$

[479] http://lambda.gsfc.nasa.gov/data/map/powspec/wmap_lcdm_pl_model_yr1_ v1.txt

[4710] See, e.g., QTF, Eq. (8.7.40). This was first calculated by O. Klein and Y. Nishina, Z. Phys. 52, 853 (1929).

where η_γ is the shear viscosity due to photon momentum transport, given in terms of the photon mean free time t_γ by [12] $\eta_\gamma = \frac{16}{45}\bar{\rho}_\gamma t_\gamma$, and $\delta\tilde{u}(\mathbf{x}, t)$ is the gauge-invariant velocity potential [13]

$$\delta\tilde{u} \equiv \delta u - aF + \frac{a^2\dot{B}}{2},$$

with F and B metric perturbations defined by Eqs. (5.1.32) and (5.1.33). Also, Eq. (7.1.45) (which was derived for a gauge in which the metric component F vanishes) gives the coefficient of the stochastic parameter $\alpha(\mathbf{q})$ in the Fourier transform of the gauge-invariant velocity potential in terms of the form factor $G(q)$ as $\delta\tilde{u}_q(t_L) = -a(t_L)\,G(q)/q$. We conclude then that

$$\Pi(q, t_L) \simeq \frac{16q^2}{45a^2(t_L)}\bar{t}_\gamma\delta\tilde{u}_q(t_L) \simeq -\frac{16q}{45a(t_L)}\bar{t}_\gamma G(q)$$

where $\bar{t}_\gamma$ is some appropriate average of $t_\gamma(t)$ during the era of recombination. Thus the peaks and valleys in $C^S_{EE,\ell}$ and in $\left|C^S_{TE,\ell}\right|$ are more-or-less the same as those in $|G(q/r_L)|^2$ and $|F(q/r_L)\,G(q/r_L)|$, respectively, and we recall that the peaks and valleys in $C^S_{TT,\ell}$ are essentially those in $|F(q/r_L)|^2$. Eqs. (7.2.23) and (7.2.24) give

$$F(q) \simeq \frac{\mathcal{R}^o_q}{5}\left[3R_L\mathcal{T}(\kappa) - (1+R_L)^{-1/4}\,e^{-\int_0^{t_L}\Gamma dt}\mathcal{S}(\kappa)\cos\chi(q)\right]$$

$$G(q) = -\frac{\sqrt{3}\mathcal{R}^o_q}{5\,(1+R_L)^{3/4}}e^{-\int_0^{t_L}\Gamma dt}\mathcal{S}(\kappa)\sin\chi(q),$$

where $\chi(q)$ is an approximately linear function of q :

$$\chi(q) \equiv \int_0^{t_L}\frac{qdt}{a\sqrt{3(1+R)}} + \Delta(\kappa)$$

and the argument κ of the transfer functions S, T, and Δ is proportional to q. As remarked in Section 7.2, the peaks in $C^S_{TT,\ell}$ are thus roughly at $\chi(\ell/r_L) = \pi, 2\pi, 3\pi, \ldots$ On the other hand, we expect the peaks in $C^S_{EE,\ell}$ to be roughly at the peaks in $|\sin(\chi(\ell/r_L))|$, or in other words, at $\chi(\ell/r_L) = \pi/2, 3\pi/2, 5\pi/2, \ldots$, and the peaks in $\left|C^S_{TE,\ell}\right|$ to be roughly at the peaks in $|\cos(\chi(\ell/r_L))\sin(\chi(\ell/r_L))|$, or in other words, at $\chi(\ell/r_L) = \pi/4, 3\pi/4, 5\pi/4, \ldots$ In particular, we expect a peak in $C^S_{EE,\ell}$ before the first peak in $C^S_{TT,\ell}$, followed by one peak in $C^S_{EE,\ell}$ between successive peaks in $C^S_{TT,\ell}$, while there should be two peaks in $\left|C^S_{TE,\ell}\right|$ before the first peak in $C^S_{TT,\ell}$, followed by two peaks in $\left|C^S_{TE,\ell}\right|$ between successive peaks in $C^S_{TT,\ell}$. This is precisely the pattern seen in computer calculations. [14] For instance, for a plausible set of cosmological parameters there are indeed two peaks in $\left|C^S_{TE,\ell}\right|$ (at $\ell = 36$ and $\ell = 144$) below the first peak in $C^S_{TT,\ell}$ at $\ell = 230$.

The E-type polarization and temperature-polarization correlation were first detected by the Degree Angular Scale Interferometer (DASI) collaboration. [15] Then the Wilkinson Microwave Anisotropy Probe (WMAP) collaboration measured the coefficients $C_{TE,\ell}$ with good accuracy over a range of multipole orders from $\ell = 2$ to $\ell \simeq 570$.[16] The results for $\ell > 10$ are in good agreement with the concordance model, assuming primordial fluctuations with a spectrum close to the Harrison-Zel'dovich form $\mathcal{R}^o_q \propto q^{-3/2}$, and a constant vacuum energy along with cold dark

[4711] Here we use the relation $\delta u^i{}_{;j} = \delta u_j{}^i = a^{-2}\delta u_{i;j} - a^{-4}h_{ij}a\dot{a}$.

[12] N. Kaiser, Mon. Not. Roy. Astron. Soc. 202, 1169 (1983). For a discussion, see footnote 5 of Section 6.4.

[13] We recognize $\delta\tilde{u}$ as the same velocity potential that appears in Eq. (7.1.37) as the Doppler contribution to the temperature shift in the class of gauges with vanishing metric component F. In the special case of Newtonian gauge we have $F = B = 0$, and $\delta\tilde{u} = \delta u$.

matter, baryonic matter, photons, and massless neutrinos, and using the same values of the cosmological parameters $\Omega_B h^2, \Omega_M h^2, h$, and N as used in fitting the model to measurements of temperature anisotropies. In particular, there is clear evidence of the second expected peak in $|C_{TE,\ell}|$, at around $\ell \simeq 140$, and the third expected peak, at $\ell \simeq 300$. More recently, the QUaD collaboration [17] has reported preliminary results from the first season of operation of a microwave polarimeter at the South Pole, which show clear evidence for the first peak in $C_{EE,\ell}$ at $\ell \simeq 350$, about where it is expected from scattering around the time of recombination. But if reionization were not included in the cosmological model, there would be a clear discrepancy between its predictions and the data for $C_{TE,\ell}$ for $\ell < 10$. This discrepancy is attributed to polarization caused by scattering in a plasma that has been reionized by ultraviolet light from a first generation of stars.

The most important application of polarization measurements so far has been in working out the history of the reionization of intergalactic matter at redshifts much less than $z_L \simeq 1090$, which contributes to the last-scattering probability distribution $P(t)$. Because nearby events subtend large angles, the additional terms in $C_{EE,\ell}$ and $C_{TE,\ell}$ due to reionization are negligible except for relatively small ℓ, in fact $\ell < 10$. After three years of operation, the WMAP collaboration [18] found on the basis of the EE correlation alone that if reionization is sudden and complete at a redshift z_r, then $z_r = 10.9^{+2.7}_{-2.3}$, corresponding to an optical depth $\tau = 0.09 \pm 0.03$. This result is in line with expectations of the onset of star formation, and it has played an essential role in the use of temperature correlations to determine the magnitude of $\mathcal{R}_q^o$ and to set an upper limit on $\mathcal{D}_q^o$, discussed in Sections 7.2 and 7.3, respectively. Scattering by the reionized plasma produces a peak in $C_{EE,\ell}$ much like the peak at $\ell \simeq 140$, but shifted to much smaller ℓ, around $\ell \simeq 4$, because it occurs at an angular diameter distance much smaller than the angular diameter distance of the era of recombination. The WMAP measurements and other measurements of $C_{EE,\ell}$ for larger values of ℓ are in good agreement with the theoretical formula (7.4.38) for $C^S_{EE,\ell}$, as shown in Figure 7.6, and future measurements are expected to further reduce the uncertainties in this comparison.

B. Tensor modes

For tensor modes, the matrix J_{ij} in Eq. (7.4.4) is given by the Fourier integral (6.6.9) and line-of-sight integral (6.6.24):

$$J_{ij}(\mathbf{x}, \hat{p}, t) = \sum_{\lambda=\pm 2} \int d^3q e^{i\mathbf{q}\cdot\mathbf{x}} \beta(\mathbf{q}, \lambda) \times \int_{t_1}^{t} dt' \exp\left(-i\mathbf{q}\cdot\hat{p} \int_{t'}^{t} \frac{dt''}{a\,(t'')} - \int_{t'}^{t} dt'' \omega_c\,(t'') \right)$$

[4714] For instance, see http://lambda.gsfc.nasa.gov/data/map/powspec/wmap_lcdm_pl_model_vr1_v1.txt.

[15] J. Kovac et al., Nature 420, 772 (2002).

[16] A. Kogut et al., Astrophys. J. Suppl. 148, 161 (2003).

[17] P. Ade et al., 0705.2359.

[4718] L. Page et al., Astrophys. J. Suppl. Ser. 170, 335 (2007) [astro-ph/0603450].

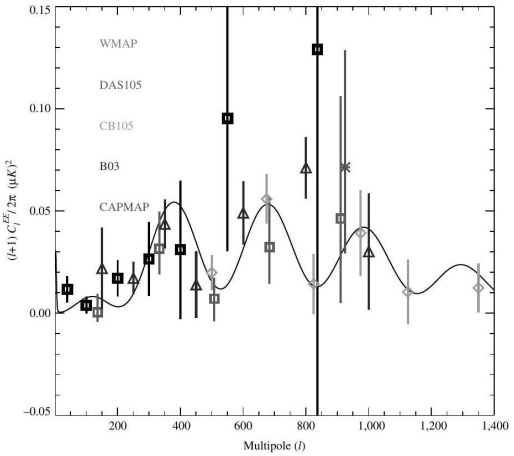


Figure 7.6: Observed values of the multipole coefficient $(\ell+1) C_{E E, \ell} / 2 \pi$ in square microKelvin, vs. ℓ, from L. Page et al., astro-ph/0603450. (Note that $C_{EE,\ell}/2\pi$ is multiplied here with $\ell+1$, rather than the usual $\ell(\ell+1)$.) Dark squares are the WMAP data; the triangles are BOOMERanG data; the lighter squares are the DASI data; the diamonds are the CBI data; and the asterisk is the CAPMAP datum. The solid curve is the theoretical curve for $(\ell+1)C^S_{EE,\ell}/2\pi$, with cosmological parameters taken from a fit to WMAP temperature and low- ℓ polarization data.

$$
\begin{aligned}
&\times \Big[-\hat{p}_k\hat{p}_l\left(\delta_{ij}-\hat{p}_i\hat{p}_j\right)e_{kl}(\hat{q},\lambda)\dot{\mathcal{D}}_q\left(t'\right)\\
&-\omega_c\left(t'\right)\Psi\left(q,t'\right)\left(e_{ij}(\hat{q},\lambda)-\hat{p}_i\hat{p}_k e_{kj}(\hat{q},\lambda)\right.\\
&\left.-\hat{p}_j\hat{p}_k e_{ik}(\hat{q},\lambda)+\hat{p}_i\hat{p}_j\hat{p}_k\hat{p}_l e_{kl}(\hat{q},\lambda)\right)\Big]
\end{aligned}
$$

Again using the conditions $\hat{n}_i e_{+i}(\hat{n}) = 0$ and $e_{+i}(\hat{n})e_{+i}(\hat{n}) = 0$, we see that the only term in the integrand here that contributes to the Stokes parameters in Eq. (7.4.4) is the one proportional to $e_{ij}\Psi$. This term gives

$$
\begin{aligned}
Q^T(\hat{n}) + iU^T(\hat{n}) = &-\frac{1}{2}T_0 e_{+i}(\hat{n})e_{+j}(\hat{n})\sum_{\lambda=\pm 2}\int d^3q\beta(\mathbf{q},\lambda)\\
&\times e_{ij}(\hat{q},\lambda)\int_{t_1}^{t_0}dt\exp\left(i\mathbf{q}\cdot\hat{n}\int_t^{t_0}\frac{dt'}{a\left(t'\right)}\right)P(t)\Psi(q,t)
\end{aligned}
\tag{7.4.42}
$$

where $\beta(\mathbf{q},\lambda)$ is the stochastic parameter for a gravitational wave of wave number $\mathbf{q}$ and helicity λ; $P(t)$ is the last-scattering distribution (7.4.32); and $\Psi(q,t)$ is a source function, given by Eq. (6.6.21) in terms of the partial wave amplitudes of the temperature and polarization perturbation amplitudes:

$$
\begin{aligned}
\Psi(q,t) = &\frac{1}{10}\Delta^{(T)}_{T,0}(q,t)+\frac{1}{7}\Delta^{(T)}_{T,2}(q,t)+\frac{3}{70}\Delta^{(T)}_{T,4}(q,t)-\frac{3}{5}\Delta^{(T)}_{P,0}(q,t)\\
&+\frac{6}{7}\Delta^{(T)}_{P,2}(q,t)-\frac{3}{70}\Delta^{(T)}_{P,4}(q,t)
\end{aligned}
\tag{7.4.43}
$$

Also, $e_{ij}(\hat{q},\lambda)$ is the polarization tensor

$$
e_{ij}(\hat{q},\lambda) = S_{ik}(\hat{q})S_{jl}(\hat{q})e_{kl}(\hat{z},\lambda),
\tag{7.4.44}
$$

where $S_{ij}(\hat{q})$ is the standard three-dimensional rotation (7.4.12) that takes the three-axis into the direction of $\hat{q}$, and $e_{kl}(\hat{z},\lambda)$ is the polarization tensor (5.2.15) for waves traveling in the three-direction, with non-vanishing components

$$e_{11}(\hat{z},\pm 2) = -e_{22}(\hat{z},\pm 2) = \mp i e_{12}(\hat{z},\pm 2) = \mp i e_{21}(\hat{z},\pm 2) = \frac{1}{\sqrt{2}}. \tag{7.4.45}$$

We will first show how to find the tensor multipole coefficients $C^T_{EE,\ell}$, $C^T_{BB,\ell}$, and $C^T_{TE,\ell}$ for a given source function $\Psi(q,t)$, and then report the results of a numerical evaluation using $\Psi(q,t)$ calculated as described in Section 7.3.

We begin by deriving a formula for the coefficients $a^T_{P,\ell m}$ in the expansion of the Stokes parameters in spin-weighted spherical harmonics:

$$Q^T(\hat{n}) + iU^T(\hat{n}) = \sum_{\ell m} a^T_{P,\ell m} \mathcal{Y}^m_\ell(\hat{n}). \tag{7.4.46}$$

Using Eq. (7.4.40) and the orthonormality property (7.4.21) of the $\mathcal{Y}^m_\ell$, we have

$$\begin{aligned} a^T_{P,\ell m} &= \int d^2\hat{n} \mathcal{Y}^{m*}_\ell(\hat{n})(Q(\hat{n}) + iU(\hat{n}))^T \\ &= -\frac{T_0}{2} \sum_\lambda \int d^3q \beta(\mathbf{q},\lambda) \int_{t_1}^{t_0} dt P(t) \Psi(q,t) g_{\ell m}(qr(t),\hat{q},\lambda) \end{aligned} \tag{7.4.47}$$

where $r(t)$ is the unperturbed co-moving radial coordinate of a photon at time t that reaches us at time t_0 :

$$r(t) \equiv \int_t^{t_0} \frac{dt'}{a\,(t')} \tag{7.4.48}$$

and

$$g_{\ell m}(\rho,\hat{q},\lambda) \equiv \int d^2\hat{n} \mathcal{Y}^{m*}_\ell(\hat{n}) e^{i\rho\hat{q}\cdot\hat{n}} \sum_{ij} e_{+i}(\hat{n}) e_{+j}(\hat{n}) e_{ij}(\hat{q},\lambda) \tag{7.4.49}$$

It is convenient first to calculate the amplitude $g_{\ell m}(\rho,\hat{q},\lambda)$ for $\hat{q}$ along the direction $\hat{z}$ of the three-axis. Using the graviton polarization tensor (7.4.43) and the photon polarization vector (7.4.9), a straightforward calculation gives
$e_{+i}(\hat{n}) e_{+j}(\hat{n}) e_{ij}(\hat{z},\pm 2) = \frac{1}{\sqrt{2}} \left(e_{+1}(\hat{n}) \pm i e_{+2}(\hat{n})\right)^2 = \frac{1}{2\sqrt{2}} e^{\pm 2i\phi}(1 \mp \cos\theta)^2$,
where as usual ϕ and θ are the azimuthal and polar angles of the direction $\hat{n}$. Eq. (7.4.22) shows that $\mathcal{Y}^{m*}_\ell(\hat{n})$ is $e^{-im\phi}$ times a function of θ, so the ϕ integral in Eq. (7.4.47) vanishes unless $m = \lambda = \pm 2$, in which case it just gives a factor 2π. For $\mathcal{Y}^{\pm 2}_\ell$ we will use the top line of Eq. (7.4.22), which gives
$\mathcal{Y}^{\pm 2}_\ell(\theta,\phi) = \sqrt{\frac{(\ell-2)!}{(\ell+2)!}} \partial_\pm Y^{\pm 2}_\ell(\theta,\phi) = e^{\pm 2i\phi} \sqrt{\frac{2\ell+1}{4\pi}} \frac{(\ell-2)!}{(\ell+2)!} \partial_\pm P^2_\ell(\cos\theta)$,
where $\partial_\pm$are the differential operators

$$\begin{aligned} \partial_\pm &\equiv \left(\frac{d}{d\theta} \mp 2\csc\theta\right)^2 - \cot\theta \left(\frac{d}{d\theta} \mp 2\csc\theta\right) \\ &= \left(1-\mu^2\right) \frac{d}{d\mu^2} \pm 4\frac{d}{d\mu} + \frac{4(1\pm\mu)}{1-\mu^2} \end{aligned} \tag{7.4.52}$$

where $\mu \equiv \cos\theta$, and $P^2_\ell(\cos\theta)$ is the usual associated Legendre polynomial

$$P_\ell^2(\mu) = \left(1-\mu^2\right)\frac{d^2}{d\mu^2}P_\ell(\mu). \tag{7.4.53}$$

Putting this together, for $\mathbf{q}$ in the three-direction $\hat{z}$, Eq. (7.4.47) becomes

$$g_{\ell m}(\rho,\hat{z},\pm 2) = \delta_{m,\pm 2}\sqrt{\frac{\pi(2\ell+1)}{8}}\frac{(\ell-2)!}{(\ell+2)!}\int_{-1}^{+1} d\mu\left[\partial_\pm P_\ell^2(\mu)\right]e^{i\rho\mu}(1\mp\mu)^2 \tag{7.4.54}$$

We use (7.4.51) for the associated Legendre polynomial, use integration by parts to let the derivatives in $\partial_\pm$act on $(1\mp\mu)^2 e^{i\rho\mu}$, and apply the formula

$$\begin{aligned}&\frac{\partial^2}{\partial\mu^2}\left(\left(1-\mu^2\right)(1\mp\mu)^2 e^{i\rho\mu}\right) + \left(\mp 4\frac{\partial}{\partial\mu} + \frac{4(1\pm\mu)}{1-\mu^2}\right)(1\mp\mu)^2 e^{i\rho\mu}\\ &\quad = \left(1-\mu^2\right)\left(12\mp 8i\rho(1\mp\mu) - \rho^2(1\mp\mu)^2\right)e^{i\rho\mu}.\end{aligned}$$

Replacing μ with $-i\partial/\partial\rho$, we find

$$\begin{aligned}g_{\ell m}(\rho,\hat{z},\pm 2) =& \delta_{m,\pm 2}\sqrt{\frac{\pi(2\ell+1)}{8}}\frac{(\ell-2)!}{(\ell+2)!}\int_{-1}^{+1} d\mu P_\ell(\mu)\\ &\times\frac{\partial^2}{\partial\mu^2}\left[12\mp 8i\rho(1\mp\mu)-\rho^2(1\mp\mu)^2\right]\left(1-\mu^2\right)^2 e^{i\rho\mu}\\ =& -\delta_{m,\pm 2}\sqrt{\frac{\pi(2\ell+1)}{8}}\frac{(\ell-2)!}{(\ell+2)!}\int_{-1}^{+1} d\mu P_\ell(\mu)\\ &\times\left[12+8\rho\frac{\partial}{\partial\rho}-\rho^2+\rho^2\frac{\partial^2}{\partial\rho^2}\mp 8i\rho\mp 2i\rho^2\frac{\partial}{\partial\rho}\right]\\ &\times\left(1+\frac{\partial^2}{\partial\rho^2}\right)^2\rho^2 e^{i\rho\mu}\end{aligned}$$

The integral over μ is now simply

$$\int_{-1}^{+1} d\mu e^{i\rho\mu}P_\ell(\mu) = 2i^\ell j_\ell(\rho).$$

A straightforward though tedious calculation using the defining differential equation $j_\ell''(\rho)+(2/\rho)j_\ell'(\rho)+$ 0 gives

$$\left(1+\frac{\partial^2}{\partial\rho^2}\right)^2\rho^2 j_\ell(\rho) = \frac{(\ell+2)!}{(\ell-2)!}\frac{j_\ell(\rho)}{\rho^2},$$

so finally

$$\begin{aligned}g_{\ell m}(\rho,\hat{z},\pm 2) =& -2i^\ell\delta_{m,\pm 2}\sqrt{\frac{\pi(2\ell+1)}{8}}\\ &\times\left[12+8\rho\frac{\partial}{\partial\rho}-\rho^2+\rho^2\frac{\partial^2}{\partial\rho^2}\mp 8i\rho\mp 2i\frac{\partial}{\partial\rho}\right]\frac{j_\ell(\rho)}{\rho^2}\end{aligned} \tag{7.4.55}$$

.

Just as in Eq. (7.3.13) for temperature fluctuations, the amplitude for $\mathbf{q}$ in a general direction is given by applying the standard rotation $S(\hat{q})$ that takes the three-axis into the direction of $\mathbf{q}$:

$$g_{\ell m}(\rho,\hat{q},\pm 2)=\sum_{m'}D^{(\ell)}_{m,m'}(S(\hat{q}))g_{\ell m'}(\rho,\hat{z},\pm 2) \tag{7.4.56}$$

where $D^{(\ell)}$ is the spin- ℓ unitary representation of the rotation group. Using Eqs. (7.4.45), (7.4.54), and (7.4.53), we have then

$$\begin{aligned}a^T_{P,\ell m} =& T_0 i^\ell\sqrt{\frac{\pi(2\ell+1)}{8}}\sum_{\pm}\int d^3q\beta(\mathbf{q},\pm 2)D^{(\ell)}_{m,\pm 2}(S(\hat{q}))\\ &\times\int_{t_1}^{t_0}dtP(t)\Psi(q,t)\\ &\times\left\{\left[12+8\rho\frac{\partial}{\partial\rho}-\rho^2+\rho^2\frac{\partial^2}{\partial\rho^2}\mp 8i\rho\mp 2i\rho^2\frac{\partial}{\partial\rho}\right]\frac{j_\ell(\rho)}{\rho^2}\right\}_{\rho=qr(t)}\end{aligned} \tag{7.4.57}$$

In order to separate the E and B terms in $a^T_{P,\ell m}$, we need the reality properties of β and $D^{(\ell)}$. First, note that

$$e_{ij}(\hat{q},\pm 2)=\frac{1}{\sqrt{2}}\left(S_{i1}(\hat{q})\pm iS_{i2}(\hat{q})\right)\left(S_{j1}(\hat{q})\pm iS_{j2}(\hat{q})\right)$$

and

$$S_{i1}(\hat{q})\pm iS_{i2}(\hat{q})=(\cos\theta\cos\phi\mp i\sin\phi,\cos\theta\sin\phi\pm i\cos\phi,-\sin\theta),$$

so

$$e^*_{ij}(\hat{q},\pm 2)=e_{ij}(\hat{q},\mp 2)=e_{ij}(-\hat{q},\pm 2).$$

With the gravitational field dominated by a single mode, $\mathcal{D}^*_q(t)$ must be proportional to $\mathcal{D}_q(t)$, so that by absorbing any phase in $\beta(\mathbf{q},\pm 2)$ we can choose $\mathcal{D}_q(t)$ to be real. The reality of $D_{ij}(\mathbf{x},t)$ then requires that

$$\beta^*(\mathbf{q},\pm 2)=\beta(-\mathbf{q},\pm 2). \tag{7.4.58}$$

Also, by writing $D^{(\ell)}(S(\hat{q}))=\exp\left(-i\phi J^{(\ell)}_3\right)\exp\left(-i\theta J^{(\ell)}_2\right)$ where $J^{(\ell)}_i$ are the angular momentum matrices [19] for angular momentum ℓ, we can
show that

$$D^{(\ell)}_{m,\pm 2}(S(\hat{q}))^*=(-1)^\ell D^{(\ell)}_{-m,\pm 2}(S(-\hat{q})). \tag{7.4.59}$$

The sign $(-1)^\ell$ here cancels the same sign in $i^{-\ell}=(-1)^\ell i^\ell$. It follows that the whole effect of taking the complex conjugate of $a^T_{P,\ell m}$, and changing m to $-m$, is to replace the differential operator acting on $j_\ell(\rho)/\rho^2$ with its complex conjugate:

[4719] As noted in footnote 4, the phases in these angular momentum matrices depend on the phase convention chosen for the spherical harmonics. They are related by $\int d^2\hat{n}Y^{m'}_\ell(\hat{n})^*\mathbf{L}Y^m_\ell(\hat{n})=\mathbf{J}^\ell_{m'm}$, where $\mathbf{L}\equiv -i\mathbf{x}\times\nabla$ is the orbital angular momentum operator.

$$
\begin{aligned}
a^{T*}_{P,\ell-m} =& T_0 i^\ell \sqrt{\frac{\pi(2\ell+1)}{8}} \sum_{\pm} \int d^3q \beta(\mathbf{q},\pm 2) D^{(\ell)}_{m,\pm 2}(S(\hat{q})) \\
&\times \int_{t_1}^{t_0} dt P(t) \Psi(q,t) \\
&\times \left\{ \left[12 + 8\rho\frac{\partial}{\partial\rho} - \rho^2 + \rho^2 \frac{\partial^2}{\partial\rho^2} \pm 8i\rho \pm 2i\rho^2 \frac{\partial}{\partial\rho} \right] \frac{j_\ell(\rho)}{\rho^2} \right\}_{\rho=qr(t)}
\end{aligned}
\tag{7.4.60}
$$

The definitions (7.4.25) then give

$$
\begin{aligned}
a^T_{E,lm} =& -i^\ell T_0 \sqrt{\frac{\pi(2\ell+1)}{8}} \sum_{\pm} \int d^3q \beta(\mathbf{q},\pm 2) D^{(\ell)}_{m,\pm 2}(S(\hat{q})) \\
&\times \int_{t_1}^{t_0} dt P(t) \Psi(q,t) \\
&\times \left\{ \left[12 + 8\rho\frac{\partial}{\partial\rho} - \rho^2 + \rho^2 \frac{\partial^2}{\partial\rho^2} \right] \frac{j_\ell(\rho)}{\rho^2} \right\}_{\rho=qr(t)}
\end{aligned}
\tag{7.4.61}
$$

$$
\begin{aligned}
a^T_{B,lm} =& i^\ell T_0 \sqrt{\frac{\pi(2\ell+1)}{8}} \sum_{\pm} \pm \int d^3q \beta(\mathbf{q},\pm 2) D^{(\ell)}_{m,\pm 2}(S(\hat{q})) \\
&\times \int_{t_1}^{t_0} dt P(t) \Psi(q,t) \left\{ \left[8\rho + 2\rho^2 \frac{\partial}{\partial\rho} \right] \frac{j_\ell(\rho)}{\rho^2} \right\}_{\rho=qr(t)}
\end{aligned}
\tag{7.4.62}
$$

We can now calculate the multipole coefficients, using the stochastic averages

$$
\langle \beta(\mathbf{q},\lambda) \beta^* (\mathbf{q}',\lambda') \rangle = \delta^3 (\mathbf{q}-\mathbf{q}') \, \delta_{\lambda\lambda'},
\tag{7.4.63}
$$

and the unitarity relation (7.3.17):

$$
\int d^2\hat{q} D^{(\ell)}_{m,\pm 2}(S(\hat{q})) D^{(\ell')}_{m',\pm 2}(S(\hat{q}))^* = \frac{4\pi}{2\ell+1} \delta_{mm'} \delta_{\ell\ell'}
$$

The sum over helicities ± 2 just gives a factor 2 in the EE and BB correlations, while as expected in the EB correlation there is a complete
cancelation, due to the $\pm$ sign in Eq. (7.4.60). The tensor mode contributions to the polarization multipole coefficients are then [20]

$$
\begin{aligned}
C^T_{EE,\ell} =& \pi^2 T_0^2 \int_0^\infty q^2 dq \\
&\times \left| \int_{t_1}^{t_0} dt P(t) \Psi(q,t) \left\{ \left[12 + 8\rho\frac{\partial}{\partial\rho} - \rho^2 + \rho^2 \frac{\partial^2}{\partial\rho^2} \right] \frac{j_\ell(\rho)}{\rho^2} \right\}_{\rho=qr(t)} \right|^2
\end{aligned}
\tag{7.4.64}
$$

$$
\begin{aligned}
C^T_{BB,\ell} =& \pi^2 T_0^2 \int_0^\infty q^2 dq \\
&\times \left| \int_{t_1}^{t_0} dt P(t) \Psi(q,t) \left\{ \left[8\rho + 2\rho^2 \frac{\partial}{\partial\rho} \right] \frac{j_\ell(\rho)}{\rho^2} \right\}_{\rho=qr(t)} \right|^2
\end{aligned}
\tag{7.4.65}
$$

The results (7.4.59) and (7.4.60), together with Eq. (7.3.14), allow us also to calculate the multipole coefficients in the correlation between temperature and polarization. In C^T_{TB} the $\pm$

sign in front of the integral over wave numbers in Eq. (7.4.60) produces a cancelation between the two terms in the sum over helicities ± 2, so that as expected $C^T_{TB} = 0$. On the other hand, in C^T_{TE} the sum over helicities ± 2 just gives a factor 2, so that Eqs. (7.3.14), (7.4.61) and (7.4.59) give

$$\begin{aligned} C^T_{TE,\ell} = & -2\pi^2 T_0^2 \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \int_0^\infty q^2 dq \\ & \times \int_{t_1}^{t_0} dt P(t) \Psi(q,t) \left\{ \left[12 + 8\rho \frac{\partial}{\partial \rho} - \rho^2 + \rho^2 \frac{\partial^2}{\partial \rho^2} \right] \frac{j_\ell(\rho)}{\rho^2} \right\}_{\rho = qr(t)} \\ & \times \int_{t_1}^{t_0} dt' d(q,t') \frac{j_\ell(qr(t'))}{q^2 r^2(t')} \end{aligned} \tag{7.4.66}$$

where

$$d(q,t) \equiv \left[\dot{\mathcal{D}}_q(t) - \frac{1}{2}\omega_c(t) \right] \exp\left(-\int_t^{t_0} \omega_c(t')\, dt' \right) \tag{7.4.67}$$

The calculation of the gravitational wave amplitude $\mathcal{D}_q(t)$ and source function $\Psi(q,t)$ is described in the previous section. For the same cosmological parameters as assumed there, Eqs. (7.4.62) and (7.4.63) give the EE

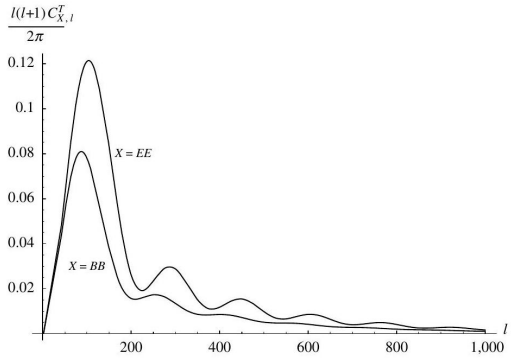


Figure 7.7: The multipole coefficients $\ell(\ell+1) C^T_{E E, \ell} / 2 \pi$ and $\ell(\ell+1)C^T_{BB,\ell}/2\pi$ in square microKelvin, vs. ℓ, for cosmological parameters given in Section 7.3, from ref. 21. The results agree to about 1% with those obtained (adopting the same cosmological parameters) by employing the computer programs CMBfast or CAMB, in which, instead of using the integral equation (7.3.22) as done here, the Boltzmann hierarchy is truncated at a maximum value of ℓ. (The results of these two computer programs also differ from each other by about the same amount.)

and BB multipole coefficients shown in Figure 7.7. (We will come to the TE coefficients below.)

The results (7.4.62)-(7.4.64) provide a basis for highly accurate computer calculations of the tensor multipole coefficients, but a casual inspection of these formulas does not provide much insight to the qualitative behavior of these coefficients as functions of ℓ. In particular, looking at Eqs. (7.4.62) and (7.4.63), we could hardly guess that $C^T_{BB,\ell} < C^T_{EE,\ell}$ for all $\ell > 15$, or that $C^T_{BB,\ell}$ and $C^T_{EE,\ell}$ approach each other for $\ell < 100$, as shown for one set of cosmological parameters in Figure 7.7.

[4720] These formulas, and also Eq. (7.4.64), are equivalent to those of Zaldarriaga and Seljak, ref. 1.

These results become much simpler and more transparent for large ℓ. Recall that for $\rho^2 - v^2 \gg v^{4/3}$ (where $v \equiv \ell + 1/2$), the spherical Bessel functions have the well-known asymptotic behavior [21]

$$j_\ell(\rho) \to \frac{\cos b}{v\sqrt{\sin b}} \cos[v(\tan b - b) - \pi/4] \tag{7.4.68}$$

where $\cos b \equiv v/\rho$, with $0 \leq b \leq \pi/2$. On the other hand, for $v^2 - \rho^2 \gg v^{4/3}$, $j_\ell(\rho)$ is exponentially small. In the range in which $|\rho^2 - v^2| < v^{4/3}$ neither approximation is valid, but $j_\ell(\rho)$ is a smooth function of ρ in this range, without the singularity at $\rho = v$ that might be suggested by the factor $1/\sqrt{\sin b}$ in Eq. (7.4.66). For $v \gg 1$ this range contributes only a small part of the range of integration, and we would expect to be able to use the approximation (7.4.66). For $\rho^2 - v^2 \gg v^{4/3} \gg 1$, the dominant contributions to derivatives of $j_\ell(\rho)/\rho^2$ come from terms in which the derivative acts only on the cosine in Eq. (7.4.66), so that

$$\left[12 + 8\rho\frac{\partial}{\partial\rho} - \rho^2 + \rho^2\frac{\partial^2}{\partial\rho^2}\right]\frac{j_\ell(\rho)}{\rho^2} \to -j_\ell(\rho) + j_\ell''(\rho)$$
$$\to -\frac{\left(1+\sin^2 b\right)\cos b}{v\sqrt{\sin b}} \cos[v(\tan b - b) - \pi/4] \tag{7.4.69}$$

$$\left[8\rho + 2\rho^2\frac{\partial}{\partial\rho}\right]\frac{j_\ell(\rho)}{\rho^2} \to 2j_\ell'(\rho) \to -\frac{2\sqrt{\sin b}\cos b}{v} \sin[v(\tan b - b) - \pi/4] \tag{7.4.70}$$

(Letting the derivatives act on $1/\sqrt{\sin b}$ would produce a non-integrable singularity at $b = 0$, but this is spurious, because the asymptotic formula (7.4.66) breaks down for ρ very near ν, where in fact there is no singularity.) Then Eqs. (7.4.62)-(7.4.64) become, for $v \equiv \ell + 1/2 \gg 1$, [22]

$$\begin{aligned} C^T_{EE,\ell} =& \frac{\pi^2 T_0^2}{v^2}\int_0^\infty q^2 dq \\ & \times \Big| \int_{r(t)>v/q} dt P(t)\Psi(q,t) \\ & \times \left\{ \frac{\left(1+\sin^2 b\right)\cos b}{\sqrt{\sin b}} \cos[v(\tan b - b) - \pi/4] \right\} \Big\}_{\cos b = v/qr(t)} \Big|^2 \\ C^T_{BB,\ell} =& \frac{\pi^2 T_0^2}{v^2}\int_0^\infty q^2 dq \Big| \int_{r(t)>v/q} dt P(t)\Psi(q,t) \\ & \times \left\{2\sqrt{\sin b}\cos b \sin[v(\tan b - b) - \pi/4]\right\}_{\cos b = v/qr(t)} \Big|^2 \end{aligned} \tag{7.4.71}$$

[4721] I. S. Gradshteyn & I. M. Ryzhik, Table of Integrals, Series, and Products, translated, corrected and enlarged by A. Jeffrey (Academic Press, New York, 1980): formula 8.453.1. The same approximation is used by J. R. Pritchard and M. Kamionkowski, Ann. Phys. 318, 2 (2005) [astro-ph/0412581], but their subsequent approximations are very different from those made here and in ref. 21.

[4722] R. Flauger and S. Weinberg, Phys. Rev. D 75, 123505 (2007) [astro-ph/0703179].

$$
\begin{aligned}
C^T_{TE,\ell} = &- \frac{2\pi^2 T_0^2}{v^2} \int_0^\infty q^2 dq \\
&\times \int_{r(t)>v/q} dt P(t) \Psi(q,t) \\
&\times \left\{ \frac{\left(1+\sin^2 b\right)\cos b}{\sqrt{\sin b}} \cos[v(\tan b - b) - \pi/4] \right\}_{\cos b = v/qr(t)} \\
&\times \int_{r(t')>v/q} dt' d\left(q,t'\right) \\
&\times \left\{ \frac{\cos^3 b'}{\sqrt{\sin b'}} \cos\left[v\left(\tan b' - b'\right) - \pi/4\right] \right\}_{\cos b' = v/qr(t')}
\end{aligned}
\tag{7.4.72}
$$

This approximate result for $C^T_{TE,\ell}$ is compared with the exact result (7.4.64) in Figure 7.8.

For $C^T_{EE,\ell}$ and $C^T_{BB,\ell}$, we can usefully make a further approximation. The quantity $b \equiv \cos^{-1}(v/qr(t))$ does not vary appreciably within the relatively narrow range of times t in which the last-scattering probability $P(t)$ is appreciable, so we can set $r(t)$ equal to $r_L \equiv r\left(t_L\right)$ everywhere except in the phase $v(\tan b - b)$, which for $v \gg 1$ does vary over a wide range in this time interval. Furthermore, because this phase varies over a wide range, the

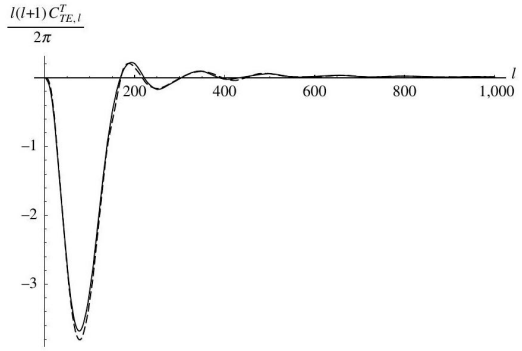


Figure 7.8: The multipole coefficient $\ell(\ell+1) C_{T E, \ell}^{T} / 2 \pi$ in square microKelvin, vs. ℓ, from ref. 21. The solid curve is taken from the exact formula (7.4.64); the dashed curve represents the approximation (7.4.71). In both cases, the method of calculating the gravitational wave amplitude and the source function, and the assumed cosmological parameters, are those described in Section 7.3.

difference between $\cos[\nu(\tan b - b) - \pi/4]$ and $\sin[\nu(\tan b - b) - \pi/4]$ is inconsequential, and we may as well replace both with $\cos[v(\tan b - b)]$. Then Eqs. (7.4.69) and (7.4.70) become [22]

$$
\begin{aligned}
\ell(\ell+1) C^T_{EE,\ell} \to \pi^2 T_0^2 &\int_{\nu/r_L}^\infty q^2 dq \left\{ \left(1+\sin^2 b_L\right)^2 \cos^2 b_L \right\}_{\cos b_L = v/qr_L} \\
&\times \left| \int_{r(t)>v/q} dt P(t) \Psi(q,t) \left\{ \frac{\cos[v(\tan b - b)]}{\sqrt{\sin b}} \right\}_{\cos b = v/qr(t)} \right|^2 ,
\end{aligned}
\tag{7.4.73}
$$

$$
\begin{aligned}
\ell(\ell+1) C^T_{BB,\ell} \to \pi^2 T_0^2 &\int_{\nu/r_L}^\infty q^2 dq \left\{ 4\sin^2 b_L \cos^2 b_L \right\}_{\cos b_L = v/qr_L} \\
&\times \left| \int_{r(t)>v/q} dt P(t) \Psi(q,t) \left\{ \frac{\cos[v(\tan b - b)]}{\sqrt{\sin b}} \right\}_{\cos b = v/qr(t)} \right|^2 .
\end{aligned}
\tag{7.4.74}
$$

(We have not set $b = b_L$ in the factors $1/\sqrt{\sin b}$ in both integrals over t, in order to avoid a divergence in the integration over q at $q = v/r_L$.) The results of using these approximate formulas are compared with the results of using the exact formulas (7.4.62) and (7.4.63) in Figures 7.9 and 7.10.

Our approximate result for $C^T_{BB,\ell}$ agrees with the exact result to about 1%, which is good enough for any practical purpose. The approximate result for $C^T_{EE,\ell}$ is not quite as good, agreeing with the exact result only to about 14%, but these approximations are evidently good enough to use them to draw qualitative conclusions. One immediate conclusion is that, since $\left(1 + \sin^2 b_L\right) \geq 4 \sin^2 b_L$ for all real b_L, we have $C^T_{EE,\ell} > C^T_{BB,\ell}$. Also, for ℓ small enough so that the wave number ℓ/r_L comes into the horizon before matter-radiation equality, say $\ell < 100$, for which $\Psi\left(\ell/r_L, t_L\right)$ is small, the integrals over q are dominated by values for which $\cos b$ is small, in which case $\left(1 + \sin^2 b\right)^2 \simeq 4 \sin^2 b$, and hence $C^T_{EE,\ell} \simeq C^T_{BB,\ell}$. As already noted, both properties are evident (for at least one set of cosmological parameters) in Figure 7.7.

The smallness of $C^T_{BB,\ell}$ makes it a difficult target for future observations, but the detection of the primordial BB mode would be of great importance to cosmology, as it would provide clear evidence of cosmological gravitational waves. The expectations for such tensor modes in theories of inflation are discussed in Chapter 10.

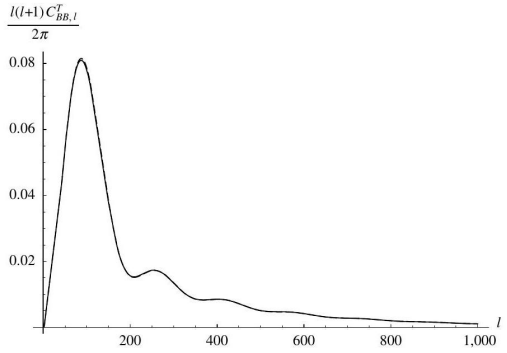


Figure 7.9: The multipole coefficient $\ell(\ell+1) C_{B B, \ell}^{T} / 2 \pi$ in square microKelvin, vs. ℓ, from ref. 22. The solid curve is taken from the exact formula (7.4.63); the dashed curve represents the approximation (7.4.70). In both cases, the method of calculating the gravitational wave amplitude and the source function, and the assumed cosmological parameters, are those described in Section 7.3.

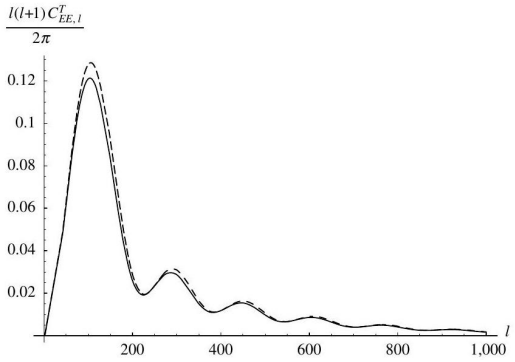


Figure 7.10: The multipole coefficient $\ell(\ell+1) C_{E E, \ell}^{T} / 2 \pi$ in square microKelvin, vs. ℓ, from ref. 22. The solid curve is taken from the exact formula (7.4.62); the dashed curve represents the approximation (7.4.69). In both cases, the method of calculating the gravitational wave amplitude and the source function, and the assumed cosmological parameters, are those described in Section 7.3.

The multipole coefficients $C_{EE,\ell}, C_{BB,\ell}$, and $C_{TE,\ell}$ may be measured by using data on correlations among the Stokes parameters and temperature fluctuations in comparison with general theoretical formulas for these
correlation functions, [23] analogous to the general formula (2.6.4) for the temperature correlation function. To derive these formulas, we start by noting from Eqs. (7.4.15) and (7.4.25) that

$$\begin{aligned} Q(\hat{n}) =& -\frac{1}{2}\sum_{\ell m} a_{E,\ell m}\left[\mathcal{Y}_{\ell}^{m}(\hat{n})+\mathcal{Y}_{\ell}^{-m}(\hat{n})^{*}\right] \\ & -\frac{i}{2}\sum_{\ell m} a_{B,\ell m}\left[\mathcal{Y}_{\ell}^{m}(\hat{n})-\mathcal{Y}_{\ell}^{-m}(\hat{n})^{*}\right] \end{aligned} \tag{7.4.75}$$

$$\begin{aligned} U(\hat{n}) =& \frac{i}{2}\sum_{\ell m} a_{E,\ell m}\left[\mathcal{Y}_{\ell}^{m}(\hat{n})-\mathcal{Y}_{\ell}^{-m}(\hat{n})^{*}\right] \\ & -\frac{1}{2}\sum_{\ell m} a_{B,\ell m}\left[\mathcal{Y}_{\ell}^{m}(\hat{n})+\mathcal{Y}_{\ell}^{-m}(\hat{n})^{*}\right] \end{aligned} \tag{7.4.76}$$

From Eqs. (7.4.28) and (7.4.29), we then find the correlation functions

$$\begin{aligned} \left\langle Q(\hat{n}) Q\left(\hat{n}'\right)\right\rangle =& \frac{1}{2}\operatorname{Re}\sum_{\ell} C_{EE,\ell}\left(F_{\ell}\left(\hat{n},\hat{n}'\right)+G_{\ell}\left(\hat{n},\hat{n}'\right)\right) \\ & +\frac{1}{2}\operatorname{Re}\sum_{\ell} C_{BB,\ell}\left(F_{\ell}\left(\hat{n},\hat{n}'\right)-G_{\ell}\left(\hat{n},\hat{n}'\right)\right), \end{aligned} \tag{7.4.77}$$

$$\begin{aligned} \left\langle U(\hat{n}) U\left(\hat{n}'\right)\right\rangle =& \frac{1}{2}\operatorname{Re}\sum_{\ell} C_{EE,\ell}\left(F_{\ell}\left(\hat{n},\hat{n}'\right)-G_{\ell}\left(\hat{n},\hat{n}'\right)\right) \\ & +\frac{1}{2}\operatorname{Re}\sum_{\ell} C_{BB,\ell}\left(F_{\ell}\left(\hat{n},\hat{n}'\right)+G_{\ell}\left(\hat{n},\hat{n}'\right)\right), \end{aligned} \tag{7.4.78}$$

$$\begin{aligned} \left\langle Q(\hat{n}) U\left(\hat{n}'\right)\right\rangle =& \frac{1}{2}\operatorname{Im}\sum_{\ell} C_{EE,\ell}\left(-F_{\ell}\left(\hat{n},\hat{n}'\right)+G_{\ell}\left(\hat{n},\hat{n}'\right)\right) \\ & +\frac{1}{2}\operatorname{Im}\sum_{\ell} C_{BB,\ell}\left(-F_{\ell}\left(\hat{n},\hat{n}'\right)-G_{\ell}\left(\hat{n},\hat{n}'\right)\right), \end{aligned} \tag{7.4.79}$$

where

$$F_\ell\left(\hat{n},\hat{n}'\right) \equiv \sum_m \mathcal{Y}_\ell^m(\hat{n})\mathcal{Y}_\ell^m\left(\hat{n}'\right)^* \tag{7.4.80}$$

$$G_\ell\left(\hat{n},\hat{n}'\right) \equiv \sum_m \mathcal{Y}_\ell^m(\hat{n})\mathcal{Y}_\ell^{-m}\left(\hat{n}'\right) \tag{7.4.81}$$

Also, using Eq. (2.6.2) (with $a_{\ell m}$ now written as $a_{T,\ell m}$) and Eqs. (7.4.27), (7.4.74) and (7.4.75), the correlation functions between temperature fluctuations and Stokes parameters are

$$\left\langle \Delta T(\hat{n})Q\left(\hat{n}'\right)\right\rangle = -\sum_\ell C_{TE,\ell}\,\mathrm{Re}\,H_\ell\left(\hat{n},\hat{n}'\right) \tag{7.4.82}$$

$$\left\langle \Delta T(\hat{n})U\left(\hat{n}'\right)\right\rangle = \sum_\ell C_{TE,\ell}\,\mathrm{Im}\,H_\ell\left(\hat{n},\hat{n}'\right) \tag{7.4.83}$$

where

$$H_\ell\left(\hat{n},\hat{n}'\right) \equiv \sum_m Y_\ell^m(\hat{n})\mathcal{Y}_\ell^m\left(\hat{n}'\right)^* \tag{7.4.84}$$

with Y_ℓ^m the ordinary spherical harmonic.
To calculate the functions $F_\ell\left(\hat{n},\hat{n}'\right)$ and $G_\ell\left(\hat{n},\hat{n}'\right)$, we note first (for instance, by inspection of the last line of Eq. (7.4.22)) that, for $\hat{n}$ in the three-direction $\hat{z}$ with $\theta = \phi = 0$,

$$\mathcal{Y}_\ell^m(\hat{z}) = \delta_{m,-2}\sqrt{\frac{2\ell+1}{4\pi}} \tag{7.4.85}$$

Then, using the definitions (7.4.16) and (7.4.11) and the transformation rule (7.4.18), we can express $\mathcal{Y}_\ell^m$ as an element of a unitary rotation matrix [24]

$$\mathcal{Y}_\ell^m(\hat{n}) = \sum_{m'} D^{(\ell)}_{m'm}\left(S^{-1}(\hat{n})\right)\mathcal{Y}_\ell^{m'}(\hat{z}) = \sqrt{\frac{2\ell+1}{4\pi}}D^{(\ell)}_{-2,m}\left(S^{-1}(\hat{n})\right) \tag{7.4.86}$$

where $S(\hat{n})$ is the rotation (7.4.12) that takes $\hat{z}$ into $\hat{n}$. Using Eq. (7.4.57), it follows then also that

$$\mathcal{Y}^{-m}(\hat{n}) = (-1)^\ell\mathcal{Y}^m(-\hat{n})^* \tag{7.4.87}$$

Hence, by using the group multiplication property $D^{(\ell)}(S)D^{(\ell)}\left(S'\right) = D^{(\ell)}\left(SS'\right)$ and the unitarity of $D^{(\ell)}(S)$, we obtain addition theorems [25] that give us the functions needed in our formulas (7.4.76)-(7.4.78) for the correlation functions

$$F_\ell\left(\hat{n},\hat{n}'\right) = \frac{2\ell+1}{4\pi}D^{(\ell)}_{-2,-2}\left(S^{-1}(\hat{n})S\left(\hat{n}'\right)\right), \tag{7.4.88}$$

$$G_\ell\left(\hat{n},\hat{n}'\right) = (-1)^\ell\frac{2\ell+1}{4\pi}D^{(\ell)}_{-2,-2}\left(S^{-1}(\hat{n})S\left(-\hat{n}'\right)\right). \tag{7.4.89}$$

[4723] General formulas for the correlation functions are presented by M. Zaldarriaga, Astrophys. J. 503, 1 (1998).

[4724] E. Newman and J. Penrose, ref. 3; J. N. Goldberg, A. J. MacFarlane, E. T. Newman, F. Rorlich, and E. C. G. Sudarshan, J. Math. Phys. 8, 2155 (1967); K. S. Thorne, Rev. Mod. Phys. 52, 299 (1980).

[25] General addition theorems for spin-weighted spherical harmonics are given by W. Hu and M. White, Phys. Rev. D 56, 596 (1997) [astro-ph/9702170].

To calculate the function $H_\ell(\hat{n},\hat{n}')$, we recall that the ordinary spherical harmonics can also be written as elements of a rotation matrix. From the formula $Y_\ell^m(\hat{z}) = \sqrt{(2\ell+1)/4\pi}\delta_{m0}$ for the case $\theta = \phi = 0$ and the transformation rule (7.4.18), we have

$$Y_\ell^m(\hat{n}) = \sqrt{\frac{2\ell+1}{4\pi}} D_{0,m}^{(\ell)}\left(S^{-1}(\hat{n})\right).$$

Hence, again using the group multiplication property of the $D^{(\ell)}$ matrices, we find

$$H_\ell(\hat{n},\hat{n}') = \frac{2\ell+1}{4\pi} D_{0,-2}^{(\ell)}\left(S^{-1}(\hat{n})S(\hat{n}')\right). \tag{7.4.90}$$

Note incidentally that

$$F_\ell(-\hat{n},-\hat{n}') = \sum_m \mathcal{Y}_\ell^{-m}(\hat{n})^*\mathcal{Y}_\ell^{-m}(\hat{n}') = F_\ell^*(\hat{n},\hat{n}'), \tag{7.4.91}$$

and

$$G_\ell(-\hat{n},-\hat{n}') = \sum_m \mathcal{Y}_\ell^{-m}(\hat{n})^*\mathcal{Y}_\ell^{m}(\hat{n}')^* = G_\ell^*(\hat{n},\hat{n}') \tag{7.4.92}$$

Also, using the property $Y_\ell^m(-\hat{n}) = (-1)^\ell Y_\ell^{-m}(\hat{n})^*$,

$$H_\ell(-\hat{n},-\hat{n}') = \sum_m Y_\ell^{-m}(\hat{n})^*\mathcal{Y}_\ell^{-m}(\hat{n}') = H_\ell^*(\hat{n},\hat{n}') \tag{7.4.93}$$

Hence the correlation functions have the reflection properties

$$\langle Q(\hat{n})Q(\hat{n}')\rangle = \langle Q(-\hat{n})Q(-\hat{n}')\rangle, \tag{7.4.94}$$
$$\langle U(\hat{n})U(\hat{n}')\rangle = \langle U(-\hat{n})U(-\hat{n}')\rangle, \tag{7.4.95}$$
$$\langle Q(\hat{n})U(\hat{n}')\rangle = -\langle Q(-\hat{n})U(-\hat{n}')\rangle, \tag{7.4.96}$$
$$\langle \Delta T(\hat{n})Q(\hat{n}')\rangle = \langle \Delta T(-\hat{n})Q(-\hat{n}')\rangle, \tag{7.4.97}$$
$$\langle \Delta T(\hat{n})U(\hat{n}')\rangle = -\langle \Delta T(-\hat{n})U(-\hat{n}')\rangle, \tag{7.4.98}$$

as expected from the space inversion properties of the Stokes parameters and temperature fluctuation and the assumed space-inversion invariance of the probability distribution over which we average.

It may be noted that, in measuring the correlation functions of the Stokes parameters observed in directions $\hat{n}$ and $\hat{n}'$, observers commonly choose the system of polar coordinates so that $\hat{n}$ and $\hat{n}'$ are on the same meridian; that is, so that $\phi = \phi'$. (This is usually expressed as the condition that the first of the two polarization vectors that are used to define the Stokes parameter should be aligned with the great circle between $\hat{n}$ and $\hat{n}$,[26] but we have defined these two polarization vectors always to lie in the directions of increasing θ and ϕ, so it amounts to the same thing.) This has the advantage that the functions $F_\ell(\hat{n},\hat{n}'), G_\ell(\hat{n},\hat{n}')$ and $H_\ell(\hat{n},\hat{n}')$ are then all real, and depend only on $\theta - \theta'$. To see this, note that by using the representation

$$D^{(\ell)}(S(\hat{n})) = \exp\left(-i\phi J_3^{(\ell)}\right)\exp\left(-i\theta J_2^{(\ell)}\right)$$

[4726] Kamionkowski, Kosowsky, and Stebbins, ref. 1.

of the standard rotation from $\hat{z}$ to $\hat{n}$ in terms of the angular momentum matrices $J_i^{(\ell)}$, we can rewrite Eqs. (7.4.87)-(7.4.89) as

$$\begin{aligned}
F_\ell(\hat{n},\hat{n}') &= \frac{2\ell+1}{4\pi}\left[\exp\left(i\theta J_2^{(\ell)}\right)\exp\left(i\phi J_3^{(\ell)}\right)\right. \\
&\quad \left.\times\exp\left(-i\phi' J_3^{(\ell)}\right)\exp\left(-i\theta' J_2^{(\ell)}\right)\right]_{-2,-2} \\
G_\ell(\hat{n},\hat{n}') &= (-1)^\ell\frac{2\ell+1}{4\pi}\left[\exp\left(i\theta J_2^{(\ell)}\right)\exp\left(i\phi J_3^{(\ell)}\right)\right. \\
&\qquad \left.\times\exp\left(-i\left[\phi'+\pi\right]J_3^{(\ell)}\right)\exp\left(-i\left[\pi-\theta'\right]J_2^{(\ell)}\right)\right]_{-2,-2} \\
H_\ell(\hat{n},\hat{n}') &= \frac{2\ell+1}{4\pi}\left[\exp\left(i\theta J_2^{(\ell)}\right)\exp\left(i\phi J_3^{(\ell)}\right)\right. \\
&\quad \left.\times\exp\left(-i\phi' J_3^{(\ell)}\right)\exp\left(-i\theta' J_2^{(\ell)}\right)\right]_{0,-2}
\end{aligned}$$

If $\phi=\phi'$, then

$$F_\ell(\hat{n},\hat{n}') = \frac{2\ell+1}{4\pi}\left[\exp\left(i\left(\theta-\theta'\right)J_2^{(\ell)}\right)\right]_{-2,-2},$$

so $F_\ell(\hat{n},\hat{n}')$ for $\phi=\phi'$ depends only on $\theta-\theta'$. Also, $iJ_2^{(\ell)}$ is a real matrix, so $F_\ell(\hat{n},\hat{n}')$ for $\phi=\phi'$ is also real. As to G_ℓ, for $\phi=\phi'$ we have

$$\begin{aligned}
G_\ell(\hat{n},\hat{n}') =& (-1)^\ell\frac{2\ell+1}{4\pi} \\
&\times\left[\exp\left(i\theta J_2^{(\ell)}\right)\exp\left(-i\pi J_3^{(\ell)}\right)\exp\left(-i\left[\pi-\theta'\right]J_2^{(\ell)}\right)\right]_{-2,-2} \\
=& (-1)^\ell\frac{2\ell+1}{4\pi} \\
&\times\left[\exp\left(i\theta J_2^{(\ell)}\right)\exp\left(+i\left[\pi-\theta'\right]J_2^{(\ell)}\right)\exp\left(-i\pi J_3^{(\ell)}\right)\right]_{-2,-2} \\
=& (-1)^\ell\frac{2\ell+1}{4\pi}\left[\exp\left(i\left(\theta-\theta'+\pi\right)J_2^{(\ell)}\right)\right]_{-2,-2}
\end{aligned}$$

so $G_\ell(\hat{n},\hat{n}')$ for $\phi=\phi'$ is real because $iJ_2^{(\ell)}$ is a real matrix, and depends only on $\theta-\theta'$. Finally, $H_\ell(\hat{n},\hat{n}')$ is real and depends only on $\theta-\theta'$, for the same reasons as for $F_\ell(\hat{n},\hat{n}')$. We conclude then that if the coordinate system is chosen so that $\phi=\phi'$, the correlation functions $\langle Q(\hat{n})Q(\hat{n}')\rangle$ and $\langle U(\hat{n})U(\hat{n}')\rangle$ and $\langle\Delta T(\hat{n})Q(\hat{n}')\rangle$ depend only on the angle $\theta-\theta'$ between $\hat{n}$ and $\hat{n}'$, while $\langle Q(\hat{n})U(\hat{n}')\rangle$ and $\langle\Delta T(\hat{n})U(\hat{n}')\rangle$ vanish. Also, $\langle Q(\hat{n})Q(\hat{n}')\rangle$ and $\langle U(\hat{n})U(\hat{n}')\rangle$ are obviously symmetric between $\hat{n}$ and $\hat{n}'$, so for $\phi=\phi'$ they actually depend only on $|\theta-\theta'|$; that is, on the angle between $\hat{n}$ and $\hat{n}'$.

10.8 The Growth of Structure

In Chapter 6 we followed the evolution of small perturbations through the radiation and matter dominated eras, up to the time of decoupling, when radiation no longer interacted effectively with matter. Now we will continue the story past the time of decoupling. In Section 8.1 we will follow the perturbations while they remained small, continuing our linear analysis. This era is increasingly becoming accessible to observation, as studies of the cosmic matter distribution are pushed to larger redshifts. As we shall see, data on the distribution of matter fluctuations already provides an important extension of results from cosmic microwave background anisotropies to smaller wavelengths, and it is hoped that eventually it may provide information of the effect of dark energy on cosmic expansion.

Of course, eventually the perturbations in the matter density became strong enough for the linear approximation to break down, as shown vividly by the existence of stars and galaxies and galaxy clusters. It is believed that these structures were formed in a two-step process. [1] First, in regions where the density was a little larger than average, the cold dark matter and baryonic matter together expanded more slowly than the universe as a whole, eventually reaching a minimum density and then recontracting. This scenario is discussed in Section 8.2. If an overdense region was sufficiently large then as shown in Section 8.3 its baryonic matter collapsed along with its cold dark matter. Then in a second stage, after this collapse, the baryonic matter lost its energy through radiative cooling, and it condensed into protogalaxies consisting of clouds of gas that eventually form stars. The cold dark matter particles could not lose their energy through radiative cooling, so they remained in large more-or-less spherical halos around these galaxies. We will not attempt a proper treatment of this second stage, which involves complications of astrophysics and mathematics that deserve a treatise to themselves.

10.8.1 8.1 Linear perturbations after recombination

After the disappearance of almost all free electrons the baryonic plasma decoupled from photons, and behaved like just another form of cold dark matter. (Effects of pressure at small wavelengths are discussed in

Section 8.3.) With the gravitational field perturbations dominated by fluctuations in the total density of cold matter, the Newtonian treatment of Appendix F is applicable, and tells us that the fractional density perturbation eventually grew as $a \propto t^{2/3}$. However, we need a relativistic analysis to connect the constant wave-number dependent factor in this growth with the strength of primordial fluctuations, characterized by the quantity $\mathcal{R}_q^0$, to describe certain small but interesting oscillations in the density fluctuations arising from the interaction of baryons and radiation before decoupling, and to carry this analysis forward to near the present, when vacuum energy became important.

In our relativistic analysis baryons and cold dark matter are treated somewhat differently, because we continue to use a synchronous gauge in which the cold dark matter velocity potential but not the baryon velocity potential vanishes. After decoupling the baryon velocity potential was no longer locked to the photon velocity potential by Thomson scattering, so to obtain the equation for baryon conservation, we may use Eq. (5.3.34) with $\bar{p}_B, \delta p_{Bq}$, and π^S_{Bq} all vanishing (because baryons move slowly), and find

$$\dot{\delta\rho}_{Bq} + 3H\delta\rho_{Bq} - \frac{q^2}{a^2}\bar{\rho}_B\delta u_{Bq} = -\bar{\rho}_B\psi_q$$

or, dividing by $\bar{\rho}_B \propto a^{-3}$,

$$\dot{\delta}_{Bq} - \left(q^2/a^2\right)\delta u_{Bq} = -\psi_q \tag{8.1.1}$$

where as before, $\delta_{Bq} \equiv \delta\rho_{Bq}/\bar{\rho}_B$. For cold dark matter there is no velocity potential, so we again have Eq. (6.2.11) for cold dark matter conservation:

$$\dot{\delta}_{Dq} = -\psi_q \tag{8.1.2}$$

It is only our choice of gauge that makes the velocity potential of the dark matter rather than the baryons vanish, so it is best to think of δu_{Bq} as the relative velocity potential of the baryons and dark matter.

The equation for baryon momentum conservation is given by Eq. (5.3.32) with zero pressure and anisotropic inertia,

[471] S. D. M. White and M. J. Rees, Mon. Not. Roy. Astron. Soc. 183, 341 (1978).

$$\partial_0 \left(\bar{\rho}_B \delta u_{Bq}\right) + 3H\bar{\rho}_B \delta u_{Bq} = 0$$

so, dividing by $\bar{\rho}_B \propto a^{-3}$, we have simply:

$$\delta \dot{u}_{Bq} = 0 \tag{8.1.3}$$

(As we will see in Section 8.3, our neglect of the baryon pressure after recombination is justified except for the smallest wavelengths.) Finally, if we neglect the contribution of photon and neutrino density fluctuations to the gravitational field perturbation, then the gravitational field equation (5.3.36) becomes

$$\frac{d}{dt}\left(a^2\psi_q\right) = -4\pi G a^2 \bar{\rho}_M \left((1-\beta)\delta_{Dq} + \beta\delta_{Bq}\right) \tag{8.1.4}$$

where β is again the constant

$$\beta \equiv \frac{\bar{\rho}_B}{\bar{\rho}_M} = \frac{\Omega_B}{\Omega_M} \approx 1/6 \tag{8.1.5}$$

We are keeping terms of order β for the present, because as we shall see they lead to effects that although small are quite distinctive.

To derive the initial conditions for these equations, we note that at times sufficiently early (say, $z > 1$), at which vacuum energy as well as curvature made a negligible contribution to the expansion rate, we have $a \propto t^{2/3}$, and $4\pi G\bar{\rho}_M = 2/3t^2$, so Eq. (8.1.4) gives

$$\frac{d}{dt}\left(t^{4/3}\psi_q\right) = -\frac{2}{3}t^{-2/3}\left((1-\beta)\delta_{Dq} + \beta\delta_{Bq}\right) \tag{8.1.6}$$

The reader can easily check that the general solution of Eqs. (8.1.1)-(8.1.3) and (8.1.6) is

$$\psi_q = a_q t^{-1/3} + b_q t^{-2} + \beta\delta u_{Bq} q^2/a^2 \tag{8.1.7}$$

$$\delta_{Bq} = -\frac{3}{2}a_q t^{2/3} + b_q t^{-1} + (1-\beta)c_q - 3(1-\beta)t\left(q^2/a^2\right)\delta u_{Bq} \tag{8.1.8}$$

$$\delta_{Dq} = -\frac{3}{2}a_q t^{2/3} + b_q t^{-1} - \beta c_q + 3\beta t\left(q^2/a^2\right)\delta u_{Bq} \tag{8.1.9}$$

where a_q, b_q, c_q, and δu_{Bq} are constants that must be found by matching these solutions to the values of $\psi_q, \delta_{Bq}, \delta_{Dq}$, and δu_{Bq} at the time of decoupling of matter and radiation. For late times, we need only keep the leading terms in this solution, with coefficient a_q; the other terms are suppressed relative to these by factors $t^{-5/3}, t^{-2/3}$, and t^{-1}. To calculate the coefficient a_q, we note that $\beta\delta_{Bq} + (1-\beta)\delta_{Dq} - t\psi_q + \beta t\left(q^2/a^2\right)\delta u_{Bq} = -5a_q t^{2/3}/2$. Setting t here equal to the time t_L of decoupling [2] (when to a good approximation

$\delta_{Bq} = \delta_{\gamma q}$ and $\delta u_{Bq} = \delta u_{\gamma q}$) then gives for $t \gg t_L$

[472] The subscript L stands for "last scattering," and indicates the time of the decoupling of radiation from matter, associated with the recombination of hydrogen. (To use a subscript R or D for this time might produce confusion with the subscripts R and D that we continue to use to denote radiation and dark matter.) Of course, this is not the moment of last scattering; some photons of the microwave background were scattered again when hydrogen became reionized at much later times. Strictly speaking, we should take t_L here as the time during recombination when a typical electron stops exchanging appreciable momentum with the photons, rather than the slightly earlier time when a typical photon stops exchanging appreciable momentum with the electrons. Because R_L is not very different from unity, there is little difference between these times.

$$\begin{aligned}\delta_{Dq}(t) \rightarrow & \delta_{Bq}(t) \rightarrow -\frac{3}{2} a_q t^{2/3} \\ & \rightarrow \frac{3}{5}\left(\frac{t}{t_L}\right)^{2/3}\left[\beta\delta_{\gamma q}(t_L) + (1-\beta)\delta_{Dq}(t_L) - t_L\psi_q(t_L)\right. \\ & \left. + \beta t_L\left(q^2/a_L^2\right)\delta u_{\gamma q}(t_L)\right]. \end{aligned} \tag{8.1.10}$$

Note that, even though the fractional density perturbations of the baryons and dark matter were quite different at decoupling, they approached each other at late times thereafter, an assumption we made in analyzing the observations of X-rays from clusters of galaxies in Section 1.9. It follows that the fractional fluctuation δ_{Mq} in the total mass density approaches

$$\delta_{Mq} \equiv \frac{\bar{\rho}_D\delta_{Dq} + \bar{\rho}_B\delta_{Bq}}{\bar{\rho}_D + \bar{\rho}_B} \rightarrow \delta_{Dq} \rightarrow \delta_{Bq} \tag{8.1.11}$$

Since these fractional density perturbations all eventually became equal, we will concentrate from now on $\delta_{Mq}(t)$. We now want to carry our calculation forward to the present, when dark energy may no longer be neglected. According to Eqs. (8.1.1)-(8.1.4), under the approximation of neglecting fluctuations in the photon and neutrino energy densities, $\delta_{Mq}(t)$ satisfies the second-order differential equation

$$\frac{d}{dt}\left[a^2\frac{d}{dt}\delta_{Mq}\right] = 4\pi G a^2\bar{\rho}_M\delta_{Mq} \tag{8.1.12}$$

The most important consequence of this equation, together with Eq. (8.1.10), is that whatever we assume about dark energy, well after recombination when the terms in $\delta_{Mq}(t)$ that decay as $1/t$ have died away, the dependence of $\delta_{Mq}(t)$ on q and t factorizes:

$$\delta_{Mq}(t) = \Delta(q)F(t) \tag{8.1.13}$$

where

$$\begin{aligned}\Delta(q) = & \beta\delta_{\gamma q}(t_L) + (1-\beta)\delta_{Dq}(t_L) \\ & - t_L\psi_q(t_L) + \beta t_L\left(q^2/a_L^2\right)\delta u_{\gamma q}(t_L), \end{aligned} \tag{8.1.14}$$

and $F(t)$ satisfies the differential equation

$$\frac{d}{dt}\left[a^2\frac{d}{dt}F\right] = 4\pi G a^2\bar{\rho}_M F \tag{8.1.15}$$

with the initial condition, that well after recombination, until dark energy becomes important,

$$F(t) \rightarrow \frac{3}{5}\left(\frac{t}{t_L}\right)^{2/3} \tag{8.1.16}$$

For instance, if we assume that the vacuum energy is constant, then the Friedmann equation for $a(t)$ reads

$$\frac{\dot{a}}{a} = \sqrt{\frac{8\pi G}{3}\left(\rho_\Lambda + \bar{\rho}_M\right)} = H_0\sqrt{\Omega_\Lambda}\sqrt{1+1/x}, \tag{8.1.17}$$

where

$$x \equiv \frac{\rho_\Lambda}{\bar{\rho}_M} = \frac{\Omega_\Lambda}{\Omega_M}\left(\frac{a}{a_0}\right)^3 \tag{8.1.18}$$

Using x instead of t as the independent variable allows us to put Eq. (8.1.15) in a parameter-free form

$$\sqrt{x(1+x)}\frac{d}{dx}\left(x^{7/6}\sqrt{1+x}\frac{dF}{dx}\right) = \frac{1}{6x^{1/3}}F. \tag{8.1.19}$$

The growing solution [3] that becomes proportional to $t^{2/3}$ for $x \ll 1$ is

$$F \propto \sqrt{\frac{1+x}{x}}\int_0^x \frac{du}{u^{1/6}(1+u)^{3/2}}$$

Eq. (8.1.16) requires that for $x \ll 1$

$$F \to \frac{3}{5}\left(\frac{a}{a_L}\right) = \frac{3}{5}(1+z_L)\left(\frac{x\Omega_M}{\Omega_\Lambda}\right)^{1/3}$$

so we can write

$$F(t) = \frac{3}{5}\left(\frac{a(t)}{a_L}\right) C\left(\frac{\Omega_\Lambda}{\Omega_M}\left(\frac{a(t)}{a_0}\right)^3\right) \tag{8.1.20}$$

where $C(x)$ is a correction factor, normalized so that $C(0) = 1$:

$$C(x) \equiv \frac{5}{6}x^{-5/6}\sqrt{1+x}\int_0^x \frac{du}{u^{1/6}(1+u)^{3/2}} \tag{8.1.21}$$

Numerical values of $C(x)$ are given in Table 8.1.

Table 8.1: Values of the function $C(x)$, giving the suppression of the growth of matter fluctuations by dark energy as a function of $x \equiv (\Omega_\Lambda / \Omega_M)(a / a_0)^3$.

x	$C(x)$	x	$C(x)$
0	1	1.0	0.8725
0.1	0.9826	1.5	0.8314
0.2	0.9667	2.0	0.7981
0.3	0.9520	2.5	0.7702
0.5	0.9256	3.0	0.7462
0.7	0.9025	3.5	0.7254

We see that at all scales and times, dark energy suppresses the growth of density fluctuations.

It is $\Delta(q)$ that contains information about conditions at decoupling. In place of $\Delta(q)$, it is conventional to introduce a power spectral function $P(k)$, defined as a function of the present value $k \equiv q/a_0$ of the physical wave number, by [4]

$$P(k) \equiv (2\pi)^3 a_0^3 F^2(t_0)\left|\Delta(a_0 k)\right|^2. \tag{8.1.22}$$

Most surveys of large scale structure report their results in terms of $P(k)$.

Now we must consider how these surveys are used to measure $\Delta(q)$ or $P(k)$. We recall that for a

[473] H. Martel, Astrophys. J. 377, 7 (1991).

single dominant scalar mode (and assuming spatial flatness), the fractional density perturbation in coordinate space takes the form of a Fourier transform like Eq. (5.2.1):

$$\delta_M(\mathbf{x},t) = \int d^3q\alpha(\mathbf{q})\delta_{Mq}(t)e^{i\mathbf{q}\cdot\mathbf{x}} = F(t)\int d^3q\alpha(\mathbf{q})\Delta(q)e^{i\mathbf{q}\cdot\mathbf{x}} \tag{8.1.23}$$

where $\alpha(\mathbf{q})$ is a stochastic variable, normalized so that

$$\langle\alpha(\mathbf{q})\alpha^*(\mathbf{q}')\rangle = \delta^3(\mathbf{q}-\mathbf{q}')\,. \tag{8.1.24}$$

The quantity $\Delta(q)$ can be found either from measurements of the correlation of matter density perturbations at different points, or more directly, from an angular average of the square of a Fourier integral of the matter density perturbation over the survey volume.

According to the ergodic theorem discussed in Appendix D, as long as the survey volume V is large compared with the volume over which density fluctuations are correlated, the ensemble averages of products of density fluctuations can be found from an average of these products over the survey volume, In particular, the two-point correlation function can be found from

$$\langle\delta_M(\mathbf{x},t)\delta_M(\mathbf{y},t)\rangle = \frac{1}{V}\int_V d^3z\delta_M(\mathbf{x}+\mathbf{z},t)\delta_M(\mathbf{y}+\mathbf{z},t) \tag{8.1.25}$$

Of course, we measure density fluctuations as functions of redshifts and angular positions rather than of three-dimensional positions and times, so the correlation function actually measured is the correlation of the fractional matter density perturbation observed at a redshift z and direction $\hat{n}$ with the perturbation observed at a redshift z' and direction $\hat{n}'$:

$$\xi(z,\hat{n};z',\hat{n}') \equiv \langle\delta_M(r(z)\hat{n},t(z))\delta_M(r(z')\hat{n}',t(z'))\rangle \tag{8.1.26}$$

where $r(z)$ and $t(z)$ are the Robertson-Walker co-moving radial coordinate and emission time associated with a redshift z. Using Eqs. (8.1.23), (8.1.24), and (8.1.13), this is

$$\begin{aligned}\xi(z,\hat{n};z',\hat{n}') &= \int d^3q\delta_{Mq}(t(z))\delta^*_{Mq}(t(z'))\exp(i\mathbf{q}\cdot(r(z)\hat{n}-r(z')\hat{n}'))\\ &= F(t(z))F(t(z'))\int d^3q|\Delta(q)|^2\exp(i\mathbf{q}\cdot(r(z)\hat{n}-r(z')\hat{n}'))\end{aligned} \tag{8.1.27}$$

In terms of $P(k)$, this reads

$$\begin{aligned}\xi(z,\hat{n};z',\hat{n}') =& \frac{1}{2\pi^2}\frac{F(t(z))F(t(z'))}{F^2(t_0)}\int_0^\infty kP(k)dk\\ &\times\frac{\sin(k|d_S(z)\hat{n}-d_S(z')\hat{n}'|)}{|d_S(z)\hat{n}-d_S(z')\hat{n}'|}\end{aligned} \tag{8.1.28}$$

where $d_S(z)$ is a convenient structure distance, related to $r(z)$ and to the angular diameter and luminosity distances at a redshift z by:

$$d_S(z) \equiv a_0 r(z) = (1+z)d_A(z) = (1+z)^{-1}d_L(z)$$

[474] The factor $(2\pi)^3$ is inserted here because the position-space perturbations are usually written as Fourier transforms with an extra factor of $(2\pi)^{-3/2}$, and the factor a_0^3 is included because these threedimensional Fourier transforms are usually written as integrals over the present value of the physical wave number, $k \equiv q/a_0$, rather than the co-moving wave number q. If the reader wishes, the co-moving coordinates could be normalized so that $a_0 = 1$, in which case all q s in this section could be replaced with k s.

Eq. (8.1.28) can be used to measure the power spectral function if we know $d_S(z)$, or to measure $d_S(z)$ if we know something about features in the power spectral function. In particular, for observations at relatively low redshift,
we can use $d_S(z) \simeq z/H_0$ and $t(z) \simeq t_0$, in which case Eqs. (8.1.28) gives the correlation function as

$$\xi\left(z,\hat{n};z',\hat{n}'\right) = \frac{H_0}{2\pi^2}\int_0^\infty kP(k)dk\frac{\sin\left(\left(k/H_0\right)\left|z\hat{n}-z'\hat{n}'\right|\right)}{\left|z\hat{n}-z'\hat{n}'\right|} \tag{8.1.29}$$

Measurements of the shape of this correlation function can evidently tell us about the dependence of the power spectral function on k/H_0, rather than k itself.

It is more common in large surveys to measure $P(k)$ from the angular average of the square of a Fourier integral of the matter density perturbation over the survey volume. Define a Fourier transform as an integral over the co-moving survey volume V :

$$\delta^V_{M\mathbf{Q}}(t) \equiv \frac{1}{\sqrt{V}(2\pi)^{3/2}}\int_V d^3x e^{-i\mathbf{Q}\cdot\mathbf{x}}\delta_M(\mathbf{x},t) \tag{8.1.30}$$

(The reason for this normalization will soon be made clear.) Using Eq. (8.1.23), this is

$$\delta^V_{M\mathbf{Q}}(t) = \int d^3q\alpha(\mathbf{q})F_V(\mathbf{q}-\mathbf{Q})\delta_{Mq}(t), \tag{8.1.31}$$

where

$$F_V(\mathbf{q}) \equiv \frac{1}{\sqrt{V}(2\pi)^{3/2}}\int_V d^3x e^{i\mathbf{q}\cdot\mathbf{x}} \tag{8.1.32}$$

It is plausible, and will be shown formally at the end of this section, that as long as the co-moving survey volume V contains many co-moving wavelengths $2\pi/q$, the angular average of $\left|\delta^V_{M\mathbf{Q}}\right|^2$ is the same as its ensemble average

$$\frac{1}{4\pi}\int d^2\hat{Q}\left|\delta^V_{M\mathbf{Q}}(t)\right|^2 = \left\langle\frac{1}{4\pi}\int d^2\hat{Q}\left|\delta^V_{M\mathbf{Q}}(t)\right|^2\right\rangle \tag{8.1.33}$$

Using Eqs. (8.1.31) and (8.1.24), this is

$$\frac{1}{4\pi}\int d^2\hat{Q}\left|\delta^V_{M\mathbf{Q}}(t)\right|^2 = \frac{1}{4\pi}\int d^2\hat{Q}\int d^3q\left|F_V(\mathbf{q}-\mathbf{Q})\right|^2\left|\delta_{Mq}(t)\right|^2 \tag{8.1.34}$$

Now, for large $V, F_V(\mathbf{q}-\mathbf{Q})$ approaches $\left((2\pi)^{3/2}/\sqrt{V}\right)\delta^3(\mathbf{q}-\mathbf{Q})$, so

$$\left|F_V(\mathbf{q}-\mathbf{Q})\right|^2 \to \frac{(2\pi)^{3/2}}{\sqrt{V}}\delta^3(\mathbf{q}-\mathbf{Q})F_V(0) = \delta^3(\mathbf{q}-\mathbf{Q}) \tag{8.1.35}$$

and therefore

$$\frac{1}{4\pi}\int d^2\hat{Q}\left|\delta^V_{M\mathbf{Q}}(t)\right|^2 = \left|\delta_{MQ}(t)\right|^2 \tag{8.1.36}$$

where $Q \equiv |\mathbf{Q}|$.
Again, we measure density fluctuations as functions of direction and redshift rather than position and time. Using Eq. (8.1.30) in Eq. (8.1.36) and expressing the result in terms of the power spectral function (8.1.22), we have

$$P(k) = \frac{1}{4\pi\mathcal{V}} \int d^2\hat{Q} \left| \int_{\mathcal{V}} d^2\hat{n} dz d_S^2(z) \right.$$
$$\left. \times\, d_S'(z) \frac{F\,(t_0)}{F(t(z))} \delta_M(r(z)\hat{n}, t(z)) e^{-i\hat{n}\cdot\hat{Q}kd_S(z)} \right|^2 \tag{8.1.37}$$

where $\mathcal{V} \equiv a_0^3 V$ is the physical survey volume. As in the case of Eq. (8.1.28) for the correlation function, we can use this formula either to calculate $P(k)$ from data for surveys with relatively low redshifts, or to find $d_S(z)$ from larger redshift surveys combined with information about $P(k)$ from other sources.

Let's now consider the calculation of $\Delta(q)$ and the power spectral function. We begin by neglecting all terms of order β in Eq. (8.1.14), so that

$$\Delta(q) = \delta_{Dq}\,(t_L) - t_L \psi_q\,(t_L) \tag{8.1.38}$$

We will return at the end of this section to the very interesting effects associated with the small baryon density. Using Eqs. (6.5.5) and (6.5.6), this is

$$\Delta(q) = \frac{3q^2 t_L^2 \mathcal{R}_q^o \mathcal{T}(\kappa)}{2a_L^2} = \frac{2q^2 \mathcal{R}_q^o \mathcal{T}(\kappa)}{3H_L^2 a_L^2} \tag{8.1.39}$$

where $\mathcal{T}$ is the dimensionless transfer function given in Table 6.1, and $\kappa/\sqrt{2}$ is the ratio of the wave number q to the wave number $q_{\rm EQ}$ that comes into the horizon just at matter-radiation equality. (The second expression is derived using $t_L = 2/3H_L$, where $H_L = \sqrt{\Omega_M} H_0\,(1+z_L)^{3/2}$ is the Hubble rate at decoupling if radiation is neglected. This is more accurate than using the actual age of the universe at decoupling, because Eqs. (8.1.38) and (8.1.39) were derived using a definition of the zero of time for which $a \propto t^{2/3}$ during the matter-dominated era.) Using Eq. (8.1.20), the power spectral function (8.1.22) is then

$$P(k) = \frac{4(2\pi)^3 a_0^3 C^2\,(\Omega_\Lambda/\Omega_M)}{25\Omega_M^2 H_0^4} \mathcal{R}_{ka_0}^o k^4 \mathcal{T}^2\left(\sqrt{2}k/k_{\rm EQ}\right), \tag{8.1.40}$$

where, according to Eq. (6.4.58),

$$k_{\rm EQ} \equiv q_{\rm EQ}/a_0 = \sqrt{2} H_0 \Omega_M / \sqrt{\Omega_R} = [13.6{\rm Mpc}]^{-1} \Omega_M h^2. \tag{8.1.41}$$

In particular, if we suppose that $\mathcal{R}_q \simeq N q^{-3/2}\,(q/q_*)^{(n_S-1)/2}$ with some spectral index n_S and constant N (which for $n_S \neq 1$ depends on the arbitrary choice of the reference wave number q_*), we would have

$$P(k) = \frac{4(2\pi)^3 N^2 C^2\,(\Omega_\Lambda/\Omega_M)}{25\Omega_M^2 H_0^4 k_*^{n_S-1}} k^{n_S} \mathcal{T}^2\left(\sqrt{2}k/k_{\rm EQ}\right), \tag{8.1.42}$$

where $k_* \equiv q_*/a_0$. The shape of this function for $n_S = 1$ is shown in Figure 8.1.

Knowledge of the power spectral function allows us to calculate the mean square value σ^2 of the fractional density fluctuation:

$$\sigma^2(z) = \xi(z,\hat{n};z,\hat{n}) = \frac{1}{2\pi^2}\left(\frac{F(t(z))}{F\,(t_0)}\right)^2 \int_0^\infty P(k) k^2 dk \tag{8.1.43}$$

With $P(k)$ given by Eq. (8.1.42), for any plausible n_S this integral is convergent at $k = 0$, where $\mathcal{T}(\kappa) \to 1$, but for $n_S \geq 1$ it diverges at $k \to \infty$ like $\int k^{n_S-2} \ln^2 k dk$. (Recall that, as we saw in Section 6.5, $\mathcal{T}(\kappa) \propto \ln k/k^2$ for $k \to \infty$. On the other hand, the sine factor in the integrand in Eq. (8.1.27) makes $\xi\,(z,\hat{n};z',\hat{n}')$ finite for $z' \neq z$ or $\hat{n}' \neq \hat{n}$.)

In order to avoid the ultraviolet divergence in σ^2, it is common instead to express the intensity of the primordial fluctuations in terms of the mean square value σ_R^2 of the average of the fractional density perturbation over a

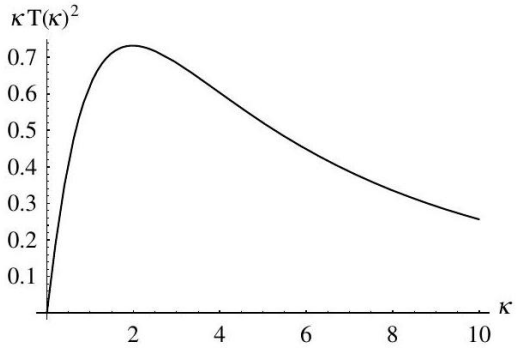


Figure 8.1: Shape of the power spectral function. This plot gives $\kappa \mathcal{T}^{2}(\kappa)$ as a function of κ, where $\mathcal{T}(\kappa)$ is the transfer function discussed in Section 6.5, $\kappa = \sqrt{2}k/k_{\rm EQ}$, and $k_{\rm EQ} = \left[13.6{\rm Mpc}^{-1}\Omega_M h^2\right.$ is the wave number that just comes into the horizon at radiationmatter equality. Eq. (8.1.42) shows that for $n_S = 1$, the power spectral function $P(k)$ is proportional to $\kappa\mathcal{T}^2(\kappa)$.

sphere of co-moving radius R/a_0 :

$$\sigma_R^2(z) \equiv \left\langle \left(\frac{3a_0^3}{4\pi R^3} \int_{a_0|\mathbf{x}|<R} d^3x \delta_M(\mathbf{x}, t(z)) \right)^2 \right\rangle \tag{8.1.44}$$

(Because of the translation invariance of the average, it would make no difference if we wrote the argument of δ_M as $\mathbf{x}+\mathbf{y}$, with $\mathbf{y}$ any fixed coordinate vector.) Again using Eqs. (8.1.22) and (8.1.23), and now also Eq. (8.1.20), we find

$$\sigma_R^2(z) = \frac{1}{2\pi^2(1+z)^2} \left(\frac{C\left(\Omega_\Lambda/\Omega_M(1+z)^3\right)}{C\left(\Omega_\Lambda/\Omega_M\right)} \right)^2 \int_0^\infty P(k)|f(kR)|^2 k^2 dk \tag{8.1.45}$$

where $f(kR)$ is the top hat distribution function

$$f(kR) \equiv \frac{3a_0^3}{4\pi R^3} \int_{a_0|\mathbf{x}|<R} d^3x e^{i\mathbf{k}\cdot\mathbf{x}a_0} = \frac{3}{(kR)^3}(\sin kR - kR\cos(kR)) \tag{8.1.46}$$

In particular, if we take $P(k)$ to be given by Eq. (8.1.42), then

$$\begin{aligned} \sigma_R^2(z) =& \frac{16\pi N^2}{25 k_*^{n_S-1}\Omega_M^2 H_0^4} \frac{C^2\left(\Omega_\Lambda/\Omega_M(1+z)^3\right)}{(1+z)^2} \\ & \times \int_0^\infty \left| \mathcal{T}\left(\sqrt{2}k/k_{\rm EQ}\right) \right|^2 |f(kR)|^2 k^{2+n_S} dk \end{aligned} \tag{8.1.47}$$

The top hat function has $f(0) = 1$, so there is no change in the infrared convergence of the integral for the mean square fluctuation, but $|f(kR)|^2$ decays as $9\cos^2(kR)/(kR)^4$ for $k \to \infty$, which is fast enough to remove the ultraviolet divergence for $n_S < 5$.

Many observations of the distribution of matter in the universe are commonly expressed in terms of a quantity called σ_8, which is the value of $\sigma_R(z)$ for $z = 0$ and $R = 8h^{-1}$Mpc. In calculating σ_8, we can evaluate the transfer function using the Dicus fitting formula (6.5.12), with the Eisenstein-Hu baryonic correction (6.5.22). Using the parameters $h = 0.72$, $\Omega_M h^2 =$

$0.14, \Omega_B h^2 = 0.024$ and $|N|^2 = 2.1 \times 10^{-10}$ which as discussed in Section 7.2 were found at the end of the first year of WMAP observations, and taking the present radiation temperature as $T_{\gamma,0} = 2.725$ K which yields $\Omega_R h^2 = 4.15 \times 10^{-5}$, Eq. (8.1.47) gives $\sigma_8 = 0.92$, in agreement with the result $\sigma_8 = 0.919$ found by the WMAP collaboration, which was calculated using the CMBfast computer program. [5] (The three-year WMAP data [6] gives a smaller value, $\sigma_8 = 0.761^{+0.049}_{-0.048}$, because the reduced optical depth inferred from three-year polarization data requires a smaller value of $|N|^2$ to give temperature correlations about the same magnitude. The reduction in $|N|^2$ also yields a corresponding decrease in the value of σ_8 calculated from Eq. (8.1.47).)

Even though Eq. (8.1.43) gives a divergent integral for $n_S \geq 1$, it suggests that $P(k)k^3$ can be used as a measure of the strength of the fluctuations of co-moving wave number $q = a_0 k$. With $P(k)$ given by Eq. (8.1.42) with $n_S \simeq 1$, we have $F^2(t)P(k)k^3 \propto t^{4/3}k^4 \left|T\left(k/k_{\rm EQ}\right)\right|^2$. This is a monotonically increasing function of k, so we can conclude that it is the perturbations of large co-moving wave number and hence small mass that become strong first, with small condensations merging into pre-galactic dark matter haloes and then ultimately into clusters of galaxies. This "bottom-up" picture of structure formation is the reverse of the "top-down" picture long advocated by Zel'dovich, [7] according to which very large condensations form first and then fragment into condensations on the scale of clusters of galaxies and finally into individual galaxies. The bottom-up picture is supported by the observation that the commonest galaxies are dwarf spheroidals, and that our galaxy and the Andromeda Nebula M31 each have about 20 smaller satellite galaxies.

Returning now to the power spectral function, inspection of Table 6.1 shows that the function $\kappa|\mathcal{T}(\kappa)|^2$ has a maximum value of 0.74 , reached at $\kappa = 2.0$. Since Eq. (6.4.58) gives k proportional to κ, with $k = \Omega_M h^2 \kappa /$ 19.3 Mpc, $P(k)$ for $n_S = 1$ has a maximum at this value of κ, corresponding to

$$k_{\rm max} = 0.10 \Omega_M h^2 {\rm Mpc}^{-1}. \tag{8.1.48}$$

Taking $C\left(\Omega_\Lambda/\Omega_M\right) = 0.767$ and $n_S = 1$, the value of the spectral function at its maximum is

$$P_{\rm max} = 7.2 \times 10^{13} \left(\Omega_M h^2\right)^{-1} |N|^2 {\rm Mpc}^3. \tag{8.1.49}$$

There are a number of measurements of the matter distribution at low redshift that have been used to calculate $P(k)$:

- There have been several surveys of galaxy positions and redshifts, of which the two most recent and detailed are the Sloan Digital Sky Survey, [8] which will include about 800,000 galaxies, and the 2 dF Survey, [9] which at its completion in 2003 had included 220,000 galaxies. Results of the Sloan Digital Sky Survey are shown in Figure 8.2.
  These two surveys give values of $P(k)$ in fair agreement with each other, for values of k ranging from about $0.015h{\rm Mpc}^{-1}$ to $0.2h{\rm Mpc}^{-1}$. (Wave numbers are given in these surveys in units of $h{\rm Mpc}^{-1}$ because distances are inferred from redshifts, and are therefore proportional to h^{-1}.) The measured $P(k)$ has a shape consistent with the result $\propto k|\mathcal{T}(\kappa)|^2$ expected from Eq. (8.1.42) for $n_S \simeq 1$, and in particular seems to reach a maximum at $k \approx 0.02h{\rm Mpc}^{-1}$, in good agreement with Eq. (8.1.48) for $\Omega_M \approx 0.3$ and $h \approx 0.7$. The value at this maximum is measured to be $P_{\rm max} \approx 5 \times 10^4 h^{-3}{\rm Mpc}^3$.

[475] D. N. Spergel et al., Astrophys. J. Suppl. 148, 175 (2003). The result $\sigma_8 = 0.9$ quoted in this reference was rounded off from $\sigma_8 = 0.919$, in order to reflect uncertainties in the input parameters.
[6] D. N. Spergel et al., Astrophys. J. Suppl. Ser. 170, 377 (2007) [astro-ph/0603449].
[7] Ya. B. Zel'dovich, Soviet Scientific Reviews, Section E: Astrophys. and Space Physics Reviews 3, 1 (1984); S. F. Shandarin and Ya. B. Zel'dovich, Rev. Mod. Phys. 61, 185 (1989); and earlier references cited therein.

Comparison of this result with Eq. (8.1.49) indicates a value $|N| \approx 2 \times 10^{-5}$ for the strength of primordial fluctuations, if we take $\Omega_M \approx 0.3$ and $h \approx 0.7$.

- The counts of numbers of "virialized" clusters of galaxies (like the Coma cluster) as a function of redshift gives information about the distribution of mass with distance. [10] (See Section 1.11.) Their results yield values for σ_8 ranging from 0.66 to about 1 , and can be interpreted as giving a value $P(k) \approx 6 \times 10^3 h^{-3}\text{Mpc}^3$ for $k \approx 0.1h\text{Mpc}^{-1}$, which falls on the curve provided by the above galaxy surveys.

- Other information about the distribution of mass with distance comes from correlations between the positions of intergalactic regions of higher than average density, revealed through Lyman α absorption of light from distant quasars that passes through these regions. [11] (See Section 1.10.) Their results yield values for $P(k)$ for k between about $0.1h\text{Mpc}^{-1}$ and $6h\text{Mpc}^{-1}$, which lie on the curve provided by the

[478] D. G. York et al., Astron. J. 120, 1579 (2000); M. Tegmark et al., Astrophys. J. 606, 702 (2004) [astro-ph/0310725]; Phys. Rev. D 69, 103501 (2004) [astro-ph/0310723]. The latest data release at the time of writing is analyzed by W. J. Percival et al., Astrophys. J. 657, 645 (2007) [astro-ph/0608636] and M. Tegmark et al., Phys. Rev. D 74, 123507 (2006) [astro-ph/0608632].

[9] W. J. Percival et al., Mon. Not. Roy. Astron. Soc. 327, 1297 (2001); M. Colless et al., Mon. Not. Roy. Astron. Soc. 328, 1039 (2001); M. Colless et al. (the 2dFGRS team), astro-ph/0306581. The final data set is analyzed in S. Cole et al. (the 2dF-GRS Team), Mon. Not. Roy. Astron. Soc. 362, 505 (2005) [astro-ph/0501174].

[10] For a summary with references to the original literature, see Table V of M. Tegmark et al., Phys. Rev. D 69, 103501 (2004).

[11] U. Seljak et al., Phys. Rev. D 71, 103515 (2005).

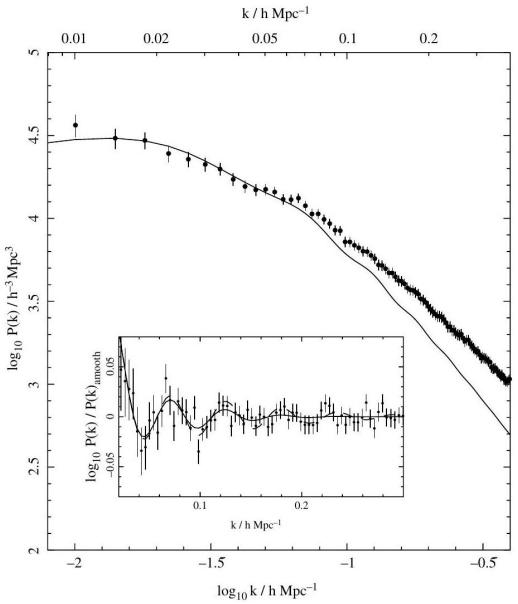


Figure 8.2: Measurement of the power spectral function by the Sloan Digital Sky Survey, from W. J. Percival et al., astro-ph/0608636. Dark circles show values of $h^{3} P(k)$, inferred from the survey of galaxy positions and redshifts, with distances calculated from redshifts using assumed cosmological parameters $\Omega_M = 0.24, \Omega_\Lambda = 0.76$. Vertical bars indicate $1 - \sigma$ errors. The solid curve is calculated using linear perturbation theory, with cosmological parameters taken from the WMAP third-year temperature and polarization data: $h = 0.73, \Omega_M = 0.24, \Omega_\Lambda = 0.76, \Omega_B/\Omega_M = 0.174$. The normalization of this curve is taken from a fit to the data for k between $0.01h\text{Mpc}^{-1}$ and $0.06h\text{Mpc}^{-1}$. The departure of the data from this theoretical curve for large k is attributed to non-linear effects on the growth of perturbations. The inset shows the effect of baryon acoustic oscillations, discussed at the end of this section. Data points give the ratio of the measured power spectral function to its value when smoothed to eliminate the oscillations. The solid curve shows the expected ratio, calculated using parameters from the third year WMAP data.

galaxy surveys for $k < 0.2h\text{Mpc}^{-1}$, and extend this curve to larger values of k, again in agreement with the expected shape $\propto k|\mathcal{T}(\kappa)|^2$.

- Yet more information about the distribution of mass comes from weak lensing of galactic images. [12] (See Section 9.5). Their results yield

values of σ_8 ranging from 0.67 to 0.97 , and can be interpreted as giving values of $P(k)$ for k in the neighborhood of $0.3h\text{Mpc}^{-1}$, that fall on the curve provided by the galaxy surveys.

These measurements are subject to various uncertainties, which raise highly technical issues of astrophysics. In relying on redshifts to give distances, these measurements are vulnerable to complications arising from the peculiar velocities of galaxies or galaxy clusters or Lyman α clouds. The theory used to interpret these measurements assumes that the concentrations of dark matter can be treated as small perturbations, so it is necessary to avoid using data for values of k where fluctuations have become non-linear. Finally, in interpreting the distribution of baryonic matter in galaxies or clusters of galaxies or Lyman α clouds in terms of the total mass density, it is necessary either to assume that this introduces no bias, or else to know what the bias is. Only the weak lensing technique directly measures fluctuations in the total mass density. For this reason, although the other techniques can give good information about

[4712] See, e.g., H. Hoekstra et al., Astrophys. J. 647, 116 (2006).

the shape of the function $P(k)$, they leave its overall normalization rather uncertain. The results from the galaxy surveys quoted above are for "no bias," that is, with the distribution of baryonic matter before perturbations become nonlinear assumed to trace the distribution of cold dark matter.

One of the striking things to emerge from the study of the cosmic microwave background described in the previous chapter is that these results for $P(k)$ obtained from studies of large scale structure agree with the strength of primordial fluctuations found from the cosmic microwave background. That is, the primordial fluctuation strength $q^3 |\mathcal{R}_q|^2$ seems to be roughly constant over a wide range of physical wave numbers $k \equiv q/a_0$, from the values $\approx 10^{-3}\text{Mpc}^{-1}$ probed by observations of the cosmic microwave background down to all but the smallest multipole orders, to values more than 1Mpc^{-1} probed by studies of large scale structure. Indeed, even before the advent of the COBE measurements discussed in Section 2.6, the study of the large scale structure of matter had led to the expectation that the cosmic microwave background would show fractional temperature fluctuations of order 10^{-5}, corresponding to $|N| \approx 10^{-5}$.

The measurements discussed so far were all at small or moderate redshifts, where the structure distance in Eq. (8.1.28) is not very different from the linear approximation $d_S(z) \simeq z/H_0$. It is widely hoped that measurements of the matter correlation function ξ at larger redshifts will provide a determination of the functional form of $d_S(z)$ (or equivalently of $d_A(z)$) beyond this approximation, which could illuminate the time dependence of the vacuum energy. But Eq. (8.1.28) shows that in the integral for $\xi(z, \hat{n}; z', \hat{n}')$ it is only wave numbers $k \gtrsim 1/r$ for which the integrand is
sensitive to the value of $r \equiv |d_S(z)\hat{n} - d_S(z')\hat{n}'|$. If $r \ll 1/k_{\rm EQ}$ (where $k_{\rm EQ}$ is the wave number that comes into the horizon just at radiationmatter equality) then for $k \gtrsim 1/r$ and $n_S \simeq 1$, the approximate formula (8.1.42) for $P(k)$ together with the asymptotic formula (6.5.11) for the transfer function give $P(k) \sim k^{-3}[\ln k + O(1)]^2$, so the part of the integral (8.1.28) for the correlation function that depends on r has the r-dependence

$$\int_{1/r}^{\infty} \frac{dk}{rk^2} \sin(ka_0 r) [\ln k + O(1)]^2$$

which varies only logarithmically with r. Thus in order to use measurements of the matter correlation function to learn something about the z-dependence of $d_S(z)$, we need either to carry the measurements of the correlation function to redshifts and angles that are sufficiently different so that $|d_S(z)\hat{n} - d_S(z')\hat{n}'|$ is at least of order $1/k_{\rm EQ} = 19.3\text{Mpc}/\Omega_M h^2$, or take advantage of the small departures of the transfer function from the asymptotic formula (6.5.11) for $k > k_{\rm EQ}$, or else take advantage of small departures from our formula (8.1.42) for $P(k)$ at $k > k_{\rm EQ}$.

Such departures from Eq. (8.1.42) for $P(k)$ are provided by terms of order $\beta \equiv \Omega_B/\Omega_M$, produced by baryon acoustic oscillations before the time of decoupling. [13] To see the effect of these oscillations, we must return to Eq. (8.1.14), and now look at the "fast" terms in the various perturbations, which oscillate with wave number for large wave number. In estimating the order of magnitude of these terms, we take $\bar{\rho}_B(t_L)$ and $\bar{\rho}_\gamma(t_L)$ to be of the same order of magnitude, so that $R(t_L)$ is of order unity, and $8\pi G\bar{\rho}_\gamma(t_L)$ is of order $\beta H^2(t_L)$. In estimating orders of magnitude we also do not distinguish between a_L and $a_{\rm EQ}$, which only differ by a factor of order 3 . Then, keeping track only of the small factors β and $a_L H_L/q$, Eqs. (6.4.45)(6.4.48) show that, aside from a common factor $\mathcal{R}_q^o \exp\left(-\int_0^{t_L} \Gamma dt\right)$, the

[4713] The influence of baryon acoustic oscillations before decoupling on the matter distribution after decoupling was recognized by P. J. E. Peebles and J. T. Yu, Astrophys. J. 162, 815 (1970); R. A. Sunyaev and Ya. B. Zel'dovich, Astrophys. Space Sci. 7, 3 (1970); J. R. Bond and G. Efstathiou, Astrophys. J. 285, L45 (1984); J. A. Holtzmann, Astrophys. J. Suppl. 71, 1 (1989); W. Hu and N. Sugiyama, Astrophys. J. 471, 30 (1996); D. J. Eisenstein and W. Hu, Astrophys. J. 496, 605 (1998). I believe that the first to suggest using redshift

fast terms in $\beta\delta_{\gamma q}(t_L), (1-\beta)\delta_{Dq}(t_L)$, and $t_L\psi_q(t_L)$ are of the order of $\beta, \beta H_L^2 a_L^2/q^2$, and $\beta H_L a_L/q$, respectively, while $\beta t_L (q^2/a_L^2)\,\delta u_{\gamma q}(t_L)$ is of order $\beta\,(q/a_L H_L)$, and is therefore dominant (aside from Silk damping) for $q/a_L H_L \gg 1$. Keeping only this term, we see that the oscillating part of $\Delta(q)$ is approximately

$$\begin{aligned}\Delta^{\rm fast}(q) &\simeq \beta t_L \left(q^2/a_L^2\right)\delta u_{\gamma q}^{\rm fast}(t_L) \\ &\simeq \frac{2\beta q \mathcal{R}_q^o}{\sqrt{3}a_L H_L (1+R_L)^{3/4}} \exp\left(-\int_0^{t_L}\Gamma dt\right)\sin(q d_H/a_L)\end{aligned} \tag{8.1.50}$$

where as usual $R \equiv 3\bar\rho_B/\bar\rho_\gamma$; d_H is the acoustic horizon distance (7.2.39) at decoupling

$$\begin{aligned}d_H &\equiv a_L\int_0^{t_L}\frac{dt}{a\sqrt{3(1+R)}} \\ &= \frac{2}{H_0\sqrt{3R_L\Omega_M}\,(1+z_L)^{3/2}}\ln\left(\frac{\sqrt{1+R_L}+\sqrt{R_{\rm EQ}+R_L}}{1+\sqrt{R_{\rm EQ}}}\right)\end{aligned} \tag{8.1.51}$$

and, for reasons explained earlier, we have replaced t_L in Eq. (8.1.14) with $2/3H_L$. Thus the ratio of $P(k)$ to the smooth curve given by Eq. (8.1.42) will have bumps at $k \simeq \pi k_H/2, 3\pi k_H/2, 5\pi k_H/2\ldots$, where $k_H = 1/d_H(1+z_L)$. These bumps have been seen in the spectral function inferred from observations by both the Sloan Digital Sky Survey [14] and the 2dF Galaxy Redshift Survey. [15] (See the inset in Figure 8.2.) The proportionality of d_H to $\Omega_M^{-1/2}$ has allowed a determination of Ω_M by matching the observed and predicted positions of these bumps [14,15]. The result [16] is $\Omega_M = 0.256^{+0.029}_{-0.024}$, or with a larger sample, [17] $\Omega_M = 0.24 \pm 0.02$. Unfortunately, these measurements are at moderate redshifts ($0.16 < z < 0.47$ for the Sloan survey, and $z < 0.3$ for the 2dF Survey), and although they give evidence for dark energy, they do not yet provide information about its time dependence.

As promised, we close by estimating the cosmic variance incurred in using Eq. (8.1.33). The mean square fractional error in this formula is

$$\begin{aligned}\Delta_V(Q,t) \equiv &\left\langle\left[\int\frac{d^2\hat Q}{4\pi}\left(\left|\delta^V_{MQ\hat Q^{(t)}}\right|^2 - \langle|\delta^V_{MQ\hat Q^{(t)}}\right|^2\rangle\right)\right]^2\right\rangle \\ &\times\left(\left\langle|\,\delta^V_{MQ\hat Q^{(t)}}\right\rangle\right)^2 = \delta_{MQ}^{-4}(t)\int\frac{d^2\hat Q}{4\pi}\int\frac{d^2\hat Q'}{4\pi} \\ &\times\left[\langle|\delta^V_{MQ\hat Q^{(t)}}\right|^2\left|\delta^V_{MQ\hat Q'}(t)\right|^2\rangle - \langle|\delta^V_{MQ\hat Q^{(t)}}\right|^2\rangle\,\langle|\delta^V_{MQ\hat Q'}(t)\right|^2\rangle\right]\end{aligned} \tag{8.1.52}$$

Assuming Gaussian statistics, this is

$$\begin{aligned}\Delta_V(Q,t) =& \delta_{MQ}^{-4}(t)\int\frac{d^2\hat Q}{4\pi}\int\frac{d^2\hat Q'}{4\pi} \\ &\times\left[|\left\langle\delta^V_{MQ\hat Q'}(t)\delta^{V*}_{MQ\hat Q'}(t)\right|^2 + \left|\left\langle\delta^V_{MQ\hat Q'}(t)\delta^V_{MQ\hat Q'}(t)\right|^2\right]\end{aligned}$$

surveys to measure $d_A(z)$ by observation of baryon acoustic oscillations were D. J. Eisenstein, W. Hu, and M. Tegmark, Astrophys. J. 504, L57 (1998), and that the suggestion to use this method to measure the evolution of dark energy is due to D. J. Eisenstein, in Next Generation Wide-Field Multi-Object Spectroscopy, eds. M. Brown and A. Dey (ASP Conference Series, vol. 280, 2002): p. 35.

[4714] D. J. Eisenstein et al., Astrophys. J. 633, 560 (2005) [astro-ph/0501171].

[15] S. Cole et al., Mon. Not. Roy. Astron. Soc. 362, 505 (2005).

[16] W. J. Percival et al., Astrophys. J. 657, 51 (2007) [astro-ph/0608635].

[17] M. Tegmark et al., Phys. Rev. 74, 123507 (2006) [astro-ph/0608632].

Using Eqs. (8.1.31) and (8.1.24) (and recalling that $\alpha(\mathbf{q}) = \alpha^*(-\mathbf{q})$ and $\delta_{Mq}(t)$ is real) then gives

$$\Delta_V(Q,t) = 2\delta_{MQ}^{-4}(t) \int \frac{d^2\hat{Q}}{4\pi} \int \frac{d^2\hat{Q}'}{4\pi} \times \left| \int d^3q F_V(\mathbf{q} - Q\hat{Q}) F_V^* \left(\mathbf{q} - Q\hat{Q}'\right) \delta_{Mq}^2(t) \right|^2$$

The large volume limit $F_V(\mathbf{q}-\mathbf{Q}) \to \left((2\pi)^{3/2}/\sqrt{V}\right) \delta^3(\mathbf{q}-\mathbf{Q})$ then gives the cosmic variance as

$$\Delta_V(Q,t) = \frac{2(2\pi)^3}{V} \int \frac{d^2\hat{Q}}{4\pi} \int \frac{d^2\hat{Q}'}{4\pi} F_V^2 \left(Q\hat{Q} - Q\hat{Q}'\right)$$

Now, $F_V^2 \left(Q\hat{Q} - Q\hat{Q}'\right)$ takes the value $V/(2\pi)^3$ for $Q\left|\hat{Q} - \hat{Q}'\right| \ll V^{-1/3}$, and vanishes exponentially for $Q\left|\hat{Q} - \hat{Q}'\right| \gg V^{-1/3}$, so aside from factors of order unity,

$$\Delta_V(Q,t) \approx \frac{1}{Q^2 V^{2/3}} \tag{8.1.53}$$

Thus the mean square fractional error in using Eq. (8.1.33) vanishes as $Q^{-2}V^{-2/3}$, and becomes negligible when the survey volume contains many
wavelengths. The numerical factor depends on the shape of the survey volume. For instance, if the survey volume is a sphere of co-moving radius R, then for $QR \gg 1$,

$$\Delta_V(Q,t) \to \frac{9}{4Q^2R^2} \tag{8.1.54}$$

10.8.2 8.2 Nonlinear growth

The study of the growth of perturbations beyond the linear approximation presents formidable mathematical difficulties. It is usually pursued by the use of computer simulations, which are beyond the scope of this book. To allow an analytic treatment, we can adopt a generalization of an idealization of nonlinear growth originally due to Peebles. [1] With Peebles, we will consider a fluctuation to have an overdensity $\Delta\rho_M$ (that is, a total density $\bar{\rho}_M + \Delta\rho_M$ greater than the cosmic average $\bar{\rho}_M$) that is uniform within a finite sphere. [2] According to the Birkhoff theorem, [3] the metric and the equations of motion of a freely falling test particle inside the sphere are independent of what is happening outside the sphere, and are therefore the same as in a homogeneous isotropic universe, described by a RobertsonWalker metric, with a density $\bar{\rho}_M(t) + \Delta\rho_M(t)$, and a curvature constant that is not in general equal to the cosmological curvature constant K. In a Robertson-Walker metric with curvature constant K, the scale factor $a(t)$ satisfies the Friedmann equation (1.5.19):

$$\dot{a}^2(t) + K = \frac{8\pi G a^2(t)}{3} \left(\bar{\rho}_M(t) + \rho_V\right) \tag{8.2.1}$$

Likewise, the scale factor $\mathcal{A}(t)$ of the Robertson-Walker metric inside the fluctuation will satisfy a Friedmann equation:

$$\dot{\mathcal{A}}^2(t) + K + \Delta K = \frac{8\pi G \mathcal{A}^2(t)}{3} \left(\bar{\rho}_M(t) + \Delta\rho_M(t) + \rho_V\right) \tag{8.2.2}$$

where $K + \Delta K$ is the curvature constant of the interior metric. We are including in the total density both a non-relativistic mass density, and a

vacuum density ρ_V (not included by Peebles) which for want of any contrary evidence is taken to be time-independent.

We will only consider the case of vanishing cosmological curvature, [4]$K = 0$. We are primarily interested in fluctuations that have a chance to stop expanding and recollapse to high density, so we shall take $\Delta K > 0$. The total matter density obeys the conservation law

$$\bar{\rho}_M(t) + \Delta\rho_M(t) \propto \mathcal{A}^{-3}(t) \tag{8.2.3}$$

while the unperturbed density satisfies

$$\bar{\rho}_M(t) \propto a^{-3}(t) \tag{8.2.4}$$

With $K = 0$, the normalization of a is arbitrary; we will find it convenient to use Eqs. (8.2.3) and (8.2.4) to normalize a so that

$$\mathcal{A}^3(t)\,(\bar{\rho}_M(t) + \Delta\rho_M(t)) = a^3(t)\bar{\rho}_M(t) \tag{8.2.5}$$

In order to provide initial conditions for this problem, we must first consider times that are sufficiently early in the matter-dominated era so that $\Delta\rho_M(t)$ and $\Delta\mathcal{A}(t) \equiv \mathcal{A}(t) - a(t)$ can be treated as small perturbations, and the vacuum energy may be neglected. As we saw in Section 1.5, in this case Eqs. (8.2.1) and (8.2.4) have the solution

$$a \propto t^{2/3}, \quad \bar{\rho}_M = \frac{1}{6\pi G t^2} \tag{8.2.6}$$

To first order in the perturbations around this solution, Eq. (8.2.2) becomes

$$2\dot{a}\Delta\dot{\mathcal{A}} + \Delta K = \frac{8\pi G a^2 \bar{\rho}_M}{3}\left(\frac{2\Delta\mathcal{A}}{a} + \frac{\Delta\rho_M}{\bar{\rho}_M}\right)$$

or, using Eq. (8.2.6)

$$\frac{\Delta K}{a^2} = \frac{4}{9t^2}\left(\frac{2\Delta\mathcal{A}}{a} + \frac{\Delta\rho_M}{\bar{\rho}_M}\right) - \frac{4}{3t}\frac{\Delta\dot{\mathcal{A}}}{a}.$$

Also, to first order in perturbations, Eq. (8.2.5) gives at early times

$$\frac{\Delta\mathcal{A}}{a} = -\frac{\Delta\rho_M}{3\bar{\rho}_M},$$

so $\Delta\mathcal{A}$ is governed at early times by the first-order differential equation

$$\frac{\Delta K}{a^2} = -\frac{4}{9t^2}\frac{\Delta\mathcal{A}}{a} - \frac{4}{3t}\frac{\Delta\dot{\mathcal{A}}}{a} \tag{8.2.7}$$

This has solutions for $t \to 0$ of the form $\Delta\mathcal{A} \propto t^{4/3}$ and $\Delta\mathcal{A} \propto t^{-1/3}$. Assuming that enough time has passed for the solution $\Delta\mathcal{A} \propto t^{-1/3}$ to die away, we have $\Delta\mathcal{A} \propto t^{4/3}$, and so

[471] P. J. E. Peebles, Astrophys. J. 147, 859 (1967).

[2] J. E. Gunn and J. R. Gott, Astrophys. J. 176, 1 (1972) took into account the infall of matter from outside a sphere of uniform overdensity, which amounts to treating the overdensity of the sphere and its surroundings as a step function of radius. A non-vanishing vacuum energy was incorporated in the Gunn-Gott model by H. Martel, P. R. Shapiro, and S. Weinberg, Astrophys. J. 492, 29 (1998), who considered fluctuations consisting of a ball with a uniform overdensity surrounded by a spherical shell with a uniform underdensity.

[3] G&C, Sec. 11.7.

[474] The case of negative cosmological curvature is considered by B. Freivogel, M. Kleban, M. N. Martinez, and L. Susskind, J. High Energy Phys. 0603, 039 (2006) [hep-th/0505232].

$$\Delta K = -\frac{20}{9}\lim_{t\to 0} t^{-2}a(t)\Delta A(t) = \frac{40\pi G}{9}\lim_{t\to 0} a^2(t)\Delta\rho_M(t) \tag{8.2.8}$$

In characterizing the initial strength of a fluctuation within a co-moving radius r, we note from Eq. (8.2.8) that at early times $\Delta\rho_M(t) \propto a^{-2}(t) \propto \bar{\rho}_M^{2/3}(t)$, so we can define a time-independent quantity, which we shall call the initial fluctuation strength:

$$\rho_1 \equiv \lim_{t\to 0}\frac{\Delta\rho_M^3(t)}{\bar{\rho}_M^2(t)} \tag{8.2.9}$$

Then Eq. (8.2.8) can be written

$$\begin{aligned}\Delta K &= \frac{40\pi G}{9}a^2(t)\bar{\rho}_M^{2/3}(t)\rho_1^{1/3} \\ &= \frac{40\pi G}{9}\mathcal{A}^2(t)\left(\bar{\rho}_M(t)+\Delta\rho_M(t)\right)^{2/3}\rho_1^{1/3}\end{aligned} \tag{8.2.10}$$

Note that we do not have to take the limit $t \to 0$ here, because $a^2\bar{\rho}_M^{2/3}$ is time-independent.

Now let's consider the development of the fluctuation at later times, when it can no longer be treated as a small perturbation. Using Eq. (8.2.10) allows us to write the Friedmann equation (8.2.2) (with $K = 0$) as

$$\dot{\mathcal{A}}^2 = \frac{8\pi G\mathcal{A}^2}{9}\left[3\left(\bar{\rho}_M+\Delta\rho_M+\rho_V\right) - 5\left(\bar{\rho}_M+\Delta\rho_M\right)^{2/3}\rho_1^{1/3}\right]$$

or, using Eq. (8.2.3),

$$\begin{aligned}&\left(\bar{\rho}_M+\Delta\rho_M\right)^{-2}\left(\frac{d}{dt}\left(\bar{\rho}_M+\Delta\rho_M\right)\right)^2 = 8\pi G \\ &\quad\times\left(3\left(\bar{\rho}_M+\Delta\rho_M+\rho_V\right) - 5\left(\bar{\rho}_M+\Delta\rho_M\right)^{2/3}\rho_1^{1/3}\right)\end{aligned} \tag{8.2.11}$$

The right-hand side vanishes at a total mass density $\bar{\rho}_M + \Delta\rho_M = \rho_c$ satisfying

$$3\left(\rho_c+\rho_V\right) = 5\rho_c^{2/3}\rho_1^{1/3} \tag{8.2.12}$$

For $\rho_c > 0$, the quantity $5\rho_c^{2/3}\rho_1^{1/3} - 3\rho_c$ takes values from $-\infty$ to $500\rho_1/243$, so Eq. (8.2.12) has a solution if and only if the fluctuation is strong enough so that [5]

$$\rho_1 \geq \frac{729}{500}\rho_V \tag{8.2.13}$$

As long as Eq. (8.2.12) has a solution, the density $\bar{\rho}_M(t) + \Delta\rho_M(t)$ will drop until it reaches the value ρ_c, and then increase again to infinity. The total time elapsed for this expansion and collapse is given by Eq. (8.2.11) as

$$t_c = \frac{2}{\sqrt{8\pi G}}\int_{\rho_c}^{\infty}\frac{d\rho}{\rho\sqrt{3\left(\rho+\rho_V\right) - 5\rho^{2/3}\rho_1^{1/3}}} \tag{8.2.14}$$

(The factor 2 appears here because it takes the same time to contract from the minimum density ρ_c to infinite density as it does to expand from infinite to minimum density.) For example, in the limit $\rho_1 \gg \rho_V$, this gives

$$t_c \to \frac{9\pi}{5^{3/2}\sqrt{8\pi G\rho_1}} \tag{8.2.15}$$

If linear perturbation theory held up to this time then instead of becoming infinite, in this case the fractional density perturbation at time t_c would be

$$\frac{\Delta\rho_M}{\bar{\rho}_M} = \left(\frac{\rho_1}{\bar{\rho}_M}\right)^{1/3} = \rho_1^{1/3}\left(6\pi G t_c^2\right)^{1/3} = \left(\frac{243\pi^2}{500}\right)^{1/3} = 1.686$$

We see that gravitational collapse occurs at a time when linear perturbation theory would predict a fractional density perturbation large enough to make obvious its own invalidity.

Of course, different fluctuations will have different initial strengths ρ_1 and co-moving radii R. In order to make contact between this analysis and observation, it is convenient to make use of an approach due to Press and Schechter. [6] By inverting Eq. (8.2.14), we can calculate the minimum initial fluctuation strength $\rho_1(t_c)$ required for collapse at or before a time t_c. For instance, for fluctuations that are sufficiently strong to collapse before vacuum energy becomes important, Eq. (8.2.15) gives $\rho_1(t_c) = 81\pi/1000Gt_c^2$. To calculate the probability that a random point in space will be in a fluctuation this strong, we assume that at an early time t, before non-linearities become significant, the probability $P_{t,R}(\Delta\rho_M)\, d\Delta\rho_M$ that the average density within a co-moving sphere of radius R is increased by a density excess

between $\Delta\rho_M$ and $\Delta\rho_M + d\Delta\rho_M$ is given by the Gaussian distribution

$$P_{t,R}(\Delta\rho_M)\, d\Delta\rho_M = \frac{d\Delta\rho_M}{\sqrt{2\pi}\sigma_R(z_t)\,\bar{\rho}_M(t)} \exp\left(-\frac{\Delta\rho_M^2}{2\sigma_R^2(z_t)\,\bar{\rho}_M^2(t)}\right) \tag{8.2.16}$$

For a Harrison-Zel'dovich spectrum of fluctuations with $\mathcal{R}_q = Nq^{-3/2}$, the standard deviation σ_R for the fractional fluctuation $\Delta\rho_M/\bar{\rho}_M$ averaged over a radius R is given by Eq. (8.1.47), which for $n_S = 1$ reads

$$\begin{aligned}\sigma_R^2(z) =& \frac{16\pi N^2}{25\Omega_M^2 H_0^4} \frac{C^2\left(\Omega_\Lambda/\Omega_M(1+z)^3\right)}{(1+z)^2} \\ & \int_0^\infty \left|\mathcal{T}\left(\sqrt{2}k/k_{\rm EQ}\right)\right|^2 |f(kR)|^2 k^3 dk\end{aligned} \tag{8.2.17}$$

where f is the top hat distribution function (8.1.46); $\mathcal{T}$ is the scalar transfer function, defined and calculated in Section 6.5; and $k_{\rm EQ}$ is the wave number (8.1.41) that comes into the horizon at matter-radiation equality. Using Eq. (8.2.9), this can be expressed as a time-independent probability $P_R(\rho_1)\, d\rho_1$ for fluctuations averaged over a co-moving sphere of radius R to have a initial strength between ρ_1 and $\rho_1 + d\rho_1$:

$$\tilde{P}_R(\rho_1)\, d\rho_1 = \frac{d\rho_1}{3\sqrt{2\pi}\rho_1^{2/3}\tilde{\sigma}_R} \exp\left(-\frac{\rho_1^{2/3}}{2\tilde{\sigma}_R^2}\right) \tag{8.2.18}$$

where $\tilde{\sigma}_R$ is a time-independent quantity, given by

$$\begin{aligned}\tilde{\sigma}_R^2 &\equiv \lim_{z\to\infty} \bar{\rho}_M^{2/3}(t(z))\sigma_R^2(z) \\ &= \frac{16\pi N^2}{25}\left(\frac{3}{8\pi G}\right)^2 \bar{\rho}_{M0}^{-4/3} \int_0^\infty \left|\mathcal{T}\left(\sqrt{2}k/k_{\rm EQ}\right)\right|^2 |f(kR)|^2 k^3 dk\end{aligned} \tag{8.2.19}$$

Because we are averaging over a sphere of co-moving radius R, it is only fluctuations whose co-moving radii are greater than R that contribute to this probability. Integrating, we see that

[475] S. Weinberg, Phys. Rev. Lett. 59, 2607 (1987).

[6] W. H. Press and P. Schechter, Astrophys. J. 239, 1 (1974).

at early times, before the fluctuations become strong, the probability that a random point in space is in a fluctuation with initial strength greater than ρ_1 and co-moving radius greater than R is

$$P\,(>\rho_1, >R) = \int_{\rho_1}^{\infty} \frac{d\rho}{3\sqrt{2\pi}\rho^{2/3}\tilde{\sigma}_R} \exp\left(-\frac{\rho^{2/3}}{2\tilde{\sigma}_R^2}\right) = \frac{1}{2}\left[1 - \mathrm{erf}\left(\rho_1^{1/3}/\sqrt{2}\tilde{\sigma}_R\right)\right] \tag{8.2.20}$$

where $\mathrm{erf}(y)$ is the usual error function

$$\mathrm{erf}(y) \equiv \frac{1}{\sqrt{2\pi}} \int_{-y\sqrt{2}}^{y\sqrt{2}} dx \exp\left(-x^2/2\right) \tag{8.2.21}$$

At early times the universe has uniform matter density, so Eq. (8.2.20) also gives the fraction of all matter that is in fluctuations with initial strength greater than ρ_1 and co-moving radius greater than R. If we now set ρ_1 in Eq. (8.2.20) equal to the critical initial strength $\rho_1(t)$ for collapse by a time t, then we find that at time t, when nonlinearities have become important, the fraction of all matter that is in collapsed structures with co-moving radii greater than R will be

$$F(>R,t) = \frac{1}{2}\left[1 - \mathrm{erf}\left(\rho_1^{1/3}(t)/\sqrt{2}\tilde{\sigma}_R\right)\right] \tag{8.2.22}$$

The mass in a sphere of co-moving radius R is the time-independent quantity $4\pi\bar{\rho}_M(t)a^3(t)R^3/3$, so at time t the number density $n(M,t)dM$ of collapsed structures with mass between M and $M + dM$ is given by

$$n(M,t) = -\frac{\bar{\rho}_M(t)}{2M}\frac{d}{dM}\,\mathrm{erf}\left(\rho_1^{1/3}(t)/\sqrt{2}\tilde{\sigma}_{R(M)}\right) = \frac{\rho_1^{1/3}(t)\bar{\rho}_M(t)}{M\sqrt{2\pi}}\left|\frac{d\tilde{\sigma}_{R(M)}^{-1}}{dM}\right| \exp\left(-\frac{\rho_1^{2/3}(t)}{2\tilde{\sigma}_{R(M)}^2}\right) \tag{8.2.23}$$

where $R(M) \equiv (3M/4\pi a^3\bar{\rho}_M)^{1/3}$. (Press and Schechter somewhat arbitrarily multiplied this by a factor 2 , to take account of the matter in regions with a negative density fluctuation. [7]) These collapsed structures eventually furnish the halos of cold dark matter that surround galaxies in the present universe. [8]

In using Eq. (8.2.23), we need to know the mass dependence of $\tilde{\sigma}_{R(M)}$. For large masses and radii, the integral (8.2.19) is dominated by low values of the wave number k, for which the transfer function $T(\kappa)$ is close to unity, so Eq. (8.2.19) gives $\tilde{\sigma}_R^2 \propto R^{-4}$, and so $\tilde{\sigma}_{R(M)}^2 \propto M^{-4/3}$. (For a primordial fluctuation spectrum $\mathcal{R}_q \propto q^{(-4+n_S)/2}$, we would have a factor $k^{2+n_S}dk$ in place of k^3dk in Eq. (8.2.19), which would give $\tilde{\sigma}_R^2 \propto R^{-3-n_S}$, and so $\tilde{\sigma}_{R(M)}^2 \propto M^{-(3+n_S)/3}$.) The argument of the exponential in Eq. (8.2.23) in this case is thus proportional for $n_S = 1$ to $M^{4/3}$, giving a rapid fall-off of the number density for large mass. On the other hand, for small masses and radii, the integral (8.2.19) is dominated by large values

[477] For a derivation of this factor of 2, see J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser, Astrophys. J. 379, 440 (1991).

[8] The effects of non-spherical collapse on the mass function $n(M,t)$ is considered by P. Monaco, Astrophys. J. 447, 23 (1995); J. Lee and S. F. Shandarin, Astrophys. J. 500, 14 (1998); R. K. Sheth and G. Tormen, Mon. Not. Roy. Astron. Soc. 308, 119 (1999); R. K. Sheth, H. J. Mo, and G. Tormen, Mon. Not. Roy. Astron. Soc. 323, 1 (2001).

of the wave number k, for which according to Eq. (6.5.11) the transfer function $T(\kappa)$ falls off like $\ln k/k^2$. The integral (8.2.19) then varies only logarithmically with R, and so the number density $n(M)$ goes for small M more-or-less like M^{-2}. The detailed M-dependence predicted by Eq. (8.2.23) is in reasonable agreement with the results of large computer simulations of the evolution of cold dark matter. [9]

10.8.3 8.3 Collapse of baryonic matter

Until now we have supposed that the baryonic matter of the universe, which after recombination consisted chiefly of neutral hydrogen and helium, had negligible pressure. In this case baryonic matter just followed along with cold dark matter in its expansion and possible recontraction. Actually, as we saw in Section 2.3, the baryonic matter retained a small residual ionization even after the nominal era of recombination, providing enough electrons for Compton scattering of photons of the cosmic microwave background to keep the temperature of baryonic matter equal to the temperature of the microwave background until the redshift dropped below about 150 . Small overdense regions did not have enough of a gravitational field to overcome the baryonic pressure, so their baryonic matter did not collapse along with the cold dark matter. This led to relatively small clumps of cold dark matter that now do not contain galaxies, and are therefore undetectable except for their gravitational effects. The question is, how small did a clump of matter have to be for its baryonic component to have resisted gravitational collapse?

Before the existence of cold dark matter was generally accepted, this question was addressed in a simple theory due originally to James Jeans. [1] According to this theory, small perturbations either oscillate or grow according to whether their wave number is greater or less than a critical wave number $k_J = \sqrt{4\pi G\bar{\rho}_B}/v_s$ (where v_s is the speed of sound) so a clump is too small to collapse if its mass is less than the Jeans mass, given by $\bar{\rho}_B\,(2\pi/k_J)^3$. This theory naturally (given its date) was not originally set in the context of an expanding universe, but not much changes

when the expansion is taken into account. [2] But as we shall see, it turns out that results are quite different when we include the effects of cold dark matter.

We are considering only non-relativistic matter during an era in which radiation contributed little to the gravitational field, so we can apply the Newtonian cosmological theory described in Appendix F. By following the same reasoning that led there to Eqs. (F.15) and (F.16), but now including both baryonic matter with a squared sound speed $\partial p_B/\partial\rho_B = v_s^2$ and cold dark matter with zero pressure, we find that the velocity potential perturbations δu_B and δu_D and the density perturbations $\delta\rho_B$ and $\delta\rho_D$ for co-moving wave number $\mathbf{q}$ are governed by the equations of continuity

$$\frac{d\delta\rho_D}{dt} + 3H\delta\rho_D - a^{-1}\bar{\rho}_D\mathbf{q}^2\delta u_D = 0 \tag{8.3.1}$$

$$\frac{d\delta\rho_B}{dt} + 3H\delta\rho_B - a^{-1}\bar{\rho}_B\mathbf{q}^2\delta u_B = 0 \tag{8.3.2}$$

and the Euler equations

$$\frac{d\delta u_D}{dt} + H\delta u_D = \frac{4\pi G a}{\mathbf{q}^2}\left[\delta\rho_D + \delta\rho_B\right] \tag{8.3.3}$$

$$\frac{d\delta u_B}{dt} + H\delta u_B = \frac{4\pi G a}{\mathbf{q}^2}\left[\delta\rho_D + \delta\rho_B\right] - \frac{v_s^2}{a\bar{\rho}_B}\delta\rho_B \tag{8.3.4}$$

[479] V. Springel et al., Nature 435, 629 (2005) [astro-ph/0504097].

[1] J. Jeans, Phil. Trans. Roy. Soc. 199A, 49 (1902), and Astronomy and Cosmogony (2nd ed., first published by Cambridge University Press in 1928; reprinted by Dover Publications, New York, 1961), pp. 345-350.

As in the relativistic case, it is convenient to introduce the fractional density perturbations $\delta_n \equiv \delta\rho_n/\bar{\rho}_n$. Using the Friedmann result that both unperturbed densities $\bar{\rho}_n$ go as $a^{-3} \propto t^{-2}$, and eliminating the velocity potentials, these equations then become

$$\ddot{\delta}_D + \frac{4}{3t}\dot{\delta}_D = \frac{2}{3t^2}\left[\beta\delta_B + (1-\beta)\delta_D\right] \tag{8.3.5}$$

$$\ddot{\delta}_B + \frac{4}{3t}\dot{\delta}_B = -\frac{2\alpha}{3t^2}\delta_B + \frac{2}{3t^2}\left[\beta\delta_B + (1-\beta)\delta_D\right] \tag{8.3.6}$$

where α and β are defined by

$$\alpha \equiv \frac{3\mathbf{q}^2 v_s^2 t^2}{2a^2} = \frac{\mathbf{q}^2 v_s^2}{4\pi G\bar{\rho}_M a^2}, \quad \beta \equiv \frac{\bar{\rho}_B}{\bar{\rho}_M} = \frac{\Omega_B}{\Omega_M} \simeq 0.17 \tag{8.3.7}$$

and $\bar{\rho}_M \equiv \bar{\rho}_D + \bar{\rho}_B$.
Note that α was constant during the era (roughly for $z > 150$) when baryonic matter had the same temperature as radiation, because $v_s^2 \propto T \propto a^{-1}$. Thus we can find power-law solutions [3] to Eqs. (8.3.5) and (8.3.6) that
apply during this era. We set

$$\delta_D \propto t^\nu, \quad \delta_B = \xi\delta_D, \tag{8.3.8}$$

where v and ξ are time-independent (but $\mathbf{q}$-dependent) quantities to be determined. Eqs. (8.3.5) and (8.3.6) then become

$$v(v-1) + \frac{4v}{3} = \frac{2}{3}[\beta\xi + (1-\beta)], \tag{8.3.9}$$

$$v(v-1) + \frac{4v}{3} + \frac{2\alpha}{3} = \frac{2}{3}[\beta + (1-\beta)/\xi]. \tag{8.3.10}$$

Eliminating ξ yields a quartic equation for v. This generically has four different solutions, so there are four independent power-law solutions of the fourth-order system of differential equations (8.3.5) and (8.3.6), which therefore form a complete set of solutions.

The general solutions of Eqs. (8.3.9) and (8.3.10) are too complicated to be illuminating, but we can find useful approximate solutions if we take into account the small value of β. In the limit of very small β, there are two baryon-poor solutions for which $\xi < 1$:

$$\begin{array}{ll} v = 2/3, & \xi = \frac{1}{1+\alpha} \\ v = -1, & \xi = \frac{1}{1+\alpha} \end{array} \tag{8.3.12}$$

and two baryon-rich solutions for which $\xi \gg 1$:

$$v = -\frac{1}{6} \pm \sqrt{\frac{1}{36} - \frac{2\alpha}{3}}, \quad \xi = -\frac{1+\alpha}{\beta} \tag{8.3.13}$$

All of these solutions decay with time, except for the first solution (8.3.11), in which the baryon and cold dark matter fractional density perturbations both grow as $t^{2/3}$. This solution therefore dominates at late times. To first order in β, the power-law exponent and baryon fraction in this mode are

$$v = 2/3 - \frac{2\beta\alpha}{5(1+\alpha)}, \quad \xi = \frac{1}{1+\alpha} - \frac{\beta\alpha^2}{(1+\alpha)^3}$$

[472] G&C, Sec. 15.9.
[3] Approximate power law solutions were given by P. J. E. Peebles, Astrophys. J. 277, 470 (1984).

The fractional baryonic corrections to v and ξ are a maximum for very short wavelengths, for which they take values $3\beta/5 \simeq 10\%$ and $\beta \simeq 17\%$, respectively.

There is nothing here like a critical wave number marking a transition from growth to oscillation, as in the classic one-component Jeans theory. The growing mode (8.3.11) grows for all wave numbers, with a growth rate that for small β depends very little on wave number. This of course is because the pressure of baryonic matter could not prevent clumps of cold dark matter from becoming increasingly denser than average. What does depend on the wave number is the fraction ξ of the baryons that follow the growing condensation of the cold dark matter. This can most conveniently be expressed in terms of the total mass M (dark and baryonic matter) in a cubic physical wavelength $2\pi a/q$:

$$M \equiv \bar{\rho}_M \left(\frac{2\pi a}{q}\right)^3 \tag{8.3.14}$$

According to the formulas for ξ and α given by Eqs. (8.3.11) and (8.3.7), the fraction of baryons that collapsed along with the cold dark matter can be written

$$\frac{\delta\rho_B/\bar{\rho}_B}{\delta\rho_D/\bar{\rho}_D} \equiv \xi = \frac{1}{1+(M_J/M)^{2/3}}, \tag{8.3.15}$$

where M_J is a sort of Jeans mass

$$M_J = \left(\frac{\pi}{G}\right)^{3/2} \frac{v_s^3}{\bar{\rho}_M^{1/2}}. \tag{8.3.16}$$

Baryons collapsed along with cold dark matter for clumps of mass much greater than M_J, while clumps of mass much less than M_J are largely free of baryonic matter.

The speed of sound for a gas of hydrogen and helium atoms at temperature T is $v_s = (5k_{\mathcal{B}}T/3\mu m_N)^{1/2}$, where μ is the mean molecular weight, which for a helium abundance by weight of 24% is $\mu = 1.22$. Since we are considering an era in which $T = T_\gamma$, and $T_\gamma^3/\bar{\rho}_M$ equals its present value $T_{\gamma 0}^3/\bar{\rho}_{M0}$, the Jeans mass can be written

$$M_J = \left(\frac{5\pi k_{\mathcal{B}} T_{\gamma 0}}{3\mu m_N G}\right)^{3/2} \bar{\rho}_{M0}^{-1/2} = 2.02\times 10^5 \left(\Omega_M h^2\right)^{-1/2} M_\odot \tag{8.3.17}$$

For $\Omega_M h^2 = 0.13$, this is $6\times 10^5 M_\odot$, corresponding to a baryonic mass $\beta M_J \simeq 10^5 M_\odot$.

After the redshift dropped to about 150, the Compton scattering of the cosmic microwave background by the residual ionized hydrogen no longer kept the temperature of baryonic matter equal to the radiation temperature. According to Eq. (1.1.23), the kinetic energy and hence the temperature of the hydrogen and helium atoms then dropped like a^{-2}. Additional baryons then began to fall into the already growing clumps of cold dark matter, until the baryons were heated again by the energy released in gravitational collapse, and eventually by the first generation of stars. The increased baryonic temperature inhibited the further accretion of baryonic matter, and may have resulted in the expulsion of some baryonic matter from the clump. This is all quite complicated, [4] but it does not change the conclusion, that a clump of cold dark matter whose mass is less than M_J will not contain a full complement of baryonic matter.

10.9 Gravitational Lenses

In 1979 Walsh, Carswell, and Weymann [1] noticed a pair of quasars at the same redshift, about 1.4 , separated by just 6″. The similarity of the spectrum of the two quasars suggested to

[474] P. R. Shapiro, M.L. Giroux, and A. Babul, Astrophys. J. 427, 25 (1994).

them that there was really just one quasar, now known as Q0957+561, split into two images by the deflection of light in the gravitational field of an intervening massive body. This suggestion was verified with the discovery of a galaxy with a redshift 0.36 between the lines of sight to the quasar. Such gravitational lenses had already been studied theoretically by many authors, and now a serious search for them was put in train. [2] Many more lensing galaxies were discovered, generally elliptical field galaxies, and also some lensing clusters of galaxies, which generally produce arc-like images. [3]

Strong gravitational lensing has been used to search for dark objects, to explore the structure of galaxy clusters, and to measure the Hubble constant. Weak lensing offers great promise in measuring the correlation function of density fluctuations. Lensing is a large subject; in this chapter we will give only an overview of its cosmological applications.

10.9.1 9.1 Lens equation for point masses

We will first consider the gravitational lens provided by a point mass, and later take up more detailed models. To analyze the splitting of images by a point mass, suppose that the lines from the earth to a point source and the earth to the lensing mass in a Robertson-Walker coordinate system centered on the earth are separated by a small angle α. (See Figure 9.1.)

This is the angle that there would be between the images of the source and lens, if their were no gravitational deflection of light. Because of the deflection of light by the lens, there is a different angle β between the actual images of the source and lens. We need to derive a lens equation, which gives the relation between β and α.

In the coordinate system centered on the earth, the light ray from the source follows a path from the source to the neighborhood of the lens that

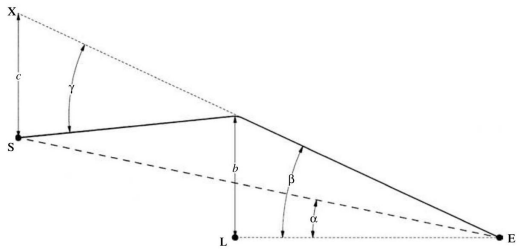


Figure 9.1: Quantities referred to in the derivation of the lens equation, (9.1.5). The bent solid line is the path of a photon from the source S past the lens L to the Earth E . The point X is the apparent position of the source. The transverse distances b and c from the lens to the light path and from the source to its apparent position are greatly exaggerated, as are the angles α, β, and γ. This is drawn for the case $K = 0$; in a Robertson-Walker coordinate system centered on the Earth, the path from the source to the bend would be curved for $|K| \neq 0$.

is curved for $K \neq 0$; is bent by the gravitational field of the lens; and then follows a straight line to the earth. The proper distance b of this path from the lens at its closest approach (which we assume to be much less than the cosmological scale $1/H_0$) is

$$b = \beta d_A(EL), \tag{9.1.1}$$

[471] D. Walsh, R. F. Carswell, and R. J. Weymann, Nature 279, 381 (1979).

[2] For a comprehensive review, see Gravitational Lenses, by P. Schneider, J. Ehlers, and E. E. Falco (Springer-Verlag, Berlin, 1992). A more recent review is given by G. Soucail, Proceedings of the XX Texas Symposium on Relativistic Astrophysics, Austin, December 2000.

[3] R. Lynds and V. Petrosian, Bull. Amer. Astron. Soc. 18, 1014 (1986); G. Soucail, B. Fort, Y. Mellier, and J. P. Picat, Astron. Astrophys. 172, L14 (1987); G. Soucail et al., Astron. Astrophys. 191, L19 (1988).

where $d_A(EL)$ is the "angular diameter distance" of the lens as seen from the earth. As discussed in Section 1.4, in general the angular diameter distance $d_A(PQ)$ of a point Q as seen from a point P is the ratio h/θ of a proper length h at Q (normal to the line PQ) to the angle θ subtended at P by this length. It is given by

$$d_A(PQ) = a\,(t_Q)\,r_P(Q), \tag{9.1.2}$$

where $r_P(Q)$ is the radial coordinate of Q in the Robertson-Walker coordinate system centered at P (in which rays of light received at P all travel on straight lines), and t_Q is the time the light leaves or arrives at Q. From Figure 9.1, we can see that the line segment SX from the true to the apparent position of the source has a proper length c given by

$$\gamma d_A(LS) = c = (\beta - \alpha) d_A(ES), \tag{9.1.3}$$

where γ is the angle of deflection of the light ray near the lens as seen from the lens, given by general relativity as [4]

$$\gamma = \frac{4MG}{b}, \tag{9.1.4}$$

and $d_A(LS)$ and $d_A(ES)$ are the angular diameter distances of the source from the lens and from the earth, respectively. From Eqs. (9.1.1), (9.1.3) and (9.1.4) we have

$$(\beta - \alpha)\beta = \frac{\gamma d_A(LS)}{d_A(ES)} \frac{b}{d_A(EL)} = \frac{4MG d_A(LS)}{d_A(ES) d_A(EL)} \equiv \beta_E^2. \tag{9.1.5}$$

This is our lens equation. It is a quadratic equation for β, the two solutions giving the directions of the two images into which the point source is split. All the effects of the large scale spacetime geometry and the expansion of the universe are contained in the angular diameter distances $d_A(EL), d_A(ES)$, and $d_A(LS)$. (In general these are independent distances; for $K \neq 0$ we do not have $d_A(ES) = d_A(EL) + d_A(LS)$.)

The lens equation (9.1.5) has two roots,

$$\beta_\pm = \frac{\alpha}{2} \pm \sqrt{\frac{\alpha^2}{4} + \beta_E^2}. \tag{9.1.6}$$

The angle α is not observed, because we cannot remove the lens. If all we measure is the angular separation between the two images of the source, then all we can learn is an upper bound on the mass of the lensing galaxy:

$$|\beta_+ - \beta_-|^2 \geq 4\beta_E^2 \tag{9.1.7}$$

For instance, if $d_A(EL) = d_A(LS) = 100\text{Mpc}, d_A(ES) = 200\text{Mpc}$, and the sources are separated by $1''$, then (remembering that $M_\odot G = 1.475$ km and $1\text{Mpc} = 3.09 \times 10^{19}$ km), we find that $M \leq 6 \times 10^9$ solar masses.

On the other hand, if the lensing galaxy and the two images are all observed, we can measure the angles $\beta_\pm$between each image and the lensing galaxy, but with α unknown the best use we can make of Eq. (9.1.6) is to eliminate α by multiplying the roots

$$\beta_+ \beta_- = -\beta_E^2 \tag{9.1.8}$$

(The minus sign just means that the two images are on opposite sides of the lensing galaxy.) This allows us to calculate the mass M, but if the distances

[474] G & C, Section 8.5.

are calculated from measurements of redshifts, then they scale with $1/H_0$, so what we really calculate in this way is MGH_0. (As we saw in Section 1.9, velocity dispersions and angular diameters also only tell us the value of MGH_0.)

In the special case where the lensing galaxy lies directly on the line between the source and the earth, the problem has cylindrical symmetry around this line of sight, and we get an Einstein ring rather than a pair of images. The angular radius of the Einstein ring is the value of β given by setting $\alpha = 0$ in Eq. (9.1.5), so it is just the angle β_E, which is why we label it with a subscript E. Einstein rings have been observed for a number of radio sources, starting with the source MG1131+0456. [5]

10.9.2 9.2 Magnification: Strong lensing and microlensing

The various images that are produced by a gravitational lens will not all have the same apparent luminosity. The apparent luminosity is the power received per receiving area, so now we need to consider light paths that end at various points on the telescope receiving area. For this purpose it is helpful to refer positions in the receiving area relative to some fixed point Y, which we can conveniently take on the axis of symmetry of the problem the line extending from a point source (or a luminous point on an extended source) through the lens and past the earth. (See Figure 9.2.) We can think of the distance h of a point on the telescope mirror from this line as a function of θ, the angle at the source between the light ray to the point on the mirror and the fixed line (in the Robertson-Walker coordinate system centered on the source) from the source through the lens. The fraction of all light that is emitted between polar angles θ and $\theta + d\theta$ (with $\theta \ll 1$) and azimuthal angles ϕ and $\phi + d\phi$ (measured at the source, around the fixed line to the lens) is $\theta d\theta d\phi/4\pi$, while the receiving area between in the rectangle with height dh and width $hd\phi$ is $hdhd\phi$, so the apparent luminosity is

$$\ell = \left| \frac{L\theta d\theta d\phi/4\pi}{hdhd\phi\,(1+z_S)^2} \right| = \frac{L}{4\pi\,(1+z_S)^2} \left| \frac{\theta d\theta}{hdh} \right| . \tag{9.2.1}$$

(The factor $(1+z_S)^{-2}$ accounts for the reduction of energy of individual photons and the reduction in the rate at which photons are emitted from the source.) From Figure 9.2, we see that

$$h = d_A(SE)\chi = d_A(SE)d_A(EL)\alpha/d_A(SL), \tag{9.2.2}$$

[475] G. H. Chen, C. S. Kochanek, and J. N. Hewitt, Astrophys. J. 447, 62 (1995).

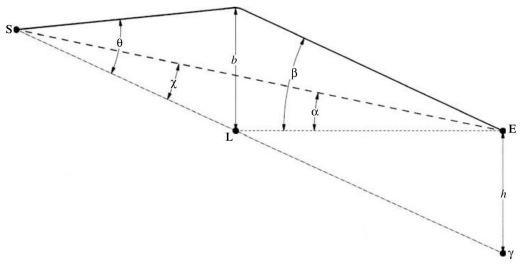


Figure 9.2: Quantities referred to in the derivation of the magnification formula, (9.2.4). The bent solid line is the path of a photon from the source S past the lens L to a point E on a telescope mirror. The point Y is a fixed point near the earth on the line from the source through the lens. The transverse distances b and h from the lens to the light path and from the point Y to the point E where the photon arrives at the telescope mirror are greatly exaggerated, as are the angles α, β, χ, and θ. As in Figure 9.1, this is drawn for the case $K = 0$; in a Robertson-Walker coordinate system centered on the earth, the path from the source to the bend would be curved for $|K| \neq 0$.

where χ is the angle between the line from the source to the point on the mirror and the line from the source to the lens, and

$$\theta = b/d_A(SL) = \beta d_A(EL)/d_A(SL). \tag{9.2.3}$$

(As in Section 9.1, for any points P and $Q, d_A(PQ)$ is the angular diameter distance of Q as seen from P.) Hence Eq. (9.2.1) gives

$$\ell = \ell_0 \left| \frac{\beta d\beta}{\alpha d\alpha} \right|, \tag{9.2.4}$$

where ℓ_0 is the luminosity that would be observed in the absence of the lens:

$$\ell_0 = \frac{L}{4\pi \left(1 + z_S\right)^2 d_A^2(SE)}. \tag{9.2.5}$$

(Note that $r_S(E) = r_E(S)$, so Eq. (9.1.2) gives $d_A(SE) = (1 + z_S)\, d_A(ES)$. According to Eq. (1.4.12), the luminosity distance of the source as seen from the earth is $d_L(S) = (1 + z_S)^2\, d_A(ES) = (1 + z_S)\, d_A(SE)$, which is the distance whose square appears in Eq. (9.2.5).)

Eq. (9.2.4) is a very general result, which applies to lenses and sources of all types. It has a particularly useful consequence for extended sources. If a
small part of the image of such a source subtends a solid angle $\beta\Delta\beta\Delta\phi$ and has an apparent luminosity $\Delta\ell$, then it has a surface brightness $\Delta\ell/\beta\Delta\beta\Delta\phi$. According to Eq. (9.2.4), this equals $\Delta\ell_0/\alpha\Delta\alpha\Delta\phi$, which is the surface brightness that this part of the image would have if there were no lens. Thus surface brightness is unaffected by lensing. This of course is because gravitational lensing changes neither the number nor the energy of photons.

Let us now specialize to the case of a point lens. The lens equation (9.1.5) gives here

$$\alpha = \beta - \beta_E^2/\beta, \tag{9.2.6}$$

so

$$\frac{\alpha d\alpha}{\beta d\beta} = 1 - \beta_E^4/\beta^4 \tag{9.2.7}$$

The luminosity (9.2.4) is then

$$\ell = \frac{\ell_0}{|1 - \beta_E^4/\beta^4|}. \tag{9.2.8}$$

When the distance [1]a of the lens from the line joining the source and observer is small, we have $\alpha \ll \beta_E$, so the two solutions (9.1.6) for β become

$$\beta_\pm \to \pm\beta_E + \frac{1}{2}\alpha. \tag{9.2.9}$$

In this case both images are amplified by a factor $|\beta_E/2\alpha|$. On the other hand, when the distance a of the lens from the line joining the source and observer is large, we have $\alpha \gg \beta_E$, so the two solutions (9.1.6) for β become

$$\beta_+ \to \alpha \gg \beta_E, \quad -\beta_- \to \beta_E^2/\alpha \ll \beta_E. \tag{9.2.10}$$

Under these conditions, the "-" ray becomes invisible, while the "+" ray has the normal brightness expected without gravitational deflection.

Roughly speaking, therefore, a point mass can only produce noticeable strong lensing, with more than one image, if it has $\alpha \leq \beta_E$, i.e., if it lies within a proper distance $a_{\max}$ of the line between the source and the observer, given by

$$a_{\max} = \beta_E d_A(EL) = \sqrt{\frac{4MGd_A(LS)d_A(EL)}{d_A(ES)}}. \tag{9.2.11}$$

Using Eq. (9.1.2), this is

$$a_{\max} = \sqrt{\frac{4MGr_L(S)r_E(L)a\left(t_L\right)}{r_E(S)}} = \sqrt{\frac{4MGr_L(S)r_E(L)a\left(t_E\right)}{r_E(S)}}\frac{1}{\sqrt{1+z_L}}, \tag{9.2.12}$$

where t_E is the time the light signal reaches the earth (elsewhere in this book called t_0), while z_L is the redshift of the lens

$$1 + z_L \equiv a\left(t_E\right)/a\left(t_L\right) \tag{9.2.13}$$

The radial coordinate $r_L(S)$ of the source in the Robertson-Walker coordinate system centered on the lens can be calculated in terms of the radial coordinates of both the source and lens in the Robertson-Walker coordinate system centered on the earth by using the condition that the threedimensional proper distance between the source and the lens is equal in these two coordinate systems

$$\int_0^{r_L(S)} \frac{dr}{\sqrt{1-Kr^2}} = \int_{r_E(L)}^{r_E(S)} \frac{dr}{\sqrt{1-Kr^2}} \tag{9.2.14}$$

This gives

$$r_L(S) = r_E(S)\sqrt{1 - Kr_E^2(L)} - r_E(L)\sqrt{1 - Kr_E^2(S)}. \tag{9.2.15}$$

If at time t there are $n(t, M)$ objects per proper volume with mass between M and $M + dM$, then (using Eq. (1.1.12)) the total number of objects that can produce a detectable splitting of the image of the source S is

[471] Here a is a proper distance transverse to the line of sight, and is not related to the Robertson-Walker scale factor, which is distinguished in this section by always writing it with a time argument, as $a(t)$.

$$N_S = \int_0^{r_E(S)} \frac{dr_E(L) a(t_L)}{\sqrt{1-Kr_E^2(L)}} \int_0^\infty \pi a_{\max}^2 n(t_L, M)\, dM$$
$$= 4\pi G a^2(t_E) \int_0^{r_E(S)} \frac{dr_E(L)\rho(t_L)}{(1+z_L)^2 \sqrt{1-Kr_E^2(L)}} \frac{r_L(S) r_E(L)}{r_E(S)} \tag{9.2.16}$$

where t_L is the time that the light from the source reaches the lens, and $\rho(t)$ is the mass density of possible lenses at time t :

$$\rho(t) \equiv \int_0^\infty n(t, M) M dM \tag{9.2.17}$$

Also, in the integrand $r_E(L)$ and t_L are related as usual by

$$\int_0^{r_E(L)} \frac{dr}{\sqrt{1-Kr^2}} = \int_{t_L}^{t_E} \frac{dt}{a(t)} \tag{9.2.18}$$

and $1 + z_L \equiv a(t_E)/a(t_L)$. If there is no evolution of the density of lensing galaxies, then $\rho(t_L) = \rho(t_E)(1+z_L)^3$, and Eq. (9.2.16) reads

$$N_S = 4\pi G a^2(t_E)\rho(t_E) \int_0^{r_E(S)} \frac{dr_E(L)(1+z_L)}{\sqrt{1-Kr_E^2(L)}} \left(\frac{r_L(S) r_E(L)}{r_E(S)} \right) \tag{9.2.19}$$

Where $N_S \ll 1$ we can ignore the possibility of multiple lensing, and interpret N_S as the probability that the image of the source S is appreciably modified by a lens near the line of sight.

For sources with $z_S \ll 1$, Eq. (9.2.19) simplifies to

$$N_S = 4\pi G a^2(t_E)\rho(t_E) \int_0^{r_E(S)} dr_E(L) \left(\frac{(r_E(S) - r_E(L)) r_E(L)}{r_E(S)} \right)$$
$$= \frac{2\pi G r_E^2(S) a^2(t_E)\rho(t_E)}{3} = \frac{2\pi G z_S^2 \rho(t_E)}{3H_0^2} = \frac{z_S^2 \Omega_L}{4} \tag{9.2.20}$$

where $\Omega_L \equiv 8\pi\rho(t_E)/3H_0^2$ is the fraction of the critical mass that at present is provided by the lensing objects. We see that even for Ω_L as large as $\Omega_M \approx 0.3$, the probability of strong lensing is small for nearby sources with $z_S \ll 1$, but the probability should become appreciable for sources at cosmological distances. [2]

For z_S of order unity or larger, the strong lensing probability turns out to depend sensitively on the cosmological model. Since the integral (9.2.19) is complicated in general, we will consider here just two extreme cases, of a flat universe dominated either by vacuum energy or by non-relativistic matter.

De Sitter model ($\Omega_\Lambda = 1, \Omega_K = \Omega_M = \Omega_R = 0$)
Here $a(t)/a(t_E) = \exp(H_0(t - t(E)))$, so

$$r_E(L) = \int_{t_L}^{t_E} \frac{dt}{a(t)} = \frac{1}{a(t_E) H_0} \left[\exp(H_0(t_E - t_L)) - 1 \right]$$

We can invert this, and find

$$a(t_L) = \frac{a(t_E)}{1 + a(t_E) H_0 r_E(L)}$$

Then Eq. (9.2.19) becomes

$$N_S = 4\pi G\rho\left(t_E\right) a^2\left(t_E\right) \times \int_0^{r_E(S)} \left(\frac{r_E(L)\left(r_E(S)-r_E(L)\right)}{r_E(S)}\right)\left(1+a\left(t_E\right)H_0 r_E(L)\right) dr_E(L)$$

It is convenient to change the variable of integration from $r_E(L)$ to the redshift z_L of the lens, which is given by

$$z_L = \left(\frac{a\left(t_E\right)}{a\left(t_L\right)}\right) - 1 = a\left(t_E\right) H_0 r_E(L)$$

so

$$\begin{aligned} N_S &= \frac{4\pi G\rho\left(t_E\right)}{H_0^2}\int_0^{z_S}\left(\frac{z_L\left(z_S-z_L\right)}{z_S}\right)\left(1+z_L\right)dz_L \\ &= \frac{\pi G\rho\left(t_E\right)}{3H_0^2}\left(z_S^3+2z_S^2\right) = \frac{\Omega_L}{8}\left(z_S^3+2z_S^2\right) \end{aligned} \tag{9.2.21}$$

Einstein-de Sitter model ($\Omega_M = 1, \Omega_K = \Omega_\Lambda = \Omega_R = 0$)
Here $a(t) = a\left(t_E\right)\left(t/t_E\right)^{2/3}$ and $1+z_L = \left(t_E/t_L\right)^{2/3}$, so

$$\begin{aligned} r_E(L) &= \int_{t_L}^{t_E}\frac{dt}{a(t)} = 3a^{-1}\left(t_E\right)\left(t_E - t_E^{2/3}t_L^{1/3}\right) \\ &= 3a^{-1}\left(t_E\right)t_E\left(1-\left(1+z_L\right)^{-1/2}\right) \end{aligned}$$

Hence Eq. (9.2.19) here becomes

$$\begin{aligned} N_S =& \frac{18\pi G\rho\left(t_E\right)t_E^2}{1-\left(1+z_S\right)^{-1/2}}\int_0^{z_S} dz_L\left(1-\left(1+z_L\right)^{-1/2}\right)\left(1+z_L\right)^{-1/2} \\ &\times\left(\left(1+z_L\right)^{-1/2}-\left(1+z_S\right)^{-1/2}\right) \\ =& \frac{8\pi G\rho\left(t_E\right)}{H_0^2}\left(-4+\frac{\sqrt{1+z_S}+1}{\sqrt{1+z_S}-1}\ln\left(1+z_S\right)\right) \\ =& 3\Omega_L\left(-4+\frac{\sqrt{1+z_S}+1}{\sqrt{1+z_S}-1}\ln\left(1+z_S\right)\right) \end{aligned} \tag{9.2.22}$$

This grows like z_S^2 for $z_S \ll 1$, in agreement with Eq. (9.2.20), but it flattens out for $z_S \gg 1$, growing only logarithmically with z_S.

We see that the probability of a given source being lensed increases much more rapidly with the redshift of the source for a vacuum-energy dominated model than for a matter-dominated model. For instance, for $z_S = 2$, Eq. (9.2.21) gives $N_S = 4\Omega_L$ for a vacuum-dominated model, while Eq. (9.2.22) gives $N_S = 0.30\Omega_L$ for a matter-dominated model.

Comparison of lensing theory (under the assumptions $K = 0$ and constant vacuum energy) with the number density of strong gravitational lenses measured as a function of redshift by the Cosmic Lens All Sky Survey [3] has given the result that

$$\Omega_M = 0.31^{+0.27}_{-0.14}(68\%\text{ stat. })^{+0.12}_{-0.1}\text{ (syst.)}$$

[472] W. H. Press and J. E. Gunn, Astrophys. J. 185, 397 (1973).

This is not yet competitive with other measurements of Ω_M, but it shows definite evidence of the effect of dark energy on the number of observed strong lenses.

The assumption that the lens can be approximated as a point mass is appropriate for so-called microlensing by stars in our galaxy. Note that for $d_A(EL) \approx 10^3\text{pc}, d_A(LS) \approx d_A(ES)$, and $MG \approx 1$ km $\approx 3\times10^{-14}$pc, the effective radius $\sqrt{4MGd_A(EL)d_A(LS)/d_A(ES)}$ of the point mass lens is of order 10^{-5}pc, which is much larger than the size of even a large star. Typically in microlensing observations a change is detected in the luminosity of a distant point source (such as a star outside our galaxy) as the star moves past the line of sight to the source. For a star that moves at 100 km/sec transverse to the line of sight, the time within which the source will be within the effective radius 10^{-5}pc of the star is about a month, which is convenient for monitoring changes in apparent luminosity. Between 13 and 17 microlensing events toward the Large Magellanic Cloud (LMC) have been seen by the MACHO collaboration, [4] and another three toward less crowded fields of the LMC by the EROS collaboration. [5] Of these, only four of the lenses have been identified, and all of them are in the LMC itself, rather than the halo of our galaxy. [6] This suggests though it does not prove that the mass of the halo does not consist of dark objects with the masses of typical stars. [7] Microlensing has also been extensively used in searches for dark stars and extra-solar planets by the Microlensing

10.9.3 9.3 Extended lenses

Observations in Astrophysics collaboration [8] and the Optical Gravitational Lensing Experiment collaboration. [9]

10.9.4 9.3 Extended lenses

The point source model is not at all valid for clusters of galaxies, and it is only marginally valid for individual galaxies. (Note that for $d_A(LS) \approx d_A(EL) \approx 10^{10}\text{pc}, d_A(ES) \approx 2\times10^{10}$pc, and $MG \approx 10^{12}$ km ≈ 0.03pc, the effective radius $\sqrt{4MGd_A(EL)d_A(LS)/d_A(ES)}$ of the point mass lens is of order 5×10^4pc, which is close to the size of the spherical halo of our own galaxy.) To deal with light rays that pass through the galaxy, it has become common to approximate the massive halos of these galaxies as spheres of matter in "isothermal" equilibrium, with a ratio of pressure p to mass density ρ given by the mean square value $\langle v^2\rangle$ of any one component of star velocity, assumed to be equal throughout the lens:

$$p(r) = \rho(r)\left\langle v^2\right\rangle. \tag{9.3.1}$$

We have already seen in Section 1.9 that the solution of the equation of hydrostatic equilibrium with this equation of state behaves at a large proper distance r from the center of the lens as

$$\rho(r) = \frac{2\rho(0)}{9\left(r/r_0\right)^2} = \frac{\langle v^2\rangle}{2\pi G r^2}, \tag{9.3.2}$$

which gives the mass contained within a sphere of radius r as

[473] K.-H. Chae et al., Phys. Rev. Lett. 89, 151301 (2002) [astro-ph/0209602].

[4] C. Alcock et al., Astrophys. J. 542, 281 (2000).

[5] T. Laserre et al., Astron. and Astrophys. 355, L39 (2000).

[6] N. W. Evans, in IDM 2002: The 4th International Workshop on the Identification of Dark Matter, eds. N. Spooner and V. Kudryavtsev (World Scientific) [astro-ph/0211302].

[7] For discussion of this issue, see A. F. Zakharov, Publ. Astron. Obs. Belgrade 74, 1 (2002) [astroph/0212009]; K. C. Sahu, proceedings of the STSci symposium Dark Universe: Matter, Energy, and Gravity [astro-ph/0302325].

$$\mathcal{M}(r) = \frac{2r\,\langle v^2\rangle}{G}. \tag{9.3.3}$$

The solution departs from this result for small r, but most of the mass of the lens is at distances from the center where it is a good approximation.

The rate of change of the unit vector $\hat{\mathbf{u}}$ giving the direction of a ray of light in a gravitational field $\mathbf{g} = -\nabla\phi$ is given in the post-Newtonian approximation by [1]

$$\frac{d\hat{\mathbf{u}}}{dt} = -2\hat{\mathbf{u}} \times (\hat{\mathbf{u}} \times \mathbf{g}) \tag{9.3.4}$$

For any spherically symmetric non-relativistic distribution of matter, we have

$$\mathbf{g}(\mathbf{r}) = -\frac{G\mathcal{M}(r)}{r^3}\mathbf{r}, \tag{9.3.5}$$

which for the outer parts of an isothermal sphere becomes

$$\mathbf{g}(\mathbf{r}) = -\frac{2\,\langle v^2\rangle}{r^2}\mathbf{r}, \tag{9.3.6}$$

For a light ray that passes the center of the lens at a distance b of closest approach, the light ray direction $\hat{\mathbf{u}}$ remains very close to a fixed direction. The small change in direction then has a magnitude

$$\gamma = 4\,\langle v^2\rangle \int_{-\infty}^{\infty} \frac{b dx}{x^2 + b^2} = 4\pi\,\langle v^2\rangle \tag{9.3.7}$$

and is independent of b. (We are taking the speed of light to be unity, so that $\langle v^2\rangle^{1/2}$ is dimensionless, equal to the rms velocity in units of the speed of light.) The direction of this change is toward the center of the lens, so light can arrive at the earth from the source along two different rays, which pass on opposite sides of the lens center, and are separated in angle by $8\pi\,\langle v^2\rangle$. The "+" ray, which passes the lens on the side toward the source, makes an angle β_+at the earth with the ray from the earth to the lens given again by Eq. (9.1.3), while the '-' ray makes an angle with the earth-lens ray β_- (now taken positive), given by replacing α with $-\alpha$ in Eq. (9.1.3), so

$$\gamma d_A(LS) = (\beta_\pm \mp \alpha)\, d_A(ES) \tag{9.3.8}$$

which with Eq. (9.3.7) gives us our lens equation

$$\beta_\pm = \pm\alpha + \beta_E \tag{9.3.9}$$

where now

$$\beta_E = \frac{4\pi\,\langle v^2\rangle\, d_A(LS)}{d_A(ES)} \tag{9.3.10}$$

Here again for a lens on the line of sight from the earth to the source we have $\alpha = 0$, and the two images become an Einstein ring with angular radius β_E. Because we have now defined $\beta_\pm$to be positive the lens equation gives two images only if $\alpha < \beta_E$, which requires that the proper distance $a = d_A(EL)\alpha$ of the lens from the line between the earth and the source be less than a maximum value

[478] P. Yock et al., Proceedings of the ninth Marcel Grossman meeting, Rome, July 2000 [astroph/0007317].
[9] A. Udalski et al., Acta Astron. 54, 313 (2004) [astro-ph/0411543].
[1] See, e.g., G&C, Eq. (9.2.7). A factor -2 that was missing on the right-hand side of this equation has been supplied.

$$a_{\max} = \beta_E d_A(EL) = \frac{4\pi \langle v^2 \rangle d_A(LS) d_A(EL)}{d_A(ES)} \tag{9.3.11}$$

10.9.5 9.3 Extended lenses

The magnification of the two images is again given by Eq. (9.2.4)

$$\ell = \ell_0 \left| \frac{\beta d\beta}{\alpha d\alpha} \right|,$$

but now we must use this with the lens equation (9.3.9), which gives

$$\ell_\pm = \ell_0 \left| \frac{\beta_\pm}{\alpha} \right| = \ell_0 \left| \frac{\beta_\pm}{\beta_\pm - \beta_E} \right|. \tag{9.3.12}$$

The magnification becomes infinite as the lens approaches the line of sight between the earth and the source, in which case $\alpha \to 0, \beta_\pm \to \beta_E$, and the image pair becomes an Einstein ring. On the other hand, for a lens approaching the maximum distance (9.3.11) from the earth-source line, we have $\alpha \to \beta_E$, so $\beta_+ \to 2\alpha$ and $\beta_- \to 0$. According to Eq. (9.3.12), the ' - ' ray disappears in this limit, while the apparent luminosity of the '+' ray is doubled.

The number of possible isothermal spherical lenses that are actually able to split the image of a point source is now given by

$$\begin{aligned} N_S &= \int_0^{r_E(S)} \frac{dr_E(L) a(t_L)}{\sqrt{1 - K r_E^2(L)}} \pi a_{\max}^2 n(t_L) \\ &= 16\pi^3 \langle v^2 \rangle^2 \int_0^{r_E(S)} dr_E(L) \frac{n(t_L) a^3(t_L)}{\sqrt{1 - K r_E^2(L)}} \left(\frac{r_L(S) r_E(L)}{r_E(S)} \right)^2 \end{aligned} \tag{9.3.13}$$

where $n(t)$ is the proper number density of these lenses at time t. For simplicity, we have here taken $\langle v^2 \rangle$ the same for all lenses. If we also now take $K = 0$, so that Eq. (9.2.15) gives $r_S(L) = r_E(S) - r_E(L)$, and also assume no evolution of the population of lensing galaxies, so that $n(t)a^3(t)$ is constant, then

$$\begin{aligned} N_S &= 16\pi^3 \langle v^2 \rangle^2 n(t_E) a(t_E)^3 \int_0^{r_E(S)} \left(\frac{(r_E(S) - r_E(L)) r_E(L)}{r_E(S)} \right)^2 dr_E(L) \\ &= \frac{8\pi^2 \langle v^2 \rangle^2 n(t_E) a(t_E)^3 r_E^3(S)}{15} \\ &= \frac{8\pi^2 \langle v^2 \rangle^2 n(t_E)}{15 H_0^3} \mathcal{F}(z_S) \end{aligned} \tag{9.3.14}$$

where

$$\mathcal{F}(z) \equiv (a(t_E) H_0 r_E(S))^3 = \left(\int_1^{1+z} \frac{dy}{\sqrt{\Omega_\Lambda + \Omega_M y^3 + \Omega_R y^4}} \right)^3 \tag{9.3.15}$$

This shows that the chance of a source being lensed depends strongly on Ω_Λ. For instance, for $\Omega_\Lambda = \Omega_R = 0$ and $\Omega_M = 1$ we have $\mathcal{F}(z) = 8(1 - 1/\sqrt{1+z})^3$, which approaches the constant 8 for large z, while for $\Omega_M = \Omega_R = 0$ and $\Omega_\Lambda = 1$ we have $\mathcal{F}(z) = z^3$. The lensing probability for a source at redshift $z = 3$ is 27 times greater if the cosmic energy density is

vacuum dominated than if it is matter dominated. From direct observations of galaxies, the constant in Eq. (9.3.15) has been estimated as [2]

$$\frac{8\pi^2 \left\langle v^2 \right\rangle^2 n\left(t_E\right)}{15 H_0^3} \simeq 0.02 \tag{9.3.16}$$

so for $\Omega_M = \Omega_R = 0$ and $\Omega_\Lambda = 1$ the probability of lensing should become large when $z \approx 3.7$.

Lensing probabilities for quasars of large redshift can provide sensitive limits on Ω_Λ.[3] At one time it was thought on the basis of early surveys that strong lensing statistics ruled out the possibility that a constant vacuum energy made a dominant contribution to producing a spatially flat universe. [4] Since then, many groups have carried out such studies, with corrections for finite core radii, selection effects, etc., to set limits on Ω_Λ in flat cosmologies with $\Omega_\Lambda + \Omega_M = 1$. They find $\Omega_\Lambda < 0.9$,[5] $\Omega_\Lambda < 0.7$[6] $\Omega_\Lambda = 0.64^{+0.157}_{-0.26} \Omega_\Lambda < 0.74$,[8] $\Omega_\Lambda < 0.79$,[9] $\Omega_\Lambda = 0.7^{+0.1}_{-0.2}$,[10] and $\Omega_\Lambda \approx 0.45 - 0.75$.[11]

The deflection $\Delta\theta$ calculated in the isothermal sphere model of gravitational lenses is relevant only if the light path actually passes through the sphere. For galaxies of relatively small mass, we may have one or both of the deflected rays passing outside the galactic radius R. The deflection of light depends only on the Newtonian gravitational potential, and, as Newton showed, the gravitational potential outside a spherically symmetric distribution of matter is just the same as if all the mass were concentrated at the center of symmetry, so the motion of a light ray outside the galaxy is described by the point mass lens model considered earlier. If the light ray passes far outside the effective radius of the galaxy we have so-called weak lensing, which is typically discovered statistically, through a correlation in

the orientation of the images of several lensed galaxies, rather than by the study of individual sources. This is the subject of Section 9.5.

10.9.6 9.4 Time delay

In addition to light rays being bent by the gravitational fields of intervening objects, they are also delayed, [1] so that a fluctuation in the distant light source appears at different times on earth in the several lensed images of the source. [2] There are two effects here.

First, there is a geometrical time delay, caused by the increased length of the total light path from the source to the earth. Since this arises over very long distances, it can be calculated by idealizing the light path as a geodesic of the Robertson-Walker metric from the source to the point P of closest approach to the lens, where the light path is bent, followed by a similar geodesic from that point to the earth. The time t_E that a light signal that leaves the source at time t_S arrives at the earth is given by

$$\int_{t_S}^{t_E} \frac{dt}{a(t)} = \sigma_{SP} + \sigma_{PE} \tag{9.4.1}$$

[472] E. L. Turner, J. P. Ostriker, and J. R. Gott, Astrophys. J. 284, 1 (1984).
[3] E. L. Turner, Astrophys. J. 242, L135 (1980).
[4] E. L. Turner, Astrophys. J. 365, L43 (1990).
[5] M. Fukugita and E. L. Turner, Mon. Not. Royal Astron. Soc. 253, 99 (1991).
[6] D. Maoz and H-W. Rix, Astrophys. J. 416, 425 (1993)
[7] M. Im, R. E. Griffiths, and K. U. Ratnatunga, Astrophys. J. 475, 457 (1997).
[8] E. E. Falco, C. S. Kochanek, and J. A. Muñoz, Astrophys. J. 494, 47 (1998).
[9] A. R. Cooray, J. M. Quashnock, and M. C. Miller, Astrophys. J. 511, 562 (1999) [astro-ph/9806080]; A. R. Cooray, Astron. Astrophys. 342, 353 (1999) [astro-ph/9811448].
[10] M. Chiba and Y. Yoshii, Astrophys. J. 510, 42 (1999) [astro-ph/9808321].
[11] Y-C. N. Cheng and L. M. Krauss, Astrophys. J. 511, 612 (1999) [astro-ph/9810392].

where σ_{SP} and σ_{PE} are the proper lengths (the integrals of $(\tilde{g}_{ij}dx^i dx^j)^{1/2}$) along the paths from the source to P and from P to the earth, respectively. The time t_{E0} that the light would arrive at the earth if there were no gravitational deflection is given simply by

$$\int_{t_S}^{t_{E0}} \frac{dt}{a(t)} = \sigma_{SE} \tag{9.4.2}$$

where σ_{SE} is the proper length along the geodesic from the source to the earth. Hence the geometric time delay (which is always very short compared to a Hubble time) is

$$\Delta t_{\text{geom}} \equiv t_E - t_{E0} = a\left(t_E\right)\left(\sigma_{SP} + \sigma_{PE} - \sigma_{SE}\right) \tag{9.4.3}$$

Now, it is easy to calculate proper lengths along geodesics that end at the earth:

$$\sigma_{SE} = \int_0^{r_E(S)} \frac{dr}{\sqrt{1-Kr^2}}, \quad \sigma_{PE} = \int_0^{r_E(P)} \frac{dr}{\sqrt{1-Kr^2}} \tag{9.4.4}$$

where $r_E(S)$ and $r_E(P)$ are the usual radial coordinates (in a RobertsonWalker coordinate system centered on the earth) of the source S and the point P of closest approach of the light to the lens. To calculate σ_{SP} we have to do a little more work. By using Eq. (1.1.17) for the spatial affine connection Γ^i_{jl}, we see that the equation (1.1.26) of a spatial geodesic is

$$\frac{d^2x^i}{d\sigma^2} + Kx^i = 0 \tag{9.4.5}$$

The general solution is

$$x^i(\sigma) = \begin{cases} A^i \cos\sigma + B^i \sin\sigma, & K = +1 \\ A^i + B^i\sigma, & K = 0 \\ A^i \cosh\sigma + B^i \sinh\sigma, & K = -1 \end{cases} \tag{9.4.6}$$

where A^i and B^i are constants characterizing different paths. These constants are subject to a normalization condition

$$1 = \tilde{g}_{ij}\frac{dx^i}{d\sigma}\frac{dx^j}{d\sigma} = \left(\frac{d\mathbf{x}}{d\sigma}\right)^2 + K\frac{(\mathbf{x}\cdot d\mathbf{x}/d\sigma)^2}{1-K\mathbf{x}^2}$$

which gives

$$\left(1-K\mathbf{A}^2\right)\left(1-\mathbf{B}^2\right) = K(\mathbf{A}\cdot\mathbf{B})^2. \tag{9.4.7}$$

We can determine the constants $\mathbf{A}$ and $\mathbf{B}$ for the geodesic from the source to the point P by requiring that $x^i(0) = r_E(S)\hat{n}_{ES}$ and $x^i\left(\sigma_{SP}\right) = r_E(P)\hat{n}_{EP}$, where $\hat{n}_{ES}$ and $\hat{n}_{EP}$ are the unit vectors from the earth to the source or point P, respectively. Imposing the normalization condition (9.4.7) then gives

$$\sigma_{SP} = \begin{cases} \cos^{-1}\left[r_E(P)r_E(S)\cos\theta + \sqrt{1-r_E(P)^2}\sqrt{1-r_E(S)^2}\right] & K = +1 \\ \sqrt{r_E^2(P) + r_E^2(S) - 2r_E(P)r_E(S)\cos\theta} & K = 0 \\ \cosh^{-1}\left[-r_E(P)r_E(S)\cos\theta + \sqrt{1+r_E(P)^2}\sqrt{1+r_E(S)^2}\right] & K = -1 \end{cases} \tag{9.4.8}$$

[471] The time delay of radar reflections from planets and of radio signals from artificial satellites caused by the gravitational potential of the sun has provided a fourth test of general relativity, as first proposed and measured by I. Shapiro, Phys. Rev. Lett. 13, 789 (1964). For a discussion, see G&C, Section 8.7.

[2] S. Refsdal, Mon. Not. Roy. Astron. Soc. 128, 307 (1964).

where θ is the angle between the directions from the earth to the source and to the point P. This result applies for a geodesic triangle with arbitrary angles, but of course we are interested here in the case where θ is very small. In this case, Eq. (9.4.8) becomes

$$\sigma_{SP} = \begin{cases} \sigma_{ES} - \sigma_{EP} + \frac{r_E(P) r_E(S) \theta^2}{2 \sin(\sigma_{ES} - \sigma_{EP})} & K = +1 \\ \sigma_{ES} - \sigma_{EP} + \frac{r_E(P) r_E(S) \theta^2}{2 (r_E(S) - r_E(P))} & K = 0 \\ \sigma_{ES} - \sigma_{EP} + \frac{r_E(P) r_E(S) \theta^2}{2 \sinh(\sigma_{ES} - \sigma_{EP})} & K = -1 \end{cases}$$

In the terms proportional to θ^2 we can ignore the separation of the lens from the point P of closest approach, taking $r_E(P) = r_E(L); \theta = \beta - \alpha$; and (using Eq. (9.2.14)) $\sigma_{ES} - \sigma_{EP} = \sigma_{LS}$, with errors that would introduce terms of higher order in θ. Thus, in all three cases

$$\sigma_{SP} = \sigma_{ES} - \sigma_{EP} + \frac{r_E(L) r_E(S) \theta^2}{2 r_L(S)}. \tag{9.4.9}$$

The geometric time delay (9.4.3) is then

$$\Delta t_{\text{geom}} = \frac{a\left(t_E\right) r_E(L) r_E(S) (\beta - \alpha)^2}{2 r_L(S)}. \tag{9.4.10}$$

Using Eq. (9.1.2), we can write this in terms of angular diameter distances and the redshift of the lens:

$$\Delta t_{\text{geom}} = \frac{\left(1 + z_L\right) d_A(EL) d_A(ES) (\beta - \alpha)^2}{2 d_A(LS)}. \tag{9.4.11}$$

The details of the lens enter here only in the lens equation, which for each image gives the unobservable angle α in terms of the observable angle β.

There is also a potential time delay, caused directly by the motion of the light through the gravitational potential of the lens. The calculation of these time differences is generally done on a case-by-case basis, using detailed models of the lensing galaxy rather than either the point source or isothermal sphere models. Here we will consider only the case of a general spherically symmetric lens. In a coordinate system centered on the lens with line element in the 'standard' form

$$d\tau^2 = B(r) dt^2 - A(r) dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta d\varphi^2, \tag{9.4.12}$$

the time required for light to travel from a large coordinate distance r to a distance b of closest approach to the lens and then out again to r is [3]

$$2t(r, b) = 2 \int_b^r \left(\frac{A(r)/B(r)}{1 - (b/r)^2 (B(r)/B(b))} \right)^{1/2} dr \tag{9.4.13}$$

For the weak gravitational fields that concern us here, we can use the postNewtonian approximation, [4] which gives

$$A(r) = 1 + 2r\phi'(r), \quad B(r) = 1 + 2\phi(r), \tag{9.4.14}$$

where $\phi(r)$ is the Newtonian gravitational potential. Then to first order in ϕ

[3] 473 G&C, Eq. (8.7.2)

[4] G&C, Eqs. (9.1.57) and (9.1.60). It is necessary to redefine the radial coordinate in order to put the line element given by the post-Newtonian approximation in the form (9.4.12).

$$\begin{aligned} 2t(r,b) =& 2\sqrt{r^2-b^2} + 2\int_b^r \left(1-\left(\frac{b}{r}\right)^2\right)^{-1/2} (r\phi'(r)-\phi(r))\,dr \\ &+ 2\int_b^r \left(1-\left(\frac{b}{r}\right)^2\right)^{-3/2} \left(\frac{b}{r}\right)^2 (\phi(r)-\phi(b))dr \end{aligned} \tag{9.4.15}$$

We will be interested in this in the case $r \to \infty$, because the effects of the finite distance between the lens and the source and earth are already included in the geometric time delay. The second integral in Eq. (9.4.15) converges for $r \to \infty$, but the first integral diverges. Noting that for any kind of lens of mass M, the Newtonian potential at large distances goes as $-MG/r$, we can put the limit of Eq. (9.4.15) for $r \to \infty$ in the form

$$2t(t,b) \to 2r + 2MG\ln(2r) + f(b) \tag{9.4.16}$$

where

$$\begin{aligned} f(b) =& -2MG\ln b + 2\int_b^\infty \left(1-\left(\frac{b}{r}\right)^2\right)^{-1/2} \left(r\phi'(r)-\phi(r)-\frac{2MG}{r}\right) dr \\ &+ 2\int_b^\infty \left(1-\left(\frac{b}{r}\right)^2\right)^{-3/2} \left(\frac{b}{r}\right)^2 (\phi(r)-\phi(b))dr \end{aligned} \tag{9.4.17}$$

The first two terms in Eq. (9.4.16) diverge for $r \to \infty$, but they are independent of b, so when we calculate the time difference between the arrival of fluctuations in different images of the source we need only to take account of the differences in $f(b)$ for various values of b. But this would give the delay in the time used in the metric (9.4.12), which is the time told by clocks that are far enough from the lens to ignore the lens's gravitational potential, but close enough to ignore cosmological effects. Because of the cosmological redshift, the time delay observed on earth is lengthened by a factor $1+z_L$, so the potential time delay is

$$\Delta t_{\rm pot} = (1+z_L)\,f(b). \tag{9.4.18}$$

For instance, if the lens is a point mass then $\phi = -MG/r$, so the first integral in Eq. (9.4.17) vanishes and the second integral is independent of b, so the gravitational potential of the lens produces a difference in time between the arrival of fluctuations in images of the source at angles β_1 and β_2 to the image of the lens, given by

$$\begin{aligned} \Delta t_{\rm pot}(\beta_1) - \Delta t_{\rm pot}(\beta_2) &= 2MG(1+z_L)\ln(b_2/b_1) \\ &= 2MG(1+z_L)\ln(\beta_2/\beta_1)\,. \end{aligned}$$

Evidently, the potential time delay will generally be of order MG. As shown for instance in Eq. (9.1.5), if $d_A(EL), d_A(ES)$, and $d_A(LS)$ are all of the same order d, then for strong lensing the squared angle $(\beta-\alpha)^2$ is generally of order MG/d, so the geometric time delay (9.4.11) is also of order MG. For a galaxy of 10^{11} solar masses this is 5.7 days. Measurement of this delay can be used to measure the mass of the lens, without needing to know the Hubble constant. We have seen that the measurement of the angles between the source image and the lens can tell us H_0MG, while the measurement of these angles and time delays can tell us MG, so the combination of angular and time-delay measurements can yield a value of H_0. Here is a list of several lenses that have been used in this way:

- QSO 0957+561: Two images separated by $6.1''$, $z_L = 1.41, z_S = 0.36$, with time difference 417 ± 3 days (95%), gave $H_0 = 64 \pm 13$ kmsec^{-1}Mpc^{-1}, [5] subsequently recalculated [6] as 77^{+29}_{-24} kmsec^{-1}Mpc^{-1}.

- B 0218+357 : Two images separated by $0.335''$ and an Einstein ring, $z_S = 0.96, z_L = 0.68$, time difference 10.5 ± 0.4 days, gave $H_0 = 69^{+13}_{-19}$kmsec^{-1}Mpc^{-1}.7

- PKS 1830-211: Two images separated by $1.0''$, plus an Einstein ring, $z_L = 0.89$, time difference 26^{+4}_{-5} days (8.6 GHz). [8] Source redshift measured as $2.507 \pm .002$, gave $H_0 = 65^{+15}_{-9}$ kmsec^{-1}Mpc^{-1} for $\Omega_M = 1$, $\Omega_\Lambda = 0, H_0 = 76^{+19}_{-10}$ kmsec^{-1}Mpc^{-1} for $\Omega_M = 0.3, \Omega_\Lambda = 0.7$.[9]

- B 1608+656: four lensed images, with three time differences from four to ten weeks, gave $H_0 = 75$ kmsec^{-1}Mpc$^{-1} \pm 10\%$.[10]

A table of time delay measurements up to 2003 is given in the review article of Kochanek and Schecter. [11] In 2006, a survey [12] of time delays in ten gravitational lenses (under the assumptions that $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$, and with a number of parameters for each lens found by requiring that all give the same Hubble constant) gave $H_0 = 75^{+7}_{-12}$ kmsec^{-1}Mpc^{-1}.

10.9.7 9.5 Weak lensing

For cosmology, the most promising application of gravitational lenses probably lies in surveys of weak lensing, the study of the distortion of the images of distant galaxies by numerous small deflections of light as it passes from the galaxies to us through a slightly inhomogeneous distribution of matter. Consider a ray of light from a point source on a distant galaxy at co-moving coordinates $r_S\hat{n}$, where $\hat{n}$ is a unit vector. Whether the intervening lenses are galaxies or clusters of galaxies or concentrations of intergalactic matter, we can think of the total deflection of the light ray as the sum of deflections caused by encounters with point lenses L of mass m_L, at positions $\mathbf{x}_L$. (Since we will be considering only terms of first order in the total deflection, we can think of the lenses as individual particles, even if they are aggregated into extended objects.) For weak lensing, the angle α_L between the directions to the source and the lens is much larger than the "Einstein ring" parameter β_{EL} for lens L, so as we have seen there is only a single image, given by taking the + sign in Eq. (9.1.6), which with $\alpha \ll \beta_E$ gives the amount of the deflection caused by the lens as $\beta_{+L} - \alpha_L = \beta_{EL}^2/\alpha_L$. We decompose the co-moving lens coordinate vector $\mathbf{x}_L$ into components parallel and perpendicular to the light ray:

$$\mathbf{x}_L = r_L\hat{n} + \mathbf{y}_L, \quad r_L \equiv \hat{n} \cdot \mathbf{x}_L, \quad \mathbf{y}_L \equiv \mathbf{x}_L - \hat{n}\,(\hat{n} \cdot \mathbf{x}_L) \tag{9.5.1}$$

The lens L deflects the light ray in a direction $-\hat{y}_L$, so the total deflection caused by all the lenses is

$$\Delta\hat{n} = -\sum_L \hat{y}_L\beta_{EL}^2/\alpha_L = -\sum_L \hat{y}_L \frac{4M_L G d_A(LS)}{d_A(ES)d_A(EL)\alpha_L} \tag{9.5.2}$$

[475] T. Kundic et al., Astrophys. J. 482, 75 (1997) [astro-ph/9610162].
[6] G. Bernstein and P. Fischer, Astron. J. 118, 14 (1999).
[7] A. D. Biggs et al., Mon. Not. Roy. Astron. Soc. 304, 349 (1999) [astro-ph/9811282.
[8] J. E. J. Lovell et al., Astrophys. J. 508, L51 (1998) [astro-ph/9809301]
[9] C. Lidman et al., Astrophys. J. 514, L57 (1999) [astro-ph/9902317].
[10] L. V. E. Koopmans, T. Treu, C. D. Fassnacht, R. D. Blandford, and G. Surpi, Astrophys. J. 599, 70 (2003).
[11] C. S. Kochanek and P. L. Schecter, Carnegie Observatories Astrophysics Series, Vol. 2, ed. W.L. Freedman (Cambridge University Press, 2003) [astro-ph/0306040]: Table 1.1.
[12] P. Saha, J. Coles, A. V. Maccio', and L. L. R. Williams, Astrophys. J. 660, L17 (2006) [astroph/0607240].

where for any points P and Q, $d_A(PQ)$ is the angular diameter distance of Q as seen from P. According to the definition of the angular diameter distance $d_A(EL)$ of the lens as seen from the earth, the angle between the directions to the source and the lens is $\alpha_L = a_L |\mathbf{y}_L| / d_A(EL)$, where a_L is the Robertson-Walker scale factor at the time that the light passes the lens L. Also, it is convenient to use Eq. (9.1.2) to write $d_A(LS) = a(t_S)\, r(r_L, r_S)$ and $d_A(ES) = a(t_S)\, r_S$, where t_S is the time the light leaves the source, and $r(r_L, r_S)$ and r_S are the radial coordinates of the source in RobertsonWalker coordinate systems centered on the lens and the earth, respectively. According to Eq. (9.2.15),

$$r(r_L, r_S) = r_S\sqrt{1 - Kr_L^2} - r_L\sqrt{1 - Kr_S^2} \tag{9.5.3}$$

The deflection is then

$$\Delta\hat{n} = -\sum_L \mathbf{y}_L \frac{4M_L G r(r_L, r_S)}{r_S a_L |\mathbf{y}_L|^2} \tag{9.5.4}$$

This is not in itself a useful result, because we do not generally know where the image of a point on the source would be if there were no lenses. It is more interesting to consider the variation of the deflection with position of the ray origin. Suppose we consider a small change $\boldsymbol{\theta}$ in the undeflected direction $\hat{n}$ to the ray origin, with $\boldsymbol{\theta}$ perpendicular to $\hat{n}$ and $|\boldsymbol{\theta}| \ll 1$. The change in the vector $\mathbf{y}_L$ from the light ray to lens L at the point of closest approach due to the displacement of the source is then

$$\delta\mathbf{y}_L = -\boldsymbol{\theta}(\hat{n} \cdot \mathbf{x}_L) - \hat{n}(\boldsymbol{\theta} \cdot \mathbf{x}_L)$$

Dropping the term proportional to $\hat{n}$, we see that to first order the change in the deflection of the image normal to the line of sight is

$$\Delta\theta_a = \sum_a M_{ab}(r_S, \hat{n})\, \theta_b \tag{9.5.5}$$

with a and b running over two orthogonal directions normal to $\hat{n}$, and $M_{ab}(r_S, \hat{n})$ the 2×2 shear matrix for images of a source at a distance r_S :

$$M_{ab}(r_S, \hat{n}) = \sum_L \frac{4M_L G r(r_L, r_S)\, r_L}{r_S a_L} \left(\frac{\delta_{ab}}{|\mathbf{y}_L|^2} - \frac{2y_{La}y_{Lb}}{|\mathbf{y}_L|^4} \right) \tag{9.5.6}$$

It is conventional to write this matrix as

$$M = \begin{pmatrix} \kappa + \gamma_1 & \gamma_2 \\ \gamma_2 & \kappa - \gamma_1 \end{pmatrix}, \tag{9.5.7}$$

with κ called the convergence and γ_i known as the shear field.
We may conveniently rewrite the shear matrix in terms of the perturbation to the total Newtonian potential due to the lenses: [1]
$\delta\phi(\mathbf{x}, t) = -\sum_L \frac{M_L G}{a_L |\mathbf{x} - \mathbf{x}_L(t)|} = -\sum_L \frac{M_L G}{a_L \left((r - r_L(t))^2 + (\mathbf{y} - \mathbf{y}_L(t))^2\right)^{1/2}}$,

where $\mathbf{x} = r\hat{n} + \mathbf{y}$ with $\hat{n} \cdot \mathbf{y} = 0$. (We are here explicitly taking into account the fact that the lenses may have a time dependence, as for instance from peculiar motion or cosmological evolution. In Eq. (9.5.6) we are using the abbreviations

$$r_L \equiv r_L(t_{r_L}), \quad \mathbf{y}_L \equiv \mathbf{y}_L(t_{r_L})$$

[471] We can neglect the effect of lenses at a cosmological distance from the light ray, so the denominator, including the factor a_L, is the proper distance to the lens L. We are here representing the complete perturbation to the mass density as a set of lensing point masses, but since these may be particles as well as galaxies, this introduces no loss of generality.

where t_r is the time that a light wave that reaches us at the present moment was at a radial coordinate r, given by $r = \int_0^{t_r} dt/a(t)$.) In particular, the second transverse derivatives of the potential on the light ray are

$$\left[\frac{\partial^2}{\partial y_a \partial y_b}\delta\phi(r\hat{n}+\mathbf{y},t)\right]_{\mathbf{y}=0} = \sum_L \frac{M_L G}{a_L}\left(\frac{\delta_{ab}}{\left((r-r_L(t))^2+|\mathbf{y}_L(t)|^2\right)^{3/2}} - \frac{3y_{La}(t)y_{Lb}(t)}{\left((r-r_L(t))^2+|\mathbf{y}_L(t)|^2\right)^{5/2}}\right)$$

The functions $\left((r-r_L)^2+|\mathbf{y}_L|^2\right)^{-3/2}$ and $\left((r-r_L)^2+|\mathbf{y}_L|^2\right)^{-5/2}$ are sharply peaked at $r = r_L$, so the integral of this second derivative times any smooth function $f(r)$ (smooth, in the sense that it varies little when r varies by an amount of order $|\mathbf{y}_L|$) is

$$\int f(r)dr\left[\frac{\partial^2}{\partial y_a \partial y_b}\delta\phi(r\hat{n}+\mathbf{y},t)\right]_{\mathbf{y}=0,t=t_r} = \sum_L \frac{f(r_L)\,M_L G}{a_L}\left(\frac{2\delta_{ab}}{|\mathbf{y}_L|^2} - \frac{4y_{La}y_{Lb}}{|\mathbf{y}_L|^4}\right)$$

Comparing this with Eq. (9.5.6), we see that the shear matrix M_{ab} for a source at radial coordinate r_S is

$$M_{ab}(r_S,\hat{n}) = 2\int_0^{r_S}\frac{r(r,r_S)\,r}{r_S}\left[\frac{\partial^2}{\partial y_a \partial y_b}\delta\phi(r\hat{n}+\mathbf{y},t)\right]_{\mathbf{y}=0,t=t_r} dr \tag{9.5.9}$$

So measurement of the shear matrix can yield information about perturbations to the gravitational potential by masses al a the line of sight. In particular, by contracting the indices a and b, we see that

$$\kappa = \frac{1}{2}\,\mathrm{Tr}\,M = \int_0^{r_S}\frac{r(r,r_S)\,r}{r_S}\left[\left(\nabla^2 - \frac{\partial^2}{\partial r^2}\right)\delta\phi(r\hat{n}+\mathbf{y},t)\right]_{\mathbf{y}=0,t=t_r} dr$$

If the lensing is due to a collection of bodies (such as a cluster of galaxies), all at about the same radial coordinate r_L, then $\delta\phi$ falls off rapidly for large values of $|r-r_L|$, so that the factor $r(r,r_S)\,r$ can be replaced with $r(r_L,r_S)\,r_L$, and the integral of $\partial^2\delta\phi/\partial^2 r$ can then be set equal to zero. Using the Poisson equation $a^{-2}\nabla^2\delta\phi = 4\pi G\delta\rho_M$, we can write the resulting expression in terms of the density fluctuation $\delta\rho_M$:

$$\kappa = \frac{4\pi G a^2(t_{r_L})\,d_A(LS)d_A(EL)}{d_A(ES)}\int \delta\rho_M(r\hat{n},t_{r_L})\,a(t_{r_L})\,dr \tag{9.5.10}$$

the integral being taken over a range of r passing through the lens. Hence, assuming that we know the angular diameter distances appearing in Eq. (9.5.10), a measurement of the value of κ for sources seen in some direction can tell us the total mass density of a cluster of lensing masses that lie along that line of sight, projected onto the plane normal to the line of sight.

Now let's consider how the lensing of the images of galaxies can be used to measure the shear matrix. In order to deal with galaxies that are irregularly shaped, it is convenient to describe the shape of a galaxy by a quadrupole matrix:

$$Q_{ab} \equiv \frac{\int d^2\theta\mathcal{L}(\boldsymbol{\theta})\theta_a\theta_b}{\int d^2\theta\mathcal{L}(\boldsymbol{\theta})} \tag{9.5.11}$$

Here the integral is over the transverse displacement $\boldsymbol{\theta}$ of the direction of points on the image of the galaxy from some central point; the indices a, b, etc. run over two orthogonal

directions in the plane of this image; and $\mathcal{L}$ is the surface brightness-the apparent luminosity per solid angle-of the image. (In order for the integral in the numerator to converge, it may be necessary to replace $\mathcal{L}$ by some function of $\mathcal{L}$, such as one that equals $\mathcal{L}$ when $\mathcal{L}$ is above some threshold brightness, and otherwise vanishes. This has no effect on the following analysis.) As remarked in Section 9.2, the surface brightness of any point is the same as would be seen in the absence of lensing from the same point on the source, so

$$\mathcal{L}((1+M)\boldsymbol{\theta}) = \mathcal{L}^s(\boldsymbol{\theta}) \tag{9.5.12}$$

where $\mathcal{L}^s$ is the surface brightness of the source in the absence of lensing. Introducing a new variable of integration $\boldsymbol{\theta}'$ in Eq. (9.5.11) by writing $\theta_a = (1+M)_{ab}\theta'_b$ and using Eq. (9.5.12), we have

$$\int d^2\theta \mathcal{L}(\boldsymbol{\theta})\theta_a\theta_b = (1+M)_{ac}(1+M)_{bd}\,\mathrm{Det}(1+M)\int d^2\theta' \mathcal{L}^s\left(\boldsymbol{\theta}'\right)\theta'_c\theta'_d$$

and likewise

$$\int d^2\theta \mathcal{L}(\boldsymbol{\theta}) = \mathrm{Det}(1+M)\int d^2\theta' \mathcal{L}^s\left(\boldsymbol{\theta}'\right)$$

The determinant cancels in the ratio, so (dropping primes) we have

$$Q_{ab} = (1+M)_{ac}(1+M)_{bd}Q^s_{cd} \tag{9.5.13}$$

where Q^s is the quadrupole matrix in the absence of lensing

$$Q^s_{ab} \equiv \frac{\int d^2\theta \mathcal{L}^s(\boldsymbol{\theta})\theta_a\theta_b}{\int d^2\theta \mathcal{L}^s(\boldsymbol{\theta})} \tag{9.5.14}$$

For simplicity of presentation, we will limit ourselves here to the case of lensing that is sufficiently weak so that all elements of M have $|M_{ab}| \ll 1$, the case of greatest interest in cosmology. In this case (9.5.13) becomes

$$Q_{ab} = Q^s_{ab} + M_{ac}Q^s_{cb} + M_{db}Q^s_{ad} \tag{9.5.15}$$

Of course, we would not know the unlensed quadrupole matrix for any particular galaxy. But if we have a sample of galaxies in about the same direction and with about the same redshift, so that the shear matrix is the same for all the galaxies in this sample, then we can learn something about this shear matrix by making the reasonable assumption that the orientations of the galaxies are uncorrelated. There are at least two ways that this can be done.

Standard method [2]

In the standard method, one considers the ellipticity matrices

$$X_{ab} \equiv Q_{ab}/\,\mathrm{Tr}\,Q, \quad X^s_{ab} \equiv Q^s_{ab}/\,\mathrm{Tr}\,Q^s \tag{9.5.16}$$

which are normalized to have unit trace. To first order in M, Eq. (9.5.15) gives

$$X_{ab} = X^s_{ab} + M_{ac}X^s_{cd} + M_{bd}X^s_{ad} - 2X^s_{ab}\,\mathrm{Tr}\left(MX^s\right). \tag{9.5.17}$$

If the orientation of galaxies is random, then the average of X^s_{ab} over a sufficiently large sample will be proportional to δ_{ab}, and since these matrices are defined to have unit trace, it follows that the coefficient of proportionality must be just $1/2$:

$$\langle X^s_{ab} \rangle = \frac{1}{2}\delta_{ab}. \tag{9.5.18}$$

To deal with the term in Eq. (9.5.17) that is quadratic in X^s, we also need the average of $X^s_{ab}X^s_{cd}$. The random orientation of the galaxies and the symmetry of the product requires this average to be a linear combination of $\delta_{ab}\delta_{cd}$

and $\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}$, and since X^s has unit trace, this linear combination must take the form

$$\langle X^s_{ab}X^s_{cd} \rangle = \frac{1}{4}\left[\delta_{ab}\delta_{cd} + \xi\left(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc} - \delta_{ab}\delta_{cd}\right)\right], \tag{9.5.19}$$

where, contracting a, b with c, d in Eq. (9.5.19),

$$\xi = \left\langle \mathrm{Tr}\left(X^{s2}\right)\right\rangle - 1/2. \tag{9.5.20}$$

(For instance, if the galaxies in our sample are all spheres, then X^s has both eigenvalues equal to 1/2, and $\xi = 0$, while if they are all extremely prolate ellipsoids then one eigenvalue is unity and the other is zero, and $\xi = 1/2$. More generally, $0 \leq \xi \leq 1/2$.) Taking the average of Eq. (9.5.17) and using Eqs. (9.5.18) and (9.5.19), we have

$$\langle X_{ab} \rangle = \frac{1}{2}\delta_{ab} + (1-\xi)\left(M_{ab} - \frac{1}{2}\delta_{ab}\,\mathrm{Tr}\,M\right). \tag{9.5.21}$$

We do not observe the average of $\mathrm{Tr}\,(X^{s2})$ that appears in our formula (9.5.20) for ξ, but since ξ appears only in (9.5.21) multiplying a term that is already of first order in M, in calculating $\langle X_{ab} \rangle$ to first order in M we only need ξ to zeroth order, and to this order we can replace X^s in Eq. (9.5.20) with the observed ellipticity matrix X :

$$\xi = \left\langle \mathrm{Tr}\left(X^2\right)\right\rangle - 1/2. \tag{9.5.22}$$

This can be measured from the observed shapes of the galaxies in our sample, and Eq. (9.5.21) can then be used to find the traceless part of the shear matrix from a measurement of the average of the observed ellipticity matrix X.

A word on formalism: In the literature, in place of the real ellipticity matrix, one often encounters a complex ellipticity parameter

$$\chi \equiv X_{11} - X_{22} + 2iX_{12}, \tag{9.5.23}$$

and in place of Eq. (9.5.21) one finds a formula relating the complex shear parameter $\gamma \equiv \gamma_1 + i\gamma_2$ to averages of functions of the ellipticity parameter [3]

$$\langle \chi \rangle = 2\gamma - 2\left\langle \chi\,\mathrm{Re}\left(\chi^*\gamma\right)\right\rangle. \tag{9.5.24}$$

Using Eq. (9.5.19), it is easy to see that the average on the right is

$$\left\langle \chi\,\mathrm{Re}\left(\chi^*\gamma\right)\right\rangle = \xi\gamma, \tag{9.5.25}$$

and that therefore the relation (9.5.24) between χ and γ reads

$$\left\langle \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} + Q_{22}} \right\rangle = 2(1-\xi)\left(\gamma_1 + i\gamma_2\right), \tag{9.5.26}$$

which is the same as Eq. (9.5.21).

[472] See, e.g., M. Bartelmann and P. Schneider, Phys. Rep. 340, 291 (2001).

[473] P. Schneider and C. Seitz, Astron. Astrophys. 294, 411 (1995). This is Eq. (4.16) of ref. 2, in the limit $|\gamma_i| \ll 1$ and $|\kappa| \ll 1$.

Alternative method

It is possible to avoid the need to measure the parameter ξ for a sample of source galaxies, by considering averages of the quadrupole matrix Q_{ab} itself, rather than the matrix X_{ab} normalized to have unit trace. The random orientation of galaxies in our sample of sources tells us that the average of unlensed quadrupole matrices takes the form

$$\langle Q^s_{ab} \rangle = \frac{1}{2} \bar{Q} \delta_{ab} \tag{9.5.27}$$

with $\bar{Q}$ an unknown positive constant. Eq. (9.5.15) then gives

$$\langle Q_{ab} \rangle = \bar{Q} \left[\frac{1}{2} \delta_{ab} + M_{ab} \right] \tag{9.5.28}$$

We can eliminate the unknown $\bar{Q}$ by dividing by the average of the trace of the quadrupole moment. Eq. (9.5.28) and its trace give, to first order in M,

$$\frac{\langle Q_{ab} \rangle}{\langle \mathrm{Tr}\, Q \rangle} = \frac{1}{2} \delta_{ab} + M_{ab} - \frac{1}{2} \delta_{ab}\, \mathrm{Tr}\, M \tag{9.5.29}$$

That is,

$$\gamma_1 + i\gamma_2 = \frac{\langle Q_{11} - Q_{22} + 2iQ_{12} \rangle}{2 \langle Q_{11} + Q_{22} \rangle} \tag{9.5.30}$$

Thus we can calculate the traceless part of M_{ab} - that is, the shear field - from a measurement of the average quadrupole tensor, with no need to make a separate measurement of a parameter like ξ.

The difference between the standard and alternative methods is not just one of formalism, for they involve different kinds of averages. In the alternative method galaxies are weighted proportionally to the area of their images, while in the standard method all galaxies are weighted equally. Thus in the alternative method one does not have to worry so much about missing galaxies of small apparent area.

Frequently it is not practical to measure the redshifts of the individual galaxies whose images are distorted by weak lensing. In such cases we must assume that there is some probability distribution $\mathcal{N}(r_S)$ for these galaxies to be at a radial coordinate r_S, with $\int_0^\infty \mathcal{N}(r_S)\, dr_S = 1$. Using the alternative method described above, we can still use Eq. (9.5.29) to calculate the traceless part of the shear matrix from the average observed quadrupole matrix, but
now the average of the quadrupole matrix is calculated using all galaxies along a given line of sight, and the shear matrix is replaced with an effective shear matrix

$$\begin{aligned} M_{ab}(\hat{n}) &= \int_0^\infty \mathcal{N}(r_S)\, M_{ab}(r_S, \hat{n})\, dr_S \\ &= 2 \int_0^\infty \mathcal{N}(r_S)\, dr_S \int_0^{r_S} \frac{r(r, r_S)\, r}{r_S} \left[\frac{\partial^2}{\partial y_a \partial y_b} \delta\phi(r\hat{n} + \mathbf{y}, t_r) \right]_{\mathbf{y}=0} dr \\ &= \int_0^\infty \left[\frac{\partial^2}{\partial y_a \partial y_b} \delta\phi(r\hat{n} + \mathbf{y}, t_r) \right]_{\mathbf{y}=0} g(r) dr \end{aligned} \tag{9.5.31}$$

where

$$g(r) \equiv 2 \int_r^\infty \frac{r(r, r_S)\, r}{r_S} \mathcal{N}(r_S)\, dr_S \tag{9.5.32}$$

Similarly, using the standard method we can still calculate the traceless part of the effective shear matrix from the mean ellipticity matrix, provided there is no evolution in the shape of galaxies. In this case, Eq. (9.5.21) still applies, but with the shear matrix given by Eq. (9.5.31), and with the average squared ellipticity matrix in our formula (9.5.22) for the ξ parameter calculated by averaging over all galaxies along the line of sight. But if galaxy shapes evolve, then Eq. (9.5.31) must be modified accordingly: Instead of a factor $1-\xi$ multiplying the shear matrix in Eq. (9.5.21), it is necessary to use Eq. (9.5.22) to calculate a function $\xi(r_S)$ from the average of the squared ellipticity matrix of galaxies at a radial coordinate r_S, and include a factor $1-\xi(r_S)$ in the integrand of Eq. (9.5.31). This complication does not occur with the alternative method.

Whichever method we use to extract the shear matrix from measurements of weak lensing, we now must face a problem. We would like to use weak lensing measurements to learn about the density perturbation $\delta\rho$, but as shown in Eq. (9.5.10), it is only the convergence κ that is related in any simple way to the density perturbation, while Eqs. (9.5.21) and (9.5.29) show that measurements of weak lensing tell us about the shear field γ_i, not κ. Fortunately, although there is no simple relation between γ_i and κ, there is a simple relation among their gradients. [4] From Eq. (9.5.9), we have

$$\frac{\partial}{\partial\hat{n}_c}M_{ab}(r_S,\hat{n}) = 2\int_0^{r_S}\frac{r(r,r_S)}{r_S}\left[\frac{\partial^3}{\partial y_a\partial y_b\partial y_c}\delta\phi(r\hat{n}+\mathbf{y},t)\right]_{\mathbf{y}=0,t=t_r}dr \tag{9.5.33}$$

Similarly, from Eq. (9.5.31), we have

$$\frac{\partial}{\partial\hat{n}_c}M_{ab}(\hat{n}) = \int_0^{\infty}\left[\frac{\partial^3}{\partial y_a\partial y_b\partial y_c}\delta\phi(r\hat{n}+\mathbf{y},t_r)\right]_{\mathbf{y}=0}\frac{g(r)}{r}dr \tag{9.5.34}$$

The important point is that both expressions are completely symmetric among a, b, and c. Thus the divergence of the traceless part of the shear matrix is proportional to the gradient of the trace:

$$\frac{\partial}{\partial\hat{n}_a}\left(M_{ab}-\frac{1}{2}\delta_{ab}\,\mathrm{Tr}\,M\right) = \frac{1}{2}\frac{\partial}{\partial\hat{n}_b}\,\mathrm{Tr}\,M \tag{9.5.35}$$

where M can be either $M(r_S,\hat{n})$ or $M(\hat{n})$. In other words,

$$\frac{\partial\kappa}{\partial\hat{n}_1} = \frac{\partial\gamma_1}{\partial\hat{n}_1}+\frac{\partial\gamma_2}{\partial\hat{n}_2},\quad \frac{\partial\kappa}{\partial\hat{n}_2} = \frac{\partial\gamma_2}{\partial\hat{n}_1}-\frac{\partial\gamma_1}{\partial\hat{n}_2}. \tag{9.5.36}$$

Thus κ can be calculated from the shear field γ_i, up to an $\hat{n}$-independent constant. As mentioned earlier, the values of κ obtained in this way can be used to measure the projected mass density of lensing bodies along the line of sight. Starting in the late 1990s, numerous groups using telescopes of moderate size have detected shear due to weak lensing. [5] These measurements, together with the above analysis, have been used to map out the distribution of all matter, dark as well as baryonic, in various clusters of galaxies, such as the bullet cluster 1E0657-558 described in Section 3.4.

Now let us turn to the application of weak lensing to find the distribution of inhomogeneities in the cosmological mass density. In using weak lensing to study the large scale distribution of matter in the universe, we are not so much interested in the shear matrix in any one direction, as in its distribution around the sky. Fluctuations in the gravitational potential have random sign, so the average over the sky of the shear matrix is zero. It is most useful to consider instead the average of the product of shear matrix elements along two different lines of sight, with "average" understood in the same sense as used in the analysis of fluctuations in the cosmic

[474] N. Kaiser, Astrophys. J. 439, L1 (1995).

microwave background: either an average over positions from which the sky might be observed, or an average over the particular series of accidents that lead to a particular distribution of cosmic matter. Neither average corresponds to what we observe, but the difference vanishes for large multipole orders, for

the same reason as explained in Section 2.6 for the multipole coefficients of anisotropies in the cosmic microwave background.

Since we are now considering different directions, it is convenient to write the definition (9.5.7) of the convergence and shear fields more explicitly. If we write $\hat{n}$ in terms of polar and azimuthal angles as

$$\hat{n} = (\sin\vartheta\cos\varphi, \sin\vartheta\sin\varphi, \cos\vartheta), \tag{9.5.37}$$

and take the two orthogonal directions normal to $\hat{n}$ as

$$\hat{\vartheta} = (\cos\vartheta\cos\varphi, \cos\vartheta\sin\varphi, -\sin\vartheta), \quad \hat{\varphi} = (-\sin\varphi, \cos\varphi) \tag{9.5.38}$$

then

$$\kappa(\hat{n}) + \gamma_1(\hat{n}) = \hat{\vartheta}_i\hat{\vartheta}_j M_{ij}(\hat{n}), \tag{9.5.39}$$

$$\kappa(\hat{n}) - \gamma_1(\hat{n}) = \hat{\varphi}_i\hat{\varphi}_j M_{ij}(\hat{n}), \tag{9.5.40}$$

$$\gamma_2(\hat{n}) = \hat{\vartheta}_i\hat{\varphi}_j M_{ij}(\hat{n}). \tag{9.5.41}$$

We are here using ϑ and φ instead of the usual θ and ϕ for the polar and azimuthal angles, to avoid confusion with other uses of the symbols θ and ϕ in this section. Also, since we will now be considering different directions $\hat{n}$, we have promoted the two-valued indices a, b, etc. to three-valued indices i, j, etc. Eq. (9.5.31) for the 2×2 shear matrix M_{ab} has accordingly been extended to give a 3×3 matrix

$$M_{ij}(\hat{n}) \equiv \int_0^\infty \left[\frac{\partial^2}{\partial x_i \partial x_j}\delta\phi(\mathbf{x}, t)\right]_{\mathbf{x}=r\hat{n}, t=t_r} g(r)dr \tag{9.5.42}$$

Equivalently, we can write

$$\gamma(\hat{n}) \equiv \gamma_1(\hat{n}) + i\gamma_2(\hat{n}) = e_{+i}(\hat{n})e_{+j}(\hat{n})M_{ij}(\hat{n}), \tag{9.5.43}$$

where, as in Section 7.4, $\mathbf{e}_+ \equiv (\hat{\vartheta} + i\hat{\varphi})/\sqrt{2}$. Also, since $\hat{\vartheta}$ and $\hat{\varphi}$ span the space normal to $\hat{n}$,

$$\kappa(\hat{n}) = \frac{1}{2}\left(\delta_{ij} - \hat{n}_i\hat{n}_j\right) M_{ij}(\hat{n}) \tag{9.5.44}$$

The expansion of the convergence in spherical harmonics follows along almost the same lines as the expansion of microwave background temperature fluctuations described in Sections 2.6 and 7.2. We write the perturbation in the gravitational potential as

$$\delta\phi(\mathbf{x}, t) = \int d^3q\,\alpha(\mathbf{q})\delta\phi_q(t)e^{i\mathbf{q}\cdot\mathbf{x}} \tag{9.5.45}$$

where $\alpha(\mathbf{q})$ is the stochastic parameter for scalar perturbations, normalized so that

[475] For references to these observations and a brief general review, see D. Munshi and P. Valageas, Roy. Soc. London Trans. Ser. A 363, 2675 (2001) [astro-ph/0509216]. Comprehensive reviews of weak lensing are given by P. Schneider and M. Bartelmann, ref. 2; A. Refrigier, Ann. Rev. Astron. Astrophys. 41, 645 (2003); P. Schneider, in Gravitational Lensing: Strong, Weak, and Micro, eds. G. Meylan et al. (Springer-Verlag, Berlin, 2006): 269 [astro-ph/0509252].

$$\langle \alpha(\mathbf{q})\alpha^*(\mathbf{q}')\rangle = \delta^3(\mathbf{q}-\mathbf{q}'). \tag{9.5.46}$$

Then Eqs.(9.5.42), (9.5.44), and (9.5.45) give

$$\kappa(\hat{n}) = -\frac{1}{2}\int d^3q\alpha(\mathbf{q})\int_0^\infty g(r)\delta\phi_q(t_r)\left(\mathbf{q}^2-(\mathbf{q}\cdot\hat{n})^2\right)e^{i\mathbf{q}\cdot\hat{n}r}dr$$

The quantity $-(\mathbf{q}\cdot\hat{n})^2$ can be replaced with $\partial^2/\partial r^2$, and for the exponential we can use the familiar formula

$$e^{i\mathbf{q}\cdot\hat{n}r} = 4\pi\sum_{\ell m} i^\ell j_\ell(qr)Y_\ell^m(\hat{n})Y_\ell^{m*}(\hat{q}). \tag{9.5.47}$$

This gives the partial wave expansion for the convergence

$$\kappa(\hat{n}) = \sum_{\ell m} a_{\kappa,\ell m}Y_\ell^m(\hat{n}), \tag{9.5.48}$$

with coefficients

$$\begin{aligned} a_{\kappa,\ell m} = &-2\pi i^\ell\int d^3qq^2\alpha(\mathbf{q})Y_\ell^{m*}(\hat{q}) \\ &\times\int_0^\infty g(r)\delta\phi_q(t_r)\left(j_\ell(qr)+j_\ell''(qr)\right)dr \end{aligned} \tag{9.5.49}$$

This can be used together with Eq. (9.5.46) to calculate the correlation of the convergence with itself, or with microwave background temperature fluctuations or any other scalar perturbations. In particular, for the correlation of the convergence with itself, we have

$$\left\langle a_{\kappa,\ell m}a_{\kappa,\ell'm'}^*\right\rangle = \delta_{\ell\ell'}\delta_{mm'}C_{\kappa\kappa,\ell}, \tag{9.5.50}$$

where $C_{\kappa\kappa,\ell}$ is the multipole coefficient

$$C_{\kappa\kappa,\ell} = 4\pi^2\int_0^\infty q^6dq\left|\int_0^\infty g(r)\delta\phi_q(t_r)\left(j_\ell(qr)+j_\ell''(qr)\right)dr\right|^2. \tag{9.5.51}$$

On the other hand, comparing Eqs. (9.5.43) and (7.4.6), we see that the shear components $\gamma_1(\hat{n})$ and $\gamma_2(\hat{n})$ involve the polarization $\mathbf{e}_+(\hat{n})$ in much the same way as the Stokes parameters $Q(\hat{n})$ and $U(\hat{n})$, so for the same reason as in Section 7.4, it is necessary to expand the complex shear parameter $\gamma_1(\hat{n})+i\gamma_2(\hat{n})$ in the spin 2 spherical harmonics $\mathcal{Y}_\ell^m(\hat{n})$, defined by Eq. (7.4.16):

$$\mathcal{Y}_\ell^m(\hat{n}) \equiv 2\sqrt{\frac{(\ell-2)!}{(\ell+2)!}}e_{+i}(\hat{n})e_{+j}(\hat{n})\tilde{\nabla}_i\tilde{\nabla}_jY_\ell^m(\hat{n}), \tag{9.5.52}$$

where $\tilde{\nabla}$ is the angular part of the gradient operator:

$$\tilde{\nabla} \equiv \hat{\vartheta}\frac{\partial}{\partial\vartheta}+\frac{\hat{\varphi}}{\sin\vartheta}\frac{\partial}{\partial\varphi}. \tag{9.5.53}$$

Using Eqs. (9.5.42), (9.5.43), (9.5.45), and (9.5.47), we have

$$\gamma_1(\hat{n})+i\gamma_2(\hat{n}) = \sum_{\ell m}a_{\gamma,\ell m}\mathcal{Y}_\ell^m(\hat{n}) \tag{9.5.54}$$

with coefficients

$$a_{\gamma,\ell m} = 2\pi i^\ell \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \int d^3q \alpha(\mathbf{q}) Y_\ell^m(\hat{q})^* \int_0^\infty \delta\phi_q(t_r)\, j_\ell(qr) g(r) r^{-2} dr \tag{9.5.55}$$

The reality of $\delta\phi(\mathbf{x},t)$ requires that $\alpha^*(\mathbf{q})\delta\phi_q^*(t) = \alpha(-\mathbf{q})\delta\phi_q(t)$, while $Y_\ell^{m*}(\hat{q}) = (-1)^\ell Y_\ell^{-m}(-\hat{q})$, so $a_{\gamma,\ell m} = a^*_{\gamma,\ell-m}$. That is, the shear produced by perturbations in the mass density is of " E " rather than " B " type, in the sense of Eq. (7.4.25). Thus any observation of B-type shear, for which $a_{\gamma,lm} = -a^*_{\gamma,l-m}$, would be a sign of lensing caused by something other than density perturbations, such as gravitational waves. Unfortunately, the lensing due to gravitational waves produced in inflation is much too small to be observed. [6]

Using (9.5.46) and the orthonormality of the spin 2 spherical harmonics,

$$\left\langle a_{\gamma,\ell m} a^*_{\gamma,\ell' m'} \right\rangle = \delta_{\ell\ell'}\delta_{mm'} C_{\gamma\gamma,\ell}, \tag{9.5.56}$$

where the multipole coefficients in shear-shear correlations are

$$C_{\gamma\gamma,\ell} = \frac{4\pi^2(\ell+2)!}{(\ell-2)!} \int_0^\infty q^2 dq \left| \int_0^\infty \delta\phi_q(t_r)\, j_\ell(qr) g(r) r^{-2} dr \right|^2 . \tag{9.5.57}$$

The Poisson equation $\nabla^2\delta\phi = 4\pi G a^2 \delta\rho_M$ relates the quantity $\delta\phi_q(t)$ appearing in Eqs. (9.5.51) and (9.5.57) the power spectral function $P(q/a_0)$ of fractional density fluctuations introduced in Section 8.1:

$$\begin{aligned} q^4 |\delta\phi_q(t)|^2 &= (4\pi G)^2 a^4(t) \bar{\rho}_M^2(t) |\delta_{Mq}(t)|^2 \\ &= (4\pi G)^2 a^4(t) \bar{\rho}^2(t) P(q/a_0) \left(\frac{F(t)}{F(t_0)} \right)^2 \\ &= \frac{9\Omega_M^2 H_0^4 a_0^3}{4a^2(t)(2\pi)^3} P(q/a_0) \left(\frac{F(t)}{F(t_0)} \right)^2 , \end{aligned} \tag{9.5.58}$$

with $F(t)$ the function (8.1.20), and $a_0 \equiv a(t_0)$.

It is possible to give less complicated formulas for the multipole coefficients in the most interesting case, of $\ell \gg 1$. Let us first consider the $\kappa\kappa$ correlation. We can evaluate the integral over r in Eq. (9.5.51) by recalling that for $\ell \gg 1$, if qr differs from $\ell + 1/2$ by more than an amount of order $\ell^{-1/3}$, the spherical Bessel function $j_\ell(qr)$ is exponentially small for $qr < \ell + 1/2$ and rapidly oscillating for $qr > \ell + 1/2$, while (unlike the source functions in microwave background anisotropies) all other ingredients in the integral vary slowly with r. For $\ell \gg 1$, we can therefore replace r with ℓ/q in $g(r)$ and t_r :

$$C_{\kappa\kappa,\ell} \to 4\pi^2 \int_0^\infty dq q^6 \left| \delta\phi_q\left(t_{\ell/q}\right) \right|^2 g^2(\ell/q) \left| \int_0^\infty \left(j_\ell(qr) + j_\ell''(qr) \right) dr \right|^2 \tag{9.5.59}$$

The integral of $j_\ell(qr)$ gives [7]

$$\int_0^\infty j_\ell(qr) dr = \frac{\sqrt{\pi}}{2q} \frac{\Gamma\left(\frac{\ell+1}{2}\right)}{\Gamma\left(\frac{\ell+2}{2}\right)} \to \frac{1}{q}\sqrt{\frac{\pi}{2\ell}} \tag{9.5.60}$$

On the other hand, since for large ℓ the contributions from $r \ll \ell/q$ and $r \gg \ell/q$ are strongly suppressed, we can drop the integral over r of $j_\ell''(qr)$. Thus Eq. (9.5.58) gives, for $\ell \gg 1$,

[476] S. Dodelson, E. Rozo, and A. Stebbins, Phys. Rev. Lett. 91, 021301 (2003).

$$
\begin{aligned}
C_{\kappa\kappa,\ell} &\to \frac{2\pi^3}{\ell}\int_0^\infty dqq^4\left|\delta\phi_q\left(t_{\ell/q}\right)\right|^2 g^2(\ell/q) \\
&= 2\pi^3\ell^4\int_0^\infty drr^{-5}\left|\delta\phi_{\ell/r}\left(t_r\right)\right|^2 g^2(r) \\
&= \frac{9\Omega_M^2 H_0^4 a_0^3}{16}\int_0^\infty dr\frac{g^2(r)}{a^2\left(t_r\right)r^2}P\left(\ell/a_0 r\right)\left(\frac{F\left(t_r\right)}{F\left(t_0\right)}\right)^2
\end{aligned}
\tag{9.5.61}
$$

The large ℓ limit of the $\gamma\gamma$ multipole coefficient is precisely the same. In this limit, the factor $(\ell+2)!/(\ell-2)!$ in Eq. (9.5.57) becomes ℓ^4, and the factor r^{-2} in the integral over r in Eq. (9.5.57) can be replaced with $(q/\ell)^2$, so for large ℓ the ratio of the integrand of the q-integral in Eq. (9.5.57) to that in Eq. (9.5.51) (dropping $j_\ell''(qr)$) is

$$
\frac{\ell^4 q^2(q/\ell)^4}{q^6} = 1
$$

Hence for large ℓ,

$$
C_{\gamma\gamma,\ell} \to C_{\kappa\kappa,\ell} \to \frac{9\Omega_M^2 H_0^4 a_0^3}{16}\int_0^\infty dr\frac{g^2(r)}{a^2\left(t_r\right)r^2}P\left(\ell/a_0 r\right)\left(\frac{F\left(t_r\right)}{F\left(t_0\right)}\right)^2 \tag{9.5.62}
$$

Thus we get just the same information by measuring the shear multipole coefficients at large ℓ as we would get if we could measure the convergence multipole coefficients at the same ℓ.

For a given weighting function $g(r)$ defined by Eq. (9.5.32), observation of the ℓ dependence of $C_{\gamma\gamma,\ell}$ (or $C_{\kappa\kappa,\ell}$) provides a measurement of the k dependence of the power spectral function $P(k)$. It should be noted that if the integral receives its main contribution from radial coordinates r such that vacuum energy has not yet become important by the times t_r, then according to Eq. (8.1.20), $F(t) \propto a(t)$, so Eq. (9.5.62) simplifies further, to

$$
C_{\gamma\gamma,\ell} \to C_{\kappa\kappa,\ell} \to \frac{9\Omega_M^2 H_0^4 a_0}{16}\int_0^\infty dr\frac{g^2(r)}{r^2}P\left(\ell/a_0 r\right) \tag{9.5.63}
$$

This is the formula usually quoted. [8]

We can express the correlation functions for the shear components and microwave background temperature fluctuations by taking over results we have already found for the correlation functions of microwave background polarization and temperature fluctuations. Comparing Eq. (9.5.54) with Eqs. (7.4.15) and (7.4.25), and comparing Eq. (9.5.56) with Eq. (7.4.28), we see that we can obtain the correlation functions for the shear components from Eqs. (7.4.76)-(7.4.78) by replacing Q and U with γ_1 and γ_2 and replacing $C_{EE,\ell}$ with $C_{\gamma\gamma,\ell}$ and (since shear is purely of E-type) dropping $C_{BB,\ell}$:

$$
\left\langle\gamma_1(\hat{n})\gamma_1\left(\hat{n}'\right)\right\rangle = \frac{1}{2}\sum_\ell C_{\gamma\gamma,\ell}\,\mathrm{Re}\left(F_\ell\left(\hat{n},\hat{n}'\right)+G_\ell\left(\hat{n},\hat{n}'\right)\right), \tag{9.5.64}
$$

$$
\left\langle\gamma_2(\hat{n})\gamma_2\left(\hat{n}'\right)\right\rangle = \frac{1}{2}\sum_\ell C_{\gamma\gamma,\ell}\,\mathrm{Re}\left(F_\ell\left(\hat{n},\hat{n}'\right)-G_\ell\left(\hat{n},\hat{n}'\right)\right), \tag{9.5.65}
$$

$$
\left\langle\gamma_1(\hat{n})\gamma_2\left(\hat{n}'\right)\right\rangle = \frac{1}{2}\sum_\ell C_{\gamma\gamma,\ell}\,\mathrm{Im}\left(-F_\ell\left(\hat{n},\hat{n}'\right)+G_\ell\left(\hat{n},\hat{n}'\right)\right), \tag{9.5.66}
$$

[477] This is a special case of formula 6.561.14 of I. S. Gradshteyn and I. M. Ryzhik, Table of Integrals, Series, and Products, ed. A. Jeffrey (Academic Press, New York, 1980).

where F_ℓ and G_ℓ are functions given by Eqs. (7.4.87) and (7.4.88):

$$F_\ell\left(\hat{n},\hat{n}'\right)=\frac{2\ell+1}{4\pi}D^{(\ell)}_{-2,-2}\left(S^{-1}(\hat{n})S\left(\hat{n}'\right)\right), \tag{9.5.67}$$

$$G_\ell\left(\hat{n},\hat{n}'\right)=(-1)^\ell\frac{2\ell+1}{4\pi}D^{(\ell)}_{-2,-2}\left(S^{-1}(\hat{n})S\left(-\hat{n}'\right)\right), \tag{9.5.68}$$

with $D^{(\ell)}$ the irreducible unitary representation of the rotation group for angular momentum ℓ, and $S(\hat{n})$ the standard rotation (7.4.12) that takes the three-direction with $\vartheta=\varphi=0$ into the direction $\hat{n}$. Also, if we write the microwave background temperature fluctuation as in Eq. (2.6.1)

$$\Delta T(\hat{n})=\sum_{\ell m}a_{T,\ell m}Y^m_\ell(\hat{n}),$$

and use rotational invariance to define

$$\left\langle a_{T,\ell m}a^*_{\gamma,\ell' m'}\right\rangle=\delta_{\ell\ell'}\delta_{mm'}C_{T\gamma,\ell}, \tag{9.5.69}$$

then we can obtain the correlation functions for the shear components with microwave background temperature fluctuations from Eqs. (7.4.81) and (7.4.82) by again replacing Q and U with γ_1 and γ_2 and replacing $C_{TE,\ell}$ with $-C_{T\gamma,\ell}$:

$$\langle\Delta T(\hat{n})\gamma_1\left(\hat{n}'\right)\rangle=\sum_\ell C_{T\gamma,\ell}\,\mathrm{Re}\,H_\ell\left(\hat{n},\hat{n}'\right), \tag{9.5.70}$$

$$\langle\Delta T(\hat{n})\gamma_2\left(\hat{n}'\right)\rangle=-\sum_\ell C_{T\gamma,\ell}\,\mathrm{Im}\,H_\ell\left(\hat{n},\hat{n}'\right), \tag{9.5.71}$$

where H_ℓ is the function (7.4.89):

$$H_\ell\left(\hat{n},\hat{n}'\right)=\frac{2\ell+1}{4\pi}D^{(\ell)}_{0,-2}\left(S^{-1}(\hat{n})S\left(\hat{n}'\right)\right) \tag{9.5.72}$$

For comparison, we mention also that the convergence correlation function is given by the analog of the temperature correlation function (2.6.4):

$$\langle\kappa(\hat{n})\kappa\left(\hat{n}'\right)\rangle=\sum_\ell\frac{2\ell+1}{4\pi}C_{\kappa\kappa,\ell}P_\ell\left(\hat{n}\cdot\hat{n}'\right). \tag{9.5.73}$$

The measurement of $C_{\gamma\gamma,\ell}$ or (with the aid of Eq. (9.5.36) or (9.5.62)) of $C_{\kappa\kappa,\ell}$ can be used not only to learn about the power spectral function $P(k)$, but also to put constraints on the cosmological parameters that enter in

Eq. (9.5.32) for $g(r)$. So far, shear measurements from the Canada-FranceHawaii Telescope Wide Synoptic Legacy Survey have been used [9] to set a value for the parameter σ_8 discussed in Section 8.1: under the assumption that $\Omega_M=0.3$, it is found that $\sigma_8=0.85\pm0.06$. Although this parameter is also measured in studies of large scale structure, the weak lensing measurement has the advantage of not depending on the use of luminous sources as a tracer of dark matter. The same group also finds from cosmic shear data alone that the ratio $w\equiv\bar{p}/\bar{\rho}$, if assumed constant, is less than -0.8 at 68% confidence. Another group, [10] combining shear measurements from the CTIO lensing survey with cosmic microwave background data and measurements of Type Ia supernova redshift and luminosities, has found that $\sigma_8=0.81^{+0.15}_{0.21}$, and that (if constant) $w=-0.89^{+0.16}_{-0.21}$, with 95% confidence. The application of weak lensing surveys to cosmology has really just begun.

[478] D. J. Bacon, A. R. Refrigier and R. S. Ellis, Mon. Not. Roy. Astron. Soc. 318, 625 (2000). Correlation functions for elements of the shear matrix were studied by N. Kaiser, Astrophys. J. 498, 26 (1998).

10.9.8 9.6 Cosmic strings

The spontaneous breakdown of symmetries in the early universe can produce linear discontinuities in fields, known as cosmic strings. [1] Cosmic strings are also common in modern string theories. [2] Unless we are unlucky enough to have a cosmic string slice through the solar system, the only way that a cosmic string can be discovered seems to be through its action as a gravitational lens. [3]

At a sufficient distance r from a string, the gravitational field becomes Newtonian. For a long straight string the solution of Laplace's equation with cylindrical symmetry has a gravitational potential of the form

$$\phi = -2G\sigma \ln r + C \tag{9.6.1}$$

where σ and C are constants of integration. If the string is non-relativistic, then σ is the string's mass per length. At a sufficient distance from the string the post-Newtonian approximation applies, and the direction $\hat{u}$ of a light

ray is governed by the equation [4]

$$\frac{d\hat{u}}{dt} = -2\hat{u} \times (\hat{u} \times \nabla\phi) \tag{9.6.2}$$

As long as the deflection of the light ray is small, it can be calculated as

$$\delta\hat{u} = -2\hat{u}_0 \times \left(\hat{u}_0 \times \int dt \nabla\phi\right) \tag{9.6.3}$$

where $\hat{u}_0$ is the photon's initial direction. An elementary integration shows that the light ray is deflected toward the string in a direction perpendicular to both the string and the light ray, by an angle

$$|\delta\hat{u}| = \frac{4\pi\sigma G}{\sin\theta} \tag{9.6.4}$$

where θ is the angle between the directions of the string and the light ray. Because the deflection is toward the string, it is in the opposite direction for rays passing the string on opposite sides, so that the image of a source behind the string is split into two parts. It is noteworthy that neither the direction nor the magnitude of this deflection depends on the distance of the light ray from the string, as long as the distance is large enough to allow the use of the non-relativistic formulas (9.6.1) and (9.6.2).

In 2003, general interest in cosmic strings was heightened by the discovery of what seemed at first to be a plausible candidate for lensing by a cosmic string. A pair of images of elliptical galaxies separated by about 1.8 arcseconds was found to have substantially the same redshift, $z = 0.46$, and the same spectra. The images were not distorted in the way that would be expected for lensing of a single galaxy by a more-or-less spherical source, but are consistent with lensing by a cosmic string. [5] Subsequently an excess of gravitationally lensed objects was found in the neighborhood of this string candidate, [6] lending further support to the view that

[479] H. Hoekstra et al., Astrophys. J. 647, 116 (2006) [astro-ph/0511089].

[10] M. Jarvis, B. Jain, G. Bernstein and D. Dolney, Astrophys. J. 644, 71 (2006) [astro-ph/0502243].

[1] T. W. B. Kibble, J. Phys. A 9, 1387 (1976); A. Vilenkin and E. P. S. Shellard, Cosmic Strings and Other Topological Defects (Cambridge University press, Cambridge, 1994).

[2] E. Witten, Phys. Lett. B 153, 243 (1985); E. J. Copeland, R. C. Myers, and J. Polchinski, J. High Energy Phys. 0406, 013 (2005) [hep-th/0312067].

[3] The relativistic calculation of lensing by a cosmic string is due to J. R. Gott, Astrophys. J. 288, 422 (1985). The anisotropy in the cosmic microwave background due to cosmic strings was calculated by N. Kaiser and A. Stebbins, Nature 310, 391 (1984).

this was the image of a single galaxy lensed by a cosmic string. But in 2006 this interpretation had to be abandoned, when observations at the Hubble Space Telescope revealed that this was in fact a pair of interacting elliptical galaxies, not the result of any sort of lensing. [7] This episode illustrates how difficult it will be to detect cosmic strings through their lensing action.

10.10 Inflation as the Origin of Cosmological Fluctuations

The most exciting aspect of the inflationary cosmological theories described in Chapter 4 is that they provide a natural quantum mechanical mechanism for the origin of the cosmological fluctuations observed in the cosmic microwave background and in the large scale structure of matter, [1] and that may in the future be observed in gravitational waves. [2] We have seen in Chapter 6 that the magnitude and wavelength dependence of adiabatic scalar and tensor fluctuations depend on initial conditions only through the quantities $\mathcal{R}_q^o$ and $\mathcal{D}_q^o$, respectively. As given by Eq. (5.4.24), $\mathcal{R}_q^o$ is the value of the gauge-invariant quantity $\mathcal{R}_q(t) \equiv A_q(t)/2 + H(t)\delta u_q(t)$ outside the horizon, that is for $q/a(t) \ll H(t)$, where for adiabatic fluctuations $\mathcal{R}_q$ is time-independent. Likewise, $\mathcal{D}_q^o$ is the time-independent value of the gravitational wave amplitude $\mathcal{D}_q(t)$ for $q/a(t) \ll H(t)$. During the matter- or radiation-dominated eras $a(t)$ increased like $t^{2/3}$ or $t^{1/2}$, respectively, while $H(t) \equiv \dot{a}(t)/a(t)$ decreased like $1/t$, so any wavelength will be found outside the horizon if we go back early enough in one or the other of these eras. But during the period of inflation that is supposed to precede the radiationdominated era, $H(t)$ was roughly (perhaps very roughly) constant, while $a(t)$ increased more-or-less exponentially, so even if a perturbation was outside the horizon at the end of inflation it would inevitably have been found deep inside the horizon sufficiently early in the era of inflation. At these very early times fields oscillated much more quickly than the universe expanded, and their quantum fluctuations therefore were essentially just what they would be in ordinary Minkowski spacetime. In this chapter we will follow adiabatic scalar fluctuations and tensor fluctuations from this very early era, through the epochs when fluctuations of various wavelengths exited the horizon, to the time when $\mathcal{R}_q(t)$ and $\mathcal{D}_q(t)$ reached the constant values $\mathcal{R}_q^o$ and $\mathcal{D}_q^o$ that are measured in the cosmic microwave background and in the large scale structure of matter, and that may some day be measured through the direct detection of cosmological gravitational radiation.

10.10.1 10.1 Scalar fluctuations during inflation

We will first consider the simplest model of inflation, with a single real "inflaton" field $\varphi(x)$, and an action (B.63):

$$I_\varphi = \int d^4x \sqrt{-\operatorname{Det} g}\left[-\frac{1}{2}g^{\mu\nu}\frac{\partial\varphi}{\partial x^\mu}\frac{\partial\varphi}{\partial x^\nu} - V(\varphi)\right], \tag{10.1.1}$$

involving an arbitrary real potential $V(\varphi)$. This is not the only possibility; we will take up the question of its plausibility at the end of this section.

[474] See footnote 1 of Section 9.3.

[5] M. V. Sazhin et al., Mon. Not. Roy. Astron. Soc. 343, 353 (2003) [astro-ph/0302547].

[6] M. V. Sazhin et al., astro-ph/0406516.

[7] M. V. Sazhin et al., Mon. Not. Roy. Astron. Soc. 376, 1731 (2007) [astro-ph/0611744].

[471] S. V. Mukhanov and G. V. Chibisov, Sov. Phys. JETP Lett. 33, 532 (1981); S. Hawking, Phys. Lett. 115B, 295 (1982); A. A. Starobinsky, Phys. Lett. 117B, 175 (1982); A. Guth and S.-Y. Pi, Phys. Rev. Lett. 49, 1110 (1982); J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, Phys. Rev. D 28, 679 (1983); W. Fischler, B. Ratra, and L. Susskind, Nucl. Phys. B 259, 730 (1985).

[2] The cosmological generation of gravitational waves was considered by L. P. Grishchuk, Sov. Phys. JETP 40, 409 (1974); A. A. Starobinsky, Sov. Phys. JETP Lett. 30, 682 (1979), and calculated in the context of inflation by V. A. Rubakov, M. V. Sazhin, and A. V. Veryaskin, Phys. Lett. 115B, 189 (1982).

In line with the observed isotropy and homogeneity of the universe on the average, we take the scalar field as an unperturbed term $\bar{\varphi}(t)$ that depends only on time, plus a small perturbation $\delta\varphi(\mathbf{x},t)$:

$$\varphi(\mathbf{x},t) = \bar{\varphi}(t) + \delta\varphi(\mathbf{x},t). \tag{10.1.2}$$

Similarly, as in Chapters 5-7, the metric is given by the unperturbed Robertson-Walker metric $\bar{g}_{\mu\nu}(t)$ (with $K = 0$) plus a small perturbation $h_{\mu\nu}(\mathbf{x},t)$

$$g_{\mu\nu}(\mathbf{x},t) = \bar{g}_{\mu\nu}(t) + h_{\mu\nu}(\mathbf{x},t). \tag{10.1.3}$$

The energy momentum tensor of the scalar field is shown in Appendix B to take the perfect fluid form, with an energy density, pressure, and velocity four-vector given by Eqs. (B.66)-(B.68) as

$$\begin{aligned}
\rho &= -\frac{1}{2}g^{\mu\nu}\frac{\partial\varphi}{\partial x^\mu}\frac{\partial\varphi}{\partial x^\nu} + V(\varphi) \\
p &= -\frac{1}{2}g^{\mu\nu}\frac{\partial\varphi}{\partial x^\mu}\frac{\partial\varphi}{\partial x^\nu} - V(\varphi) \\
u^\mu &= -\left[-g^{\rho\sigma}\frac{\partial\varphi}{\partial x^\rho}\frac{\partial\varphi}{\partial x^\sigma}\right]^{-1/2} g^{\mu\tau}\frac{\partial\varphi}{\partial x^\tau}.
\end{aligned}$$

The energy-density and pressure of the unperturbed scalar field are then

$$\bar{\rho} = \frac{1}{2}\dot{\bar{\varphi}}^2 + V(\bar{\varphi}), \quad \bar{p} = \frac{1}{2}\dot{\bar{\varphi}}^2 - V(\bar{\varphi}), \tag{10.1.4}$$

while the unperturbed velocity four-vector has components

$$\bar{u}^i = 0, \quad \bar{u}^0 = 1. \tag{10.1.5}$$

The Friedmann equation (with zero spatial curvature) is here

$$H^2 = \frac{8\pi G}{3}\left(\frac{1}{2}\dot{\bar{\varphi}}^2 + V(\bar{\varphi})\right). \tag{10.1.6}$$

and the energy conservation condition $\dot{\bar{\rho}} = -3H(\bar{\rho} + \bar{p})$ yields the field equation of the unperturbed scalar field

$$\ddot{\bar{\varphi}} + 3H\dot{\bar{\varphi}} + V'(\bar{\varphi}) = 0, \tag{10.1.7}$$

where as usual $H(t) \equiv \dot{a}(t)/a(t)$.
The perturbation $\delta g_{\mu\nu} \equiv h_{\mu\nu}$ to the metric will be taken in the Newtonian gauge form (5.3.18):

$$h_{00} = -2\Psi, \quad h_{0i} = 0, \quad h_{ij} = -2a^2\delta_{ij}\Psi \tag{10.1.8}$$

(The scalar anisotropic inertia term π^S, as well as π^V and π^T, vanishes in scalar field theories, so according to Eq. (5.3.20) the other scalar gravitational perturbation Φ here equals Ψ.) The perturbations to the pressure, energy density, and velocity three-vector are given by the terms in Eqs. (B.66)-(B.68) that are of first order in perturbations:

$$\delta\rho = \dot{\bar{\varphi}}\delta\dot{\varphi} + V'(\bar{\varphi})\delta\varphi - \Psi\dot{\bar{\varphi}}^2 \tag{10.1.9}$$

$$\delta p = \dot{\bar{\varphi}}\delta\dot{\varphi} - V'(\bar{\varphi})\delta\varphi - \Psi\dot{\bar{\varphi}}^2 \tag{10.1.10}$$

$$\delta u_i = \frac{\partial\delta u}{\partial x^i} \text{ where } \delta u = -\frac{\delta\varphi}{\dot{\bar{\varphi}}} \tag{10.1.11}$$

The field equations may be taken as the Einstein equation Eq. (5.3.21) and the energy conservation condition (5.3.24), which here take the form

$$\dot{\Psi} + H\Psi = 4\pi G\dot{\bar{\varphi}}\delta\varphi \tag{10.1.12}$$

$$\delta\ddot{\varphi} + 3H\delta\dot{\varphi} + \frac{\partial^2 V(\bar{\varphi})}{\partial\bar{\varphi}^2}\delta\varphi - \left(\frac{\nabla^2}{a^2}\right)\delta\varphi = -2\Psi\frac{\partial V(\bar{\varphi})}{\partial\bar{\varphi}} + 4\dot{\Psi}\dot{\bar{\varphi}} \tag{10.1.13}$$

while the constraint (5.3.26) is here

$$\left(\dot{H} - \frac{\nabla^2}{a^2}\right)\Psi = 4\pi G(-\dot{\bar{\varphi}}\delta\dot{\varphi} + \ddot{\bar{\varphi}}\delta\varphi) \tag{10.1.14}$$

(In deriving Eq. (10.1.14), we make use of the convenient relation $\dot{H} = -4\pi G\dot{\bar{\varphi}}^2$, which follows from Eqs. (10.1.6) and (10.1.7).) The field equation (5.3.20) has already been accounted for by setting $\Phi = \Psi$, while the remaining field equations (5.3.19) and (5.3.22) and the momentum conservation equation (5.3.23) just repeat the information contained in Eqs. (10.1.12)(10.1.14).

Let us first consider the plane wave solutions of Eqs. (10.1.12)-(10.1.14), in which $\delta\varphi$ and Ψ are of the form $\exp(i\mathbf{q}\cdot\mathbf{x})\delta\varphi_q(t)$ and $\exp(i\mathbf{q}\cdot\mathbf{x})\Psi_q(t)$, respectively. We will return later to the issue of how to put these solutions
together in forming the perturbations $\delta\varphi(\mathbf{x},t)$ and $\Psi(\mathbf{x},t)$. The time dependence of the plane wave solutions is given by Eqs. (10.1.12) and (10.1.13), with $-\nabla^2$ replaced with $q^2 \equiv \mathbf{q}^2$:

$$\dot{\Psi}_q + H\Psi_q = 4\pi G\dot{\bar{\varphi}}\delta\varphi_q, \tag{10.1.15}$$

$$\begin{aligned}&\delta\ddot{\varphi}_q + 3H\delta\dot{\varphi}_q + \frac{\partial^2 V(\bar{\varphi})}{\partial\bar{\varphi}^2}\delta\varphi_q + \left(\frac{q^2}{a^2}\right)\delta\varphi_q \\ &= -2\Psi_q\frac{\partial V(\bar{\varphi})}{\partial\bar{\varphi}} + 4\dot{\Psi}_q\dot{\bar{\varphi}},\end{aligned} \tag{10.1.16}$$

while the constraint (10.1.14) now reads

$$\left(\dot{H} + \frac{q^2}{a^2}\right)\Psi_q = 4\pi G\left(-\dot{\bar{\varphi}}\delta\dot{\varphi}_q + \ddot{\bar{\varphi}}\delta\varphi_q\right). \tag{10.1.17}$$

At sufficiently early times q/a will be much larger than H or $\partial^2 V(\bar{\varphi})/\partial\bar{\varphi}^2$, so we can look for WKB solutions, for which time derivatives of fields yield factors of order q/a, of the form

$$\begin{aligned}\delta\varphi_q(t) &\to f(t)\exp\left(-iq\int_{t_*}^{t}\frac{dt'}{a\,(t')}\right) \\ \Psi_q(t) &\to g(t)\exp\left(-iq\int_{t_*}^{t}\frac{dt'}{a\,(t')}\right)\end{aligned} \tag{10.1.18}$$

where $f(t)$ and $g(t)$ vary much more slowly than the argument of the exponential, and t_* is arbitrary. The exponential factor is chosen so that the terms in Eq. (10.1.16) of second order in q/a should cancel. Eqs. (10.1.15) and (10.1.17) are both satisfied to leading order in q/a if we take

$$g/f = 4i\pi G\dot{\bar{\varphi}}a/q. \tag{10.1.19}$$

The terms in Eq. (10.1.16) of first order in q/a then give $2\dot{f} + 2Hf = 0$, so $f \propto 1/a$. For reasons that will soon become clear, we will choose the common constant factor in both f and

g so that $f = 1/(2\pi)^{3/2}\sqrt{2q}a$. With this normalization, $\delta\varphi_q(t)$ and $\Psi_q(t)$ are defined as the solution of Eqs. (10.1.15)-(10.1.17) that satisfies the initial condition, that for $a(t) \to 0$,

$$\delta\varphi_q(t) \to \frac{1}{(2\pi)^{3/2}a(t)\sqrt{2q}} \exp\left(-iq\int_{t_*}^{t}\frac{dt'}{a\,(t')}\right) \tag{10.1.20}$$

$$\Psi_q(t) \to \frac{4i\pi G\dot{\bar{\varphi}}(t)}{(2\pi)^{3/2}\sqrt{2q^3}} \exp\left(-iq\int_{t_*}^{t}\frac{dt'}{a\,(t')}\right) \tag{10.1.21}$$

The complex conjugate $\delta\varphi_q^*(t), \Psi_q^*(t)$ is another independent solution, and since the system of equations (10.1.15)-(10.1.17) is second order, these are the only solutions.

In general, the fields $\delta\varphi(\mathbf{x},t)$ and $\Psi(\mathbf{x},t)$ satisfying Eqs. (10.1.12)(10.1.14) can be written as superpositions of these two solutions, which the reality of the fields requires to take the form

$$\delta\varphi(\mathbf{x},t) = \int d^3q\left[\delta\varphi_q(t)e^{i\mathbf{q}\cdot\mathbf{x}}\alpha(\mathbf{q}) + \delta\varphi_q^*(t)e^{-i\mathbf{q}\cdot\mathbf{x}}\alpha^*(\mathbf{q})\right] \tag{10.1.22}$$

$$\Psi(\mathbf{x},t) = \int d^3q\left[\Psi_q(t)e^{i\mathbf{q}\cdot\mathbf{x}}\alpha(\mathbf{q}) + \Psi_q^*(t)e^{-i\mathbf{q}\cdot\mathbf{x}}\alpha^*(\mathbf{q})\right] \tag{10.1.23}$$

Now we must say something about the coefficients $\alpha(\mathbf{q})$ and $\alpha^*(\mathbf{q})$. For this purpose, we use the canonical commutation relations of the fields. The interaction of the scalar field with gravitation makes these commutation relations rather complicated, but they become simple at very early times. For any given q, we can find a time sufficiently early so that the expansion rate H is negligible compared with q/a. For such early times, both $\dot{\bar{\varphi}}$ and $\ddot{\bar{\varphi}}$ on the right-hand side of Eq. (10.1.17) become negligible, so Ψ_q becomes negligible, as can also be seen by noting from Eqs. (10.1.18) and (10.1.19) that the ratio of Ψ_q to $\delta\varphi_q$ vanishes like $a(t)$ for $a(t) \to 0$. At such early times, as far as $\alpha(\mathbf{q})$ and $\alpha^*(\mathbf{q})$ are concerned, we can find the canonical commutation relations by using Eq. (10.1.1) with $g_{\mu\nu}$ taken as the unperturbed Robertson-Walker metric:

$$I_\varphi = \int d^4x\mathcal{L}_\varphi, \quad \mathcal{L}_\varphi = a^3(t)\left[\frac{1}{2}\left(\frac{\partial\varphi}{\partial t}\right)^2 - \frac{1}{2a^2(t)}\frac{\partial\varphi}{\partial x^i}\frac{\partial\varphi}{\partial x^i} - V(\varphi)\right]$$

The canonical conjugate to the field φ is then $\pi = \partial\mathcal{L}_\varphi/\partial\dot{\varphi} = a^3\dot{\varphi}$. Since the unperturbed fields are c-numbers, the commutators of perturbations are the same as the commutators of the fields themselves. This gives, for very early times,

$$[\delta\varphi(\mathbf{x},t),\delta\varphi(\mathbf{y},t)] = 0, \quad [\delta\varphi(\mathbf{x},t),\delta\dot{\varphi}(\mathbf{y},t)] = ia^{-3}(t)\delta^3(\mathbf{x}-\mathbf{y}) \tag{10.1.24}$$

With $\delta\varphi_q(t)$ normalized to satisfy Eq. (10.1.20) for $a(t) \to 0$, these commutation relations imply that $\alpha(\mathbf{q})$ and $\alpha^*(\mathbf{q})$ behave as conventionally normalized annihilation and creation operators

$$[\alpha(\mathbf{q}),\alpha\,(\mathbf{q}')] = 0, \quad [\alpha(\mathbf{q}),\alpha^*\,(\mathbf{q}')] = \delta^3\,(\mathbf{q}-\mathbf{q}') \tag{10.1.25}$$

It may come as a surprise that the same annihilation and creation operators appear in the scalar field and gravitational perturbations, but it should be
kept in mind that Ψ does not represent gravitational radiation, whose quanta are created and annihilated by independent operators. Rather, Ψ is an auxiliary field, given by Eq. (10.1.14) as a functional of the inflaton field $\delta\varphi$, in much the same way that in the Coulomb gauge quantization of quantum electrodynamics, the time-component of the vector potential is a functional of the charged matter fields. [3]

It may be noted in passing that the scalar field we have constructed is in accord with the Principle of Equivalence of general relativity. As long as we do not concern ourselves with co-moving wave numbers below some infrared limit Q (as for example by confining measurements to a cube of co-moving volume less than $(2\pi/Q)^3$), then at times for which $Q/a \gg H$ the form of the scalar field $\delta\varphi(\mathbf{x},t)$ ought to be essentially the same as a free massless real scalar field $\varphi(\mathbf{x},t)$ in ordinary Minkowskian space-time:
$\int \frac{d^3k}{(2\pi)^{3/2}\sqrt{2k}} \left[A(\mathbf{k})e^{i\mathbf{k}\cdot a\mathbf{x}} \exp\left(-i\int k dt\right) + A^*(\mathbf{k})e^{-i\mathbf{k}\cdot a\mathbf{x}} \exp\left(i\int k dt\right)\right]$,
where $A(\mathbf{k})$ and $A^*(\mathbf{k})$ are annihilation and creation operators, satisfying the familiar commutation relations

$$[A(\mathbf{k}), A(\mathbf{k}')] = 0, \quad [A(\mathbf{k}), A^*(\mathbf{k}')] = \delta^3(\mathbf{k}-\mathbf{k}').$$

Note that the space coordinate appearing in the exponentials is $a\mathbf{x}$, because in the $K=0$ Robertson-Walker metric this is the vector that measures proper distances. It follows that $\mathbf{k}$ is related to the time-independent co-moving wave number vector $\mathbf{q}$ by $\mathbf{k} = \mathbf{q}/a$, so it is time-dependent, which is why we had to write the time-dependence factors as $\exp\left(\mp i\int k dt\right)$ rather than $\exp(\mp ikt)$. Also, $\delta^3(\mathbf{k}-\mathbf{k}') = a^3\delta^3(\mathbf{q}-\mathbf{q}')$, so we can define operators $\alpha(\mathbf{q})$ that satisfy (10.1.25) as $\alpha(\mathbf{q}) = a^{-3/2}A(\mathbf{k})$. Changing the variable of integration from $\mathbf{k}$ to $\mathbf{q}$ then gives the field for $a(t) \to 0$ as

$$\int \frac{d^3q/a^3}{(2\pi)^{3/2}\sqrt{2q/a}} \left[a^{3/2}\alpha(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{x}} \exp\left(-i\int q dt/a\right) + a^{3/2}\alpha^*(\mathbf{q})e^{-i\mathbf{q}\cdot\mathbf{x}} \exp\left(i\int q dt/a\right)\right]$$

in agreement with Eqs. (10.1.20) and (10.1.22).
Finally, we have to choose the quantum state of the inflaton field during inflation. Though there are other possibilities, the simplest and most natural

assumption is that the state of the universe during inflation is the vacuum $|0\rangle$, defined so that [4]

$$\alpha(\mathbf{q})|0\rangle = 0, \quad \langle 0 \mid 0\rangle = 1 \tag{10.1.26}$$

One other possibility is that it is a linear combination of α and α^* rather than α that annihilates the state $|0\rangle$.[5] Another is that inflation takes place in the presence of a thermal distribution of inflatons. [6]

But although not certain, the assumption (10.1.26) is at least plausible. The initial condition (10.1.20), which picks out a particular solution of the second-order system (10.1.15)-(10.1.17), is imposed at an early time, at which $q/a \gg H$. At such times, we can treat the action as if it were nearly time-independent, so there exists a Hamiltonian operator H which to a good approximation generates the time-dependence of the fields:

$$[H, \delta\varphi(\mathbf{x},t)] \simeq -i\dot{\varphi}(\mathbf{x},t)$$

According to Eq. (10.1.18), for $q/a \gg H$ the time dependence of the coefficient function $\varphi_q(t)$ is given approximately by

$$\dot{\varphi}_q \simeq -i(q/a)\varphi_q$$

so

[473] See QTF, Vol. I, Eq. (8.2.9).

$$[H, \alpha(\mathbf{q})] \simeq -(q/a)\alpha(\mathbf{q}).$$

Hence if a state $|\psi\rangle$ is an eigenstate of H with energy E, then $\alpha(\mathbf{q})|\psi\rangle$ is an eigenstate of H with a lower energy $\simeq E - q/a$, unless $|\psi\rangle$ is the state $|0\rangle$ for which $\alpha(\mathbf{q})|0\rangle = 0$, which is therefore the state of lowest energy. Just as in ordinary laboratory physics, we expect any other state to decay into the state $|0\rangle$ of lowest energy, although there remains a question whether the decay occurs rapidly enough to be effective in the period before horizon exit. [7]

In the state satisfying Eq. (10.1.26), quantum averages $\langle 0|\delta\varphi\left(\mathbf{x}_1, t\right)\delta\varphi\left(\mathbf{x}_2, t\right)\cdots|0\rangle$ (as well as those also involving $\Psi(\mathbf{x}, t)$) may be calculated by moving all annihilation operators $\alpha(\mathbf{q})$ to the right and all creation operators $\alpha^*(\mathbf{q})$ to the left, picking up commutators when a $\alpha(\mathbf{q})$ is moved to the right past a $\alpha^*\left(\mathbf{q}'\right)$ or a $\alpha^*(\mathbf{q})$ is moved to the left past a $\alpha\left(\mathbf{q}'\right)$. The result is then given by Wick's theorem: [8] The quantum averages of products of $\delta\varphi$ s and/or

Ψ s are Gaussian in the sense of Appendix E (except that we must keep track of the order of operators), with the pairings given by vacuum expectation values of products of the paired fields:

$$\langle 0|\delta\varphi(\mathbf{x}, t)\delta\varphi(\mathbf{y}, t)|0\rangle = \int d^3q \left|\delta\varphi_q(t)\right|^2 e^{i\mathbf{q}\cdot(\mathbf{x}-\mathbf{y})} \tag{10.1.27}$$

$$\langle 0|\Psi(\mathbf{x}, t)\Psi(\mathbf{y}, t)|0\rangle = \int d^3q \left|\Psi_q(t)\right|^2 e^{i\mathbf{q}\cdot(\mathbf{x}-\mathbf{y})} \tag{10.1.28}$$

$$\langle 0|\delta\varphi(\mathbf{x}, t)\Psi(\mathbf{y}, t)|0\rangle = \int d^3q \delta\varphi_q(t)\Psi_q^*(t) e^{i\mathbf{q}\cdot(\mathbf{x}-\mathbf{y})} \tag{10.1.29}$$

$$\langle 0|\Psi(\mathbf{x}, t)\delta\varphi(\mathbf{y}, t)|0\rangle = \int d^3q \Psi_q(t)\delta\varphi_q^*(t) e^{i\mathbf{q}\cdot(\mathbf{x}-\mathbf{y})} \tag{10.1.30}$$

For instance,

$$\begin{aligned}
&\langle 0|\delta\varphi(\mathbf{w}, t)\Psi(\mathbf{x}, t)\Psi(\mathbf{y}, t)\delta\varphi(\mathbf{z}, t)|0\rangle \\
&\quad = \langle 0|\delta\varphi(\mathbf{w}, t)\Psi(\mathbf{x}, t)|0\rangle\langle 0|\Psi(\mathbf{y}, t)\delta\varphi(\mathbf{z}, t)|0\rangle \\
&\qquad + \langle 0|\delta\varphi(\mathbf{w}, t)\Psi(\mathbf{y}, t)|0\rangle\langle 0|\Psi(\mathbf{x}, t)\delta\varphi(\mathbf{z}, t)|0\rangle \\
&\qquad + \langle 0|\delta\varphi(\mathbf{w}, t)\delta\varphi(\mathbf{z}, t)|0\rangle\langle 0|\Psi(\mathbf{x}, t)\Psi(\mathbf{y}, t)|0\rangle
\end{aligned}$$

These are quantum averages, not averages over an ensemble of classical field configurations. We see this most clearly in Eqs. (10.1.29) and (10.1.30), which give complex results for the averages of products of real fields, and consequently also depend on the order of the fields. [9] Just as in the measurement of a spin in the laboratory, some sort of decoherence must set in; the field configurations must become locked into one of an ensemble of classical configurations, with ensemble averages given by the quantum expectation values calculated as in Eqs. (10.1.27)-(10.1.30) It is not apparent just how this happens, but it is clear that decoherence cannot occur until expectation values of products of real fields become real, which for free fields will also imply that the expectation values do not depend on the order of the fields. As we shall see, this happens after perturbations leave the horizon, when the various functions $\varphi_q(t)$ and $\Psi_q(t)$ become dominated by a single solution of the field equations, which (since the field equations are real) is necessarily real up to a possible complex factor. Once the universe becomes classical

[474] This state is often called the Bunch-Davies vacuum: see T. S. Bunch and P. C. W. Davies, Proc. Roy. Soc. Ser. A 360, 117 (1978).

[5] E. Mottola, Phys. Rev. D 31, 754 (1985); B. Allen, Phys. Rev. D 32, 3136 (1985).

[6] K. Bhattacharya, S. Mohanty, and R. Rangarajan, Phys. Rev. Lett. 96, 121302 (2006).

[7] This is studied by C. Armendariz-Picon, J. Cosm. & Astropart. Phys. 0702, 031 (2007) [astroph/0612288].

[8] See QTF, Vol. I, Sec. 9.1.

in this sense, we can invoke the Ergodic Theorem of Appendix D to interpret averages over ensembles of possible classical universes as averages over the position of the observer in our universe.

For any given potential $V(\bar{\varphi})$, it is always possible to find $\varphi_q(t)$ and $\Psi_q(t)$ by numerically solving Eqs. (10.1.15)-(10.1.17), subject to the initial

conditions (10.1.20) and (10.1.21). But this is complicated, and gives more information than we can use. Between the time of inflation and the present there intervenes a so-called reheating period when the energy of the inflaton field φ is converted into ordinary matter and radiation. We know essentially nothing about this process, so the solutions for $\delta\varphi$ and Ψ during inflation do not have an immediate interpretation in terms of observations of the present universe. Fortunately, the reheating era (and other ill-understood eras) occur when all cosmological fluctuations of observational interest are outside the horizon. The one use that we can make of the solutions for the fields during inflation is to calculate some quantity that becomes conserved outside the horizon, and that thus provides an initial condition for the evolution of perturbations after they re-enter the horizon.

Here we will concentrate on the quantity $\mathcal{R}$ discussed in Section 5.4, defined in Newtonian gauge by

$$\mathcal{R} \equiv -\Psi + H\delta u \tag{10.1.31}$$

With a single scalar field this is conserved outside the horizon during inflation, because as shown in Section 5.4 there are always two solutions for which $\mathcal{R}$ is constant for $q/a \ll H$, and as already noted the equations for inflation with a single scalar field only have two independent solutions. Using Eqs. (10.1.11) and (10.1.22)-(10.1.23) we find that during inflation

$$\mathcal{R}(\mathbf{x},t) = \int d^3q \left[\mathcal{R}_q(t)e^{i\mathbf{q}\cdot\mathbf{x}}\alpha(\mathbf{q}) + \mathcal{R}_q^*(t)e^{-i\mathbf{q}\cdot\mathbf{x}}\alpha^*(\mathbf{q})\right] \tag{10.1.32}$$

where

$$\mathcal{R}_q = -\Psi_q - H\delta\varphi_q/\dot{\bar{\varphi}} \tag{10.1.33}$$

Of course, the quantum averages of products of $\mathcal{R}(\mathbf{x},t)$ s are Gaussian in the same sense as those of $\Psi_q(t)$ and $\delta\varphi_q(t)$, with pairings given by the expectation value

$$\langle 0|\mathcal{R}(\mathbf{x},t)\mathcal{R}(\mathbf{y},t)|0\rangle = \int d^3q e^{i\mathbf{q}\cdot(\mathbf{x}-\mathbf{y})}\,|\mathcal{R}_q(t)|^2 \tag{10.1.34}$$

Instead of calculating $\Psi_q(t)$ and $\delta\varphi_q(t)$ and then using the results to calculate $\mathcal{R}_q(t)$, it is much more convenient to solve a differential equation for $\mathcal{R}_q(t)$ itself. This equation can be derived in Newtonian gauge with some trouble from Eqs. (10.1.15)-(10.1.17), but it is more easily derived in a different gauge, defined by the conditions

$$\delta\varphi_q = 0, \quad B_q = 0$$

Inspection of Eqs. (B.66)-(B.68) shows that in this gauge

$$\delta\rho = \delta p = \frac{1}{2}h_{00}\dot{\bar{\varphi}}^2 = -\frac{1}{2}E\dot{\bar{\varphi}}^2, \quad \delta u = 0$$

Also, as mentioned in Appendix B, the energy-momentum tensor for a single real scalar field has the perfect fluid form, so the anisotropic inertia π_{ij} vanishes. The gravitational field equations (5.1.44) and (5.1.46) and the energy conservation equation (5.1.49) then give

[479] For a discussion of this point, see D. H. Lyth and D. Seery, astro-ph/0607647.

$$0 = H\dot{E} + 2\left(3H^2 + \dot{H}\right)E + a^{-2}\nabla^2 A - \ddot{A} - 6H\dot{A} + 2a^{-1}H\nabla^2 F,$$
$$0 = -HE + \dot{A}$$
$$0 = -\frac{1}{2}\frac{\partial}{\partial t}(E\dot{H}) - 3H\dot{H}E - a^{-1}\dot{H}\nabla^2 F + \frac{3}{2}\dot{H}\dot{A},$$

in which we have again used the relation $\dot{H} = -4\pi G|\dot{\bar{\varphi}}|^2$. Eliminating E and F yields a differential equation for A :

$$\ddot{A} + \left(3H - \frac{2\dot{H}}{H} + \frac{\ddot{H}}{\dot{H}}\right)\dot{A} - \frac{1}{a^2}\nabla^2 A = 0.$$

The gauge invariant formula (5.4.22) tells us that in this gauge $\mathcal{R} = A/2$, so the same equation applies to $\mathcal{R}$. Going over to its Fourier transform, this gives what is sometimes known as the Mukhanov-Sasaki equation: [10]

$$\frac{d^2\mathcal{R}_q}{d\tau^2} + \frac{2}{z}\frac{dz}{d\tau}\frac{d\mathcal{R}_q}{d\tau} + q^2\mathcal{R}_q = 0, \tag{10.1.35}$$

where τ is the conformal time

$$\tau \equiv \int_{t_*}^{t} \frac{dt'}{a\,(t')} \tag{10.1.36}$$

with t_* an arbitrary time, to be chosen later, and

$$z \equiv \frac{a\dot{\bar{\varphi}}}{H}. \tag{10.1.37}$$

The initial condition is provided by returning to Newtonian gauge, and using Eqs. (10.1.20) and (10.1.21) in Eq. (10.1.33). Assuming that $a(t)\dot{\bar{\varphi}}^2(t)/H(t)$ vanishes in the limit $a(t) \to 0$, only the term in Eq. (10.1.33) proportional to $\delta\varphi_q$ contributes in this limit, and we find that for $a(t) \to 0$:

$$\mathcal{R}_q(t) \to -\frac{H(t)}{(2\pi)^{3/2}\sqrt{2q}a(t)\dot{\bar{\varphi}}(t)}\exp(-iq\tau) \tag{10.1.38}$$

For a given potential we must integrate Eq. (10.1.35) out from $a = 0$ to beyond the horizon, where $q/a \ll H$. In this limit we can drop the $q^2\mathcal{R}_q$ term in Eq. (10.1.35), which then has two solutions, a dominant solution with $\mathcal{R}_q$ a non-zero constant, and a solution for which $\mathcal{R}_q$ approaches zero with $d\mathcal{R}_q/d\tau$ decaying as $1/z^2$. It is the constant limit $\mathcal{R}_q^o$ of $\mathcal{R}_q(t)$ outside the horizon that we need. (As we saw in Section 5.3, the value of the quantity ζ_q that is sometimes used in analyses of cosmological fluctuations is given far outside the horizon by $\zeta_q^o = \mathcal{R}_q^o$.)

It is striking that within the scope of the general assumptions made here, it is not necessary to make any arbitrary assumptions about the strength of cosmological fluctuations. For any given potential, we need only solve equation (10.1.35) with the initial condition (10.1.38), and carry the solution forward in time to when $\mathcal{R}_q(t)$ reaches its constant value $\mathcal{R}_q^o$ outside the horizon.

On what features of the potential does $\mathcal{R}_q^o$ depend? The initial behavior (10.1.38) of $\mathcal{R}_q(t)$ deep inside the horizon is independent of the nature of the potential, while outside the horizon $\mathcal{R}_q(t)$ simply becomes constant. Thus $\mathcal{R}_q^o$ can depend only on the behavior of the potential

[4][7][10] V. S. Mukhanov, JETP Lett. 41, 493 (1986); S. Sasaki, Prog. Theor. Phys. 76, 1036 (1986); V. S. Mukhanov, H. A. Feldman, and R.H. Brandenberger, Phys. Rep. 215, 203 (1992); E. D. Stewart and D. H. Lyth, Phys. Lett. B 302, 171 (1993).

$V(\varphi)$ for values of φ near the value taken by $\bar{\varphi}(t)$ at the time the perturbation leaves the horizon.

This has an important implication for the part of the era of inflation that can be revealed through observations of scalar fluctuations. As we have just seen, in observing a fluctuation with co-moving wave number q, we learn about the time t_q of horizon exit during inflation, when $q/a\left(t_q\right) = H\left(t_q\right)$. To put this another way, the number of e-foldings $\mathcal{N}(q)$ between the time t_q that we learn about in observing a perturbation of wave number q and the beginning of the radiation-dominated era at a time t_1 is [11]

$$\mathcal{N}(q) \equiv \ln\left(\frac{a\left(t_1\right)}{a\left(t_q\right)}\right) = \mathcal{N}_0 + \ln\left(\frac{a\left(t_0\right) H\left(t_0\right)}{q}\right) + \ln\left(\frac{H\left(t_q\right)}{H\left(t_1\right)}\right) \tag{10.1.39}$$

where

$$\mathcal{N}_0 \equiv \ln\left(\frac{a\left(t_1\right) H\left(t_1\right)}{a\left(t_0\right) H\left(t_0\right)}\right) \tag{10.1.40}$$

and as usual t_0 is the present time. In particular, we cannot observe any fluctuation unless q is large enough to have entered the horizon by the present time, which requires that the present physical wavelength $a\left(t_0\right)/q$ be less than the present horizon distance $\approx 1/H\left(t_0\right)$, so the maximum number of e-foldings before the beginning of the radiation-dominated era that can

ever be observed is

$$\mathcal{N}_{\max} = \mathcal{N}_0 + \ln\left(\frac{H(t)}{H\left(t_1\right)}\right)_{\max}, \tag{10.1.41}$$

in which the second term on the right is the maximum value of $\ln[H(t)/\ H\left(t_1\right)]$ for t within the last $\mathcal{N}_{\max}$ e-foldings of inflation. In general we expect energy to be lost during inflation and in the reheating phase at the end of inflation, so $H\left(t_{\text{exit}}\right) > H\left(t_1\right)$ and hence $\mathcal{N}_{\max} > \mathcal{N}_0$, but for slow roll inflation it may not be a bad approximation to neglect this energy loss, in which case the second term in Eq.(10.1.41) can be neglected, and

$$\mathcal{N}_{\max} \simeq \mathcal{N}_0. \tag{10.1.42}$$

The reason that we have chosen to write the formula for $\mathcal{N}(q)$ as in Eq. (10.1.39) is that we have already calculated $\mathcal{N}_0$ in Section 4.1; it is

$$\mathcal{N}_0 = \ln\left(\frac{\rho_1^{1/4}}{0.037h\text{eV}}\right), \tag{10.1.43}$$

where ρ_1 is the energy density at the beginning of the radiation-dominated era. For instance, if we take $h = 0.7$ and $\rho_1 = G^{-2} = [1.2 \times 10^{19}\text{GeV}]^4$, then $\mathcal{N}_0 \simeq 68$, and according to Eq.(10.1.42) we can only explore the final 68 e-foldings of inflation. We will re-evaluate this bound in Section 10.3, with a better estimate of ρ_1.

There is one form of the potential for which the constant $\mathcal{R}_q^o$ can be calculated analytically, with no further approximations. [12] It is the exponential potential

$$V(\varphi) = ge^{-\lambda\varphi} \tag{10.1.44}$$

(with g and λ arbitrary real constants) which we have already considered in Section 4.2. Of course, the potential cannot have this form for all φ, or inflation would never end, but

[4711] A. R. Liddle and S. M. Leach, Phys. Rev. D 68, 103503 (2003) [astro-ph/0305263].

as remarked in the previous paragraph, in the calculation of $\mathcal{R}_q^o$ it is only relevant that the potential should take this form for values of the field near the value it takes at horizon crossing.

The solution of Eqs. (10.1.6) and (10.1.7) for the exponential potential is

$$\bar{\varphi}(t) = \frac{1}{\lambda} \ln\left(\frac{8\pi G g \epsilon^2 t^2}{3-\epsilon}\right), \tag{10.1.45}$$

and

$$H = 1/\epsilon t \tag{10.1.46}$$

where ϵ is the positive dimensionless quantity

$$\epsilon \equiv -\frac{\dot{H}}{H^2} = \frac{\lambda^2}{16\pi G} \tag{10.1.47}$$

Inflation with this potential is often called power-law inflation, because $a \propto t^{1/\epsilon}$. It is convenient for the moment to normalize the co-moving coordinates so that

$$a = t^{1/\epsilon}. \tag{10.1.48}$$

Note that $a\dot{\bar{\varphi}}^2/H \propto t^{(1-\epsilon)/\epsilon}$, so our previous assumption that $a\dot{\bar{\varphi}}^2/H$ vanishes for $a(t) \to 0$ is satisfied if $\epsilon < 1$, as we will assume.

For this potential it is convenient to take the constant t_* in the definition (10.1.36) of conformal time as $t_* = \infty$, in which case the conformal time is negative

$$\tau = -\left(\frac{\epsilon}{1-\epsilon}\right) t^{-(1-\epsilon)/\epsilon} \tag{10.1.49}$$

As t and $a(t)$ run from zero to infinity, τ runs from $-\infty$ to zero. In terms of τ, Eq. (10.1.35) now reads

$$\frac{d^2\mathcal{R}_q}{d\tau^2} - \frac{2}{(1-\epsilon)\tau}\frac{d\mathcal{R}_q}{d\tau} + q^2\mathcal{R}_q = 0 \tag{10.1.50}$$

whose solutions are proportional to τ^ν times a Hankel function $H_\nu^{(1)}(-q\tau)$ or $H_v^{(2)}(-q\tau)$, where

$$v = \frac{1}{2}\left(1 + \frac{2}{1-\epsilon}\right) = \frac{3}{2} + \frac{\epsilon}{1-\epsilon}. \tag{10.1.51}$$

For large real x the first Hankel function has the asymptotic behavior

$$H_v^{(1)}(x) \to \sqrt{\frac{2}{\pi x}} \exp(ix - iv\pi/2 - i\pi/4)$$

and $H_\nu^{(2)}(x) = H_\nu^{(1)*}(x)$, so the initial condition (10.1.38) picks out the solution $\propto \tau^\nu H_\nu^{(1)}(-q\tau)$, and fixes its normalization so that

$$\begin{aligned}\mathcal{R}_q(t) = &-\frac{\lambda\sqrt{\pi}}{4(2\pi)^{3/2}\epsilon}\left(\frac{\epsilon}{1-\epsilon}\right)^{-1/(1-\epsilon)} \\ &\times \exp\left(\frac{i\pi v}{2} + \frac{i\pi}{4}\right)(-\tau)^v H_v^{(1)}(-q\tau)\end{aligned} \tag{10.1.52}$$

[4712] L. F. Abbott and M. B. Wise, Nucl. Phys. B 244, 541 (1984); F. Lucchin and S. Matarrese, Phys. Rev. D 32, 1316 (1985); Phys. Lett. B 164, 282 (1985); D. H. Lyth and E. D. Stewart, Phys. Lett. B 274, 168 (1992).

For small real x

$$H_v^{(1)}(x) \to \frac{-i\Gamma(v)}{\pi}\left(\frac{x}{2}\right)^{-v}$$

so outside the horizon, in the limit $q/a \ll H$ where $-q\tau \ll 1$, the quantity $\mathcal{R}_q$ approaches a constant $\mathcal{R}_q^o$, given by Eq. (10.1.52) as

$$\mathcal{R}_q^o = i\frac{\lambda\Gamma(\nu)}{8\sqrt{2}\pi^2\epsilon}\left(\frac{\epsilon}{1-\epsilon}\right)^{-1/(1-\epsilon)}\exp\left(\frac{i\pi\nu}{2}+\frac{i\pi}{4}\right)\left(\frac{q}{2}\right)^{-\nu}. \tag{10.1.53}$$

For purposes of comparison with results for other potentials, it is convenient to rewrite this in terms of quantities at the time t_q of horizon crossing, defined by

$$q/a\,(t_q) = H\,(t_q)\,. \tag{10.1.54}$$

Solving this gives $t_q = (\epsilon q)^{\epsilon/(1-\epsilon)}$ and $H\,(t_q) = q^{-\epsilon/(1-\epsilon)}\epsilon^{-1/(1-\epsilon)}$. It is also convenient to use Eq. (10.1.47) to express λ in terms of ϵ and G. Then Eq. (10.1.53) can be written
$\zeta_q^o = \mathcal{R}_q^o = i\sqrt{16\pi G}q^{-3/2}H\,(t_q)\,\frac{\Gamma(v)2^{\nu-3/2}}{4\pi^2\sqrt{\epsilon}}(1-\epsilon)^{1/(1-\epsilon)}\exp\left(\frac{i\pi\nu}{2}+\frac{i\pi}{4}\right)$.

Note that this formula does not depend on the convention we have chosen in Eq. (10.1.48) for the constant factor in $a(t)$.

To have a sufficient number of e-foldings in inflation, it is necessary for the potential to be fairly flat, so we are chiefly interested in the case where ϵ is small. For $\epsilon \ll 1$, we use $\Gamma(3/2) = \sqrt{\pi}/2$, and write Eq. (10.1.55) as

$$\zeta_q^o = \mathcal{R}_q^o = -i\sqrt{16\pi G}q^{-3/2}H\,(t_q)\,\frac{1}{8\pi^{3/2}\sqrt{\epsilon}}. \tag{10.1.56}$$

The most important point is that for a nearly flat potential, for which $H\,(t_q)$ is nearly q-independent, $\mathcal{R}_q^o$ is nearly proportional to $q^{-3/2}$. This result is not limited to potentials of exponential form, as long as the potential is fairly flat for fields near $\bar{\varphi}\,(t_q)$.

The approximate $q^{-3/2}$ dependence of $\mathcal{R}_q^o$ on q is well supported by observation. We saw in Section 8.1 that the power spectral function $P(k)$ that governs the distribution of dark matter has a dependence on the physical wave number k that for small k is given by a factor $k^4\left|\mathcal{R}_{ka_0}^o\right|^2$, so any potential that is fairly flat for fields near $\bar{\varphi}\,(t_q)$ gives a spectrum close to the HarrisonZel'dovich spectrum, $P(k) \propto k$ for small k. Also, as we saw in Section 6.3, the spectral function $\mathcal{P}_\phi(q)$ for the Newtonian gravitational potential is
proportional for long wavelengths to $\left|\mathcal{R}_q^o\right|^2$, which as shown in Section 2.6 implies that (aside from the integrated Sachs-Wolfe effect and the effects of reionization) the quantity $\ell(\ell+1)C_\ell$ becomes independent of ℓ for small ℓ. And we saw in Sections 7.2 and 7.4 that the near proportionality of $\mathcal{R}_q^o$ to $q^{-3/2}$ gives results for the correlations of temperature and polarization fluctuations in the cosmic microwave background at larger values of ℓ in good agreement with observation.

To be more precise, Eq. (10.1.53) shows that for the exponential potential, the scalar spectral index defined in Section 7.2 by $\left|\mathcal{R}_q^o\right|^2 \propto q^{-4+n_S}$ has the constant value

$$n_S = 4 - 2v = 1 - \frac{2\epsilon}{1-\epsilon}$$

The third-year WMAP result $n_S = 0.958 \pm 0.016$ quoted in Section 7.2 thus shows that for the exponential potential, $\epsilon = 0.021 \pm 0.008$. This is small enough so that we can use the approximate formula (10.1.56). Writing $\left|\mathcal{R}_q^o\right|^2 \simeq |N|^2 q^{-3}$, we see that for the exponential potential

$$|N|^2 \simeq \frac{GH_{\rm exit}^2}{4\pi^2\epsilon}$$

where $H_{\rm exit}$ is the expansion rate at horizon exit, now ignoring its weak dependence on q. The third-year WMAP result $|N|^2 = (1.93 \pm 0.12) \times 10^{-10}$ quoted in Section 7.2 thus shows that for the exponential potential, $H_{\rm exit} \simeq 2\pi|N|\sqrt{\epsilon/G} \simeq 1.5 \times 10^{14}$GeV, corresponding to an energy density $3H_{\rm exit}^2/8\pi G \simeq [2.6 \times 10^{16}{\rm GeV}]^4$.

Of course, the exponential potential is just one special case, with the special property that $V''/V = (V'/V)^2$, so there is no reason to expect these results to apply in detail for other potentials. In Section 10.3 we will see what to expect for general potentials, under the slow-roll approximation. But within the slow-roll approximation, the results for general potentials are similar in order of magnitude to those we have already found for the exponential potential.

We are now in a position to consider the plausibility of the assumption that the action has the simple form (10.1.1). [13] With a single real scalar field χ, there is no loss of generality in taking the action in the form (10.1.1) if

we assume that the action contains at most two spacetime derivatives. The most general action in this case is

$$I_\chi = \int d^4x \sqrt{-\,{\rm Det}\, g}\left[-\frac{1}{2}K(\chi)g^{\mu\nu}\frac{\partial\chi}{\partial x^\mu}\frac{\partial\chi}{\partial x^\nu} - U(\chi)\right],$$

where K and U are real functions of χ that are arbitrary, except that unitarity requires $K \geq 0$. This can be put in the form (10.1.1) by introducing $\varphi = \int d\chi K^{1/2}(\chi)$.

But why should the terms in the action contain no more than two spacetime derivatives? This is only one of a number of similar questions: Why should the action of gravitation contain only the Einstein term, proportional to $\int d^4x\sqrt{-\,{\rm Det}g}R$, and not other generally covariant terms with more than two derivatives of the metric? And why should the action of the standard electroweak model or quantum chromodynamics contain only renormalizable terms?

We do not know the answers to any of these questions with any certainty, but we have at least a plausible possible answer. Any of these additional terms in the action would involve operators of higher dimensionality, but dimensional analysis requires that such terms must be accompanied by coefficients containing additional negative powers of some fundamental mass. If that mass is large enough enough, then the additional terms in the action are suppressed under ordinary conditions. For the success of general relativity, it is only necessary that the length $1/M$ be sub-macroscopic, but unless we impose a condition of baryon and lepton conservation, the fundamental mass appearing in the action of quarks and leptons has to be at least of order 10^{16}GeV, in order to suppress proton decay below experimental bounds.

The inflaton field is considered to take values of the order of the Planck mass, so there is nothing to suppress arbitrary powers of the scalar field in the inflaton action, which is why we take the potential $V(\varphi)$ as an arbitrary function. But spacetime derivatives of the inflaton field yield factors of order q/a, so each additional spacetime derivative would introduce a factor of order q/aM. At horizon exit, q/a equals H, which we have seen is of order 10^{14}GeV, while we would expect M to be much larger, somewhere

[47][13] There is a large literature on other possibilities. Here is a partial list: C. Armendariz-Picon, T. Damour, and V. Mukhanov, Phys. Lett. B 458, 209 (1999) [hep-th/9904075]; J. Martin and R. H. Brandenberger, Phys. Rev. D 63, 1235012 (2001) [hep-th/0005209]; R. H. Brandenberger and J. Martin, Mod. Phys. Lett. A 16, 999 (2001) [astro-ph/0005432]; J. C. Niemeyer, Phys. Rev. D 63, 123502 (2001) [astro-ph/0005533]; J. C. Niemeyer and R. Parentani, Phys. Rev. D 64, 101301 (2001) [astro-ph/0101451]; A. Kempf and J. C. Niemeyer, Phys. Rev. D 64, 103501 (2001) [astro-ph/0103225]; R. Easther, B. R. Greene, W. H. Kinney, and G. Shiu, Phys. Rev. D 64, 103502 (2001) [hep-th/0104102];

[47]Phys. Rev. D 66, 023518 (2002); Phys. Rev. D 67, 063508 (2003) [hep-th/0110226]; R. H. Brandenberger, S. E. Joras, and J. Martin, Phys. Rev. D 66, 083514 (2002) [hep-th/0112122]; N. Kaloper, M. Kleban,

in the range from $|V|^{1/4} \approx 10^{16}$GeV to $1/\sqrt{G} \simeq 10^{19}$GeV. Hence it is plausible that at and after horizon exit, and for some time before horizon exit, terms with more than the minimum number of spacetime derivatives are suppressed. This is all we need to justify the calculations of this section, at least as a good first approximation, as long as we stick to a single inflaton field. The possibility of more than one inflaton field will be discussed in Section 10.4.

10.10.2 10.2 Tensor fluctuations during inflation

The fluctuations in the tensor field $D_{ij}(\mathbf{x}, t)$ during inflation can be treated in much the same way as the scalar fluctuations considered in the previous section. Since the tensor anisotropic inertia π^T_{ij} vanishes for scalar field theories, the field equation (5.1.53) for the tensor modes takes the simple form

$$\nabla^2 D_{ij} - a^2 \ddot{D}_{ij} - 3a\dot{a}\dot{D}_{ij} = 0. \tag{10.2.1}$$

We recall also that D_{ij} satisfies conditions that eliminate any vector or scalar contributions:

$$D_{ij} = D_{ji}, \quad D_{ii} = 0, \quad \partial_i D_{ij} = 0 \tag{10.2.2}$$

The plane wave solutions have the form $e_{ij}\mathcal{D}_q(t)e^{i\mathbf{q}\cdot\mathbf{x}}$, where $\mathcal{D}_q(t)$ satisfies the differential equation

$$\ddot{\mathcal{D}}_q + 3H\dot{\mathcal{D}}_q + \left(q^2/a^2\right)\mathcal{D}_q = 0 \tag{10.2.3}$$

and e_{ij} is a time-independent polarization tensor satisfying the conditions

$$e_{ij} = e_{ji}, \quad e_{ii} = 0, \quad q_i e_{ij} = 0 \tag{10.2.4}$$

We recall from Section 5.2 that for a given unit vector $\hat{q}$, there are two independent polarization tensors satisfying these conditions. For $\hat{q}$ in the three-direction, these can be chosen to have components

$$e_{11} = -e_{22} = 1/\sqrt{2}, \quad e_{12} = e_{21} = \pm i/\sqrt{2}, \quad e_{3i} = e_{i3} = 0. \tag{10.2.5}$$

For $\hat{q}$ in any other direction we define e_{ij} to be the tensor obtained by applying to (10.2.5) the standard rotation (7.4.12) that takes the three-axis into the direction of $\hat{q}$. The polarization tensors constructed in this way are called $e_{ij}(\hat{q}, \pm 2)$, because Eq. (10.2.5) describes a wave of helicity ± 2.

At early times, when $q/a \gg H$, Eq. (10.2.3) has WKB solutions of the form

$$\mathcal{D}_q(t) \to h(t) \exp\left(-iq \int_{t_*}^{t} \frac{dt'}{a\,(t')}\right) \tag{10.2.6}$$

where $h(t)$ varies much more slowly than the argument of the exponential, and t_* is arbitrary. Then the terms in Eq. (10.2.3) of second order in q/a cancel, while the terms of first order give $2\dot{h} + 2Hh = 0$, so that $h \propto a^{-1}$. From now on we will define $\mathcal{D}_q(t)$ as the solution of Eq. (10.2.3) which is normalized so that, for $a(t) \to 0$,

A. Lawrence, and S. Shenker, Phys. Rev. D 66, 123510 (2002) [hep-th/0201158]; N. Kaloper, M. Kleban, A. Lawrence, S. Shenker, and L. Susskind, J. High Energy Phys. 0211, 037 (2002) [hep-th/0209231]; U. H. Danielsson, Phys. Rev. D 66, 023511 (2002); L. Bergström and U. H. Danielsson, J. High Energy Phys. 12, 38 (2002); C. P. Burgess, J. M. Cline, F. Lemieux, and R. Holman, J. High Energy Phys. 2, 48 (2003); J. Martin and R. Brandenberger, Phys. Rev. D 68, 063513 (2003); J. Martin and C. Ringeval, Phys. Rev. D 69, 083515 (2004); T. Okamoto and E. A. Lim, Phys. Rev. D 69, 083519 (2004). R. de Putter and E. V. Linder, 0705.0400.

$$\mathcal{D}_q(t) \to \frac{\sqrt{16\pi G}}{(2\pi)^{3/2}\sqrt{2q}a(t)} \exp\left(-iq\int_{t_*}^{t}\frac{dt'}{a\left(t'\right)}\right). \tag{10.2.7}$$

Since $\mathcal{D}_q(t)$ and $\mathcal{D}_q^*(t)$ are a complete set of solutions of Eq. (10.2.3), the most general real tensor field satisfying the conditions (10.2.1) and (10.2.2) takes the form

$$\begin{aligned} D_{ij}(\mathbf{x},t) = \sum_{\lambda=\pm 2}\int d^3q \left[\mathcal{D}_q(t)e^{i\mathbf{q}\cdot\mathbf{x}}\beta(\mathbf{q},\lambda)e_{ij}(\hat{q},\lambda)\right. \\ \left.+\mathcal{D}_q^*(t)e^{-i\mathbf{q}\cdot\mathbf{x}}\beta^*(\mathbf{q},\lambda)e_{ij}^*(\hat{q},\lambda)\right] \end{aligned} \tag{10.2.8}$$

(That is, instead of characterizing two independent solutions $\mathcal{D}_{1q}(t)$ and $\mathcal{D}_{2q}(t)$ of Eq. (10.2.3) by their behavior at late times, as we did in Section 5.2, we now take them as $\mathcal{D}_q(t)$ and $\mathcal{D}_q^*(t)$, characterized by their behavior at early times.) With $\mathcal{D}_q(t)$ normalized to satisfy Eq. (10.2.7), the canonical commutation relations require that

$$\left[\beta(\mathbf{q},\lambda),\beta\left(\mathbf{q}',\lambda'\right)\right]=0, \quad \left[\beta(\mathbf{q},\lambda),\beta^*\left(\mathbf{q}',\lambda'\right)\right]=\delta^3\left(\mathbf{q}-\mathbf{q}'\right)\delta_{\lambda\lambda'}. \tag{10.2.9}$$

Thus $\beta(\mathbf{q},\lambda)$ and $\beta^*(\mathbf{q},\lambda)$ can be interpreted as the annihilation and creation operators for a graviton of helicity λ.

As in the case of scalar perturbations, we assume that during inflation the universe is in a quantum state $|0\rangle$ satisfying the vacuum condition $\beta(\mathbf{q},\lambda)|0\rangle = 0$. Then expectation values of products of D s are Gaussian, with pairings

$$\langle 0|D_{ij}(\mathbf{x},t)D_{kl}(\mathbf{y},t)|0\rangle = \int d^3q e^{i\mathbf{q}\cdot(\mathbf{x}-\mathbf{y})}\left|\mathcal{D}_q(t)\right|^2\Pi_{ij,kl}(\hat{q}) \tag{10.2.10}$$

where $\Pi_{ij,kl}(\hat{q})$ is the helicity sum given by Eq. (5.2.25):

$$\begin{aligned} \Pi_{ij,k\ell}(\hat{q}) \equiv& \sum_{\lambda} e_{ij}(\hat{q},\lambda)e_{k\ell}^*(\hat{q},\lambda) \\ =&\delta_{ik}\delta_{j\ell}+\delta_{i\ell}\delta_{jk}-\delta_{ij}\delta_{k\ell}+\delta_{ij}\hat{q}_k\hat{q}_\ell+\delta_{k\ell}\hat{q}_i\hat{q}_j-\delta_{ik}\hat{q}_j\hat{q}_\ell-\delta_{i\ell}\hat{q}_j\hat{q}_k \\ &-\delta_{jk}\hat{q}_i\hat{q}_\ell-\delta_{j\ell}\hat{q}_i\hat{q}_k+\hat{q}_i\hat{q}_j\hat{q}_k\hat{q}_\ell \end{aligned} \tag{10.2.11}$$

For any given potential, the function $\mathcal{D}_q(t)$ is to be calculated by integrating the differential equation (10.2.3) with the initial condition (10.2.7).

Outside the horizon, for $q/a \ll H$, the solution becomes a constant $\mathcal{D}_q^o$, which provides an initial condition for the gravitational wave when it re-enters the horizon. In carrying this out, it is again useful to replace ordinary time as the independent variable with conformal time

$$\tau \equiv -\int_t^\infty \frac{dt'}{a\left(t'\right)} \tag{10.2.12}$$

and take $t_* = \infty$ in Eq. (10.2.7). Eq. (10.2.3) then becomes

$$\frac{d^2\mathcal{D}_q}{d\tau^2}+2Ha\frac{d\mathcal{D}_q}{d\tau}+q^2\mathcal{D}_q=0 \tag{10.2.13}$$

Note that

$$Ha=\frac{da}{dt}=\frac{1}{a}\frac{da}{d\tau}$$

Thus Eq. (10.2.13) is the same as the Mukhanov-Sasaki equation (10.1.35) except that $z \equiv a\dot{\bar{\varphi}}/H$ is replaced with a.

The initial condition (10.2.7) for $q/aH \gg 1$ is independent of the details of the potential, while for $q/aH \ll 1$ the tensor amplitude $\mathcal{D}_q(t)$ simply approaches a constant $\mathcal{D}_q^o$. Hence, just as for the scalar amplitude $\mathcal{R}_q^0$, the tensor amplitude $\mathcal{D}_q^o$ outside the horizon can only depend on the behavior of the potential at values of $\bar{\varphi}(t)$ at around the time of horizon exit, when q/aH is of order unity. Thus the measurements of the tensor amplitude after horizon re-entry can only tell us about the last $\mathcal{N}e$-foldings of inflation, with $\mathcal{N}$ bounded by Eq. (10.1.40), just as for scalar modes.

As for scalar modes, it is useful to consider as a test case the one potential for which $\mathcal{D}_q^o$ can be calculated analytically without relying on the slow-roll approximation, the exponential potential (10.1.44). For this potential $\dot{\bar{\varphi}}/H$ is the constant $2\epsilon/\lambda$, so Eq. (10.1.37) gives $z \propto a$, and in this case Eq. (10.2.13) is precisely the same as the Mukhanov-Sasaki equation. Also, the initial condition (10.2.7) is the same as the initial condition (10.1.38), except for a factor $-2\epsilon\sqrt{16\pi G}/\lambda = -2\sqrt{\epsilon}$. Hence for the exponential potential:

$$\mathcal{D}_q(t)/\mathcal{R}_q(t) = -2\epsilon\sqrt{16\pi G}/\lambda = -2\sqrt{\epsilon} \tag{10.2.14}$$

Of course, the values $\mathcal{D}_q^o$ and $\mathcal{R}_q^o$ outside the horizon then also have this ratio. This is usually expressed in terms of a scalar tensor ratio, conventionally defined as

$$r_q \equiv 4\left|\mathcal{D}_q^o/\mathcal{R}_q^o\right|^2 \tag{10.2.15}$$

We see that for the exponential potential, r_q has the wavelength-independent value

$$r = 16\epsilon \tag{10.2.16}$$

We saw in the previous section that for the exponential potential the thirdyear WMAP results give $\epsilon = 0.021 \pm 0.008$, so that $r = 0.34 \pm 0.13$. This is almost incompatible with the upper bound $r < 0.3$ on r set (for this value of ϵ) by the third-year WMAP results, [1] so a potential $V(\varphi)$ that is exponential around the value that φ takes at the time of horizon exit is almost ruled out. (Of course, we already knew that the potential could not be exponential over the whole range of φ, for then inflation would never end, but we are here not relying on any assumption about the form of the potential except for the values of φ taken around the time of horizon exit.) To analyze tensor as well as scalar perturbations for more general potentials, we need to invoke the slow-roll approximation, to which we turn in the next section.

10.10.3 10.3 Fluctuations during inflation: The slow-roll approximation

It is not possible to calculate the scalar and tensor perturbations $\mathcal{R}_q(t)$ and $\mathcal{D}_q(t)$ analytically for general potentials. However, the need (discussed in Chapter 4) for a substantial number of e-foldings of expansion during inflation suggests that H should have been slowly varying during an era long compared with $1/H$. We will therefore now assume that H varies little throughout a "slow-roll" era, during which q/aH goes from much less to much greater than unity. We saw in Sections 10.1 and 10.2 that the asymptotic values $\mathcal{R}_q^o$ and $\mathcal{D}_q^o$ depend only on the evolution of the fields around the time of horizon crossing, when $q/a \approx H$, so it will not be necessary for us to assume that the slow-roll era extends back to the beginning of the expansion, or forward to the end of inflation.

We will work with the Mukhanov-Sasaki equation (10.1.35): [1]

$$\frac{d^2\mathcal{R}_q}{d\tau^2} + \frac{2}{z}\frac{dz}{d\tau}\frac{d\mathcal{R}_q}{d\tau} + q^2\mathcal{R}_q = 0, \tag{10.3.1}$$

where τ is the conformal time, and

$$z \equiv \frac{a\dot{\bar{\varphi}}}{H}. \tag{10.3.2}$$

Recalling once again that $\dot{H} = -4\pi G\dot{\bar{\varphi}}^2$, we can write

$$\frac{1}{z}\frac{dz}{d\tau} = aH(1+\delta+\epsilon) \tag{10.3.3}$$

where for a general potential

$$\epsilon \equiv -\dot{H}/H^2, \quad \delta \equiv \ddot{H}/2H\dot{H} \tag{10.3.4}$$

We also need a formula for aH in terms of τ. For this purpose, we note that

$$\frac{d}{d\tau}\left(\frac{1}{aH}\right) = -1+\epsilon \tag{10.3.5}$$

So far, everything is exact. The slow-roll approximation requires that ϵ and δ are small during the era of horizon crossing, which has the consequence that ϵ varies little during this era, because

$$\dot{\epsilon} = 2\epsilon(\epsilon+\delta)H \tag{10.3.6}$$

We will also assume that δ varies little. [2] (Of course, ϵ and δ cannot be strictly constant except for an exponential potential, for which $H \propto 1/t$ and therefore $\delta = -\epsilon$.) Integrating Eq. (10.3.5) then gives, for a suitable choice of an additive constant in τ,

$$aH = -\frac{1}{(1-\epsilon)\tau} \tag{10.3.7}$$

As in the case of the exponential potential, τ is negative, and its magnitude goes from $-\tau \gg 1/q$ early in inflation when $q/aH \gg 1$, to $-\tau \ll 1/q$ for $q/aH \ll 1$. Using Eqs. (10.3.3) and (10.3.7) in Eq. (10.3.1) then gives, to first order in ϵ and δ,

$$\frac{d^2\mathcal{R}_q}{d\tau^2} - \frac{2(1+\delta+2\epsilon)}{\tau}\frac{d\mathcal{R}_q}{d\tau} + q^2\mathcal{R}_q = 0 \tag{10.3.8}$$

The general solution of this equation for constant δ and ϵ is a linear combination of $\tau^\nu H_\nu^{(1)}(-q\tau)$ and $\tau^\nu H_\nu^{(2)}(-q\tau)$, where now

$$v = \frac{3}{2} + 2\epsilon + \delta \tag{10.3.9}$$

This agrees to first order in ϵ with the result (10.1.51) for the exponential potential, for which $\delta = -\epsilon$.

Now, we are not assuming that the slow-roll approximation applies all the way back to the beginning of the expansion, but fortunately this is not necessary. Eq. (10.1.38) should

[471] See Figure 14 of D. Spergel et al., Astrophys. J. Suppl. Ser. 170, 377 (2007) [astro-ph/0603449].

[1] This is usually given in the equivalent form

$$\frac{d^2u_q}{d\tau^2} + \left[q^2 - \frac{1}{z}\frac{d^2z}{d\tau^2}\right]u_q = 0$$

where $u_q \equiv z\mathcal{R}_q$. It is easier to work with it in the form (10.3.1), if only because in this way we only need to calculate the first derivative of z.

[472] This is the case for a power-law potential, under the same condition $|\varphi| \gg 1/\sqrt{4\pi G}$ that was found necessary in Section 4.2 to justify the slow-roll approximation. For the consequences of dropping this assumption, see S. Dodelson and E. D. Stewart, Phys. Rev. D 65, 101301 (2002) [astro-ph/0109354]; E. D. Stewart, Phys. Rev. D 65, 103508 (2002) [astro-ph/0110322].

still apply by the beginning of the slow-roll era, when q/aH is still very large, so this initial condition fixes the solution during the whole of the slow-roll era as

$$\mathcal{R}_q(t) = -\frac{\sqrt{-\pi\tau}}{2(2\pi)^{3/2}z(\tau)}e^{i\pi v/2+i\pi/4}H_v^{(1)}(-q\tau). \tag{10.3.10}$$

(Note that Eqs. (10.3.3) and (10.3.7) give $z(\tau) \propto \tau^{-v+1/2}$ during the slowroll era.) Then late in the slow-roll era, when $q/aH \ll 1$, Eq. (10.3.10) has the asymptotic value

$$\mathcal{R}_q^o = i\frac{\sqrt{-\tau}\Gamma(v)}{2\sqrt{\pi}(2\pi)^{3/2}z(t)}e^{i\pi v/2+i\pi/4}\left(\frac{-q\tau}{2}\right)^{-v}, \tag{10.3.11}$$

which is constant because Eqs. (10.3.3) and (10.3.7) give $z \propto \tau^{-v+1/2}$. Thus $\mathcal{R}_q^o$ has the q-dependence [3]

$$\mathcal{R}_q^o \propto q^{-\nu} = q^{-3/2-2\epsilon-\delta} \tag{10.3.12}$$

This may be regarded as a generalization of the result $\mathcal{R}_q^o \propto q^{-3/2-\epsilon}$ that we found for the exponential potential, to the case where $\delta \neq -\epsilon$.

Because $\mathcal{R}_q^o$ is time-independent, it can be calculated by setting t in Eq. (10.3.11) to any convenient value. We shall evaluate it at the time t_q of horizon crossing, defined as in Sections 10.1 and 10.2 by

$$q/a\left(t_q\right) = H\left(t_q\right). \tag{10.3.13}$$

(But note that Eq. (10.3.10) with $t = t_q$ does not give the correct value of $\mathcal{R}_q^o$.) Ignoring corrections of order ϵ or δ except in exponents, Eq. (10.3.7) gives

$$\tau\left(t_q\right) = -\frac{1}{(1-\epsilon)q} \simeq -\frac{1}{q}. \tag{10.3.14}$$

Also, $\dot{H} = -4\pi G\dot{\bar{\varphi}}^2$, so

$$\dot{\bar{\varphi}}\left(t_q\right) = \pm\sqrt{-\dot{H}\left(t_q\right)}/\sqrt{4\pi G} = \pm H\left(t_q\right)\sqrt{\epsilon\left(t_q\right)}/\sqrt{4\pi G}.$$

and so

$$z\left(t_q\right) = \frac{\pm\sqrt{\epsilon\left(t_q\right)}q}{H\left(t_q\right)\sqrt{4\pi G}} \tag{10.3.15}$$

The slow-roll approximation thus gives [4]

$$\mathcal{R}_q^o = \mp i\frac{\sqrt{16\pi G}q^{-3/2}H\left(t_q\right)}{8\pi^{3/2}\sqrt{\epsilon\left(t_q\right)}} \tag{10.3.16}$$

This is the same as the result (10.1.56) for an exponential potential, except that the factor $1/\sqrt{\epsilon\left(t_q\right)}$ now contributes to the q-dependence of $\mathcal{R}_q^o$. (To check that the q dependence of $H\left(t_q\right)$ and $\epsilon\left(t_q\right)$ gives the extra factor $q^{-2\epsilon-\delta}$, we differentiate Eq. (10.3.13) with respect to q, and find

[473] E. D. Stewart and D. H. Lyth, Phys. Lett. B 302, 171 (1993). This calculation was carried to the next order of the slow-roll approximation by A. R. Liddle and M. S. Turner, Phys. Rev. D 50, 758 (1994). For a review, see J. E. Lidsey, A. R. Liddle, E. W. Kolb, E. J. Copeland, T. Barreiro, and M. Abney, Rev. Mod. Phys. 69, 373 (1997).

$$\frac{dt_q}{dq} = \frac{1}{a\left(t_q\right)\left(H^2\left(t_q\right) + \dot{H}\left(t_q\right)\right)}$$

Then

$$\frac{q}{H\left(t_q\right)}\frac{dH\left(t_q\right)}{dq} = \frac{\dot{H}\left(t_q\right)}{H^2\left(t_q\right) + \dot{H}\left(t_q\right)} = -\frac{\epsilon\left(t_q\right)}{1-\epsilon\left(t_q\right)}$$

and Eq. (10.3.6) gives

$$\frac{q}{\epsilon\left(t_q\right)}\frac{d\epsilon\left(t_q\right)}{dq} = \frac{2\left(\epsilon\left(t_q\right) + \delta\left(t_q\right)\right)}{1-\epsilon\left(t_q\right)}$$

Hence, replacing the denominators $1-\epsilon$ with unity, $H\left(t_q\right) \propto q^{-\epsilon}$ and $\epsilon\left(t_q\right) \propto q^{2\epsilon+2\delta}$, and so $H\left(t_q\right)/\sqrt{\epsilon\left(t_q\right)} \propto q^{3/2-v}$, which gives Eq. (10.3.16) the q-dependence (10.3.12).)

Now let us apply the slow-roll approximation to the tensor modes. In general, the tensor wave equation (10.2.3) can be written as

$$\frac{d^2\mathcal{D}_q}{d\tau^2} + 2aH\frac{d\mathcal{D}_q}{d\tau} + q^2\mathcal{D}_q = 0 \tag{10.3.17}$$

During the slow-roll era, we can use Eq. (10.3.7) to put this in the form

$$\frac{d^2\mathcal{D}_q}{d\tau^2} - \frac{2}{(1-\epsilon)\tau}\frac{d\mathcal{D}_q}{d\tau} + q^2\mathcal{D}_q = 0 \tag{10.3.18}$$

The general solution for constant ϵ is a linear combination of $\tau^\mu H_\mu^{(1)}(-q\tau)$ and $\tau^\mu H_\mu^{(2)}(-q\tau)$, with

$$\mu = \frac{3}{2} + \frac{\epsilon}{1-\epsilon}. \tag{10.3.19}$$

The solution during the slow-roll era that satisfies the initial condition (10.2.7) is

$$\mathcal{D}_q(t) = \frac{\sqrt{16\pi G}}{(2\pi)^{3/2}\sqrt{2q}a(t)}\sqrt{\frac{-q\tau\pi}{2}}\exp(i\mu\pi/2 + i\pi/4)H_\mu^{(1)}(-q\tau). \tag{10.3.20}$$

(Note that Eq. (10.3.7) gives $(\tau/a)da/d\tau = aH\tau = -1/(1-\epsilon)$, so $a \propto \tau^{-1/(1-\epsilon)}$ and therefore $\sqrt{-\tau}/a \propto \tau^\mu$.) The asymptotic solution for $q/a \ll H$ is then

$$\mathcal{D}_q^o = -i\frac{\sqrt{16\pi G}\Gamma(\mu)\sqrt{-\tau}}{2\sqrt{\pi}(2\pi)^{3/2}a(t)}\exp(i\mu\pi/2 + i\pi/4)\left(\frac{-q\tau}{2}\right)^{-\mu}. \tag{10.3.21}$$

Thus $\mathcal{D}_q^o$ has the q-dependence [3]

$$\mathcal{D}_q^o \propto q^{-\mu} \simeq q^{-3/2-\epsilon}. \tag{10.3.22}$$

In contrast with the case of the exponential potential, the asymptotic q-dependence of the tensor modes is in general different from that of the scalar modes.

We can give a more convenient expression for $\mathcal{D}_q^o$ by setting $t = t_q$ in Eq. (10.3.21). Then using Eq. (10.3.14) and $a\left(t_q\right) = q/H\left(t_q\right)$ and taking $\epsilon \to 0$ everywhere but in the q-dependence of $H\left(t_q\right)$, Eq. (10.3.21) gives [5]

[474] S. W. Hawking, Phys. Lett. B 115, 295 (1982); A. A. Starobinsky, Phys. Lett. B 117, 175 (1982); A. Guth and S.-Y. Pi, Phys. Rev. Lett. 49, 1110 (1982); J. M. Bardeen, P. J. Steinhardt and M. S. Turner, Phys. Rev. D 28, 679 1983); D. H. Lyth, Phys. Lett. B 147, 403 (1984); B 150, 465 (1985); Phys. Rev. D 31, 1792 (1985).

$$\mathcal{D}_q^o = i\frac{\sqrt{16\pi G}H\left(t_q\right)}{4\pi^{3/2}q^{3/2}}. \tag{10.3.23}$$

It is conventional to write the q-dependence of the squared magnitudes of the tensor and scalar amplitudes outside the horizon as

$$\left|\mathcal{D}_q^o\right|^2 \propto q^{-3+n_T(q)}, \quad \left|\mathcal{R}_q^o\right|^2 \propto q^{-4+n_S(q)}. \tag{10.3.24}$$

Then in the slow-roll approximation, Eqs. (10.3.12) and (10.3.22) give

$$n_T(q) = -2\epsilon\left(t_q\right), \quad n_S(q) = 1 - 4\epsilon\left(t_q\right) - 2\delta\left(t_q\right). \tag{10.3.25}$$

(As a check, recall that for the exponential potential $\delta = -\epsilon$, so Eq. (10.3.25) gives $n_S = 1 - 2\epsilon$, in agreement with the result of Section 10.1.) Also,
comparison of Eqs. (10.3.16) and (10.3.23) yields the relation

$$r(q) = 16\epsilon\left(t_q\right) = -8n_T(q) \tag{10.3.26}$$

where, with the conventional definition of $r, r(q) \equiv 4\left|\mathcal{D}_q^o/\mathcal{R}_q^o\right|^2$. This relation among measurable quantities is known in the literature as the slow-roll consistency condition. For any potential other than the exponential potential the scalar/tensor ratio r depends on q.

For inflation with a single inflaton field, the relation $\dot{H} = -4\pi G\dot{\bar{\varphi}}^2$ tells us that $\epsilon(t)$ is always positive, but $\delta(t)$ can have either sign, so in general Eq. (10.3.25) gives $n_T(q) < 0$, while $n_S(q)$ can be greater or less than unity. Nevertheless, experience with many models shows [6] that physically plausible potentials that are not finely tuned tend to have $n_S(q)$ less than unity, and even less than 0.98 . But for slow-roll inflation, ϵ and δ are small, so $n_S(q)$ cannot be very much less than unity. Thus the general picture of slow-roll inflation received some support from the third-year WMAP result quoted in Section 7.2, that $n_S = 0.958 \pm 0.016$.

As we saw in Chapter 7, the quantities $q^3\left|\mathcal{D}_q^o\right|^2$ and $q^3\left|\mathcal{R}_q^o\right|^2$ provide a measure of the contribution of tensor and scalar fluctuations to the multipole coefficients $C_{TT,\ell}$ in the angular distribution of the cosmic microwave background temperature. From Eq. (10.3.26) we can see that the tensor modes are likely to contribute much less to the $C_{TT,\ell}$ than the scalar modes. Also, Eq. (10.3.16) and the fact that anisotropies in the cosmic microwave background temperature are small but not too small to be observed indicates that the Hubble constant during the slow-roll era must be small compared with the Planck mass $1/\sqrt{G}$, but not too small. In the slow-roll limit, where $v \simeq 3/2$, we can write $\left|\mathcal{R}_q^o\right|^2 = |N|^2 q^{-3}$, and Eq. (10.3.16) shows that

$$|N|^2 = \frac{16\pi G H_{\rm exit}^2}{64\pi^3\left|\epsilon_{\rm exit}\right|} = \frac{(8\pi G)^2\bar{\rho}_{\rm exit}}{96\pi^3\left|\epsilon_{\rm exit}\right|},$$

where the subscript "exit" denotes the time of horizon exit, and in accordance with the slow-roll approximation we here ignore the weak dependence of this time on q. As we saw in Section 7.2, the factor $|N|^2$ has the value $(1.93 \pm 0.12) \times 10^{-10}$, so

$$\frac{\bar{\rho}_{\rm exit}}{\left|\epsilon_{\rm exit}\right|} = \left[(6.70 \pm 0.10) \times 10^{16}{\rm GeV}\right]^4 \tag{10.3.27}$$

Measurements of cosmic microwave background anisotropies have so far been sensitive only to the spectral index of the scalar rather than the tensor modes, and therefore have yielded

[475] A. A. Starobinsky, JETP Lett. 30, 683 (1979).

[476] M. B. Hoffman and M. S. Turner, Phys. Rev D 64, 023506 (2001); W. H. Kinney, Phys. Rev. D 66, 083508 (2002); H. V. Peiris et al., Astrophys. J. Suppl. Ser. 1148, 213 (2003); G. Efstathiou and K. J. Mack, J. Cosm. Astropart. Phys. 05, 008 (2005); L. A. Boyle, P. J. Steinhardt, and N. Turok, Phys. Rev. Lett. 96, 111301 (2006).

information only on $\delta + 2\epsilon$, not ϵ, but unless there is a cancelation between δ and 2ϵ, these measurements suggest that $|\epsilon|$ is probably not much greater than a few percent. If for instance we take $|\epsilon|_{\rm exit} = 0.05$, then $\bar\rho_{\rm exit} \simeq [3.2 \times 10^{16}{\rm GeV}]^4$. In any case, we now see that in inflationary theories, the smallness of cosmic fluctuations before horizon re-entry is simply a reflection of the fact that, for reasons that are still mysterious, the energy scale defined by the energy density of the universe at horizon exit is a few orders of magnitude less than the Planck energy scale, $(8\pi G)^{-1/2} = 2.4 \times 10^{18}{\rm GeV}$.

This mystery is strongly reminiscent of another mystery encountered in elementary particle physics: the unification energy scale, where the three coupling constants of the electroweak and strong interactions all come together, [7] is about $2 \times 10^{16}{\rm GeV}$, also a few orders of magnitude less than the Planck energy scale $(8\pi G)^{-1/2}$. Perhaps they are the same mystery.

The measured values of n_S and $|N|$ and the observational upper limit on the tensor/scalar ratio r already allow us to put useful constraints on the inflaton potential. We saw in the previous section that this data is close to ruling out any potential $V(\bar\varphi)$ with an exponential dependence on $\bar\varphi$ for the values that $\bar\varphi(t)$ takes around the time of horizon exit. To go further, it is useful first to express ϵ and δ in terms of the potential. Using the general relation (4.2.3)

$$\dot H = -4\pi G\dot{\bar\varphi}^2,$$

and the slow-roll formula (4.2.8)

$$\dot{\bar\varphi} = -\frac{V'(\varphi)}{3H} = -\frac{V'(\varphi)}{\sqrt{24\pi G V(\varphi)}},$$

gives

$$\epsilon(t) = \frac{1}{16\pi G}\left(\frac{V'(\bar\varphi(t))}{V(\bar\varphi(t))}\right)^2. \tag{10.3.28}$$

Also, the time-derivative of Eq. (4.2.3) gives $\ddot H = -8\pi G\dot{\bar\varphi}\ddot{\bar\varphi}$, and using Eq. (4.2.12) then gives

$$\delta(t) = \frac{1}{16\pi G}\left(\frac{V'^2(\bar\varphi(t))}{V^2(\bar\varphi(t))} - \frac{2V''(\bar\varphi(t))}{V(\bar\varphi(t))}\right). \tag{10.3.29}$$

10.10.4 10.3 Fluctuations during inflation: The slow-roll approximation

For instance, for a power-law potential $V(\varphi) \propto \varphi^\alpha$, we have

$$\epsilon(t) = \frac{\alpha^2}{16\pi G\bar\varphi^2(t)}, \quad \delta(t) = \frac{2\alpha - \alpha^2}{16\pi G\bar\varphi^2(t)}, \tag{10.3.30}$$

so $\delta(t) = (2/\alpha - 1)\epsilon(t)$, and therefore

$$n_S(q) = 1 - \frac{(\alpha+2)r(q)}{8\alpha} \tag{10.3.31}$$

The experimental bound on n_S depends on the value assumed for r, so observations define an allowed region in the $n_S - r$ plane. [8] At present, the straight line (10.3.31) intersects the (68% confidence level) allowed region for all positive α, even when WMAP three year data is combined with data from the CBI and VSA microwave backgrounds, or from the Sloan or 2 dF

[477] See e.g. QTF, Vol. II, Sec. 28.2.

sky surveys. However, low values of α are favored, and even a modest shrinking of the allowed area would rule out high values of α.

To go further, we need to say something about the value of the scalar field at horizon exit. For this purpose, we can make use of the relation (4.2.14), which gives the number $\Delta\mathcal{N}$ of e-foldings when the scalar field goes from φ_1 to φ_2, under the assumption that the slow-roll approximation holds over this period, as

$$\Delta\mathcal{N} = -\int_{\varphi_1}^{\varphi_2} \left(\frac{8\pi G V(\varphi)}{V'(\varphi)}\right) d\varphi \tag{10.3.32}$$

If $|V'/V| \simeq \sqrt{16\pi G\epsilon}$ is essentially constant over the range of φ from φ_1 to φ_2, then the number of e-foldings associated with this change in φ is

$$\Delta\mathcal{N} = \Delta\varphi\sqrt{4\pi G/\epsilon} \tag{10.3.33}$$

Lyth [9] has used this relation in the case where φ_1 and φ_2 are the field values at horizon exit for wave numbers corresponding to $\ell \simeq 1$ and $\ell \simeq 100$, for which $\Delta\mathcal{N} = \ln 100 = 4.6$, to show that if ϵ is large enough to give a detectable tensor mode, then the scalar field must change by an amount that is at least as large as the Planck scale $1/\sqrt{4\pi G}$.

If we make the strong assumption that the slow roll approximation holds over the whole era from horizon exit to the end of inflation, but do not now assume that ϵ is necessarily constant through this era, then for a power-law

potential $V(\varphi) \propto \varphi^\alpha$, Eq. (10.3.32) gives the number of e-foldings from horizon exit for a wave number q to the end of inflation as

$$\mathcal{N}(q) = \frac{4\pi G}{\alpha}\left[\bar{\varphi}^2(t_q) - \bar{\varphi}^2(t_{\rm end})\right], \tag{10.3.34}$$

where t_q and $t_{\rm end}$ are the times of horizon exit and the end of inflation, respectively. If we further assume that $|\bar{\varphi}(t_{\rm end})| \ll |\bar{\varphi}(t_q)|$, then

$$\bar{\varphi}^2(t_q) \simeq \frac{\alpha\mathcal{N}(q)}{4\pi G}, \tag{10.3.35}$$

so Eq. (10.3.30) gives

$$\epsilon(t_q) \simeq \frac{\alpha}{4\mathcal{N}(q)}, \quad \delta(t_q) \simeq \frac{2-\alpha}{4\mathcal{N}(q)}$$

and therefore [10]

$$n_S(q) \simeq 1 - \frac{\alpha+2}{2\mathcal{N}(q)}, \quad r(q) = \frac{4\alpha}{\mathcal{N}(q)}. \tag{10.3.36}$$

We noted in Section 10.1 that if the energy density at the beginning of the radiation-dominated era is the Planck density G^{-2}, then $\mathcal{N}(q) \simeq 68$ for wave numbers that are just coming into the horizon at the present, and correspondingly less for larger wave numbers; for instance, for wave numbers corresponding to $\ell = 100, \mathcal{N}$ would be less by an amount $\ln 100 = 4.6$. To derive a better estimate of $\mathcal{N}$, we can use Eq. (10.3.27). Under the risky assumption that the energy density ρ_1 at the beginning of the radiationdominated era is the same as at horizon exit, and taking $\epsilon = O(.02)$, we have $\rho_1 \simeq [2.5 \times 10^{16}\text{GeV}]^{1/4}$, so Eq. (10.1.43) with $h = 0.7$ shows that for a wave number that just enters the horizon at the present, $\mathcal{N} \simeq 62$, while for the

[478] See Figure 14 of D. Spergel et al., Astrophys. J. Suppl. Ser. 170, 377 (2007) [astro-ph/0603449].

[9] D. H. Lyth, Phys. Rev. Lett. 78, 1861 (1997) [hep-ph/9606387]. Also see G. Efstathiou and K. J. Mack, J. Cosm. & Astropart. Phys. 05, 008 (2005) [astro-ph/0503360]; R. Easther, W. H. Kinney, and B. A. Powell, J. Cosm. Astropart. Phys. 08, 004 (2006) [astro-ph/0601276].

wave number corresponding to $\ell \simeq 100, \mathcal{N} \simeq 57$. Taking $\mathcal{N} = 60$, for a quadratic potential Eq. (10.3.36) gives $n_S = .97$ and $r = 0.13$, which is consistent with the WMAP third-year results, while for a quartic potential Eq. (10.3.35) gives $n_S = 0.95$ and $r = 0.26$, which is barely outside the range allowed by WMAP. [11] Any $\alpha > 4$ is ruled out. But this conclusion is contingent on the assumption of a slow-roll inflation from the time of horizon exit until the end of inflation, with $V(\bar{\varphi}(t)) \propto \bar{\varphi}^\alpha(t)$ over this whole period.

10.10.5 10.4 Multifield inflation

Observations of the cosmic microwave background and large scale structure indicate that the primordial scalar fluctuations outside the horizon are

- nearly Gaussian,
- adiabatic,
- nearly scale invariant, in the sense that $\mathcal{R}_q^o$ is nearly proportional to $q^{-3/2}$,
- weak, in the sense that $q^{3/2}\mathcal{R}_q^o \ll 1$. (We saw in Section 7.2 that $q^{3/2}\mathcal{R}_q^o$ is of order 10^{-5}.)

We have seen in Sections 10.1 and 10.3 that these properties of primordial scalar fluctuations follow under the assumptions that

1. The energy density during inflation receives appreciable contributions from just a single real "inflaton" scalar field. This implies that the fluctuations are adiabatic during inflation, in which case they remain so thereafter.

2. During the era of horizon exit (say, for q/aH falling from 10 to 0.1), H is sufficiently small so that q/a is less than whatever fundamental scale (such as the grand unification scale or the Planck scale) characterizes the theory, not only during this era but for some time before it. (For the case considered in Section 10.1, we estimated that $H \approx 10^{14}$GeV, which is probably small enough.) This implies that during this era the scalar field is described by a simple effective action, involving no more than two spacetime derivatives. It follows that for some time before the era of horizon exit the inflaton behaves like a free field, so that the fluctuations are Gaussian.

3. For observed fluctuations, in the era of horizon exit inflation is "slowroll," in the sense that $|\dot{H}|/H^2 \ll 1$ and $|\ddot{H}/H\dot{H}| \ll 1$. Together with assumptions 1 and 2, this implies that the fluctuations are nearly scale invariant.

But there is no particular reason to believe that the energy density during inflation is dominated by a single scalar field, so we are naturally led to consider the case of several inflaton fields $\varphi^n(x)$. We shall show that the same properties of primordial scalar fluctuations follow if we make assumptions 2 and 3 , but replace assumption 1 with

1*. The energy in all the scalar fields is converted at the end of inflation into ordinary matter and radiation in local thermal equilibrium, in which all chemical potentials vanish. (Baryon and lepton number would then have to be generated later, as discussed in Section 3.3.)

Whatever the number of scalar fields, under assumption 2, the effective action of the scalar fields is dominated by terms with a minimum number of spacetime derivatives during and after

[4710] D. H. Lyth and A. Riotto, Phys. Rep. 314, 1 (1999).

[11] D. Spergel, ref. 8.

the era of horizon exit and for some time before it. For arbitrary numbers of scalar fields, the most general such action takes the form

$$I_\varphi = \int d^4x \sqrt{-\operatorname{Det} g}\left[-\frac{1}{2}g^{\mu\nu}\gamma_{nm}(\varphi)\frac{\partial\varphi^n}{\partial x^\mu}\frac{\partial\varphi^m}{\partial x^\nu} - V(\varphi)\right], \tag{10.4.1}$$

where $V(\varphi)$ is an arbitrary real potential, repeated scalar field indices are summed, and $\gamma_{nm}(\varphi)$ is an arbitrary real symmetric positive-definite matrix, which we shall call the field metric. (This matrix must be positive-definite to give the right sign to commutators of fields and their time derivatives.) The energy-momentum tensor, which serves as the source of the gravitational field, is derived as described in Appendix B from the action (10.4.1), and takes the form

$$T_{\mu\nu} = g_{\mu\nu}\left[-\frac{1}{2}g^{\rho\sigma}\gamma_{nm}(\varphi)\frac{\partial\varphi^n}{\partial x^\rho}\frac{\partial\varphi^m}{\partial x^\sigma} - V(\varphi)\right] + \gamma_{nm}(\varphi)\frac{\partial\varphi^n}{\partial x^\mu}\frac{\partial\varphi^m}{\partial x^\nu}. \tag{10.4.2}$$

The scalar field equations are derived from the principle that I_φ must be stationary with respect to infinitesimal variations in the the scalar fields, and take the Euler-Lagrange form

$$\begin{aligned}\frac{\partial}{\partial x^\mu}\left(\sqrt{-\operatorname{Det} g}\,g^{\mu\nu}\gamma_{nm}(\varphi)\frac{\partial\varphi^n}{\partial x^\nu}\right) &= \sqrt{-\operatorname{Det} g}\\ \times\left(\frac{1}{2}g^{\mu\nu}\frac{\partial\gamma_{lm}(\varphi)}{\partial\varphi^n}\frac{\partial\varphi^l}{\partial x^\mu}\frac{\partial\varphi^m}{\partial x^\nu} + \frac{\partial V(\varphi)}{\partial\varphi^n}\right).\end{aligned} \tag{10.4.3}$$

We take each scalar field $\varphi^n(x)$ as an unperturbed term $\bar\varphi^n(t)$ that depends only on time, plus a small perturbation $\delta\varphi^n(\mathbf{x},t)$:

$$\varphi^n(\mathbf{x},t) = \bar\varphi^n(t) + \delta\varphi^n(\mathbf{x},t). \tag{10.4.4}$$

Similarly, as in Chapters 5-8, the metric is assumed to be given by the unperturbed Robertson-Walker metric $\bar g_{\mu\nu}(t)$ (with $K = 0$) plus a small perturbation $h_{\mu\nu}(\mathbf{x},t)$

$$g_{\mu\nu}(\mathbf{x},t) = \bar g_{\mu\nu}(t) + h_{\mu\nu}(\mathbf{x},t). \tag{10.4.5}$$

The unperturbed energy-momentum tensor is of the perfect fluid form (5.1.35), with unperturbed energy density, pressure, and velocity

$$\bar\rho = \frac{1}{2}\gamma_{nm}(\bar\varphi)\dot{\bar\varphi}^n\dot{\bar\varphi}^m + V(\bar\varphi) \tag{10.4.6}$$

$$\bar p = \frac{1}{2}\gamma_{nm}(\bar\varphi)\dot{\bar\varphi}^n\dot{\bar\varphi}^m - V(\bar\varphi) \tag{10.4.7}$$

$$\bar u^0 = 1, \quad \bar u^i = 0 \tag{10.4.8}$$

The scalar field equation (10.4.3) for the unperturbed fields is

$$\ddot{\bar\varphi}^n + \gamma^n_{ml}(\bar\varphi)\dot{\bar\varphi}^m\dot{\bar\varphi}^l + 3H\dot{\bar\varphi}^n + \gamma^{nm}(\bar\varphi)\frac{\partial V(\bar\varphi)}{\partial\bar\varphi^m} = 0 \tag{10.4.9}$$

where γ^{nm} is the reciprocal of the matrix γ_{nm}, γ^n_{ml} is the affine connection in field space:

$$\gamma^n_{ml}(\bar\varphi) = \frac{1}{2}\gamma^{nk}(\bar\varphi)\left(\frac{\partial\gamma_{km}(\bar\varphi)}{\partial\bar\varphi^l} + \frac{\partial\gamma_{kl}(\bar\varphi)}{\partial\bar\varphi^m} - \frac{\partial\gamma_{ml}(\bar\varphi)}{\partial\bar\varphi^k}\right) \tag{10.4.10}$$

and H is the expansion rate $H \equiv \dot{a}/a = \sqrt{8\pi G\bar{\rho}/3}$. The reader may check that Eq. (10.4.9) guarantees that the energy-conservation equation $\dot{\bar{\rho}} = -3H(\bar{\rho}+\bar{p})$ is satisfied by the quantities (10.4.6) and (10.4.7). From Eqs. (10.4.6) and (10.4.9) we find the convenient formula

$$\dot{H} = -4\pi G\gamma_{nm}(\bar{\varphi})\dot{\bar{\varphi}}^n\dot{\bar{\varphi}}^m \tag{10.4.11}$$

For more than one scalar field $T_{\mu\nu}$ is not of the perfect fluid form (5.1.35) to all orders in perturbations, but by comparing the first-order terms in Eq. (10.4.2) with Eqs. (5.1.39)-(5.1.41), we see that to first order the anisotropic inertia vanishes, and the perturbations to the energy density, pressure, and velocity potential are

$$\begin{aligned}\delta\rho =&\gamma_{nm}(\bar{\varphi})\dot{\bar{\varphi}}^n\delta\dot{\varphi}^m + \frac{1}{2}\dot{\bar{\varphi}}^n\dot{\bar{\varphi}}^m\frac{\partial\gamma_{nm}(\bar{\varphi})}{\partial\bar{\varphi}^k}\delta\varphi^k\\ &+\frac{\partial V(\bar{\varphi})}{\partial\bar{\varphi}^n}\delta\varphi^n + \frac{1}{2}h_{00}\gamma_{nm}(\bar{\varphi})\dot{\bar{\varphi}}^n\dot{\bar{\varphi}}^m \qquad (10.4.12)\\ \delta p =&\gamma_{nm}(\bar{\varphi})\dot{\bar{\varphi}}^n\delta\dot{\varphi}^m + \frac{1}{2}\dot{\bar{\varphi}}^n\dot{\bar{\varphi}}^m\frac{\partial\gamma_{nm}(\bar{\varphi})}{\partial\bar{\varphi}^k}\delta\varphi^k\\ &-\frac{\partial V(\bar{\varphi})}{\partial\bar{\varphi}^n}\delta\varphi^n + \frac{1}{2}h_{00}\gamma_{nm}(\bar{\varphi})\dot{\bar{\varphi}}^n\dot{\bar{\varphi}}^m \qquad (10.4.13)\\ \delta u =&-\frac{\gamma_{nm}(\bar{\varphi})\dot{\bar{\varphi}}^n\delta\varphi^m}{\gamma_{kl}(\bar{\varphi})\dot{\bar{\varphi}}^k\dot{\bar{\varphi}}^l} \qquad (10.4.14)\end{aligned}$$

The reader can easily check that these formulas reduce to Eqs. (10.1.9)(10.1.11) in the single-field case, with $\gamma_{11} = 1$.

Since there is no first-order anisotropic inertia, in Newtonian gauge we have $\Phi = \Psi$, so $h_{00} = -2\Psi$, and the Einstein field equation Eq. (5.3.21) takes the form

$$\dot{\Psi} + H\Psi = 4\pi G\gamma_{nm}(\bar{\varphi})\dot{\bar{\varphi}}^n\delta\varphi^m, \tag{10.4.15}$$

The terms of first order in the field equation Eq. (10.4.3) are much simplified if we now adopt a notation that reveals the transformation of quantities under redefinitions $\varphi^n \to \varphi'^n(\varphi)$ of the scalar fields. Under such transformations, quantities like $\dot{\bar{\varphi}}^n$ and $\delta\varphi^n$ transform as contravariant vectors, in the sense that

$$\delta\varphi'^n = \frac{\partial\bar{\varphi}'^n}{\partial\bar{\varphi}^m}\delta\varphi^m, \quad \dot{\bar{\varphi}}'^n = \frac{\partial\bar{\varphi}'^n}{\partial\bar{\varphi}^m}\dot{\bar{\varphi}}^m \tag{10.4.16}$$

For any vector v^n that transforms in this way, we can define a rate of change that is also a vector:

$$\frac{D}{Dt}v^n \equiv \frac{\partial}{\partial t}v^n + \gamma^n_{lm}(\bar{\varphi})\dot{\bar{v}}^l v^m. \tag{10.4.17}$$

With this notation, the first-order terms in the field equation (10.4.3) give

$$\begin{aligned}&\frac{D^2}{Dt^2}\delta\varphi^n + 3H\frac{D}{Dt}\delta\varphi^n + \gamma^{nm}(\bar{\varphi})\frac{\partial^2 V(\bar{\varphi})}{\partial\bar{\varphi}^m\partial\bar{\varphi}^l}\delta\varphi^l - \left(\frac{\nabla^2}{a^2}\right)\delta\varphi^n\\ &\quad = -2\gamma^{nm}(\bar{\varphi})\Psi\frac{\partial V(\bar{\varphi})}{\partial\bar{\varphi}^m} + 4\dot{\Psi}\dot{-}^n + \gamma^n{}_{lmk}(\bar{\varphi})\dot{\bar{\varphi}}^l\dot{\bar{\varphi}}^m\delta\varphi^k \qquad (10.4.18)\end{aligned}$$

where $\gamma^n{}_{lmk}(\bar{\varphi})$ is the Riemann-Christoffel tensor in field space:

$$\gamma^n_{lmk}(\bar\varphi) \equiv \frac{\partial\gamma^n_{ml}(\bar\varphi)}{\partial\bar\varphi^k} - \frac{\partial\gamma^n_{mk}(\bar\varphi)}{\partial\bar\varphi^l} + \gamma^r_{lm}(\bar\varphi)\gamma^n_{kr}(\bar\varphi) - \gamma^r_{lk}(\bar\varphi)\gamma^n_{mr}(\bar\varphi). \tag{10.4.19}$$

(The scalar fields can be redefined to make $\gamma_{nm} = \delta_{nm}$ if and only if $\gamma^n{}_{lmk} = 0$. We are not assuming that this is the case.) Also, the constraint (5.3.26) is here

$$\left(\dot H - \frac{\nabla^2}{a^2}\right)\Psi = 4\pi G\gamma_{nm}(\bar\varphi)\left(-\dot{\bar\varphi}^n\frac{D}{Dt}\delta\varphi^m + \delta\varphi^m\frac{D}{Dt}\dot{\bar\varphi}^n\right). \tag{10.4.20}$$

The solutions are written as superpositions of plane waves

$$\delta\varphi^n(\mathbf{x},t) = \sum_N\int d^3q\left[\delta\varphi^n_{Nq}(t)e^{i\mathbf{q}\cdot\mathbf{x}}\alpha(\mathbf{q},N) + \delta\varphi^{n*}_{Nq}(t)e^{-i\mathbf{q}\cdot\mathbf{x}}\alpha^*(\mathbf{q},N)\right] \tag{10.4.21}$$

$$\Psi(\mathbf{x},t) = \sum_N\int d^3q\left[\Psi_{Nq}(t)e^{i\mathbf{q}\cdot\mathbf{x}}\alpha(\mathbf{q},N) + \Psi^*_{Nq}(t)e^{-i\mathbf{q}\cdot\mathbf{x}}\alpha^*(\mathbf{q},N)\right] \tag{10.4.22}$$

Here N labels different solutions of the coupled equations (10.4.15)(10.4.17), with ∇^2 replaced with $-q^2$:

$$\dot\Psi_{Nq} + H\Psi_{Nq} = 4\pi G\gamma_{nm}(\bar\varphi)\dot{\bar\varphi}^n\delta\varphi^m_{Nq} \tag{10.4.23}$$

$$\frac{D^2}{Dt^2}\delta\varphi^n_q + 3H\frac{D}{Dt}\delta\varphi^n_q + \gamma^{nm}(\bar\varphi)\frac{\partial^2V(\bar\varphi)}{\partial\bar\varphi^m\partial\bar\varphi^l}\delta\varphi^l_q + \left(\frac{q^2}{a^2}\right)\delta\varphi^n_q$$
$$= -2\gamma^{nm}(\bar\varphi)\Psi_q\frac{\partial V(\bar\varphi)}{\partial\bar\varphi^m} + 4\dot\Psi_q\dot{\bar\varphi}^n + \gamma^n{}_{lmk}(\bar\varphi)\dot{\bar\varphi}^l\dot{\bar\varphi}^m\delta\varphi^k_q \tag{10.4.24}$$

$$\left(\dot H + \frac{q^2}{a^2}\right)\Psi_q = 4\pi G\gamma_{nm}(\bar\varphi)\left(-\dot{\bar\varphi}^n\frac{D}{Dt}\delta\varphi^m_q + \delta\varphi^m_q\frac{D}{Dt}\dot{\bar\varphi}^n\right) \tag{10.4.25}$$

There is one second-order equation for each scalar field, and one firstorder equation for Ψ, so with one constraint on first derivatives the number of independent solutions equals twice the number of scalar fields. Since $\left(\delta\varphi^n_{Nq}, \Psi_{Nq}\right)$ and $\left(\delta\varphi^{n*}_{Nq}, \Psi^*_{Nq}\right)$ are all independent solutions, the index N takes as many values as the index n.

To find the initial conditions for Eqs. (10.4.23)-(10.4.25), and to find the commutation relations for the operators $\alpha(\mathbf{q})$ and $\alpha^*(\mathbf{q})$ in Eqs. (10.4.21) and (19.4.22), we note that for some time before the era of horizon exit we will have $q/a \gg H$, and q^2/a^2 much greater than any element of $\partial^2V/\partial\bar\varphi^n\partial\bar\varphi^m$. Hence the solutions up to the beginning of the era of horizon exit take the WKB form

$$\delta\varphi^n_{Nq}(t) \to f^n_{Nq}(t)\exp\left(-iq\int_{t_1}^t\frac{dt'}{a\left(t'\right)}\right)$$
$$\Psi_{Nq}(t) \to g_{Nq}(t)\exp\left(-iq\int_{t_1}^t\frac{dt'}{a\left(t'\right)}\right) \tag{10.4.26}$$

where $f^n_{Nq}(t)$ and $g_{Nq}(t)$ vary much more slowly than the argument of the exponential, and t_1 is arbitrary. Eqs. (10.4.23) and (10.4.25) are both satisfied to leading order in q/a if we take

$$g_{Nq} = \frac{4i\pi Ga}{q}\gamma_{nm}(\bar{\varphi})\dot{\bar{\varphi}}^n f^m_{Nq} \tag{10.4.27}$$

The terms in Eq. (10.4.24) of first order in q/a then give

$$\frac{D}{Dt}f^n_{Nq} + Hf^n_{Nq} = 0 \tag{10.4.28}$$

To solve this, we note that, because $\gamma^{nm}(\bar{\varphi}(t))$ is positive-definite, it can be written in terms of a set of vielbein vectors $e^n_N(t)$ (with N running over as many values as n), as

$$\gamma^{nm}(\bar{\varphi}(t)) = \sum_N e^n_N(t)e^m_N(t). \tag{10.4.29}$$

These vielbeins can be defined to satisfy the equation of parallel transport [1]

$$\frac{D}{Dt}e^n_{Nq} = 0, \tag{10.4.30}$$

so the solution of Eq. (10.4.28) is

$$f^n_{Nq}(t) \propto a^{-1}(t)e^n_N(t) \tag{10.4.31}$$

For reasons that will soon become apparent, we shall normalize these solutions so that, for $q/a \gg H$,

$$f^n_{Nq} = (2\pi)^{-3/2}(2q)^{-1/2}a^{-1}e^n_N.$$

With this normalization, at the beginning of the era of horizon exit we have [2]

$$\delta\varphi^n_{Nq}(t) = \frac{1}{(2\pi)^{3/2}a(t)\sqrt{2q}}e^n_N(t)\exp\left(-iq\int_{t_1}^t\frac{dt'}{a\,(t')}\right) \tag{10.4.32}$$

$$\Psi_{Nq}(t) = \frac{4i\pi G\gamma_{nm}(\bar{\varphi})e^m_N(t)\dot{\bar{\varphi}}^n(t)}{(2\pi)^{3/2}\sqrt{2q^3}}\exp\left(-iq\int_{t_1}^t\frac{dt'}{a\,(t')}\right) \tag{10.4.33}$$

For times early enough so that $q/a \gg H$ the commutation relations for φ^n can be obtained from the action (10.4.1) with Ψ neglected, and therefore take the form:

$$[\varphi^n(\mathbf{x},t),\varphi^m(\mathbf{y},t)] = 0$$
$$[\varphi^n(\mathbf{x},t),\dot{\varphi}^m(\mathbf{y},t)] = ia^{-3}(t)\gamma^{nm}(\bar{\varphi}(t))\delta^3(\mathbf{x}-\mathbf{y})$$

[471] We define the vectors $e^n_{Nq}(t)$ to satisfy the first-order differential equation (10.4.30), and, as an initial condition, to satisfy Eq. (10.4.29) at some initial time $t = t_1$. From Eq. (10.4.30) and the definition (10.4.10) it follows that for all times

$$\dot{D}^n_m = \left[-\gamma^n_{lk}D^k_m + \gamma^k_{lm}D^n_k\right]\dot{\bar{\varphi}}^l,$$

where

$$D^n_m \equiv \sum_N e^n_N e^k_N\gamma_{km}$$

This differential equation for D^n_m has a solution $D^n_m = \delta^n_m$, and our initial condition tells us that $D^n_m = \delta^n_m$ at $t = t_1$, so this is the solution for all times. It follows that the vectors $e^n_{Nq}(t)$ defined in this way satisfy the condition (10.4.29) for vielbeins for all times.

[2] A result equivalent to Eq. (10.4.32) is given in Eq. (4.4) of H.-C. Lee, M. Sasaki, E. D. Stewart, T. Tanaka, and S. Yokoyama, J. Cosm. & Astropart. Phys. 0510, 004 (2005) [astro-ph/0506262], using a " δN " formalism due to M. Sasaki and E. D. Stewart, Prog. Theor. Phys. 95, 71 (1996) [astroph/9507001]. But their paper does not reach the conclusion (10.4.41) found in this section.

With $\delta\varphi^n$ normalized as in Eq. (10.4.32), this implies that the timeindependent operator coefficients in Eqs. (10.4.21) and (10.4.22) satisfy the commutation relations

$$[\alpha(\mathbf{q},N),\alpha\left(\mathbf{q}',N'\right)]=0,\quad [\alpha(\mathbf{q},N),\alpha^*\left(\mathbf{q}',N'\right)]=\delta^3\left(\mathbf{q}-\mathbf{q}'\right)\delta_{NN'}. \tag{10.4.34}$$

Assuming that there is enough time before horizon exit for the state of the world to decay into the Bunch-Davies vacuum $|0\rangle$, with $\alpha(\mathbf{q},N)|0\rangle=0$, it follows then from Eqs. (10.4.21) and (10.4.22) that the observed perturbations will be Gaussian, just as in the single-field case.

According to assumption 3, during the era when q/a drops from being somewhat larger to somewhat smaller than H - say, from $10H$ to $0.1H$, the scalar fields are rolling slowly down the potential hill. We assume that $V(\varphi)$ satisfies whatever flatness conditions are necessary to allow us to drop all terms in Eqs. (10.4.23)-(10.4.25) proportional to $\dot{\bar{\varphi}}^n_N$ or $\ddot{\bar{\varphi}}^n_N$ and to ignore the second derivative of the potential in Eq. (10.4.24). Then during the era of horizon exit Eq. (10.4.24) is approximately

$$\delta\ddot{\varphi}^n_{Nq}+3H\delta\dot{\varphi}^n_{Nq}+\left(\frac{q^2}{a^2}\right)\delta\varphi^n_{Nq}=0. \tag{10.4.35}$$

H is roughly constant during this era, [3] so the independent solutions of Eq. (10.4.35) are proportional to $(1+iq\tau)\exp(-iq\tau)$ and its complex conjugate, where τ is again the conformal time

$$\tau=\int_\infty^t\frac{dt'}{a\left(t'\right)}\simeq-\frac{1}{\mathrm{Ha}(t)}$$

The scalar field perturbations at the beginning of the era of horizon exit are given by Eq. (10.4.32), so during this era we have

$$\delta\varphi^n_{Nq}\simeq\frac{1}{(2\pi)^{3/2}\sqrt{2q}}\left(\frac{1}{a}+\frac{iH}{q}\right)e^{+iq/aH}e^n_N. \tag{10.4.36}$$

By the end of the era of horizon exit we have $q/a\ll H$, and the scalar field perturbations approach the quantities

$$\delta\varphi^n_{Nq}\to\frac{iH}{(2\pi)^{3/2}\sqrt{2q^3}}e^n_N. \tag{10.4.37}$$

Since H and e^n_N are slowly varying, they can be evaluated at the time of horizon exit, and depend weakly on q.

At some time after the era of horizon exit the slow-roll conditions must become violated, if only in order that the energy in the inflaton fields can eventually be converted into ordinary matter and radiation. The potential term on the left-hand side of Eq. (10.4.24) is then no longer negligible, and things get complicated. But once q/a becomes much less than H, the subsequent evolution of the scalar fields during inflation cannot depend on q, so until horizon re-entry all scalar field perturbations $\delta\varphi^n_{Nq}$ will have the same wave number dependence, close to $q^{-3/2}$, as given by Eq. (10.4.37) at the end of the era of horizon exit. The same applies to the scalar metric perturbations; when a becomes sufficiently large so that $q^2/a^2\ll|\dot{H}|$, Eq. (10.4.25) gives Ψ_{Nq} approximately proportional to $q^{-3/2}$.

In general, it is not easy to see what these results imply for the perturbations observed in the cosmic microwave background or large scale structure. However, there is one case where an important conclusion can be reached. If according to assumption 1* the energy in all the scalar fields is converted at the end of inflation into ordinary matter and radiation in local thermal

[473] The fractional change in H during the era of horizon exit is $|\dot{H}/H|\times\ln(100)/H$, which is small if $|\dot{H}|/H^2\ll 1/\ln(100)=0.22$.

equilibrium, and if at this time all conserved quantities like electric charge have zero density, then as remarked in Section 5.4, the perturbations become adiabatic, with $\mathcal{R}_q$ taking a constant value $\mathcal{R}_q^o$ until horizon re-entry. For small fluctuations, $\mathcal{R}_q^o$ will be some linear combination of the perturbations $\delta\varphi_{Nq}^n$ and Ψ_{Nq} at the end of the era of horizon exit. We do not know the coefficients in this linear combination, which in general will depend on the shape of the potential experienced as the field evolves until the end of inflation, as well as on the mechanism of energy transfer to matter and radiation. But we can be sure that these coefficients are independent of wave number, because once the era of horizon exit is over the perturbations are far outside the horizon. Hence we can conclude that in this case $\mathcal{R}_q^o$ will have the same wave number dependence as $\delta\varphi_{Nq}^n$ and Ψ_{Nq} at the end of the era of horizon exit, which for slow-roll inflation will be close to $q^{-3/2}$.

The ubiquity of the $q^{-3/2}$ wavelength dependence can be understood on very general grounds. For negligible spatial curvature, nothing should depend on how the co-moving coordinate vector $\mathbf{x}$ is normalized, so suppose we change its scale by a transformation $\mathbf{x} \to \lambda\mathbf{x}$, with λ constant. To keep $\mathbf{q}\cdot\mathbf{x}$ unchanged, we must then change the scale of co-moving wave numbers by a transformation $\mathbf{q} \to \lambda^{-1}\mathbf{q}$. Consider any perturbation $Z(\mathbf{x},t)$ that, like $\delta\varphi^n(\mathbf{x},t)$ and $\Psi(\mathbf{x},t)$, is given by a Fourier integral

$$Z(\mathbf{x},t) = \sum_N \int d^3q \left[e^{i\mathbf{q}\cdot\mathbf{x}} a(\mathbf{q},N) z_{Nq}(t) + e^{-i\mathbf{q}\cdot\mathbf{x}} a^*(\mathbf{q},N) z_{Nq}^*(t)\right]$$

where N labels the various solutions of the field equations, and $a(\mathbf{q},N)$ and $a^*(\mathbf{q},N)$ are annihilation and creation operators satisfying the commutation relations

$$[a(\mathbf{q},N), a^*(\mathbf{q}',N')] = \delta_{NN'}\delta^3(\mathbf{q}-\mathbf{q}')$$

Under the transformation $\mathbf{q} \to \lambda^{-1}\mathbf{q}$, the delta function in this commutation relation transforms as $\delta^3(\mathbf{q}-\mathbf{q}') \to \lambda^3\delta^3(\mathbf{q}-\mathbf{q}')$, so we must have $a(\mathbf{q},N) \to \lambda^{3/2}a(\mathbf{q},N)$. Also, of course $d^3q \to \lambda^{-3}d^3q$. Hence in order for $Z(\mathbf{x},t)$ to be unaffected by this change of scale, we must have $z_{Nq}(t) \to \lambda^{3/2}z_{Nq}(t)$. This condition is satisfied if $z_{Nq}(t)$ has a q-dependence $\propto q^{-3/2}$, and if it does not depend on the scale of $a(t)$. Of course, to keep the physical coordinate vector $\mathbf{x}a(t)$ independent of the normalization chosen for the co-moving coordinates, $a(t)$ has the scale transformation $a(t) \to \lambda^{-1}a(t)$, so $q/a(t)$ is scale-invariant, and if $z_{Nq}(t)$ depends on the scale of the function $a(t)$ then we cannot conclude that it is proportional to $q^{-3/2}$. But if $z_{Nq}(t)$, like $\delta\varphi_{Nq}^n(t)$ and $\Psi_{Nq}(t)$, takes a nearly time-independent value z_{Nq}^o after horizon exit, then outside the horizon it will not depend strongly on $a(t)$. It could still depend on $H(t_q), \dot{H}(t_q)$, etc., where t_q is the time of horizon exit, defined by the scale-invariant condition $q/a(t_q) = H(t_q)$, but in the limit of very slow roll inflation $H, \dot{H}$, etc. depend only weakly on time, so z_{Nq}^o can depend only weakly on the scale of the function a, and the scale-invariance of $Z(\mathbf{x},t)$ requires that z_{Nq}^o be nearly proportional to $q^{-3/2}$.

The intensity of observed adiabatic fluctuations is related to the quantity $\mathcal{R}$ defined by Eq. (5.4.1). Using Eqs. (10.4.14), (10.4.11), (10.4.21), and (10.4.22), during inflation this is

$$\begin{aligned}
\mathcal{R}(\mathbf{x},t) &\equiv -\Psi(\mathbf{x},t) + H(t)\delta u(\mathbf{x},t) \\
&= -\Psi(\mathbf{x},t) + \frac{4\pi G H(t)}{\dot{H}(t)} \gamma_{nm}(\bar{\varphi}(t))\dot{\bar{\varphi}}^n(t)\delta\varphi^m(\mathbf{x},t) \\
&= \sum_N \int d^3q \left[e^{i\mathbf{q}\cdot\mathbf{x}} a(\mathbf{q},N)\mathcal{R}_{Nq}(t) + e^{-i\mathbf{q}\cdot\mathbf{x}} a^*(\mathbf{q},N)\mathcal{R}_{Nq}^*(t)\right]
\end{aligned} \tag{10.4.38}$$

where

$$\mathcal{R}_{Nq}(t) = -\Psi_{Nq}(t) + \frac{4\pi G H(t)}{\dot{H}(t)} \gamma_{nm}(\bar{\varphi}(t)) \dot{\bar{\varphi}}^n(t) \delta\varphi^m_{Nq}(t) \tag{10.4.39}$$

Because of the factor $\dot{H}$ in the denominator, at the end of the era of horizon exit the second term in Eq. (10.4.39) dominates over the term $-\Psi_{Nq}$, and Eq. (10.4.37) then gives at this time

$$\mathcal{R}_{Nq} \to i \frac{4\pi G H^2}{(2\pi)^{3/2}\sqrt{2q^3}\dot{H}} \gamma_{nm}(\bar{\varphi}) \dot{\bar{\varphi}}^n e^m_N. \tag{10.4.40}$$

The definition (10.4.29) of the vielbeins and the formula (10.4.11) then lead immediately to a sum rule

$$\left(\sum_N |\mathcal{R}_{Nq}|^2 \right)^{1/2} = \frac{\sqrt{G} H^2}{2\pi q^{3/2} \sqrt{|\dot{H}|}}. \tag{10.4.41}$$

That is, the root-mean-square value of the quantities $|\mathcal{R}_{Nq}|$ at the end of the era of horizon exit is the same as the value of $|\mathcal{R}_q|$ outside the horizon in the single-field case, given by Eq. (10.3.16). But Eqs. (10.4.38) and (10.4.34) tell us that the correlation function of $\mathcal{R}$ in the Bunch-Davies vacuum is

$$\int d^4x e^{-i\mathbf{q}\cdot(\mathbf{x}-\mathbf{y})} \langle \mathcal{R}(\mathbf{x},t) \mathcal{R}(\mathbf{y},t) \rangle = (2\pi)^3 \sum_N |\mathcal{R}_{Nq}|^2 \tag{10.4.42}$$

So from Eq. (10.4.41) it follows that the correlation function $\langle \mathcal{R}(\mathbf{x},t) \mathcal{R}(\mathbf{y},t) \rangle$ is the same at the end of the era of horizon exit as in the single field case. With more than one scalar field the correlation function (10.4.42) is not in general time-independent outside the horizon, but it is plausible that the value of $|\mathcal{R}_q|$ during a period of thermal equilibrium after inflation will not be orders of magnitude different from Eq. (10.4.41) at the end of the era of horizon exit. Thus the observed strength and spectral shape of anisotropies in the cosmic microwave background suggests a value of H at horizon exit of order 10^{14}GeV, as in the single field case.

One can derive stronger results in the case where all but one of the eigenvalues of the matrix $\gamma^{nm}(\bar{\varphi}) \partial^2 V(\bar{\varphi}) / \partial\bar{\varphi}^m \partial\bar{\varphi}^l$ in Eq. (10.4.23) are large and positive. In this case, the unperturbed scalar fields roll along the direction of the eigenvector for the small eigenvalue, and the only significant perturbations lie in that direction. The problem then reduces to the single field case; we have an essentially adiabatic perturbation, with the amplitude $\mathcal{R}^o_q$ given in the slow-roll approximation by Eq. (10.3.16), and the slope $n_S(q)$ given in this approximation by Eq. (10.3.25), where ϵ and δ are to be calculated in terms of the time-derivatives of the expansion rate by Eq. (10.3.4). But there is no known reason why the potential should have the properties needed to justify these results. On the other hand, tensor perturbations during inflation are governed by Eq. (10.3.17) however many scalar fields there are, and the only role played by the scalar fields is to contribute to the

Hubble rate H. Thus for slow-roll inflation with any number of scalar fields, the tensor amplitude $\mathcal{D}^o_q$ is given by Eq. (10.3.23), and the slope parameter $n_T(q)$ is given by Eq. (10.3.25), just as in the single field case.

The fact that observations of cosmic microwave background anisotropies and large scale structure indicate that scalar fluctuations outside the horizon are adiabatic and Gaussian, with $\mathcal{R}^o_q$ approximately proportional to $q^{-3/2}$, and with $q^{3/2}\mathcal{R}^o_q \ll 1$, evidently is consistent with a very large class of models of inflation. This is encouraging, because it supports the general idea of slow-roll inflation, but also disappointing, because it shows that these observations so far do not really tell us anything specific about the details of inflation. With further improvements in experimental precision, we can look forward to a more decisive test of theories of the early universe.

10.11 Appendix A Some Useful Numbers

Numerical Constants

$$\pi = 3.1415927 \quad 1'' = 4.84814 \times 10^{-6} \text{ radians}$$
$$e = 2.7182818 \quad \ln 10 = 2.3025851$$
$$\gamma = 0.5772157 \quad \zeta(3) = 1.2020569$$

Physical Constants [1]

Speed of light in vacuum	$c \equiv 2.99792458 \times 10^{10}$ cmsec^{-1}
Planck constant	$h = 6.6260693(11) \times 10^{-27}$ erg-sec
Reduced Planck constant	$\hbar \equiv h/2\pi = 1.05457168(18) \times 10^{-27}$ergsec
Electronic charge (unrat.)	$e = 4.80320441(41) \times 10^{-10}$ esu
Electron volt	$1\text{eV} = 1.60217653(14) \times 10^{-12}$erg
	$\hbar c = 197.326968(17) \times 10^{-13}$MeVcm
Fine structure constant	$\alpha \equiv e^2/\hbar c = 1/137.03599911\ (46)$
Electron mass	$m_e = 9.1093826(16) \times 10^{-28}$ g
	$m_e c^2 = 0.510998918(44)$MeV
Rydberg energy	$hcR \equiv m_e e^4/2\hbar^2 = 13.6056923(12)$eV
Thomson cross section	$\sigma_{\mathcal{T}} = 8\pi e^4/3m_e^2c^4 = 0.665245873(13)\times 10^{-24}$ cm^2
Proton mass	$m_p = 1.67262171(29) \times 10^{-24}$ g $m_p c^2 = 938.272029(80)$MeV
Neutron mass	$m_n c^2 = 939.565360(81)$MeV
Deuteron mass	$m_d c^2 = 1875.61282(16)$MeV
Atomic mass unit	$m\left(\text{C}^{12}\right)/12 = 1.66053886(28) \times 10^{-24}$ g $m\left(\text{C}^{12}\right)c^2/12 = 931.494043(80)$MeV
Avogadro's number	$N_A = 6.0221415(10) \times 10^{23}$/mole
Boltzmann constant	$k_{\mathcal{B}} = 1.3806505(24) \times 10^{-16}$erg/K $= 8.617343(15) \times 10^{-5}$eV/K
Radiation energy constant	$a_{\mathcal{B}} = \dfrac{8\pi^5 k_{\mathcal{B}}^4}{15h^3c^3} = 7.56577(5) \times 10^{-15}$erg cm^{-3} K^{-4}

Weak coupling constant $G_{wk} = 1.16637(1) \times 10^{-5}\text{GeV}^{-2}$
Gravitational constant $G = 6.6742(10) \times 10^{-8}$ dyn cm^2 g^{-2}
Planck energy $\sqrt{\hbar c/G} = 1.22090(9) \times 10^{19}$GeV

Astronomical Constants [2]

Julian year 1 year $\equiv 365.25$ days $= 3.1557600 \times 10^7$sec
Light year 1 light (Julian) year $= 9.460730472 \times 10^{17}$ cm
Mean earth-sun distance 1 A.U. $= 1.4959787066 \times 10^{13}$ cm
Parsec 1pc $\equiv 648000/\pi$ A.U. $= 3.0856776 \times 10^{18}$ cm $= 3.2615638$ light (Julian) year
Solar mass $M_\odot = 1.9891 \times 10^{33}$ g

[471] From Review of Particle Physics, S. Eidelman et al. (Particle Data Group), Phys. Lett. B 592, 1 (2004).

Solar luminosity $L_\odot = 3.845(8) \times 10^{33}\text{ergsec}^{-1}$
Apparent luminosity for apparent magnitude m

$$\ell = 2.52 \times 10^{-5}\text{erg cm}^{-2}\text{sec}^{-1} \times 10^{-2m/5}$$

Absolute luminosity for absolute magnitude M

$$\mathcal{L} = 3.02 \times 10^{35}\text{ergsec}^{-1} \times 10^{-2M/5}$$

For a Hubble constant $H_0 = h \times 100$ kmsec^{-1}Mpc^{-1} :
Hubble time

$$H_0^{-1} = 3.0857h^{-1} \times 10^{17}\text{sec} = 9.778h^{-1} \times 10^9 years$$

Hubble distance
c / H_{0}=2997.92458 h^{-1} \mathrm{Mpc}
$$
Critical density

$$\rho_{\text{crit}} \equiv \frac{3H_0^2}{8\pi G} = 1.878h^2 \times 10^{-29}\text{ g cm}^{-3}$$
$$= [0.00300\text{eV}]^4 h^2$$

10.12 Appendix B Review of General Relativity

In this appendix we offer a brief introduction to the General Theory of Relativity, Einstein's theory of gravitation. This appendix is not a substitute for a thorough treatment of the theory, but it outlines the parts of the theory that are used in this book, and it serves to establish our notation.

1 The Equivalence Principle

General Relativity is based on the Principle of the Equivalence of Gravitation and Inertia, or the Equivalence Principle for short. The Equivalence Principle is a generalization of the familiar observation that, because of the equality of gravitational and inertial mass, freely falling observers do not feel the effects of gravitation. According to the Equivalence Principle, at any spacetime point in an arbitrary gravitational field there is a "locally inertial" coordinate system in which the effects of gravitation are absent in a sufficiently small spacetime neighborhood of that point. This Principle allows us to write the equations governing any sufficiently small physical system in a gravitational field if we know the equations governing it in the absence of gravitation: it is only necessary to write the equations in a form which is generally covariant - that is, whose form is independent of the spacetime coordinates used - and which reduce to the correct equations in the absence of gravitation. Such equations will be true in the presence of a gravitational field, because general covariance guarantees that they are true in any set of coordinates if they are true in any other set of coordinates, and the Equivalence Principle tells us that there is a set of coordinates in which the equations are true - the set of coordinates that is locally inertial at the spacetime location of the system in question. In general there will be more than one set of generally covariant equations that reduce to the correct equations in the absence of gravitation, but the differences between these equations always involve terms with extra spacetime derivatives, which become negligible if we restrict ourselves to a spacetime

[472] From Allen's Astrophysical Quantities, ed. A. N. Cox (AIP Press, New York, 2000).

region that is small compared with the scale of distances and times over which the gravitational and other fields vary appreciably.

2 The metric: Ticking clocks

As an example of this procedure, consider the equation that governs the rate at which clocks tick in a gravitational field. Special Relativity tells us that if
a clock ticks once in every time interval dT when at rest in the absence of a gravitational field, then the separation $d\xi^\alpha$ between the spacetime locations of successive ticks when the clock is moving in the absence of a gravitational field is governed by the relation

$$\eta_{\alpha\beta} d\xi^\alpha d\xi^\beta = -dT^2. \tag{B.1}$$

(Here ξ^1, ξ^2, and ξ^3 are the Cartesian space coordinates, using units of length in which the speed of light c is unity; $\xi^0 \equiv t; \eta_{\alpha\beta}$ is the Minkowski metric, the diagonal matrix with $\eta_{11} = \eta_{22} = \eta_{33} = 1$ and $\eta_{00} = -1$; and repeated indices are summed.) The correct equation governing the ticking of the clock in a general gravitational field is then

$$g_{\mu\nu}(x) dx^\mu dx^\nu = -dT^2, \tag{B.2}$$

where $g_{\mu\nu}(x)$ is the metric, a field defined by the two properties that, first, a transformation to a coordinate system x'^μ changes the metric to

$$g'_{\rho\sigma}(x') = g_{\mu\nu}(x) \frac{\partial x^\mu}{\partial x'^\rho} \frac{\partial x^v}{\partial x'^\sigma}, \tag{B.3}$$

and, second, in coordinates that are locally inertial and Cartesian at a point x, the metric at x is $\eta_{\alpha\beta}$ and its first derivatives at x vanish. (In Eq. (B.3), x^μ and x'^μ are the coordinates of the same physical point in two different coordinate systems.) Eq. (B.2) is generally covariant, because the coordinate differentials have the obvious transformation property

$$dx'^\rho = \frac{\partial x'^\rho}{\partial x^\mu} dx^\mu \tag{B.4}$$

so that
$g'_{\rho\sigma}(x')\, dx'^\rho dx'^\sigma = g_{\mu\nu}(x) \frac{\partial x^\mu}{\partial x'^\rho} \frac{\partial x^\nu}{\partial x'^\sigma} \frac{\partial x'^\rho}{\partial x^\kappa} dx^\kappa \frac{\partial x'^\sigma}{\partial x^\lambda} dx^\lambda = g_{\mu\nu}(x) dx^\mu dx^\nu$.
To repeat our general argument, Eq. (B.2) is true, because in locally inertial Cartesian coordinates it reduces to the equation (B.1) that describes clocks in the absence of gravitation, and its general covariance means that if it is true in one set of coordinates then it is true in any other set of coordinates. In the same way, the spacetime separation dx^μ of the ends of a small ruler whose length is dL when measured at rest in the absence of gravitation will in general be given by

$$g_{\mu\nu}(x) dx^\mu dx^\nu = +dL^2.$$

Likewise, the differences dx^μ between the spacetime coordinates of two successive positions along a ray of light are governed by the equation

$$g_{\mu\nu}(x) dx^\mu dx^\nu = 0.$$

3 Tensors, vectors, scalars

Quantities that transform as in (B.3) and (B.4) are known as covariant tensors and contravariant vectors, respectively. In general, contravariant and covariant quantities are labeled

with upper and lower indices, respectively, and for each such index there is a factor in the transformation rule of $\partial x'/\partial x$ or $\partial x/\partial x'$, respectively. It is also possible to have mixed quantities, with some upper and some lower indices. For instance, there is a mixed tensor δ^μ_ν, defined in any coordinate system by

$$\delta^\mu_v \equiv \begin{cases} 1 & \mu = v \\ 0 & \mu \neq v \end{cases} . \tag{B.5}$$

Even though its components are the same in all coordinate systems this is a tensor because

$$\delta^\mu_\nu \frac{\partial x'^\rho}{\partial x^\mu}\frac{\partial x^\nu}{\partial x'^\sigma} = \frac{\partial x'^\rho}{\partial x^\mu}\frac{\partial x^\mu}{\partial x'^\sigma} = \delta^\rho_\sigma.$$

There is also a contravariant tensor $g^{\mu\nu}$, defined as the inverse of the metric

$$g^{\mu\lambda}g_{\lambda v} = \delta^\mu_v. \tag{B.6}$$

To see that this is a tensor, just note that

$$\left(g^{\rho\sigma}(x)\frac{\partial x'^\mu}{\partial x^\rho}\frac{\partial x'^\lambda}{\partial x^\sigma}\right) g'_{\lambda\nu}(x') = g^{\rho\sigma}(x)\frac{\partial x'^\mu}{\partial x^\rho}\frac{\partial x'^\lambda}{\partial x^\sigma}\frac{\partial x^\eta}{\partial x'^\lambda}\frac{\partial x^\tau}{\partial x'^\nu}g_{\eta\tau}(x)$$
$$= g^{\rho\sigma}(x)\frac{\partial x'^\mu}{\partial x^\rho}\delta^\eta_\sigma\frac{\partial x^\tau}{\partial x'^\nu}g_{\eta\tau}(x) = \delta^\rho_\tau\frac{\partial x'^\mu}{\partial x^\rho}\frac{\partial x^\tau}{\partial x'^\nu} = \delta^\mu_\nu$$

Thus the quantity in parenthesis in the first line is the reciprocal of the transformed metric

$$g^{\rho\sigma}(x)\frac{\partial x'^\mu}{\partial x^\rho}\frac{\partial x'^\lambda}{\partial x^\sigma} = g'^{\mu\lambda}(x')$$

verifying that the reciprocal of the metric is a contravariant tensor.
A scalar $s(x)$ is a quantity whose value at a physical spacetime point is not changed by a coordinate transformation; that is, using a set of coordinates x'^μ it is

$$s'(x') = s(x) \tag{B.7}$$

The derivative $v_\mu \equiv \partial s/\partial x^\mu$ of a scalar $s(x)$ is a covariant vector:

$$v'_\rho \equiv \frac{\partial s'}{\partial x'^\rho} = \frac{\partial s}{\partial x^\mu}\frac{\partial x^\mu}{\partial x'^\rho} = v_\mu\frac{\partial x^\mu}{\partial x'^\rho}. \tag{B.8}$$

Scalars and vectors may be regarded as tensors with no indices or one index, respectively. We can make tensors out of other tensors by taking the direct product; for instance, if $A^\mu{}_\nu$ and $B_{\rho\sigma}$ are tensors, then so is the direct product $C^\mu{}_{\nu\rho\sigma} \equiv A^\mu{}_\nu B_{\rho\sigma}$. We can also make tensors with fewer indices out of other tensors by contracting upper and lower indices, as Eqs. (B.2) and (B.6). As a special case, we often lower (or raise) an index on a tensor by taking the direct product of the tensor with the metric (or its inverse) and then contracting an upper (or lower) index on the tensor with an index on the metric (or its inverse). For instance, if $A^{\mu\nu}$ is a tensor, then so is $A_\rho{}^\nu \equiv g_{\mu\rho}A^{\mu\nu}$, while if $B_\rho{}^\nu$ is a tensor then so is $B^{\sigma\nu} = g^{\sigma\rho}B_\rho{}^\nu$. Note that raising and lowering the same index just gives back the original tensor; for instance $g^{\sigma\rho}(g_{\mu\rho}A^{\mu\nu}) = A^{\sigma\nu}$. Any equation that states the equality of two tensors of the same type or that a tensor of any type vanishes is generally covariant.

4 The affine connection: Falling bodies

But not everything is a tensor. For instance, Eq. (B.4) tells us that the first derivative of the coordinate x^μ of a particle with respect to some scalar quantity u that parameterizes position along the particle's trajectory (such as the time on some fixed clock) is a vector

$$\frac{dx'^\rho}{du} = \frac{\partial x'^\rho}{\partial x^\mu}\frac{dx^\mu}{du}, \tag{B.9}$$

but the second derivative is not a vector

$$\begin{aligned}\frac{d^2x'^\rho}{du^2} &= \frac{d}{du}\left(\frac{\partial x'^\rho}{\partial x^\mu}\frac{dx^\mu}{du}\right)\\ &= \frac{\partial x'^\rho}{\partial x^\mu}\frac{d^2x^\mu}{du^2} + \frac{\partial^2 x'^\rho}{\partial x^\mu \partial x^\nu}\frac{dx^\mu}{du}\frac{dx^\nu}{du}.\end{aligned} \tag{B.10}$$

This means that the correct generalization of the equation $d^2\xi^\alpha/du^2 = 0$ for the motion of a particle in the absence of gravitation is not $d^2x^\mu/du^2 = 0$, because this equation is not generally covariant. Instead, to cancel the second term on the right-hand side of Eq. (B.10) we must introduce a quantity $\Gamma^\lambda_{\mu\nu}(x)$ defined by the transformation property

$$\Gamma'^\tau_{\sigma\rho} = \frac{\partial x'^\tau}{\partial x^\lambda}\frac{\partial x^\mu}{\partial x'^\sigma}\frac{\partial x^\nu}{\partial x'^\rho}\Gamma^\lambda_{\mu\nu} - \frac{\partial^2 x'^\tau}{\partial x^\mu \partial x^\nu}\frac{\partial x^\mu}{\partial x'^\sigma}\frac{\partial x^\nu}{\partial x'^\rho}, \tag{B.11}$$

and the proviso that $\Gamma^\lambda_{\mu\nu}(x)$ vanishes in a coordinate system that is locally inertial and Cartesian at x. The correct equation of motion for a particle
that is freely falling in a gravitational field is then

$$\frac{d^2x^\lambda}{du^2} + \Gamma^\lambda_{\mu\nu}\frac{dx^\mu}{du}\frac{dx^\nu}{du} = 0, \tag{B.12}$$

because this is generally covariant, and reduces to the correct equation of motion $d^2\xi^\alpha/du^2 = 0$ in the absence of gravitation. The field $\Gamma^\lambda_{\mu\nu}$ is known as the affine connection, and of course is not a tensor.

There is a simple formula for the affine connection in terms of the metric

$$\Gamma^\lambda_{\mu\nu} = \frac{1}{2}g^{\lambda\rho}\left[\frac{\partial g_{\rho\mu}}{\partial x^\nu} + \frac{\partial g_{\rho\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial x^\rho}\right]. \tag{B.13}$$

It is straightforward to check that this is generally covariant, and it is true in a locally inertial Cartesian coordinate system because in such a system both sides vanish, so it is correct for general gravitational fields and coordinate systems.

A trajectory that satisfies Eq. (B.12) is called a spacetime geodesic, because on such a trajectory the integral

$$\int_{u_1}^{u_2}\sqrt{g_{\mu\nu}(x(u))\frac{dx^\mu(u)}{du}\frac{dx^\nu(u)}{du}}du$$

is stationary under variations that leave $x^\mu(u)$ fixed at the endpoints u_1 and u_2. Often instead of specifying the metric we specify the line element $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$ for arbitrary differentials dx^μ.

The equation of motion (B.12) is not valid for just any choice of the parameter u. To see this, note that Eq. (B.12) implies a conservation law

$$\frac{d}{du}\left[g_{\mu\nu}\frac{dx^\mu}{du}\frac{dx^\nu}{du}\right] = \frac{\partial g_{\mu\nu}}{\partial x^\lambda}\frac{dx^\mu}{du}\frac{dx^\nu}{du}\frac{dx^\lambda}{du} + g_{\mu\nu}\frac{d^2x^\mu}{du^2}\frac{dx^\nu}{du} + g_{\mu\nu}\frac{dx^\mu}{du}\frac{d^2x^\nu}{du^2}$$
$$= \left(\frac{\partial g_{\mu\nu}}{\partial x^\lambda} - g_{\mu\kappa}\Gamma^\kappa_{\nu\lambda} - g_{\nu\kappa}\Gamma^\kappa_{\mu\lambda}\right)\frac{dx^\mu}{du}\frac{dx^\nu}{du}\frac{dx^\lambda}{du}$$
$$= 0 \tag{B.14}$$

It follows that u must be a linear function of the proper time τ, defined by

$$d\tau \equiv \sqrt{-g_{\mu\nu}dx^\mu dx^\nu}, \tag{B.15}$$

which according to Eq. (B.2) is the time told by a clock that falls freely along with the particle. The only exception to this conclusion is for massless particles like photons, whose spacetime trajectory satisfies the same equation of motion (B.12) as for a massive particle, but for which the conserved quantity
$d\tau/du$ vanishes. For a massless particle or a ray of light we need to choose the parameter u as the time told by some other freely falling clock.

In the case of non-zero mass m, it is convenient to take the affine parameter u as $u = \tau/m$, for then we can define the energy-momentum four-vector as

$$p^\mu = m\frac{dx^\mu}{d\tau} = \frac{dx^\mu}{du},$$

and, using Eq. (B.15),

$$g_{\mu\nu}p^\mu p^\nu = -m^2.$$

For massless particles we have $g_{\mu\nu}\left(dx^\mu/du\right)\left(dx^\nu/du\right) = 0$ however u is normalized, so we can simply suppose that it is normalized in such a way that the energy-momentum four-vector is $p^\mu = dx^\mu/du$.

5 Gravitational time dilation

These results allow us to derive one of the most important consequences of the Equivalence Principle. For a slowly moving particle, dx^i/du is much less than dx^0/du, so Eq. (B.12) becomes

$$\frac{d^2x^i}{du^2} + \Gamma^i_{00}\frac{dx^0}{du}\frac{dx^0}{du} = 0. \tag{B.16}$$

For a weak gravitational field, the metric $g_{\mu\nu}$ is nearly equal to the Minkowski metric $\eta_{\mu\nu}$, so

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{B.17}$$

with the components of $h_{\mu\nu}$ much less than unity. We then can take $u = \tau \simeq x^0 \equiv t$, so the equation of motion of a freely falling slowly moving particle in a weak gravitational field is

$$\frac{d^2x^i}{dt^2} = -\Gamma^i_{00}, \tag{B.18}$$

where i runs over the values $1, 2, 3$, labeling spatial directions in a Cartesian coordinate system. The affine connection for a weak gravitational field is

$$\Gamma^\lambda_{\mu\nu} \simeq \frac{\eta^{\lambda\rho}}{2}\left[\frac{\partial h_{\rho\mu}}{\partial x^\nu} + \frac{\partial h_{\rho\nu}}{\partial x^\mu} - \frac{\partial h_{\mu\nu}}{\partial x^\rho}\right]. \tag{B.19}$$

In particular, for a weak time-independent gravitational field we have

$$\Gamma^i_{00} \simeq -\frac{1}{2}\frac{\partial h_{00}}{\partial x^i}. \tag{B.20}$$

Eqs. (B.18) and (B.20) allow us to identify $-h_{00}/2$ as the Newtonian gravitational potential ϕ, so in a weak static gravitational field we have

$$g_{00} \simeq -1 - 2\phi. \tag{B.21}$$

Now consider a clock at rest in such a field. According to Eq. (B.2), if the time between ticks in the absence of a gravitational field is dT, then in the presence of the field it is dt, where

$$(-1 - 2\phi)dt^2 \simeq -dT^2.$$

Hence the time between ticks is no longer dT, but rather

$$dt \simeq (1 - \phi)dT. \tag{B.22}$$

In the negative gravitational potential at the surface of a star clocks therefore tick more slowly than in interstellar space, or in the much weaker gravitational potential at the surface of the earth. This could not be observed on the star's surface, since all physical processes would be slowed there by the same factor, but it is observed at a distance, by measuring the decrease in the frequency of photons emitted from atomic transitions on the star's surface. The gravitational time dilation is measured most accurately by observing the shift of spectral lines as photons rise or fall in the earth's gravitational field.

6 Covariant derivatives

Although the spacetime derivative of a scalar field is a vector, the derivative of a vector or a tensor field is in general not a tensor. For instance, a contravariant vector field v^μ has the transformation property

$$v'^\rho = \frac{\partial x'^\rho}{\partial x^\mu} v^\mu, \tag{B.23}$$

so that

$$\frac{\partial v'^\rho}{\partial x'^\sigma} = \frac{\partial x'^\rho}{\partial x^\mu}\frac{\partial x^\nu}{\partial x'^\sigma}\frac{\partial v^\mu}{\partial x^\nu} + \frac{\partial^2 x'^\rho}{\partial x^\mu \partial x^\nu}\frac{\partial x^\nu}{\partial x'^\sigma} v^\mu. \tag{B.24}$$

To construct a tensor we must add a term that cancels the second term in this transformation law. In this way we are led to introduce a covariant derivative

$$v^\mu_{;v} \equiv \frac{\partial v^\mu}{\partial x^v} + \Gamma^\mu_{v\lambda} v^\lambda, \tag{B.25}$$

which does transform as a mixed tensor

$$v'^\rho_{;\sigma} = \frac{\partial x'^\rho}{\partial x^\mu}\frac{\partial x^\nu}{\partial x'^\sigma} v^\mu_{;\nu}. \tag{B.26}$$

Likewise, the covariant derivative of a covariant vector is defined by

$$v_{v;\mu} \equiv \frac{\partial v_\nu}{\partial x^\mu} - \Gamma^\lambda_{\mu\nu} v_\lambda, \tag{B.27}$$

and is a covariant tensor. More generally, the covariant derivative of a tensor with any number of upper and/or lower indices is given by similar formulas, with a $+\,\Gamma$ for every upper

index and a $-\Gamma$ for every lower index. It is easy to check that the covariant derivatives $g_{\mu\nu;\lambda}$ of the metric tensor all vanish, as they must, since in a locally inertial Cartesian frame the covariant derivative is an ordinary derivative and the first derivatives of the metric vanish, so that the covariant derivatives vanish, and a tensor $g_{\mu\nu;\lambda}$ that vanishes in one coordinate system must vanish in all coordinate systems.

7 Effects of gravitation: The Maxwell equations

Given the equations that govern some set of fields in the absence of gravitation, we can find the equations that apply (at least in sufficiently small regions) in a gravitational field by replacing all Minkowski metrics η with metric tensors and all derivatives with covariant derivatives. As mentioned earlier, the procedure does not give a unique result, since there are tensors formed from second and higher derivatives of the metric that vanish in the absence of gravitation, but the effect of including such tensors in the generally covariant field equations would be negligible in a sufficiently small spacetime region.

For instance, in Cartesian coordinates in the absence of gravitation electric and magnetic fields are governed by Maxwell's equations

$$\partial_\alpha F^{\alpha\beta} = -J^\beta, \tag{B.28}$$

$$\partial_\alpha F_{\beta\gamma} + \partial_\beta F_{\gamma\alpha} + \partial_\gamma F_{\alpha\beta} = 0, \tag{B.29}$$

where $F^{\alpha\beta}$ is the electromagnetic field strength tensor (with $F^{01} = E_1$, $F^{23} = B_1$, etc.), J^α is the electric current four-vector (with J^0 the charge density and J^1, J^2, J^3 the electric current density), and

$$F_{\alpha\beta} \equiv \eta_{\alpha\gamma}\eta_{\beta\delta}F^{\gamma\delta}. \tag{B.30}$$

Hence in the presence of gravitation the field equations are

$$F^{v\mu}_{\ ;v} = -J^\mu \tag{B.31}$$

$$F_{\mu v;\lambda} + F_{v\lambda;\mu} + F_{\lambda\mu;\nu} = 0, \tag{B.32}$$

with

$$F_{\mu\nu} \equiv g_{\mu\lambda}g_{\nu\kappa}F^{\lambda\kappa} \tag{B.33}$$

We use the same letter of the alphabet for tensors like $F_{\mu\nu}$ and $F^{\mu\nu}$ that are related by raising and lowering indices by contraction with the metric, because they represent the same physical quantity.

8 Currents and conservation laws

For a system of particles labeled by an index n, with spacetime coordinates $x^\mu_n(u)$ and electric charges e_n, the electric current four-vector is given by

$$\begin{aligned} J^\mu(x) &\equiv \frac{1}{\sqrt{-\operatorname{Det} g(x)}} \int du \sum_n e_n \delta^4\left(x_n(u) - x\right) \frac{dx^\mu(u)}{du} \\ &= \frac{1}{\sqrt{-\operatorname{Det} g(x)}} \sum_n e_n \delta^3\left(\mathbf{x}_n(t) - \mathbf{x}\right) \frac{dx^\mu(t)}{dt} \end{aligned} \tag{B.34}$$

Here $\delta^4(z)$ is a fictitious function with an infinitely narrow and infinitely high peak at $z = 0$, normalized so that, for any smooth function $f(z)$,

$$\int d^4x f(y)\delta^4(y-x) = f(x)$$

and $\delta^3(\mathbf{z})$ is the same in three dimensions. In particular, the integral of $\sqrt{-\text{ Detg }}J^0$ over a finite three-dimensional volume equals the total electric charge within that volume. (In cosmology we would be more interested in the baryon current, with the baryon number of the nth particle or the nth galaxy appearing instead of e_n.) The factor $1/\sqrt{-\text{Det}\, g(x)}$ is needed here, because the four-dimensional delta function is not a scalar. We can see this by noting that, under a transformation from coordinates x^μ to coordinates x'^μ, the differential spacetime volume element is changed to

$$d^4x' = \left|\frac{\partial x'}{\partial x}\right| d^4x \tag{B.35}$$

where $|\partial x'/\partial x|$ is the Jacobian of the coordinate transformation - that is, the determinant of the matrix whose components are $\partial x'^\mu/\partial x^\nu$. The Jacobian can be expressed in terms of the determinants of the metrics; by taking the determinant of Eq. (B.3), we find

$$\text{Det}\, g' = \left|\frac{\partial x'}{\partial x}\right|^{-2} \text{Det}\, g. \tag{B.36}$$

Hence it is the spacetime volume $d^4x\sqrt{-\text{Det}\, g}$ rather than d^4x that transforms as a scalar. (A minus sign is inserted in front of the determinant of the metric, because in physical spacetimes this determinant is negative.) From the defining equation of the delta function

$$\begin{aligned} f(y) &= \int d^4x f(x)\delta^4(x-y) \\ &= \int \left(d^4x\sqrt{-\text{Det}\, g(x)}\right) f(x) \left(\delta^4(x-y)/\sqrt{-\text{Det}\, g(x)}\right) \end{aligned}$$

we see that it is the ratio $\delta^4\left(x_n - x\right)/\sqrt{-\text{Det}\, g(x)}$ appearing in the current (B.34) rather than the delta function itself that transforms as a scalar. This current satisfies the conservation law

$$\partial_\mu\left(\sqrt{-\text{Det}\, g(x)}J^\mu(x)\right) = \int du \frac{d}{du}\sum_n e_n\delta^4\left(x_n(u) - x\right) = 0, \tag{B.37}$$

provided x is not at the value of any $x_n(u)$ at either endpoint of the integral. This is the same as the generally covariant conservation law

$$0 = J^\mu_{;\mu} \equiv \frac{\partial}{\partial x^\mu}J^\mu + \Gamma^\mu_{\nu\mu}J^\nu \tag{B.38}$$

because

$$\Gamma^\nu_{\mu\nu} = \frac{1}{2}g^{\nu\lambda}\partial_\mu g_{\nu\lambda} = \frac{1}{2}\partial_\mu \ln(-\text{Det}\, g).$$

This is the correct conservation condition, because in the absence of gravitation there is a current that in Cartesian coordinate systems satisfies the conservation law $\partial_\mu J^\mu = 0$, and therefore in general coordinates in a gravitational field must satisfy the generally covariant generalization (B.38).

9 The energy-momentum tensor

Likewise, in the absence of gravitation any set of particles and/or fields will have a symmetric energy-momentum tensor $T^{\alpha\beta}$, which is conserved in the sense that

$$\frac{\partial T^{\alpha\beta}}{\partial x^\beta} = 0. \tag{B.39}$$

Just as J^β is the β component of the current of electric charge, we can think of $T^{\alpha\beta}$ as the β component of the current of p^α. In the presence of a gravitational field the conservation law becomes

$$T^{\mu\nu}{}_{;\nu} \equiv \frac{\partial T^{\mu\nu}}{\partial x^\nu} + \Gamma^\mu_{\kappa\nu} T^{\kappa\nu} + \Gamma^\nu_{\kappa\nu} T^{\mu\kappa} = 0. \tag{B.40}$$

The final Γ term here is a geometric effect, similar to what we found for conserved currents, but the other Γ term represents the exchange of energy and momentum between gravitation and the other fields.

For instance, for an ideal gas of particles that move freely except for gravitational forces and perhaps for collisions that are localized in space, the energy-momentum tensor is given by replacing e_n in Eq. (B.34) with the energy-momentum four-vector $p^\mu_n = E_n dx^\mu_n/dt$, whose spatial components are the components of the three-momentum, and $p^0_n = E_n$:

$$\begin{aligned} T^{\mu v}(x) &= \frac{1}{\sqrt{-\operatorname{Det} g(x)}} \int du \sum_n \delta^4\left(x_n(u) - x\right) \frac{dx^\mu_n(u)}{du} p^v_n(u) \\ &= \frac{1}{\sqrt{-\operatorname{Det} g(x)}} \sum_n \delta^3\left(\mathbf{x}_n(t) - \mathbf{x}\right) p^\mu_n(t) p^v_n(t)/E_n(t). \end{aligned} \tag{B.41}$$

In particular, T^{00} is the energy density. Direct calculation using the equation of motion (B.12) shows that this satisfies the covariant conservation law (B.40).

10 Perfect and imperfect fluids

A perfect fluid is defined as a medium for which at every point there is a locally inertial Cartesian frame of reference, moving with the fluid, in which the fluid appears the same in all directions. In such a locally inertial co-moving frame the components of the energy momentum tensor must take the form

$$T^{ij} = \delta_{ij} p, \quad T^{i0} = T_{0i} = 0, \quad T^{00} = \rho,$$

where i and j run over the three Cartesian coordinate directions 1,2,3. (This is because non-zero value of T^{i0} and any term in T^{ij} other than one proportional to δ_{ij} would select out special directions in space, such as the direction of T^{i0}, or of one of the non-degenerate eigenvectors of T^{ij}.) The coefficients p and ρ are known as the pressure and energy density, respectively. Then in a locally inertial Cartesian frame with an arbitrary velocity, the energymomentum tensor takes the form

$$T^{\alpha\beta} = p\eta^{\alpha\beta} + (p + \rho) u^\alpha u^\beta, \tag{B.42}$$

where ρ and p are defined to be the same as in the co-moving inertial frame, and u^α is defined by the conditions that it transforms as a fourvector under Lorentz transformations, and that in the locally Cartesian co-moving inertial frame it has components $u^0 = 1$ and $u^i = 0$. This fourvector, known as the velocity vector, is normalized so that, in any inertial

frame, $\eta_{\alpha\beta}u^\alpha u^\beta = -1$. It follows that in a general gravitational field the energy-momentum tensor of a perfect fluid is

$$T^{\mu\nu} = pg^{\mu\nu} + (p+\rho)u^\mu u^\nu, \quad g_{\mu\nu}u^\mu u^\nu = -1 \tag{B.43}$$

where p and ρ are defined by the condition that they are equal to the coefficients in the energy-momentum tensor in a locally co-moving inertial coordinate system, so that they are scalars, and u^μ is defined by the conditions that it transforms as a four-vector under general coordinate transformations and has the components $u^0 = 1$ and $u^i = 0$ in the locally co-moving Cartesian inertial frame. This formula for $T^{\mu\nu}$ is correct because it is generally covariant and it is true in locally inertial Cartesian coordinate systems. The equations of relativistic hydrodynamics in a gravitational field are derived by imposing the conservation condition (B.40) on this tensor. In addition, if the pressure depends on the density n of some conserved quantity such as baryon number as well as on the energy density ρ, then we need the equation of conservation, which in locally inertial Cartesian frames reads

$$\frac{\partial}{\partial x^\alpha}(nu^\alpha) = 0. \tag{B.44}$$

Thus in a general coordinate system in an arbitrary gravitational field, we have

$$(nu^\mu)_{;\mu} = 0. \tag{B.45}$$

For an imperfect fluid, there is a small correction $\Delta T^{\alpha\beta}$ to formula (B.42) for the energy-momentum tensor in locally inertial Cartesian coordinate system:

$$T^{\alpha\beta} = p\eta^{\alpha\beta} + (p+\rho)u^\alpha u^\beta + \Delta T^{\alpha\beta} \tag{B.46}$$

and a small correction ΔN^α to whatever current may be conserved

$$\frac{\partial}{\partial x^\alpha}(nu^\alpha + \Delta N^\alpha) = 0. \tag{B.47}$$

The scalar ρ is defined as the energy density observed in a co-moving frame in which $u^i = 0$, so that in this frame $\Delta T^{00} \equiv 0$. This implies that in all locally inertial Cartesian frames $u_\alpha u_\beta \Delta T^{\alpha\beta} = 0$, since this quantity is a scalar and vanishes in a co-moving frame. The scalar n can be defined as the value of the conserved density observed in such a co-moving frame, so by the same reasoning in all locally inertial Cartesian frames we have $u_\alpha \Delta N^\alpha = 0$. The pressure can be defined as whatever function of ρ and perhaps n gives
the pressure in a static homogeneous fluid. (A different definition is used in Chapter 5 et seq.) But the definition of the velocity four-vector u^α remains somewhat ambiguous. We could define u^i to be the velocity of particle transport, [1] in which case in a co-moving frame with $u^i = 0$ we would have $\Delta N^i = 0$ as well as $\Delta N^0 = 0$, so that in general locally inertial Cartesian frames $\Delta N^\alpha = 0$. Instead we will define u^i to be the velocity of energy transport, [2] so that in co-moving frames we also have $T^{i0} = \Delta T^{i0} = 0$, and so $u_\beta \Delta T^{\alpha\beta} = \Delta T^{\alpha 0} = 0$ in this frame, which implies that $u_\beta \Delta T^{\alpha\beta} = 0$ in all locally inertial Cartesian frames, but in general $\Delta N^\alpha \neq 0$. With this definition of velocity, the second law of thermodynamics together with the conditions $u_\beta \Delta T^{\alpha\beta} = 0$ and $u_\alpha \Delta N^\alpha = 0$ gives [3]

$$\Delta T_{\alpha\beta} = -\eta\left(\frac{\partial u_\alpha}{\partial x^\beta} + \frac{\partial u_\beta}{\partial x^\alpha} + u_\beta u^\gamma \frac{\partial u_\alpha}{\partial x^\gamma} + u_\alpha u^\gamma \frac{\partial u_\beta}{\partial x^\gamma}\right) - \left(\zeta - \frac{2}{3}\eta\right)\frac{\partial u^\gamma}{\partial x^\gamma}\left(\eta_{\alpha\beta} + u_\alpha u_\beta\right) \tag{B.48}$$

$$\Delta N_\alpha = -\chi\left(\frac{nT}{\rho + p}\right)^2 \left[\frac{\partial}{\partial x^\alpha}\left(\frac{\mu}{T}\right) + u_\alpha u^\beta \frac{\partial}{\partial x^\beta}\left(\frac{\mu}{T}\right)\right]. \tag{B.49}$$

Here η, ζ, and χ are the positive coefficients of shear viscosity, bulk viscosity, and heat conduction, respectively, and μ is the chemical potential associated with the conserved quantum number, defined by the condition that the entropy density is $(p + \rho - \mu n)/T$. It is then an immediate consequence of the Equivalence Principle that in general coordinate systems in arbitrary gravitational fields that vary little over a mean free path or mean free time,

$$\Delta T_{\mu\nu} = -\eta\left(u_{\mu;\nu} + u_{\nu;\mu} + u_\nu u^\kappa u_{\mu;\kappa} + u_\mu u^\kappa u_{\nu;\kappa}\right) - \left(\zeta - \frac{2}{3}\eta\right) u^\kappa_{;\kappa}\left(g_{\mu\nu} + u_\mu u_\nu\right), \tag{B.50}$$

$$\Delta N_\mu = -\chi\left(\frac{nT}{\rho + p}\right)^2 \left[\frac{\partial}{\partial x^\mu}\left(\frac{\mu}{T}\right) + u_\mu u^\nu \frac{\partial}{\partial x^\nu}\left(\frac{\mu}{T}\right)\right]. \tag{B.51}$$

11 The action principle

There is a general algorithm for deriving the energy-momentum tensor for systems that may be more complicated than an ideal gas or perfect fluid, provided only that they are governed by an action principle. According to the action principle, the differential equations governing the behavior of particles and fields can be expressed as the statement that the "matter action," a functional I_m of the fields and particle trajectories, is stationary with respect to infinitesimal variations of the fields and particles. The Equivalence Principle tells us to include the metric in the matter action in such a way that I_m is invariant under general coordinate transformations. Then the change in the action when we make an infinitesimal change $\delta g_{\mu\nu}$ in the metric (leaving all other dynamical variables unchanged) must be of the form

$$\delta I_m = \frac{1}{2}\int d^4x \sqrt{-\operatorname{Det} g(x)}\, T^{\mu v}(x)\delta g_{\mu v}(x) \tag{B.52}$$

where $T^{\mu\nu}(x)$ is a symmetric tensor, which we identify as the energymomentum tensor.

For instance, the action for a gas of charged particles with masses m_n, charges e_n, and trajectories $x^\mu_n(u)$ interacting with electromagnetic fields is taken as

[471] This is the option adopted by C. Eckart, Phys. Rev. 58, 919 (1940), and also in Secs. 2.11 and 15.10 of G&C.

[2] This is the definition used by L. D. Landau and E. M. Lifschitz, Fluid Mechanics, trans. by J. B. Sykes and W. H. Reid (Pergamon Press, London, 1959), Section 127. We will adopt this definition of velocity in this book, because it imposes the maximum possible constraint on the energy-momentum tensor at the cost of putting less of a constraint on the current of conserved quantities, and in cosmology we frequently have to do with situations in which there are either no non-zero conserved quantities at all, such as in the early universe before cosmological leptogenesis or baryongenesis, or no conserved quantities that are large enough to seriously affect the relation between pressure and density, as in the radiation-dominated era at temperatures above about 10^4 K.

[3] Landau & Lifschitz, op. cit.

$$I_m = -\frac{1}{4}\int d^4x\sqrt{-\operatorname{Det} g}F_{\mu\nu}F_{\rho\sigma}g^{\mu\rho}g^{\nu\sigma}$$
$$-\sum_n m_n\int du\left[-g_{\mu\nu}\left(x_n(u)\right)\frac{dx_n^\mu(u)}{du}\frac{dx_n^\nu(u)}{du}\right]^{1/2}$$
$$+\sum_n e_n\int du\frac{dx_n^\mu(u)}{du}A_\mu(u)$$

in which the homogeneous Maxwell equations (B.29) are enforced by writing the field strength tensor in terms of a vector potential A_μ as $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. The reader can verify that the conditions for the matter action to be stationary with respect to arbitrary small variations in $A_\mu(x)$ and $x_n^\mu(u)$ (arbitrary, except that they vanish for $x^\mu \to \infty$ or $u \to \pm\infty$) are the inhomogeneous Maxwell equations (B.28) together with the equations of motion of charged particles in a combined gravitational and electromagnetic fields

$$m_n\left[\frac{d^2x_n^\mu}{d\tau_n^2} + \Gamma^\mu_{\nu\rho}\frac{dx_n^\nu}{d\tau_n}\frac{dx_n^\rho}{d\tau_n}\right] = e_n\frac{dx_n^\nu}{d\tau_n}F^\mu_\nu\left(x_n\right),$$

where τ_n is the invariant proper time along the nth particle trajectory, defined by

$$\frac{d\tau_n}{du} \equiv \left[-g_{\mu\nu}\left(x_n(u)\right)\frac{dx_n^\mu(u)}{du}\frac{dx_n^\nu(u)}{du}\right]^{1/2}.$$

Eq. (B.52) gives an energy-momentum tensor

$$T^{\mu\nu}(x) = [-\operatorname{Det} g(x)]^{-1/2}\sum_n m_n\int d\tau_n\frac{dx_n^\mu}{d\tau_n}\frac{dx_n^\nu}{d\tau_n}\delta^4\left(x - x_n\right)$$
$$+ F^{\rho\mu}(x)F_\rho{}^\nu(x) - \frac{1}{4}g^{\mu\nu}(x)F^{\rho\sigma}(x)F_{\rho\sigma}(x)$$

which is the same as Eq. (B.41), with extra terms representing the energy and momentum in the electromagnetic field. (In the derivation we use the relations $\delta\operatorname{Det} g = \operatorname{Det} g g^{\mu\nu}\delta g_{\mu\nu}$ and $\delta g^{\mu\rho} = -g^{\mu\nu}\delta g_{\nu\sigma}g^{\sigma\rho}$.)

The justification of the identification of $T^{\mu\nu}$ in Eq. (B.52) is that this tensor is conserved, in the sense of Eq. (B.40). To show this, note that in general our assumption that I_m is a scalar tells us that it is unchanged if we simultaneously make the replacements

$$d^4x \to d^4x', \quad \frac{\partial}{\partial x^\mu} \to \frac{\partial}{\partial x'^\mu}, \tag{B.53}$$

$$x_n^\mu(u) \to x_n'^\mu(u), \quad A_\mu(x) \to A'_\mu\left(x'\right) = \frac{\partial x^\nu}{\partial x'^\mu}A_\nu(x), \tag{B.54}$$

$$g_{\mu\nu}(x) \to g'_{\mu\nu}\left(x'\right) = \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\sigma}{\partial x'^\nu}g_{\rho\sigma}(x), \tag{B.55}$$

and likewise for any other fields entering in the action. But the coordinate x'^μ (unlike $x_n'^\mu(u)$) is just a variable of integration, so we can change x'^μ back to x^μ everywhere without changing I_m. It follows that I_m is unchanged by the replacements

$$x_n^\mu(u) \to x_n'^\mu(u) \tag{B.56}$$

$$A_\mu(x) \to A'_\mu(x) = \frac{\partial x^\nu}{\partial x'^\mu}A_\nu(x) - \left[A'_\mu\left(x'\right) - A'_\mu(x)\right] \tag{B.57}$$

$$g_{\mu\nu}(x) \to g'_{\mu\nu}(x) = \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\sigma}{\partial x'^\nu}g_{\rho\sigma}(x) - \left[g'_{\mu\nu}\left(x'\right) - g'_{\mu\nu}(x)\right] \tag{B.58}$$

with x^μ and $\partial/\partial x^\mu$ now left unchanged. (This combination of a coordinate transformation and a relabeling of coordinates is sometimes called a gauge transformation.) For a general infinitesimal coordinate transformation we have $x^\mu \to x^\mu + \epsilon^\mu(x)$, with $\epsilon^\mu(x)$ arbitrary infinitesimal functions of x.

Then the transformations (B.56)-(B.58) become

$$x_n^\mu(u) \to x_n^\mu(u) + \epsilon^\mu\left(x_n\right) \tag{B.59}$$

$$A_\mu(x) \to A_\mu(x) - \frac{\partial\epsilon^\nu(x)}{\partial x^\mu}A_\nu(x) - \frac{\partial A_\mu(x)}{\partial x^\nu}\epsilon^\nu(x) \tag{B.60}$$

$$\begin{aligned} g_{\mu\nu}(x) \to & g_{\mu\nu}(x) - \frac{\partial\epsilon^\rho(x)}{\partial x^\mu}g_{\rho\nu}(x) - \frac{\partial\epsilon^\sigma(x)}{\partial x^\nu}g_{\mu\sigma}(x) \\ & - \frac{\partial g_{\mu\nu}(x)}{\partial x^\rho}\epsilon^\rho(x) \end{aligned} \tag{B.61}$$

Now, as long as the equations of motion of particles and the field equations for "matter" (including electromagnetic) fields are satisfied, the action is unaffected by any infinitesimal changes in particle trajectories and matter fields. On the other hand, unlike the total action of matter plus gravitation, the matter action is not stationary under variations in the metric, even when the field equations are satisfied. For a general infinitesimal coordinate transformation, using Eq. (B.61) in Eq. (B.52) lets us write the condition that I_m is a scalar as

$$\begin{aligned} 0 = \delta I_m = \int d^4x\sqrt{-\operatorname{Det} g(x)}T^{\mu v}(x)\Bigg[& -\frac{\partial\epsilon^\rho(x)}{\partial x^\mu}g_{\rho v}(x) - \frac{\partial\epsilon^\rho(x)}{\partial x^\nu}g_{\mu\rho}(x) \\ & -\frac{\partial g_{\mu\nu}(x)}{\partial x^\rho}\epsilon^\rho(x)\Bigg] \end{aligned} \tag{B.62}$$

Integrating by parts and setting the coefficient of $\epsilon^\rho(x)$ equal to zero then yields the conservation condition (B.40).

12 Scalar field theory

We will frequently encounter cosmological models involving a scalar field $\varphi(x)$, with an action [4]

$$I_\varphi = -\int d^4x\sqrt{-\operatorname{Det} g}\left[\frac{1}{2}g^{\mu\nu}\frac{\partial\varphi}{\partial x^\mu}\frac{\partial\varphi}{\partial x^\nu} + V(\varphi)\right], \tag{B.63}$$

where $V(\varphi)$ is a function known as the potential. The field equations for φ in a gravitational field are given by the condition that this be stationary with respect to variations in φ :

$$\frac{1}{\sqrt{-\operatorname{Det} g}}\frac{\partial}{\partial x^\mu}\left[\sqrt{-\operatorname{Det} g}g^{\mu\nu}\frac{\partial\varphi}{\partial x^\nu}\right] = \frac{\partial V(\varphi)}{\partial\varphi}. \tag{B.64}$$

The energy-momentum tensor for the scalar field is found by varying the metric and comparing with Eq. (B.40):

$$T_\varphi^{\mu\nu} = -g^{\mu\nu}\left[\frac{1}{2}g^{\rho\sigma}\frac{\partial\varphi}{\partial x^\rho}\frac{\partial\varphi}{\partial x^\sigma} + V(\varphi)\right] + g^{\mu\rho}g^{\nu\sigma}\frac{\partial\varphi}{\partial x^\rho}\frac{\partial\varphi}{\partial x^\sigma}. \tag{B.65}$$

[474] In general, the coefficient of the first term in square brackets might depend on φ, but such a fielddependent coefficient can always be eliminated by a redefinition of the scalar field. This simplification is not generally possible, however, with more than one scalar field, the case discussed in Section 10.4.

This has the same form as the energy-momentum tensor (B.43) for a perfect fluid, with energy density, pressure, and velocity four-vector given by [5]

$$\rho = -\frac{1}{2} g^{\mu\nu} \frac{\partial\varphi}{\partial x^\mu} \frac{\partial\varphi}{\partial x^\nu} + V(\varphi) \tag{B.66}$$

$$p = -\frac{1}{2} g^{\mu\nu} \frac{\partial\varphi}{\partial x^\mu} \frac{\partial\varphi}{\partial x^\nu} - V(\varphi) \tag{B.67}$$

$$u^\mu = -\left[-g^{\rho\sigma} \frac{\partial\varphi}{\partial x^\rho} \frac{\partial\varphi}{\partial x^\sigma} \right]^{-1/2} g^{\mu\tau} \frac{\partial\varphi}{\partial x^\tau} \tag{B.68}$$

(The sign of u^μ does not affect the energy-momentum tensor, so it cannot be found by comparing Eqs. (B.65) and (B.43). It is chosen here so that u^0 should have the value $u^0 = +1$ in the case considered in Chapter 4− a spacetime with $g^{00} = -1$ and a scalar field that does not depend on position and increases with time - provided the square root is understood to be positive.) The reader can check that $T^{\mu\nu}_{\varphi}$ is conserved in the sense of Eq. (B.40) as a consequence of the field equation (B.64) for φ.

13 Parallel transport

A body carried along an orbit $x^\mu = x^\mu(t)$ may be characterized by one or more t-dependent vectors or tensors. If these vectors or tensors do not change at time t in a frame of reference that is locally inertial at $x^\mu(t)$, then

in an arbitrary frame of reference they will undergo a change known as parallel transport. For instance, a contravariant vector $v^\mu(t)$ will have a rate of change

$$\frac{dv^\mu(t)}{dt} = -\Gamma^\mu_{\nu\lambda}(x(t)) v^\nu(t) \frac{dx^\lambda(t)}{dt}. \tag{B.69}$$

To check that this is valid, we note that it is assumed to be true at time t in frames of reference in which the affine connection $\Gamma^\mu_{\nu\lambda}(x(t))$ vanishes, and it is generally covariant, and it is therefore true in all frames if it is true in such locally inertial frames. In particular, Eq. (B.12) shows that the momentum $p^\mu = dx^\mu/du$ satisfies the equation of parallel transport. Similarly, any tensor carried along an orbit that does not change in a frame that is locally inertial at a point along the orbit will have a rate of change in general frames at that point given by a sum of terms like the right-hand side of Eq. (B.69), with $a - \Gamma$ for every contravariant index and $a + \Gamma$ for every covariant index. For instance, for a covariant tensor $J_{\mu\nu}$, the equation of parallel transport is

[475] This result for pressure is different from that given (without explanation) by E. W. Kolb and M. S. Turner, The Early Universe (Addison-Wesley, Redwood City, CA, 1990), Eq. (8.21), according to whom the pressure of a scalar field in a Robertson-Walker metric is

$$p = \dot{\varphi}^2/2 - V(\varphi) - (\nabla\varphi)^2/6a^2,$$

instead of the result for a Robertson-Walker metric given by Eq. (B.67)

$$p = \dot{\varphi}^2/2 - V(\varphi) - (\nabla\varphi)^2/2a^2.$$

The Kolb-Turner result is obtained if we define the pressure as the value of $T^i{}_i/3$ measured by an observer in a locally inertial coordinate system moving with four-velocity v_μ

$$p \equiv \frac{1}{3} \left(g_{\mu\rho} + v_\mu v_\rho \right) T^{\rho\sigma} \left(\delta^\mu_\sigma + v_\sigma v^\mu \right),$$

but take the velocity v^μ to have components $v^0 = 1, v^i = 0$. It seems more natural in using this formula for pressure to take v^μ as the velocity u^μ given by Eq. (B.68), for this is the velocity that appears in the perfect fluid formula (B.43) for the energy-momentum tensor. This choice of v^μ leads to Eq. (B.67).

$$\frac{dJ_{\mu\nu}(t)}{dt} = +\Gamma^{\rho}_{\mu\lambda}(x(t))J_{\rho\nu}(t)\frac{dx^{\lambda}(t)}{dt} + \Gamma^{\rho}_{\nu\lambda}(x(t))J_{\mu\rho}(t)\frac{dx^{\lambda}(t)}{dt}.$$

14 The gravitational field equations

It remains only to give the equations that govern the gravitational field itself. These must satisfy two requirements: they must be generally covariant, and for weak slowly changing gravitational fields they must yield the Poisson equation for the Newtonian potential ϕ in Eq. (B.21)

$$\nabla^2\phi = 4\pi G T^{00}, \tag{B.70}$$

where G is Newton's constant. If we limit ourselves to partial differential equations that (like the Poisson equation) have just two spacetime derivatives of the metric, then the field equations are unique:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}g^{\lambda\kappa}R_{\lambda\kappa} = -8\pi G T_{\mu\nu}, \tag{B.71}$$

where $T_{\mu\nu}$ is the energy-momentum tensor with lowered indices

$$T_{\mu\nu} \equiv g_{\mu\lambda}g_{\nu\kappa}T^{\lambda\kappa} \tag{B.72}$$

and $R_{\mu\nu}$ is the Ricci tensor:

$$R_{\mu\nu} \equiv \frac{\partial\Gamma^{\lambda}_{\mu\lambda}}{\partial x^{\nu}} - \frac{\partial\Gamma^{\lambda}_{\mu\nu}}{\partial x^{\lambda}} + \Gamma^{\kappa}_{\mu\lambda}\Gamma^{\lambda}_{\nu\kappa} - \Gamma^{\kappa}_{\mu\nu}\Gamma^{\lambda}_{\lambda\kappa}. \tag{B.73}$$

For weak static fields the time-time component of this equation is the same as the Poisson equation (B.70), provided g_{00} is related to ϕ by Eq. (B.21). The tensor appearing on the left-hand side of Eq. (B.71) satisfies a set of differential Bianchi identities:

$$\left[g^{\lambda\nu}\left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}g^{\rho\kappa}R_{\rho\kappa}\right)\right]_{;\lambda} = 0. \tag{B.74}$$

This is why this is the linear combination of the tensors $R_{\mu\nu}$ and $g_{\mu\nu}g^{\lambda\kappa}R_{\lambda\kappa}$ that appears on the left-hand side of the field equations; otherwise the field equations would not be consistent with the energy-momentum conservation law (B.40). The field equations could have been derived more easily by including a gravitational term in the action

$$I_g = -\frac{1}{16\pi G}\int d^4x\sqrt{-\operatorname{Det} g(x)}\,g^{\lambda\kappa}(x)R_{\lambda\kappa}(x). \tag{B.75}$$

The Bianchi identities (B.74) can be derived from the fact that I_g is automatically invariant under general coordinate transformations of the metric, and the field equations (B.71) can be derived from the condition that the total action $I_g + I_m$ be stationary with respect to arbitrary variations of the metric.

If we allow terms in the gravitational field equation with fewer than two spacetime derivatives, then it is possible to include a term on the lefthand side of the field equation (B.71) proportional to $g_{\mu\nu}$. This is the so-called cosmological constant term, discussed in Section 1.5. It can be regarded as a "vacuum-energy" correction to $T_{\mu\nu}$. Aside from a cosmological constant, the only other possible modification of the left-hand side of Eq. (B.71) would involve terms with more than two spacetime derivatives. (They can be derived, for instance, by including terms in the integrand of the gravitational action proportional to $\sqrt{-\operatorname{Det} g}R^{\mu\nu}R_{\mu\nu}$ or $\sqrt{-\operatorname{Det} g}\,(R^{\mu}{}_{\mu})^2$.) Dimensional analysis tells us that such terms would have coefficients whose

dimensionality (relative to that of the factor $1/16\pi G$ in Eq. (B.75)) is a positive power of length. The experimental success of General Relativity shows that this length is much smaller than the scale of the solar system, so the effect of such terms would be completely negligible at cosmological distance scales.

10.13 Appendix C Energy Transfer between Radiation and Electrons

One often needs to know the rate at which a photon will lose or gain energy when passing through an ionized gas. For instance, we need this information in calculating the Sunyaev-Zel'dovich effect, discussed in Section 2.5, and in understanding the preservation of thermal equilibrium between matter and radiation, treated in Section 2.2. This appendix will first derive formulas for the mean and the mean square change in energy of a photon in striking a single electron, and will then use these results to derive the rate of change of the photon energy distribution function in passing through an ionized gas. For the purposes of this section, we take $\hbar = c = 1$.

Suppose that an electron traveling in the three-direction with momentum four-vector

$$p = (0, 0, p_e, E_e)\,, \quad E_e \equiv \sqrt{p_e^2 + m_e^2}, \tag{C.1}$$

is struck by a photon with energy ω moving along a direction with polar and azimuthal angles η and ϕ, giving the photon a new energy ω' and a new direction with polar and azimuthal angles η' and ϕ'. That is, the initial four-momentum q and final four-momentum q' of the photon take the forms

$$q = (\sin\eta\cos\phi, \sin\eta\sin\phi, \cos\eta, 1)\omega, \tag{C.2}$$
$$q' = (\sin\eta'\cos\phi', \sin\eta'\sin\phi', \cos\eta', 1)\,\omega'. \tag{C.3}$$

To calculate the cross section for this scattering event, we perform a Lorentz transformation to a frame of reference in which the electron is initially at rest, in which case the cross section takes a simple and well-known form. In the electron rest-frame, the initial and final photon four-momenta are Lq and Lq', where $L^\mu{}_\nu$ is the Lorentz transformation

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \gamma & -\beta\gamma \\ 0 & 0 & -\beta\gamma & \gamma \end{pmatrix}, \tag{C.4}$$

where $\beta \equiv p_e/E_e$ is the electron velocity, and $\gamma \equiv (1-\beta^2)^{-1/2}$. This gives the initial and final photon four-momenta in the electron rest frame as

$$Lq = (\sin\alpha\cos\phi, \sin\alpha\sin\phi, \cos\alpha, 1)k \tag{C.5}$$
$$Lq' = (\sin\alpha'\cos\phi', \sin\alpha'\sin\phi', \cos\alpha', 1)\,k' \tag{C.6}$$

where the initial and final photon energies in the electron rest frame are

$$k = (-\beta\gamma\cos\eta + \gamma)\omega, \quad k' = (-\beta\gamma\cos\eta' + \gamma)\,\omega' \tag{C.7}$$

while the initial and final polar angles of the photon's velocity in this frame are given by

$$\cos\alpha = \frac{\cos\eta - \beta}{1 - \beta\cos\eta}, \quad \cos\alpha' = \frac{\cos\eta' - \beta}{1 - \beta\cos\eta'} \tag{C.8}$$

and there is no change in the azimuthal angles. We also recall that the conservation of energy and momentum in the electron rest frame gives the final photon energy in this frame as [1]

$$k' = \frac{k}{1 + (k/m_e)\,(1 - \cos\theta)}, \tag{C.9}$$

where θ is the scattering angle in the electron rest frame

$$\cos\theta \equiv \widehat{Lq} \cdot \widehat{Lq'} = \cos\alpha\cos\alpha' + \cos(\phi - \phi')\sin\alpha\sin\alpha'. \tag{C.10}$$

The fractional change in the photon energy in the original frame of reference can then be expressed in terms of quantities in the electron rest frame by

$$\frac{\omega' - \omega}{\omega} = \frac{1}{1 + \beta\cos\alpha}\left(\frac{1 + \beta\cos\alpha'}{1 + (k/m_e)\,(1 - \cos\theta)} - \beta\cos\alpha - 1\right). \tag{C.11}$$

In most cases of interest k/m_e and β are much less than unity, so we will keep only terms of first order in k/m_e and zeroth order in β together with terms of (for the moment) arbitrary order in β and zeroth order in k/m_e :

$$\frac{\omega' - \omega}{\omega} \simeq -(k/m_e)\,(1 - \cos\theta) + \frac{\beta\,(\cos\alpha' - \cos\alpha)}{1 + \beta\cos\alpha} \tag{C.12}$$

For an electron at rest and for photon energy $k \ll m_e$, the cross-section differential (summed over final spins and polarizations, and averaged over initial spins and polarizations) is [2]

$$d\sigma = \frac{3\sigma_{\mathcal{T}}}{16\pi}\left(1 + \cos^2\theta\right) d\,(\cos\alpha')\,d\phi' \tag{C.13}$$

where $\sigma_{\mathcal{T}}$ is the Thomson cross section $e^4/6\pi m_e^2$. The cross-section differential is itself Lorentz invariant, though the variables on which it depends are not, so the average photon energy change per collision is

$$\begin{aligned}
\langle\omega' - \omega\rangle &= \frac{1}{\sigma_{\mathcal{T}}}\int(\omega' - \omega)\,d\sigma \\
&= \frac{3}{16\pi}\int_{-1}^{1} d\,(\cos\alpha')\int_0^{2\pi} d\phi'\,(\omega' - \omega)\left(1 + \cos^2\theta\right) \\
&= -\frac{\beta\omega\cos\alpha}{1 + \beta\cos\alpha} - \frac{k\omega}{m_e}
\end{aligned} \tag{C.14}$$

We must average this over the electron direction of motion, or equivalently over α. In calculating this average we must keep in mind that it is $\cos\eta$ rather than $\cos\alpha$ whose distribution function in the interval from -1 to +1 is flat. Also, the transition probability is proportional not only to $d\sigma$ but also to the relative speed [3]

$$u = \frac{|p \cdot q|}{E_e\omega} = 1 - \beta\cos\eta = \frac{1 - \beta^2}{1 + \beta\cos\alpha}. \tag{C.15}$$

[471] QTF, Vol. I, Eq. (8.7.14).
[2] QTF, Vol. I, Eq. (8.7.42).

Hence the average of the energy change over the relative direction of the initial electron and photon is

$$\begin{aligned}\langle\langle\omega'-\omega\rangle\rangle &= \int_{-1}^{+1}\langle\omega'-\omega\rangle\, ud\cos\eta \Big/ \int_{-1}^{+1} ud\cos\eta \\ &= \frac{1}{2}\int_{-1}^{+1}\frac{(1-\beta^2)^2}{(1+\beta\cos\alpha)^3}\left[-\frac{\beta\omega\cos\alpha}{1+\beta\cos\alpha}-\frac{k\omega}{m_e}\right]d\cos\alpha \end{aligned} \tag{C.16}$$

The terms of first order in β make no contribution to the integral over α, so the leading terms in the fractional energy transfer are those of second order in β or first order in ω/m_e :

$$\langle\langle\omega'-\omega\rangle\rangle/\omega \simeq \frac{4}{3}\beta^2-\frac{k}{m_e}\simeq\frac{4}{3}\beta^2-\frac{\omega}{m_e} \tag{C.17}$$

Since the mean photon fractional energy change contains terms of order β^2 and ω/m_e, we will be interested in any terms in the average squared fractional energy change of the same order. Inspection of Eq. (C.12) shows that to this order, we have

$$(\omega'-\omega)^2/\omega^2 \simeq \beta^2(\cos\alpha'-\cos\alpha)^2 \tag{C.18}$$

To this order we can neglect the difference between u and unity and between $d\cos\eta$ and $d\cos\alpha$, so that

$$\begin{aligned}\left\langle\left\langle(\omega'-\omega)^2\right\rangle\right\rangle \simeq& \frac{3\beta^2\omega^2}{32\pi}\int_{-1}^{+1}d\cos\alpha\int_{-1}^{+1}d\cos\alpha' \\ &\times\int_0^{2\pi}d\phi'\,(\cos\alpha'-\cos\alpha)^2\left(1+\cos^2\theta\right) \\ =& \frac{2}{3}\beta^2\omega^2 \end{aligned} \tag{C.19}$$

All this is for an electron of a fixed speed β. If the number of electrons with speed between β and $\beta+d\beta$ is given by the Maxwell-Boltzmann distribution with electron temperature T_e, and is hence proportional to $\beta^2\exp\left(-m_e\beta^2/2k_{\mathcal{B}}T_e\right)d\beta$, the average value of β^2 is $3k_{\mathcal{B}}T_e/m_e$, and Eqs. (C.17) and (C.19) become

$$\langle\langle\omega'-\omega\rangle\rangle \simeq \frac{4k_{\mathcal{B}}T_e}{m_e}\omega-\frac{\omega^2}{m_e}, \tag{C.20}$$

and

$$\left\langle\left\langle(\omega'-\omega)^2\right\rangle\right\rangle = \frac{2k_{\mathcal{B}}T_e}{m_e}\omega^2 \tag{C.21}$$

Now suppose that photons with a number $N(\omega)d\omega$ with energies between ω and $\omega+d\omega$ interact with a gas of non-relativistic electrons. For photon energies $\omega \ll m_e$, the rate of change of the distribution function is

$$\begin{aligned}\dot{n}(\omega)d\omega =& n_e\int d\omega' n(\omega')\,R(\omega'\to\omega)\,d\omega(1+N(\omega)) \\ &- n_e n(\omega)d\omega\int d\omega' R(\omega\to\omega')\,(1+N(\omega')) \end{aligned} \tag{C.22}$$

[473] QTF, Vol. I, Eq. (3.4.17).

Here n_e is the number density of electrons; $R(\omega \to \omega')\, d\omega'$ is the average of $ud\sigma$ over initial electron velocities for collisions in which a photon of initial energy ω is given an energy between ω' and $\omega' + d\omega'$; $N(\omega) = (2\pi\hbar)^3 n(\omega) d\omega / 8\pi\omega^2 d\omega$ is the number of photons per quantum state of energy ω (the denominator is $8\pi\omega^2 d\omega$ instead of $4\pi\omega^2 d\omega$ because of the two polarization states of photons); and the factors $1 + N(\omega)$ and $1 + N(\omega')$ are included to take account of the stimulated emission of photons into states that are already occupied. The first term in Eq. (C.22) gives the increase in $n(\omega)$ due to scattering of photons of any initial energy ω' into energy ω, while the second term gives the decrease in $n(\omega)$ due to scattering of photons of energy ω into any final energy ω'. We saw that collisions of photons of energy $\omega \ll m_e$ with a non-relativistic gas of electrons at
temperature T typically change photon energies by only small fractional amounts, of order ω/m_e or $k_{\mathcal{B}}T/m_e$, so the rate constants $R(\omega' \to \omega)$ and $R(\omega \to \omega')$ are sharply peaked around $\omega' \simeq \omega$. It is therefore convenient to change the variable of integration from ω' to $\Delta \equiv \omega - \omega'$ in the first term and to $\Delta \equiv \omega' - \omega$ in the second term of Eq. (C.22). Also canceling the unintegrated differential $d\omega$ and multiplying with $(2\pi\hbar)^3/8\pi\omega^2$, this formula now reads

$$\begin{aligned}\dot{N}(\omega) =& \frac{n_e}{\omega^2} \int_{-\infty}^{\infty} d\Delta (\omega - \Delta)^2 N(\omega - \Delta) R(\omega - \Delta \to \omega)(1 + N(\omega)) \\ & - n_e N(\omega) \int d\Delta R(\omega \to \omega + \Delta)(1 + N(\omega + \Delta)) \end{aligned} \tag{C.23}$$

Since the rate coefficients $R(\omega - \Delta \to \omega)$ and $R(\omega \to \omega + \Delta)$ are sharply peaked around $\Delta = 0$, we can expand the other factors in Eq. (C.23) (and the difference between $R(\omega - \Delta \to \omega)$ and $R(\omega \to \omega + \Delta)$) to second order in Δ :

$$\begin{aligned}\dot{N}(\omega) =& -\frac{n_e}{\omega^2} \int_{-\infty}^{\infty} d\Delta \Delta (1 + N(\omega)) \frac{\partial}{\partial \omega} \left[\omega^2 N(\omega) R(\omega \to \omega + \Delta)\right] \\ & + \frac{n_e}{2\omega^2} \int_{-\infty}^{\infty} d\Delta \Delta^2 (1 + N(\omega)) \frac{\partial^2}{\partial \omega^2} \left[\omega^2 N(\omega) R(\omega \to \omega + \Delta)\right] \\ & - n_e N(\omega) \int_{-\infty}^{\infty} d\Delta \Delta \frac{\partial N(\omega)}{\partial \omega} R(\omega \to \omega + \Delta) \\ & - \frac{n_e}{2} N(\omega) \int_{-\infty}^{\infty} d\Delta \Delta^2 \frac{\partial^2 N(\omega)}{\partial \omega^2} R(\omega \to \omega + \Delta). \end{aligned}$$

This can be rewritten as

$$\begin{aligned}\dot{N}(\omega) =& -\frac{n_e}{\omega^2} \int_{-\infty}^{\infty} d\Delta \Delta \frac{\partial}{\partial \omega} \left[\omega^2 N(\omega) R(\omega \to \omega + \Delta)(1 + N(\omega))\right] \\ & + \frac{n_e}{2\omega^2} \int_{-\infty}^{\infty} d\Delta \Delta^2 \frac{\partial^2}{\partial \omega^2} \left[\omega^2 N(\omega) R(\omega \to \omega + \Delta)(1 + N(\omega))\right] \\ & - \frac{n_e}{\omega^2} \int_{-\infty}^{\infty} d\Delta \Delta^2 \frac{\partial}{\partial \omega} \left[\omega^2 N(\omega) R(\omega \to \omega + \Delta) \frac{\partial N(\omega)}{\partial \omega}\right]. \end{aligned}$$

Inverting the order of integration and differentiation, this becomes

$$\begin{aligned}\dot{N}(\omega) =& -\frac{n_e \sigma_{\mathcal{T}}}{\omega^2} \left[\frac{\partial}{\partial \omega} \left(\omega^2 N(\omega)(1 + N(\omega)) \left\langle\left\langle \omega' - \omega \right\rangle\right\rangle\right) \right. \\ & - \frac{1}{2} \frac{\partial^2}{\partial \omega^2} \left(\omega^2 N(\omega)(1 + N(\omega)) \left\langle\left\langle (\omega' - \omega)^2 \right\rangle\right\rangle\right) \\ & \left. + \frac{\partial}{\partial \omega} \left(\omega^2 N(\omega) \left\langle\left\langle (\omega' - \omega)^2 \right\rangle\right\rangle \frac{\partial N(\omega)}{\partial \omega}\right)\right], \end{aligned} \tag{C.24}$$

in which we use

$$\int_{-\infty}^{\infty} d\Delta \Delta R(\omega \to \omega + \Delta) = \sigma_{\mathcal{T}} \langle\langle \omega' - \omega \rangle\rangle$$
$$\times \int_{-\infty}^{\infty} d\Delta \Delta^2 R(\omega \to \omega + \Delta) = \sigma_{\mathcal{T}} \left\langle \left\langle (\omega' - \omega)^2 \right\rangle \right\rangle \tag{C.25}$$

Inserting the values (C.20) and (C.21) of $\langle\langle \omega' - \omega \rangle\rangle$ and $\left\langle \left\langle (\omega' - \omega)^2 \right\rangle \right\rangle$ gives the Kompaneets equation: [4]

$$\dot{N}(\omega) = \frac{n_e \sigma_{\mathcal{T}} k_{\mathcal{B}} T_e}{m_e \omega^2} \frac{\partial}{\partial \omega} \left[\omega^4 \frac{\partial N(\omega)}{\partial \omega} \right] + \frac{n_e \sigma_{\mathcal{T}}}{m_e \omega^2} \frac{\partial}{\partial \omega} \left[\omega^4 N(\omega)(1 + N(\omega)) \right]. \tag{C.26}$$

As a check, we may note that if the photon distribution function is given by the Planck formula with photon temperature equal to the electron temperature, so that $N(\omega) = [\exp(\omega/k_{\mathcal{B}} T_e) - 1]^{-1}$, then the two terms in Eq. (C.26) cancel, giving no change in $N(\omega)$, as of course must be the case for photons in equilibrium with electrons. As a further check, we note that (as long as $\omega^4 N(\omega)$ and $\omega^4 \partial N/\partial \omega$ both vanish at $\omega = 0$ and $\omega \to \infty$) the total number density of photons $\int_0^\infty 4\pi\omega^2 N(\omega) d\omega$ does not change with time, as could also be seen by integrating Eq. (C.22) over ω, and interchanging ω and ω' in the second term.

10.14 Appendix D The Ergodic Theorem

In cosmology we often have to deal with position-dependent variables like temperature or density or scalar fields whose fluctuations are governed by some sort of probability distribution. In this appendix we will consider a general real random variable $\varphi(x)$ depending on a D-dimensional Euclidean coordinate x. The generalization to several random variables will be obvious.

We will assume that the distribution function giving the probabilities of various functional forms for $\varphi(x)$ is homogeneous, in the sense that the average of any product of φ s with different arguments depends only on the differences of the arguments. That is, for arbitrary z,

$$\langle \varphi(x_1) \varphi(x_2) \cdots \varphi(x_n) \rangle = \langle \varphi(x_1 + z) \varphi(x_2 + z) \cdots \varphi(x_n + z) \rangle. \tag{D.1}$$

In cases where x is a time coordinate, such distributions are said to be stationary.

We will further assume that the φ s at distant arguments are uncorrelated. To put this formally, we assume that for $|u| \to \infty$,

$$\begin{aligned} &\langle \varphi(x_1 + u) \varphi(x_2 + u) \cdots \varphi(y_1 - u) \varphi(y_2 - u) \cdots \rangle \\ &\quad \to \langle \varphi(x_1 + u) \varphi(x_2 + u) \cdots \rangle \langle \varphi(y_1 - u) \varphi(y_2 - u) \cdots \rangle \\ &\quad = \langle \varphi(x_1) \varphi(x_2) \cdots \rangle \langle \varphi(y_1) \varphi(y_2) \cdots \rangle \end{aligned} \tag{D.2}$$

with the final expression justified by Eq. (D.1).

Under these conditions, we can prove an important result known as the Ergodic Theorem: If the limit in Eq. (D.2) is approached sufficiently rapidly, then the root mean square difference between any product $\varphi(x_1 + z) \varphi(x_2 + z) \cdots$, averaged over a range R of z values around an arbitrary point z_0, and the ensemble average of the same product, vanishes as $R^{-D/2}$ for large R. That is, if we define

[474] A. Kompaneets, Zh. Exper. Teor. Fiziki 312+, 876 (1956).

$$\Delta_R^2(x_1, x_2, \ldots) \equiv \left\langle \left(\int d^D z N_R(z) \varphi(x_1 + z)\, \varphi(x_2 + z) \cdots - \langle \varphi(x_1)\, \varphi(x_2) \cdots \rangle \right)^2 \right\rangle \tag{D.3}$$

where [1]

$$N_R(z) \equiv (\sqrt{\pi} R)^{-D} \exp\left(- |z - z_0|^2 / R^2 \right), \tag{D.4}$$

then for $R \to \infty$,

$$\Delta_R \to O\left(R^{-D/2} \right). \tag{D.5}$$

To prove this theorem, we first use the condition $\int N_R(z) d^D z = 1$ to rewrite Eq. (D.3) as

$$\Delta_R^2(x_1, x_2, \ldots) = \left\langle \left(\int d^D z N_R(z) \left[\varphi(x_1 + z)\, \varphi(x_2 + z) \cdots - \langle \varphi(x_1)\, \varphi(x_2) \cdots \rangle \right] \right)^2 \right\rangle$$

Expanding the square and again using Eq. (D.1) and the normalization condition for N_R, and introducing new integration variables $u \equiv (z - w)/2$ and $v \equiv (z + w)/2$, we have

$$\begin{aligned}
\Delta_R^2 &= \int d^D z N_R(z) \int d^D w N_R(w) \\
&\quad \times \left[\langle \varphi(x_1 + z)\, \varphi(x_2 + z) \cdots \varphi(x_1 + w)\, \varphi(x_2 + w) \cdots \rangle \right. \\
&\quad \left. - \langle \varphi(x_1)\, \varphi(x_2) \cdots \rangle^2 \right] \\
&= \left(\frac{2}{\pi R^2} \right)^D \int d^D v \exp\left(-2 |v - z_0|^2 / R^2 \right) \int d^D u \exp\left(-2|u|^2 / R^2 \right) \\
&\quad \times \left[\langle \varphi(x_1 + u)\, \varphi(x_2 + u) \cdots \varphi(x_1 - u)\, \varphi(x_2 - u) \cdots \rangle \right. \\
&\quad \left. - \langle \varphi(x_1)\, \varphi(x_2) \cdots \rangle^2 \right] \\
&= \left(\frac{2}{\pi R^2} \right)^{D/2} \int d^D u \exp\left(-2|u|^2 / R^2 \right) \\
&\quad \times \left[\langle \varphi(x_1 + u)\, \varphi(x_2 + u) \cdots \varphi(x_1 - u)\, \varphi(x_2 - u) \cdots \rangle \right. \\
&\quad \left. - \langle \varphi(x_1)\, \varphi(x_2) \cdots \rangle^2 \right].
\end{aligned}$$

Assuming that the limit in Eq. (D.2) is approached sufficiently rapidly, the u integral would converge even without the factor $\exp(-2|u|^2/R^2)$, so for $R \to \infty$ we can take this factor as unity, and find our final expression

$$\begin{aligned}
\Delta_R^2 \to \left(\frac{2}{\pi R^2} \right)^{D/2} & \int d^D u \\
& \times \left[\langle \varphi(x_1 + u)\, \varphi(x_2 + u) \cdots \varphi(x_1 - u)\, \varphi(x_2 - u) \cdots \rangle \right. \\
& \left. - \langle \varphi(x_1)\, \varphi(x_2) \cdots \rangle^2 \right]
\end{aligned} \tag{D.6}$$

thus confirming Eq. (D.5).

[471] The specific form (D.4) for the function N_R is not essential. It is only important that the function be constant for $|z - z_0|^2 \ll R^2$, vanish rapidly for $|z - z_0|^2 \gg R^2$, and be normalized so that $\int N_R(z) d^D z = 1$.

10.15 Appendix E Gaussian Distributions

Consider a random variable $\varphi(x)$ depending on a D-dimensional coordinate x. We will define it to have zero average value, by subtracting from φ any non-zero average it may have. The distribution function governing φ is said to be Gaussian if the average of a product of an even number of φ s is the sum over all ways of pairing φ s with each other of a product of the average values of the pairs:

$$\langle \varphi(x_1)\,\varphi(x_2)\cdots\rangle = \sum_{\text{pairings}}\prod_{\text{pairs}}\langle\varphi\varphi\rangle, \tag{E.1}$$

(with the sum over pairings not distinguishing those which interchange coordinates in a pair, or which merely interchange pairs), while the average of the product of any odd number of φ s vanishes. For instance

$$\begin{aligned}\langle \varphi(x_1)\,\varphi(x_2)\,\varphi(x_3)\,\varphi(x_4)\rangle &= \langle \varphi(x_1)\,\varphi(x_2)\rangle\,\langle \varphi(x_3)\,\varphi(x_4)\rangle \\ &+ \langle \varphi(x_1)\,\varphi(x_3)\rangle\,\langle \varphi(x_2)\,\varphi(x_4)\rangle + \langle \varphi(x_1)\,\varphi(x_4)\rangle\,\langle \varphi(x_2)\,\varphi(x_3)\rangle\,.\end{aligned} \tag{E.2}$$

(Of course, there is no way of pairing all of an odd number of φ s, which is why for Gaussian distributions the average of any odd number of φ s must vanish.) More generally, in the average of a product of $2n$ factors of φ, each of $(\ 2n\ )!$ permutations of the coordinates defines a pairing (with the first permuted φ paired with the second, the third with the fourth, and so on), but $2^n n!$ of these differ only by permutations of the two coordinates within a pair or permutations of the n pairs, so in general this average contains $(\ 2n\ )!/2^n n!$ terms.

If $\varphi(x)$ is governed by a Gaussian distribution, then so is any linear functional of $\varphi(x)$. For instance, consider a set of linear functionals of the form

$$F_i[\varphi] \equiv \int d^D x f_i(x)\varphi(x) \tag{E.3}$$

By multiplying Eq. (E.2) by $f_1(x_1)\,f_2(x_2)\,f_3(x_3)\,f_4(x_4)$ and integrating, we find

$$\langle F_1F_2F_3F_4\rangle = \langle F_1F_2\rangle\,\langle F_3F_4\rangle + \langle F_1F_3\rangle\,\langle F_2F_4\rangle + \langle F_1F_4\rangle\,\langle F_2F_3\rangle\,. \tag{E.4}$$

Such distributions are called Gaussian because of the form taken by the probability distribution of general linear functionals $F[\varphi] = \int d^D x f(x)\varphi(x)$ of φ. In $\langle F^{2n}[\varphi]\rangle$, each of the $(\ 2n\ )!\ /2^n n\ !$ terms makes the same contribution
$\langle F^2[\varphi]\rangle^n$, so

$$\left\langle F^{2n}[\varphi]\right\rangle = \left\langle F^2[\varphi]\right\rangle^n (2n)!/2^n n!. \tag{E.5}$$

This implies that the probability $P(F)dF$ that the functional is between F and $F+dF$ is the Gaussian function

$$P(F)dF = \frac{1}{\sqrt{2\pi\,\langle F^2\rangle}}\exp\left(-\frac{F^2}{2\,\langle F^2\rangle}\right)dF. \tag{E.6}$$

Distributions of this sort arise commonly when $\varphi(x)$ is the sum of a large number of independently fluctuating terms. The central limit theorem tells us that in this case the distribution of the sum is Gaussian. In the application that most concerns us here, the fluctuations in the temperature of the cosmic microwave background are believed to be nearly Gaussian because they arise (as discussed in Chapter 10) from the quantum fluctuations of one or more nearly free quantum fields.

10.16 Appendix F Newtonian Cosmology

During the era when the energy density of the universe is dominated by cold dark matter, the behavior of perturbations can be adequately treated by the methods of Newtonian mechanics. This has applications both for our introductory study of anisotropies in the cosmic microwave background in Section 2.6, and for the treatment of the large scale structure of matter in Section 8.1.

The equations of non-relativistic hydrodynamics and Newtonian gravitational theory for a fluid with mass density ρ, velocity $\mathbf{v}$, zero pressure, and gravitational potential ϕ are the equation of continuity

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\mathbf{v}\rho) = 0 \tag{F.1}$$

the Euler equation

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\nabla\phi \tag{F.2}$$

and the Poisson equation

$$\nabla^2 \phi = 4\pi G\rho. \tag{F.3}$$

These equations have an unperturbed solution (distinguished by bars) of the form

$$\bar{\rho} = \rho_0 \left(a_0/a\right)^3, \quad \overline{\mathbf{v}} = H\mathbf{X}, \quad \bar{\phi} = 2\pi G\bar{\rho}\mathbf{X}^2/3 \tag{F.4}$$

where $H \equiv \dot{a}/a; a_0$ and ρ_0 are constants; and $a(t)$ satisfies the equation

$$\dot{a}^2 + K = 8\pi G\bar{\rho}a^2/3, \tag{F.5}$$

with K a constant. (We use $\mathbf{X}$ to denote the ordinary Euclidean coordinate vector, to distinguish it from the co-moving coordinate vector $\mathbf{x}$.) This of course corresponds to a cosmological theory with a RobertsonWalker metric, and indeed we have already encountered this solution in Section 1.5, as an alternative approach to the derivation of the Friedmann equation (F.5). In particular, the solution of the equation $d\mathbf{X}/dt = \overline{\mathbf{v}} = H\mathbf{X}$ for the motion of a co-moving object is $\mathbf{X}(t) = [a(t)/a\,(t_0)]\,\mathbf{X}\,(t_0)$, in agreement with Eq. (1.5.22). The co-moving coordinate is thus $\mathbf{x} = \mathbf{X}/a$.

To this unperturbed solution for $\rho, \mathbf{v}$, and ϕ we now add small perturbations $\delta\rho, \delta\mathbf{v}$, and $\delta\phi$. The terms in Eqs. (F.1)-(F.3) of first
order in these perturbations are

$$\frac{\partial \delta\rho}{\partial t} + 3H\delta\rho + H\mathbf{X} \cdot \nabla\delta\rho + \bar{\rho}\nabla \cdot \delta\mathbf{v} = 0 \tag{F.6}$$

$$\frac{\partial \delta\mathbf{v}}{\partial t} + H\mathbf{X} \cdot \nabla\delta\mathbf{v} + H\delta\mathbf{v} = -\nabla\delta\phi \tag{F.7}$$

$$\nabla^2 \delta\phi = 4\pi G\delta\rho \tag{F.8}$$

Eqs. (F.6)-(F.8) do not appear translation-invariant, but the underlying spatial homogeneity of this problem can be restored by writing the equations in terms of the co-moving coordinate $\mathbf{X}/a$. This is brought out most conveniently if we write the perturbations as Fourier transforms in the co-moving coordinate:

$$\delta\rho(\mathbf{X}, t) = \int d^3q \exp\left(\frac{i\mathbf{q} \cdot \mathbf{X}}{a(t)}\right) \delta\rho_{\mathbf{q}}(t), \tag{F.9}$$

and likewise for $\delta\mathbf{v}$ and $\delta\phi$. The partial differential equations (F.6)-(F.8) then become the ordinary differential equations

$$\frac{d\delta\rho_{\mathbf{q}}}{dt} + 3H\delta\rho_{\mathbf{q}} + ia^{-1}\bar{\rho}\mathbf{q}\cdot\delta\mathbf{v}_{\mathbf{q}} = 0 \tag{F.10}$$

$$\frac{d\delta\mathbf{v}_{\mathbf{q}}}{dt} + H\delta\mathbf{v}_{\mathbf{q}} = -ia^{-1}\mathbf{q}\delta\phi_{\mathbf{q}} \tag{F.11}$$

$$\mathbf{q}^2\delta\phi_{\mathbf{q}} = -4\pi G a^2\delta\rho_{\mathbf{q}} \tag{F.12}$$

The solutions of Eqs. (F.10)-(F.12) can be classified according to the transformation properties of the dependent variables under three-dimensional rotations:

Vector modes: In these modes, all scalars vanish: not only $\delta\rho_{\mathbf{q}}$ and $\delta\phi_{\mathbf{q}}$, but also $\mathbf{q}\cdot\delta\mathbf{v}_{\mathbf{q}}$. Then Eqs. (F.10) and (F.12) are automatically satisfied, while Eq. (F.11) becomes

$$\frac{d\delta\mathbf{v}_{\mathbf{q}}}{dt} + H\delta\mathbf{v}_{\mathbf{q}} = 0, \tag{F.13}$$

whose solution is simply $\delta\mathbf{v}_{\mathbf{q}} \propto 1/a$. Because the vector modes simply decay, they are widely ignored.

Scalar Modes: In these modes, the velocity perturbation $\delta\mathbf{v}(\mathbf{X},t)$ can be expressed as the gradient (with respect to $\mathbf{x} = \mathbf{X}/a$) of a scalar potential perturbation $\delta u(\mathbf{X},t)$. For the Fourier transforms, this means that

$$\delta\mathbf{v}_{\mathbf{q}} = i\mathbf{q}\delta u_{\mathbf{q}}. \tag{F.14}$$

Then, using Eq. (F.12) to eliminate $\delta\phi_{\mathbf{q}}$, Eq. (F.11) becomes

$$\frac{d\delta u_{\mathbf{q}}}{dt} + H\delta u_{\mathbf{q}} = -a^{-1}\delta\phi_{\mathbf{q}} = \frac{4\pi G a}{\mathbf{q}^2}\delta\rho_{\mathbf{q}}, \tag{F.15}$$

while Eq. (F.10) gives

$$\frac{d\delta\rho_{\mathbf{q}}}{dt} + 3H\delta\rho_{\mathbf{q}} - a^{-1}\bar{\rho}\mathbf{q}^2\delta u_{\mathbf{q}} = 0. \tag{F.16}$$

For $K = 0$ we have $4\pi G\bar{\rho} = 3H^2/2$. Using $\bar{\rho} \propto a^{-3}$ and the definition of H, we can eliminate $\delta u_{\mathbf{q}}$ from these coupled equations and write them as a second-order differential equation for the fractional density perturbation

$$\frac{d}{dt}\left(a^2\frac{d}{dt}\left(\frac{\delta\rho_{\mathbf{q}}}{\bar{\rho}}\right)\right) - \frac{3}{2}H^2a^2\left(\frac{\delta\rho_{\mathbf{q}}}{\bar{\rho}}\right) = 0, \tag{F.17}$$

or, recalling that for $K = 0$ we have $a \propto t^{2/3}$ and $H = 2/3t$,

$$\frac{d}{dt}\left(t^{4/3}\frac{d}{dt}\left(\frac{\delta\rho_{\mathbf{q}}}{\bar{\rho}}\right)\right) - \frac{2}{3}t^{-2/3}\left(\frac{\delta\rho_{\mathbf{q}}}{\bar{\rho}}\right) = 0. \tag{F.18}$$

The general solution of this equation is a linear combination of the powers $t^{2/3}$ and t^{-1}. It is reasonable to suppose that by the time of last scattering only the leading mode with $\delta\rho_{\mathbf{q}}/\bar{\rho} \propto t^{2/3} \propto a$ will have survived. Since the mean density $\bar{\rho}(t)$ is proportional to $a^{-3}(t)$, we have $\delta\rho_{\mathbf{q}}(t) \propto a^{-2}(t)$. Eq. (F.12) then shows that $\delta\phi_{\mathbf{q}}$ is time independent, a result used in Section 2.6.

To verify that these are the most general solutions of Eqs. (F.10)-(F.12), we need only count equations and solutions. With $\delta\phi_{\mathbf{q}}(t)$ eliminated by use of Eq. (F.12), Eqs. (F.10) and (F.11) are a set of $1+3=4$ coupled first-order differential equations, so they have four linearly

independent solutions. We have found two independent vector modes (corresponding to the two directions perpendicular to $\boldsymbol{q}$) and two independent scalar modes (with $\delta\rho_{\mathbf{q}}/\bar{\rho}$ proportional to $t^{2/3}$ or t^{-1}), so these are indeed the most general solutions.

The cases of non-zero pressure or $K \neq 0$ are treated in detail in Section 15.9 of G&C.

10.17 Appendix GPhoton Polarization

The polarization of photons is of great interest in cosmology, not only because it can be observed in the cosmic microwave background, but also because it affects the anisotropic inertia terms appearing in the gravitational field equations for both scalar and tensor perturbations. This appendix will review the description of the spin state of an individual photon in terms of photon polarization vectors and the related density matrix, and the somewhat unusual parallel transport equation that governs the time dependence of photon polarization vectors and density matrices in gravitational fields.

Let us first recall how we describe the polarization of a photon in the absence of gravitational fields. The most general pure state of a single photon is a linear superposition $\alpha_+\Psi_+ + \alpha_-\Psi_-$, where $\Psi_\pm$are states of helicity ∓ 1 (that is, eigenstates of the component of angular momentum in the direction $\hat{n}$ from which the photon is coming [1] with eigenvalues $\sigma = \pm\hbar$, and normalization $(\Psi_\sigma, \Psi_{\sigma'}) = \delta_{\sigma\sigma'}$) and $\alpha_\pm$are complex numbers satisfying the normalization condition $|\alpha_+|^2 + |\alpha_-|^2 = 1$. We represent such a state by a polarization vector $\mathbf{e} = \alpha_+\mathbf{e}_+ + \alpha_-\mathbf{e}_-$with $\mathbf{e}^* \cdot \mathbf{e} = 1$, where $\mathbf{e}_\pm$is the polarization vector for photons of helicity ∓ 1. For instance, for a photon that is seen coming from the direction [2]

$$\hat{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \tag{G.1}$$

we can take

$$\mathbf{e}_\pm = (\hat{\theta} \pm i\hat{\phi})/\sqrt{2}, \tag{G.2}$$

where $\hat{\theta}$ and $\hat{\phi}$ are orthogonal unit vectors in the plane perpendicular to $\hat{n}$:

$$\hat{\theta} \equiv (\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta), \hat{\phi} \equiv (-\sin\phi, \cos\phi, 0). \tag{G.3}$$

A linearly polarized photon has $|\alpha_+| = |\alpha_-|$, and hence a polarization vector that is real up to an unimportant over-all phase factor. The opposite case is circular polarization, for which either $|\alpha_+|$or $|\alpha_-|$vanishes. Between these

two extremes is elliptic polarization, for which $|\alpha_+|$and $|\alpha_-|$are unequal, but neither vanishes.

An individual photon will in general have probabilities P_m of being in any one of various pure states $\Psi_m = \alpha_{m+}\Psi_+ + \alpha_{m-}\Psi_-$, so according to the rules of quantum mechanics, the probability of finding it in a particular state $\Psi = \alpha_+\Psi_+ + \alpha_-\Psi_-$represented by a polarization vector $\mathbf{e} = \alpha_+\mathbf{e}_+ + \alpha_-\mathbf{e}_-$, will be

$$P(\mathbf{e}) = \sum_m P_m \left|(\Psi_m, \Psi)\right|^2 = \sum_m P_m \left|\alpha^*_{m+}\alpha_+ + \alpha^*_{m-}\alpha_-\right|^2 = \mathbf{e}^{*i}\mathbf{e}^j N^{ij} \tag{G.4}$$

where N^{ij} is the density matrix for that photon

[471] This is the opposite of the direction of the photon's motion, so the helicity, which is defined as the component of angular momentum along the direction of the photon's motion, is the negative of the component along the direction from which it is coming. This is why we use a label $\pm$ to indicate photons of helicity ∓ 1.

[2] For correlations between polarizations in a small patch of sky, one can choose the three-axis to be in the direction $\hat{n}$ of this patch, in which case the polarization vectors for helicity ∓ 1 could be chosen simply as $(1, \pm i, 0)/\sqrt{2}$.

$$N^{ij} \equiv \sum_m P_m e^i_m e^{j*}_m \tag{G.5}$$

(Repeated indices i, j, etc. are summed over the coordinate indices 1,2,3.) Because the probabilities are real and positive, the matrix N^{ij} is hermitian and positive:

$$N^{ij*} = N^{ji}, \quad N^{ij}\xi^i\xi^{j*} \geq 0 \text{ for all } \xi^i, \tag{G.6}$$

and because $\sum_m P_m = 1$ and $e^i_m e^{i*}_m = 1$ (not summed over m), this matrix has unit trace

$$N^{ii} = 1. \tag{G.7}$$

Of course, the photon polarization vectors are all orthogonal to the photon's direction of motion $-\hat{n}$, so also

$$N^{ij}\hat{n}^i = N^{ij}\hat{n}^j = 0, \tag{G.8}$$

The scattering of light by non-relativistic electrons does not produce circular polarization, and therefore we expect that all microwave background photons are linearly polarized, in which case N^{ij} is real.

So far, we have defined the polarization vector e^i in the absence of gravitation as a unit three-vector, transverse to the photon's direction of motion. We can if we like define it in a gravitational field as the spatial part of a four-component object e^μ that happens to have $e^0 = 0$, and that satisfies

$$p_i e^i = p_\mu e^\mu = 0, \quad g_{ij}e^i e^{j*} = g_{\mu\nu}e^\mu e^{\nu*} = 1 \tag{G.9}$$

where $p^\mu = g^{\mu\nu}p_\nu$ is the photon four-momentum. [3] But since this object obviously does not transform as a four-vector there is no reason to think that its time dependence would be given by the equation (B.69) of parallel transport. Indeed, it cannot evolve by parallel transport, for if it did then the condition $e^0 = 0$ would not be preserved along the photon trajectory $x^i = x^i(t)$ unless it so happened that $\Gamma^0_{ij}(\mathbf{x}(t), t)$ vanishes, which is not usually the case. Instead, we expect that in a locally inertial frame of reference in which the affine connection $\Gamma^\mu_{\nu\lambda}(x(t))$ vanishes, the polarization vector of a photon will be time-independent only up to a gauge transformation $e^\mu \to e^\mu + \alpha p^\mu$, so that gauge-invariant quantities like the field strength $p^\mu e^\nu - p^\nu e^\mu$ will be time-independent. Then in a general frame of reference, a photon polarization vector will undergo parallel transport up to a gauge transformation:

$$\frac{de^\mu(t)}{dt} = -\Gamma^\mu_{v\lambda}(x(t))e^v(t)\frac{dx^\lambda(t)}{dt} + \alpha(t)\frac{dx^\mu(t)}{dt} \tag{G.10}$$

The gauge transformation parameter $\alpha(t)$ can then be determined from the condition that $e^0 = 0$, so that $de^0/dt = 0$:

$$\alpha(t) = \Gamma^0_{\nu\lambda}(x(t))e^\nu(t)\frac{dx^\lambda(t)}{dt} \tag{G.11}$$

With $e^0 = 0$, Eq. (G.10) then reads

$$\frac{de^j(t)}{dt} = \left[-\Gamma^j_{i\lambda}(x(t)) + \Gamma^0_{i\lambda}(x(t))\frac{dx^j(t)}{dt}\right] e^i(t)\frac{dx^\lambda(t)}{dt} \tag{G.12}$$

Note that the conditions (G.9) are preserved with time, for $p_\mu e^\mu$ and $g_{\mu\nu}e^\mu e^\nu$ are preserved by ordinary parallel transport because they are scalars, and are preserved under gauge transformations because $p_\mu p^\mu = 0$ and $p_\mu e^\mu = 0$.

In a gravitational field we continue to define the statistical matrix N^{ij} by Eq. (G.5), and it continues to satisfy the hermiticity and positivity conditions (G.6), but now instead of (G.7) and (G.8) we have

$$g_{ij}N^{ij} = 1, \quad p_i N^{ij} = p_j N^{ij} = 0 \tag{G.13}$$

It follows from Eq. (G.12) that the statistical matrix of a photon moving in a gravitational field satisfies the parallel transport equation

$$\begin{aligned}\frac{dN^{ij}(t)}{dt} =& \left[-\Gamma^i_{k\lambda}(x(t)) + \Gamma^0_{k\lambda}(x(t))\frac{dx^i(t)}{dt}\right] N^{kj}(t)\frac{dx^\lambda(t)}{dt} \\ &+ \left[-\Gamma^j_{k\lambda}(x(t)) + \Gamma^0_{k\lambda}(x(t))\frac{dx^j(t)}{dt}\right] N^{ik}(t)\frac{dx^\lambda(t)}{dt}\end{aligned} \tag{G.14}$$

10.18 Appendix H The Relativistic Boltzmann Equation

In this appendix we derive the Boltzmann equations for neutrinos and photons, that govern the evolution of the distribution of these relativistic particles in phase space. This will provide the basis for our account in Section 6.1 of the widely used numerical calculations of the evolution of scalar perturbations. In addition, these results will be needed in Section 6.6, where we evaluate the damping of tensor modes due to the anisotropic inertia of free-streaming neutrinos, and in Section 7.4, where we calculate the polarization of the microwave background produced by the scattering of an anisotropic distribution of photons by non-relativistic electrons. This appendix will first consider the simpler case of neutrinos, for which scattering may be neglected, and then turn to the more complicated case of photons, for which scattering plays an essential role. In an elementary application of the results for photons, we will derive a formula for the rate of damping of acoustic waves in a medium of photons and charged particles, used in Section 6.4.

Throughout this appendix we shall adopt a coordinate system for which

$$g_{00} = -1, \quad g_{i0} = 0 \tag{H.1}$$

while $g_{ij}(\mathbf{x}, t)$ is unconstrained. In the linear approximation, this form of the metric is automatic for tensor modes, and follows if we work in synchronous gauge for scalar modes. However, we will not specialize to the case of linear perturbations until later. With this metric, the only non-vanishing components of the affine connection are

$$\Gamma^k_{ij} = \frac{1}{2}g^{kl}\left(\frac{\partial g_{li}}{\partial x^j} + \frac{\partial g_{lj}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^l}\right), \tag{H.2}$$

$$\Gamma^j_{i0} = \frac{1}{2}g^{jk}\dot{g}_{ki}, \quad \Gamma^0_{ij} = \frac{1}{2}\dot{g}_{ij}. \tag{H.3}$$

[473] We can construct e^μ in a gravitational field in a general coordinate system x^μ from the polarization vector $e^i_{\rm flat}$ of the photon in a locally inertial frame with coordinates ξ^α, as

$$e^\mu = \frac{\partial x^\mu}{\partial \xi^i} e^i_{\rm flat} + \beta p^\mu$$

with β adjusted to make $e^0 = 0$.

1 Neutrinos

It will be convenient to work with a neutrino distribution function $n_\nu(\mathbf{x}, \mathbf{p}, t)$, defined by

$$n_v(\mathbf{x}, \mathbf{p}, t) \equiv \sum_r \left(\prod_{i=1}^{3} \delta\left(x^i - x_r^i(t)\right) \right) \left(\prod_{i=1}^{3} \delta\left(p_i - p_{ri}(t)\right) \right), \tag{H.4}$$

with r labeling trajectories of individual neutrinos (or antineutrinos). This expression has the defining property of a number density, that the integral
of n_ν over any volume of phase space equals the number of neutrinos in that volume. According to Section 4 of Appendix B, the momentum variable in Eq. (H.4) is defined by $p_{ri} = g_{ij}p_r^j$, where p^μ is the momentum four-vector

$$p_r^\mu \equiv dx_r^\mu / du_r \tag{H.5}$$

where u_r is a suitably normalized affine parameter for which the spacetime trajectory satisfies Eq. (B.12):

$$\frac{d^2 x_r^\lambda}{du_r^2} + \Gamma_{\mu\nu}^\lambda (x_r) \frac{dx_r^\mu}{du_r} \frac{dx_r^\nu}{du_r} = 0 \tag{H.6}$$

Then between collisions the rate of change of the momentum is simply given by

$$\dot{p}_{ri} = \frac{1}{2p_r^0} p_r^j p_r^k \left(\frac{\partial g_{jk}}{\partial x^i} \right)_{\mathbf{x}=\mathbf{x}_r} \tag{H.7}$$

while the rate of change of the coordinate is

$$\dot{x}_r^i = p_r^i / p_r^0 \tag{H.8}$$

It follows then directly from Eqs. (H.4), (H.7), and (H.8) that in the absence of collisions, n_ν satisfies a Boltzmann equation

$$\frac{\partial n_v}{\partial t} + \frac{\partial n_v}{\partial x^i} \frac{p^i}{p^0} + \frac{\partial n_v}{\partial p_i} \frac{p^j p^k}{2p^0} \frac{\partial g_{jk}}{\partial x^i} = 0. \tag{H.9}$$

It should be understood that p^i and p^0 are expressed here in terms of the independent variable p_i by $p^i = g^{ij}(\mathbf{x}, t)p_j$ and $p^0 = \left(g^{ij}(\mathbf{x}, t)p_i p_j\right)^{1/2}$, so they depend on position and time as well as on p_i.

We now specialize to the case of a small perturbation. The spatial metric is then of the form

$$g_{ij}(\mathbf{x}, t) = a^2(t)\delta_{ij} + \delta g_{ij}(\mathbf{x}, t) \tag{H.10}$$

with δg_{ij} small. With only a small perturbation to the metric, the neutrino distribution function never gets very different from its equilibrium form, so we write

$$n_v(\mathbf{x}, t) = \bar{n}_v \left(a(t) \sqrt{g^{ij}(\mathbf{x}, t) p_i p_j} \right) + \delta n_v(\mathbf{x}, t), \tag{H.11}$$

where

$$\bar{n}_v(p) \equiv \frac{1}{(2\pi)^3} \left[\exp\left(p / k_{\mathcal{B}} a(t) \bar{T}(t) \right) + 1 \right]^{-1} \tag{H.12}$$

and δn_ν is a small perturbation. (Note that the factor $a(t)$ in the argument of $\bar{n}$ in Eq. (H.11) is canceled by the factor $a(t)$ multiplying $\bar{T}(t)$ in Eq. (H.12); this factor has been

included in Eqs. (H.11) and (H.12) because at times of interest, when $\bar{T} < 10^{10}$ K, we have $\bar{T}(t) \propto 1/a(t)$, so that Eq. (H.12) defines $\bar{n}_\nu$ as a time-independent function of its argument.) The first term in Eq. (H.11) is the neutrino distribution we would expect according to the Principle of Equivalence in a perturbed gravitational field, if the distribution in locally inertial frames were just the equilibrium distribution $\bar{n}_\nu(p)$; the second term δn_ν thus represents the departure of the neutrino distribution from its equilibrium form.

We can derive an initial condition for δn_ν by noting that, at a time t_1 corresponding to a temperature $\bar{T}_1 \approx 10^9$ K, the neutrino scattering rate had already dropped well below the expansion rate, but there had not yet been time for the perturbations in the metric to distort the neutrino distribution away from local thermal equilibrium, but with a perturbed temperature. Thus as a convenient initial condition we may take

$$\delta n_\nu\left(\mathbf{x}, \mathbf{p}, t_1\right) = -p\bar{n}'_\nu(p)\left[\frac{\delta T\left(\mathbf{x}, t_1\right)}{\bar{T}\left(t_1\right)} + \frac{\hat{p}_k \cdot \delta u_k\left(\mathbf{x}, t_1\right)}{a\left(t_1\right)}\right] \tag{H.13}$$

where $\bar{T}\left(t_1\right) \approx 10^9$ K. The second term in square brackets represents the Doppler shift due to a possible neutrino streaming velocity $\delta\mathbf{u}$, analogous to that given for microwave background photons by Eqs. (2.4.5) and (2.4.6). In local thermal equilibrium this is the same velocity perturbation as for baryons and photons. (Note that the metric position vector is ax^k, so the metric velocity vector is $au^k = a^{-1}u_k$.)

To first order in metric and density perturbations, Eq. (H.9) reads

$$\begin{aligned} 0 =& \frac{\partial \delta n_v(\mathbf{x}, \mathbf{p}, t)}{\partial t} + \frac{\partial \delta n_v(\mathbf{x}, \mathbf{p}, t)}{\partial x^i}\frac{p_i}{a(t)p} \\ &+ \frac{\bar{n}'_v(p)}{2p}\frac{\partial}{\partial t}\left(a^2(t)\delta g^{ij}(\mathbf{x}, t)\right) p_i p_j \\ &+ \frac{a(t)\bar{n}'_v(p)}{2p^2}\frac{\partial \delta g^{ij}(\mathbf{x}, t)}{\partial x^k} p_i p_j p_k \\ &+ \bar{n}'_v(p)\frac{p_i p_j p_k}{2a^3(t)p^2}\frac{\partial \delta g_{ij}(\mathbf{x}, t)}{\partial x^k} \end{aligned}$$

where here $p \equiv \sqrt{p_i p_i}$. (The third from last and penultimate terms on the right-hand side arise from the dependence of the argument of $\bar{n}_\nu$ in Eq. (H.11) on t and $\mathbf{x}$ through the combination $a^2(t)g^{ij}\left(\mathbf{x}, t_1\right) p_i p_j$.) To first order, we have $g^{ij} = a^{-2}\delta_{ij} - a^{-4}\delta g_{ij}$, so the penultimate and last terms on
the right-hand side cancel, leaving us with the much simpler result

$$\frac{\partial \delta n_v(\mathbf{x}, \mathbf{p}, t)}{\partial t} + \frac{\hat{p}_i}{a(t)}\frac{\partial \delta n_v(\mathbf{x}, \mathbf{p}, t)}{\partial x^i} = \frac{p\bar{n}'(p)}{2}\hat{p}_i\hat{p}_j\frac{\partial}{\partial t}\left(a^{-2}(t)\delta g_{ij}(\mathbf{x}, t)\right), \tag{H.14}$$

where again $p \equiv \sqrt{p_k p_k}$, and as usual we write $\hat{p}_i \equiv p_i/p$.

2 Photons

The Boltzmann equation for photons is considerably more complicated than for neutrinos, because of the necessity of taking photon scattering into account. Since scattering can change the polarization of photons, we can no longer write a separate Boltzmann equation for each helicity state of photons, as we did for neutrinos. Instead, we now define a number density matrix:

$$n_\gamma^{ij}(\mathbf{x}, \mathbf{p}, t) \equiv \sum_r \left(\prod_{k=1}^{3} \delta\left(x^k - x_r^k(t)\right)\right)\left(\prod_{k=1}^{3} \delta\left(p_k - p_{rk}(t)\right)\right) N_r^{ij}(t) \tag{H.15}$$

with r here labeling trajectories of individual photons, and $N_r^{ij}(t)$ the polarization density matrix of the r th photon. As discussed in Appendix G , if an individual photon can have any one of several polarization vectors e_m^i, with probabilities P_m, then it has a polarization density matrix

$$N^{ij} \equiv \sum_m P_m e_m^i e_m^{j*} \tag{H.16}$$

If we observe whether the polarization of a photon with polarization density matrix N^{ij} is in a particular direction e^i rather than in an orthogonal direction, we find a probability $g_{ik}g_{jl}e^{*i}e^j N^{kl}$. Recall that for a photon of three-momentum p_i in a general three-metric g_{ij}, the photon polarization vectors are defined so that

$$p_i e^i = 0, \quad g_{ij}e^i e^{j*} = 1 \tag{H.17}$$

so the polarization density matrix of photon r satisfies

$$p_{ri}(t)N_r^{ij}(t) = p_{ri}(t)N_r^{ji}(t) = 0, \quad g_{ij}\left(\mathbf{x}_r(t),t\right)N_r^{ij}(t) = 1 \tag{H.18}$$

and the number density matrix correspondingly satisfies

$$p_i n_\gamma^{ij}(\mathbf{x},\mathbf{p},t) = p_i n_\gamma^{ji}(\mathbf{x},\mathbf{p},t) = 0, \quad g_{ij}(\mathbf{x},t)n_\gamma^{ij}(\mathbf{x},\mathbf{p},t) = n_\gamma(\mathbf{x},\mathbf{p},t), \tag{H.19}$$

where $n_\gamma(\mathbf{x},\mathbf{p},t)$ is the phase space number density of photons, defined just as in Eq. (H.4). By their definition, both $N_r^{ij}(t)$ and $n_\gamma^{ij}(\mathbf{x},\mathbf{p},t)$ are Hermitian matrices.

As discussed in Appendix G, between collisions the time dependence of the polarization vectors is given by a combination of parallel transport with a gauge transformation that keeps the time component of $e_r^\mu(t)$ vanishing:

$$\frac{de_r^j(t)}{dt} = \left[-\Gamma_{i\lambda}^j\left(\mathbf{x}_r(t),t\right) + \Gamma_{i\lambda}^0\left(\mathbf{x}_r(t),t\right)\frac{dx_r^j(t)}{dt}\right]e_r^i(t)\frac{dx_r^\lambda(t)}{dt} \tag{H.20}$$

and consequently the time dependence of the density matrix of the r th photon is

$$\begin{aligned}\frac{dN_r^{ij}(t)}{dt} = &\left[-\Gamma_{k\lambda}^i\left(\mathbf{x}_r(t),t\right) + \Gamma_{k\lambda}^0\left(\mathbf{x}_r(t),t\right)\frac{dx_r^i(t)}{dt}\right]N_r^{kj}(t)\frac{dx_r^\lambda(t)}{dt} \\ &+ \left[-\Gamma_{k\lambda}^j\left(\mathbf{x}_r(t),t\right) + \Gamma_{k\lambda}^0\left(\mathbf{x}_r(t),t\right)\frac{dx_r^j(t)}{dt}\right]N_r^{ik}(t)\frac{dx_r^\lambda(t)}{dt}.\end{aligned} \tag{H.21}$$

The time dependence of the variables $p_{ri}(t)$ and $x_r^i(t)$ is given by the same equations (H.7) and (H.8) as for neutrinos. It follows then directly from Eqs. (H.15), (H.21), (H.7) and (H.8) that $n_\gamma^{ij}(\mathbf{x},\mathbf{p},t)$ satisfies a Boltzmann equation:

$$\begin{aligned}&\frac{\partial n_\gamma^{ij}}{\partial t} + \frac{\partial n_\gamma^{ij}}{\partial x^k}\frac{p^k}{p^0} + \frac{\partial n_\gamma^{ij}}{\partial p_k}\frac{p^l p^m}{2p^0}\frac{\partial g_{lm}}{\partial x^k} \\ &\quad + \left(\Gamma_{k\lambda}^i - \frac{p^i}{p^0}\Gamma_{k\lambda}^0\right)\frac{p^\lambda}{p^0}n_\gamma^{kj} \\ &\quad + \left(\Gamma_{k\lambda}^j - \frac{p^j}{p^0}\Gamma_{k\lambda}^0\right)\frac{p^\lambda}{p^0}n_\gamma^{ik} = C^{ij}\end{aligned} \tag{H.22}$$

where C^{ij} is a term representing the effect of photon scattering. (As a reminder, we note that in Eq. (H.22) p^i and p^0 are functions of $\mathbf{x},t$, and the p_j, given by $p^i \equiv g^{ij}p_j$ and $p^0 = \left[g^{ij}p_i p_j\right]^{1/2}$.)

To evaluate the collision term C^{ij} in Eq. (H.22), let us first consider the case of flat spacetime, where $g_{ij} = \delta_{ij}$. Since the Thomson scattering of an unpolarized or linearly polarized photon can only produce linearly polarized photons, we will limit ourselves here to the case of linear polarization, for which polarization vectors can be taken to be real, and the polarization density matrix and number density matrices are real and symmetric. We will first consider the case of electrons at rest, and later (in dealing with scalar perturbations) take up the case of a non-zero plasma velocity.

The scattering of photons of momentum $\mathbf{p}$ into some other direction causes a decrease in $n_\gamma^{ij}(\mathbf{x}, \mathbf{p}, t)$ which is simply given by a term in C^{ij} :

$$C_-^{ij}(\mathbf{x}, \mathbf{p}, t) = -\omega_c(t) n_\gamma^{ij}(\mathbf{x}, \mathbf{p}, t), \tag{H.23}$$

where $\omega_c(t)$ is the total collision rate. (We hold here to the convention of writing n_γ^{ij} and C^{ij} with upper indices, though of course in flat spacetime there is no distinction between contravariant and covariant spatial tensors.) There is also an increase in $n_\gamma^{ij}(\mathbf{x}, \mathbf{p}, t)$ caused by the scattering of photons with some initial momentum $\mathbf{p}_1$ into momentum $\mathbf{p}$, with energy conservation requiring (assuming that $|\mathbf{p}| \ll m_e$) that $|\mathbf{p}_1| = |\mathbf{p}|$. The Klein-Nishina formula [1] tells us that when a photon with momentum p_{1i} and real polarization vector e_1^i is scattered by an electron at rest, the probability of finding a photon in the final state with real polarization vector e^i is proportional to $(e \cdot e_1)^2$, with no dependence on the initial photon momentum p_{1i} or the final photon momentum p_i, except of course that e_1^i and e^i must be orthogonal to p_{1i} and p_i, respectively. If the initial photon can have various polarizations e_{1n}^i with probabilities P_n, then correspondingly the probability of finding the photon in the final state with polarization e^i is proportional to $e^i e^j N_1^{ij}$, where $N_1^{ij} = \sum_n P_n e_{1n}^i e_{1n}^j$. As already mentioned, this probability must equal $e^i e^j N^{ij}$, where N^{ij} is the polarization density matrix of the photon in the final state. Together with the conditions that $p_i N^{ij} = 0$ and $N^{ii} = 1$, this tells us that the scattered photon has polarization density matrix

$$N^{ij}(\hat{p}) = S^{-1}(\hat{p}) \left[N_1^{ij} - \hat{p}_i \hat{p}_k N_1^{kj} - \hat{p}_j \hat{p}_k N_1^{ik} + \hat{p}_i \hat{p}_j \hat{p}_k \hat{p}_l N_1^{kl} \right], \tag{H.24}$$

where

$$S(\hat{p}) \equiv 1 - \hat{p}_i \hat{p}_j N_1^{ij} \tag{H.25}$$

The rate of increase in $n_\gamma^{ij}(\mathbf{x}, \mathbf{p}, t)$ due to scattering of photons from an arbitrary direction $\hat{p}_1$ into direction $\hat{p}$ is then

$$\begin{aligned} C_+^{ij}(\mathbf{x}, \mathbf{p}, t) = n_e \int d^2\hat{p}_1 \frac{d^2\sigma}{d^2\hat{p}} \\ \times S^{-1}(\hat{p}) \left[n_\gamma^{ij}(\mathbf{x}, |\mathbf{p}|\hat{p}_1, t) - \hat{p}_i \hat{p}_k n_\gamma^{kj}(\mathbf{x}, |\mathbf{p}|\hat{p}_1, t) \right. \\ \left. - \hat{p}_j \hat{p}_k n_\gamma^{ik}(\mathbf{x}, |\mathbf{p}|\hat{p}_1, t) + \hat{p}_i \hat{p}_j \hat{p}_k \hat{p}_l n_\gamma^{kl}(\mathbf{x}, |\mathbf{p}|\hat{p}_1, t) \right] \end{aligned} \tag{H.26}$$

where n_e is the electron density, and $d^2\sigma/d^2\hat{p}$ is the differential scattering cross section. Summing $(e \cdot e_1)^2$ over final polarizations gives a differential cross section proportional to $S(\hat{p})$, and $\int d^2\hat{p} S(\hat{p}) = 8\pi/3$, so equating
$\int d^2\hat{p} \, (d^2\sigma/d^2\hat{p})$ to the Thomson scattering cross section $\sigma_\mathcal{T}$, we have

$$\frac{d^2\sigma}{d^2\hat{p}} = \frac{3\sigma_\mathcal{T}}{8\pi} S(\hat{p}) \tag{H.27}$$

[471] See, e.g., QTF, Vol. I, Sec. 8.7.

Then Eqs. (H.23), (H.26), and (H.27) give a net change in $n_\gamma^{ij}(\mathbf{x},\mathbf{p},t)$ equal to

$$\begin{aligned} C^{ij}(\mathbf{x},\mathbf{p},t) &= C_-^{ij}(\mathbf{x},\mathbf{p},t) + C_+^{ij}(\mathbf{x},\mathbf{p},t) \\ &= -\omega_c(t) n_\gamma^{ij}(\mathbf{x},\mathbf{p},t) + \frac{3\omega_c(t)}{8\pi}\int d^2\hat{p}_1 \\ \times \left[n_\gamma^{ij}\left(\mathbf{x},|\mathbf{p}|\hat{p}_1,t\right) - \hat{p}_i\hat{p}_k n_\gamma^{kj}\left(\mathbf{x},|\mathbf{p}|\hat{p}_1,t\right) - \hat{p}_j\hat{p}_k n_\gamma^{ik}\left(\mathbf{x},|\mathbf{p}|\hat{p}_1,t\right)\right. \\ \left. + \hat{p}_i\hat{p}_j\hat{p}_k\hat{p}_l n_\gamma^{kl}\left(\mathbf{x},|\mathbf{p}|\hat{p}_1,t\right)\right] \end{aligned} \tag{H.28}$$

in which we use $\omega_c(t) = n_e(t)\sigma_{\mathcal{T}}$.
As a check, it is easy to see that the collision term (H.28) vanishes in local thermal equilibrium. In equilibrium photons are unpolarized, so $n_\gamma^{ij} \propto \delta_{ij} - \hat{p}_i\hat{p}_j$, and their momentum distribution is homogeneous and isotropic, so the coefficient of proportionality f depends only on $|\mathbf{p}|$ and t :

$$n_{\rm eq}^{ij}(\mathbf{x},\mathbf{p},t) = f(|\mathbf{p}|,t)\left[\delta_{ij} - \hat{p}_i\hat{p}_j\right].$$

Then

$$\int d^2\hat{p}_1 n_{\rm eq}^{ij}\left(\mathbf{x},|\mathbf{p}|\hat{p}_1,t\right) = \frac{8\pi}{3} f(|\mathbf{p}|,t)\delta_{ij}$$

Using this in Eq. (H.28) shows that in equilibrium the two terms in C^{ij} cancel, as of course they must.

To find the result for an arbitrary three-metric g_{ij}, we must simply write Eq. (H.28) in a form that is invariant under general three-dimensional coordinate transformations, and reduces to Eq. (H.28) in the case of a flat three-metric with $g_{ij} = \delta_{ij}$. In this way, we find

$$\begin{aligned} C^{ij}(\mathbf{x},\mathbf{p},t) &= -\omega_c(t) n_\gamma^{ij}(\mathbf{x},\mathbf{p},t) \\ &+ \frac{3\omega_c(t)}{8\pi}\int \frac{d^3p_1\sqrt{\mathrm{Det}\, g(\mathbf{x},t)}}{p^{02}(\mathbf{x},\mathbf{p},t)}\delta\left(p^0(\mathbf{x},\mathbf{p},t) - p^0\left(\mathbf{x},\mathbf{p}_1,t\right)\right) \\ &\times \left[n_\gamma^{ij}\left(\mathbf{x},\mathbf{p}_1,t\right) - \frac{g^{ik}(\mathbf{x},t)p_k p_l}{p^{02}(\mathbf{x},\mathbf{p},t)} n_\gamma^{lj}\left(\mathbf{x},\mathbf{p}_1,t\right)\right. \\ &- \frac{g^{jk}(\mathbf{x},t)p_k p_\ell}{p^{02}(\mathbf{x},\mathbf{p},t)} n_\gamma^{il}\left(\mathbf{x},\mathbf{p}_1,t\right) \\ &\left. + \frac{g^{ik}(\mathbf{x},t)g^{jl}(\mathbf{x},t)p_k p_l p_m p_n}{p^{04}(\mathbf{x},\mathbf{p},t)} n_\gamma^{mn}\left(\mathbf{x},\mathbf{p}_1,t\right)\right] \end{aligned} \tag{H.29}$$

where

$$p^0(\mathbf{x},\mathbf{p},t) \equiv \sqrt{g^{ij}(\mathbf{x},t)p_i p_j} \tag{H.30}$$

We now again specialize to the case of a small perturbation, writing the metric as in Eq. (H.10), and putting the photon number density matrix in a form analogous to Eq. (H.11):

$$\begin{aligned} n_\gamma^{ij}(\mathbf{x},\mathbf{p},t) = \frac{1}{2}\bar{n}_\gamma\left(a(t)p^0(\mathbf{x},\mathbf{p},t)\right)\left[g^{ij}(\mathbf{x},t) - \frac{g^{ik}(\mathbf{x},t)g^{jl}(\mathbf{x},t)p_k p_l}{p^{02}(\mathbf{x},\mathbf{p},t)}\right] \\ + \delta n_\gamma^{ij}(\mathbf{x},\mathbf{p},t) \end{aligned} \tag{H.31}$$

where $\bar{n}_\gamma(p)$ is the equilibrium phase space number density, a timeindependent function of its argument

$$\bar{n}_\gamma(p) \equiv \frac{1}{(2\pi)^3}\left[\exp\left(p/k_{\mathcal{B}}a(t)T(t)\right)-1\right]^{-1} \tag{H.32}$$

and δn^{ij} is a small perturbation. Fortunately, to first order in perturbations the Boltzmann equation (H.22) is greatly simplified by the fact that the quantity in square brackets in Eq. (H.31) satisfies a collisionless Boltzmann equation

$$\begin{aligned} 0 = &\left(\frac{\partial}{\partial t}+\frac{p^k}{p^0}\frac{\partial}{\partial x^k}+\frac{p^l p^m}{2p^0}\frac{\partial g_{lm}}{\partial x^k}\frac{\partial}{\partial p^k}\right)\left[g^{ij}-\frac{p^i p^j}{p^{02}}\right] \\ &+\frac{p^\lambda}{p^0}\left(\Gamma^i_{k\lambda}-\frac{p^i}{p^0}\Gamma^0_{k\lambda}\right)\left[g^{kj}-\frac{p^k p^j}{p^{02}}\right] \\ &+\frac{p^\lambda}{p^0}\left(\Gamma^j_{k\lambda}-\frac{p^j}{p^0}\Gamma^0_{k\lambda}\right)\left[g^{ik}-\frac{p^i p^k}{p^{02}}\right]. \end{aligned} \tag{H.33}$$

(This can be proved directly, or more easily by noting that if we set $p_i = p_{ri}(t)$ and $x^i = x^i_r(t)$, then the quantity in square brackets is the sum over two orthogonal polarizations of the quantity $e^i_r(t)e^j_r(t)$, and therefore satisfies Eq. (H.33) as a consequence of Eq. (H.20). Since Eq. (H.33) holds in this sense for any photon trajectory, it is necessary for it to hold for arbitrary x^i and p_i.) Acting on the factor $\bar{n}_\gamma\left(ap^0\right)$, the derivative operators in Eq. (H.22) give to first order

$$\begin{aligned} \left(\frac{\partial}{\partial t}+\frac{p^k}{p^0}\frac{\partial}{\partial x^k}+\frac{p^l p^m}{2p^0}\frac{\partial g_{lm}}{\partial x^k}\frac{\partial}{\partial p_k}\right)\bar{n}_\gamma\left(ap^0\right) &= -\frac{1}{2}\sqrt{p_i p_i} \\ \times\bar{n}'_\gamma\left(\sqrt{p_i p_i}\right)\hat{p}_k\hat{p}_l\frac{\partial}{\partial t}\left(a^{-2}\delta g_{kl}\right) & \end{aligned} \tag{H.34}$$

with the effect of the position and momentum derivative terms in parentheses on the left-hand side canceling. There is another simplification provided by the fact that the the first term in Eq. (H.31) does not contribute to the collision term, since even in a gravitational field collisions by themselves do not alter equilibrium distribution functions. Keeping only terms of first order in perturbations, the Boltzmann equation (H.22) then becomes

$$\begin{aligned} &\frac{\partial\delta n^{ij}_\gamma(\mathbf{x},\mathbf{p},t)}{\partial t}+\frac{\hat{p}_k}{a(t)}\frac{\partial\delta n^{ij}_\gamma(\mathbf{x},\mathbf{p},t)}{\partial x^k}+\frac{2\dot{a}(t)}{a(t)}\delta n^{ij}_\gamma(\mathbf{x},\mathbf{p},t) \\ &-\frac{1}{4a^2(t)}p\bar{n}'_\gamma(p)\hat{p}_k\hat{p}_l\frac{\partial}{\partial t}\left(a^{-2}\delta g_{kl}(\mathbf{x},t)\right)\left(\delta_{ij}-\hat{p}_i\hat{p}_j\right) \\ &=-\omega_c(t)\delta n^{ij}_\gamma(\mathbf{x},\mathbf{p},t)+\frac{3\omega_c(t)}{8\pi}\int d^2\hat{p}_1 \\ &\times\left[\delta n^{ij}_\gamma\left(\mathbf{x},p\hat{p}_1,t\right)-\hat{p}_i\hat{p}_k\delta n^{kj}_\gamma\left(\mathbf{x},p\hat{p}_1,t\right)-\hat{p}_j\hat{p}_k\delta n^{ik}_\gamma\left(\mathbf{x},p\hat{p}_1,t\right)\right. \\ &\left.+\hat{p}_i\hat{p}_j\hat{p}_k\hat{p}_l\delta n^{kl}_\gamma\left(\mathbf{x},p\hat{p}_1,t\right)\right] \end{aligned} \tag{H.35}$$

where again $p \equiv \sqrt{p_i p_i}$, and $\hat{p} \equiv \mathbf{p}/p$. As a check, note that Eq. (H.35) is consistent with the condition $p_i\delta n^{ij} = 0$. As a further check, note that the first-order perturbation to the photon phase space number density $g_{ij}n^{ij}_\gamma$ is $a^2\delta n^{ii}$, and according to the trace of Eq. (H.35), the Boltzmann equation for $a^2\delta n^{ii}_\gamma$ is the same as the Boltzmann equation (H.14) for the perturbation to the neutrino phase space density, aside from the presence of the collision term and the appearance of $\bar{n}'_\gamma$ instead of $\bar{n}'_\nu$.

For tensor modes, Eq. (H.35) can be used as it stands, but in dealing with scalar (or vector) modes, we need to take up a complication that has been ignored until now: in general the plasma has a small velocity $\delta\mathbf{u}(\mathbf{x},t)$ (which in Section 6.1 et seq. is denoted $\delta\mathbf{u}_B(\mathbf{x},t)$). Since $\delta\mathbf{u}$ is itself a first-order perturbation, in calculating its effect on the collision term C^{ij}, to first order we can ignore any perturbations to the gravitational field or the photon number density matrix. The effect of the plasma velocity is to shift the energy $|\mathbf{p}_1|$ of the incident photon that when scattered yields a final photon with momentum $\mathbf{p}$. To calculate this effect, let's first consider the case of flat spacetime. Since photon energy is conserved in the rest frame of the plasma, in which the plasma metric four-velocity is simply $v^\mu = (0,0,0,1)$, we have $(p_1 - p)\cdot v = 0$; evaluating this scalar in the "lab" frame in which the plasma velocity four-vector is $v^\mu = \left[\delta\mathbf{v}, \sqrt{1-(\delta\mathbf{v})^2}\right]$ we have

$$0 = v_\mu (p_1 - p)^\mu = \delta\mathbf{v}\cdot(\mathbf{p}_1 - \mathbf{p}) - \sqrt{1-(\delta\mathbf{v})^2}\left(p_1^0 - p^0\right),$$

or, to first order in velocity,

$$|\mathbf{p_1}| = |\mathbf{p}|\left[1 + (\hat{p}_1 - \hat{p})\cdot\delta\mathbf{v}\right].$$

As already noted, the flat-space velocity components v^k are related to the velocity components in a Robertson-Walker three-metric $g_{ij} = a^2\delta_{ij}$ by $v^k = au^k = a^{-1}u_k$, so

$$|\mathbf{p_1}| = |\mathbf{p}|\left[1 + \frac{(\hat{p}_1 - \hat{p})_k\,\delta u_k}{a}\right]. \tag{H.36}$$

As a result, the two terms in the collision term (H.28) do not cancel even when we set the three-metric g_{ij} equal to $a^2\delta_{ij}$ and set the photon number density matrix equal to the form it would have in equilibrium with this three-metric:

$$n^{ij}_{\gamma,\mathrm{eq}}(\mathbf{x},\mathbf{p},t) = \frac{1}{2a^2(t)}\bar{n}_\gamma(|\mathbf{p}|)\left[\delta_{ij} - \hat{p}_i\hat{p}_j\right].$$

Since in this case these two terms in C^{ij} would cancel if $|\mathbf{p_1}|$ were equal to $|\mathbf{p}|$, we find a new term in $C^{ij}(\mathbf{x},\mathbf{p},t)$, equal to the difference between the term $-\omega_c(t)n^{ij}_{\gamma,\,\mathrm{eq}}(\mathbf{x},\mathbf{p},t)$ and the same term with $|\mathbf{p}|$ replaced with $|\mathbf{p_1}|$, averaged over $\hat{p}_1$. The linearized Boltzmann equation (H.35) now becomes

$$\begin{aligned}
&\frac{\partial\delta n^{ij}_\gamma(\mathbf{x},\mathbf{p},t)}{\partial t} + \frac{\hat{p}_k}{a(t)}\frac{\partial\delta n^{ij}_\gamma(\mathbf{x},\mathbf{p},t)}{\partial x^k} + \frac{2\dot{a}(t)}{a(t)}\delta n^{ij}_\gamma(\mathbf{x},\mathbf{p},t)\\
&\quad -\frac{1}{4a^2(t)}p\bar{n}'_\gamma(p)\hat{p}_k\hat{p}_l\frac{\partial}{\partial t}\left(a^{-2}(t)\delta g_{kl}(\mathbf{x},t)\right)(\delta_{ij} - \hat{p}_i\hat{p}_j)\\
&\qquad = -\omega_c(t)\delta n^{ij}_\gamma(\mathbf{x},\mathbf{p},t) + \frac{3\omega_c(t)}{8\pi}\int d^2\hat{p}_1\\
&\times\left[\delta n^{ij}_\gamma(\mathbf{x},p\hat{p}_1,t) - \hat{p}_i\hat{p}_k\delta n^{kj}_\gamma(\mathbf{x},p\hat{p}_1,t) - \hat{p}_j\hat{p}_k\delta n^{ik}_\gamma(\mathbf{x},p\hat{p}_1,t)\right.\\
&\qquad \left. + \hat{p}_i\hat{p}_j\hat{p}_k\hat{p}_l\delta n^{kl}_\gamma(\mathbf{x},p\hat{p}_1,t)\right]\\
&\quad -\frac{\omega_c(t)}{2a^3(t)}\left(p_k\delta u_k(\mathbf{x},t)\right)\bar{n}'_\gamma(p)\left[\delta_{ij} - \hat{p}_i\hat{p}_j\right]
\end{aligned} \tag{H.37}$$

(The term in Eq. (H.36) proportional to $\hat{p}_{1k}\delta u_k$ gives no contribution to the integral over $\hat{p}_1$.)

In an important application of Eq. (H.37), we may derive the decay rate of acoustic waves in a homogeneous and time-independent plasma in flat

spacetime (with $a = 1$ and ω_c time-independent) when the collision rate is much larger than the sound frequency. (This result will be used in Section 6.4.) We seek a solution of the Boltzmann equation (H.37) of the form

$$\delta n_\gamma^{ij}(\mathbf{x}, \mathbf{p}, t) = e^{i\mathbf{k}\cdot\mathbf{x}} e^{-i\omega t} \delta n_\gamma^{ij}(\mathbf{p}), \quad \delta u_j(\mathbf{x}, t) = e^{i\mathbf{k}\cdot\mathbf{x}} e^{-i\omega t} \delta u_j. \tag{H.38}$$

Then with $a = 1$ and ω_c constant, Eq. (H.37) becomes

$$\begin{aligned}
&[\omega_c - i\omega + i\hat{p}\cdot\mathbf{k}]\, \delta n_\gamma^{ij}(\mathbf{p}) \\
&= \frac{3\omega_c}{8\pi} \int d^2\hat{p}_1 \\
&\quad \times \left[\delta n_\gamma^{ij}(p\hat{p}_1) - \hat{p}_i\hat{p}_k \delta n_\gamma^{kj}(p\hat{p}_1) - \hat{p}_j\hat{p}_k \delta n_\gamma^{ik}(p\hat{p}_1) + \hat{p}_i\hat{p}_j\hat{p}_k\hat{p}_l \delta n_\gamma^{kl}(p\hat{p}_1)\right] \\
&\quad - \frac{\omega_c}{2}(\mathbf{p}\cdot\delta\mathbf{u})\bar{n}_\gamma'(p)\left[\delta_{ij} - \hat{p}_i\hat{p}_j\right].
\end{aligned} \tag{H.39}$$

We now need a formula for the plasma velocity δu_j. For this purpose, we note that in flat space the first-order photon and baryonic plasma contributions to the energy-momentum tensor have

$$\delta T_{\gamma j}^i = \int d^3p a^2 n_\gamma^{kk}(\mathbf{p}) |\mathbf{p}| \hat{p}_i \hat{p}_j, \quad \delta T_{\gamma j}^0 = \int d^3p a^2 n_\gamma^{kk}(\mathbf{p}) |\mathbf{p}| \hat{p}_j, \tag{H.40}$$

and

$$\delta T_{Bj}^i = 0, \quad \delta T_{Bj}^0 = \bar{\rho}_B \delta u_j, \tag{H.41}$$

so the equation of momentum conservation may be written

$$\omega \bar{\rho}_B \delta u_j = \int d^3p a^2 \delta n_\gamma^{kk}(\mathbf{p}) p_j (\mathbf{k}\cdot\hat{p} - \omega). \tag{H.42}$$

To evaluate the right-hand side, multiply Eq. (H.39) with p_j and integrate over $\mathbf{p}$; this gives

$$i \int d^3p (\mathbf{k}\cdot\hat{p} - \omega) p_j \delta n_\gamma^{kk}(\mathbf{p}) = -\omega_c \int d^3p p p_j \delta n_\gamma^{kk}(\mathbf{p}) + \frac{4\omega_c}{3} \delta u_j \bar{\rho}_\gamma,$$

where $\bar{\rho}_\gamma = \int d^3p p \bar{n}_\gamma(p) = -(1/4) \int d^3p p^2 \bar{n}_\gamma'(p)$. Using this in Eq. (H.42) gives our formula for $\delta\mathbf{u}$:

$$\left(\omega\bar{\rho}_B + \frac{4i\omega_c}{3}\bar{\rho}_\gamma\right) \delta u_j = i\omega_c \int d^3p p p_j \delta n_\gamma^{kk}(\mathbf{p}). \tag{H.43}$$

For scalar modes, $\mathbf{k}$ is the only parameter in the problem with a sense of direction, so we can express the integrals appearing in Eqs. (H.39) and
(H.43) as

$$\int d^3p |\mathbf{p}| \delta n_\gamma^{ij}(\mathbf{p}) = \bar{\rho}_\gamma \left[X\delta_{ij} + Y\hat{k}_i\hat{k}_j\right] \tag{H.44}$$

$$\int d^3p \delta n_\gamma^{ii}(\mathbf{p}) p_j = \bar{\rho}_\gamma Z \hat{k}_j \tag{H.45}$$

where X, Y, and Z are here functions of $k \equiv |\mathbf{k}|$ and ω. Eq. (H.43) then lets us express $\delta\mathbf{u}$ in terms of Z :

$$\delta\mathbf{u} = \frac{3Z}{4\left[1 - i\omega t_c R\right]}\hat{k} \tag{H.46}$$

where $R \equiv 3\bar{\rho}_B/4\bar{\rho}_\gamma$, and $t_c \equiv 1/\omega_c$ is the mean time between collisions. Using Eqs. (H.44) and (H.46) in Eq. (H.39) then yields a formula for the perturbed intensity:

$$\begin{aligned} 4\pi \int \delta n_r^{ij}(p\hat{p})p^3 dp = &\frac{3\bar{\rho}_\gamma}{2\left[1 - i\omega t_c + i\hat{p}\cdot\mathbf{k}t_c\right]}\left[\left(X + \frac{Z(\hat{p}\cdot\hat{k})}{1 - i\omega t_c R}\right)(\delta_{ij} - \hat{p}_i\hat{p}_j)\right. \\ &\left. + Y\left(\hat{k}_i - \hat{p}_i(\hat{p}\cdot\hat{k})\right)\left(\hat{k}_j - \hat{p}_j(\hat{p}\cdot\hat{k})\right)\right] \end{aligned} \tag{H.47}$$

By inserting this back in the definitions (H.44) and (H.45), we find three homogeneous linear relations among X, Y, and Z. For these to be consistent, the determinant of the coefficients must vanish, which yields a relation between ω and $k = |\mathbf{k}|$.

The resulting dispersion relation is quite complicated, but it becomes much simpler in the case of small mean free time t_c. As an ansatz, we can try taking Y/X of first order in t_c, while Z/X is of order unity, leaving it for later to check whether this leads to a consistent solution. To second order in t_c, the terms in Eq. (H.44) proportional to δ_{ij} and $\hat{k}_i\hat{k}_j$ and the coefficient of $\hat{k}_i$ in Eq. (H.45) yield the homogeneous linear relations

$$\begin{aligned} 0 = &\left(i\omega t_c - \omega^2 t_c^2 - \frac{2}{5}k^2 t_c^2\right)X + \frac{1}{10}\left(1 + i\omega t_c\right)Y \\ &+\frac{2}{5}\left(-ikt_c + \omega k t_c^2(2 + R)\right)Z \end{aligned} \tag{H.48}$$

$$\begin{aligned} 0 = &\frac{1}{5}k^2 t_c^2 X + \left(-\frac{3}{10} + \frac{7i}{10}\omega t_c\right)Y \\ &+\left(\frac{i}{5}kt_c - \frac{(2 + R)}{5}\omega k t_c^2\right)Z \end{aligned} \tag{H.49}$$

$$\begin{aligned} 0 = &\left(-ikt_c + 2\omega k t_c^2\right)X - \frac{1}{5}ikt_c Y \\ &+\left(i\omega t_c(1 + R) - \omega^2 t_c^2\left(1 + R + R^2\right) - \frac{3}{5}k^2 t_c^2\right)Z. \end{aligned} \tag{H.50}$$

(As anticipated, these relations are consistent with the assumption that Y is of order t_c relative to X and Z.) To first order in t_c, the vanishing of the determinant of the coefficients in these relations yields the condition

$$0 = 15(1 + R)\omega^2 - 5k^2 + i\omega^3 t_c\left[-5 - 5R + 15R^2\right] + 7i\omega k^2 t_c$$

This has the solution, again to first order in t_c :

$$\omega = \pm\frac{k}{\sqrt{3(1 + R)}} - i\Gamma \tag{H.51}$$

where Γ is the decay rate

$$\Gamma = \frac{k^2 t_c}{6(1 + R)}\left\{\frac{16}{15} + \frac{R^2}{1 + R}\right\} \tag{H.52}$$

This is the same as the formula (6.4.25) for the damping rate given in Section 6.4, originally derived by Kaiser. [2] As discussed there, it is equivalent to the formulas (6.4.24) for the coefficients of shear viscosity and heat conduction of a non-relativistic plasma in which momentum and energy are transported by photons.

10.19 Glossary of Symbols

(These are symbols used in more than one section. Numbers following the symbol give the section in which the symbol is first used.)

a	1.1	scale factor in Robertson-Walker metric
$a_{\mathcal{B}}$	2.1	radiation energy constant
$a_{\ell m}$	2.6	temperature partial wave amplitude (same as $a_{T,\ell m}$)
A	5.1	one of scalar perturbations to metric
B	3.3	baryon number
	5.1	one of scalar perturbations to metric
	6.1	subscript denoting baryonic plasma
C_i	5.1	one of vector perturbations to metric
C_ℓ	2.6	temperature multipole coefficient (same as $C_{TT,\ell}$)
d	1.2	proper distance
d_A	1.4	angular diameter distance
d_H	2.6	horizon distance
d_L	1.4	luminosity distance
D	6.1	subscript denoting cold dark matter
$D^{(\ell)}$	7.4	representative of the rotation group for spin ℓ
D_{ij}	5.1	tensor perturbation to metric
$\mathcal{D}_q$	5.2	gravitational wave amplitude
$\mathcal{D}_q^o$	6.6	constant gravitational wave amplitude outside horizon
e_{ij}	5.2	graviton polarization tensor
E	5.1	one of scalar perturbations to metric
EQ	6.2	subscript indicating the time of matter-radiation equality
F	5.1	one of scalar perturbations to metric
	6.6	function appearing in tensor anisotropic inertia
	7.1	form factor in scalar temperature fluctuation
g	3.1	multiplicity of states
$g_{\mu\nu}$	1.1	spacetime metric
$\tilde{g}_{ij}$	1.1	spatial part of Robertson-Walker metric
G	1.5	Newton's gravitational constant
	7.1	form factor in scalar temperature fluctuation
G_i	5.1	one of vector perturbations to metric
h	1.3	Hubble constant in units of 100 kmsec^{-1}Mpc^{-1}
	2.4	Planck's constant $= 2\pi\hbar$
$h_{\mu\nu}$	5.1	perturbation to the metric
H	2.2	expansion rate $\dot{a}(t)/a(t)$
H_0	1.2	Hubble constant $\dot{a}\left(t_0\right)/a\left(t_0\right)$

[472] N. Kaiser, Mon. Not. Roy. Astron. Soc. 202, 1169 (1983).

$H_v()$	10.1	Hankel functions
I	1.12	action
ISW	2.6	contribution of integrated Sachs-Wolfe effect
j_ℓ	2.6	spherical Bessel function
J	6.1	dimensionless neutrino intensity
J^μ	1.1	current four-vector
J_{ij}	6.1	dimensionless photon intensity matrix
k	8.1	physical wave number at present $= q/a_0$
$k_{\mathcal{B}}$	2.2	Boltzmann constant
K	1.1	curvature constant in Robertson-Walker metric
	6.6	function appearing in tensor anisotropic inertia
ℓ	1.3	apparent luminosity
	2.6	multipole order
L	1.3	absolute luminosity
	2.1	subscript indicating time of last scattering
	3.3	lepton number
m	1.3	apparent magnitude
	1.5	particle mass
M	1.3	absolute magnitude
	1.9	total mass
Mpc	1.3	million parsecs
n	1.10	number density
n^{ij}	6.1	photon number density matrix
n_S	7.2	slope parameter in $\mathcal{R}_q^o$
N	7.2	coefficient of $q^{-3/2}$ in $\mathcal{R}_q^o$
$\mathcal{N}$	3.1	effective number of particle types
	4.1	number of e-foldings in inflation
p	1.1	pressure
pc	1.3	parsec
P_ℓ	2.4	Legendre polynomials
P	8.1	power spectral function of density fluctuations
$\mathcal{P}_\phi$	2.6	power spectral function of gravitational potential
q	2.6	co-moving wave number vector
q_0	1.4	deceleration parameter
r	1.1	radial coordinate in Robertson-Walker metric
	7.3	tensor/scalar ratio
R	2.6	3/4 the ratio of baryon to photon energy density
$R_{\mu\nu}$	1.5	Ricci tensor
$\mathcal{R}$	5.4	curvature perturbation
$\mathcal{R}^o$	6.2	constant curvature perturbation outside horizon
s	1.1	proper length
	3.1	entropy density

S	7.1	superscript indicating scalar mode
$S_{\mu\nu}$	1.5	source term in Einstein field equation
$S(\hat{n})$	7.4	standard rotation from three-axis to direction $\hat{n}$
$\mathcal{S}$	6.5	a scalar transfer function
SW	2.6	contribution of Sachs-Wolfe effect
t	1.1	time
t_0	1.5	present age of universe
t_1	6.1	initial time in solution of Boltzmann equation
T	2.1	temperature
	7.1	superscript indicating tensor mode
$T_{\mu\nu}$	1.1	energy-momentum tensor
$\mathcal{T}$	6.5	the main scalar transfer function
u	1.1	affine parameter
	5.1	velocity potential
u_μ	5.1	velocity four-vector
$\mathcal{U}$	6.6	a tensor transfer function
v_s	2.6	sound speed
V	1.12	scalar field potential
w	1.1	ratio of pressure to density
x	1.1	co-moving spatial coordinate
X	2.3	fractional ionization
X	1.9	Newtonian spatial coordinate
y	6.3	ratio of matter and radiation densities $= a/a_{\rm EQ}$
Y_ℓ^m	2.6	ordinary spherical harmonics
$\mathcal{Y}_\ell^m$	7.4	spin +2 spherical harmonics
z	1.2	redshift
	10.1	quantity $a\dot{\bar{\varphi}}/H$ in Mukhanov-Sasaki equation
α	5.2	stochastic parameter of scalar modes
	9.1	angle between gravitational lens and true source positions
β	5.2	stochastic parameter of tensor modes
	6.4	ratio of baryon and total matter densities
	9.1	angle between gravitational lens and apparent source positions
β_E	9.1	angular radius of Einstein ring
γ	6.1	subscript denoting photons
	6.4	Euler constant
Γ	6.4	acoustic decay rate
$\Gamma^\mu_{\nu\kappa}$	1.1	affine connection
$\tilde{\Gamma}^i_{jk}$	1.1	spatial affine connection for Robertson-Walker metric
δ_{ij}	1.1	Kronecker delta
δ_{Xq}	6.2	$\delta\rho_{Xq}/\left(\bar{\rho}_X + \bar{p}_X\right)$ (for $X = B, D, \gamma, \nu$)
Δ	5.3	change due to gauge transformation

	6.5	a scalar transfer function
$\Delta_T^{(S)}$	6.1	scalar photon temperature amplitude
$\Delta_P^{(S)}$	6.1	scalar photon polarization amplitude
$\Delta_T^{(T)}$	6.6	tensor photon temperature amplitude
$\Delta_P^{(T)}$	6.6	tensor photon polarization amplitude
ϵ	10.1	$-\dot{H}/H^2$
ϵ_0, ϵ^S	5.3	gauge transformation parameters
ζ	5.4	perturbation conserved outside horizon
η	3.2	baryon/photon ratio
$\eta_{\mu\nu}$	1.5	Minkowski spacetime metric
θ	1.1	polar angle (except in Sec. 9.5)
κ	6.2	dimensionless rescaled wave number
μ	3.1	chemical potential
ν	1.2	frequency
	6.1	subscript denoting neutrinos & antineutrinos
Ξ	6.6	a tensor transfer function
π_{ij}	1.5	total anisotropic inertia tensor
π^S	1.5	scalar anisotropic inertia
π_{iT}	1.5	vector anisotropic inertia
π_{ij}^T	1.5	tensor anisotropic inertia
Π	6.1	a scalar source function
ρ	1.1	energy density
σ	1.10	cross section
	2.2	entropy per baryon
	8.1	standard deviation of density perturbations
τ	1.1	proper time
	1.10	optical depth
	5.3	parameter for transformation between synchronous gauges
	10.1	conformal time
ϕ	1.1	azimuthal angle (except in Sec. 9.5)
	2.6	gravitational potential
φ	1.12	scalar field
Φ	5.3	a scalar metric perturbation in Newtonian gauge
	6.1	a scalar source function
ψ	5.3	a scalar metric perturbation in synchronous gauge
Ψ	5.3	a scalar metric perturbation in Newtonian gauge
	6.6	tensor source function
Ω_K	1.5	$-K/a_0^2H_0^2$
Ω_M	1.5	ratio of matter density to critical density
Ω_R	1.5	ratio of radiation density to critical density
Ω_Λ	1.5	ratio of vacuum density to critical density

10.20 Assorted Problems

1. Consider a universe described by a Robertson-Walker metric with $K = +1$. Give a transformation of co-moving space coordinates that leaves the metric unchanged, and that takes the point $\mathbf{x} = (0, 0, r)$ into a point $\mathbf{x} = (0, 0, r')$, with no change in the time. (Hint: Consider the three-dimensional space as the surface of a four-dimensional ball, construct this transformation as a rotation in four dimensions, and then express it in

terms of the Robertson-Walker coordinates.) Also give the corresponding transformation for $K = -1$.

2. Suppose that the total pressure and energy density of the universe are related by $p = -\rho + \rho^2/\rho_1$, where ρ_1 is a constant. Assume zero spatial curvature. Calculate ρ as a function of the Robertson-Walker scale factor a, taking $a = a_1$ when $\rho = \infty$. Calculate a and ρ as functions of time, taking $t = 0$ as the time when $\rho = \infty$. Calculate the age of the universe and the deceleration parameter q_0 for a given present density ρ_0.

3. Consider the empty cosmology, with $\Omega_M = \Omega_R = \Omega_\Lambda = 0$. Calculate the luminosity distance and angular diameter distance as functions of redshift in this cosmological model. What is the age of the universe as a function of the present Hubble constant?

4. Suppose that astronomers measure the age of a galaxy with redshift $z = 2.5$. How old would this galaxy have to be (at the time the light from it was emitted) in order to rule out the hypothesis that $\Omega_M = 1$ with negligible vacuum and radiation energy density. Use $H_0 = 70$km/sec/Mpc.

5. Suppose that $\Omega_M = 0.25$ and $\Omega_V = 0.75$, with Ω_R negligible. What is the redshift at which the expansion of the universe stopped decelerating and began to accelerate?

6. Suppose that the gravitational potential energy of any pair of galaxies with separation r decreases as r^{-n} instead of r^{-1}. What combination of the mass of a virialized cluster of galaxies and the Hubble constant could be calculated from measurements of angular separations and velocity dispersions of its individual galaxies?

7. Suppose that the fluctuations in the temperature of the cosmic microwave background are governed by a Gaussian isotropic
translation-invariant probability distribution. Calculate the quantity

$$\left\langle \left(C_\ell^{\rm obs} - C_\ell\right)^3 \right\rangle / C_\ell^3$$

as a function of ℓ.

8. Suppose that there is some new massless elementary particle that effectively decouples from everything else at a temperature 10^{16} K. Suppose that the particles with mass less (and in fact much less) than 1 TeV are those of the minimal Standard Model: three generations of quarks, leptons, antiquarks, and antileptons with helicity $\pm 1/2$; the $W^\pm$and Z^0 with helicity ± 1 and 0 ; the photon and eight massless gluons with helicity ± 1; and a single neutral boson of helicity zero; together with their antiparticles (where distinct from the particles). What is the ratio at present of the temperature of the new massless particles and the temperature of photons?

9. Suppose that the binding energy of deuterium were 5% less. Estimate the effect that this would have on the abundance of He^4 produced in the early universe.

10. Suppose that instead of $B - L$, only the combination $3B - L$ of baryon number and lepton number were conserved, along with the usual conserved quantities of the Standard Model. Taking account of the particles of the minimal (non-supersymmetric) Standard Model, what is the density of baryon number and lepton number in thermal equilibrium for a given density of $3B - L$?

11. Suppose that the Robertson-Walker scale factor increases during inflation by a factor e^{50}; that the scale factor at the beginning of the radiation-dominated era was the same as at the end of inflation; that the energy density at the end of inflation was $[2 \times 10^{16}\text{GeV}]^4$; and that the energy density at the beginning of the radiation-dominated era was $[2 \times 10^{15}\text{GeV}]^4$. Over what range of angles would you expect the cosmic microwave background now to have a nearly

uniform temperature?
12. Use the "slow-roll" approximation to calculate the inflaton field $\varphi(t)$ as a function of time for a potential $V(\varphi) = g\varphi^n$, where g and n are positive constants.
13. Find the two independent solutions of the field equations and conservation equations for scalar perturbations far outside the horizon in Newtonian gauge in the case of a perfect fluid having $p = w\rho$, with
w constant. Use the results to calculate $\mathcal{R} \equiv -\Psi + H\delta u$ for each solution.
14. Consider the equations for scalar fluctuations in Newtonian gauge in a universe containing only cold dark matter. Find the two solutions of the coupled equations for Φ_q and δu_q. (You can normalize these solutions any way you like, but keep it simple.) For each solution, calculate $\mathcal{R}_q, \delta\rho_q$, and ζ_q. Do not make any assumption about the magnitude of the wave number.
15. Consider a vector field

$$V^\mu(\mathbf{x}, t) = \bar{V}^\mu(t) + \delta V^\mu(\mathbf{x}, t)$$

where the unperturbed space components $\bar{V}^i$ are zero, and the unperturbed time component $\bar{V}^0$ is some non-zero function only of time. What is the change in δV^0 and δV^i induced by a general gauge transformation?
16. Re-do the analysis of Section 5.2, but for vector fluctuations.
17. Suppose that the differential scattering cross section of neutrinos on all other particles were a constant, independent of direction or energy. (It isn't.) For a given total neutrino collision rate $\omega_c(t)$, derive the Boltzmann equation that would be satisfied by the neutrino phase space distribution function $n_\nu(\mathbf{x}, \mathbf{p}, t)$ in a spacetime metric with $g_{00} = -1$, $g_{i0} = 0$, and $g_{ij}(\mathbf{x}, t)$ arbitrary. For the case where $g_{ij}(\mathbf{x}, t) = a^2(t)\delta_{ij} + \delta g_{ij}(\mathbf{x}, t)$, derive the Boltzmann equation for the phase space density perturbation $\delta n_\nu(\mathbf{x}, t)$ defined by Eq. (6.1.39), to first order in δg_{ij} and δn_ν.
18. Suppose that H_0 were doubled, and Ω_B and Ω_M reduced by a factor four. What would be the qualitative effect on the plot of $\ell^2 C_\ell$ vs. ℓ for $\ell \gg 1$?
19. Assume that $C^S_{TT,\ell}$ is dominated by a co-moving wave number at last scattering equal to ℓ/r_L, where r_L is the co-moving radial coordinate of the surface of last scattering. Also, neglect the Doppler effect, and use $\mathcal{S} = 5$ and $\mathcal{T} = 1$ for the transfer functions. With these assumptions, give an approximate formula for the ratio of the heights of the first and second acoustic peaks in $C^S_{TT,\ell}$.
20. What value of the multipole order ℓ corresponds to a co-moving wavelength equal to the diameter of a co-moving sphere that encloses the total mass $10^{15} M_\odot$ (including cold dark matter) of a cluster of galaxies?
21. Define a mean square matter density fluctuation as

$$\sigma_R^2 \equiv \left\langle \left(\frac{1}{(\sqrt{\pi}R)^3} \int d^3x \delta_M(\mathbf{x}, t) e^{-\mathbf{x}^2/R^2} \right)^2 \right\rangle$$

where the integral extends over all space. At what time does σ_R^2 become equal to 0.1 for R equal to the diameter of a co-moving sphere that encloses the total mass $10^{12} M_\odot$ (including cold dark matter) of a large galaxy? Assume a Harrison-Zel'dovich primordial spectrum with $|N| = 2 \times 10^{-5}$.
22. Derive a formula for the multipole coefficients in the correlation between fluctuations in the cosmic microwave background temperature in a direction $\hat{n}$ and the cosmic mass density fluctuation at a redshift z and direction $\hat{n}'$. Use any approximation that you think is reasonable, but make it clear what approximations you are using.
23. What is the effect of a general infinitesimal rotation on the polarization tensor $e_{ij}(\hat{n}, \pm 2)$ of a gravitational wave travelling in a direction $\hat{n}$ along the 1 -axis?

24. Suppose that the probability distribution of tensor fluctuations respected the symmetries of translation and rotation invariance, but not invariance under space inversion. What would that imply for the average $\langle\beta(\mathbf{q},\lambda)\beta^*(\mathbf{q}',\lambda')\rangle$, where $\beta(\mathbf{q},\lambda)$ is the stochastic parameter for wave number $\mathbf{q}$ and helicity λ ? Give a formula for the multipole coefficient $C^T_{EB,\ell}$ in this case.
25. Suppose that light passing a galaxy is deflected toward the center of the galaxy by an angle A/b^n, where A and n are constants, and b is the distance of closest approach of the light to the galaxy center. Find the lens equation, i.e. a relation between the observed angle β between the lines of sight from the earth to the lensing galaxy and a more distant source of light and the angle α that there would be between these lines of sight if light were not deflected by the galaxy. Give a formula for the apparent luminosity of the image of the source in terms of β.
26. Define multipole coefficients $C_{\gamma E,\ell}$ by

$$\left\langle a_{\gamma,\ell m} a^*_{E,\ell' m'}\right\rangle = \delta_{\ell\ell'}\delta_{mm'}C_{\gamma E,\ell},$$

where $a_{\gamma,\ell m}$ and $a_{E,\ell m}$ are the coefficients in the partial wave expansion of the shear and the Stokes parameters, defined in Sections 9.5 and 7.4. Give a formula for the correlation functions $\langle\gamma_1(\hat{n})Q(\hat{n}')\rangle$,
$\langle\gamma_2(\hat{n})Q(\hat{n}')\rangle, \langle\gamma_1(\hat{n})U(\hat{n}')\rangle, \langle\gamma_2(\hat{n})U(\hat{n}')\rangle$ of the shear and Stokes parameters in terms of these multiipole coefficients.
27. Assume that inflation is driven by a single inflation field φ, with a potential $V(\varphi) = g\exp(-\lambda\varphi^2)$. What are the conditions on λ and φ to allow the use of the slow-roll approximation? Assuming these conditions to be satisfied, give values for the scalar and tensor slope parameters n_S and n_T. By how many e-foldings does the universe expand when φ changes by a factor 100 ?
28. In multi-field inflation, what is the dependence of scalar field perturbations on wave number outside the horizon when you take account of a slight variation of the potential during the era of horizon exit?

11 Challenges for ΛCDM: An update by Perivolaropoulos, Skara

Abstract

A number of challenges to the standard ΛCDM model have been emerging during the past few years as the accuracy of cosmological observations improves. In this review we discuss in a unified manner many existing signals in cosmological and astrophysical data that appear to be in some tension (2σ or larger) with the standard ΛCDM model as specified by the Cosmological Principle, General Relativity and the Planck18 parameter values. In addition to the well-studied 5σ challenge of ΛCDM (the Hubble H_0 tension) and other well known tensions (the growth tension, and the lensing amplitude A_L anomaly), we discuss a wide range of other less discussed less-standard signals which appear at a lower statistical significance level than the H_0 tension some of them known as 'curiosities' in the data) which may also constitute hints towards new physics. For example such signals include cosmic dipoles (the fine structure constant α, velocity and quasar dipoles), CMB asymmetries, BAO Lyα tension, age of the Universe issues, the Lithium problem, small scale curiosities like the core-cusp and missing satellite problems, quasars Hubble diagram, oscillating short range gravity signals etc. The goal of this pedagogical review is to collectively present the current status (2022 update) of these signals and their level of significance, with emphasis on the Hubble tension and refer to recent resources where more details can be found for each signal. We also briefly discuss theoretical approaches that can potentially explain some of these signals.

Contents

11.1 Introduction

The concordance or standard Λ Cold Dark Matter (ΛCDM) cosmological model [1418, 1422, 449] is a well defined, predictive and simple cosmological model [see 391, for a review]. It is defined by a set of simple assumptions:

- The Universe consists of radiation (photons, neutrinos), ordinary matter (baryons and leptons), cold (non-relativistic) dark matter (CDM) [1912, 1913, 739, 1547, 1546, 343, 292] being responsible for structure formation and cosmological constant Λ [451, 449], a homogeneous form of energy which is responsible for the late time observed accelerated expansion. The cosmological constant is currently associated with a dark energy or vacuum energy whose density remains constant even in an expanding background [see 1422, 1393, 1394, 1819, for a review].

- General Relativity (GR) [659] is the correct theory that describes gravity on cosmological scales. Thus, the action currently relevant on cosmological scales reads

$$S = \int d^4x\sqrt{-g}\left[\frac{1}{16\pi G}(R - 2\Lambda) + \frac{1}{4\alpha}F_{\mu\nu}F^{\mu\nu} + \mathcal{L}_m(\psi, A)\right] \tag{33}$$

  where α is the fine structure constant, G is Newton's constant, $F_{\mu\nu}$ is the electromagnetic field-strength tensor and $\mathcal{L}_m$ is the Lagrangian density for all matter fields ψ_m.

- The Cosmological Principle (CP) states that the Universe is statistically homogeneous and isotropic in space and matter at sufficiently large scales ($\gtrsim 100\,Mpc$).

- There are six independent (free) parameters: the baryon $\omega_b = \Omega_{0b}h^2$ and cold dark matter $\omega_c = \Omega_{0c}h^2$ energy densities (where $h = H_0/100\ km\ s^{-1}\ Mpc^{-1}$ is the dimensionless Hubble constant and $\Omega_X \equiv \rho_X/\rho_{crit}$ is the density of component X relative to the critical density, $\rho_{crit} = 3H^2/8\pi G$), the angular diameter distance to the sound horizon at last scattering θ_s, the amplitude A_s and tilt n_s of primordial scalar fluctuations and the reionization optical depth τ.

- The spatial part of the cosmic metric is assumed to be flat described by the Friedmann-Lemaitre-Roberson-Walker (FLRW) metric

$$ds^2 = dt^2 - a(t)^2(dr^2 + r^2d\theta^2 + r^2sin^2\theta d\phi^2) \tag{34}$$

  which emerges from the CP.

  Assuming this form of the metric and Einstein's field equations with a Λ-term we obtain the Friedmann equations which may be written as

$$H^2 \equiv \frac{\dot{a}^2}{a^2} = \frac{8\pi G\rho + \Lambda c^2}{3} \tag{35}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + \frac{3p}{c^2}) + \frac{\Lambda c^2}{3} \tag{36}$$

where a is the scale factor $a = \frac{1}{1+z}$ (with z the redshift). The cosmological constant may also be viewed as a cosmic dark energy fluid with equation of state parameter

$$w = \frac{p_\Lambda}{\rho_\Lambda} = -1 \tag{37}$$

where ρ_Λ and p_Λ are the energy density and the pressure of the dark energy respectively.

- A primordial phase of cosmic inflation (a period of rapid accelerated expansion) is also assumed in order to address the horizon and flatness problems [260, 119, 170, 15]. During this period, Gaussian scale invariant primordial fluctuations are produced from quantum fluctuations in the inflationary epoch.

Fundamental generalizations of the standard ΛCDM model may be produced by modifying the defining action (33) by generalizing the fundamental constants to dynamical variables in the existing action or adding new terms. Thus the following extensions of ΛCDM emerge:

- Promoting Newton's constant to a dynamical degree of freedom by allowing it to depend on a scalar field Φ as $G \to G(\Phi(r,t))$ where the dynamics of Φ is determined by kinetic and potential terms added to the action. This class of theories is known as 'scalar-tensor theories' with its most general form with second order dynamical equations the Horndeski theories [909, 573] [see also 1014, 1078, for a comprehensive review].

- Promoting the cosmological constant to a dynamical degree of freedom by the introduction of a scalar field (quintessence) with $\Lambda \to V(\Phi(r,t))$ and the introduction of a proper kinetic term.

- Allowing for a dynamical Fine Structure Constant (Maxwell Dilaton theories) with $\alpha \to \alpha(\Phi(r,t))$ [248, 1582, 225, 224, 223] [see also 1248, for a review].

- Addition of new terms to the action which may be functions of the Ricci scalar, the torsion scalar or other invariants $((f(R), f(T), \dots))$ [260, 1360, 1362, 569, 1674, 708, 1342, 413, 435].

The ΛCDM model has been remarkably successful in explaining most properties of a wide range of cosmological observations including the accelerating expansion of the Universe [1524, 1445], the power spectrum and statistical properties of the cosmic microwave background (CMB) anisotropies [1395], the spectrum and statistical properties of large scale structures of the Universe [288, 391] and the observed abundances of different types of light nuclei hydrogen, deuterium, helium, and lithium [1605, 1683, 944, 539].

Despite of its remarkable successes and simplicity, the validity of the cosmological standard model ΛCDM is currently under intense investigation [see 30, 386, 615, 1604, 139, 1599, for a review]. This is motivated by a range of profound theoretical and observational difficulties of the model.

The most important theoretical difficulties that plague ΛCDM are the fine tuning [1819, 1242, 396] and coincidence problems [1684, 1769]. The first fundamental problem is associated with the fact that there is a large discrepancy between observations and theoretical expectations on the value of the cosmological constant Λ (at least 60 orders of magnitude) [1819, 508, 1242, 1662] and the second is connected to the coincidence between the observed vacuum energy density Ω_Λ and the matter density Ω_m which are approximately equal nowadays despite their dramatically different evolution properties. The anthropic principle has been considered as a possible solution to these problems. It states that these 'coincidences' result from a selection bias towards the existence of human life in the context of a multiverse [1692, 1818].

In addition to the above theoretical challenges, there are signals in cosmological and astrophysical data that appear to be in some tension (2σ or larger) with the standard ΛCDM model as specified by the Planck18 parameter values [72, 68]. The most intriguing large scale tensions are the following[48] [30] [see also 620, 619, for a recent overview of the main tensions]:

- **The Hubble tension ($> 4\sigma$):** (see Section 11.2) Using a distance ladder approach, the local (late or low redshift) measurements of the Hubble constant H_0 are measured to values that are significantly higher than those inferred using the angular scale of fluctuations of the CMB in the context of the ΛCDM model. Combined local direct measurements of H_0 are in 5σ tension (or more if combinations of local measurements are used) with CMB indirect measurements of H_0 [366, 596, 1519]. The Planck/ΛCDM best fit value is $H_0 = 67.4 \pm 0.5\ km\ s^{-1} Mpc^{-1}$ [72] while the local measurements using Cepheid calibrators by the SH0ES Team indicate $H_0 = 73.04 \pm 1.04\ km\ s^{-1} Mpc^{-1}$ ($\sim 5\sigma$) [1529] [see 615, 1624, 1585, for a review]. In the previous analysis by the SH0ES Team [1521] using the Gaia Early Data Release 3 (EDR3) parallaxes [751] a value of $H_0 = 73.2 \pm 1.3$ is obtained, at a 4.2σ tension with the prediction from Planck18 CMB observations. A wide range of local observations appear to be consistently larger than the Planck/ΛCDM measurement of H_0 at various levels of statistical significance [366, 596, 1519]. Theoretical models addressing the Hubble tension utilize either a recalibration of the Planck/ΛCDM standard ruler (the sound horizon) assuming new physics before the time of recombination [1011, 301, 11] or a deformation of the Hubble expansion rate $H(z)$ at late times [102, 612] or a transition/recalibration of the SnIa absolute luminosity due to late time new physics [1238] [see in 1436, for a relevant talk]. Also, for more detailed discussions of the proposed new-physics models see in [1772, 615, 1604, 139].

- **The growth tension ($2 - 3\sigma$):** (see Subsection 11.3.1) Direct measurements of the growth rate of cosmological perturbations (Weak Lensing, Redshift Space Distortions (peculiar velocities), Cluster Counts) indicate a lower growth rate than that indicated by the Planck/ΛCDM parameter values at a level of about $2 - 3\sigma$ [995, 106, 234]. In the context of General Relativity such lower growth rate would imply a lower matter density and/or a lower amplitude of primordial fluctuation spectrum than that indicated by Planck/ΛCDM [1220, 1344, 1027, 1646].

- **CMB anisotropy anomalies ($2-3\sigma$):** (see Subsection 11.3.2) These anomalies include lack of power on large angular scales, small vs large scales tension (different best fit values of cosmological parameters), cold spot anomaly, hints for a closed Universe (CMB vs BAO), anomaly on super-horizon scales, quadrupole-octopole alignment, anomalously strong ISW effect, cosmic hemispherical power asymmetry, lensing anomaly, preference for odd parity correlations, parity violating rotation of CMB linear polarization (cosmic birefringence) etc. [see 1608, 88, for a review].

- **Cosmic dipoles ($2 - 5\sigma$):** (see Subsection 11.3.3) The large scale velocity flow dipole [1808, 1017], the Hubble flow variance in the cosmic rest frame [1831], the dipole anisotropy in radio source count [263], the quasar density dipole [1614] and the fine structure constant dipole (quasar spectra) [1065, 1809] indicate that the validity of the cosmological principle may have to be reevaluated.

- **Baryon Acoustic Oscillations (BAO) curiosities ($2.5-3\sigma$):** (see Subsection 11.3.4) There is a discrepancy between galaxy and Lyman-α (Lyα) BAO at an effective redshift of $z \sim 2.34$ [532, 683, 9].

[48] We use the term 'curiosity' as a term describing a discrepancy between datasets in ΛCDM best fit parameter values at a level with a statistical significance $\lesssim 3\sigma$.

- **Parity violating rotation of CMB linear polarization (Cosmic Birefringence):** (see Subsection 11.3.5) The recent evidence of the non zero value of birefringence poses a problem for standard ΛCDM cosmology and indicates a hint of a new ingredient beyond this standard model. In particular using a novel method developed in Minami *et al.* [1294], Minami [1291], Minami and Komatsu [1293], a non-zero value of the isotropic cosmic birefringence $\beta_a = 0.35 \pm 0.14$ deg (68% C.L) was recently detected in the Planck18 polarization data at a 2.4σ statistical significance level by Minami and Komatsu [1292].

- **Small-scale curiosities:** (see Subsection 11.3.6) Observations on galaxy scales indicate that the ΛCDM model faces several problems (core-cusp problem, missing satellite problem, too big to fail problem, angular momentum catastrophe, satellite planes problem, baryonic Tully-Fisher relation problem, void phenomenon etc.) in describing structures at small scales [see 575, 393, for a review].

- **Age of the Universe:** (see Subsection 11.3.7) The age of the Universe as obtained from local measurements using the ages of oldest stars in the Milky Way (MW) appears to be marginally larger and in some tension with the corresponding age obtained using the CMB Planck18 data in the context of ΛCDM cosmology [1771].

- **The Lithium problem $(2-4\sigma)$:** (see Subsection 11.3.8) Measurements of old, metal-poor stars in the Milky Way's halo find 5 times less lithium than that BBN predicts [712].

- **Quasars Hubble diagram $(\sim 4\sigma)$:** (see Subsection 11.3.9) The distance modulus-redshift relation for the sample of 1598 quasars at higher redshift ($0.5 < z < 5.5$) is in some tension with the concordance ΛCDM model indicating some hints for phantom late time expansion [1532, 210, 1213].

- **Oscillating signals in short range gravity experiments:** (see Subsection 11.3.10) A reanalysis of short range gravity experiments has indicated the presence of an oscillating force signal with sub-millimeter wavelength [1435, 146].

- **Anomalously low baryon temperature $(\sim 3.8\sigma)$:** (see Subsection 11.3.11) The Experiment to Detect the Global Epoch of Reionization Signature (EDGES) collaboration [349] using global (sky-averaged) 21-cm absorption signal, reports anomalously low baryon temperature $T_b \approx 4K$ at $z \approx 17$ (half of its expected value).

- **Colliding clusters with high velocity $(\sim 6\sigma)$:** (see Subsection 11.3.12) The El Gordo (ACT-CL J0102-4915) galaxy cluster at $z = 0.87$ is in its formation process which occurs by a collision of two subclusters with mass ratio 3.6 merging at a very high velocity $V_{infall} \simeq 2500 km/s$. Such cluster velocities at such a redshift are extremely rare in the context of ΛCDM as demonstrated by Asencio *et al.* [172] using the estimation of Kraljic and Sarkar [1102] for the expected number of merging clusters from interrogation of the DarkSky simulations.

The well known Hubble tension and the other less discussed curiosities of ΛCDM at a lower statistical significance level may hint towards new physics [see 933, for a review].

In the context of the above observational puzzles the following strategic questions emerge

- What are the current cosmological and astrophysical datasets that include the above non-standard signals?

- What is the statistical significance of each signal?

- Is there a common theoretical framework that may explain simultaneously many non-standard signals?

These questions will be discussed in what follows. There have been previous works [1431, 1432] collecting and discussing signals in data that are at some statistical level in tension with the standard ΛCDM model but these are by now outdated and the more detailed and extended update provided by the present review may be a useful resource.

The plan of this review is the following: In the next section (11.2) we focus on the Hubble tension. We provide a list of observational probes that can lead to measurements of the Hubble constant, point out the current tension level among different probes and discuss some of the possible generic extensions of ΛCDM model that can address this tension. In section 11.3 we present the current status of other less significant tensions, their level of significance and refer to recent resources where more details can be found for each signal. We also discuss possible theoretical approaches that can explain the non-standard nature of these signals. Finally, in section 11.4 we conclude and discuss potential future directions of the reviewed research.

In Table 23 of of the Appendix 11.5 we list the acronyms used in this review.

11.2 Hubble tension

11.2.1 Methods for measuring H_0 and data

The measurement of the Hubble constant H_0 which is the local expansion rate of the Universe, is of fundamental importance to cosmology. This measurement has improved in accuracy through number of probes [see 1817, for a review of most well established probes].

Distances to cosmological objects constitute the most common way to probe the cosmic metric and the expansion history of the Universe. In this subsection we review the use of the two main cosmological distances used to probe the cosmic expansion history.

- **Luminosity distance**

  Consider a luminous cosmological source of absolute luminosity L (emitted power) and an observer (Fig. 116) at a distance d_L from the luminous source. In a static cosmological setup, the power radiated by the luminous source is conserved and distributed in the spherical shell with area $4\pi d_L^2$ and therefore the apparent luminosity l (energy flux) detected by the observer is

  $$l = \frac{L}{4\pi d_L^2} \tag{38}$$

  Eq. (38) defines the quantity d_L known as *luminosity distance*. It is straightforward to show that in an expanding flat Universe, where the energy is not conserved due to the increase of the photon wavelength and period with time, the luminosity distance can be expressed as [629, 1430]

  $$d_L(z)_{th} = c(1+z)\int_0^z \frac{dz'}{H(z')} \tag{39}$$

  The luminosity distance is an important cosmological observable that is measured using standard candles (see Subsection 11.2.1)

- **Angular diameter distance**

  Consider a source (standard ruler) with a physical scale r that subtends an angle θ in the sky (Fig. 117). In Euclidean space, assuming that θ is small, the physical angular

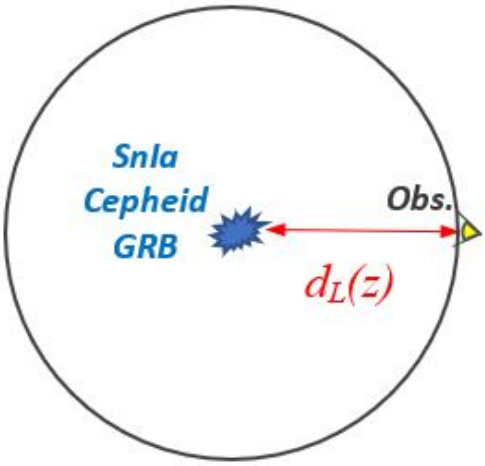


Figure 116: The luminosity distance is obtained from the apparent and absolute luminosities.

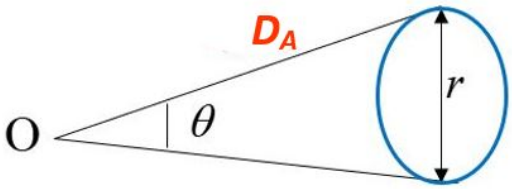


Figure 117: The angular diameter distance is obtained from the angular and physical scales.

diameter distance D_A is defined as [e.g. 629, 901]

$$D_A(z) = \frac{r}{\theta} \tag{40}$$

A particularly useful standard ruler is the sound horizon at recombination calibrated by the peaks of the CMB anisotropy spectrum and observed either directly through the CMB anisotropies or through its signatures in the large scale structure (Baryon Acoustic Oscillations (BAO)) (see Subsection 11.2.1).

It is straightforward to show that in an expanding flat Universe the physical angular diameter distance can be expressed as [e.g. 629]

$$D_A(z)_{th} = \frac{c}{(1+z)} \int_0^z \frac{dz'}{H(z')} \tag{41}$$

The luminosity and angular diameter distances can be measured using standard candles and standard rulers thus probing the cosmic expansion rate at both the present time ($H(z=0) \equiv H_0$) and at higher redshifts ($H(z)$).

Standard candles as probes of luminosity distance

The luminosity distance to a source may be probed using standardizable candles like Type Ia supernovae (SnIa) ($z < 2.3$) [1524, 1445, 293, 329] and gamma-ray bursts (GRBs) ($0.1 < z \lesssim 9$) [122, 121, 1573, 1704, 1578, 1593, 531, 1799, 577, 1701, 439, 1048, 1047, 551, 550, 627, 578, 428, 1212, 427, 1199, 918, 547, 1211].

Surveys can indicate the distance-redshift relation of SnIa by measuring their peak luminosity that is tightly correlated with the shape of their characteristic light curves (luminosity as a function of time after the explosion) and the redshifts of host galaxies. The latest and largest SnIa dataset available that incorporates data from six different surveys is the Pantheon sample consisting of a total of 1048 SnIa in the redshift range $0.01 < z < 2.26$ (the number of SnIa with $z > 1.4$ is only six) [329]. More recently, the Pantheon+ sample which comprises 18 different samples has been released [380, 1611] [see also 1448, 384]. Brout *et al.* [380], Scolnic *et al.* [1611] present 1701 light curves of 1550 distinct SnIa in the redshift range $0.001 < z < 2.26$ including SnIa which are in very nearby galaxies ($z \lesssim 0.01$) with measured Cepheid distances. For determination of H_0 the SH0ES team [1529] use as calibrator sample 42 SnIa in the 37 Cepheid hosts and 277 SnIa in the Hubble flow ($0.0233 < z < 0.15$) from the Pantheon+ sample.

The apparent magnitude[49] m_{th} of SnIa in the context of a specified form of $H(z)$, is related to their luminosity distance $d_L(z)$ of Eq. (39) in Mpc as

$$m(z)_{th} = M + 5\log_{10}\left[\frac{d_L(z)}{Mpc}\right] + 25 \tag{43}$$

Using now the dimensionless Hubble free luminosity distance

$$D_L(z) = \frac{H_0 d_L(z)}{c} \tag{44}$$

the apparent magnitude can be written as

$$m(z)_{th} = M + 5\log_{10}\left[D_L(z)\right] + 5\log_{10}\left[\frac{c/H_0}{Mpc}\right] + 25 \tag{45}$$

The use of Eq. (45) to measure H_0 using the measured apparent magnitudes of SnIa requires knowledge of the value of the SnIa absolute magnitude M which can be obtained using calibrators of local SnIa at $z < 0.01$ (closer than the start of the Hubble flow) in the context of a distance ladder [e.g. 1579] using calibrators like Cepheid stars.

In the cosmic distance ladder approach each step of the distance ladder uses parallax methods and/or the known intrinsic luminosity of a standard candle source to determine the absolute (intrinsic) luminosity of a more luminous standard candle residing in the same galaxy. Thus highly luminous standard candles are calibrated for the next step in order to reach out to high redshift luminosity distances.

SnIa standard candles and their calibration

- **SnIa-Cepheid:** Geometric anchors may be used to calibrate the Cepheid variable star standard candles at the local Universe (primary distance indicators) whose luminosities are correlated with their periods of variability[50]. The MW, the Large Magellanic Cloud (LMC) and NGC 4258 are used as distance geometric anchor galaxies. For Cepheids in the anchor galaxies there are three different ways of geometric distance calibration of their luminosities: trigonometric parallaxes in the MW [258, 1146, 452, 1520, 1187,

[49] The apparent magnitude m of an astrophysical source detected with flux l is defined as

$$m = -2.5\, log_{10}\left(\frac{l}{l_0}\right) \tag{42}$$

where l_0 is a reference flux (zero point). The absolute magnitude M of an astrophysical source is the apparent magnitude the source would have if it was placed at a distance of $10\,pc$ from the observer.

[50] The period–luminosity (PL) relation is also called the Leavitt law [1141, 1142].

1528, 311, 1521], Detached Eclipsing Binary Stars (DEBs) in the LMC [1456] and water masers (see Subsection 11.2.1) in NGC 4258 [1885, 1502]. The DEBs technique relies on surface-brightness relations and is a one-step distance determination to nearby galaxies independent from Cepheids [1457].

Using the measured distances of the calibrated Cepheid stars the intrinsic luminosity of nearby SnIa residing in the same galaxies as the Cepheids is obtained. This SnIa calibration which fixes M is then used for SnIa at distant galaxies to measure H_0 ($z \in [0.01, 0.1]$) and $H(z)$ ($z \in [0.01, 2.3]$).

- **SnIa-TRGB:** Instead of Cepheid variable stars, the Tip of the Red Giant Branch (TRGB) stars in the Hertzsprung-Russell diagram [245, 735] and Miras [924, 923] [see also 543, for a review] can be used as calibrators of SnIa. The Red Giant stars have nearly exhausted the hydrogen in their cores and have just began helium burning (helium flash phase). Their brightness can be standardized using parallax methods and they can serve as bright standard candles visible in the local Universe for the subsequent calibration of SnIa.

- **SnIa-Miras:** Miras (named for the prototype star Mira) are highly evolved low mass variable stars at the tip of asymptotic giant branch (AGB) stars [e.g. 936]. The water megamaser as distance indicator (see Subsection 11.2.1) can be used to calibrate the Mira period–luminosity (PL) relation [924]. Miras with short period (< 400 days) have low mass progenitors and are present in all galaxy types or in the halos of galaxies, eliminating the necessity for low inclination SnIa host galaxies.

- **SBF:** Another method to determine the Hubble constant based on calibration of the peak absolute magnitude of SnIa is the Surface Brightness Fluctuations (SBF) method [974, 426, 1052]. SBF is a secondary[51] luminosity distance indicator that uses stars in the old stellar populations (II) and can reach larger distances than Cepheids even inside the Hubble flow region where the recession velocity is larger than local peculiar velocities ($z > 0.01$) [1719, 1720, 313, 1277, 300, 315, 312]. For SBF calibration Blakeslee *et al.* [314] use both Cepheids and TRGB demonstrating that these calibrators are consistent with each other.

  Assume that a galaxy includes a finite number of stars covering a range of luminosity. Using SBF in the galaxy image for the determination of its distance, the ratio $\bar{L}$ of the second and first moments of the stellar luminosity function in the galaxy is used along with the mean flux per star $\bar{l}$ as follows [1719, 313]

$$d^2 = \frac{\bar{L}}{4\pi \bar{l}} \tag{46}$$

  where

$$\bar{L} = \frac{\int n(L) L^2 dL}{\int n(L) L dL} = \frac{\sigma_L^2}{\langle L \rangle} \tag{47}$$

  where $n(L)$ is the expectation number of stars with luminosity L. Thus SBF can be viewed as providing an average brightness. A galaxy with double distance appears with double smoothness due to the effect of averaging.

[51] Nearby Cepheids or stellar population models are used for the empirical or theoretical calibration of the SBF distances respectively.

Alternative cosmological standard candles

- **SneII:** An independent method to determine the Hubble constant utilizes Type II supernovae (SneII) as cosmic distance indicators [959]. SneII are characterised by the presence of hydrogen lines in their spectra [713, 714]. This feature distinguishes SneII from other types of supernovae. Their light curve shapes include a plateau of varying steepness and length differ significantly from those of SnIa. The use of SneII as standard candles is motivated by the fact that they are more abundant than SnIa [1168, 817] [although 1-2 mag fainter 1517] and are produced by different stellar populations than SnIa which are more difficult to standardize. The SneII progenitors (red super giant stars) however are better understood than those of SnIa.

  Different SneII distance-measurement techniques have been proposed and tested. These include, the expanding photosphere method [1067, 649, 585], the spectral-fitting expanding atmosphere method [221, 586], the standardized candle method [843], the photospheric magnitude method [1534] and the photometric color method [960]. For example, the standardized candle method is based on the relation between the luminosity and the expansion velocity of the photosphere [843, 1385, 961, 962].

- **GRBs:** In addition to SnIa and SneII, GRBs are widely proposed as standard candles to trace the Hubble diagram at high redshifts [1132, 235, 1795, 1048, 1047]. However GRBs distance calibration is not easy and various cosmology independent methods [e.g. 1195] or phenomenological relations [e.g. 123, 782] have been proposed for their calibration.

  Furthermore GRBs can be combined with other probes to study the redshift evolution of Hubble constant [553] [see 1309, for a review].

Using SnIa to measure H_0 and $H(z)$:

The best fit values of the parameter H_0 and the deceleration parameter q_0 may be obtained[52] [420] using local distance ladder measurements (e.g. Cepheid calibration up to $z \simeq 0.01$) to measure directly M, low z measurements of the SnIa apparent magnitude $m(z)$ and a kinematic local expansion of $D_L(z)$ ($z < 0.1$) as [e.g. 1820]

$$D_L(z, q_0) = z\left[1 + \frac{1}{2}(1 - q_0)z\right] \tag{48}$$

Alternatively, q_0 may be fixed to its ΛCDM value $q_0 = -0.55$ and H_0 may be fit as a single parameter [308, 1526, 306].

Using higher z SnIa the best fit parameters of ΛCDM may be obtained by fitting the ΛCDM expansion rate $H(z)$

$$H^2(z) = H_0^2\left[\Omega_{0m}(1+z)^3 + (1 - \Omega_{0m})\right] \tag{49}$$

where Ω_{0m} is the matter density parameter today. Using Eqs. (39), (44) and (49), the Hubble free luminosity distance can be written as

$$D_L(z, \Omega_{0m}) = (1+z)\int_0^z \frac{dz'}{\left[\Omega_{0m}(1+z')^3 + (1 - \Omega_{0m})\right]^{1/2}} \tag{50}$$

A key assumption in the use of SnIa in the measurement of H_0 and $H(z)$ is that they are standardizable and after proper calibration they have a fixed absolute magnitude independent of

[52] q_0 is the deceleration parameter $q_0 \equiv -\frac{1}{H_0^2}\frac{d^2a(t)}{dt^2}\Big|_{t=t_0}$

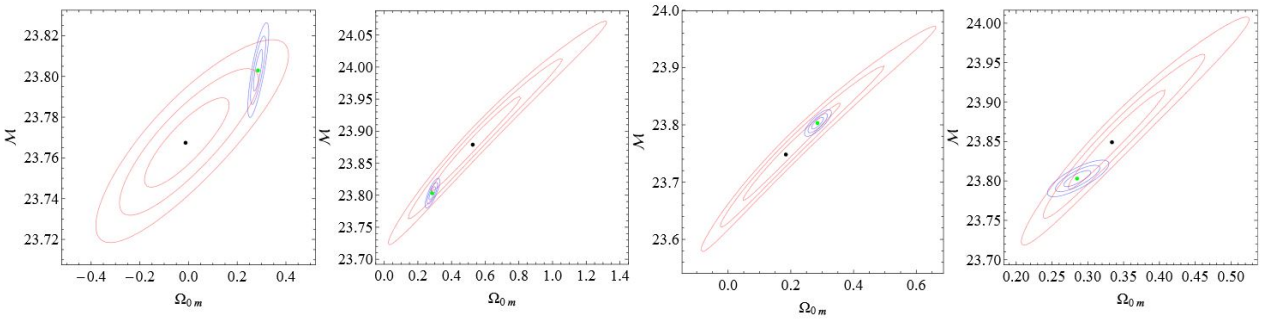


Figure 118: The $1\sigma - 3\sigma$ confidence contours in the parametric space (Ω_{0m}, $\mathcal{M}$). The blue contours correspond to the $1\sigma - 3\sigma$ full Pantheon dataset (1048 SnIa datapoints) best fit, while the red contours describe the $1\sigma - 3\sigma$ confidence contours of the four bins (from left to right). The black points represent the best fit of each bin, while the green dot represents the best fit value indicated by the full Pantheon dataset ($\Omega_{0m} = 0.285$ and $\mathcal{M} = 23.803$) [from 1029].

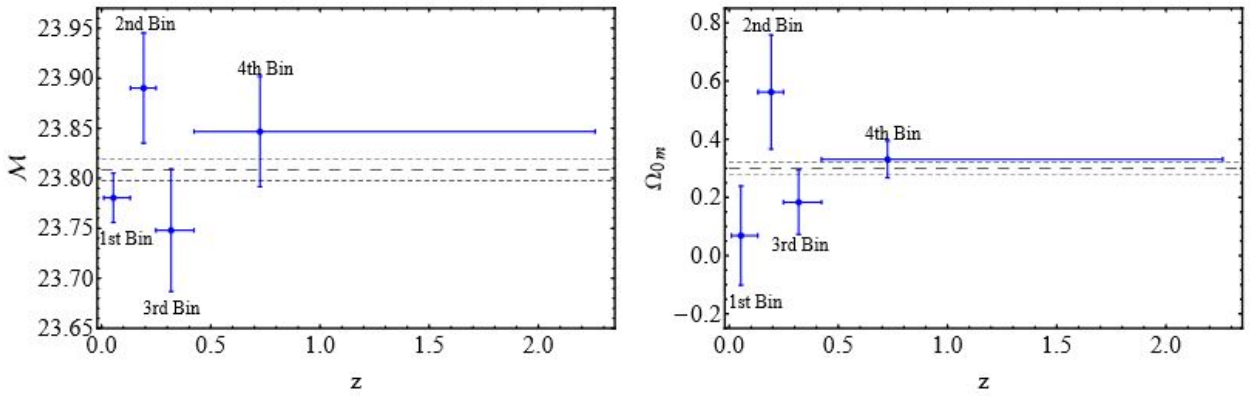


Figure 119: The best fit values of $\mathcal{M}$ (left panel) and Ω_{0m} (right panel) as well as the 1σ errors for the four bins, including the systematic uncertainties. This oscillating behaviour relatively improbable in the context of constant underlying $\mathcal{M}$ and Ω_{0m} [from 1029].

redshift[53]. This assumption has been tested in Colgáin [494], Kazantzidis and Perivolaropoulos [1028, 1029], Sapone *et al.* [1584], Koo *et al.* [1095], Kazantzidis *et al.* [1026], Luković *et al.* [1208], Tutusaus *et al.* [1743, 1744], Drell *et al.* [636].

Using the degenerate combination

$$\mathcal{M} = M + 5\log_{10}\left[\frac{c/H_0}{Mpc}\right] + 25 \tag{51}$$

into Eq. (45) we obtain

$$m(z, M, H_0, \Omega_{0m})_{th} = \mathcal{M}(M, H_0) + 5\log_{10}\left[D_L(z, \Omega_{0m})\right] \tag{52}$$

The theoretical prediction (52) may now be used to compare with the observed m_{obs} data and to obtain the best fits for the parameters $\mathcal{M}$ and Ω_{0m}. Using the maximum likelihood analysis the best fit values for these parameters may be found by minimizing the quantity

$$\chi^2(\mathcal{M}, \Omega_{0m}) = \sum_i \frac{\left[m_{obs,i} - m_{th}(z_i; \mathcal{M}, \Omega_{0m})\right]^2}{\sigma_i^2} \tag{53}$$

The results from the recent analysis by Kazantzidis and Perivolaropoulos [1029] using the latest SnIa (Pantheon) data [329] (consisting of 1048 datapoints in the redshift range $0.01 < z < 2.3$ sorting them from lowest to highest redshift and dividing them in four equal uncorrelated bins) in the context of a ΛCDM model are shown in Figs. 118 and 119[54]. An oscillating signal for $\mathcal{M}$ and Ω_{0m} (2σ) is apparent in Fig. 119 and its statistical significance may be quantified using simulated data [1026, 1095].

The presence of large scale inhomogeneities at low z including voids or a supercluster [814] can be a plausible physical explanation for this curious behavior. In the context of a local void model the analysis by Kazantzidis and Perivolaropoulos [1029] indicated that the value of H_0 increases by $2-3\%$ which is less than the 9% required to address the Hubble tension. The bias and systematics induced by such inhomogeneities on the Hubble diagram within a well-posed fully relativistic framework [light cone averaging formalism 765] has been discussed in Fanizza *et al.* [689].

Marra and Perivolaropoulos [1238] have pointed out that this H_0 tension is related to the mismatch between the SnIa absolute magnitude calibrated by Cepheids at $z < 0.01$ [420, 421]

$$M^{<} = -19.2334 \pm 0.0404 \tag{54}$$

and the SnIa absolute magnitude using the parametric free inverse distance ladder calibrating SnIa absolute magnitude using the sound horizon scale [419]

$$M^{>} = -19.401 \pm 0.027 \tag{55}$$

Thus a transition in the absolute magnitude with amplitude $\Delta M \simeq 0.2$ may provide a solution to this tension [see Subsection 11.2.2 and in 103, 1238]. Alternatively if this discrepancy is not due to systematics [135, 1772], it could be an indication of incorrect estimate of the sound horizon scale due e.g. to early dark energy [480] or to late phantom dark energy [102].

Note also that Rameez and Sarkar [1492] find discrepancies between 'Joint Light-curve Analysis' (JLA) SnIa and Pantheon SnIa datasets which imply an uncertainty in the calibration of the absolute magnitude or equivalently of the Hubble constant which is large enough to undermine the claim for Hubble tension.

[53]The possibility for intrinsic luminosity evolution of SnIa with redshift was first pointed out by Tinsley [1713]. Also, the assumption that the luminosity of SnIa is independent of host galaxy properties (e.g. host age, host morphology, host mass) and local star formation rate has been discussed in Kang *et al.* [1006], Rose *et al.* [1535], Jones *et al.* [987], Rigault *et al.* [1530], Kim *et al.* [1064].

[54]For $M = -19.24$ as indicated by Cepheid calibrators [420] of SnIa at $z < 0.01$ and the SnIa local determination $H_0 = 74\,km\,s^{-1}Mpc^{-1}$ [306] Kazantzidis and Perivolaropoulos [1029] find $\mathcal{M} = 23.80$ which is consistent with the full Pantheon SnIa best fit shown in Fig. 118.

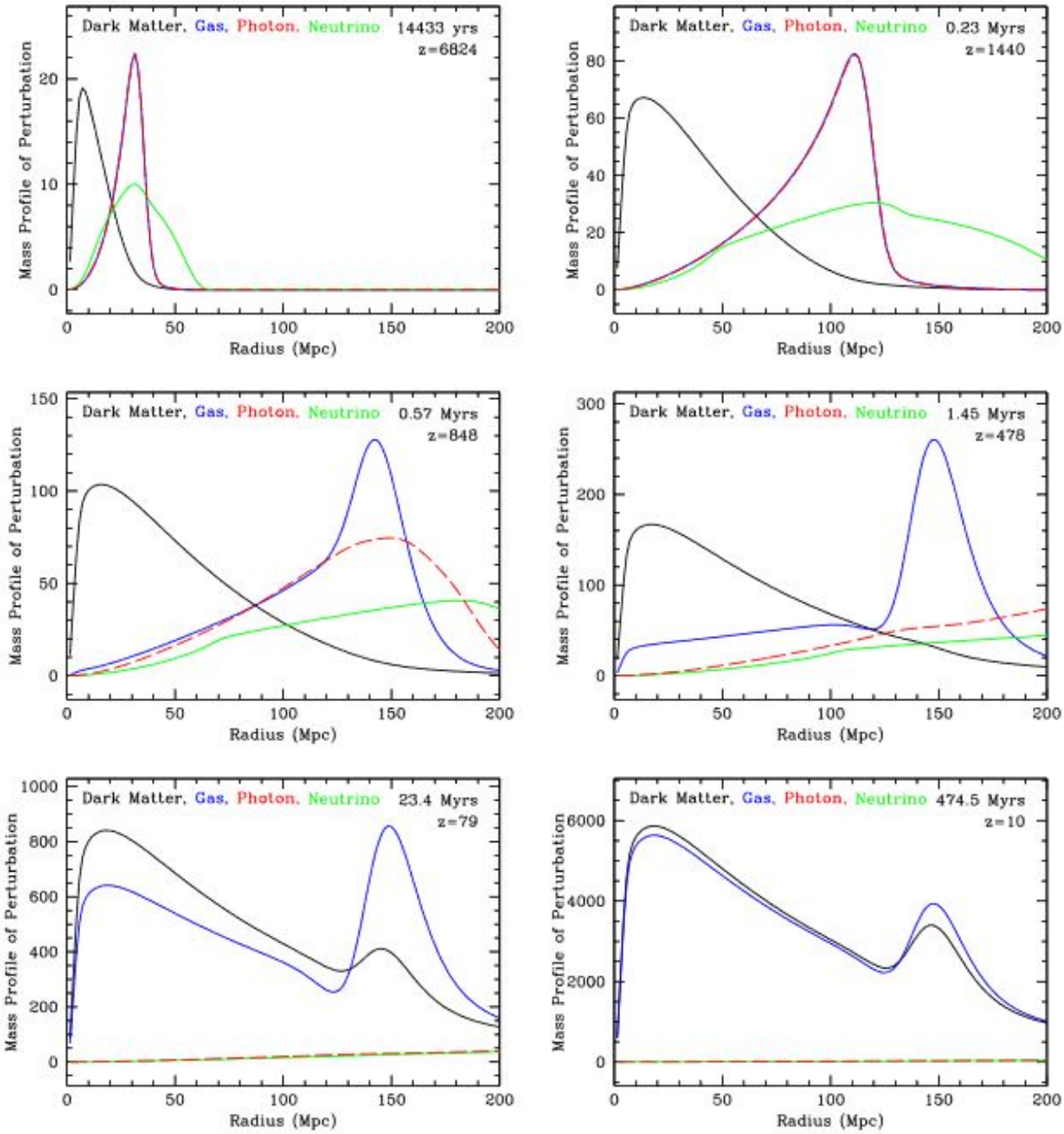


Figure 120: The snapshots show the radial mass profile of perturbation as a function of the comoving radius of an initially point-like overdensity located at the origin for redshifts $z = 6824, 1440, 848, 478, 79, 10$. The time after the Big Bang are given in each snapshots. The black, blue, red, and green lines correspond to the dark matter, baryons, photons, and neutrinos (all perturbations are fractional for that species), respectively. The top snapshots are for the early time before recombination where the overdensities in photons and baryons evolve together, the middle snapshots for soon after but close to recombination where the baryons freeze at the location reached with the photons forming a thick spherical shell, and the bottom snapshots are for long after recombination where the baryon overdensities start to gravitationally grow like dark matter overdensities [from 662].

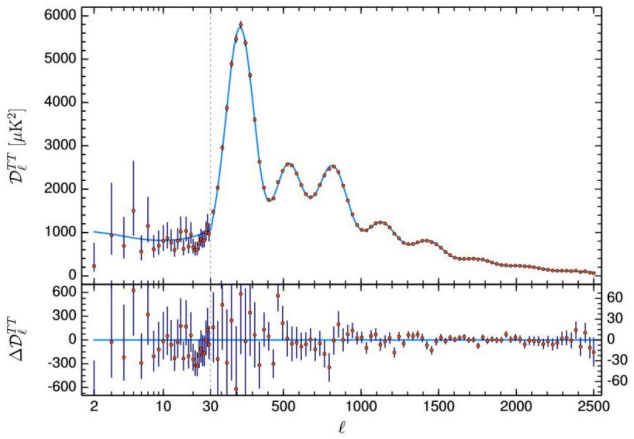


Figure 121: The Planck18 CMB angular power spectrum $\mathcal{D}_l^{TT} \equiv l(l+1)/(2\pi)C_l^{TT}$ (top) and residual angular power spectrum (bottom) of temperature fluctuations as a function of multipole moment l. The light blue line in the upper panel is the best-fitting to the Planck TT, TE, EE+lowE+lensing likelihoods assuming the base-ΛCDM cosmology. The red points correspond to the binned Planck data. The lowest multipole range ($l \leq 30$) is dominated by cosmic variance (approximated as Gaussian), while positions and amplitudes of the acoustic peaks are accurately constrained [from 72].

Observational data - Constraints:

- **SnIa-Cepheid:** Using the analysis of the Hubble Space Telescope (HST) observations [1579] the Hubble constant H_0 value has been measured from Cepheid-calibrated supernovae (using 70 long-period Cepheids in the LMC) by the Supernovae H_0 for the Equation of State (SH0ES) of dark energy collaboration [308, 1526, 306]. The analysis by the SH0ES Team using this local model-independent measurement refers $H_0 = 73.04 \pm 1.04\ km\ s^{-1}Mpc^{-1}$ [1529], which results in 5σ tension with the value estimated by CMB Planck18 [72] assuming the ΛCDM model while in previous analysis by SH0ES Team [1521] using the Gaia Early Data Release 3 (EDR3) parallaxes [751] and reaching 1.8% precision by improving the calibration a value of $H_0 = 73.2 \pm 1.3$ is obtained, a 4.2σ tension with the prediction from Planck18 CMB observations. Riess *et al.* [1528] analysing the HST data, using Cepheids as distance calibrators reports $H_0 = 73.48 \pm 1.66\ km\ s^{-1}Mpc^{-1}$. A reanalysis of the SH0ES collaboration results using a cosmographic method allowing also the deceleration parameter q_0 to be a free parameter by [421] leads to $H_0 = 74.30 \pm 1.45\ km\ s^{-1}Mpc^{-1}$.

  Breuval *et al.* [370] considered companion and average cluster parallaxes instead of direct Cepheid parallaxes and obtained $H_0 = 72.8 \pm 1.9\,(statistical + systematics) \pm 1.9$ (ZP) $km\ s^{-1}Mpc^{-1}$ when all Cepheids are considered and $H_0 = 73.0 \pm 1.9\,(statistical + systematics) \pm 1.9$ (ZP) $km\ s^{-1}Mpc^{-1}$ for fundamental mode pulsators only (where ZP is the second Gaia data release (GDR2) [381] parallax zero point).

  Various other previous estimates of H_0 have been obtained by treatments of the distance ladder [591, 397, 698]. In particular, Dhawan *et al.* [591] find $H_0 = 72.8 \pm 1.6$ (statistical)± 2.7 (systematic) $km\ s^{-1}Mpc^{-1}$ using SnIa as standard candles in the near-infrared (NIR), Burns *et al.* [397] find $H_0 = 73.2 \pm 2.3\ km\ s^{-1}Mpc^{-1}$ analysing the final data release of the Carnegie Supernova Project[55] (CSP) I [1108] and Feeney *et al.* [698] find $H_0 = 73.15 \pm 1.78\ km\ s^{-1}Mpc^{-1}$ using a Bayesian hierarchical model of the local

[55] https://csp.obs.carnegiescience.edu

distance ladder.

- **SnIa-TRGB:** The Carnegie–Chicago Hubble Program[56] (CCHP) [245] using calibration of SnIa with the TRGB method estimates $H_0 = 69.8 \pm 0.8\,(\pm 1.1\%stat) \pm 1.7\,(\pm 2.4\%sys)$ $km\ s^{-1}Mpc^{-1}$ [738] and a revision of their measurements has lead to $H_0 = 69.6 \pm 0.8\,(\pm 1.1\%stat) \pm 1.7\,(\pm 2.4\%sys)$ $km\ s^{-1}Mpc^{-1}$ [735]. Recently, the updated TRGB calibration applied to a distant sample of SnIa from the CCHP lead to a value of the Hubble constant of $H_0 = 69.8 \pm 0.6\,(\text{stat}) \pm 1.6\,(\text{sys})\,\text{km s}^{-1}\,\text{Mpc}^{-1}$ [734]. Using the LMC and the NGC 4258 as TRGB calibration of the SnIa distance ladder, the SH0ES team finds $H_0 = 72.4 \pm 2$ $km\ s^{-1}Mpc^{-1}$ [1886] and $H_0 = 71.1 \pm 1.9\,\text{km s}^{-1}\,\text{Mpc}^{-1}$ [1502] respectively. Freedman [734], Freedman *et al.* [735], Hoyt [915] argue that the difference in the derived value of H_0 by SH0ES team compared to CCHP was due to incorrect assumptions regarding calibration of the TRGB in the LMC made by Yuan *et al.* [1886]. A value of $H_0 = 65.8 \pm 3.5\,(\text{stat}) \pm 2.4\,(\text{sys})\,\text{km s}^{-1}\,\text{Mpc}^{-1}$ is obtained by Kim *et al.* [1063] using peculiar velocities and TRGB distances of 33 galaxies located between the Local Group and the Virgo cluster ($\sim 16.5\,Mpc$) [mainly the sample of Virgo infall galaxies from 1009].

  More recently, Soltis *et al.* [1672] have reported a measurement of $H_0 = 72.1 \pm 2.0$ $km\ s^{-1}Mpc^{-1}$ using the TRGB distance indicator calibrated from the European Space Agency (ESA) Gaia mission Early Data Release 3 (EDR3) trigonometric parallax of Omega Centauri [751]. Anand *et al.* [133] find $H_0 = 71.5 \pm 1.8\,km\,s^{-1}\,Mpc^{-1}$ combining TRGB measurements with either the Pantheon or CSP samples of supernova. Finally, Jones *et al.* [988] using NIR only cosmological analysis and TRGB distances to calibrate the SnIa luminosity of the CSP and RAISIN (an anagram for "SnIa in the IR") samples [986, 379] and Dhawan *et al.* [592] using TRGB calibration of SnIa observed by the Zwicky Transient Facility (ZTF) [253, 813] report $H_0 = 72.4 \pm 3.3\,\text{km s}^{-1}\,\text{Mpc}^{-1}$ and $H_0 = 76.94 \pm 6.4\,\text{km s}^{-1}\,\text{Mpc}^{-1}$ respectively.

- **SnIa-Miras:** Calibration of SnIa in the host NGC 1559 galaxy with the Miras method using a sample of 115 oxygen-rich Miras[57] discovered in maser host NGC 4258 galaxy, has lead to a measurement of the Hubble constant as $H_0 = 73.3 \pm 4$ $km\ s^{-1}Mpc^{-1}$ [923].

- **SBF:** Calibrating the SnIa luminosity with SBF method and extending it into the Hubble flow by using a sample of 96 SnIa in the redshift range $0.02 < z < 0.075$, extracted from the Combined Pantheon Sample has lead to the measurement $H_0 = 70.50 \pm 2.37$ (stat)$\pm$3.38 (sys) $km\,s^{-1}Mpc^{-1}$ by Khetan *et al.* [1052]. Previously Cantiello *et al.* [426] combining distance measurement with the corrected recession velocity of NGC 4993 reported a Hubble constant $H_0 = 71.9 \pm 7.1\,km\,s^{-1}Mpc^{-1}$. A new measurement of the Hubble constant $H_0 = 73.3 \pm 0.7 \pm 2.4 km\,s^{-1}Mpc^{-1}$ has recently been obtained based on a set of 63 SBF [314] distances extending out to $100\,Mpc$.

- **SneII:** SneII have also been used for the determination of H_0. Using 7 SneII as cosmological standardisable candles with host-galaxy distances measured from Cepheid variables or the TRGB the Hubble constant was measured to be $H_0 = 75.8^{+5.2}_{-4.9}\,km\,s^{-1}Mpc^{-1}$ [959]. More recently, de Jaeger *et al.* [958] find $H_0 = 75.4^{+3.8}_{-3.7}\,km\,s^{-1}\,Mpc^{-1}$ using 13 SneII.

Sound horizon as standard ruler: early time calibrators

Before recombination ($z > 1100$), the primeval plasma of coupled baryons to photons (baryon-photon fluid) oscillates as spherical sound waves emanating from baryon gas perturbations are driven by photon pressure. At recombination when the Universe has cooled enough the

[56] https://carnegiescience.edu/projects/carnegie-hubble-program
[57] Miras can be divided into oxygen- and carbon-rich Miras.

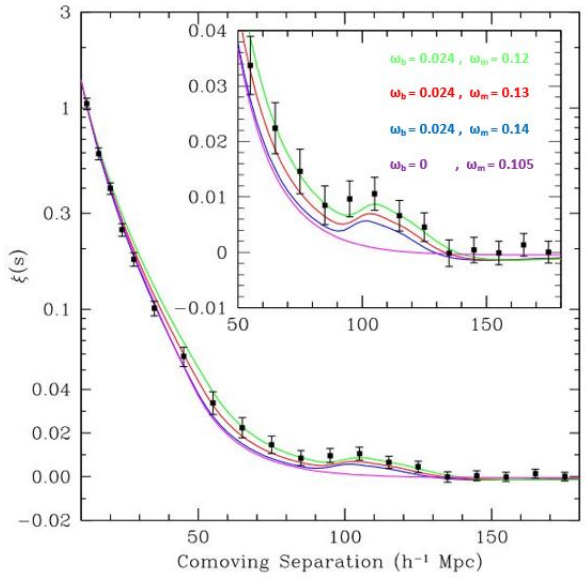


Figure 122: The signature of baryonic acoustic oscillations in galaxy two-point correlation function $\xi(s)$ as measured by Eisenstein *et al.* [663] using the luminous red galaxy samples of the Sloan Digital Sky Survey. The data show the existence of a baryonic acoustic peak in the galaxy correlation function $\xi(s)$ around the comoving separation scale $100\,h^{-1}Mpc$. The solid green, red, and blue lines correspond to model predictions with $\Omega_{0m}h^2 = 0.12, 0.13$ and 0.14, respectively. All models are taken to have the same $\Omega_{0b}h^2 = 0.024$ and $n = 0.98$. The magenta line corresponds to a model with no baryons and $\Omega_{0m}h^2 = 0.105$, which has no acoustic peaks [from 663].

electrons and protons combine to form hydrogen [e.g. 1415], photons decouple from baryons and propagate freely since the pressure becomes negligible. Thus the spherical sound wave shells of baryons become frozen. This process which was first detected in the galaxy power spectrum by Cole *et al.* [493], Eisenstein *et al.* [663] is illustrated in Fig. 120. It inflicts a unique Baryon Acoustic Oscillations (BAO) scale on the CMB anisotropy spectrum peaks shown in Fig. 121 and on the matter large scale structure (LSS) power spectrum on large scales at the radius of the sound horizon (the distance that the sound waves have traveled before recombination). This scale emerges as a peak in the correlation function $\xi(s)$[58] as illustrated in Fig. 122 or equivalently as damped oscillations in the LSS power spectrum [661, 663, 1278, 1254]. The characteristic BAO scale is also imprinted in the Lyman-α (Lyα) forest absorption lines of neutral hydrogen in the intergalactic medium (IGM) detected in quasar (QSO) spectra.

The measured angular scale of the sound horizon θ_s at the drag epoch when photon pressure vanishes can be used to probe the Hubble expansion rate using the standard ruler relation [e.g. 1417, 126]

$$\theta_s = \frac{r_s}{d_A} \tag{57}$$

where $d_A \equiv \frac{D_A}{a} \equiv (1+z)\,D_A = c\int_0^z \frac{dz'}{H(z')}$ is the comoving angular diameter distance to last

[58]The correlation function is defined as the excess probability of one galaxy to be found within a given distance of another. Using the Landy-Szalay estimator [1136] this function can be computed [663]

$$\xi(s) \equiv \langle \delta(x)\delta(x+s)\rangle = \frac{DD(s) - 2DR(s) + RR(s)}{RR(s)} \tag{56}$$

where s is the comoving galaxy separation distance and $DD(s)$, $RR(s)$ and $DR(s)$ correspond to the number of galaxy pairs with separations s in real-real, random-random and real-random catalogs, respectively.

scattering (at redshift $z \approx 1100$) and r_s is the radius of sound horizon at last scattering.

The radius r_s of the sound horizon at last scattering can be calculated by the distance that sound can travel from the Big Bang, $t = 0$, to time t_d at the drag epoch when the photon pressure can no longer prevent gravitational instability in baryons. This happens shortly after the time t_s of the last scattering when the optical depth due to Thomson scattering reaches unity [661]. Thus [29]

$$r_s = \int_0^{t_d} \frac{c_s(a)}{a(t)} dt = \int_{z_d}^{\infty} \frac{c_s(z)}{H(z;\rho_b,\rho_\gamma,\rho_c)} dz = \\ = \int_0^{a_d} \frac{c_s(a)}{a^2 H(a;\rho_b,\rho_\gamma,\rho_c)} da \tag{58}$$

where the drag redshift z_d corresponds to time t_d, ρ_b, ρ_c and ρ_γ denote the densities for baryon, cold dark matter and radiation (photons) respectively and c_s is the sound speed in the photon-baryon fluid given by [653, 157]

$$c_s = \frac{c}{\sqrt{3\left(1+\frac{3\rho_b}{4\rho_\gamma}\right)}} = \frac{c}{\sqrt{3\left(1+\frac{3\omega_b}{4\omega_\gamma}a\right)}} \tag{59}$$

The expansion rate $H(z)$ depends on the ratio of the matter density to radiation density and the sound speed determined by the baryon-to-photon ratio. Both the matter-to-radiation ratio and the baryon-to-photon ratio can be estimated from the details of the acoustic peaks in CMB anisotropy power spectrum [e.g. 921]. Thus the CMB is possible to lead to an independent determination of the radius of sound horizon. Alternatively an independent determination of the radius of sound horizon can obtained using primordial deuterium measurements [49, 9]. Now using the Eqs. (41) and (57) we can write the angular size of the sound horizon as

$$\theta_s = \frac{H_0 r_s}{c\int_0^{z_d} \frac{dz'}{E(z')}} \tag{60}$$

where $E(z)$ is the dimensionless normalized Hubble parameter and for a flat ΛCDM model is given by

$$E(z) \equiv \frac{H(z)}{H_0} = \left[\Omega_{0m}(1+z)^3 + (1-\Omega_{0m})\right]^{1/2} \tag{61}$$

Eq. (60) indicates that there is a degeneracy between r_s, H_0 and $E(z)$. Thus H_0 can not be derived using the BAO data alone which constrain $E(z)$ and the degeneracy is broken when r_s is fixed using either CMB power spectra [373] or deuterium abundance [49, 9].

For example $r_s = 147.05 \pm 0.30 Mpc$ is inferred from Planck18 TT,TE,EE+lowE CMB data [72]. Using the independent determination of r_s, measuring the angular acoustic scale θ_s from the location of the first acoustic peak in the CMB spectrum and fitting the integral in Eq. (60) using low z BAO or SnIa data, the Hubble constant H_0 can be derived. This is the 'inverse distance ladder' approach [29, 88, 407] which uses the sound horizon scale calibrated by the CMB peaks or by Big Bang Nucleosynthesis (BBN) [1600] instead of the SnIa absolute magnitude M calibrated by Cepheid stars to obtain H_0.

The deformation of the expansion rate $H(a)$ before recombination using additional components like early dark energy that increase $H(a)$ in Eq. (58) and thus decrease r_s and increase the predicted value of H_0 for fixed measured θ_s in Eq. (60), has been used as a possible approach to the solution of the Hubble tension. A challenge for this class of models is the required fine-tuning so that the evolution of $H(z)$ returns quickly to its standard form after recombination for consistency with lower z cosmological probes and growth measurements [1468]. The assumed increase of $H(z)$ at early times has been claimed to lead to a worsened growth tension [971] as discussed below even though the issue is under debate [1651, 481].

Observational data - Constraints

- **CMB:** The measurement of the Hubble constant H_0 using the sound horizon at recombination as standard ruler calibrated by the CMB anisotropy spectrum is model dependent and is based on assumptions about the nature of dark matter and dark energy as well as on an uncertain list of relativistic particles [see 459, for a review]. The best fit value obtained by the Planck18/ΛCDM CMB temperature, polarization, and lensing power spectra is $H_0 = 67.36 \pm 0.54\ km\ s^{-1}Mpc^{-1}$ [72]. The measurements of the CMB from the combination Atacama Cosmology Telescope (ACT)[59] and Wilkinson Microwave Anisotropy Probe (WMAP) estimated the Hubble constant to be $H_0 = 67.6 \pm 1.1\,km\,s^{-1}Mpc^{-1}$ and from ACT alone to be $H_0 = 67.9 \pm 1.5\,km\,s^{-1}Mpc^{-1}$ [80]. Note that the analysis of the nine-year data release of WMAP [895] alone prefers a value for the Hubble constant $H_0 = 70.0 \pm 2.2\,km\,s^{-1}Mpc^{-1}$. More recently, Balkenhol *et al.* [205] obtain CMB-based constraints on Hubble parameter $H_0 = 67.49 \pm 0.53\,km\,s^{-1}Mpc^{-1}$ using combined South Pole Telescope[60] (SPT), Planck, and ACT DR4 datasets. Dutcher *et al.* [643] find $H_0 = 68.8 \pm 1.5\,km\,s^{-1}Mpc^{-1}$ using SPT-3G data alone, while a previous analysis of SPT data by Henning *et al.* [876] results in $H_0 = 71.3 \pm 2.1\,km\,s^{-1}Mpc^{-1}$.

- **BAO:** The analysis of the wiggle patterns of BAO is an independent way of measuring cosmic distance using the CMB sound horizon as a standard ruler. This measurement has improved in accuracy through a number of galaxy surveys which detect this cosmic distance scale: the Sloan Digital Sky Survey (SDSS) supernova survey [1882, 1707] encompassing the Baryon Oscillation Spectroscopic Survey (BOSS) which has completed three different phases [566]. Its fourth phase (SDSS-IV) [317] encompasses the Extended Baryon Oscillation Spectroscopic Survey (eBOSS) [567] [see also 94, 792, 1350, 1491, 911], the WiggleZ Dark Energy Survey [638, 308, 53], the 2-degree Field Galaxy Redshift Survey (2dFGRS) [500, 493], the 6-degree Field Galaxy Survey (6dFGS) [985, 47, 48].

  More recently, BAO measurements have been extended in the context of quasar redshift surveys and Lyα absorption lines of neutral hydrogen in the IGM detected in QSO spectra using the complete eBOSS survey. The measurement of BAO scale using first the auto-correlation of Lyα function [576, 96, 37] and then the Lyα-quasar cross-correlation function [724, 57] or both the auto- and cross-correlation functions [346] pushed BAO measurements to higher redshifts ($z \sim 2.4$). Recent studies present BAO measurements from the Lyα using the eBOSS sixteenth data release (DR16) [78] of the SDSS IV [e.g. 346].

  As discussed in subsection 11.2.1 BAO data alone cannot constrain H_0 because BAO observations measure the combination $H_0 r_s$ rather than H_0 and r_s individually (where r_s is the radius of sound horizon). Using the CMB calibrated physical scale of the sound horizon and the combination of BAO with SnIa data (i.e inverse distance ladder) the value $H_0 = 67.3 \pm 1.1\ km\ s^{-1}Mpc^{-1}$ was reported which is in agreement with the value obtained by CMB data alone [29]. The analysis by Wang *et al.* [1805] using a combination of BAO measurements from 6dFGS [47], Main Galaxy Sample (MGS) [313], BOSS DR12 and eBOSS DR14 quasar sample in a flat ΛCDM cosmology reports $H_0 = 69.13 \pm 2.34\ km\ s^{-1}Mpc^{-1}$. Using BAO measurements and CMB data from WMAP, Zhang and Huang [1901] reported the constraints of $H_0 = 68.36^{+0.53}_{-0.52}$. The analysis by Addison et al. [9] combining galaxy and Lyα forest BAO with a precise estimate of the primordial deuterium abundance (BBN) results in $H_0 = 66.98 \pm 1.18\,km\,s^{-1}Mpc^{-1}$ for the flat ΛCDM model. Alam *et al.* [95] find $H_0 = 67.35 \pm 0.97\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ using BOSS galaxy and eBOSS,

[59] https://act.princeton.edu
[60] https://pole.uchicago.edu

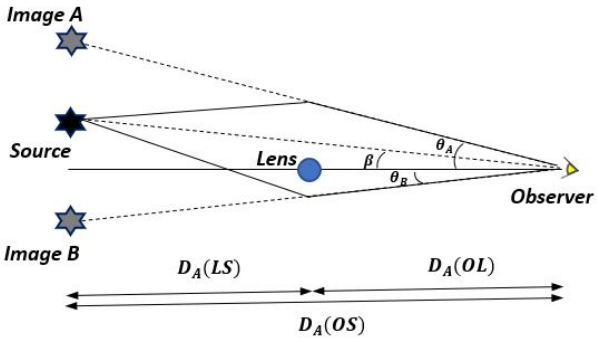


Figure 123: Schematic illustration of a typical gravitational lens system.

with the BBN prior independent from the CMB anisotropies. D'Amico *et al.* [556] obtain $H_0 = 68.5 \pm 2.2\, km\, s^{-1} Mpc^{-1}$ performing a analysis for the cosmological parameters of the DR12 BOSS data using the Effective Field Theory of Large-Scale Structure (EFTofLSS) formalism[61] and Colas *et al.* [492] obtain $H_0 = 68.7 \pm 1.5\, km\, s^{-1} Mpc^{-1}$ assuming a BBN prior on the baryon fraction of the energy density instead of the baryon/dark-matter ratio.

Recently, Pogosian *et al.* [1469] report the constraints of $H_0 = 69.6 \pm 1.8\, km\, s^{-1} Mpc^{-1}$ using BAO data, including the released eBOSS DR16, and CMB data from Plank. Zhang *et al.* [1898] infer $H_0 = 68.19 \pm 0.99$ imposing BBN priors on the baryon density and combining the BOSS Full Shape with the BAO measurements from BOSS and eBOSS. Also, a new analysis of galaxy 2-point functions in the BOSS survey, including full-shape information and post-reconstruction BAO by Chen *et al.* [468] results in $H_0 = 69.23 \pm 0.77\, km\, s^{-1} Mpc^{-1}$ and a full-shape analysis of BOSS DR12 by Philcox and Ivanov [1453] results in $H_0 = 68.31^{+0.83}_{-0.86}\, km\, s^{-1} Mpc^{-1}$. A previous analysis of BOSS DR12 on anisotropic galaxy clustering in Fourier space by Ivanov *et al.* [953] gives $H_0 = 67.9 \pm 1.1\, km\, s^{-1}\, Mpc^{-1}$. Finally, analyzing the BOSS DR12 galaxy power spectra using a new approach based on the horizon scale at matter-radiation equality Farren *et al.* [694] find $H_0 = 69.5^{+3.0}_{-3.5}\, km\, s^{-1} Mpc^{-1}$ and adding Planck lensing Philcox *et al.* [1454] find $H_0 = 70.6^{+3.0}_{-5.4}\, km\, s^{-1} Mpc^{-1}$.

Time delays: gravitational lensing

Gravitational lensing time-delay cosmography can be used to measure H_0. This approach was first proposed by Refsdal [1500] and recently implemented by Shajib *et al.* [1625], Wong et al. [366], Birrer *et al.* [299] [see also 1724, 1693, for clear reviews]. Strong gravitational lensing [1500] arises from the gravitational deflection of light rays of a background source when an intervening lensing mass distribution (e.g. a massive galaxy or cluster of galaxies) exists along the line of sight. The light rays go through different paths such that multiple images of the background source appear around the intervening lens [1381].

The time delay Δt_{AB} between two images $\theta_\mathbf{A}$ and $\theta_\mathbf{B}$ by a single deflector originating from the same source at angle β shown in Fig. 123 is given as [1694]

$$\Delta t_{AB} = \frac{1 + z_L}{c} \frac{D_A(OL) D_A(OS)}{D_A(LS)} \left[\phi(\theta_\mathbf{A}, \beta) - \phi(\theta_\mathbf{B}, \beta)\right] \tag{62}$$

where z_L is the lens redshift, $D_A(OL)$ is the angular diameter distance to the lens, $D_A(OS)$ is the angular diameter distance to the source, $D_A(LS)$ is the angular diameter distance between

[61] The EFTofLSS formalism can provide a prediction of the LSS clustering in the mildly non-linear regime [558, 239, 444, 1471, 1444].

the lens and the source and $\phi(\theta, \beta)$ is the Fermat potential [e.g. 1694]

$$\phi(\theta, \beta) = \frac{(\theta - \beta)^2}{2} - \psi(\theta) \tag{63}$$

with $\psi(\theta)$ the lensing potential at the image direction. The time delay Δt_{AB} in Eq. (62) is thus connected to the time delay distance defined as [e.g. 1694, 1696]

$$D_{\Delta t} = \frac{1 + z_L}{c} \frac{D_A(OL) D_A(OS)}{D_A(LS)} \tag{64}$$

This distance is inversely proportional to H_0

$$D_{\Delta t} \propto \frac{1}{H_0} \tag{65}$$

and thus its measurement constrains H_0. Strongly lensed quasars (bright and time variable sources) lensed by a foreground lensing mass are used to measure the above observable time delay on cosmologically interesting scales [1037, 1080, 1380, 340, 366]. Active galactic nuclei (AGN) constitute another background source which may be used to measure the time delay [1083, 696, 658]. Recently, Qi *et al.* [1488] proposed the strongly lensed SnIa as a precise late-universe probe to improve the measurements on the Hubble constant and cosmic curvature. The inference of H_0 from $D_{\Delta t}$ is relatively insensitive to the assumed background cosmology.

Note that a source of systematic effects in time delay cosmography is the uncertainty of the mass along the line of sight modeling with respect to the mass sheet transformation (MST). This is a mathematical degeneracy [e.g. 687, 808, 1558, 1602, 1079, 1081] and can bias the strong lensing determination of Hubble constant [1082].

Observational data - Constraints:

Strong gravitational lensing time delay measurements of H_0 are consistent with the local measurements using late time calibrators and in mild tension with Planck [e.g. 340]. The method of the measurement of H_0 Lenses in COSMOGRAIL's Wellspring (H0LiCOW) collaboration [366] is independent of the cosmic distance ladder and is based on time delays between multiple images of the same source, as occurs in strong gravitational lensing.

Using joint analysis of six gravitationally lensed quasars with measured time delays from the COSmological MOnitoring of GRAvItational Lenses (COSMOGRAIL) project, the value $H_0 = 73.3^{+1.7}_{-1.8}\ km\ s^{-1} Mpc^{-1}$ was obtained which is in 3.1σ tension with Planck CMB. Assuming the Universe is flat and using lensing systems from the lensing program H0LiCOW and the Pantheon supernova compilation a value of $H_0 = 72.2 \pm 2.1\ km\ s^{-1} Mpc^{-1}$ was reported by the analysis of Ref. [1174]. A similar value of $H_0 = 72.8^{+1.6}_{-1.7}\ km\ s^{-1} Mpc^{-1}$ was found using updated H0LiCOW dataset consisting of six lenses [1175]. The reanalysis of the four publicly released lenses distance posteriors from the H0LiCOW by Yang *et al.* [1852] leads to $H_0 = 73.65^{+1.95}_{-2.26}\ km\ s^{-1} Mpc^{-1}$. The analysis of the strong lens system DES $J0408 - 5354$ by Shajib *et al.* [1625] for strong lensing insights into dark energy survey collaboration (STRIDES), infers $H_0 = 74.2^{+2.7}_{-3.0}\ km\, s^{-1} Mpc^{-1}$ in the ΛCDM cosmology. The analysis by Birrer *et al.* [299] based on the strong lensing and using Time-Delay COSMOgraphy (TDCOSMO[62, 63]) data set alone infers $H_0 = 74.5^{+5.6}_{-6.1}\ km\ s^{-1} Mpc^{-1}$ and using a joint hierarchical analysis of the TDCOSMO and Sloan Lens ACS (SLACS) [333] sample reports $H_0 = 67.4^{+4.1}_{-3.2}\ km\ s^{-1} Mpc^{-1}$. Chen *et al.*

[62] TDCOSMO collaboration [1290] was formed by members of H0LiCOW, STRIDES, COSMOGRAIL and SHARP.

[63] http://www.tdcosmo.org/

[465] based on a joint analysis of 3 strong lensing system, using ground-based adaptive optics (AO) from SHARP AO effort and the HST find $H_0 = 76.8 \pm 2.6\ km\ s^{-1} Mpc^{-1}$. A reanalysis of six of the TDCOSMO lenses using a power-law mass profile model results in $H_0 = 74.2 \pm 1.6\ km\ s^{-1} Mpc^{-1}$ [1290]. Analysing 8 strongly, quadruply lensing systems Denzel *et al.* [579] present a determination of the Hubble constant $H_0 = 71.8^{+3.9}_{-3.3}\ km\ s^{-1} Mpc^{-1}$ which is consistent with both early and late Universe observations. The value $H_0 = 73.6^{+1.8}_{-1.6}\ km\, s^{-1}\, Mpc^{-1}$ was reported by Qi *et al.* [1489] by combining the observations of ultra-compact structure in radio quasars and strong gravitational lensing with quasars acting as background source.

Standard sirens: gravitational waves

An independent and potentially highly effective approach for the measurement of $H(z)$ and the Hubble constant is the use of gravitational wave (GW) observations and in particular those GW bursts that have an electromagnetic counterpart (standard sirens) [1607, 180, 554, 1358, 1357]. In analogy with the traditional standard candles, it is possible to use standard sirens to directly measure the luminosity distance d_L of the GW source.

Standard sirens involve the combination of a GW signal and its independently observed electromagnetic (EM) counterpart. Such counterpart may involve short gamma-ray bursts (SGRBs) signal from binary neutron star mergers [657] or associated isotropic kilonova emission [525, 1653] and enables the immediate identification of the host galaxy. In contrast to traditional standard candles such as SnIa calibrated by Cepheid variables, standard sirens do not require any form of cosmological distance ladder. Instead they are calibrated in the context of general relativity through the observed GW waveform.

The simultaneous observations of the GW signal and its EM counterpart (multi-messenger observations) of nearby compact-object merger leads to a measurement of the luminosity distance which depends on the inclination angle of the binary orbit with respect to the line of sight and the redshift (measured using photons) of the host galaxy respectively. An EM counterpart detected with a GW observation can further constrain the inclination angle and may also indicate the source's sky position and the GW merger's time and phase [e.g. 1357].

In the case of GW events with small enough localization volumes without an observed EM counterpart (dark sirens) [466] a statistical analysis over a set of potential host galaxies within the event localization region may provide redshift information. A candidate for such statistical method is a merger of stellar-mass binary black holes[64] (BBH) which is usually not expected to result in bright EM counterparts unless it takes place in significantly gaseous environment [812]. For example, GW190521 [20] is a possible candidate with EM counterpart corresponding to a stellar-origin BBH merger in active galactic nucleus (AGN) disks [1272] detected by ZTF [253, 813].

Alternatively, in the absence of an EM counterpart the redshift can be determined by exploiting information on the properties of the source (e.g. the knowledge of neutron star equation of state) to derive frequency-dependent features in the waveform [1285] or using the gravitational waveform to determine the redshift of the mass distribution of the sources [1706, 693]. Also, Leandro *et al.* [1140] use an alternative method, presented in Ding *et al.* [626], for redshift determination by the statistical knowledge of the redshift distribution of sources. Trott and Huterer [1725] argue that any absolute determination of H_0 may be biased due to the fundamental degeneracy between redshift and H_0 and therefore can not lead to reliable determination of H_0. According to Trott and Huterer [1725] the reliable determination of H_0 with GW can only be achieved using standard sirens.

The luminosity distance-redshift relation Eq. (39) determines the Universe's expansion

[64] The stability analysis of the structures around black holes have been widely employed in the literature [see e.g. 1069, 249, 535, 159, 104].

history and the associated cosmological parameters including the Hubble constant H_0 [10, 715]. In particular using the mergers of binary neutron stars (BNS), or a binary of a neutron star with a stellar-mass black hole (NS-BH), which are excellent standard sirens, both the luminosity distance (from the gravitational wave waveform) and redshift of the host galaxy (from the electromagnetic counterpart) can be measured.

Using a BNS or a NS-BH merger, the distance to the source can be estimated from the detected amplitude $\langle h \rangle$ (r.m.s. - averaged over detector and source orientations) of the GW signal by the expression [1607, 141, 966, 1091, 1090]

$$d = C f^{-2} \langle h \rangle^{-1} \tau^{-1} \tag{66}$$

where f is the gravitational wave frequency, $\tau \equiv f/\dot{f}$ is the timescale of frequency change, C is a known numerical constant. Assuming a flat[65] Universe the luminosity distance can then be obtained from the relation

$$d(z) = \frac{1}{1+z} d_L(z) \tag{67}$$

For nearby sources, the recession velocity using the Hubble's law is determined by the expression

$$v(z) = H_0 d(z) \tag{68}$$

and using Eqs. (39), (44) and (67) is given by

$$v(z) = \frac{H_0 d_L(z)}{1+z} = \frac{c D_L(z)}{1+z} = c H_0 \int_0^z \frac{dz'}{H(z')} \tag{69}$$

At low redshifts using the local expansion Eq. (48) we obtain

$$v(z) = \frac{cz}{1+z} \left[1 + \frac{1}{2}(1-q_0) z \right] \tag{70}$$

which is approximated for $d \leq 100\ Mpc$ (or $z \leq 0.03$) as

$$v(z) = cz = H_0 d \tag{71}$$

Using Eqs. (68) and (70), the equation for the determination of H_0 as a function of observables, z and d is [e.g. 1895]

$$H_0 = \frac{cz}{d(1+z)} \left[1 + \frac{1}{2}(1-q_0) z \right] \tag{72}$$

where the deceleration parameter may be set by a fit to the GW data or may be fixed to its Planck/ΛCDM best fit form ($q_0 = -0.55$).

Observational data - Constraints:

The first multi-messenger detection of a BNS merger, GW170817, by LIGO [2] and Virgo [41] interferometers enabled the first standard siren measurement of the Hubble constant H_0. Using the BNS merger GW170817, the distance to the source was estimated to be $d = 43.8^{+2.9}_{-6.9}$ Mpc (i.e. at redshift $z \sim 0.01$) from the detected amplitude $\langle h \rangle$ (r.m.s. - averaged over detector and source orientations) of the GW signal by the Eq. (66) [10]. Also using the Hubble flow velocity $v_H = 3017 \pm 166\ km\ s^{-1}$ inferred from measurement of the redshift of the host galaxy,

[65] In an open (closed) Universe the distance in Hubble's law is given $d(z) = \frac{1}{1+z} \frac{\chi}{\sinh \chi} d_L(z)$ ($d(z) = \frac{1}{1+z} \frac{\chi}{\sin \chi} d_L(z)$)

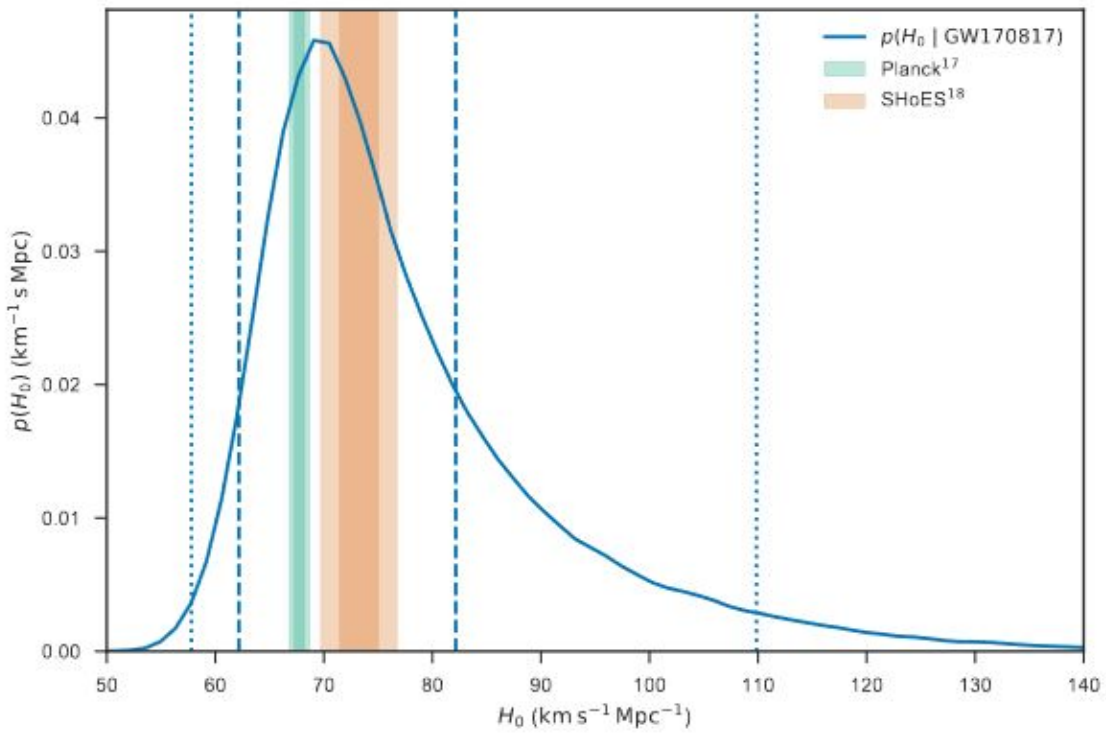


Figure 124: The probability of different values of H_0 with the maximum at $H_0 = 70.0^{+12.0}_{-8.0}$ $km\ s^{-1}Mpc^{-1}$ (solid blue curve) derived by BNS event GW170817. The dashed and dotted lines show minimal 68.3% (1σ) and 95.4% (2σ) credible intervals. The shaded green and orange bands show the 1σ and 2σ constraints from the analysis of the CMB data obtained by the Planck [50] and from the analysis of the SnIa data obtained by SH0ES [1526] respectively [from 10].

NGC 4993 (NGC 4993 was identified as the unique host galaxy), the Hubble constant was determined to be $H_0 = 70.0^{+12.0}_{-8.0}$ $km\ s^{-1}Mpc^{-1}$ [10] (see Fig. 124) by using Eq. (71).

Using continued monitoring of the the radio counterpart of GW17081 combining with earlier GW and EM data Hotokezaka *et al.* [910] obtain a improved measurement of $H_0 = 68.9^{+4.7}_{-4.6}$ $km\ s^{-1}Mpc^{-1}$. Note that using the BNS merger GW170817 in Fishbach *et al.* [715] and a statistical analysis (as first proposed in Schutz [1607]) over a catalog of potential host galaxies, the Hubble constant was determined to be $H_0 = 77.0^{+37.0}_{-18.0}$ $km\ s^{-1}Mpc^{-1}$. Using density-estimation Likelihood-Free Inference (LFI) Gerardi *et al.* [780] focused on the inference of the cosmological expansion H_0 from GW-selected catalogues of BNS mergers with EM counterparts.

Also using the BBH merger GW170814 as a standard (dark) siren in the absence of an electromagnetic counterpart, combined with a photometric redshift catalog from the Dark Energy Survey (DES) [25] the analysis by Soares-Santos *et al.* [1654] results in $H_0 = 75^{+40}_{-32}$ $km\ s^{-1}Mpc^{-1}$. Using multiple GW observations (the BNS event GW170817 and the BBH events observed by advanced LIGO and Virgo in their first and second observing runs) in Abbott *et al.* [17] the Hubble constant was constrained as $H_0 = 69.0^{+16.0}_{-8.0}$ $km\ s^{-1}Mpc^{-1}$. Using the event GW190814 from merger of a black hole with a lighter compact object the Hubble constant was measured to be $H_0 = 75^{+59}_{-13}$ $km\ s^{-1}Mpc^{-1}$ [18]. In Mukherjee *et al.* [1327] the BBH merger GW190521 was analysed choosing the NRSur7dq4 waveform[66] for the estimation of luminosity distance, after marginalizing over matter density Ω_{0m} when the ΛCDM model is considered and using its EM counterpart ZTF19abanrhr[67] as identified in Graham *et al.* [812] the Hubble constant was measured to be $H_0 = 50.4^{+28.1}_{-19.5}$ $km\ s^{-1}Mpc^{-1}$. The same study [1327] choosing different types of waveform finds $H_0 = 43.1^{+24.6}_{-11.4}$ $km\ s^{-1}Mpc^{-1}$ and $H_0 = 62.2^{+29.5}_{-19.7}$ $km\ s^{-1}Mpc^{-1}$. Combining their results with the binary neutron star event GW170817 leads to $H_0 = 67.6^{+4.3}_{-4.2}$ $km\ s^{-1}Mpc^{-1}$. In Chen *et al.* [467] for the same GW-EM event, assuming a flat wCDM model has obtained $H_0 = 48^{+23}_{-10}$ $km\ s^{-1}Mpc^{-1}$.

The analysis by Abbott *et al.* [21] using 47 gravitational-wave sources from the Third LIGO–Virgo–KAGRA Gravitational-Wave Transient Catalog (GWTC–3), infers $H_0 = 68^{+12}_{-7}$ $km\ s^{-1}Mpc^{-1}$. Palmese *et al.* [1397] find $H_0 = 72.77^{+11.0}_{-7.55}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ using the best available gravitational wave events, uniform galaxy catalog from the Dark Energy Spectroscopic Instrument (DESI) [64, 65] Legacy Survey and combining with the GW170817. The value $H_0 = 88.6^{+17.1}_{-34.3}$ $km\ s^{-1}Mpc^{-1}$ for GW190521 event was reported, and $H_0 = 73.4^{+6.9}_{-10.7}$ $km\ s^{-1}Mpc^{-1}$ was obtained when combing the GW190521 with the results of the neutron star merger GW170817 [767]. More recently, Mukherjee *et al.* [1328] report $H_0 = 67^{+6.3}_{-3.8}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ combining the bright standard siren measurement from GW170817 with a better measurement of peculiar velocity.

Megamaser technique

Observations of water megamasers which are found in the accretion disks around supermassive black holes (SMBHs) in active galactic nuclei (AGN) have been demonstrated to be powerful one-step geometric probes for measuring extragalactic distances [880, 931, 1504].

Assuming a Keplerian circular orbit around the SMBH, the centripetal acceleration and the

[66] NRSur7dq4 is a numerical relativity surrogate 7-dimensional approximate waveform model of binary black hole merger with mass ratios $q \equiv \frac{m_1}{m2} \leq 4$ [1765]. This model is made publicly available through the gwsurrogate (see `https://pypi.org/project/gwsurrogate`) and surfinBH (see `https://pypi.org/project/surfinBH`) Python packages.

[67] The ZTF19abanrhr event was reported by ZTF [253]. This candidate EM counterpart is flare after a kicked BBH merger in the accretion disk of an AGN [1272] with peak luminosity occurred 50 days after the BBH event GW19052. The ZTF19abanrhr was first observed after 34 days from the GW detection at the sky direction ($RA = 192.42625^0$, $Dec = 34.82472^0$) and was associated with an AGN J124942.3 + 344929 at redshift $z = 0.438$ [812].

velocity of a masing cloud are given as [1504]

$$A = \frac{V^2}{r} \tag{73}$$

$$V = \sqrt{\frac{GM}{r}} \tag{74}$$

where G is the Newton's constant, M is the mass of the central supermassive black hole, and r is the distance of a masing cloud from the supermassive black hole.

The angular scale θ subtended by r is given by

$$\theta = \frac{r}{d} \tag{75}$$

where d is the distance to the galaxy.

Thus, from the velocity and acceleration measurements obtained from the maser spectrum, the distance to the maser may be determined

$$d = \frac{V^2}{A\theta} \tag{76}$$

where A is measured from the change in Doppler velocity with time by monitoring the maser spectrum on month timescales. Using Hubble's law the Hubble constant may be approximated as [1504]

$$H_0 \approx \frac{\upsilon}{d} \tag{77}$$

where υ is the measured recessional velocity.

In order to constrain the Hubble constant the Megamaser Cosmology Project (MCP) [1503] uses angular diameter distance measurements to disk megamaser-hosting galaxies well into the Hubble flow ($50-200\,Mpc$). These distances are independent of standard candle distances and their measurements do not rely on distance ladders, gravitational lenses or the CMB [1447]. Early measurements of H_0 using masers tended to favor lower values of $H_0 \simeq 67\ km\,s^{-1}Mpc^{-1}$ while more recent measurements favor higher values $H_0 \simeq 73\ km\,s^{-1}Mpc^{-1}$ as shown e.g. in Table 17.

Observational data - Constraints:

Recently, the Megamaser Cosmology Project (MCP) [1503] using geometric distance measurements to megamaser-hosting galaxies and assuming a global velocity uncertainty of 250 $km\ s^{-1}$ associated with peculiar motions of the maser galaxies constrains the Hubble constant to be $H_0 = 73,9 \pm 3\ km\ s^{-1}Mpc^{-1}$ [1447]. Previously the MCP reported results on galaxies, UGC 3789 with $H_0 = 68.9 \pm 7.1\,km\,s^{-1}Mpc^{-1}$ [1504], NGC 6264 with $H_0 = 68.0 \pm 9.0\,km\,s^{-1}Mpc^{-1}$ [1123], NGC 6323 with $H_0 = 73^{+26}_{-22}\,km\,s^{-1}Mpc^{-1}$ [1124] and NGC 5765b with $H_0 = 66.0 \pm 6.0\,km\,s^{-1}Mpc^{-1}$ [757]. Reid *et al.* [1502] use a improved distance estimation of the maser galaxy NGC 4258 (also known as Messier 106) to calibrate the Cepheid-SN Ia distance ladder combined with geometric distances from MW parallaxes and DEBs in the LMC. The measured value of the Hubble constant is $H_0 = 73.5 \pm 1.4\,km\,s^{-1}Mpc^{-1}$.

Tully-Fisher relation (TFR) as distance indicator

The Tully-Fisher (TF) method is a historically useful distance indicator based on the empirical relation between the intrinsic total luminosity (or the stellar mass) of a spiral galaxy[68] and

[68]Similarly, in the case of a elliptical galaxy the Faber–Jackson (FJ) empirical power-law relation $L \propto \sigma^{\gamma_{FJ}}$ (where L is the luminosity of galaxy, σ the velocity dispersion of its stars and γ_{FJ} is a index close to 4) [684]

its rotation velocity (or neutral hydrogen (HI) 21 cm emission line width) [1737]. This method has been used widely in measuring extragalactic distances [e.g. 1562].

The Baryonic Tully Fisher relation (BTFR) [1270, 1773, 829, 1267] connects the rotation speed V_c and total baryonic mass M_b (stars plus gas) of a spiral galaxy as

$$M_b = A_c V_c^s \tag{78}$$

where s [with $s \approx 3-4$ 1270, 1267, 1150] is a parameter and $\log A_c$ is the zero point in a log-log BTFR plot. This relation has been measured for hundreds of galaxies. The rotation speed V_c can be measured independently of distance while the total baryonic mass M_b may be used as distance indicator since it is connected to the intrinsic luminosity. Thus, the BTFR is a useful cosmic distance indicator approximately independent of redshift and thus can be used to obtain H_0.

The BTFR has a smaller amount of scatter with a corresponding better accuracy as a distance indicator than the classic TF relation [1150]. In addition the BTFR recovers two decades in velocity and six decades in mass [1270, 1267, 1266, 945, 1151, 1603].

A simple heuristic analytical derivation for the BTFR is obtained [1] by considering a star rotating with velocity v in a circular orbit of radius R around a central mass M. Then the star velocity is connected with the central mass as

$$v^2 = G\, M_b/R \implies v^4 = (G\, M_b/R)^2 \sim M_b\, S\, G^2 \tag{79}$$

where G is Newton's constant and S the surface density $S \equiv M/R^2$ which may be shown to be approximately constant [740]. From Eqs. (78) and (79) we have

$$A_c \sim G^{-2} S^{-1} \tag{80}$$

which indicates that the zero point intercept of the BTFR can probe both galaxy formation dynamics (through e.g. S) and possible fundamental constant dynamics (through G) [100].

Observational data - Constraints:

The analysis by Kourkchi *et al.* [1097] using infrared (IR) data of sample galaxies and the Tully Fisher relation determined the value of Hubble constant to be $H_0 = 76.0 \pm 1.1(stat.) \pm 2.3(sys.)\, km\, s^{-1} Mpc^{-1}$. In Schombert *et al.* [1603] a value of $H_0 = 75.1 \pm 2.3(stat.) \pm 1.5(sys.)\, km\, s^{-1} Mpc^{-1}$ was found using Baryonic Tully Fisher relation for 95 independent Spitzer photometry and accurate rotation curves (SPARC) galaxies[69] (up to distances of $\sim 130 Mpc$).

Extragalactic background light γ-ray attenuation

This method is based on the fact that the extragalactic background light (EBL) which is a diffuse radiation field that fills the Universe from ultraviolet (UV) through infrared wavelength induces opacity for very high energy (VHE) photons ($\geq 30\, GeV$) induced by photon-photon interaction [861]. In this process a γ-ray and an EBL photon in the intergalactic medium may annihilate and produce an electron-positron pair [811]. The induced attenuation in the spectra of γ-ray sources is characterized by an optical depth $\tau_{\gamma\gamma}$ that scales as $n\sigma_T l$ (where n is the photon density of the EBL, σ_T is the Thomson cross section, and l is the distance from the

can be used as a distance indicator. The FJ relation is the projection of the fundamental plane (FP) of elliptical galaxies which defined as $R_{eff} \propto \sigma^{s_1} I_{eff}^{s_2}$ (where R_{eff} is the effective radius and I_{eff} is the mean surface brightness within R_{eff}) [628].

[69] The SPARC catalogue contains 175 nearby (up to distances of $\sim 130 Mpc$) late-type galaxies (spirals and irregulars) [1149, 1150]. The SPARC data are publicly available at http://astroweb.cwru.edu/SPARC.

γ-ray source to Earth). The cosmic evolution and the matter content of the Universe determine the γ-ray optical depth and the amount of γ-ray attenuation along the line of sight [633, 1892]. Thus a derivation of H_0 can be obtained by measuring the γ-ray optical depth with the γ-ray telescopes [632]. This derivation is independent and complementary to that based on the distance ladder and cosmic microwave background (CMB) and seems to favor lower values of H_0 as shown in Table 17.

Observational data - Constraints:

The analysis by Domínguez *et al.* [633] using extragalactic background light γ - ray attenuation data from Fermi Large Area Telescope (Fermi-LAT) derives $H_0 = 67.4^{+6.0}_{-6.2}\, km\, s^{-1} Mpc^{-1}$ and $\Omega_{0m} = 0.14^{+0.06}_{-0.07}$. The analysis by Zeng and Yan [1892] fitting the $> 10\, GeV$ extragalactic background data with modeled extragalactic background spectrum results in $H_0 = 64.9^{+4.6}_{-4.3}\, km\, s^{-1} Mpc^{-1}$ and $\Omega_{0m} = 0.31^{+0.13}_{-0.14}$.

Cosmic chronometers

Cosmic chronometers are objects whose evolution history is known. For instance such objects are some types of galaxies. The observation of these objects at different redshifts and the corresponding differences in their evolutionary state has been used to obtain the value of $H(z)$ at each redshift z.

The cosmic chronometer technique for the determination of H_0 was originally suggested in Jimenez and Loeb [980] and is based on the quasi-local ($0.07 \lesssim z \lesssim 2.36$) measurements along the Hubble flow of the Hubble parameter expressed as

$$H(z) = -\frac{1}{1+z}\frac{dz}{dt} \tag{81}$$

Thus, the expansion rate may be obtained by measuring the age difference Δt between two old and passively evolving galaxies[70] which are separated by a small redshif interval Δz, to infer the dz/dt [1308, 1307].

This approach determines the $H_0 = H(z = 0)$ independent of the early-Universe physics and is not based on the distance ladder [e.g. 980, 469, 692, 1884, 798]. The estimated H_0 values are more consistent with the values estimated from recent CMB and BAO data than those values estimated from SnIa. The value of H_0 can not be derived using the cosmic chronometers observations alone because there is a background degeneracy between H_0 and Ω_{0m} and this degeneracy is broken when these observations are combined.

Observational data - Constraints:

In Chen *et al.* [469] the value of Hubble constant was found to be $H_0 = 68.3^{+2.7}_{-2.6}$ in the flat ΛCDM model relying on 28 $H(z)$ measurements and their extrapolation to redshift zero. Analysing 31 $H(z)$ data determined by the cosmic chronometric (CCH) method, and 5 $H(z)$ data by BAO observations and using the Gaussian Process (GP) method [1616, 1622, 1846, 992] to determine a continuous $H(z)$ function the Hubble constant is estimated to be $H_0 \sim 67 \pm 4\, km\, s^{-1} Mpc^{-1}$ by Yu *et al.* [1884]. Also using the GP an extension of this analysis by Gómez-Valent and Amendola [798], including the $H(z)$ measurements obtained from Pantheon compilation and HST CANDELS and CLASH Multi-Cycle Treasury (MCT) programs, finds $H_0 = 67.06 \pm 1.68\, km\, s^{-1} Mpc^{-1}$ which is more consistent again with the lower range of values

[70] These galaxies form only a few new stars and become fainter and redder with time. The time that has elapsed since they stopped star formation can be deduced.

for H_0. The GP method [1495] is used as a 'non-parametric' technique which does not assume any parametrization or any cosmological model [see 1375, for a discussion about GP as model independent method]. The GP modeling approach has been performed by several authors to reconstruct cosmological parameters and thus to extract cosmological information directly from data [see e.g 271, 675, 1508, 906, 1034, 905, 262, 1897, 400, 1615, 1559, 1162, 1163, 1237, 1512, 268, 339, 338, 1688, 675, 589, 1326, 1035, 191, 1549, 1633, 1511, 1371, 1164, 252, 1459, 410, 849, 1896, 1791, 1623, 265, 264, 371, 1166].

Recently, a analysis by Moresco *et al.* [1309] reports $H_0 = 67.8^{+8.7}_{-7.2}\, km\, s^{-1}\, Mpc^{-1}$ and $H_0 = 66.5 \pm 5.4\, km\, s^{-1}\, Mpc^{-1}$ for a generic open wCDM and for a flat ΛCDM respectively. The analysis by Moresco *et al.* [1309] examine the possible effects that can systematically bias the measurement and can affect the CC method. It should be pointed out however that the quality and reliability of cosmic chronometer data has been challenged by some authors. This is partly due to the fact that these datapoints are not model independent and are obtained by combining several datasets [798]. This has improved significantly in the context of the aforementioned analysis by Moresco *et al.* [1309] where a detailed study of the covariance matrix and the effects of systematics has been implemented.

HII galaxy measurements

The ionized hydrogen gas (HII) galaxies (HIIG) emit massive and compact bursts generated by the violent star formation (VSF) in dwarf irregular galaxies. The HIIG measurements can be used to probe the background evolution of the Universe. This method of H_0 determination is based on the standard candle calibration provided by a $L-\sigma$ (luminosity-velocity dispersion) relation. This relation exists in HIIGs and Giant extragalactic HII regions (GEHR) in nearby spiral and irregular galaxies. The turbulent emission line ionized gas velocity dispersion σ of the prominent Balmer lines[71] H-alpha ($H\alpha$) and H-beta ($H\beta$) relates with its integrated emission line luminosity L [1282, 1637, 1466, 462, 463, 1708, 1815, 461, 1880, 805]. The relationship between $L(H\beta)$ and $\sigma(H\beta)$ has a small enough scatter to define a cosmic distance indicator (that can be utilized out to $z \sim 4$) independently of redshift and can be approximated as [462, 463, 1708, 1815, 1139, 461, 1880, 1541, 805]

$$\log L(H\beta) = \nu \log \sigma(H\beta) + \kappa \tag{82}$$

where ν and κ are constants representing the slope and the logarithmic luminosity at $\log \sigma(H\beta) = 0$.

From Eq. (38) the luminosity $L(H\beta)$ is given by

$$L(H\beta) = 4\pi d_L^2 l(H\beta) \tag{83}$$

Thus using Eq. (82), the distance modulus $\mu \equiv m - M$ of an HIIG can be obtained [1815, 1139, 461, 1880, 1541, 805]

$$\mu_{obs} = 2.5\left[\nu \log \sigma(H\beta) + \kappa - \log l(H\beta)\right] - 100.2 \tag{84}$$

This observational distance modulus can be compared with the theoretical distance modulus. From the Eq. (43) this is given

$$\mu(z)_{th} = 5 \log_{10}\left[\frac{d_L(z)}{Mpc}\right] + 25 \tag{85}$$

[71] The Balmer series, or Balmer lines is one of a set of six named series describing the spectral line emissions of the hydrogen atom. This is characterized by the electron transitioning from $n \geq 3$ to $n = 2$ (where n is the principal quantum number of the electron. The transitions $n = 3$ to $n = 2$ and $n = 4$ to $n = 2$ are called H-alpha and H-beta respectively.

Using now the dimensionless Hubble free luminosity distance Eq. (44) this can be written as

$$\mu(z)_{th} = 5\log_{10}\left[D_L(z)\right] + 5\log_{10}\left[\frac{c/H_0}{Mpc}\right] + 25 \tag{86}$$

In order to obtain the best fit values for the parameters Ω_{0m} and H_0 this theoretical prediction may now be used to compared with the observed μ_{obs} data. Using the maximum likelihood analysis the best fit values for these parameters may be found in the usual manner by minimizing the quantity

$$\chi^2(H_0, \Omega_{0m}) = \sum_i \frac{\left[\mu_{obs,i} - \mu_{th}(z_i; H_0, \Omega_{0m})\right]^2}{\varepsilon_i^2} \tag{87}$$

where ε_i is the uncertainty of the ith measurement.

Observational data - Constraints:

Using 156 HII galaxy measurements as a new distance indicator and implementing the model-independent GP, the Hubble constant was found to be $H_0 = 76.12^{+3.47}_{-3.44}\, km\, s^{-1} Mpc^{-1}$ which is more consistent with the recent local measurements [1790]. Using data of 130 giant HII regions in 73 galaxies with Cepheid determined distances the best estimate of the Hubble parameter is $H_0 = 71.0 \pm 2.8\,(random) \pm 2.1\,(systematic)\, km\, s^{-1} Mpc^{-1}$ [707].

Combinations of data

The Hubble constant H_0 values at 68% CL through direct and indirect measurements obtained by the different methods described in subsection 11.2.1 are shown in Table 17 and described in more detail below in Fig. 125. Also the relative probability density value of H_0 was derived by recently published studies in the literature are shown in Fig. 126.

Cosmological parameter degeneracies from each individual probe can be broken using combination of probes. The multi-probe analysis are crucial for independent H_0 determination and are required in order to reduce systematic uncertainties [1695, 464] [see 1309, for a review].

The analysis by Wong et al. [366] using a combination of SH0ES and H0LiCOW results reports $H_0 = 73.8 \pm 1.1\ km\ s^{-1} Mpc^{-1}$ which raises the Hubble tension to 5.3σ between late Universe determinations of H_0 and Planck. This has been discredited by Kochanek [1082] who points out that an artificial reduction of the allowed degrees of freedom can lead to very precise but inaccurate estimates of H_0 based on gravitational lens time delays.

The analysis by DES Collaboration [105] using a combination of the Dark Energy Survey (DES) [355, 106, 205] clustering and weak lensing measurements with BAO and BBN experiments assuming a flat ΛCDM model with minimal neutrino mass ($\Sigma m_\nu = 0.06\ eV$) finds $H_0 = 67.2^{+1.2}_{-1.0}\ km\ s^{-1} Mpc^{-1}$ which is consistent with the value obtained with CMB data.

Using an extension of the standard GP formalism, and a combination of low-redshift expansion rate data (SnIa+BAO+CC) the Hubble constant was estimated to be $H_0 = 68.52^{+0.94+2.51(sys)}_{-0.94}\, km\, s^{-1} M$ by Haridasu *et al.* [849]. Using an alternative method Baxter and Sherwin [243] analysing the current CMB lensing data from Planck combined with Pantheon supernovae and using conservative priors, find an r_s independent constraint of $H_0 = 73.5 \pm 5.3\ km\ s^{-1} Mpc^{-1}$. Analysing low-redshift cosmological data from SnIa, BAO, strong gravitanional lensing, $H(z)$ measurements using cosmic chronometers and growth measurements from LSS observations for ΛCDM model Dutta *et al.* [644] find $H_0 = 70.30^{+1.36}_{-1.35}\, km\, s^{-1} Mpc^{-1}$ which is in $\sim 2\sigma$ tension with various low and high redshift observations.

Table 17: The Hubble constant H_0 values at 68% CL through direct and indirect measurements by different methods.

Dataset	H_0 **[**$km\ s^{-1}Mpc^{-1}$**]**	**Year**	**Refs.**
Planck CMB	$\mathbf{67.27 \pm 0.60}$	**2020**	[72]
Planck CMB+lensing	67.36 ± 0.54	2020	[72]
Planck+SPT+ACT CMB	67.49 ± 0.53	2021	[205]
eBOSS+Planck CMB	69.6 ± 1.8	2020	[1469]
SPT-3G CMB	68.8 ± 1.5	2021	[643]
ACT CMB	67.9 ± 1.5	2020	[80]
ACT+WMAP CMB	67.6 ± 1.1	2020	[80]
SPT CMB	71.3 ± 2.1	2018	[876]
WMAP9 CMB	70.0 ± 2.2	2013	[895]
BAO+WMAP CMB	$68.36^{+0.53}_{-0.52}$	2019	[1901]
BOSS correlation function+BAO+BBN	68.19 ± 0.99	2022	[1898]
P+BAO+BBN	69.23 ± 0.77	2022	[468]
P+Bispectrum+BAO+BBN	$68.31^{+0.83}_{-0.86}$	2022	[1453]
BAO+BBN	66.98 ± 1.18	2018	[9]
BOSS DR12+BBN	68.5 ± 2.2	2020	[556]
BOSS DR12+BBN	68.7 ± 1.5	2020	[492]
BOSS DR12+BBN	67.9 ± 1.1	2020	[953]
BOSS+eBOSS+BBN	67.35 ± 0.97	2020	[95]
LSS t_{eq} standard ruler	$69.5^{+3.0}_{-3.5}$	2022	[694]
LSS t_{eq} standard ruler+lensing	$70.6^{+3.0}_{-5.4}$	2020	[1454]
BAO+RSD	69.13 ± 2.34	2017	[1805]
SnIa-Cepheid	73.04 ± 1.04	2022	[1529]
SnIa-Cepheid	74.30 ± 1.45	2021	[421]
SnIa-Cepheid	73.20 ± 1.30	2021	[1521]
SnIa-Cepheid	74.03 ± 1.42	2019	[306]
SnIa-Cepheid	73.48 ± 1.66	2018	[1528]
SnIa-Cepheid	72.80 ± 2.70	2020	[370]
SnIa-Cepheid	73.00 ± 2.70	2020	[370]
SnIa-TRGB	76.94 ± 6.4	2022	[592]
SnIa-TRGB	72.4 ± 3.3	2022	[988]
SnIa-TRGB	71.5 ± 1.8	2021	[133]
SnIa-TRGB	69.8 ± 1.7	2021	[734]
SnIa-TRGB	65.8 ± 4.2	2021	[1063]
SnIa-TRGB	72.10 ± 2.10	2020	[1672]
SnIa-TRGB	69.60 ± 1.90	2020	[735]
SnIa-TRGB	69.80 ± 1.90	2019	[738]
SnIa-TRGB	71.1 ± 1.9	2019	[1502]
SnIa-TRGB	72.40 ± 2.00	2019	[1886]
SnIa-Miras	73.30 ± 4.00	2020	[923]
SBF	73.30 ± 2.50	2021	[314]

Continued on next page

Table 17 – continued from previous page

Dataset	H_0 [$km\ s^{-1}Mpc^{-1}$]	Year	Refs.
SBF	70.50 ± 4.10	2020	[1052]
SBF	71.90 ± 7.10	2018	[426]
SneII	$75.4^{+3.8}_{-3.7}$	2022	[958]
SneII	$75.8^{+5.2}_{-4.9}$	2020	[959]
Time-delay (TD) lensing	$71.8^{+3.9}_{-3.3}$	2021	[579]
TD lensing	$73.3^{+1.7}_{-1.8}$	2020	[366]
TD lensing	$72.8^{+1.6}_{-1.7}$	2020	[1175]
TD lensing	72.2 ± 2.1	2020	[1174]
TD lensing	$73.65^{+1.95}_{-2.26}$	2020	[1852]
TD lensing	74.2 ± 1.6	2020	[1290]
TD lensing	$73.6^{+1.8}_{-1.6}$	2021	[1489]
TD lensing	$74.2^{+2.7}_{-3.0}$	2020	[1625]
TD lensing	$74.5^{+5.6}_{-6.1}$	2020	[299]
TD lensing+SLACS	$67.4^{+4.1}_{-3.2}$	2020	[299]
TD lensing+SLACS	76.8 ± 2.6	2019	[465]
GW Standard Sirens	$67^{+6.3}_{-3.8}$	2022	[1328]
GW Standard Sirens	68^{+12}_{-7}	2021	[21]
GW Standard Sirens	$72.77^{+11.0}_{-7.55}$	2021	[1397]
GW Standard Sirens	$73.4^{+6.9}_{-10.7}$	2021	[767]
GW Standard Sirens	75^{+59}_{-13}	2020	[18]
GW Standard Sirens	$50.4^{+28.1}_{-19.5}$	2020	[1327]
GW Standard Sirens	$67.6^{+4.3}_{-4.2}$	2020	[1327]
GW Standard Sirens	48^{+23}_{-10}	2020	[467]
GW Standard Sirens	$69.0^{+16.0}_{-8.0}$	2019	[17]
GW Standard Sirens	75^{+40}_{-32}	2019	[1654]
GW Standard Sirens	$68.9^{+4.7}_{-4.6}$	2019	[910]
GW Standard Sirens	$77.00^{+37.00}_{-18.00}$	2019	[715]
GW Standard Sirens	$70.0^{+12.0}_{-8.0}$	2017	[10]
Masers	73.90 ± 3.00	2020	[1447]
Masers	73.50 ± 1.40	2019	[1502]
Masers	66.0 ± 6.0	2016	[757]
Masers	$73.0^{+26.0}_{-22.0}$	2015	[1124]
Masers	68.0 ± 9.0	2013	[1123]
Masers	68.9 ± 7.1	2013	[1504]
Tully Fisher	76.00 ± 2.60	2020	[1097]
Tully Fisher	75.1 ± 2.80	2020	[1603]
γ-ray attenuation	$67.4^{+6.0}_{-6.2}$	2019	[633]
γ-ray attenuation	$64.9^{+4.6}_{-4.3}$	2019	[1892]
HII galaxy	71.00 ± 2.8	2018	[707]
HII galaxy	$76.12^{+3.47}_{-3.44}$	2017	[1790]
Cosmic chronometers, flat ΛCDM with systematics	66.5 ± 5.4	2022	[1309]
Cosmic chronometers, open wCDM with systematics	$67.8^{+8.7}_{-7.2}$	2022	[1309]
Cosmic chronometers, without systematics	67.06 ± 1.68	2018	[798]
Cosmic chronometers, without systematics	67.00 ± 4.00	2018	[1884]
Cosmic chronometers, without systematics	$68.3^{+2.7}_{-2.6}$	2017	[469]

Continued on next page

Table 17 – continued from previous page

Dataset	H_0 [$km\ s^{-1}Mpc^{-1}$]	Year	Refs.
H(z)+BAO+SN-Pantheon+SN-DES+QSO+HIIG+ GRB	69.7 ± 1.2	2022	[429]
CMB (r_s-independent)+lensing+Pantheon	73.5 ± 5.3	2021	[243]
SnIa-Cepheid and TD lensing	73.8 ± 1.1	2020	[366]
SnIa+BAO+TD lensing+cosmic chronometers+ LSS	$70.30^{+1.36}_{-1.35}$	2019	[644]
BAO+BBN+WL-CC	$67.20^{+1.2}_{-1.0}$	2018	[105]
SnIa+BAO+CC	$68.52^{+0.94+2.51(sys)}_{-0.94}$	2018	[849]

More recently, the joint analysis of lower-redshift, non-CMB, data such as BAO, $H(z)$, SnIa, QSO, HII and GRBs by Cao and Ratra [429] gives a model-independent determinations of the Hubble constant, $H_0 = 69.7 \pm 1.2\ km\ s^{-1}Mpc^{-1}$ [see also 432, 433, 430, 431, for previous joint analyses].

Many other estimates of H_0 have been obtained in the literature within the standard ΛCDM model or in alternative scenarios by using joint analysis [338, 1512]. In addition, many analyses using various combinations of data assuming a ΛCDM model or a extended model beyond ΛCDM cosmology investigate whether the H_0 tension persists (or not). For example Okamatsu *et al.* [1382] use non-CMB data and specifically adopt the data from BAO, BBN, and SnIa to study the H_0 tension. They show that this tension exists in a broad framework beyond the standard ΛCDM model.

The current status - Historic evolution

Hubble's initial value in 1929 for the expansion rate, now called the Hubble constant, was approximately $500 km\, s^{-1}Mpc^{-1}$. From the 1970s, through the 80s and into the 90s the value of H_0 was estimated to be between 50 and 100 $km\, s^{-1}Mpc^{-1}$ [1736]. Of interest is the historical Hubble constant debate between, for example, long series of papers by Gérard de Vaucouleurs, who claimed that the value of H_0 is $90 < H_0 < 100\, km\, s^{-1}Mpc^{-1}$ [e.g. 571, 570], and Allan Sandage , who claimed the value is $50 < H_0 < 55\, km\, s^{-1}Mpc^{-1}$ [1580, 1581] [see 1741, for a historical review].

During the last decades there has been remarkable progress in measuring the Hubble constant. The available technology and measurement methods determine the accuracy of this quantity. The Hubble constant as a function of publication date, using a set of different methods is shown in Fig. 127. The values of H_0 determined in the late Universe with a calibration based on the Cepheid distance scale and the derived values of H_0 from analysis of the CMB anisotropy spectrum data are shown. The uncertainties in these values have been decreasing for both methods and the recent measurements disagree beyond 4σ.

Furthermore the comoving Hubble expansion rate as a function of redshift obtained from the Planck18 CMB is shown in Fig. 128 along with a few relevant data-points demonstrating the Hubble tension.

The basic strategic questions emerge

- How can $H(z)$ derived from Cepheid late time calibrators (blue point in Fig. 128) become consistent with $H(z)$ derived from the sound horizon early time calibrator (black line in Fig. 128)?

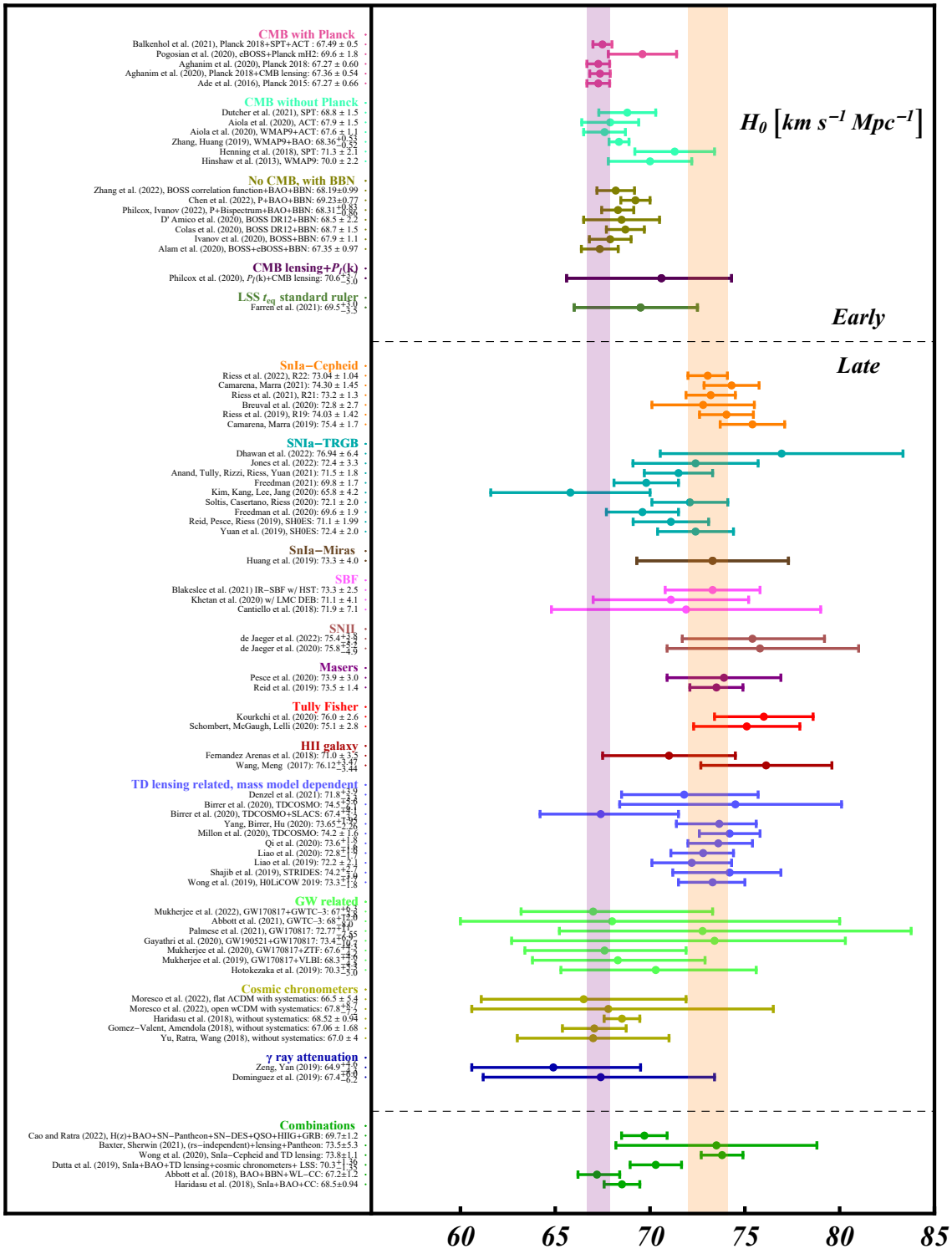


Figure 125: The Hubble constant H_0 values with the 68% CL constraints derived by recent measurements. The value of the Hubble constant H_0 is derived by early time approaches based on sound horizon, under the assumption of a ΛCDM background.

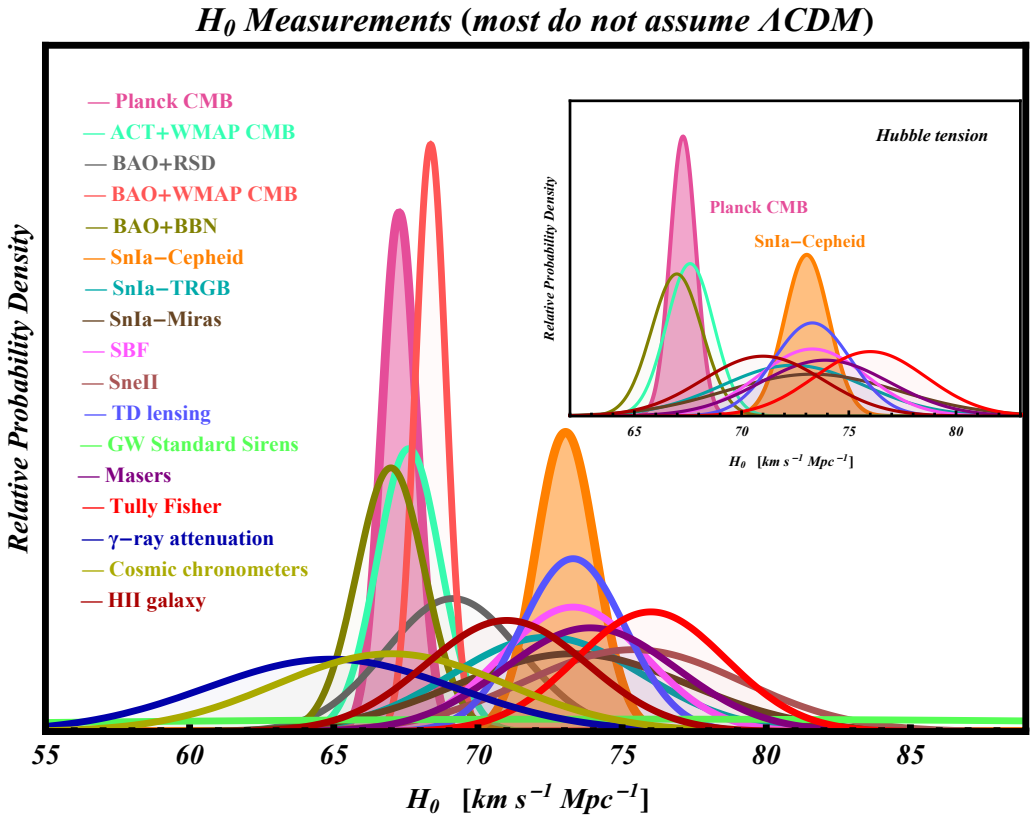


Figure 126: The one dimensional relative probability density value of H_0 derived by recent measurements (Planck CMB [72], ACT+WMAP CMB [80], BAO+RSD [1805], BAO+WMAP CMB [1901], BAO+BBN [9], SnIa-Cepheid [1529], SnIa-TRGB [988], SnIa-Miras [923], SBF [314], SneII [958], TD lensing [366], GW Standard Sirens [18], Masers [1447], Tully Fisher [1097], γ-ray attenuation [1892], cosmic chronometers [1884], HII galaxy [707]). All measurements are shown as normalized Gaussian distributions. Notice that the tension is not so much between early and late time approaches but more between approaches that calibrate based on low z ($z \lesssim 0.01$) gravitational physics and those that are independent of this assumption. For example cosmic chronometers and γ-ray attenuation which are late time but independent of late gravitational physics are more consistent with the CMB-BAO than with late time calibrators.

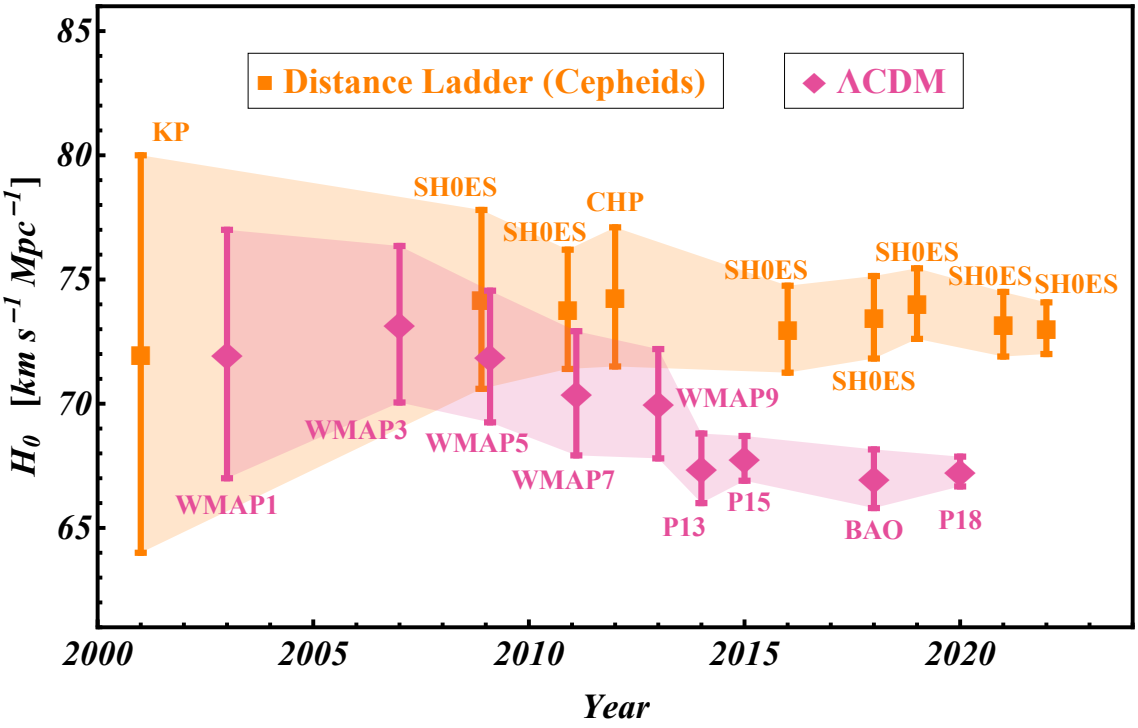


Figure 127: The Hubble constant as a function of publication date, using a set of different tools. Symbols in orange denote values of H_0 determined in the late Universe with a calibration based on the Cepheid distance scale (Key Project (KP) [737], SH0ES [1525, 308, 1526, 1528, 306, 1521, 1529], Carnegie Hubble Program (CHP) [736]). Symbols in purple denote derived values of H_0 from analysis of the CMB data based on the sound horizon standard ruler (First Year WMAP (WMAP1) [1676], Three Year WMAP (WMAP3) [1677], Five Year WMAP (WMAP5) [641], Seven Year WMAP (WMAP7) [1093], Nine Year WMAP (WMAP9) [42], Planck13 (P13) [55], Planck15 (P15) [50], Planck18 (P18) [72], BAO [9]). The orange and purple shaded regions demonstrate the evolution of the uncertainties in these values which have been decreasing for both methods. The most recent measurements disagree at greater than 4σ.

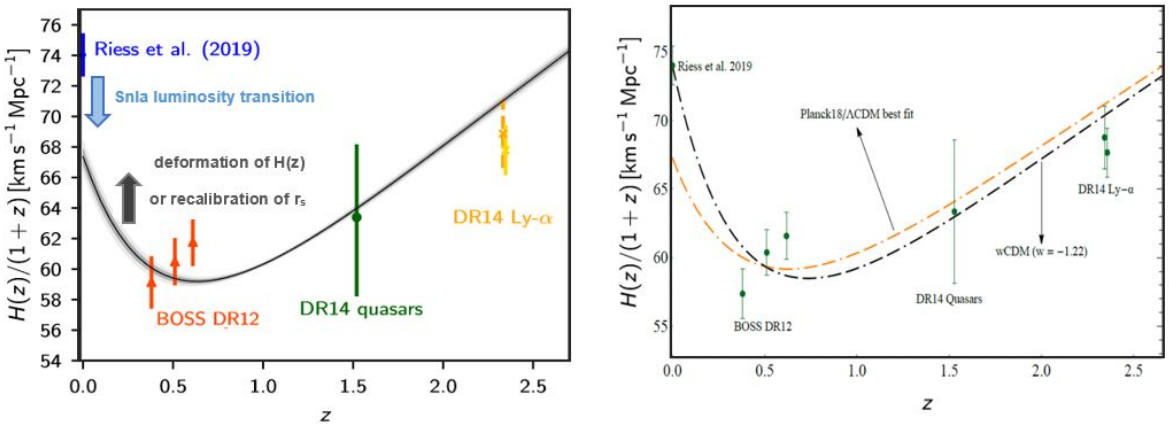


Figure 128: Left panel: The comoving Hubble parameter as a function of redshift. The black line corresponds to the best fit obtained from the Planck18 CMB when the ΛCDM model is considered, while the grey areas are the 1σ regions. The blue point at redshift zero denotes the inferred Hubble measurement by HST survey [306]. The orange points, green point, and yellow points correspond to BAO data from BOSS DR12 survey [15], BOSS DR14 quasar sample [373], and SDSS DR12 Lyα sample [1528] respectively. The arrows indicate approaches for the resolution of the Hubble tension: Down arrow (blue) corresponds to decrease of the Riess et. al. datapoint due to systematics or transition of the absolute magnitude M (light blue arrow). Up arrow (black) corresponds to recalibration of r_s which shifts the whole curve up or and late time deformation of $H(z)$ [adapted from 72]. Right panel: The comoving Hubble parameter as a function of redshift for a wCDM phantom modification of ΛCDM model which drives upward the low z part of the $H(z)$ curve shown in left panel. Thus it brings the $z = 0$ prediction of the CMB closer to the H_0 result of the local measurements (late time $H(z)$ deformation).

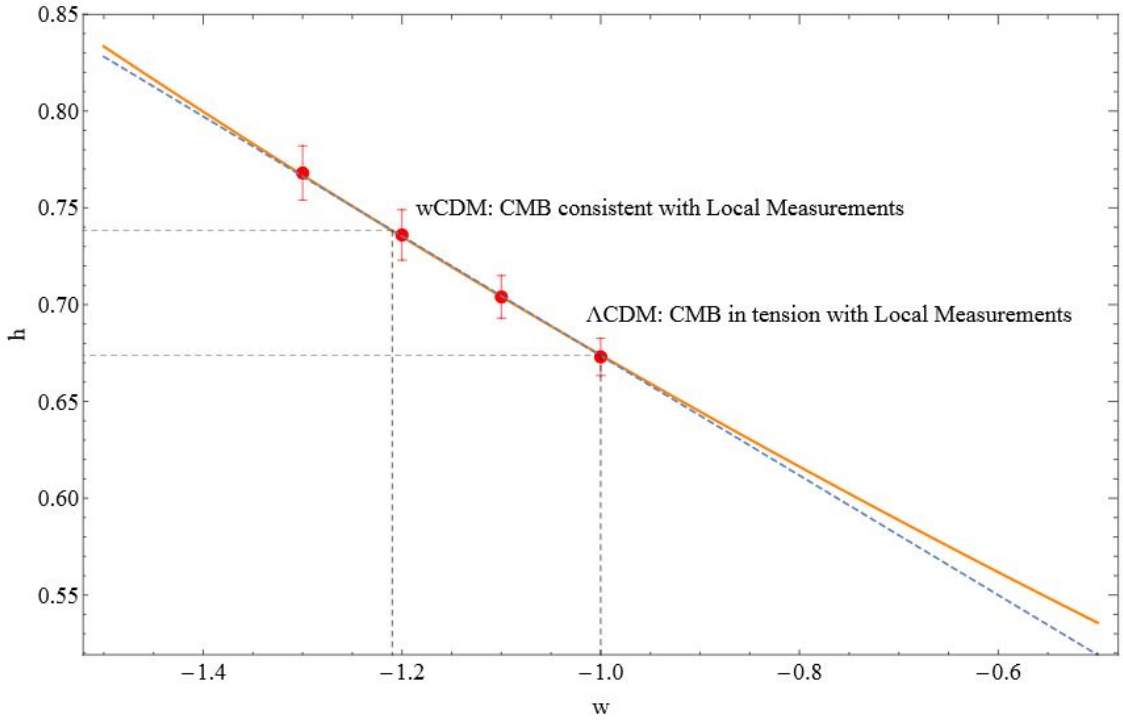


Figure 129: The predicted value of h as a function of the fixed w assuming one parameter dark energy (wCDM) model. The theoretically predicted best fit values of h for different values of w in the case of the wCDM model (orange line), whereas the linear fitting that has been made (dashed blue line). The redpoints correspond to the actual best fit values, including the errorbars, of h for specific values of w obtained by fitting these models to the CMB TT anisotropy [from 102].

- What type of systematics could move the blue point down or shift black line up in Fig. 128 in early and late time calibrators?
- To what extend can dynamical dark energy address the Hubble tension by distorting the black line in Fig. 128?

These important Hubble tension questions will be discussed in the next subsection.

11.2.2 Theoretical models

A wide range of models have been used to address the H_0 tension by introducing additional degrees of freedom to ΛCDM model where additional parameters are allowed to vary such as quintessence [746, 728, 1823, 1822, 1497, 473, 709, 710, 507, 417, 1909, 1733, 1595, 1685, 1180], in which a scalar field plays the role of dark energy or modified gravity [679, 125, 697, 1361, 1429, 330], in which General Relativity is modified on cosmological scales [see 508, 208, for a review].

The extensions of ΛCDM model which can be used to resolve the Hubble constant H_0 tension fall into two categories: models with late time and models with early time modification (in the epoch before the recombination) [see 1772, 615, 1604, for a review].

The models with late time modification can be divided in four broad classes: deformations of the Hubble expansion rate $H(z)$ at late times [e.g. late time phantom dark energy 102, 612], deformations of the Hubble expansion rate $H(z)$ with additional interactions/degrees of freedom (e.g. interacting dark energy Di Valentino *et al.* 607, Yang *et al.* 1869 and decaying dark matter Berezhiani *et al.* 283), deformations of the Hubble expansion rate $H(z)$ due to

inhomogeneous/anisotropic modifications [e.g. inhomogeneous causal horizons 729] and transition/recalibration of the SnIa absolute luminosity [1029] [or combination of the previous classes e.g. late $w - M$ phantom transition 103].

Model selection statistical tools and approaches include the Akaike Information Criterion (AIC) [81], the Bayesian Information Criterion (BIC) [1610] and the Deviance Information Criterion (DIC) [1678] and Bayesian model comparison [e.g. 1343, 1560, 1275, 1032, 1094]. These tools have been developed and used to test, discriminate and compare the proposed models [1177, 1178, 160, 1046] [see also 30, for a list of statistical tools].

Late time deformations of the Hubble expansion rate $H(z)$

These late time models for the solution of the Hubble tension use a late time smooth deformation of the Hubble expansion rate Planck18/ΛCDM $H(z)$ so that it can match the locally measured value of H_0 while keeping the radius r_s of the sound horizon at the last scattering surface (see Subsection 11.2.1). Many of these models effectively fix the comoving distance to the last scattering surface and the matter energy density $\omega_m = \Omega_{0m}h^2$ to values consistent with Planck/ΛCDM to maintain consistency with the CMB anisotropy spectrum while introducing late time phantom dark energy to deform $H(z)$ so that it matches the local measurements of $H(z)$. The required phantom behavior of such $H(z)$ deformations can not be provided by minimally coupled quintessence models and therefore such models have been shown to be unable to resolve the Hubble tension [209, 1143]. These models have three problems

- They tend to worsen the fit to low z distance probes such as BAO and SnIa [e.g. 102]
- They tend to worsen level of the growth tension [105].
- They tend to predict a lower value of SnIa absolute magnitude than the one determined by local Cepheid calibrators shown in Eq. (43) [1238, 420, 421].

Thus, these models can not fully resolve the Hubble tension [see also 101, 100, 1859, 261, 652, 157, 1709, 1110, 286, 1757, 408, 407, 675, 271].

Physical models where the deformation of $H(z)$ may be achieved include the following: phantom dark energy [e.g. 102] (see in Paragraph 11.2.2), running vacuum model [e.g. 1667] (see in Paragraph 11.2.2), phenomenologically emergent dark energy [1170] (see in Paragraph 11.2.2), vacuum phase transition [e.g. 602] (see in Paragraph 11.2.2), phase transition in dark energy [e.g. 1053] (see in Paragraph 11.2.2). Plethora of late dark energy models with an equation of state $w \neq -1$ ($w < -1$ or $w > -1$) both constant or dynamical with redshift [e.g. 1244] were proposed to address the Hubble tension. Recently, using a model-independent approach and a fully analytical analysis Heisenberg *et al.* [870, 871] derive a set of necessary conditions that any late dark energy model must satisfy in order to potentially address both the Hubble and the growth tensions. In particular, solving the H_0 tension requires $w(z) < -1$ at some z and solving both the H_0 and σ_8 tensions demands time-varying dark energy equation of state which cross the phantom divide. However Alestas and Perivolaropoulos [105] have shown that $H(z)$ deformation approaches to the Hubble tension tend to worsen the σ_8 growth tension.

The following models may be classified in this class of theories: the holographic dark energy [827, 546, 495, 1486, 1487, 1638, 877, 60], the considering Chevallier - Polarski - Linder (CPL) [471, 1190, 1068] parameterization [1865], the considering w dependence on non-vanishing spatial curvature [1286], the phantom brane dark energy [96, 197], the negative cosmological constant [1780, 416, 1618], the negative dark energy [645], the graduated dark energy [82], the simple-graduated dark energy [43], the Λ_sCDM model (sign-switching) [83], the transitional dark energy [1031], the frame dependent dark energy [62], the running H_0 with redshift [552, 1109], the varying gravitational constant [1564], the deviation from the cold dark matter

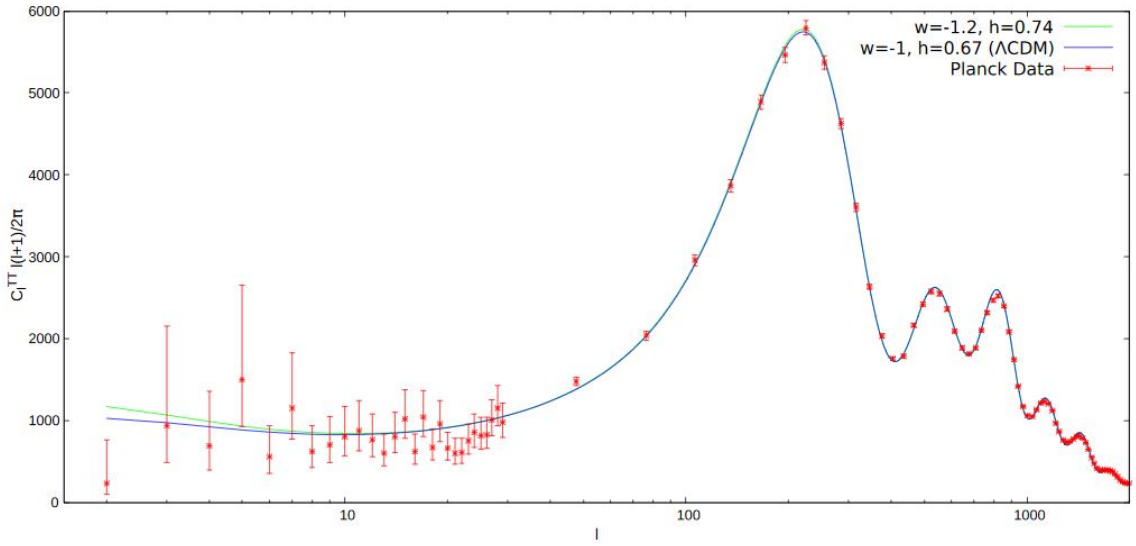


Figure 130: The predicted form of the CMB TT anisotropy spectrum with $w = -1$, $h = 0.67$, $\Omega_{0m} = 0.314$ for ΛCDM (blue line) and with $w = -1.2$, $h = 0.74$, $\Omega_{0m} = 0.263$ (green line). Red points correspond to the binned high-l and low-l Planck data [from 102].

[665] and the phantom crossing [617]. For example in the case of the holographic dark energy model [546] and phantom crossing [617] models the tension on H_0 appears to be significantly alleviated within 1σ even though the three problems mentioned above do remain.

Phantom dark energy:

The deformation of $H(z)$ through the implementation of late time phantom dark energy [102, 612, 606, 614, 1752, 926] can address the Hubble tension as shown in Fig. 128.

The analysis by Alestas *et al.* [102] indicates that mildly phantom models with mean equation of state parameter $w = -1.2$ have the potential to alleviate this tension. It was shown that the best fit value of H_0 in the context of the CMB power spectrum is degenerate with a constant equation of state parameter w. The CMB anisotropy spectrum was shown to be unaffected when changing $H(z)$ provided that specific parameter combinations remain unchanged. These cosmological parameters fix to high accuracy the form of the CMB anisotropy spectrum. The values of these parameters as determined by the Planck/ΛCDM CMB temperature power spectrum are the following [72].

$$\omega_{m,Planck} = 0.1430 \pm 0.0011 \tag{88}$$
$$\omega_{b,Planck} = 0.02237 \pm 0.00015 \tag{89}$$
$$\omega_{r,Planck} = (4.64 \pm 0.3)\ 10^{-5} \tag{90}$$
$$\omega_{k,Planck} = -0.0047 \pm 0.0029 \tag{91}$$
$$d_{A,Planck} = (4.62 \pm 0.08)\ (\ 100\ km\ s^{-1} Mpc^{-1})^{-1} \tag{92}$$

where $\omega_{i,Planck} = \Omega_{0i,Planck} h^2$ is the energy density of component i and $d_{A,Planck}$ is the comoving angular diameter distance.

Using the Eq. (41) the comoving angular diameter distance d_A to the recombination surface is ($c = 1$)

$$d_A = \int_0^{z_r} \frac{dz}{H(z)} = \int_0^{z_r} \frac{dz}{h(z)\ 100\ km\ s^{-1} Mpc^{-1}} \tag{93}$$

where $z_r \simeq 1100$ is the redshift of recombination and $h(z) = H(z)(\,100\,km\,s^{-1}Mpc^{-1})^{-1}$ is the dimensionless Hubble parameter which in general takes the form

$$h(z) = \big[\omega_r(1+z)^4 + \omega_m(1+z)^3 + \\ +(h^2 - \omega_r - \omega_m) f_{DE}(z)\big]^{1/2} \tag{94}$$

where $h = h(z=0)$ and $f_{DE}(z)$ determines the evolution of dark energy.

In the context of a simple one parameter parametrization where the equation of state w remains constant in time and redshift (wCDM model), $f_{DE}(z)$ takes the simple form

$$f_{DE}(z) = (1+z)^{3(1+w)} \tag{95}$$

If the four energy densities Eqs. (88), (89), (90) and (91) and the observed value of the comoving angular diameter Eq. (92) are fixed then they provide the analytically predicted best fit value of the Hubble parameter H_0 (or h) given the dark energy equation of state parameter $w(w_0, w_1, ..., z)$ where w_0, w_1,... are the parameters entering the $w(z)$ parametrization. Thus assuming a flat Universe ($\omega_k = 0$) and solving the following equation with respect to h

$$d_A(\omega_{m,Planck}, \omega_{r,Planck}, \omega_{b,Planck}, h = 0.674, w = -1) = \\ d_A(\omega_{m,Planck}, \omega_{r,Planck}, \omega_{b,Planck}, h, w) \tag{96}$$

it is straightforward to derive the degeneracy function $h(z = 0, w) \equiv h$ shown in Fig. 129 (continuous orange line). In the range $w \in [-1.5, -1]$, $h(w)$ is approximated as a straight line (dashed blue line in Fig. 129)

$$h(w) \approx -0.3093w + 0.3647 \tag{97}$$

For $w = -1$, this linear degeneracy equation leads to the best fit dimensionless Hubble constant $h = 0.674$ as expected while for $w = -1.217$ the corresponding predicted CMB best fit is $h = 0.74$ which is consistent with the value obtained by local distance ladder measurements. The invariance of the CMB power spectrum when the cosmological parameters are varied along the above described degeneracy directions is shown in Fig. 130. This method of Alestas *et al.* [102] can be used to find general degeneracy relations between $f_{DE}(z)$ and H_0 and fixing $h = 0.74$ gives infinite $f_{DE}(z)$, $w(z)$ forms that can potentially resolve the H_0 problem if they can also properly fit the low z date (e.g. BAO, SnIa, Cepheid value of absolute luminosity M). Low z distance data (BAO and SnIa) will determine which one of these forms is observationally favored. However, none of these forms can provide a quality of fit to low z data equally good or better than ΛCDM despite the introduced additional parameters. In addition, these models suffer from the other two problems mentioned above (worse growth tension and lower value of SnIa absolute magnitude).

Running vacuum model:

The running vacuum models (RVM) [1667, 1666, 1631, 1629, 1630, 236, 1427, 229, 1427, 1732, 213, 695, 231, 1181, 1306, 230, 1257, 1515, 1260, 1261, 1628, 232] [see 1425, 1661, 1662, 1665, 1663, 1664, 1259, 1258, for a review] attempts to address both the Hubble constant H_0 tension [1658] and the σ_8 growth tension using a mechanism that has common features with the IDE models [e.g. 1656, 1657, 795, 796, 1668, 1671] (for relaxing the growth tension, see Subsection 11.3.1).

The RVM of the cosmic evolution is well motivated by the generic idea of renormalization group formalism which is used in Quantum Field Theory (QFT) [1660, 1626, 1627] [see also

1306, 1305, for a approach using adiabatic regularization and renormalization techniques]. In the RVM the cosmological constant, the corresponding vacuum energy density and pressure are assumed to be functions of the Hubble rate e.g. a power series of the Hubble rate and its cosmic time derivative with even time derivatives of the scale factor [799]

$$\Lambda = a_0 + \sum_{k=1} a_k H^{2k} + \sum_{k=1} b_k \dot{H}^k, \tag{98}$$

$\rho_\Lambda = \rho_\Lambda(H) = \frac{\Lambda(H)}{8\pi G}$ and $p_\Lambda = p_\Lambda(H) = -\rho_\Lambda(H)$ respectively.

For the current Universe the vacuum energy density can be written in the relatively simple form [e.g. 1665, 799, 1659, 1658]

$$\rho_\Lambda(H) = \frac{\Lambda(H)}{8\pi G} = \frac{3}{8\pi G}(c_0 + \nu H^2) \tag{99}$$

where $c_0 = H_0^2(\Omega_{0\Lambda} - \nu) \simeq \frac{\Lambda_0}{3}$ (with Λ_0 the current value) is an integration constant which is fixed by the boundary condition $\rho_\Lambda(H_0) = \rho_{\Lambda,0}$ (with $\rho_{\Lambda,0}$ the current value) and ν is a dimensionless running parameter which characterizes the dynamics of the vacuum at low energy. For $\nu = 0$ the vacuum energy remains constant at all times and for $\nu > 0$ the vacuum energy density decreases with the time. In QFT the running parameter is $|\nu| \simeq 10^{-6} - 10^{-3}$ [1660] but in RVM it has been treated as a free parameter by fitting to the observational data [e.g. 799, 1659].

Phenomenologically emergent dark energy:

Phenomenologically emergent dark energy (PEDE) is a zero freedom dark energy scenario proposed by Li and Shafieloo [1170]. In this model the dark energy density has the following form

$$\tilde{\Omega}_{DE}(z) = \Omega_{0DE}\left[1 - \tanh\left(\log_{10}(1+z)\right)\right] \tag{100}$$

where $\Omega_{0DE} = 1 - \Omega_{0m} - \Omega_{0r}$.

The dark energy in this model has no effective presence in the past and emerges at the later times and with the same number (six) of parameters compared to the spatially flat ΛCDM scenario. It has the potential for alleviating the H_0 tension [1170, 1401, 1856, 1200, 1514, 1857, 600]. The generalised emergent dark energy (GEDE) model has one extra dimensionless free parameter Δ including both ΛCDM model as well as the PEDE model as two of its special limits introduced by Li and Shafieloo [1171]. In the GEDE model the dark energy density has the following form [1858]

$$\tilde{\Omega}_{DE}(z) = \Omega_{0DE}\frac{1 - \tanh\left(\Delta \log_{10}(\frac{1+z}{1+z_t})\right)}{1 + \tanh\left(\Delta \log_{10}(1+z_t)\right)} \tag{101}$$

where z_t is the transition redshift where dark energy density equals to matter density. For $\Delta = 0$ and $\Delta = 1$ this model recovers ΛCDM and PEDE model respectively. Using the latest observational Hubble dataset Hernández-Almada *et al.* [878] revisited and constrained the free parameters of the PEDE and GEDE models.

Other versions of the PEDE model are the Modified Emergent Dark Energy (MEDE) [257] and the Transitional Dark Energy (TDE) [1908] models. The MEDE model with one extra degree of freedom reduces the Hubble tension to 2.4σ [257] even though it also suffers from the three problems of the late time $H(z)$ deformation models.

Vacuum phase transition:

Vacuum phase transition [602, 603, 618, 600] based on vacuum metamorphosis (VM) or vacuum cold dark matter model (VCDM) [1407, 1408, 418] has the potential to address the H_0 tension. This mechanism with six free parameters as the spatially flat ΛCDM. It also assumes a phase transition in the nature of the vacuum similar to Sakharov's induced gravity [1563]. The phase transition occurs when the evolving Ricci scalar curvature R becomes equal to the value of scalar field mass squared m^2 [602]

$$R = 6(\dot{H} + H^2) = m^2 \tag{102}$$

where the dot corresponds to the derivative with respect to cosmic time t. After the transition the Ricci scalar curvature remains constant with $R = m^2$ and this changes the expansion rate below ($z < z_t$) due to the phase transition

$$\frac{H^2}{H_0^2} = \Omega_{0m}(1+z)^3 + \Omega_{0m}(1+z)^3 + \\ + M\left\{1 - \left[3\left(\frac{4}{3\Omega_{om}}\right)^4 M(1-M)^3\right]^{-1}\right\}, \; z > z_t \tag{103}$$

$$\frac{H^2}{H_0^2} = (1-M)(1+z)^4 + M, \; z \leq z_t \tag{104}$$

where $M = \frac{m^2}{12H_0^2}$ and $z_t = -1 + \frac{3\Omega_{0m}}{4(1-M)}$ is the transition redshift.

Phase transition in dark energy:

- Phase transition in dark energy explored by Khosravi *et al.* [1053], Banihashemi *et al.* [215, 216], Moshafi *et al.* [1315] can address the Hubble tension. Generalizing this model by assigning a more realistic time evolution of dark energy Banihashemi *et al.* [214] propose the critically emergent dark energy (CEDE) model.

  In Farhang and Khosravi [690] the form of phase transition parametrized phenomenologically by a hyperbolic tangent function. This scenario for dark energy is similar used independently as PEDE and GEDE.

- Late dark energy (LDE) transition [261] at redshifts $z \ll 0.1$ can reduce the Hubble tension. This class of $H(z)$ deformation models has a more intense form of the third problem of the deformation class as they predict a significantly lower value of the SnIa absolute magnitude than the other $H(z)$ deformation models [103, 421].

  In this scenario the true Hubble constant is given by [1311, 261]

  $$H_0^2 = \tilde{H}_0^2(1 + 2\delta) \tag{105}$$

  where $\tilde{H}_0$ is the prediction for a flat ΛCDM model in the context of a CMB sound horizon calibration.

  In Benevento *et al.* [261], Dhawan *et al.* [590] it was shown that this model can not fully resolve the Hubble problem as it would imply a transition in the SnIa apparent magnitude which is not observed. These models however become viable in the context of a SnIa absolute magnitude transition [103, 1238].

Deformations of the Hubble expansion rate $H(z)$ with additional interactions/degrees of freedom

There exist several varieties of the models for the solution of the Hubble tension which use deformations of the Hubble expansion rate $H(z)$ with additional interactions/degrees of freedom. For example the interacting dark energy models [e.g. 607, 1869] (see in Paragraph 11.2.2) with an extra non-gravitational interaction between the components of the Universe and the decaying dark matter models [e.g. 283] (see in Paragraph 11.2.2) with additional degrees of freedom are able to alleviate the Hubble constant H_0 tension.

The following models may be classified in this class of theories: multi-interacting dark energy [1205], new interacting dark energy [758], interacting vacuum energy [1118], metastable dark energy [1172, 1855], Quintom dark energy [1405], cannibal dark matter [387], baryons-dark energy interacting [978] [see also 1759, 704], swampland conjectures [1374, 497, 75], non-local gravity [39, 250], late time transitions in the quintessence field [599], Galileon gravity [1510, 1911, 1424, 744, 869], $f(R)$ gravity [1378, 545, 1789, 981], $f(T)$ gravity [1367, 1792, 1507, 414, 858, 371, 270, 201, 1508], $f(T, B)$ gravity [674], $f(Q)$ gravity [1227], Brans-Dicke gravity [1669, 1670], minimal theory of massive gravity [149], scale–dependent gravity [115], unimodular gravity [1426, 1185], the screened fifth forces [583, 584], the minimally modified gravity [568], the Lifshitz cosmology [280], the Milne cosmology [1779], 4D Gauss-Bonnet gravity [1793], the generalized Chaplygin gas [1872], the unified cosmologies [1871], the Λ-gravity [830, 831], the $\Lambda(t)$-model [260, 259], the bulk viscous cosmology [666, 1639, 1864, 1364] and the surface tension hypothesis [1390]. For instance in the case of the metastable dark energy [1855], generalized Chaplygin gas [1872] and Galileon gravity [1510] models the tension on H_0 appears to be significantly alleviated to within about 1σ even though the there problems of the $H(z)$ deformation models remain to be addressed.

Interacting dark energy:

In the cosmological interacting dark energy (IDE) models [1370, 1193, 904, 1226, 1632, 446, 445, 21, 1803, 1571, 521, 1478, 1449, 1572, 1119, 1842, 1120, 1763, 607, 1868, 1873, 1867, 1861, 1863, 1167, 131, 801, 1122, 1403, 1860, 1869, 794, 109, 609, 1243, 608, 1875, 1874, 1206, 1854, 1853, 128, 984, 1400] [see 1788, 332, for a review] the dark components of the Universe i.e dark matter (DM) and dark energy (DE) have an extra non-gravitational interaction. The IDE model was proposed to address the coincidence problem [e.g. 501, 928, 411, 1410, 1900, 424, 284, 425, 866]. In addition the interaction between the dark fluids has been shown to be effective in substantially alleviating the Hubble constant H_0 tension [1861, 1863, 607, 610, 1400, 1119, 1869, 1206, 1800, 763, 1368, 826, 1402] or in addressing the structure growth σ_8 tension between the values inferred from the CMB and the WL measurements [1479, 132, 222, 422] (see Subsection 11.3.1) or in solving the two tensions simultaneously [1120, 1122, 609].

In IDE cosmology assuming spatially flat Friedmann-Lemaitre-Roberson-Walker background and pressureless dark matter ($w_c = 0$) the equations of evolution of the dark matter and dark energy densities ρ_c and ρ_{DE} respectively are given by [766]

$$\dot{\rho}_c + 3H\rho_c = Q(t) \tag{106}$$

$$\dot{\rho}_{DE} + 3H(1 + w_{DE})\rho_{DE} = -Q(t) \tag{107}$$

where the dot corresponds to the derivative with respect to cosmic time t, $w_{DE} = \frac{p_{DE}}{\rho_{DE}}$ is the equation of state of dark energy and Q represents the interaction rate between the dark sectors (i.e. the rate of energy transfer between the dark fluids). For $Q < 0$ energy flows from dark matter to dark energy, whereas for $Q > 0$ the energy flow is opposite.

These models combine the deformation of $H(z)$ with an extra modification of the growth rate of perturbations due to the tuned evolution of $\Omega_m(z)$ induced by the interaction term Q.

This additional tuning allows for a simultaneous improvement of the growth tension in contrast to models that involve a simple $H(z)$ deformation.

Various phenomenological IDE models were proposed in the literature where the rate of the interaction Q has a variety of possible functional forms [1399]. For example in some classes of IDE models the rate of the interaction Q is proportional to the energy density of dark energy $Q = \delta\, H \rho_{DE}$ [1120, 1122, 1869, 609] or cold dark matter $Q = \delta\, H \rho_c$ [1119] (where δ is a constant and $\delta = 0$ in the ΛCDM cosmology), or some combination of the two. Note that in the case of functional form $Q = \delta\, H \rho_c$ instabilities develop in the dark sector perturbations at early times [867].

Decaying dark matter:

Decaying dark matter into dark radiation (i.e. an unknown relativistic species that is not directly detectable), which has been first analysed by Ichiki *et al.* [937] and studied by **?** Chudaykin *et al.* [483], Bjaelde *et al.* [304], Wang *et al.* [1802], Anchordoqui *et al.* [137], Xiao *et al.* [1843], Choi *et al.* [475, 474], Nygaard *et al.* [1373], provides a promising scenario to relieve the Hubble constant H_0 tension [e.g. 283]. Also, it has been shown that this scenario can resolve the σ_8 growth tension [668, 36] or the two tensions simultaneously [1404] by a similar mechanism as in the IDE models. However, using the Planck data the analysis of the model by Poulin *et al.* [1474], Chudaykin *et al.* [482] has shown that the cosmological tensions are only slightly alleviated [see 373, for a different result].

In these models assuming spatially flat Friedmann-Lemaitre-Roberson-Walker Universe, pressureless dark matter, $w_c = 0$ and equation of state of dark radiation $w_{DR} = 1/3$, the equations of evolution of the dark matter and dark radiation densities ρ_c and ρ_{DR} respectively are given by [766]

$$\dot{\rho}_c + 3H\rho_c = -\Gamma\rho_c \tag{108}$$

$$\dot{\rho}_{DR} + 4H\rho_{DR} = \Gamma\rho_c \tag{109}$$

where $\Gamma = \frac{1}{\tau}$ is the decay rate of dark matter particles (with τ the particle's lifetime). In the literature a variety of possible functional forms of the decay rate has been explored [668, 215, 1474, 373]. For example in some cases the decay rate is proportional to the Hubble rate, $\Gamma \propto H$ [1404]. Constraints on the decay rate of dark matter have been obtained by the analysis of Ando and Ishiwata [143], Enqvist *et al.* [668].

A model with decaying dark matter into dark radiation in early/late Universe ($\tau \ll t_s$ / $\tau \gg t_s$, where t_s is the time of last scattering) increases/decreases the expansion rate $H(a; \rho_b, \rho_\gamma, \rho_c, \rho_{DR}, \rho_{DE})$ at high/low redshifts as it predicts a smaller matter content and a larger radiation content as time evolves (the early/late Universe is dominated by the radiation/matter and the dark radiation density decreases more rapidly than the matter density, $\rho_{DR} \propto a^{-4}$ and $\rho_c \propto a^{-3}$). In the case of $\tau \ll t_s$, the faster cosmological expansion $H(z)$ decreases the scale of the sound horizon r_s in Eq. (58) because the baryon-to-photon ratio, and thus c_s in Eq. (59), is tightly constrained by CMB fluctuations and BBN [57]. In the context of the degeneracy $H_0 r_s$ shown in Eq. (60) the lower scale of the sound horizon r_s yields a larger value of H_0. In the case of $\tau \gg t_s$, the lower dimensionless normalized Hubble rate $E(z)$ in the late-time leads to a larger value of H_0 since θ_s and r_s must be kept fixed in Eq. (60). Accordingly, both early and late decaying dark matter model are able to alleviate the Hubble constant H_0 tension [see 135, 138, for a detailed discussion].

There are alternative decaying dark matter models such as the light dark matter [99], the dynamical dark matter [582], the many-body or 2-body decaying cold dark matter scenarios [305] and the decaying warm dark matter scenario [322]. In the 2-body decaying cold dark matter scenario the decaying dark matter produces two particles, one massive warm dark matter

particle and one massless relativistic particle (dark radiation). This scenario can address the Hubble constant H_0 tension [1766, 488, 850] and the σ_8 growth tension [35].

A self-interacting dark matter model with a light force mediator coupled to dark radiation studied by Binder *et al.* [297], Hryczuk and Jodlowski [916]. This model can simultaneously reduce the tension between CMB and low-redshift astronomical observations of H_0 and σ_8.

Jaeckel and Yin [956] pointed out that a dark particle from reheating [957] can alleviate the H_0 tension through its decay to relativistic component which contributes to the dark radiation.

Recently, Ly-α constraints on possible models of dark-matter physics have been evaluated by Dienes *et al.* [625]. In particular the Ly-α bounds on different classes of dark-matter velocity distributions have been obtained.

Deformations of the Hubble expansion rate $H(z)$ with inhomogeneous/anisotropic modifications

Models where the cosmological principle and the FLRW metric are relaxed by considering inhomogeneous/anisotropic modifications have the potential to resolve the Hubble problem [1013]. Physical models where the deformation of $H(z)$ may be achieved with inhomogeneous/anisotropic modifications, include the following: Chameleon dark energy [e.g. 406] (see in Paragraph 11.2.2), cosmic voids [1840] (see also Paragraph 11.2.2) and inhomogeneous causal horizons [729] (see also Paragraph 11.2.2), charged dark matter [977, 254, 255], Bianchi type I spacetime [84] and emerging spatial curvature [868, 331].

Chameleon dark energy:

Chameleon dark energy [1057, 1056] [see also 364, 211, 562, 366, 1798, 1748, 1055, 1758, 269] attempts to address the Hubble constant H_0 tension by introducing a cosmic inhomogeneity in the Hubble expansion rate at late-time from the chameleon field coupled to the local matter overdensities [406]. This field trapped at a higher potential energy density acts as an effective cosmological constant and results in a faster local expansion rate than that of the background with lower matter density.

Cosmic voids:

In cosmic void models the local H_0 departs significantly from the cosmic mean H_0 because of the presence of an under-dense region (local void) [1201]. However in Wu and Huterer [1840], Kenworthy *et al.* [1045] it was shown that this alternative theory is inconsistent with current observations. The analysis was based on the assumption of the validity of standard ΛCDM and a study of the sample variance in the local measurements of the Hubble constant this alternative theory has been shown inconsistent with current observations. Wu and Huterer [1840] estimated that the required radius of void to resolve the tension in H_0 is about 150 Mpc and density contrast of $\delta \equiv \frac{\rho-\bar{\rho}}{\bar{\rho}} \simeq -0.8$ which is inconsistent at $\sim 20\sigma$ with the ΛCDM cosmology [1045, 859].

In the context of this inconsistency, Haslbauer *et al.* [859] considered a cosmological Milgromian dynamics or modified Newtonian dynamics (MOND) model [1288] with the presence of $11eV/c^2$ sterile neutrinos[72] to show that the Keenan-Barger-Cowie (KBC) void[73] has the potential to resolve the Hubble tension.

[72]Sterile neutrinos are a special kind of neutrino with right handed chirality that might interact only through gravity [113, 352] [see 1127, 8, 637, for a review]. They have been proposed to resolve some anomalies in neutrino data.

[73]The KBC void [1036] is a large local underdensity between 40 and 300 Mpc (i.e. $0.01 \lesssim z \lesssim 0.07$) around the Local Group.

Inhomogeneous Causal Horizons:

Fosalba and Gaztañaga [729] proposed a simple solution to the H_0 tension based on causally disconnected regions of the CMB temperature anisotropy maps from Planck [69]. It was pointed out that CMB maps show 'causal horizons' where cosmological parameters have distinct values. This could be justified by the fact that these regions of the Universe have never been in causal contact. Thus it was shown that the Hubble constant H_0 takes values which differ up to 20% among different causally disconnected regions. These cosmological parameter inhomogeneities are in agreement with the model of the Universe proposed in Gaztanaga [768] (see also Gaztanaga 769, 772, 770 for details) where the cosmological constant is simply formulated as a boundary term in the Einstein equations and where 'Causal Horizons' naturally arise. Thus if there are similar 'causal horizons' in the local universe (i.e, $z < 1100$), then 20% variations between the local and high-z measures of H_0 are indeed to be expected [729].

Late time modifications - Transition/Recalibration of the SnIa absolute luminosity

This class of models can address the problems of the $H(z)$ deformation models (especially the low M problem) by assuming a rapid variation (transition) of the SnIa intrinsic luminosity and absolute magnitude due e.g. to a gravitational physics transition at a redshift $z_t \lesssim 0.01$ [1238, 103, 100].

Gravity and evolution of the SnIa intrinsic luminosity:

As shown in the recent analysis by Kazantzidis and Perivolaropoulos [1029] there are abnormal features which may be interpreted as evolution of the measured parameter combination $\mathcal{M}$ (see Section 11.2.1). This measured parameter combination $\mathcal{M}$ in Eq. (51) depends on the absolute magnitude M and on the Hubble constant H_0 (M and H_0 are degenerate parameters). Any variation of the parameter $\mathcal{M}$ is due to a variation of M which could be induced by a varying $\mu_G(z) \equiv \frac{G(z)}{G_0}$ (where G_0 is the local value of the Newton's constant $G(z)$). If the calibrated SnIa absolute magnitude M were truly constant then the parameter $\mathcal{M}$ should also be constant (independent of redshift).

A possible variation of the absolute magnitude M and equivalently of the absolute luminosity

$$L \sim 10^{-2M/5} \tag{110}$$

could be due to a variation of the fine structure constant α or the Newton's constant G.

If the absolute luminosity is proportional to the Chandrasekhar mass $L \sim M_{Chandr}$ we have [759, 771]

$$L \sim G^{-3/2} \tag{111}$$

Thus L will increase as G decreases[74].

Under these assumptions, we obtain

$$M(z) - M_0 = \frac{15}{4} \log \mu_G(z) \tag{112}$$

where M_0 corresponds to a reference local value of the absolute magnitude and $\mu_G \equiv \frac{G}{G_0}$ is the relative effective gravitational constant (with G the strength of the gravitational interaction and G_0 the locally measured Newton's constant).

[74] Adopting a semi-analytical model which takes into account the stretch of SnIa light curves but assumes fixed mass of Ni, obtains SnIa light curves in the context of modified gravity Wright and Li [1837] have claimed that L will increase as G increases.

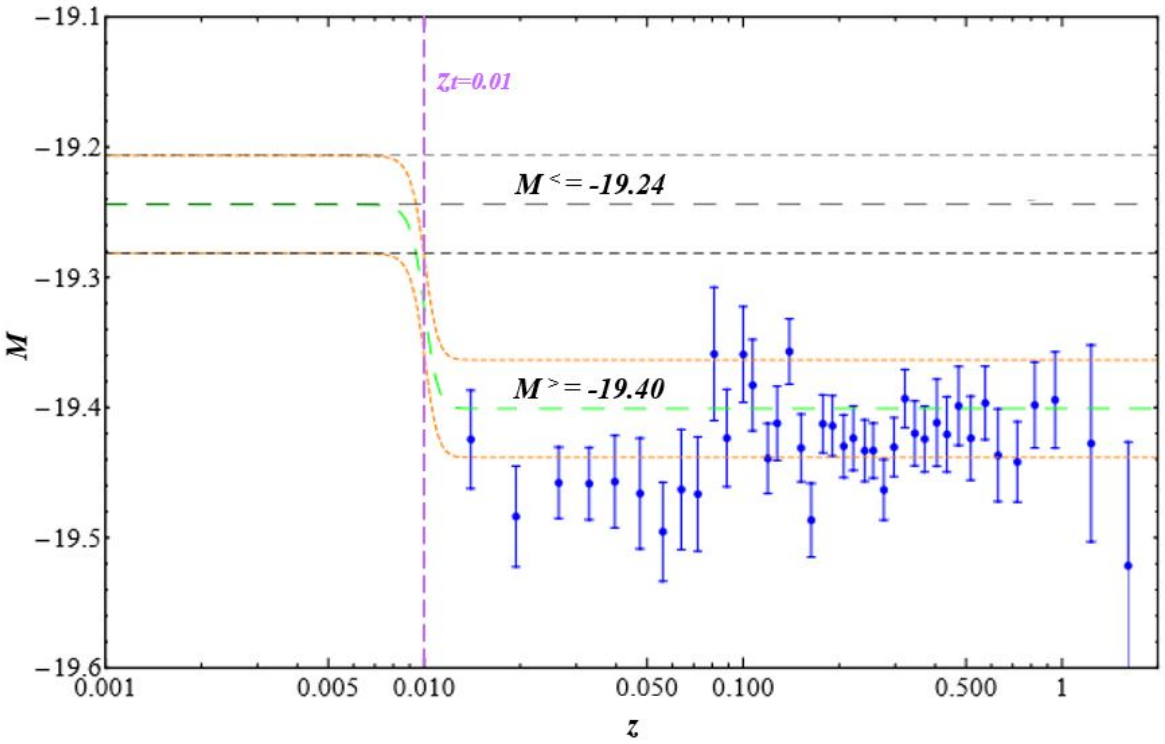


Figure 131: The Pantheon binned SnIa absolute magnitudes Eq. (117) M_i (blue points) [329] for a Planck/ΛCDM luminosity distance. The data are inconsistent with the SnIa absolute magnitude $M^< = -19.24$ calibrated by Cepheids but the inconsistency disappears if there is a transition in the absolute magnitude with amplitude $\Delta M \simeq -0.2$ at redshift $z_t \simeq 0.01$ [from 1238].

Then, the Eq. (51) takes the following form

$$\mathcal{M}(z) = M_0 + \frac{15}{4}\log\mu_G(z) + 5\log_{10}\left[\frac{c/H_0}{Mpc}\right] + 25 \tag{113}$$

and the apparent magnitude Eq. (52) can be written as

$$m(z, H_0, \Omega_{0m})_{th} = \mathcal{M}(z, H_0) + 5\log_{10}\left[D_L(z, \Omega_{0m})\right] \tag{114}$$

A mild tension at 2σ level in the best fit value of $\mathcal{M}$ was found in between the low-redshfit ($z \lesssim 0.2$) data and the full Pantheon dataset in the context of a ΛCDM model. This tension can be interpreted as [1029]

- a locally higher value of H_0 by about 2%, corresponding to a local matter underdensity.
- a time variation of Newton's constant which implies an evolving Chandrasekhar mass and thus an evolving absolute luminosity L and absolute magnitude M of low z SnIa.

In addition, the oscillating features shown in Fig. 119 hint also to the possibility of evolutionary effects of M. As discussed below such evolutionary effects if they exist in the form of a transition may provide a solution to the Hubble and growth tensions.

Transition of the SnIa absolute magnitude M at a redshift $z \simeq 0.01$:

Recently, Marra and Perivolaropoulos [1238] have proposed that a rapid transition (abrupt deformation) at a transition redshift $z_t \simeq 0.01$ in the value of the SnIa absolute magnitude M

of the form

$$M^{>}(z) = M^{<} + \Delta M\, \Theta(z - z_t) \tag{115}$$

(where Θ is the Heaviside step function) due to a rapid transition of the gravitational constant can address the Hubble tension.

In particular the analysis by Marra and Perivolaropoulos [1238] has shown that a 10% rapid transition in the value of the relative effective gravitational constant μ_G at $z_t \simeq 0.01$ is sufficient to induce the required reduction of M

$$\Delta M \equiv M^{>} - M^{<} \simeq -0.2 \tag{116}$$

where $M^{<}$ is the SnIa absolute magnitude of Eq. (54) calibrated by Cepheids at $z < 0.01$ [420, 421] and $M^{>}$ is the SnIa absolute magnitude of Eq. (55) using the parametric-free inverse distance ladder of Camarena and Marra [419]. Fig. 131 shows the Pantheon SnIa absolute magnitudes for a Planck/ΛCDM luminosity distance [329] obtained from

$$M_i = m_i - 5\log_{10}\left[\frac{d_L(z_i)}{Mpc}\right] - 25 \tag{117}$$

where m_i are the apparent magnitude datapoints.

The data are in disagreement with the SnIa absolute magnitude $M^{<}$ calibrated by Cepheids but they become consistent if there is a transition in the absolute magnitude with amplitude $\Delta M \simeq -0.2$ [1238]. Thus, this class of M-transition models avoids the M-problem of late time $H(z)$ deformation models.

Assuming the power law dependence Eq. (111) and using RSD and WL data [107, 1646, 102] reported a best fit value $\Delta\mu_G \equiv \mu_G^{>} - \mu_G^{<} = -0.19 \pm 0.09$ ($\mu_G^{>}$ corresponds to $z > 0.01$ and $\mu_G^{<}$ corresponds to $z < 0.01$) in the context of a ΛCDM background $H(z)$. The analysis by Marra and Perivolaropoulos [1238] showed that a rapid $\sim 10\%$ increase of the effective gravitational constant roughly 150 million years ago can also solve Ω_m-σ_8 growth tension.

Recently, Alestas and Perivolaropoulos [105] have demonstrated that this model has an advantage over both early time and late time deformations of $H(z)$ to fully resolve the Hubble tension while at the same time improving the level of the Ω_m-σ_8 growth tension. In addition it has the potential to provide equally good fit to low z distance probes such as BAO and SnIa as the Planck18/ΛCDM model.

More recently, Perivolaropoulos and Skara [1442] generalized the symmetron screening mechanism[75] [896, 897] by allowing for an explicit symmetry Z_2 breaking of the symmetron ϕ^4 potential. The explicit symmetry breaking can create an asymmeron wall network pinned on matter overdensities separating regions with distinct gravitational properties which could constitute a physical mechanism for the realization of gravitational transitions in redshift space that could help in the resolution of the Hubble and growth tensions. Another theoretical model leading to a gravitational transition could include a pressure non-crushing cosmological singularity in the recent past [1377].

Late (low-redshift) $w - M$ phantom transition:

The late (low-redshift) $w - M$ phantom transition [103] is a late time approach involving a combination of the previous two classes: the transition of the SnIa absolute luminosity and

[75] For reviews of modified gravity theories with screening mechanisms, such as the Vainshtein [1760, 166, 572] and the chameleon [1057, 1056, 825, 363, 362, 1747, 1318, 1319, 365, 366] models see [1054, 399, 1565, 368, 964, 564, 929, 398, 997, 367, 204, 1566] and for screening effects see [1509].

the deformation of the Hubble expansion rate $H(z)$. A rapid phantom transition of the dark energy equation of state parameter w at a transition redshift $z_t < 0.1$ of the form

$$w(z) = -1 + \Delta w\, \Theta(z_t - z) \tag{118}$$

with $\Delta w < 0$ and a similar transition in the value of the SnIa absolute magnitude M of the form

$$M(z) = M_C + \Delta M\, \Theta(z - z_t) \tag{119}$$

with $\Delta M < 0$ due to evolving fundamental constants can address the Hubble tension [103]. Where Θ is the Heaviside step function, M_C is the SnIa absolute magnitude Eq. (54) calibrated by Cepheids [421, 420] at $z < 0.01$ and ΔM, Δw are parameters to be fit by the data. Alestas *et al.* [103] find $\Delta M \simeq -0.1$, $\Delta w \simeq -4$ for $z_t = 0.02$ which imply a lower value of μ_G at $z > 0.02$ (about 6%) compared to the pure M-transition model.

The late (low-redshift) $w - M$ phantom transition (LwMPT) can lead to a resolution of the Hubble tension in a more consistent manner than smooth deformations of the Hubble tension and other types of late time transitions such as the Hubble expansion rate transition [261, 590]. Its main advantages include the consistency in the predicted value of the SnIa absolute magnitude M and the potential for simultaneous resolution of the growth tension.

Mortsell *et al.* [1312, 1313], Perivolaropoulos and Skara [1441] have analyzed the color-luminosity relation of Cepheids in anchor galaxies and SnIa host galaxies by identifying the color-luminosity relation for each individual galaxy instead of enforcing a universal color-luminosity relation to correct the NIR Cepheid magnitudes. A systematic brightening of Cepheids at distances larger than about 20 Mpc which could be enough to resolve the Hubble tension was found. In addition, Perivolaropoulos and Skara [1441] investigating the effects of variation of the Cepheid calibration empirical parameters (the color-luminosity parameter or the Cepheid absolute magnitude) find hints for the presence of a fundamental physics transition taking place at a time more recent than 100 Myrs ago. The magnitude of the transition lead to value of H_0 that is consistent with the CMB inferred value thus eliminating the Hubble tension. The distance range/timescale corresponding to this transition is consistent with solar system history data [1438] indicating an increase of the rate of impactors on the Moon and Earth surfaces by about a factor of 2-3 during the past 100Myrs which correspond to $z < 0.008$ [1636, 774, 1265, 819, 1806, 1262, 344] and low redshift galaxy surveys data [106]. Such a transition is also consistent with a recent analysis by Alestas *et al.* [100] indicating a transition in the context of the Tully-Fisher data.

In particular, using a robust dataset of 118 Tully-Fisher datapoints Alestas *et al.* [100] have demonstrated that evidence for a transition in the evolution of BTFR appears at a level of more than 3σ. Such effect could be interpreted as a transition of the effective Newton's constant. The amplitude and sign of the gravitational transition are consistent with the mechanisms for the resolution of the Hubble and growth tension discussed above [1238, 103] [see in 1437, for a talk of the tensions of the ΛCDM and a gravitational transition].

Early time modifications of sound horizon

Modifying the scale of sound horizon r_s (i.e. the scale of the standard ruler) by introducing new physics before recombination that deform $H(z)$ at prerecombination redshifts $z \gtrsim 1100$ can increase the CMB inferred value of H_0 [44, 1472, 194, 1074] and thus resolve the Hubble tension. Such deformation may be achieved by introducing various types of additional to the standard model components [see 171, for a review]. These models have the problem of predicting stronger growth of perturbations than implied by dynamical probes like redshift space distortion (RSD) and weak lensing (WL) data and thus may worsen the Ω_m-σ_8 growth tension [971, 105].

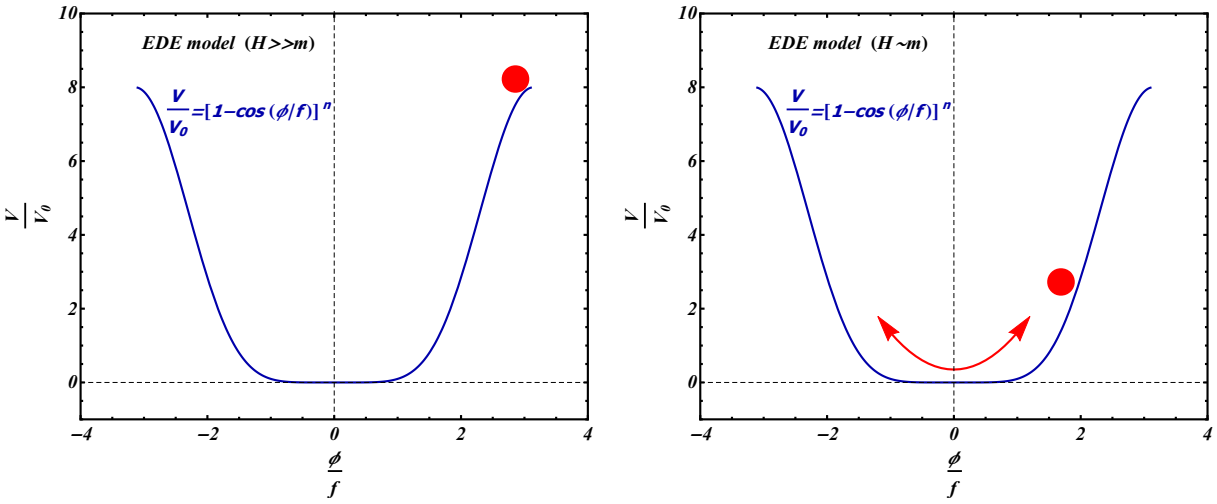


Figure 132: The potential V/V_0 (with $V_0 = m^2 f^2$, $n = 3$ in Eq. (121)) as a function of ϕ/f at early times ($H \gg m$) (left panel) when the field ϕ is initially frozen in its potential due to Hubble friction and acts as a cosmological constant with equation of state $w_\phi = -1$, and at a critical redshift z_c when the Hubble parameter drops below some value ($H \sim m$) (right panel) and the field becomes dynamical and begins to oscillate around its minimum which is locally $V \sim \phi^{2n}$.

A wide range of mechanisms has been proposed for the decrease of the the sound horizon scale at recombination. These mechanisms include the introduction of early dark energy, extra neutrinos or some other dark sector at recombination, features in the primordial power spectrum, modified scenarios of recombination etc. The following models and theories may be classified in this class of mechanisms: early dark energy [e.g. 301] (see Paragraph 11.2.2), dark radiation [e.g. 818] (see Paragraph 11.2.2), neutrino self-interactions [e.g 206] (see Paragraph 11.2.2), large primordial non-Gaussianities [59] (see Paragraph 11.2.2), Heisenberg's uncertainty principle [436] (see Paragraph 11.2.2), early modified gravity [358] (see Paragraph 11.2.2), cosmological inflation physics [1703, 158, 864, 1723, 598, 616, 828, 472, 605, 1196, 1033, 1876, 1697], dark matter - photon coupling [1121, 1845], dark matter-neutrino interactions [1409], interacting dark radiation [324], ultralight dark photon [717], primordial black holes [1348, 720], primordial magnetic fields [972, 970, 1710], non-standard recombination [1197], unparticles dark energy [170], varying fundamental constants [732, 854, 1617, 855, 856], early-time thermalization of cosmic components [1767], CMB monopole temperature shift [951], open and hotter universe [342, 266], Axi-Higgs cosmology [750, 749], string Cosmology [134, 136] and dark massive vector fields [140]. In this list of proposed cosmological models the tension on H_0 is alleviated with a significance ranging from the 1σ to the 3σ level.

Early dark energy:

In the early dark energy (EDE) model [1072, 879, 1128, 800, 723, 892, 1475, 1363, 1316, 1451, 301, 1011, 1567, 11, 222, 285, 1650, 1204, 107, 1322, 1002, 1331, 797, 851, 1649] an additional dynamical scalar field behaves like a cosmological constant at early times (near matter-radiation equality but before recombination). This field decays rapidly after recombination thus leaving the rest of the expansion history practically unaffected up to a rescaling which modifies H_0. This rescaling allows for the resolution of the Hubble constant tension. Using the Eq. (58) in a EDE model the radius of sound horizon at last scattering can be calculated by

$$r_s = \int_0^{t_d} \frac{c_s(a)}{a(t)} dt =$$

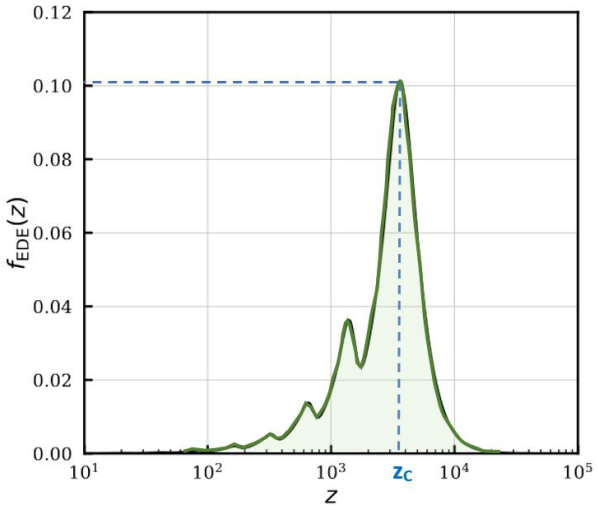


Figure 133: Fractional contribution of EDE to the cosmic energy budget as a function of redshift [adapted from 952].

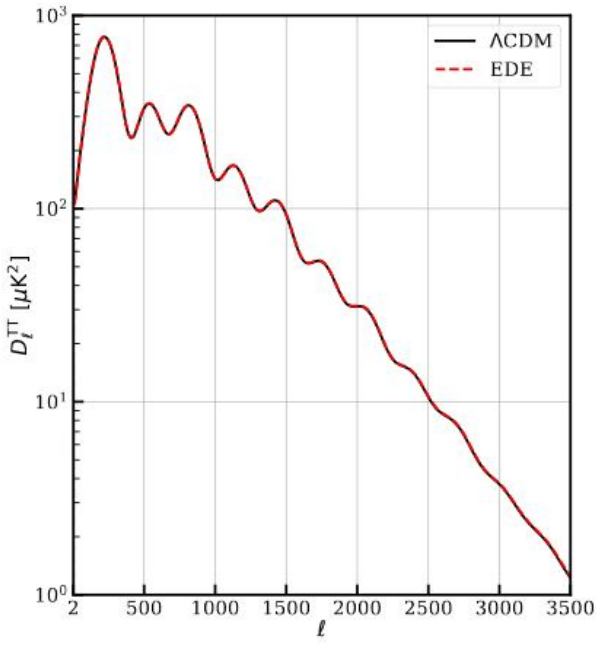


Figure 134: CMB TT power spectrum. The black solid and the red dashed lines correspond to ΛCDM model with $H_0 = 68.07\ km\ s^{-1}Mpc^{-1}$ and EDE model with $H_0 = 71.15\ km\ s^{-1}Mpc^{-1}$ respectively [from 952].

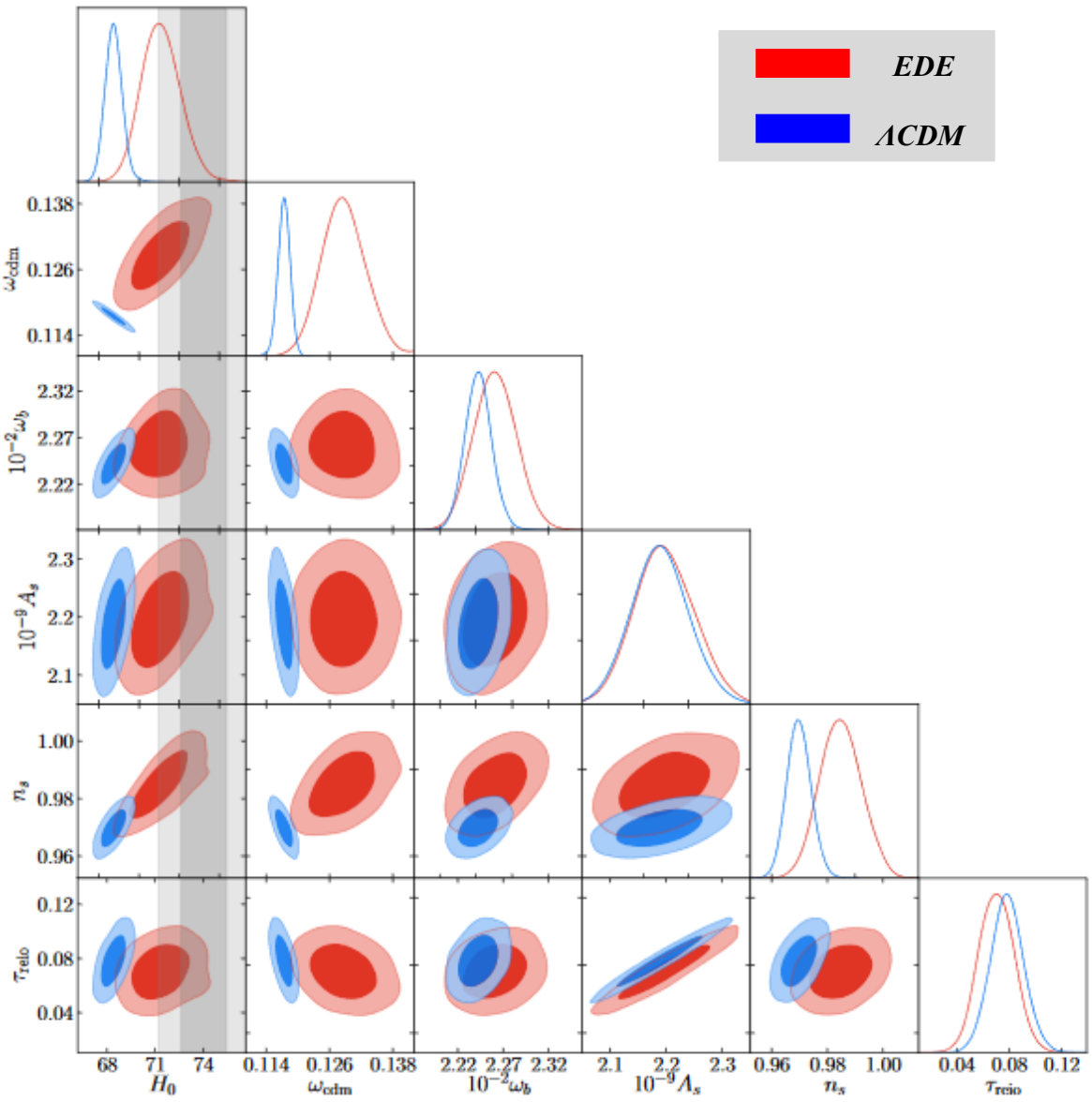


Figure 135: Posterior 1D and 2D distributions of the cosmological ΛCDM parameters reconstructed from a run to all data (including Planck high l polarization) in EDE (red) and the ΛCDM (blue) scenario. The gray bands correspond to the SH0ES determination of H_0 [adapted from 1650].

$$= \int_{z_d}^{\infty} \frac{c_s(z)}{H(z;\rho_b,\rho_\gamma,\rho_c,\rho_{DE})} dz = \\ = \int_0^{a_d} \frac{c_s(a)}{a^2 H(a;\rho_b,\rho_\gamma,\rho_c,\rho_{DE})} da \quad (120)$$

The baryon-to-photon ratio, and thus c_s in Eq. (59), is tightly constrained by CMB fluctuations and BBN [57]. As a consequence a EDE phase before and around the recombination epoch would increase $H(z)$ and thus decrease the scale of the sound horizon r_s in Eq. (120). In the context of the degeneracy $H_0 r_s$ shown in Eq. (60) this decrease of r_s leads to an increased value of H_0 for a fixed measured value of θ_s.

An EDE model can be implemented by several functional forms of scalar field which contribute to the cosmic energy shortly before matter-radiation equality. Possible functional forms of scalar field are the axion-like potential (higher-order periodic potential) inspired by string axiverse scenarios for dark energy [1239, 179, 1004, 1011, 1476], the single axion-like particle potential consisting of two cosine functions which unifies the inflaton and DM while reheating the universe [548, 549], the power-law potential [11], the acoustic dark energy [222, 1184, 1881], the α-attractor-like potential [359] and others.

Poulin et al. [301] consider two physical models. One that involves an oscillating scalar field and another with a slowly-rolling scalar field. In the case of the first model of the proposal of Poulin et al. [301], the potential of the scalar field ϕ is a generalization of the axion potential of the form

$$V(\phi) = m^2 f^2 \left(1 - cos(\phi/f)\right)^n \quad (121)$$

where m is the field mass (for ultralight scalar field $m \sim 10^{-28}\ eV$) and f is a decay constant.

Consider the time evolution of the EDE scalar field which may be written as

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0 \quad (122)$$

where the dot and the prime denote the derivatives with respect to cosmic time t and field ϕ respectively.

At early times, deep in the radiation era the field ϕ is initially frozen in its potential due to Hubble friction ($H \gg m$) and acts as a cosmological constant with equation of state $w_\phi = -1$ (hence the name Early Dark Energy), but when the Hubble parameter drops below some value ($H \sim m$) at a critical redshift z_c (for EDE this happens when $z_c \sim z_d$ for $m \sim 10^{-27}\ eV$) the field becomes dynamical and begins to oscillate around its minimum which is locally $V \sim \phi^{2n}$ (Fig. 132). It thus begins to behave like a fluid with an equation of state [1739]

$$w_\phi = \frac{n-1}{n+1} \quad (123)$$

The energy density of the field dilutes as $a^{-3(1+w_\phi)}$ and thus when $n = 1$, $n = 2$ and $n \geq 3$ dilutes as cold dark matter (a^{-3}, $w_\phi = 0$), as radiation (a^{-4}, $w_\phi = 1/3$) and faster than radiation (a^{-x} with $x > 4$, $w_\phi > 1/3$) respectively. Also when $n \to \infty$ the energy density dilutes as free scalar field (stiff matter [460]) (a^{-6}, $w_\phi = 1$) i.e. the scalar field is fully dominated by its kinetic energy.

The EDE models are parameterized by the critical redshift z_c , the dimensionless quantity $\theta_i = \phi_i/f$ (with ϕ_i the initial value of the scalar field and $0 < \theta_i < \pi$) and the peak EDE energy density fraction of the Universe $f_{EDE}(z_c)$ which is given by

$$f_{EDE}(z_c) \equiv \frac{\Omega_\phi(z_c)}{\Omega_{tot}(z_c)} = \frac{\rho_{EDE}(z_c)}{3M_{pl}^2 H(z_c)^2} \quad (124)$$

where Ω_ϕ is the EDE energy density which evolves as [1476, 301]

$$\Omega_\phi(z) = \frac{2\Omega_\phi(z_c)}{[(1+z_c)/(1+z)]^{3(w_n+1)}+1} \tag{125}$$

The fractional contribution of EDE to the cosmic energy budget as a function of redshift, i.e. $f_{EDE}(z)$, is shown in Fig. 133 [from the analysis by 952]. Clearly, for $z \simeq z_c$ the EDE contributes the most to the total energy density ($\sim 10\%$), for $z > z_c$ the EDE is not dynamically important while for $z < z_c$ decays away as radiation or faster than radiation leaving the later evolution of the Universe relatively unchanged. By construction, the EDE models can nicely match the CMB TT power spectrum of ΛCDM and therefore of Planck as illustrated in Fig. 134. The black solid and the red dashed lines (almost identical) correspond to ΛCDM model with $H_0 = 68.07\ km\ s^{-1}Mpc^{-1}$ and EDE model with $H_0 = 71.15\ km\ s^{-1}Mpc^{-1}$ respectively [952].

EDE models face the fine-tuning issues [1567] and suffer from a coincidence problem [1451]. Sakstein and Trodden [1567], Carrillo González *et al.* [447] proposed a natural explanation for this coincidence using the idea of neutrino-assisted early dark energy.

EDE modifies growth and $H(z)$ at early times (around recombination) and higher matter density is required to compensate for this effect in the CMB. Higher matter density contradicts the required low value of matter density at late times from weak lensing and growth data as shown in Fig. 135. In particular the analysis by Hill *et al.* [891], Clark *et al.* [487] has shown that an EDE model can not practically resolve the Hubble tension because it results in higher value of the late-time density fluctuation amplitude σ_8 and thus the tension with LSS dynamical probes WL, RSD and CC data can get worse. In addition Jedamzik *et al.* [971] argued that any model which attempts to reconcile the CMB inferred value of H_0 by solely reducing the sound horizon results into tension with either the BAO or the galaxy weak lensing data. Thus, a compelling and full resolution of the Hubble tension may require multiple modifications (more than just the size of the sound horizon) of the ΛCDM cosmology.

Recent studies by Smith *et al.* [1651], Chudaykin *et al.* [481] reexamining the above issue and using combined data method show that the EDE scenario remains a potential candidate solution to the Hubble tension. Future observations will provide data with improved quality and thus will enable more detailed tests of the EDE model.

Many alternative models have been proposed to implement the basic EDE scenario such as Chain EDE [741], Axion EDE [559, 976], Anti-de Sitter EDE [1796, 1877, 1879, 1878, 975, 976], assisted quintessence EDE [1552], EDE with extra radiation [1620], EDE in the framework of the ultralight scalar decay to massless fields [802] and New EDE (NEDE) [1356, 1353, 1354, 1352, 1355] which can potentially address the Hubble tension. In NEDE a vacuum first-order phase transition of the NEDE scalar field is assumed to have taken place before recombination in the early Universe. The NEDE sudden transition can be described by a scalar field whose potential at some critical point develops two non-degenerate minima (true and false vacuum)[76]. Allali *et al.* [110] develop a phenomenological dark sector with decaying dark energy and ultra-light axions which addresses the Hubble tension similarly to the EDE and NEDE scenarios and simultaneously can resolve the S_8 tension. Karwal *et al.* [1012], Sabla and Caldwell [1553], McDonough *et al.* [1264] argue that a EDE model may require a more complicated dynamics in order to soften both the H_0 and S_8 tensions. In particular, McDonough *et al.* [1264] introduced the Early Dark Sector (EDS) model considering an EDE-dependence of the mass of dark matter. The considering form of the potential is given by Eq. (121) (with

[76]It has recently been pointed out [1238, 103, 106, 1442, 1438, 1441, 100] that a similar mechanism in the context of the ultra late transition taking place at a redshift $z \lesssim 0.01$ can lead to a resolution of the Hubble tension.

$n = 3$) and the form of the field-dependent mass given by

$$m(\phi) = m_0 \exp(\frac{c\phi}{M_{pl}}) \tag{126}$$

as motivated by the the Swampland Distance Conjecture (SDC) [1389] and its extension to axions [1070, 1592, 327, 240].

Dark radiation:

Modifications in the light relic sector can relieve the tension by changing the early-time dynamics of the Universe [209, 65]. The dark radiation model assumes an increased number of light relics [786, 1841, 153, 42, 321, 503, 1781, 647, 1147, 44, 112, 702, 1752, 1621, 789, 38] which are weakly interacting components of radiation (i.e. relativistic species). For example the addition of hidden photons, sterile neutrinos [701, 443, 776, 775], Goldstone bosons, Majoron [706], axions [824, 581, 533] which are predicted in many extensions of the Standard Model (SM) increases the value in the effective number of relativistic particles N_{eff} beyond its canonical expectation value $N_{eff}^{SM} \simeq 3.044$ [1230, 100, 85, 678, 277, 743]. These extra particles modify the time of matter-radiation equality and would lead to a lower r_s sound horizon. As a consequence a lower expansion rate of the Universe and a higher value of H_0 emerges from early-time physics [818] (see Eq. (60)).

Another interesting approach was presented by Lesgourgues et al. [215], Cyr-Racine *et al.* [540], Schewtschenko *et al.* [1596], Buen-Abad *et al.* [388], Chacko *et al.* [455], Ko and Tang [1076, 1077], Tang [1702], Buen-Abad et al. [65], Krall *et al.* [1104], Archidiacono *et al.* [150], Archidiacono et al. [25], Ko *et al.* [1075], Becker *et al.* [247], Choi *et al.* [476], in which dark matter (DM) interacts with a new form of dark radiation (DR) aimed at solving H_0 tension. Assuming the Effective Theory of Structure Formation (ETHOS) paradigm [542, 1782] the interaction between the dark matter and dark radiation components is a 2-to-2 scattering DM + DR $\leftrightarrow$ DM + DR.

Neutrino self-interactions:

The strong (massive) neutrino self-interactions cosmological model can provide a larger value of H_0 and smaller σ_8, hence can resolve the tensions between cosmological datasets [206]. The strong neutrino self-interactions were proposed in Cyr-Racine and Sigurdson [541] and further studied in Lancaster et al. [209], Oldengott et al. [251]. The introduction of strong self-interacting neutrinos increases the value in the effective number of relativistic particles $N_{eff} = 4.02 \pm 0.29$ without extra neutrino species. This model modifies the standard neutrino free-streaming in the early Universe. The onset of neutrino free-streaming is delayed until close to the matter radiation equality epoch. This late-decoupling of the neutrinos shifts the CMB power spectra peaks towards smaller scales as compared to ΛCDM model. This shift modifies the scale of sound horizon r_s that can resolve the Hubble constant H_0 tension [206].

Furthermore self-interactions between the neutrinos or between other additional light relics was studied by Das and Ghosh [560], Roy Choudhury *et al.* [1539], Mazumdar *et al.* [1263], Forastieri *et al.* [727], He *et al.* [865], Berbig *et al.* [279], Lyu *et al.* [1217], Hannestad *et al.* [845], Archidiacono *et al.* [154], Chu et al. [78], Blinov *et al.* [323], Ghosh *et al.* [785], Brinckmann *et al.* [372], Archidiacono *et al.* [152], Corona *et al.* [520], Archidiacono *et al.* [155], Archidiacono et al. [24]. The strong neutrino self-interactions models are basically excluded by various existing data or experimental tests [323, 1217, 385, 580, 369]. The analysis by Brinckmann *et al.* [372] leads to conclusion that these models can not ease the Hubble tension more effectively than the ΛCDM+N_{eff} approach alone.

Models with nonstandard neutrinos - dark matter interactions were studied by Huang and Rodejohann [925], Escudero and Witte [677], Boyarsky *et al.* [351], Mangano *et al.* [1229], Wilkinson *et al.* [1829], Mosbech *et al.* [1314], Stadler *et al.* [1680], Choi *et al.* [477], Di Valentino et al. [108], Arias-Aragon *et al.* [161], Escudero and Witte [676]. These models increase the value in the effective number of relativistic particle N_{eff} and thus can provide a solution to the Hubble problem. However, in this class of models it is not possible to solve simultaneously the Hubble and growth tensions [108].

Large primordial non-Gaussianities:

The presence of large primordial non-Gaussianity in the CMB can affect the higher-order n-point correlation functions statistics. A non-vanishing primordial trispectrum ($n = 4$) which is the Fourier transform of the connected four-point correlation function leads to the non-Gaussian covariance of the angular power spectrum estimators [919, 1648, 58]. The trispectrum is nonzero when there is a strong coupling between long-wavelength (super-CMB) modes and short-wavelength modes. The non-Gaussian covariance scenario (Super-ΛCDM model) has two extra free parameters relative to those in ΛCDM and provides a larger value of H_0 reducing tension with late Universe measurements of the Hubble constant [59].

Heisenberg's uncertainty principle:

The Heisenberg's uncertainty principle [872, 1533] and the generalized uncertainty Principle [1274, 1222, 1221, 1223, 1043, 893, 1042, 1041, 1652, 1849, 1010, 61, 176, 178, 179, 175, 1365, 686, 108, 1298, 685, 1130, 1906, 1750, 1131] [see 1705, for a review] can provide constraints to the values for certain pairs of physical quantities of a particle and raise the possibility of the existence of observational signatures in cosmological data [e.g. 1434, 1645]. Capozziello *et al.* [436] have argued that the Heisenberg's uncertainty principle can provide an explanation for the Hubble constant H_0 tension. In particular the authors equate the luminosity distance (expanded for low z as in Eqs. (44) and (48)) with the photon (assumed massive) Compton wavelength

$$\lambda\!\!\!^{-}_C = \frac{\hbar}{m\,c} \tag{127}$$

and express the corresponding effective "rest mass" of the photon as a function of the cosmological redshift

$$m = \frac{\hbar H_0}{zc^2\left[1 + \frac{z}{2}(1 - q_0)\right]} \tag{128}$$

Thus, choosing $z = 1$, fixing $q_0 = -1/2$ and setting $H_0 = 74\ km\ s^{-1}Mpc^{-1}$ and $H_0 = 67\ km\ s^{-1}Mpc^{-1}$ in Eq. (128) find $m = 1.61 \times 10^{-69}kg$ and $m = 1.46 \times 10^{-69}kg$ respectively[77]. Thus using these results infer that the tension on the H_0 measurements can be the effect of the uncertainty on the photon mass i.e.

$$\frac{\Delta m}{m} = \frac{\Delta H_0}{H_0} \simeq 0.1 \tag{129}$$

Note that the non-zero photon mass could emerge through the Heisenberg's uncertainty principle and through the recent analysis of the Standard-Model Extension[78] [336, 337].

[77] The current upper limit on the photon mass is $m = 10^{-54}kg$ [1700].

[78] For studies of the massive photons in the Standard-Model Extension, see Helayël-Neto and Spallicci [873], Spallicci *et al.* [1675].

Early modified gravity:

A scalar tensor modified gravity model can be described by the following action

$$S = \int d^4x\sqrt{-g}\left[\frac{F(\sigma)}{2}R - \frac{g^{\mu\nu}}{2}\partial_\mu\sigma\partial_\nu\sigma - \Lambda - V(\sigma)\right] + S_m \tag{130}$$

where, R is the Ricci scalar, Λ is the cosmological constant, S_m is the action for matter fields, σ is a scalar field non-minimally coupled to the Ricci scalar, $F(\sigma)$ is the coupling to the Ricci scalar and $V(\sigma)$ is the potential for the scalar field. A variety of possible types of the non-minimal coupling of the scalar field to the Ricci $F(\sigma)$ and of the potential for the scalar field which can alleviate the H_0 tension by reducing the sound horizon scale through modified early cosmic expansion, has been considered in the literature [206, 357, 1537, 207, 3].

In particular Braglia *et al.* [358] introduce a model of early modified gravity[79]. This model has a non-minimal coupling of the form [358]

$$F(\sigma) = M_{pl}^2 + \xi\sigma^2 \tag{131}$$

and a quartic potential

$$V(\sigma) = \frac{\lambda\sigma^4}{4} \tag{132}$$

where λ and ξ are dimensionless parameters. For $\xi = 0$ this model reduces to the EDE model of Agrawal et al. [11]. In the early modified gravity model, gravity changes with redshift in such a way that the H_0 estimate from CMB can have larger values. Braglia *et al.* [358] have shown that this model can resolve the Hubble tension and at the same time, in contrast to an EDE model, results in lower value of the late-time density fluctuation amplitude σ_8 and thus the tension with LSS dynamical probes WL, RSD and CC data can be at least partially resolved. In general early modified gravity model compared to the EDE can provide a better fit to LSS data and can imply better predictions on LSS observables.

11.3 Other Tensions - Curiosities

In this section we provide a list of the non-standard signals in cosmological data and the curiosities of the ΛCDM cosmology beyond the Hubble tension which is currently the most widely studied and among the most statistically significant tensions. In many cases the signals are controversial and there is currently debate in the literature on the possible physical or systematic origin of these signals. For completeness we refer to all signals we could identify in the literature referring also to references that dispute the physical origin of these signals.

11.3.1 Growth tension

The Planck/ΛCDM parameter values in the context of GR indicate stronger growth of cosmological perturbations than the one implied by observational data of dynamical probes. In this section we review the observational evidence for this tension also known as the $\Omega_{0m} - \sigma_8$ tension or simply 'growth tension'.

[79] It should not be confused with the previously introduced differed model with the same name 'Early Modified Gravity' [361, 1450, 1182]. In this model gravity is allowed to be modified after BBN, before and during recombination.

Methods and data

The value of the growth parameter combination $S_8 \equiv \sigma_8(\Omega_{0m}/0.3)^{0.5}$ (where σ_8 is discussed in more detail in what follows) is found by weak lensing (WL) [838, 173, 200, 995, 174, 106, 26], cluster counts (CC) [1493, 1540, 1231, 53, 523, 524] and redshift space distortion (RSD) data [1220, 234, 1344, 1557, 1027, 1028, 1439, 1646, 162] to be lower compared to the Planck CMB (TT,TE,EE+lowE) value $S_8 = 0.834 \pm 0.016$ [72] at a level of about $2-3\sigma$ as shown[80] in Table 18 and in Fig. 136 [see 619, 30, for a recent review of this tension]. The tension is also confirmed by the latest ACT+WMAP CMB analysis [80] which finds $S_8 = 0.840 \pm 0.030$.

This is also expressed by the fact that dynamical cosmological probes (WL, RSD, CC) favor lower value of the matter density parameter $\Omega_{0m} \approx 0.26 \pm 0.04$ [94] than geometric probes (CMB, BAO, SnIa). This could be a signal of weaker gravity than the predictions of General Relativity in the context of a ΛCDM background [1220, 1734, 1344, 1027, 1646] [for a recent study on a weak gravity in the context of a ΛCDM background, see 755].

The observational evidence for weaker growth indicated by the dynamical probes of the cosmic expansion and the gravitational law on cosmological scales may be reviewed as follows:

Weak lensing:

The weak gravitational lensing from matter fluctuations along the line of sight slightly distorts the shapes (shear) and size (magnification) of distant galaxies [see 35, 1058, 1228, for a review]. This distortion is a powerful and principal cosmological probe of the mass distribution which can be predicted theoretically [195, 1001, 1832, 1785]. Using various statistical methods shape distortions can be measured by analyzing the angular shear correlation function, or its Fourier transform, the shear power spectrum [172, 174]. A special type of WL is the galaxy-galaxy lensing (GGL) [360, 927] which is the slight distortion of shapes of source galaxies in the background of a lens galaxy arising from the gravitational deflection of light due to the gravitational potential of the lens galaxy along the line of sight.

The WL surveys, the Kilo Degree Survey (KiDS) [989, 990, 991, 1116], the Subaru Hyper Suprime-Cam lensing survey (HSC) [1296, 79] and the Dark Energy Survey (DES) [23, 967] provide data useful for cosmic shear studies. In particular WL measurements of S_8 obtained from the shear catalogues by the lensing analysis of the Canada-France-Hawaii Telescope Lensing (CFHTLenS) [171, 887, 121, 882, 1289, 993] and the KiDS [174, 200] appear to be lower compared to the Planck value at a level of about 3σ. The analysis by Hildebrandt et al. [174] adopting a spatially flat ΛCDM model and using the KiDS-450 data reports $S_8 = 0.745 \pm 0.039$ which results in 2.3σ tension with the value estimated by Planck15. This KiDS-Planck discordance has also been investigated in Köhlinger et al. [200] where applying the quadratic estimator to KiDS-450 shear data reports $S_8 = 0.651 \pm 0.058$ which is in tension with the Planck2015 results at the 3.2σ level.

Using a combination of the measurements of KiDS-450 and VISTA Kilo-Degree infrared Galaxy Survey (VIKING) [169], Hildebrandt et al. [175] find $S_8 = 0.737^{+0.040}_{-0.036}$ which is discrepant with measurements from the Planck analysis at the 2.3σ level. For the KiDS+VIKING-450 (or KV450) Wright *et al.* [1836] report an updated constraint of $S_8 = 0.716^{+0.043}_{-0.038}$. Meanwhile, using the DES first year (DES-Y1) data assuming a ΛCDM model Troxel et al. [355] report $S_8 = 0.782^{+0.027}_{-0.027}$ which is in $\sim 2.3\sigma$ tension[83] with the Planck18 result. The constraint on S_8

[80]The definition $S_8 = \sigma_8(\Omega_{0m}/0.3)^{\alpha}$ with $\alpha = 1/2$ has been uniformly used for all points. In those cases where $\alpha \neq 1/2$ has been used in some references, the value of S_8 with $\alpha = 1/2$ was recalculated (along with the uncertainties) using the constraints on σ_8 and Ω_{0m} shown in those references, assuming their errors σ_{σ_8} and $\sigma_{\Omega_{0m}}$ are Gaussian. The errors of the S_8 constraints are propagated according to $\sigma^2_{S_8} = (\Omega_{0m}/0.3)^{2\alpha}\sigma^2_{\sigma_8} + \sigma^2_8\alpha^2(\Omega_{0m}/0.3)^{2\alpha-2}\sigma^2_{\Omega_{0m}}$, with $\alpha = 1/2$.

[83]This tension was calculated by Lemos *et al.* [1152]. The authors have explored a number of different methods

Table 18: The value of the structure growth parameter combination $S_8 \equiv \sigma_8(\Omega_{0m}/0.3)^{0.5}$, the matter density parameter Ω_{0m} and the the power spectrum amplitude σ_8 at 68% CL through direct and indirect measurements by different methods.

Dataset	S_8	Ω_{0m}	σ_8
CMB Planck TT,TE,EE+lowE	0.834 ± 0.016	0.3166 ± 0.0084	0.812 ± 0.007
CMB Planck TT,TE,EE+lowE+lens.	0.832 ± 0.013	0.3153 ± 0.0073	0.811 ± 0.006
CMB ACT+WMAP	0.832 ± 0.013	0.3153 ± 0.0073	0.840 ± 0.030
WL KiDS-1000	$0.759^{+0.024}_{-0.021}$	-	-
WL KiDS + VIKING + DES-Y1	$0.755^{+0.019}_{-0.021}$	-	-
WL KiDS + VIKING + DES-Y1	$0.762^{+0.025}_{-0.024}$	-	-
WL KiDS+VIKING-450	$0.716^{+0.043}_{-0.038}$	-	-
WL KiDS+VIKING-450	$0.737^{+0.040}_{-0.036}$	-	-
WL KiDS-450	0.651 ± 0.058	-	-
WL KiDS-450	0.745 ± 0.039	-	-
WL DES-Y3	$0.759^{+0.025}_{-0.023}$	$0.290^{+0.039}_{-0.063}$	$0.783^{+0.073}_{-0.092}$
WL DES-Y1	$0.782^{+0.027}_{-0.027}$	-	-
WL HSC-TPCF	$0.804^{+0.032}_{-0.029}$	$0.346^{+0.052}_{-0.100}$	$0.766^{+0.110}_{-0.098}$
WL KiDS-1000 pseudo-C_l	$0.754^{+0.027}_{-0.029}$	-	-
WL HSC-pseudo-C_l	$0.780^{+0.030}_{-0.033}$	-	-
WL CFHTLenS	$0.740^{+0.033}_{-0.038}$	-	-
WL+CMB lensing DES-Y3+SPT+Planck	$0.73^{+0.04}_{-0.03}$	$0.25^{+0.03}_{-0.04}$	$0.82^{+0.08}_{-0.07}$
WL+GC[81]	$0.795^{+0.049}_{-0.042}$	$0.383^{+0.028}_{-0.053}$	$0.718^{+0.044}_{-0.031}$
WL+GC+CMB lensing[82]	0.7781 ± 0.0094	$0.305^{+0.021}_{-0.025}$	0.774 ± 0.033
WL+GC KiDS-1000 3×2pt	$0.766^{+0.020}_{-0.014}$	$0.305^{+0.010}_{-0.015}$	$0.76^{+0.025}_{-0.020}$
WL+GC KiDS-450 3×2pt	0.742 ± 0.035	$0.243^{+0.026}_{-0.045}$	$0.832^{+0.080}_{-0.079}$
WL+GC KiDS+GAMA 3×2pt	$0.800^{+0.029}_{-0.027}$	$0.33^{+0.05}_{-0.06}$	$0.78^{+0.06}_{-0.08}$
WL+GC DES-Y3 3×2pt	$0.776^{+0.017}_{-0.017}$	$0.339^{+0.032}_{-0.031}$	$0.733^{+0.039}_{-0.049}$
WL+GC DES-Y1 3×2pt	$0.773^{+0.026}_{-0.020}$	$0.267^{+0.030}_{-0.017}$	$0.817^{+0.045}_{-0.056}$
WL+GC KiDS+VIKING-450+BOSS	0.728 ± 0.026	$0.323^{+0.014}_{-0.017}$	0.702 ± 0.029
GC BOSS DR12 bispectrum	0.751 ± 0.039	$0.32^{+0.01}_{-0.01}$	$0.722^{+0.032}_{-0.036}$
GC BOSS+eBOSS	0.72 ± 0.042	-	-
GC BOSS galaxy power spectrum	0.703 ± 0.045	0.293 ± 0.012	0.713 ± 0.045
GC BOSS power spectra	0.736 ± 0.051	0.303 ± 0.0082	0.733 ± 0.047
GC BOSS DR12	0.729 ± 0.048	$0.317^{+0.015}_{-0.019}$	0.710 ± 0.049
GC+CMB lensing DESI+Plank	0.73 ± 0.03	-	-
GC+CMB lensing unWISE+Plank	0.784 ± 0.015	0.307 ± 0.018	0.775 ± 0.029
CC AMICO KiDS-DR3	0.78 ± 0.04	$0.24^{+0.03}_{-0.04}$	0.86 ± 0.07
CC SDSS-DR8	$0.79^{+0.05}_{-0.04}$	$0.22^{+0.05}_{-0.04}$	$0.91^{+0.11}_{-0.10}$
CC ROSAT (WtG)	0.77 ± 0.05	0.26 ± 0.03	0.83 ± 0.04
CC DES-Y1	$0.65^{+0.04}_{-0.04}$	$0.179^{+0.031}_{-0.038}$	$0.85^{+0.04}_{-0.06}$
CC XMM-XXL	0.83 ± 0.11	0.40 ± 0.09	0.72 ± 0.07
CC SPT-tSZ	0.749 ± 0.055	0.276 ± 0.047	0.781 ± 0.037
CC Planck tSZ	0.785 ± 0.038	0.32 ± 0.02	0.76 ± 0.03
CC Planck tSZ	0.792 ± 0.056	0.31 ± 0.04	0.78 ± 0.04

from the combined tomographic weak lensing analysis of KiDS + VIKING + DES-Y1 adopting a flat ΛCDM model by Joudaki et al. [195] is $S_8 = 0.762^{+0.025}_{-0.024}$ which is in 2.5σ tension with Planck18 result and by Asgari *et al.* [173] is $S_8 = 0.755^{+0.019}_{-0.021}$ which is in 3.2σ tension with Planck18 result. Analysing the most recent KiDS cosmic shear data release (KiDS-1000 Kuijken *et al.* 1117) alone and assuming a spatially flat ΛCDM model the value $S_8 = 0.759^{+0.024}_{-0.021}$ was estimated by Asgari *et al.* [174]. Analysing the first-year data of HSC in the context of the flat ΛCDM model and using the pseudo-spectrum (pseudo-C_l) method[84], Hikage et al. [172] find $S_8 = 0.780^{+0.030}_{-0.033}$ and adopting the standard two-point correlation functions (TPCF) estimators, $\xi_\pm$, Hamana et al. [158] find $S_8 = 0.804^{+0.032}_{-0.029}$. Recently, a analysis of the KiDS-1000 data using pseudo-C_l method by Loureiro *et al.* [1202] has lead to $S_8 = 0.754^{+0.027}_{-0.029}$. The latest cosmic shear analysis of the DES third Year (DES-Y3) [129, 1613] in the context of the ΛCDM model constrains the clustering amplitude as $S_8 = 0.759^{+0.025}_{-0.023}$. Also, recently Chang *et al.* [458] found $S_8 = 0.73^{+0.04}_{-0.03}$ using the cross-correlations of galaxy positions and shears from DES-Y3 with CMB lensing maps from SPT and Planck.

The analysis of galaxy clustering and weak gravitational lensing of the DES-Y1 data combining three two-point functions (the so-called 3×2pt analysis) of gravitational lensing and galaxy positions (the cosmic shear correlation function, the galaxy clustering angular autocorrelation function, the galaxy-galaxy lensing cross-correlation function) by DES Collaboration [106] gives $S_8 = 0.773^{+0.026}_{-0.020}$ and $\Omega_{0m} = 0.267^{+0.030}_{-0.017}$ in flat ΛCDM model. This value is in $\sim 2.3\sigma$ tension with Planck18 result. In the latest analysis by Abbott *et al.* [29] the constraints $S_8 = 0.776^{+0.017}_{-0.017}$ and $\Omega_{0m} = 0.339^{+0.032}_{-0.031}$ in flat ΛCDM model are obtained using an improvement in signal-to-noise of the DES-Y3 3×2pt data relative to DES-Y1 by a factor of 2.1. Also, Heymans *et al.* [883] using 3×2pt analysis of KiDS-1000+BOSS+ 2-degree Field Lensing Survey[85] (2dFLenS) [311] data finds $S_8 = 0.766^{+0.020}_{-0.014}$. While previous analyses 3×2pt of KiDS+GAMA data and KiDS-450+BOSS+2dFLenS data by van Uitert et al. [358] and Joudaki *et al.* [995] obtained $S_8 = 0.800^{+0.029}_{-0.027}$ and $S_8 = 0.742 \pm 0.035$ respectively. A combined analysis of KiDS+VIKING-450+BOSS data by Tröster *et al.* [1727] resulted in $S_8 = 0.728 \pm 0.026$. Performing a Joint analysis of galaxy-galaxy weak lensing and galaxy clustering from first-year data of HSC and SDSSS-III/BOSS DR11 Miyatake *et al.* [1295] found $S_8 = 0.795^{+0.049}_{-0.042}$. Also, from a combined analysis of KiDS-1000 and DES-Y1 cosmic shear and galaxy clustering, eBOSS quasars, DESI, Planck CMB lensing data García-García *et al.* [761] obtains a constraint $S_8 = 0.7781 \pm 0.0094$.

Clearly, the tension between WL and CMB measurements is a level more than 2σ as seen in Table 18 and in Fig. 136. In addition, the tension with more recent measurements persists at the level of $\sim 2\sigma$. Finally, combined analyses of WL with galaxy clustering does not change the tension level.

Cluster counts:

Galaxy clusters which are related to peaks in the matter density field on large scales constitute a probe of the growth history of structures [682, 1421] [see 111, 1106, for a review]. Current analyses from the number counts of galaxy clusters use catalogs from surveys at different wavelengths of the electromagnetic spectrum. Such surveys include Planck[86], South Pole Telescope (SPT) and Atacama Cosmology Telescope (ACT) in the microwave (millimeter) via

to quantify the tension relative to the best-fit Planck2018 cosmology.

[84] For a realistic experiment the pseudo-C_l statistics from cut-sky maps which provide incomplete data are applied in order to obtain unbiased estimates of the angular power and cross-power spectra by correcting for the convolution with the survey window [see 899, 383, 113, 1787, for details of this method].

[85] https://2dflens.swin.edu.au

[86] https://www.cosmos.esa.int

the thermal Sunyaev-Zel'dovich (tSZ) effect[87] [1689, 1690, 1691], extended Roentgen survey with an imaging telescope array (eROSITA[88]) [1482, 1284, 1458, 902] in the X-ray that finds extended sources and measures the X-ray luminosity and temperature, Sloan Digital Sky Survey[89] (SDSS) and Dark Energy Survey[90] (DES) in the optical/NIR. These surveys find peaks in the galaxy distribution and measure the richness of the corresponding clusters. The microwave/tSZ and X-ray surveys detection techniques are based on the hot ICM [860, 319] and in some cases require auxiliary data to obtain useful constraints e.g. redshift estimates [see 1071, 318, for recent methods].

The CC method is based on the predicted halo abundance (number density) $n(M, z)$ of halos with mass less than M at redshift z which is also known as the halo mass function (HMF). This formalism was originally introduced by Press and Schechter [1483]. A general mathematical form for the comoving number density expression of haloes is [e.g. 973, 1634, 1827, 1807, 1712]

$$\frac{dn}{dM} = f(\sigma)\frac{\bar{\rho}_m}{M}\frac{d\ln\sigma^{-1}}{dM} \tag{133}$$

where $\bar{\rho}_m = \rho_{crit}\Omega_m$ is the mean matter density of the Universe, σ is the rms variance of the linear density field smoothed on a spherical volume containing a mass M, and $f(\sigma)$ is a model-dependent 'universal' halo multiplicity function[91]. There are numerous parametrizations of the multiplicity function $f(\sigma)$ based on numerical N-body simulations or theoretical models. A popular parametrization provided by Tinker *et al.* [1712] is

$$f(\sigma) = A\left[\left(\frac{\sigma}{b}\right)^{-\chi} + 1\right]e^{-c/\sigma^2} \tag{134}$$

where A, χ, b, c are four free parameters that depend on the halo definition and need to be calibrated.

Measurements of the abundance of galaxy clusters $n(M, z)$ provide consistent constraints on the density of matter Ω_{0m}, the root mean square density fluctuation σ_8, the parameter combination $S_8(\alpha) = \sigma_8(\Omega_{0m}/0.3)^\alpha$ [e.g. 1059, 172, 174] (where $\alpha \sim 0.2 - 0.6$ and $S_8 \equiv S_8(\alpha = 0.5)$), the dark energy equation-of-state w and the sum of the neutrino masses $\sum m_\nu$ (massive neutrinos can suppress the matter power spectrum on small scales and this directly affect the growth of cosmic structure) [1817]. More recently, Sabti *et al.* [1555, 1554] used a method of clustering measurements at higher redshift ($z = 4-10$) based on UV galaxy luminosity function data from the Hubble Space Telescope [see e.g. 187, 347]. They derive the large-scale matter clustering amplitude to be $\sigma_8 = 0.76^{+0.12}_{-0.14}$.

Using cluster abundance analysis in the SDSS DR8 for a flat ΛCDM cosmological model with massive neutrinos Costanzi *et al.* [523] find $S_8 = 0.79^{+0.05}_{-0.04}$. Mantz *et al.* [1232] using Weighting the Giant (WtG) [1189, 147] lensing analysis of the X-ray ROentgen SATellite (ROSAT) cluster catalogs [1728] find $S_8 = 0.77 \pm 0.05$. The analysis of the counts and weak lensing signal of of the DES-Y1 dataset by Abbott *et al.* [26] gives $S_8 = 0.65 \pm 0.04$ and $\Omega_{0m} = 0.267^{+0.030}_{-0.017}$ in flat ΛCDM. Also, assuming a flat ΛCDM model and performing a galaxy cluster abundance analysis in the AMICO KiDS-DR3 catalogue Lesci *et al.* [1154] obtains $S_8 = 0.78 \pm 0.04$.

Using galaxy clusters observed in millimeter wavelengths through the tSZ effect Salvati et al. [321] report $S_8 = 0.785 \pm 0.038$ assuming ΛCDM model. The analysis of the Planck 2015

[87] The inverse Compton scattering between CMB photons and hot electrons in the intracluster medium (ICM) [see 298, 442, 1321, for a review]

[88] https://www.mpe.mpg.de/eROSITA

[89] https://www.sdss.org/

[90] https://www.darkenergysurvey.org

[91] For a publicly available cluster toolkit Python package, see in https://cluster-toolkit.readthedocs.io/en/latest/source/massfunction.html.

cluster counts via the tSZ signal by Ade *et al.* [53] finds $S_8 = 0.792 \pm 0.056$. Recently, assuming a flat ΛCDM model, in which the total neutrino mass is a free parameter, the analysis of SPT tSZ cluster counts by Bocquet *et al.* [328] results in $S_8 = 0.749 \pm 0.055$. Using X-ray clusters detected from the XMM-XXL survey [1455] for a flat ΛCDM cosmological model Pacaud *et al.* [1391] report $S_8 = 0.83 \pm 0.10$. Also, constraints on structure growth parameter combination S_8 from cluster abundance data have been obtained by de Haan et al. [95], Abdullah *et al.* [31], Kirby *et al.* [1066], Schellenberger & Reiprich [326], Ntampaka *et al.* [1366], Zubeldia & Challinor [378]. For example using GalWCat19 [32], a catalog of 1800 galaxy clusters was derived from the SDSS-DR13 [97] and assuming a flat ΛCDM cosmology Abdullah *et al.* [31] measured the matter density and the amplitude of fluctuations to be $\Omega_m = 0.310^{+0.023}_{-0.027} \pm 0.041$ (systematic) and $\sigma_8 = 0.810^{+0.031}_{-0.036} \pm 0.035$ (systematic) respectively.

The results of S_8 from all cluster count experiments as seen in Table 18 and in Fig. 136 are in agreement with WL measurements and similarly prefer a lower value compared to the CMB measurements.

Redshift space distortion-Galaxy clustering:

Peculiar motions of galaxies falling towards overdense region generate large scale galaxy clustering, anisotropic in redshift space. Measuring this illusory anisotropy that distorts the distribution of galaxies in redshift space (i.e. RSD) we can quantify the galaxy velocity field. This important probe of LSS can be used to constrain the growth rate of cosmic structures [1000, 840, 841].

In particular the RSD is sensitive to the cosmological growth rate of matter density perturbations f which depends on the theory of gravity and is defined as [1801, 1191, 1470]

$$f(a) \equiv \frac{d \ln \delta(a)}{d \ln a} \simeq [\Omega_m(a)]^{\gamma(a)} \tag{135}$$

where $a = \frac{1}{1+z}$ is the scale factor, $\delta \equiv \frac{\delta\rho_m}{\rho_m}$ is the matter overdensity field (with ρ_m is the matter density of the background) and γ is the growth index [e.g. 1129]. The nearly constant and scale-independent value $\gamma \simeq \frac{6}{11} \simeq 0.545$ corresponds to General Relativity (GR) prediction in the context of ΛCDM [e.g. 1191].

The observable combination $f\sigma_8(a) \equiv f(a) \cdot \sigma(a)$ is measured at various redshifts by different surveys as a probe of the growth of matter density perturbations. The theoretically predicted value of this product can be obtained from the solution $\delta(a)$ of the equation [1347]

$$\delta''(a) + \left(\frac{3}{a} + \frac{H'(a)}{H(a)}\right)\delta'(a) - \frac{3}{2}\frac{\Omega_{0m} G_{eff}(a)/G}{a^5 H(a)^2/H_0^2}\delta(a) = 0 \tag{136}$$

using the definition

$$\sigma(a) \equiv \sigma_8 \frac{\delta(a)}{\delta(a=1)} \tag{137}$$

where G is Newton's constant as measured by local experiments, G_{eff} is the effective gravitational coupling which is related to the growth of matter perturbation, $\sigma(a)$ is the redshift dependent rms fluctuations of the linear density field within spheres of radius $R = 8h^{-1}Mpc$ and σ_8 is its value today.

Hence, the more robust bias free quantity $f\sigma_8$ is given by

$$f\sigma_8(a) = \frac{\sigma_8}{\delta(a=1)}\, a\, \delta'(a) \tag{138}$$

RSD growth data in the form of $f\sigma_8$[92] have been provided by wide variety of surveys including the 2-degree Field Galaxy Redshift Survey (2dFGRS) [1414, 862], VIMOS-VLT Deep Survey (VVDS) [834], SDSS [1707, 1577, 1501, 1716, 479, 182], WiggleZ [309], 6dFGS [48, 982], Galaxy and Mass Assembly (GAMA) [1640], BOSS [1825, 15, 1173, 793], Subaru Fiber Multi-Object Spectrograph (FMOS) galaxy redshift survey (FastSound) [250], VIMOS Public Extra-galactic Redshift Survey (VIPERS) [1722, 269, 1297], eBOSS [1904, 1699, 1902, 242, 1905, 1255, 94], DESI [64, 65]. Using such data the Ω_{0m} and σ_8 parameters in the context of a ΛCDM background can be constrained. Thus, Kazantzidis and Perivolaropoulos [1027] using a compilation of 63 RSD datapoints find the ΛCDM best fit value $\Omega_{0m} = 0.279 \pm 0.028$ and $\sigma_8 = 0.775 \pm 0.018$. Benisty [268] using RSD selected data and assuming ΛCDM model report $S_8 = 0.700^{+0.038}_{-0.037}$, $\Omega_{0m} = 0.201^{+0.036}_{-0.033}$ and $\sigma_8 = 0.857^{+0.044}_{-0.042}$ which are in 3σ tension with the Planck 2018 results. Recently, using RSD data and the RSD+BAO+Pantheon and RSD+BAO+Pantheon+CC dataset combinations Nunes and Vagnozzi [1369] find $S_8 = 0.739^{+0.036}_{-0.040}$, $S_8 = 0.762^{+0.030}_{-0.025}$ and $S_8 = 0.777^{+0.026}_{-0.027}$ respectively.

Galaxy clustering methods, such as the galaxy power spectrum and bispectrum have also been used to constrain S_8. Constraints from the BOSS galaxy power spectrum [953] gave $S_8 = 0.703 \pm 0.045$ and from BOSS DR12 bispectrum [1453] gave $S_8 = 0.751 \pm 0.039$. Previous analysis of the BOSS DR12 data by Tröster *et al.* [1727] gave $S_8 = 0.729 \pm 0.048$. A analysis of the power spectrum of eBOSS by Ivanov [950] resulted in $S_8 = 0.720 \pm 0.042$. Recently, using the BOSS power spectra Chen *et al.* [468] found $S_8 = 0.736 \pm 0.051$. Also, the combination of the auto- and cross-correlation signal of unWISE [93] galaxies [1598] and Planck CMB lensing maps [73] by Krolewski *et al.* [1111] gave $S_8 = 0.784 \pm 0.015$. Finally, using the luminous red galaxies of the DESI in combination with Planck CMB lensing maps White *et al.* [1826] found $S_8 = 0.73 \pm 0.03$.

Clearly, as seen in Table 18 and in Fig. 136 the analyses of RSD data gives S_8 values in tension with CMB measurements at level more than 2σ, in agreement with other dynamical cosmological probes (WL and CC).

Theoretical models

Non-gravitational mechanisms can address the S_8 tension [see 947, for a review]. Such mechanisms include the following:

- Dynamical dark energy models [1538, 1190, 506, 651, 182, 471, 1821, 700, 1345, 968, 218, 1862, 1280, 1870, 1398, 1513, 994, 639, 1754, 1903, 1133, 1388, 1866, 272, 558] and running vacuum models [1656, 1657, 795, 796, 1804, 496, 1671, 256], which modify the cosmological background $H(z)$ to a form different from ΛCDM (see Subsection 11.2.2). This modification may involve the presence of dynamical dark energy dominant at late cosmological times or at times before recombination.

- Interacting dark energy models, which modify the equation for the evolution of linear matter fluctuations as well as the $H(z)$ cosmological background [1479, 132, 222, 422] as discussed in Subsection 11.2.2. This class of models can address the structure growth σ_8 tension between the values inferred from the CMB and the WL measurements.

[92] For an extensive compilation of RSD data points $f\sigma_8$, see in Skara and Perivolaropoulos [1646] and for other compilations, see in Basilakos and Nesseris [233], Nesseris *et al.* [1344], Kazantzidis and Perivolaropoulos [1027], Sagredo *et al.* [1557], Kazantzidis *et al.* [1030]. Also for a publicly available RSD likelihood for MontePython see in Arjona *et al.* [162], Cardona *et al.* [438].

[93] Wide-field Infrared Survey Explorer (WISE) [1838] is a NASA infrared astronomy space telescope and is mapping the whole sky.

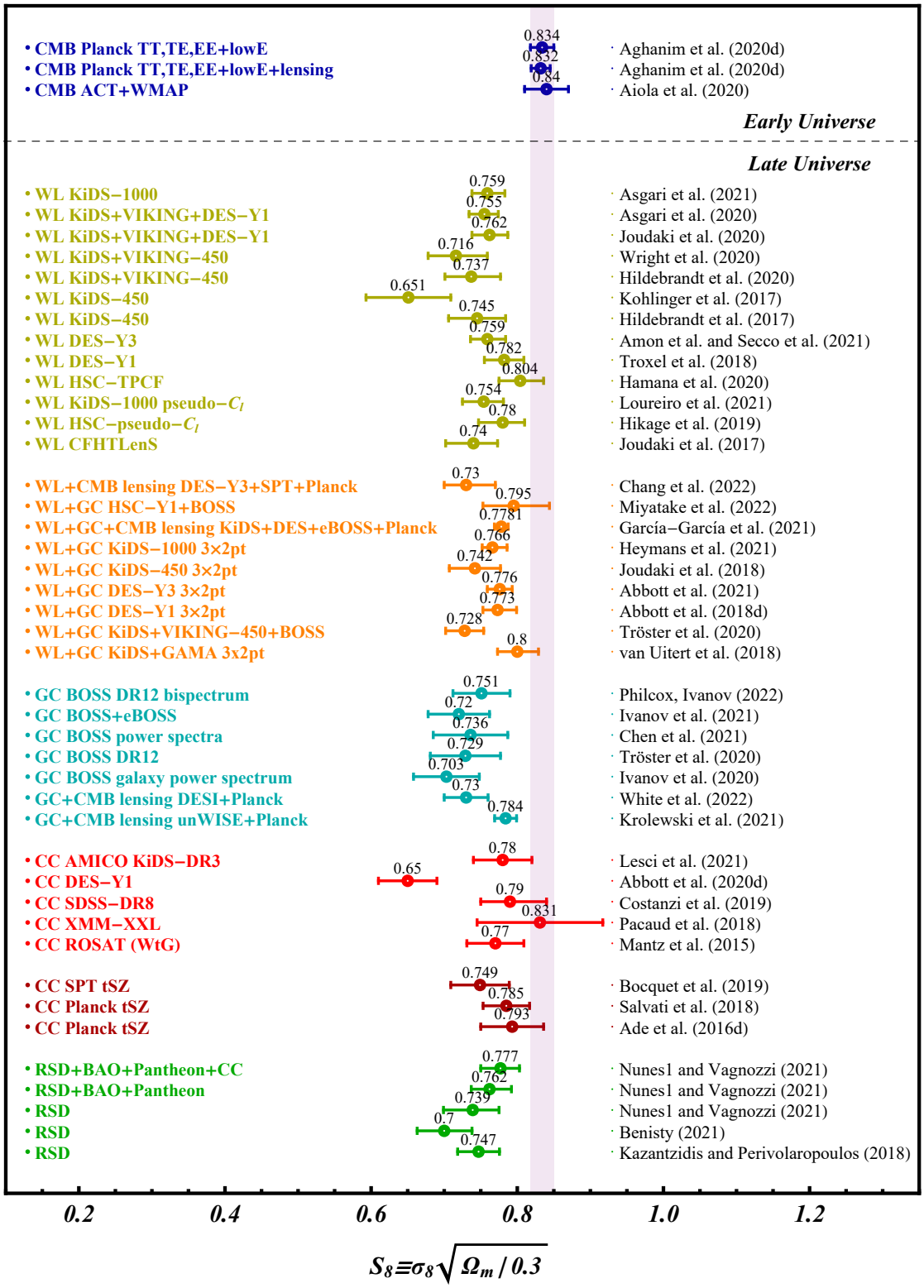


Figure 136: The value of S_8 with the 68% CL constraints derived by recent measurements.

- Effects of massive neutrinos [1250, 238, 522, 994, 1893, 1472, 623, 302] which are relativistic at early times and contribute to radiation while at late times they become non-relativistic but with significant velocities (hot dark matter) [see 216, 1835, 1157, 1158, for a review]. The change of radiation to hot dark matter affects the Hubble expansion. Simultaneously the residual streaming velocities are still large enough at late times to slow down the growth of structure [335]. This effect of massive neutrinos slows down the growth as required by the RSD data and relieves the S_8 tension coming from WL data [623].

- Primordial magnetic fields [972, 212] [see 642, 1687, 1749, for a review] induce additional mildly non-linear, small-scale baryon inhomogeneities present in the plasma before recombination. The required field results in a reduction of the sound horizon scale at recombination and has the potential to resolve both the H_0 and S_8 tension [970, 1494, 1710].

- Non-thermal dark radiation [563] seems to help alleviate the S_8 tension to a great extent. However, the inclusion of BAO data reduces significantly the quality of fit of this model.

In addition to these non-gravitational mechanisms discussed above that can slow down growth at low redshifts a possible interesting new fundamental physics approach can also reduce the S_8 tension. Such an approach is most likely to affect three basic observable parameters: the Hubble parameter $H(z, w)$ (with w the dark energy equation of state parameter), as well as the effective Newton constants for growth of perturbations

$$\mu_G(z,k) \equiv \frac{G_{eff}(z,k)}{G} \tag{139}$$

and lensing

$$\Sigma(z,k) \equiv \frac{G_L(z,k)}{G} \tag{140}$$

where G is the locally measured value of the Newton's constant. According to ΛCDM $H(z) = H(z, w = -1)$, $\mu_G = 1$, $\Sigma = 1$.

The Bardeen potentials [219] (the Newtonian potential Ψ and the spatial curvature potential Φ) appear in the scalar perturbed Friedmann-Lemaître-Robertson-Walker (FLRW) metric in the conformal Newtonian gauge [1325, 1218, 679]

$$ds^2 = -(1+2\Psi)dt^2 + a^2(1-2\Phi)d\vec{x}^2 \tag{141}$$

The LSS probes are sensitive to the Bardeen potentials Ψ and Φ. In particular the WL probe is sensitive to $\nabla^2(\Psi+\Phi)$. The galaxy clustering arises from the gravitational attraction of matter and is sensitive only to the potential Ψ. The RSD probe is sensitive to the rate of growth of matter density perturbations f (see Eq. (135)) and provides measurements of $f\sigma_8$ (see Eq. (138)) that depends on the potential Ψ.

At linear level, in modified gravity models, using the perturbed metric Eq. (141) and the gravitational field equations the following phenomenological equations in Fourier space emerge for the scalar perturbation potentials defining the functions $\mu_G(a,k)$ and $\Sigma(a,k)$ on subhorizon scales (i.e. $k/aH \gg 1$)

$$k^2(\Psi+\Phi) = -8\pi G\Sigma(a,k)a^2\rho\Delta \tag{142}$$

$$k^2\Psi = -4\pi G\mu_G(a,k)a^2\rho\Delta \tag{143}$$

where ρ_m is the matter density of the background, Δ the comoving matter density contrast defined as $\Delta \equiv \delta + 3Ha(1+w)v/k$ which is gauge-invariant [1468], $w = p/\rho$ is the equation-of-state parameter and $v^i = -\nabla^i u$ is the irrotational component of the peculiar velocity u [330].

Using the gravitational slip parameter η (or anisotropic stress parameter) which describes the possible inequality [1467, 964] of the two Bardeen potentials that may occur in modified gravity theories

$$\eta(a,k) = \frac{\Phi(a,k)}{\Psi(a,k)} \tag{144}$$

the two LSS parameters μ_G and Σ are related via

$$\Sigma(a,k) = \frac{1}{2}\mu_G(a,k)\,[1+\eta(a,k)] \tag{145}$$

The the Hubble parameter $H(z)$ is usually parametrized as wCDM

$$H(z) = H_0\left[\Omega_{0m}(1+z)^3 + (1-\Omega_{0m})(1+z)^{3(1+w)}\right]^{1/2} \tag{146}$$

while the two LSS parameters μ_G and Σ do not have a commonly accepted parametrization. A model and scale independent parametrization for μ_G and Σ which reduce to the GR value at early times and at the present time as indicated by solar system (ignoring possible screening effects) and BBN constraints ($\mu_G = 1$ and $\mu_G' = 0$ for $a = 1$ and $\mu_G = 1$ for $a \ll 1$) [1330, 1461, 756] is of the form [1346, 1344, 1027, 1646]

$$\mu_G = 1 + g_a(1-a)^n - g_a(1-a)^{2n} = 1 + g_a(\frac{z}{1+z})^n - g_a(\frac{z}{1+z})^{2n} \tag{147}$$

$$\Sigma = 1 + g_b(1-a)^m - g_b(1-a)^{2m} = 1 + g_b(\frac{z}{1+z})^m - g_b(\frac{z}{1+z})^{2m} \tag{148}$$

where g_a and g_b are parameters to be fit and n and m are integer parameters with $n \geq 2$ and $m \geq 2$.

Alternatively, a rapid transition parametrization is of the form [103, 1238]

$$\mu_G^{>}(z) = \mu_G^{<} + \Delta\mu_G\,\Theta(z-z_t) \tag{149}$$

$$\Sigma^{>}(z) = \Sigma^{<} + \Delta\Sigma\,\Theta(z-z_t) \tag{150}$$

where Θ is the Heaviside step function, z_t is a transition redshift, $\mu_G^{>}$ and $\Sigma^{>}$ correspond to $z > z_t$ and $\mu_G^{<}$ and $\Sigma^{<}$ correspond to $z < z_t$.

Various studies utilize modified gravity theories including Teleparallel theories of gravity[94] [544, 804] [see 199, for a review], Horndeski theories [1044, 1192] or theories beyond Horndeski [557] to reduce the effective Newton's constant G_{eff} at low redshifts and slow down growth at low redshifts. The above parametrizations can be realized in the context of physical models based on the above theories.

11.3.2 CMB anisotropy anomalies

There is a wide range of other less discussed no-standard signals and statistical anomalies of the large angle fluctuations in the CMB [1496] with a typical 2 to 3σ significance. As mentioned a main assumption of the ΛCDM model is that the fluctuations are Gaussian and statistically homogeneous and isotropic. Diverse anomalies have been noticed in the CMB at large angular scales by the space missions Cosmic Background Explorer (COBE) [1252], Wilkinson Microwave Anisotropy Probe (WMAP) [274] and Planck satellite [44], which appear to violate this assumption [see 1608, 88, for a review]. Perivolaropoulos [1433] presents possible explanations of the observed CMB anomalies and Cayuso and Johnson [453] explore the kinetic

[94] Many authors have studied the extensions of the Teleparallel gravity such as the scalar-torsion theories of gravity [777, 778, 803, 1396, 1647, 1085, 1153] and the Teleparallel Horndeski theories [622, 289, 198, 200, 202].

and the polarized Sunyaev-Zel'dovich effects as potential probes of physical models of these anomalies.

In what follows we discuss some of these signals. Note that some of these may not be independent[95]. Some of these signals have been attributed to the look-elsewhere effect. Based on this effect any large dataset will have a small number of peculiar features when there is a careful search for such features. However, this argument may not be applicable when the considered statistics are simple and generic as are most of the signals discussed below [see 244, 1423, 820, for a detailed discussion].

Hints for a closed Universe (CMB vs BAO)

The Universe under the assumption of the cosmological principle is described by the Friedmann-Lemaître-Roberson-Walker (FLRW) metric

$$ds^2 = dt^2 - a(t)^2 \left[\frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + sin^2\theta d\phi^2) \right] \tag{151}$$

where K characterizes the constant spatial curvature of the spatial slices with $K = -1, 0, +1$ corresponding to open hyperbolic space (negative spatial curvature), flat Euclidean space (zero spatial curvature), and closed hyperspherical space (positive spatial curvature) respectively. The curvature density parameter is defined as $\Omega_K \equiv -K/(Ha)^2$ so that a closed Universe corresponds to $\Omega_K < 0$ and an open Universe to $\Omega_K > 0$. This parameter plays a crucial role in determining the evolution of the Universe, and is closely related with the early Universe physics.

The Planck18 temperature and polarization data [72] show a preference ($\sim 3.4\sigma$) for a closed Universe ($\Omega_K < 0$) in the context of ΛCDM. In particular using these data from Planck18 the curvature density parameter was constrained to be $-0.095 < \Omega_K < -0.007$ at 99 % C.L [72, 71]. This anomaly may be connected with other asymmetries of the CMB anisotropy spectrum discussed below. The preference for closed universe however disappears when the CMB data are combined with the BAO data. Di Valentino *et al.* [621, 613], Handley [844] pointed out that Planck+BAO can give a biased result because Planck and BAO are in disagreement at more than 3σ. Combining Planck18 data with recent BAO measurements the curvature density parameter was estimated to be $\Omega_K = 0.0008 \pm 0.0019$ [621, 613, 844] in agreement with a spatially flat Universe. Using the full-shape galaxy power spectrum measurements $P(k)$, Vagnozzi *et al.* [1755] has also confirmed that the Planck data are in tension with both the full-shape power spectrum and BAO with respect to Ω_K. The recent study by Efstathiou and Gratton [655] confirms the tension between Planck and BAO data in the context of cosmic curvature. Efstathiou and Gratton [655] used a new statistical analysis (the alternative Planck CamSpec likelihood TTTEEE instead of Plik as discussed in Efstathiou and Gratton [654]) to show that Planck favors a closed Universe at more than 99% CL. However, Planck+BAO was again found to be in agreement with a spatially flat Universe with $\Omega_K = 0.0004 \pm 0.0018$ thus confirming previous studies by Di Valentino *et al.* [613], Handley [844] .

In an effort to further investigate this tension between Planck and BAO data, the analysis of Vagnozzi *et al.* [1756] combined Planck18 CMB temperature and polarization data with cosmic chronometer measurements and was lead to confirm that the Universe is consistent with spatial flatness to $\mathcal{O}(10^{-2})$ level.

A positive curvature (closed Universe) may be a plausible source of the anomalous lensing amplitude [621, 613, 844] (see Subsection 11.3.2).

[95] The covariance of CMB anomalies in the standard ΛCDM model has been studied by Muir *et al.* [1324]. This study focusing on the correlation of observed anomalies (i.e. the relationship or connection between all of them) examines the independence of large-angle CMB feature quantities.

Anomalously strong ISW effect

The decay of cosmological large-scale gravitational potential Ψ causes the integrated Sachs-Wolfe (ISW) effect [1556] which imprints tiny secondary anisotropies to the primary fluctuations of the CMB and is a complementary probe of dark energy [1086, 730, 788, 999]. Using a stacking technique in the CMB data [see 1233, 52, for a detailed discussion] anomalously strong integrated Sachs–Wolfe (ISW) signal ($> 3\sigma$) has been detected for supervoids and superclusters on scales larger than $100h^{-1}Mpc$ [815, 816]. This stronger than expected within standard ΛCDM signal of the ISW effect first emphasised in Hunt and Sarkar [932] has been studied by Nadathur *et al.* [1334], Flender *et al.* [718], Ilic *et al.* [939], Cai *et al.* [412], Kovács *et al.* [1100], Kovács [1098], Kovács *et al.* [1099], Dong *et al.* [634].

In particular the analysis by Kovács *et al.* [1099] for DES data alone found an excess ISW imprinted profile with $A_{ISW} \equiv \Delta T^{data}/\Delta T^{theory} \approx 4.1 \pm 2.0$ amplitude (where $A_{ISW} = 1$ corresponds to the ΛCDM prediction). Also a combination with independent BOSS data leads to $A_{ISW} = 5.2 \pm 1.6$. This is in 2.6σ tension with ΛCDM cosmology.

The average expansion rate approximation (AvERA) inhomogeneous cosmological simulation [1490] uses the separate Universe conjecture to calculates the spatial average of the expansion rate of local mini-Universes predicts. It indicates under the inhomogeneity assumption, about $\sim 2-5$ times higher ISW effect than ΛCDM depending on the l index of the spherical power spectrum [246]. Thus large scale spatial inhomogeneities could provide an explanation to this ISW excess signal. Giannantonio *et al.* [787] use angular cross-correlation techniques and combines several tracer catalogues to report $A_{ISW} \approx 1.38 \pm 0.32$.

Vagnozzi [1753] investigated the early Integrated Sachs-Wolfe (eISW) effect [see e.g. 348, 753] which is assumed to occur soon after recombination ($30 < z < 1100$), due to the presence of a non-negligible radiation. Constraints were thus imposed on the parameter A_{eISW} introduced by Hou *et al.* [913]. Using Planck CMB data, this parameter was constrained to $A_{eISW} = 0.988 \pm 0.027$, in perfect agreement with ΛCDM. Note that in previous studies the parameter A_{eISW} was constrained to $A_{eISW} = 0.979 \pm 0.055$ using data from WMAP7+SPT [913], to $A_{eISW} = 1.06 \pm 0.04$ from the Planck 2015 data release [404], and to $A_{eISW} = 1.064 \pm 0.042$ from the Planck 2018 temperature data alone [998].

In general the reported A_{ISW} amplitude varies in the literature depending on the dataset and the assumptions of the analysis. Further investigation of this issue is needed.

CMB cold spot

The cold (blue) spot was first found in WMAP 1-year temperature data by Vielva *et al.* [1777] and was confirmed in Planck data [56, 52, 88] in the southern hemisphere at the galactic longitude and latitude $(l, b) = (209^0, -57^0)$. It is a statistical anomaly of the large-angle fluctuations in the CMB indicating non-Gaussian features. This inconsistency with Gaussian simulations has a p-value of $\sim 1\%$.

The cold spot is an unusually large region of low temperature with the mean temperature decrement $\Delta T \approx -100\,\mu K$ and is not consistent with the prediction of gaussianity of the standard ΛCDM model [529, 530, 527].

Zhang and Huterer [1899], Nadathur *et al.* [1335], Kovács [1098] pointed out that the anomalous nature of the cold spot corresponds to a rather cold area with an angular radius in the sky of about $5^0 - 10^0$ from the centre surrounded by a hot ring.

Possible approaches for the explanation of the Cold Spot include: non-Gaussian feature due to a large statistical fluctuation [1777], an artifact of inflation [529], the foreground [530, 847], multiple voids [1336], the imprint of a supervoid (about $140-200Mpc$ radius completely empty void at $z \leq 1$) through the ISW effect [940, 941, 1548, 815], the axis of rotation of the Universe [963], cosmic texture [529, 1907], adiabatic perturbation on the last scattering surface [1762]

[see 528, 1776, for a review].

Cosmic hemispherical power asymmetry

The cosmic hemispherical power asymmetry (or dipolar asymmetry) is a directional dependency of the CMB angular power spectrum [672, 846, 673, 1392]. The continuous dipolar modulation of hemispherical power asymmetry corresponds to a hemispherical temperature variance asymmetry (signal in the CMB temperature field) [672, 846, 1301, 903, 86, 291, 52, 88, 1379].

The dipolar modulated/observed CMB temperature fluctuation $\frac{\Delta T}{T}|_{mod}$ in the direction $\hat{n}$ which appears to extend to $l_{max} \simeq 64$ can be expressed as [806, 275, 86][96]

$$\frac{\Delta T}{T}|_{mod}(\hat{n}) = [1 + A_{dm}\hat{n} \cdot \hat{p}] \frac{\Delta T}{T}|_{iso}(\hat{n}) \tag{152}$$

where $\frac{\Delta T}{T}|_{iso}$ is a statistically unmodulated/isotropic temperature fluctuation, A_{dm} denotes the amplitude of dipolar modulation and $\hat{n} \cdot \hat{p}$ corresponds to the dipolar modulation between the line-of-sight (LOS) of the observer (with unit vector $\hat{n}$) and the preferred dipolar direction (with unit vector $\hat{p}$). The amplitude of dipolar modulation A_{dm} is large at large angular scales $2 < l \lesssim 64$ ($k \lesssim 0.035 Mpc^{-1}$), small at small angular scales $l \gtrsim 64$ and vanishes by a multipole moment of $\sim 500-600$ [56, 52, 88]. The scale dependence of the hemispherical power asymmetry was suggested by Erickcek *et al.* [671], Liddle and Cortês [1179], Lyth [1216], Wang *et al.* [1794], Mukherjee and Souradeep [1329], Yang *et al.* [1850], Byrnes *et al.* [402, 403, 401], Ashoorioon and Koivisto [177], Jazayeri *et al.* [969] and was investigated by Shiraishi *et al.* [1635], Li *et al.* [1165].

According to the hemispherical asymmetry nearly aligned with the Ecliptic, the temperature fluctuations are larger on one side of the CMB sky than on the other, resulting in an unexpected dipole configuration in the CMB power spectrum with an anomalously lower value of the variance in the northern sky compared to the southern sky [56]. The preferred direction for the asymmetry from the Planck18 data is $(l, b) = (221^0, -20^0)$ in galactic coordinates and the amplitude is $A_{dm} \sim 0.07$ with statistically significant at the $\sim 3\sigma$ level [88]. This amplitude is ~ 2 times higher than expected asymmetry due to cosmic variance ($A_{dm} \sim 0.03$) and it is inconsistent with isotropy ($A_{dm} = 0$) at the $\sim 3\sigma$ level. The hemispherical power asymmetry in CMB can be explained by assuming a superhorizon perturbation [807, 806] or asymmetric initial states of the quantum perturbations [177].

Quadrupole-octopole alignment

The fluctuations in the standard ΛCDM model are Gaussian and statistically isotopic. Thus in harmonic space the quadrupole ($l = 2$) and octopole ($l = 3$) harmonics are expected to have independent and random orientations and shapes. The quadrupole and octopole have been observed to be planar and unexpectedly aligned with each other [1386, 1609, 510, 509, 512, 515]. This implies a violation of statistical isotropy.

In particular in this low multipole moment anomaly the quadrupole and octopole planes are found to be mutually aligned with the direction of the cosmic dipole or CMB dipole (see Subsection 11.3.3 and Table 19) and perpendicular to the Ecliptic [1608].

In order to study this large-angle anomaly one can use the maximum angular momentum dispersion [1386]

$$\langle\psi|(\hat{\mathbf{n}}_l \cdot \mathbf{L})^2|\psi\rangle = \sum_{m=-l}^{l} m^2 |a_{lm}(\hat{\mathbf{n}}_l)|^2 \tag{153}$$

[96] Note that the hemispherical dipole is distinct from the usual CMB dipole. In the former case the *power spectrum* is assumed modulated discontinuously across a circle on the sky and in the second the actual temperature map has a component modulated by a smooth cosine function across the sky [275].

where the CMB map is represented by a wave function

$$\frac{\Delta T}{T}(\hat{\mathbf{n}}_l) \equiv \psi(\hat{\mathbf{n}}_l) \tag{154}$$

Here $a_{lm}(\hat{\mathbf{n}}_l)$ correspond to the spherical harmonic coefficients of the CMB map in a coordinate system with its z-axis in the the $\hat{\mathbf{n}}_l$-direction.

The preferred axis $\hat{\mathbf{n}}_l$ is the axis around which the angular momentum dispersion is maximized. The directions of the quadrupole $\hat{\mathbf{n}}_2$ and the octopole $\hat{\mathbf{n}}_3$ are [1386]

$$\hat{\mathbf{n}}_2 = (-0.1145, -0.5265, 0.8424) \tag{155}$$

$$\hat{\mathbf{n}}_3 = (-0.2578, -0.4207, 0.8698) \tag{156}$$

with

$$|\hat{\mathbf{n}}_2 \cdot \hat{\mathbf{n}}_3| \simeq 0.9838 \tag{157}$$

This unexpected alignment of the $\hat{\mathbf{n}}_2$ and $\hat{\mathbf{n}}_3$ directions has only a 1/62 probability of happening.

An approach in the analysis of this large-angle anomaly may also involve the use the multipole vectors [516] (an alternative to the spherical harmonics) where each multipole order l is represented by l unit vectors i.e a dipole $l = 1$ can be constructed by a vector, a quadrupole by the product of two vectors/dipoles, an octopole from three vectors/dipoles etc.

The alignment of low multipoles indicates the existence of a preferred direction in the CMB temperature anisotropy. Furthermore possible relation between the quadrupole-octopole alignment and the dipolar asymmetry has been investigated by Hoftuft *et al.* [903], Gordon [806]. A negligible relation between these anomalies was reported. However the analysis by Marcos-Caballero and Martínez-González [1234] has shown that a particular dipolar modulation including the scale dependence may be connected with the quadrupole-octopole alignment.

Lack of large-angle CMB temperature correlations

There is a lack of large-angle CMB temperature correlations as first was observed by COBE satellite [894] and was confirmed by the WMAP [43, 1676] and Planck [52, 88] temperature maps in the range $l = 2$ to 32. This is in tension with the ΛCDM prediction.

This anomaly is directly connected to the temperature T two-point angular correlation function $C^{TT}(\theta)$ of the CMB at large angular scale ($\theta \gtrsim 60^0$) which is unexpectedly close to zero [511, 512, 514]. In angular space the two-point angular correlation function is defined as

$$C^{TT}(\theta) \equiv \langle T(\hat{n}_1) T(\hat{n}_2) \rangle = \frac{1}{4\pi} \sum_l (2l+1) C_l P_l(\cos\theta) \tag{158}$$

where the average is over all pairs of directions $\hat{n}$ with $\hat{n}_1 \cdot \hat{n}_2 = \cos\theta$, $P_l(\cos\theta)$ are the Legendre polynomials and C_l is the angular power spectrum

$$C_l \equiv \frac{1}{2l+1} \sum_{m=-l}^{l} |a_{lm}|^2 \tag{159}$$

with a_{lm} the spherical harmonic coefficients of the temperature fluctuations.

The simplest and most useful statistic is $S_{1/2}$ first introduced in the WMAP first-year release [1676] in order to measure the deviation of the angular correlation function from zero at angular scales $60^0 < \theta < 180^0$. It is defined as

$$S_{1/2} = \int_{\mu_2}^{\mu_1} \left[C^{TT}(\theta)\right]^2 d(\cos\theta) \tag{160}$$

with $\mu_1 \equiv \cos\theta_1 = \cos 60^0 = 1/2$ and $\mu_2 \equiv \cos\theta_2 = \cos 180 = -1$.

A number of alternative statistics have been proposed in the literature [837, 656, 821, 88]. For example a generalization of the $S_{1/2}$ statistic suggested by Copi *et al.* [513]. This statistic known as S^{TQ} uses the two-point angular correlation function between fluctuations in the temperature T and the Stokes parameter[97] Q, $C^{TQ}(\theta)$, which can be expressed in terms of the two-point angular power spectrum, C_l^{TE} (with E the gradient mode of polarization). The significance of a test statistic can be quantified by using the p-value[98], suggested by Ade *et al.* [56].

No sufficient explanation has yet been suggested for this large-angle anomaly. Copi *et al.* [517] study the ISW effect, Aurich *et al.* [189], Bernui *et al.* [290] explore a non-trivial spatial topology of the Universe and Pranav *et al.* [1481] study the topology of the Planck CMB temperature fluctuations in order to find a possible explanation to the suppression of large-angle CMB temperature correlations. Also the low observed power in the quadrupole is a potential explanation for the lack of correlation in the temperature maps. Copi *et al.* [511] argue that there is a cancellation between the combined contributions of C_l with multipoles $l \leq 5$ and the contributions of C_l with multipoles $l \geq 6$.

Anomaly on super-horizon scales

Pranav [1480] analysed the topological characteristics of the CMB temperature fluctuation. Using mathematical investigations on persistent homology to describe the cosmic mass distribution and performing experiments on Planck 2020 data release 4 (DR4) (based on the NPIPE data processing pipeline Akrami *et al.* 90), Pranav [1480] claimed a detection of an anomalous topological signature in the Planck CMB maps indicating non-Gaussian fluctuations. In particular Pranav [1480] reports an anomaly in the behavior of the loops (a 4σ deviation in the number of loops) in the observed sky compared to the analysis of the redshift evolution of structure on simulations when the ΛCDM model is considered.

The lensing anomaly

The recent Planck18 release by Aghanim *et al.* [72] has confirmed the higher compared to that expected in the standard ΛCDM model, anomalous, lensing contribution in the CMB power spectra which is quantified by the phenomenological parameter, A_L [66, 1889]. This weak lensing parameter A_L rescales the lensing potential power spectrum as[99]

$$C_l^{\Psi} \rightarrow A_L C_l^{\Psi} \tag{161}$$

where $A_L = 0$ corresponds to unlensed while $A_L = 1$ is the expected lensed result [66] measuring the lensing effect in the CMB temperature power spectrum.

Since the main impacts of lensing on the CMB temperature power spectrum are to add power at small scales and to smooth the structure of the acoustic peaks and troughs (the peaks are reduced slightly, and the troughs between them filled in) [220, 51] the adding of parameter A_L changes the amount of smoothing of the CMB primary spectra peaks and troughs. A higher lensing amplitude ($A_L > 1$) than predicted in the flat ΛCDM cosmology ($A_L = 1$) by roughly

[97] The Stokes parameters Q and U [for the Stokes parameters formalism see 955] are used to describe the state of CMB polarization [e.g. 1888, 629]. These parameters are directly related to the E and B modes [197, 1161, 395]. The polarization amplitude is given by $P = \sqrt{Q^2 + U^2}$.

[98] The probability value or p-value is the probability of measuring a test statistic equal to or more extreme as the observed one, considering that the null hypothesis is correct [56]. It provides the lower value of significance at which the model would be ruled out. A low p-value means that there is strong indication of new physics beyond the null hypothesis.

[99] Note that this is not the usual C_l but it is the additional contribution due to lensing.

10% (at the level of 2.8σ) has been found in the temperature power spectra by the Planck team [72].

It should be noted that the oscillatory residuals between the Planck temperature power spectra and the best-fit ΛCDM model in the multipole range $l \in [900, 1700]$ are in opposite phase compared to the CMB and thus phenomenologically similar to the effects of gravitational lensing [67, 1320].

A plausible explanation of the anomalous lensing amplitude is a positive curvature (closed Universe) which was investigated by Di Valentino *et al.* [621, 613], Handley [844]. Other possible sources which explain the lensing anomaly by mimicking a lensing effect are: a component of cold dark matter isocurvature (CDI) perturbation with a blue tilt [see 89, for a detailed discussion] and oscillations in the primordial power spectrum which have the same frequency but opposite phase with the acoustic peaks [72]. All these effects are degenerate with the smoothing effect of lensing.

Furthermore, the modified gravity models could be candidates for a solution of the lensing anomaly [57, 611, 1315, 601]. In particular the hints for $\Sigma_0 > 1$ (where Σ_0 the current value of parameter Σ which modifies the equation for the lensing potential i.e. Eq. (142)) are directly connected to the lensing anomaly as characterized by $A_L > 1$ [611, 601].

High-low l consistency

Addison et al. [6] pointed out that there are internal inconsistencies in the Planck TT power spectrum. The ΛCDM parameter values derived by the high l part of the CMB anisotropy spectrum ($l > 1000$) are in $2-3\sigma$ tension with the corresponding values of these parameters derived from the low l part of the spectrum ($1 < 1000$). For example the low l multipoles predict a lower value of the cold dark matter density parameter ω_c than the high l multipoles, with discrepancy at 2.5σ [6]. In addition it has been shown that the value of H_0 predicted by Planck from $l > 1000$, $H_0 = 64.1 \pm 1.7\ km\ s^{-1}Mpc^{-1}$, disagrees with the value predicted by Planck from $l < 1000$, $H_0 = 69.7 \pm 1.7\ km\ s^{-1}Mpc^{-1}$ at the 2.3σ level. Thus it is found that the value of H_0 depends on the CMB l-range examined.

This anomaly is probably related to the lensing anomaly i.e. the fact that ΛCDM is more consistent with the low l part of the spectrum that this not affected by the lensing anomaly [see 1526, 67, 71, for a discussion].

The preference for odd parity correlations

There is an anomalous power excess (deficit) of odd (even) l multipoles in the CMB anisotropy spectrum on the largest angular scales ($2 < l < 30$), [1135, 1061, 1060, 1062, 822, 823, 88]. A map consisting of odd (even) multipoles possesses odd (even) parity thus this effect may be considered as power (spectrum) asymmetry between even and odd parity map which is known as parity asymmetry.

In order to compare even and odd multipoles Kim and Naselsky [1060] consider the parity asymmetry statistic defined as the ratio $P \equiv P^+/P^-$ of quantities P^+ and P^- which represent the mean power in even and odd only multipoles respectively for the range $2 \leq l \leq l_{max}$

$$P^{\pm} = \sum_{2}^{l_{max}} \frac{\left[1 \pm (-1)^l\right] l(l+1)C_l}{4\pi} \tag{162}$$

A different statistic to quantify the parity asymmetry has been proposed by Aluri and Jain [114].

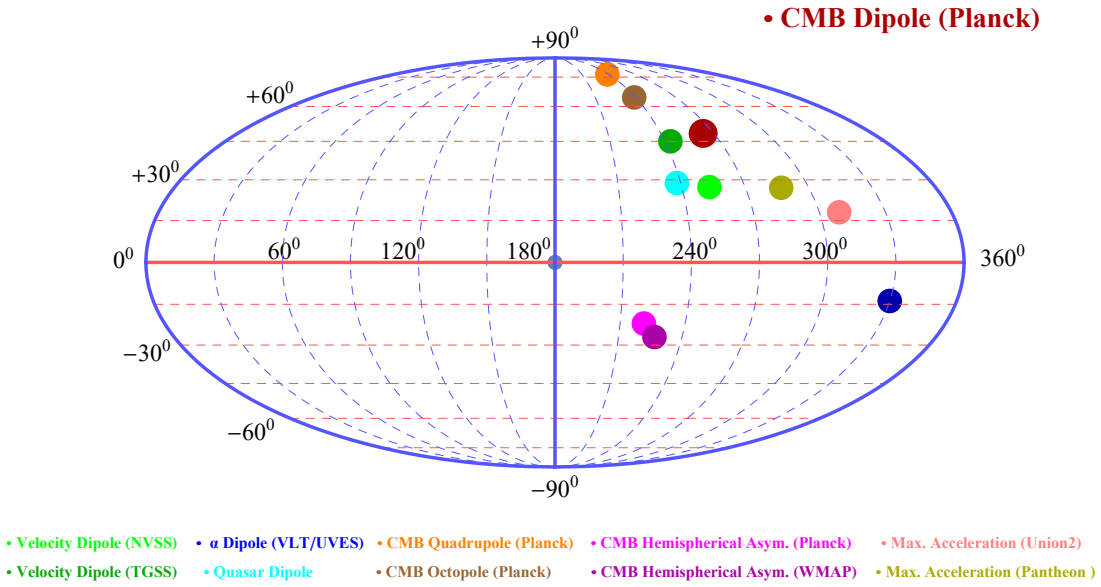


Figure 137: Mollweide-projection view of preferred directions in galactic coordinates for different cosmological observations (see Table 19).

11.3.3 Cosmic dipoles

There have been studies pointing out the presence of signals which indicate the violation of the cosmological principle. A physical mechanism producing such violation on Hubble scales is studied by Bueno Sanchez and Perivolaropoulos [390]. Various other possible mechanisms have been suggested to explain the observed violations of statistical isotropy e.g. superhorizon perturbations which introduce a preferred direction in our Universe [807, 670] [see also 1433, for a review]. The dipole amplitudes and the directions (l, b) (galactic coordinates) from the different cosmological observations described below are shown in Fig. 137 and along with the corresponding references in Table 19.

The physical origin of these dipoles is described in the following subsections.

Velocity radio dipole

A large scale velocity flow dipole[101] was pointed out in Watkins *et al.* [1808], Kashlinsky *et al.* [1017]. The dipole moment of the peculiar velocity field (dipole bulk flow) which is a sensitive probe of the amplitude and growth rate of fluctuations on large scales [1084] was investigated by Watkins *et al.* [1808], Feldman *et al.* [699], Kashlinsky *et al.* [1017, 1016], Nusser and Davis [1372], Ma and Scott [1219], Turnbull *et al.* [1738], Kashlinsky *et al.* [1015], Hong *et al.* [908]. In many cases the results are controversial and there is a debate in the literature on the consistency with the ΛCDM model.

A recent detailed analysis has indicated that 'tilted observers' within the bulk flow can be misled into inferring acceleration [1729, 1730, 1731] [see also 186, for observational constraints of the deceleration parameter in a tilted universe].

[101] For peculiar velocities as variation of the Hubble expansion produced by nearby nonlinear structures see in Wiltshire *et al.* [1831] and for dipole anisotropy in radio source count, see in Bengaly *et al.* [263].

Table 19: The amplitudes and the directions (l, b) (galactic coordinates) from different cosmological observations (Fig. 137) along with the corresponding references . The amplitude of CMB dipole has derived using the Eq. (165) [e.g. 1614].

Observations	$l\,[deg]$	$b\,[deg]$	Amplitude	Re
CMB Dipole (Planck)	264.021 ± 0.011	48.253 ± 0.005	~ 0.007	[69,
Velocity Radio Dipole (TGSS)	243.00 ± 12.00	45.00 ± 3.00	0.070 ± 0.004	[26
Velocity Radio Dipole (NVSS)	253.12 ± 11.00	27.28 ± 3.00	0.023 ± 0.004	[26
Velocity Radio Dipole (NVSS)	253.00 ± 2.00	28.71 ± 12.00	0.019 ± 0.002	[49
Velocity Radio Dipole (NVSS)	253.00	32.00 ± 12.00	0.012 ± 0.005	[17
Quasar Dipole	238.20	28.80	0.01554	[16
α Dipole (VLT/UVES)	330 ± 15	-13 ± 10	$0.97^{+0.22}_{-0.20} \times 10^{-5}$	[10
CMB Quadrupole (Planck SMICA[100])	238.5	76.6		[5
CMB Octopole (Planck SMICA)	239.0	64.3		[5
CMB Hemispher. Asym. (Planck)	221	-22	0.07	[8
CMB Hemispher. Asym. (WMAP)	227	-27	0.07	[19
Maximum Acceleration (Pantheon)	286.93 ± 18.52	27.02 ± 6.50	0.0018 ± 0.0002	[10
Maximum Acceleration (Union2)	309^{+23}_{-3}	18^{+11}_{-10}		[14

Colin *et al.* [499] use SnIa JLA data to demonstrate that the indications for cosmic acceleration found in the SnIa data disappears if a bulk flow induced anisotropy is allowed in the SnIa data. Thus, a bulk flow dipole (at 3.9σ) aligned with the local bulk flow is identified while any monopole (which can be attributed to Λ) is consistent with zero (at 1.4σ).[102]

It is usually assumed that our local (solar system) peculiar motion with respect to the CMB rest frame produces the CMB dipole anisotropy ($l = 1$) [1087, 716] [also known as solar dipole 66, 70]. In the standard model, this implies that the LSS distribution should have a similar kinematic dipole known as the velocity dipole or radio dipole which arises from the Doppler boosting of the CMB monopole and from special relativistic aberration effects [949].

In order to describe the origin of this dipole, a population of sources with power-law spectra depending on frequency ν is usually assumed

$$S_\nu \propto \nu^{-\alpha} \tag{163}$$

where S_ν is the flux density and α[103] is an individual spectral index with typically assumed value $\alpha \sim 0.75$ [196]. The integral source counts per unit solid angle above some limiting flux density S_ν can be approximated by a power law

$$\frac{dN}{d\Omega}(> S) \propto S_\nu^{-x} \tag{164}$$

where $x \sim 1$ and can be different for each survey. An observer moving with velocity $v \ll c$ with respect to the frame in which these sources are isotropically distributed sees a dipole anisotropy

[102] A recent model independent analysis of SnIa data (Pantheon) data Arjona and Nesseris [163] implementing machine learning has confirmed a $\sim 4.5\sigma$ detection of the accelerated expansion even though that analysis did not allow for anisotropic dipole effects.

[103] The individual spectral index α should not be confused with the fine-structure constant α.

$1 + D\cos\theta$ over the sky with amplitude [667]

$$D = [2 + x(1+\alpha)]\frac{\upsilon}{c} \tag{165}$$

According to the most recent measurements the inferred velocity of the Sun relative to the CMB rest frame is [69, 70]

$$\beta \equiv \frac{\upsilon}{c} = (1.23357 \pm 0.00036) \times 10^{-3} \tag{166}$$

$$or \quad \upsilon = 369.82 \pm 0.11\, km\, s^{-1} \tag{167}$$

along the direction with galactic longitude and latitude $(l, b) = (264.021^0 \pm 0.011^0, 48.253^0 \pm 0.005^0)$ or $RA \sim 168^0$, $Dec \sim -7^0$ [69, 70].

The CMB rest frame is conventionally taken to correspond to the standard of cosmic rest frame and is assumed to be statistically homogeneous and isotropic in the context of the FLRW model. In this rest frame the Hubble flow should be most uniform (minimum Hubble variation frame) and the comoving observers should not see a kinematic dipole. However it has been observed [1831, 1271] that the dipole structure of the velocity field is less in the reference frame of the Local Group of galaxies than in the CMB frame. This persistence of the dipole structure of the velocity flow in the CMB frame at large distances is not unexpected if we are located in an underdensity [1103].

According to the standard model if the Universe is isotropic our velocity with respect to the CMB rest frame and our velocity relative to the LSS should be identical. However, as was first noted by Singal [1641], while the direction of the radio dipole is consistent with that of the CMB, the velocity of our local motion obtained from the radio dipole exceeds that obtained from the CMB dipole. Radio continuum surveys which sample the Universe at intermediate redshifts ($z \sim 1$) have been used as an excellent probe to large scale isotropy and a discrepancy between the predicted and measured amplitudes of the velocity have been revealed [791, 1545, 1714, 1715, 498, 263]. In particular the analysis by Bengaly *et al.* [263] has shown that the radio dipole using the sky distribution of radio sources from the NRAO VLA Sky Survey (NVSS) dataset [502] and TIFR GMRT Sky Survey (TGSS) dataset [263, 942, 1642] is ~ 2 and ~ 5 times larger than predicted by the mock realisations within the context of ΛCDM cosmology respectively. The above observed discrepancy between the radio and CMB dipoles has been confirmed by independent groups and could imply the existence of an anisotropic Universe.

Possible explanations of the violation of statistical isotropy are: systematics due to the incomplete sky coverage of the radio continuum surveys [307, 791, 1715], intrinsic dipole in the local LSS [705], nearby nonlinear structures of voids and walls and filaments [1831], remnant of the pre-inflationary Universe [1740] and superhorizon perturbation [784, 561].

Quasar dipole

The distribution of quasars across the sky may provide independent probe of the cosmological principle [1643]. As previously discussed there is an expected anisotropy related to the CMB dipole anisotropy (about 1 part in 1000) due to our motion with respect to the CMB rest frame.

Secrest *et al.* [1614] used mid-infrared data from the Wide-field Infrared Survey Explorer (WISE) [1838] to create reliable AGN/quasar catalogs and a custom quasar sample from the new CatWISE2020 data release [660]. It was shown that there is a statistically significant dipole in the density of distant quasars with direction $(l, b) = (238.2^0, 28.8^0)$ which is 27.8^0 away from the direction of the CMB dipole. Its amplitude was found to be 0.01554, ~ 2 times larger than predicted, with statistical significance at the 4.9σ level (or with a p-value of 5×10^{-7}) for a normal distribution. This result is in conflict with the cosmological principle.

Fine structure constant α dipole

In the past 20 years there has been interest in the possibility of the variation of the fine structure constant $\alpha \equiv e^2/(4\pi\epsilon_0\hbar c)$ (where e, ϵ_0, $\hbar$, and c are the electron charge, the vacuum permittivity, the reduced Planck's constant, and the speed of light) [1332, 1809, 1299, 282, 1812, 681, 203, 917, 1287, 1144] [for review of varying fine structure constant, see 1248].

The analysis by King *et al.* [1065], Webb *et al.* [1809] uses the "many multiplet" (MM) method [1810, 648, 1811, 1333] to analyze quasar absorption line spectra obtained using the Ultraviolet and Visual Echelle Spectrograph (UVES) [574] on the Very Large Telescope (VLT). It indicates both the violation of the cosmological principle and the spatial variation of the fine structure constant α which is approximated as a spatial dipole with direction $(l, b) = (330^0 \pm 15^0, -13^0 \pm 10^0)$ and amplitude $0.97^{+0.22}_{-0.20} \times 10^{-5}$, preferred over a simple monopole model with significance at the 4.2σ [for possible systematics in this analysis, see 423, 1005, 1159].

The variation of α across the sky was shown to be well fit by an angular dipole model of the form [1065]

$$\frac{\Delta\alpha}{\alpha} \equiv \frac{\alpha - \alpha_0}{\alpha} = C_A \cos\theta + C_B \tag{168}$$

where α_0 is the present local value, θ is the angle with respect to the dipole direction, C_A is the angular amplitude of the dipole term and C_B is a monopole term.

It is worth to note that the analysis by Martins and Pinho [1249] suggests that there are no robust indications of time or space variations of α. However recent measurements of quasar absorption-line spectra indicate a spatially dependent value of fine structure constant α at a $\sim 4\sigma$ significance level over a simple monopole (no-variation) model [640, 1828]. In addition, it is found that the fine structure constant α dipole is anomalously aligned with other dipoles and the preferred direction in $\Delta\alpha/\alpha$ is correlated with the one in the distribution of SNIa [1235, 1236].

11.3.4 BAO curiosities

As mentioned above (see Subsection 11.2.1) the BAO measurements can be classified in two classes: galaxy BAO and Lyα BAO (with Lyα auto-correlation function and Lyα-quasar cross-correlation function). A $2.5 - 3\sigma$ discrepancy between the BAO peak position in the Lyα at an effective redshift of $z \sim 2.34$ and the CMB predictions from Planck/ΛCDM cosmological model has been found [724, 29, 114].

For example, Aubourg et al. [29] use the Lyα auto-correlation function and the Lyα-quasar cross-correlation function to report the measurements of the BAO scale in the line-of-sight direction

$$D_H(z = 2.40)/r_s = 8.94 \pm 0.22 \tag{169}$$

and in the transverse direction

$$D_M(z = 2.40)/r_s = 36.6 \pm 1.2 \tag{170}$$

where $D_H(z) \equiv \frac{c}{H(z)}$ is the Hubble distance and $D_M(z) \equiv (1+z)D_A(z) = d_A(z)$ is the comoving angular diameter distance. These values are in $\sim 2.3\sigma$ tension with CMB predictions $D_H(z = 2.40)/r_s = 8.586 \pm 0.021$ and $D_M(z = 2.40)/r_s = 39.77 \pm 0.09$ by Planck 2015 flat ΛCDM cosmology [50].

The galaxy BAO peak position in the matter correlation function $\xi(s)$ (see Eq. (56) and Fig. 122) and the measurements $D_H(z = 2.40)/r_s$ and $D_M(z = 2.40)/r_s$ were found to be consistent with CMB predictions. This discrepancy between galaxy and Lyα BAO constitutes the BAO anomaly which has been investigated in Cuceu *et al.* [532], Evslin [683], Addison et al. [9].

Using new Lyα BAO measurements from the BOSS survey and from its extended version eBOSS in the SDSS DR14 the tension with CMB predictions was reduced to $\sim 1.7\sigma$ [57, 96] and from eBOSS in the SDSS DR16 to only $\sim 1.5\sigma$ [346].

Evslin [683] argues that this anomaly arises by cosmological effects at $z < 2.34$ and the tension is caused by evolution of dark energy equation of state $w(z)$ for redshift range $0.57 < z < 2.34$.

11.3.5 Parity violating rotation of CMB linear polarization (Cosmic Birefringence)

In the standard model of elementary particles and fields, parity violation is observed only in the weak interaction sector [1145, 1839]. A certain class of quintessence models should generically generate such parity asymmetric physics [450, 448]. In particular a parity violating (nearly) massless axionlike scalar field ϕ (dark matter or dark energy) would rotate CMB polarisation angles of CMB photons as they travel from the last scattering surface ($z \approx 1000$) to the present by a non-zero angle β_a (cosmic birefringence).

A Chern–Simons coupling between a time-dependent axionlike field $\phi(t)$ and the electromagnetic tensor and its dual in the Lagrangian density [e.g. 1351, 1742, 450]

$$\mathcal{L} = \frac{1}{4} g_{\phi\gamma} \phi F_{\mu\nu} \tilde{F}^{\mu\nu} \tag{171}$$

induces a cosmic isotropic birefringence angle [e.g. 848, 747]

$$\beta_a = \frac{1}{2} g_{\phi\gamma} \int_{t_s}^{t_0} \dot{\phi} dt \tag{172}$$

and produces a non-zero observed EB spectrum [1207]

$$C_l^{EB} = \frac{1}{2} \sin(4\beta_a)(C_l^{EE} - C_l^{BB}) \tag{173}$$

where $g_{\phi\gamma}$ is a Chern-Simons coupling constant which has mass-dimension -1, $\tilde{F}^{\mu\nu}$ is the dual of the electromagnetic tensor of $F_{\mu\nu}$, and t_0 and t_s are the times at present and last scattering surface, respectively.

Using a novel method developed in Minami *et al.* [1294], Minami [1291], Minami and Komatsu [1293], a non-zero value of the isotropic cosmic birefringence $\beta_a = 0.35 \pm 0.14$ deg (68% C.L) was recently detected in the Planck18 polarization data at a 2.4σ statistical significance level by Minami and Komatsu [1292]. More recently, using the latest Planck public data release 4 [91] a birefringence angle of $\beta_a = 0.30 \pm 0.11$ deg (68% C.L) was reported by Diego-Palazuelos *et al.* [624].

These recent evidences of the non zero value of birefringence poses a problem for standard ΛCDM cosmology and indicates a hint of a new ingredient beyond this model.

An axion or an axion-like particle with a weak coupling to photon as a possible source of the cosmic birefringence was investigated by Fujita *et al.* [748] and a two-axion alignment model with periodic potential was investigated by Obata [1376]. Takahashi and Yin [1698] showed that if an ultralight axion coupled to photons forms domain walls due to inflationary fluctuations, the domain-wall network can explain the hint for isotropic cosmic birefringence found by Minami and Komatsu [1292]. This model predicts a testable peculiar anisotropic cosmic birefringence as well. In contrast to the approach of Fujita *et al.* [748], this scenario explains the birefringence with the photon anomalous coefficient of the axion-like particle $\sim O(1)$. Furthermore, birefringence inducing axion-like particles could be candidates for an early dark energy resolution to the Hubble tension [748]. Bianchini *et al.* [296], Namikawa *et al.* [1338] study the anisotropic birefringence and constraints are derived. The axion field fluctuations over space and time generate anisotropic birefringence.

11.3.6 Small-scale curiosities

On small scales (on scales of hundreds of *kpc* and below) the predictions of ΛCDM model are in many cases inconsistent with observations [1114, 1816, 1337]. In particular observations on galaxy scales indicate that the ΛCDM model faces several problems in describing structures at small scales ($\lesssim 1Mpc$) [see 1112, 1113, 575, 393, 1570, for a review]. Alternative models that modify the nature of dark matter have been used to solve these problems e.g. warm [329, 5, 1775, 1601, 1774], fuzzy [920, 930, 1597, 946], self-interacting [441, 320, 760, 1735] and meta-cold dark matter [1686] [see also 1247, for a review]. Other models which have the potential to provide a solution to these problems have been proposed by Foot and Vagnozzi [725, 726], Sawala *et al.* [1590], Schewtschenko *et al.* [1596], Vogelsberger *et al.* [1782], Archidiacono *et al.* [150]. In particular Foot and Vagnozzi [725, 726] argued that the existence of a dissipative hidden dark matter sector (dark matter coupled to a massless dark photon) can solve some of these problems (core-cusp, missing satellites, and plane of satellites problem).

These small scale signals include the following:

The core-cusp curiosity

The core-cusp curiosity [1302, 721, 325] (see also Cooke *et al.* 505, Lelli 1148) refers to a discrepancy between the density of a dark matter halo profile of low-mass galaxies $\rho(r) \propto r^{-x}$ in N-body simulations (an important tool for evaluating the predictions of the ΛCDM model) with $1 \lesssim x \lesssim 1.5$ (cusp profile)[104] [1341, 1339, 1304, 711] and the astronomical observed profile with $x \sim 0$ (core profile) [1302, 565, 1341, 1339, 721, 237, 1786, 130, 779]. McGaugh and de Blok [1269] probe this problem in low surface brightness galaxies.

The missing satellites problem (or dwarf galaxy problem)

The missing satellites problem (or dwarf galaxy problem) [1018, 1073, 1303, 392] refers to an over-abundance of the predicted number of halo substructures in detailed collisionless N-body simulations compared to the observed number of satellite galaxies in the Local Group. In particular the ΛCDM model predicts orders of magnitude larger number of satellites (~ 1000) than the observed number of dwarf galaxies (~ 50) [1251, 1303].

The Too Big To Fail (TBTF) problem

The Too Big To Fail (TBTF) problem [354, 355, 764, 1406, 1717, 1007] refers to an inconsistency between the predicted mass of dark matter subhaloes in ΛCDM theory and the observed central mass of brightest satellite galaxies in the Local Group [764, 1406] (also in the MW Boylan-Kolchin *et al.* 354, 355 or in the Andromeda (M31) Tollerud *et al.* 1717).

In particular the ΛCDM predicted central densities of the most massive dark matter subhalos are systematically larger than the inferred from kinematics of the brightest Local Group satellites [1498, 355, 764]. An observed bright satellite is more likely to reside in subhalos with lower mass than is expected in a ΛCDM model. The simulated massive dark matter subhalos *'failed'* to form a comparatively bright satellite galaxy.

This problem is possibly related to the missing satellites problem but it is a distinct problem which dependents on the internal structure of subhalos or the central shapes of density profiles of satellite halos [764].

Alternative models that modify the nature of dark matter have been investigated to solve this problem: non-trivial dark matter physics [1203, 356], interaction between the dark matter and dark radiation components [1596, 1782], self-interacting dark matter [1891, 1783] and fuzzy dark matter [1597]

[104] The well know Navarro–Frenk–White profile [1340, 1341] is cusped with $\rho(r \to 0) \sim r^{-1}$.

The problem of satellite planes

In the problem of satellite planes [1115, 934, 504, 1413, 1411] several satellite galaxies of the MW, of neighboring Andromeda galaxy (M31), and of Centaurus A (CenA) are part of thin plane that is approximately perpendicular to the Galactic disk. Moreover measurement of the motions of satellite galaxies has shown that their orbits appear to be correlated [1412, 935, 1655]. This flattened structure and coherent motions of satellite galaxy systems is in inconsistency with the prediction of the ΛCDM model as inferred from simulations [1411]. The simulations based on ΛCDM cosmology indicate uncorrelated and close to isotropic satellite structures [1176, 1894]. In these simulations the observed structure formations with spatial and kinematic coherence distribution are very rare with a probability $\sim 10^{-3}$ [1413, 1411].

The angular momentum catastrophe

The angular momentum catastrophe [341] concerns a catastrophic angular momentum loss of gas during disk galaxies formation in Smooth Particle Hydrodynamics (SPH) [1300] simulations. The formed disks in simulations according to the predictions of ΛCDM have smaller scale lengths by a factor of $2-3$ compared with observed ones [394]. An axion dark matter model may resolve this discrepancy between the observed and predicted angular momentum distributions of baryons (ordinary cold dark matter) in the dwarf galaxies [217].

Baryonic Tully-Fisher Relation (BTFR)

Baryonic Tully-Fisher Relation (BTFR) [1268, 1150]. As mentioned above, the well known Tully-Fisher (TF) [1737] empirical relation connects the velocity of rotation of a spiral galaxy with its intrinsic luminosity while the Baryonic Tully-Fisher Relation (BTFR) [1270, 1773, 829] Eq. (78) is a scaling relation between the observed total baryonic mass M_b (stars plus gas) of a spiral galaxy and its rotation velocity V_c (see Subsection 11.2.1). The problem for ΛCDM model as inferred from simulations [e.g. 646] is that the BTFR leads to existence of a higher intrinsic scatter ($\sim 0.15\,dex$) and a lower slope ($s = 3$) compared to the observed ($\sim 0.10\,dex$ and $s \sim 4$) [1150]. McGaugh [1266] suggests the Modified Newtonian Dynamics (MOND) [1288] as a possible solution to this problem. However some simulations or semi-analytic approaches of galaxy formation within a ΛCDM cosmological context can reproduce a realistic BTFR slope but not its small scatter [e.g. 1569, 595, 773, 1583].

The void phenomenon

The void phenomenon[1419] refers to the emptiness of voids (the number of small galaxies in the void). Cosmological N-body simulations in the context of ΛCDM have established a clear prediction [809] that many small dark matter haloes should reside in voids [1420, 1416]. This is consistent with observations on large scales but is inconsistent with observations on small scales. In particular the local void contains much fewer galaxies than expected from ΛCDM theory [1711].

11.3.7 Age of the Universe

A lower limit can be set on the age of the Universe by the ages of the oldest stars (or oldest astrophysical objects) because on cosmological timescales they form shortly after the Big Bang. In the context of ΛCDM cosmology, the standard theory [836, 376, 33, 377, 375] and cosmological numerical simulations [1883, 898, 1844] predict that the first stars, the so-called population III (Pop III), formed in dark matter minihaloes of typical mass $M \sim 10^5 - 10^6 M_\odot$

at redshifts $z \sim 20 - 30$ (about 100 million years after the Big Bang i.e. about around the end of the cosmic dark ages) [for models indicating late, $z \sim 2 - 7$, Pop III star formation, see 1721, 983].

The age of the Universe t_* as obtained from local measurements using the ages of oldest observed stars (the so-called population II (Pop II)) in the MW appears to be larger and in some tension with the corresponding age of the Universe t_U obtained using the CMB Planck data in the context of ΛCDM [1771].

The age of the Universe in the flat ΛCDM model is an observable determined by the integral

$$t(z) = \int_0^{z_t} \frac{dz'}{(1+z')H(z')} = \frac{1}{H_0}\int_0^{z_t} \frac{dz'}{(1+z')\left[\Omega_{0m}(1+z')^3 + \Omega_{0r}(1+z')^4 + (1-\Omega_{0m})\right]^{1/2}} \tag{174}$$

where t is the cosmic time corresponding to redshift z_t. Thus the age of the Universe is $t_U = t(z_t = \infty)$.

For example the age of the MW Population II halo, metal deficient, high velocity subgiant HD-140283 (also known as Methuselah star) is estimated to be $t_* = 14.46 \pm 0.31\, Gyr$ by Bond *et al.* [334] and using new sets of stellar models is estimated to be $t_* = 14.27 \pm 0.80\, Gyr$ by VandenBerg *et al.* [1764]. These estimates of the age of this star are slightly higher ($\sim 2\sigma$) than the age of Universe $t_U = 13.800 \pm 0.024\, Gyr$ inferred by CMB Planck18 data [72] but within the errors it does not conflict with this age.

Despite of the above indications the analysis by Jimenez *et al.* [979] using new parallaxes from the Gaia space mission [1446, 1484] in place of the older HST, reports a revision of the age of HD-140283 to $t_* = 13.5 \pm 0.7\, Gyr$ which is more compatible with the age t_U inferred by Planck data. Also the analysis by Valcin *et al.* [1761] using populations of stars in globular clusters (very-low-metallicity stars) reports age of the Universe constrained to be larger than $t_* = 13.5^{+0.16}_{-0.14}\, Gyr$.

Clearly, Eq. (174) indicates that in a ΛCDM Universe the quantities H_0, t_U and Ω_{0m} are related. Therefore the determination of the age of older objects based on local Universe observations provides a test of the current cosmological model and plays an important role in the studies of Hubble and spatial curvature tensions [979, 613, 621, 1814].

11.3.8 The Lithium problem

It has long been known (since the early 80's) that absorption lines in the photospheres of old, metal-poor (Population II) halo stars in the Milky Way's halo indicate ~ 3.5 times less primordial abundance of lithium isotope ^7Li compared to the prediction of the standard BBN theory [536, 181, 538]. The observed value of the lithium abundance[105] $^7Li/H = (1.6 \pm 0.3) \times 10^{-10}$ [1914] is smaller than the theoretically expected value $^7Li/H = (5.62 \pm 0.25) \times 10^{-10}$ [272] at a level $\sim 5\sigma$. This constitutes the lithium problem [712]. No such problem exists for the observed abundances of other light elements ^2H (or D), ^3He, and ^4He that are in broad quantitative agreement with BBN predictions + WMAP/Planck cosmic baryon density Ω_b which is deduced by the CMB [539, 1700].

A number of theoretical or experimental studies in the literature have attempted to address the lithium problem [e.g. 842, 1464, 1847, 1126, 1473, 1588, 810, 1575, 912, 491, 1310, 863, 948, 938]. For example the analysis by Clara and Martins [486] shows that the variations in Nature's fundamental constants on primordial nucleosynthesis provide a possible solution to the lithium problem. Specifically, they determined that if the value of the fine-structure constant α at the primordial nucleosynthesis epoch was larger than the present one by ten parts per million of relative variation, the lithium problem could be resolved.

[105] Usually in the literature the abundance of lithium is expressed by $A(^7Li) = 12 + \log_{10}[n(^7Li)/n(H)]$ where n is the number density of atoms and 12 is the solar hydrogen abundance.

It was also proposed by Di Bari *et al.* [594] that decaying dark matter into dark radiation in the early Universe can solve the long-standing lithium problem, leaving completely unaffected the abundance of other light elements. This mechanism was also proposed to alleviate the H_0 tension (see Subsection 11.2.2) but is severely constrained by the Planck data [135].

Measurements of lithium [e.g. 1591, 1279] may not be representative of the cosmological production mechanism [1679, 943]. It is thus possible that the solution to the lithium problem lies in the effects of stars in the lithium abundance. Therefore a precise knowledge of the stellar formation process and physics of stellar atmosphere is necessary to provide a fully satisfactory solution. Thus, possible solutions to this persistent problem can be classified into four categories [see 712, 539, 1253, for a review]:

- Cosmological solutions (e.g. new theory beyond the standard BBN including variations of fundamental constants) [732, 486, 1137, 1089, 281, 1025, 470, 490, 912, 1210]
- Nuclear Physics solutions (e.g. reactions destroy lithium during or after BBN) [537, 353, 456, 489, 374, 1310, 863, 948]
- Astrophysical solutions (e.g. stars destroy lithium after BBN) [1460, 1516, 1096, 745]
- Extensions of the standard model (e.g. simultaneous imposition of photon cooling after BBN, X-particle decay and a primordial magnetic field [1847, 1848], destruction of ^{7}Be due to the decay of a sterile neutrino [1575] and including new particles or interactions [810].

11.3.9 Quasars Hubble diagram

The quasar distances can be estimated from their X-ray (coronal) emission generated by a plasma of hot relativistic electrons around the accretion disk. The emission is induced through inverse-Compton scattering processes and UV emission generated by the accretion disk where the gravitational energy of the infalling material is partially converted to radiation [1532, 1213].

In recent years model independent derivation[106] of the distance modulus–redshift relation using high-z quasars ($z \lesssim 7$) as distance indicators (quasars Hubble diagram) provides a new bright standard candle in the higher redshifts and earlier times beyond SnIa. The method used is based on a non-linear relation between the X-ray and the UV emissions at low redshift which is of the form [192, 1214, 1682, 1531, 1532]

$$\log_{10} L_X = \gamma_q \log_{10} L_{UV} + \beta_q \tag{175}$$

where L_X and L_{UV} are the rest-frame monochromatic luminosities at 2 keV and at 2500 Å, respectively [192]. Also $\gamma_q \sim 0.6$ [1682, 1214, 1576, 1778, 1751, 301, 1532] and β_q are fitting parameters of the luminosities.

Extending a Hubble diagram up to redshift $z = 5.5$ shows hints for phantom dark energy [1532, 210, 1213]. In particular the distance modulus-redshift relation for a sample of 1598 quasars at higher redshift ($0.5 < z < 5.5$) is in disagreement with the concordance model at a $\sim 4\sigma$ significance level[107] [1532]. Moreover, the analysis by Lusso *et al.* [1213] building a Hubble diagram by combining three samples of Pantheon, quasars, and gamma-ray bursts (GRBs) reported tension at more than the $\sim 4\sigma$ statistical level with the flat ΛCDM model.

[106] Yang *et al.* [1851] argued that even though the data used in this approach are valid, their analysis involves significant uncertainties as it may lead to spurious artificial tensions.

[107] The analyses of the high$-z$ quasar data has lead to a wide range of conclusions [1281, 1851, 1049, 1169, 1768]. For example Yang *et al.* [1851] concludes that the log polynomial expansion generically fails to recover flat ΛCDM beyond $z \sim 2$, thus implying that the previously derived $\sim 4\sigma$ tension may be artificial.

However Khadka and Ratra [1050] using an updated, larger QSO dataset [1215] containing 2421 QSO measurements with redshifts up to $z \sim 7.5$ have demonstrated that the L_X-L_{UV} relation parameter values depend on the cosmological model thus cannot be used to constrain cosmological parameters. Recently Khadka and Ratra [1051] demonstrated that the parameter values of the largest of seven sub-samples in this QSO dataset depend on the cosmological model while the second and third biggest sub-samples appear standardizable.

11.3.10 Oscillating signals in short range gravity experiments

The most constraining test of gravity at very short distance (sub-millimeter) scales looking for departures from Newtonian gravity is implemented via torsion balance experiments. A reanalysis of short range gravity experiments has indicated the presence of an oscillating force signal with sub-millimeter wavelength [1435, 146]. In particular Perivolaropoulos [1435] has indicated the presence of a signal at 2σ level of spatially oscillating new force residuals in the torsion balance data of the Washington experiment [1008]. As an extension of the previous analysis the study by Antoniou and Perivolaropoulos [146] using Monte Carlo simulation and analysing the data of the Stanford Optically Levitated Microsphere Experiment (SOLME) which involves force measurements an optically levitated microsphere as a function of its distance z from a gold coated silicon cantilever [1518] reports a oscillating signal at about 2σ level.

The sub-millimeter scale of the quantum nature of dark energy may be written as

$$\lambda_{de} \equiv \sqrt[4]{\frac{\hbar c}{\rho_{de}}} \approx 0.085\, mm \tag{176}$$

where it is assumed that $\Omega_{0m} = 0.3$ and $H_0 = 70\ km\ s^{-1}\ Mpc^{-1}$.

Thus, if the accelerating expansion of the Universe is connected with effects of modified gravity due to quantum gravity [46] it would be natural to expect some modification of Newton's law at the submillimeter scale.

The deviations from Newton's law of gravitation is usually described in the context of scalar-tensor [679, 756] and flat extra dimension theories [164, 165, 144, 1443, 1428, 719, 1039] by a short range Yukawa type potential of the form

$$V_{eff} = -G\frac{M}{r}\left(1 + \alpha_Y e^{-mr}\right) \tag{177}$$

where α_Y and m are parameters to be constrained by the data.

Alternatively, a power law ansatz may also generalize the gravitational potential to the form

$$V_{eff} = -G\frac{M}{r}\left[1 + \bar{\beta}^k \left(\frac{1}{mr}\right)^{k-1}\right] \tag{178}$$

This power law parametrization is motivated by some brane world models [635, 267, 378, 1359].

For $m^2 < 0$ the Yukawa gravitational potential becomes oscillating and takes the form

$$V_{eff} = -G\frac{M}{r}\left[1 + \alpha_Y \cos\left(mr + \theta\right)\right] \tag{179}$$

where θ is a parameter.

Recently a reanalysis of the data of the Washington experiment searching for modifications of Newton's Law on sub-millimeter scales by Perivolaropoulos and Kazantzidis [1439] has indicated that a spatially oscillating signal is hidden in this dataset. In addition it is shown that even

though this signal cannot be explained in the context of standard modified gravity theories[108] (viable scalar tensor and $f(R)$ theories), it occurs naturally in nonlocal (infinite derivative) gravity theories [650, 1038, 742] that predict such spatial oscillations without the presence of ghosts (instabilities) and has a well-defined Newtonian limit.

The origin of oscillating signals could be due to three possible effects:

- A statistical fluctuation of the data.
- A periodic distance-dependent systematic feature in the data.
- A signal for a short distance modification of GR (e.g. non-local modified theory of gravity).

In the later case, it is important to identify modified theories that are consistent with such an oscillating signal and are not associated with instabilities [e.g. 1718, 303].

11.3.11 Anomalously low baryon temperature

The Experiment to Detect the Global Epoch of Reionization Signature (EDGES) collaboration [349] report anomalously low baryon temperature $T_b \approx 4K$ at $z \approx 17$ (half of its expected value). This temperature was inferred from the detection of global (sky-averaged) 21-cm absorption signal which is centred at a frequency of $\sim 78\ MHz$. The absorption depth of cosmic CMB photons at redshifts range $15 \lesssim z \lesssim 20$ estimated by EDGES is more than twice the maximal value expected in the ΛCDM model, at $\sim 3.8\sigma$ significance.

Possible explanations of this discrepancy were investigated and various models were proposed [e.g. 890, 733, 1323, 350]. For example Hill and Baxter [890] argue that EDE can explain this anomaly.

The EDGES observation has been used to constrain various cosmological models of dark matter and dark energy [e.g. 1872, 220, 1101, 890].

11.3.12 Colliding clusters with high velocity

Observed galaxy clusters like the massive ($\sim 10^{15} M_\odot$) high-redshift ($z = 0.87$) interacting pair known as El Gordo (ACT-CL J0102-4915) [1283] have a very high relative velocity. This implies that formation of large structures may have taken place earlier than expected in ΛCDM cosmology. Asencio *et al.* [172] based on light cone tomography estimated that the too-early formation of El Gordo rules out ΛCDM cosmology at 6.16σ confidence. The early and rapid formation of clusters which consist of two colliding massive galaxy clusters at a high redshift may constitute a problem of the ΛCDM model. Asencio *et al.* [172] argue that MOND with light sterile neutrinos model as suggested by Haslbauer *et al.* [859] can resolve this issue.

[108] For a free massive scalar Perivolaropoulos and Skara [1440] investigate the physical conditions that can eliminate the tachyonic instabilities or at least drastically change their lifetime.

Table 20: Some existing and upcoming large-scale structure missions/experiments.

Experiments	Type	Probes	Redshift	Wavelengths	Operator
Euclid	Space	WL, BAO	$z \lesssim 6$	$550\,nm - 2\,\mu m$	ESA
Vera C. Rubin Obs.	Ground	WL, BAO	$z \lesssim 7.5$	$320 - 1060\,nm$	LSST
Gaia	Space	Astrometry	$z \simeq 0$	$320 - 1000\,nm$	ESA
JWST	Space	WL	$z \lesssim 15$	$0.6 - 28.3\,\mu m$	NASA-ESA-CSA
GAUSS	Space	WL 3×2pt	$z \lesssim 5$	$0.5 - 5\,nm$	

Table 21: Some existing and upcoming CMB missions/experiments.

Experiments	Type	Detectors	Frequencies[109] (GHz)	Resolution[110] $(arcmin)$	Sensitivity[111] $(\mu K\,arcmin)$	Sky Cover
Planck	Space	74	$25 - 1000$	$5 - 33$	~ 30	All
CMB S4	Ground	$500 \cdot 10^3$	$30 - 270$	$0.8 - 11$	~ 1	70%
SO LAT	Ground	$30 \cdot 10^3$	$27 - 280$	0.1	~ 6	40%
SO SATs	Ground	$30 \cdot 10^3$	$90 - 280$	0.5	~ 2	10%
SPT-3G:	Ground	$16 \cdot 10^3$	$90 - 280$	1	$\sim 3.5, 6$	10%

Table 22: Some existing and upcoming GW experiments/observatories

Experiments	Type/Detectors	Arms	Frequencies[112] (Hz)	Location	L
Adv. LIGO	Ground/Laser interf.	$2 \times 4\,km$	$10 - 10^3$	Hanford, USA	
Adv. LIGO	Ground/Laser interf.	$2 \times 4\,km$	$10 - 10^3$	Livingston, USA	
Adv. Virgo	Ground/Laser interf.	$2 \times 3\,km$	$10 - 10^3$	Pisa, Italy	
KAGRA	Undergr./Laser interf.	$2 \times 3\,km$	$10 - 10^3$	Kamioka, Japan	
CE	Ground/Laser interf.	$2 \times 40\,km$	$5 - 4 \cdot 10^3$	USA	
LISA	Space/Laser interf.	$3 \times 2.5\,Gm$	$10^{-4} - 10^{-1}$	Heliocentric orbit	
Taiji	Space/Laser interf.	$3 \times 2\,Gm$	$10^{-4} - 10^{-1}$	Heliocentric orbit	
TianQin	Space/Laser interf.	$3 \times 0.1\,Gm$	$10^{-4} - 1$	Geocentric orbit	
DECIGO	Space/Laser interf.	$4 \times 3 \times 1\,Mm$	$1 - 10$	Heliocentric orbit	
ET	Undergr./Laser interf.	$3 \times 2 \times 10\,km$	$1 - 10^4$		

11.4 Conclusions-Discussion-Outlook

In the present review, we discussed in a unified manner many existing curiosities in cosmological and astrophysical data that appear to be in some tension (2σ or larger) with the standard ΛCDM model as specified by the Planck18 parameter values. The Hubble tension is the most significant observational indication that the current standard model ΛCDM may need to be modified after more than 20 years since its establishment.

In addition to the well known tensions (H_0 tension, S_8 tension and A_L anomaly), we provided a list of the non-standard cosmological signals in cosmological data. We presented the current status of these signals and their level of significance and also referred to recent resources where more details can be found for each signal. These signals have a lower statistical significance level than the H_0 tension but may also constitute hints towards new physics. We also briefly discussed possible theoretical approaches that have been considered in order to explain the non-standard nature of these signals.

We also discussed the possible generic extensions of ΛCDM model. Generic extensions of ΛCDM may allow for a redshift dependence of the parameters w, μ_G, Σ and α as well as a possible large scale spatial dependence of these parameters which could violate the cosmological principle. *Varying fundamental constants* can potentially address the Hubble tension, the fine structure constant α dipole, the lithium problem, the growth tension, the curious SnIa M signals (variation of the SnIa absolute magnitude $\mathcal{M}$), quasar signals and the ISW CMB signal.

The strategic approach required for the identification of new physics may include the following three steps:

- Tune current missions towards verification or rejection of non-standard signals.
- Identify favored parametrizations of $H(z, w(z), r)$, $\mu_G(z, r)$, $\Sigma(z, r)$, $\alpha(z, r)$ assuming that at least some of the non-standard signals are physical.
- Identify theoretical models (field Lagrangians) that are consistent with these parametrizations that can address simultaneously more than one of these tensions. Interestingly, for example only a small subset of modified gravity models is consistent with the weak gravity + ΛCDM background [1833, 1465, 755, 754] suggested in the context of the S_8 tension.

In the next decades new observational data from existing and upcoming missions/experiments (see Tables 20, 21 and 22) will improve measurements and open up a wide range of new directions in the explanation of the curiosities of ΛCDM cosmology and understanding of cosmological physics. Here we provide an incomplete list of these missions:

- **Euclid:** The European Space Agency (ESA) Euclid mission [1138] is planned for launch in July-December 2022. The goals of Euclid are to investigate the nature of dark matter, dark energy and gravity and thus to provide a better knowledge of the origin of the accelerated expansion of the Universe [1138, 1499, 484, 1586, 127]. The optical and NIR Euclid survey using the cosmological WL and BAO probes will detect a high number of galaxy clusters up to redshift $z \sim 2$ (and possibly higher) in a redshift range that is sensitive to dark energy [45] and will provide consistent growth rate data in both the low-z and high-z regimes. Therefore, Euclid will improve significantly the constraints on cosmological parameters such as σ_8 and the mass density parameter Ω_{0m} with a precision of $\sim 10^{-3}$ for ΛCDM [1586]. The Euclid survey will also measure the equation of state parameter of dark energy w_{DE} with higher precision ($\sim 1\%$) than precursor surveys. Stochastic inhomogeneities are expected to lead to an intrinsic uncertainty in the values of cosmological parameters obtained with such high redshift surveys. The corresponding cosmic variance in the context of Euclid for the measurement of H_0 has been shown to be limited to about 0.1% [688]. Thus Euclid and other deep surveys ($z \gtrsim 0.15$) will

provide an estimation of the H_0 which will be more precise than the low redshift surveys ($z \lesssim 0.15$). Such improved constraints from Euclid in combination with contemporary surveys [1245, 1246, 1349] will allow the verification or rejection of many of the non-standard signals discussed in this review and will also help distinguish among the favored theoretical models that have been proposed for the explanation of these signals.

- **Vera C. Rubin Observatory Legacy Survey of Space and Time:** The Large Synoptic Survey Telescope (LSST), recently renamed the Vera C. Rubin Observatory LSST [954, 34] is a future survey of the southern sky planned for the beginning in 2022. The Vera C. Rubin Observatory based in Chile with an $8.4\,m$ ($6.5\,m$ effective) telescope in six bands, targeting at least $18,000\,deg^2$ of high galactic latitude sky, will provide databases including 25 billion galaxies with $\gtrsim 0.2$ arcsecond pixel sampling [1241]. The main cosmological goals of the Vera C. Rubin Observatory ground-based project are to investigate the nature of dark matter and the dynamical behavior of dark energy by measuring WL and BAO [1241]. Vera C. Rubin observatory will detect enormous number of galaxies and in combination with the Euclid BAO survey will probe an unprecedented range of redshifts. These surveys can determine $w_{DE}(z)$ in bins of redshift and their dark energy constraining power could be orders of magnitude greater than that of precursor surveys [4]. The provided improved constraints on cosmological parameters allow to address potential systematics and to ensure that any measured tension is robust. In addition, the Vera C. Rubin project would be a useful tool in testing the models which have been used to explain these tensions.

- **CMB-S4:** The fourth generation[113] (Stage-4) ground-based CMB experiment (CMB-S4) [1, 6, 7], is planned to start observations in 2027. It is anticipated to be the definitive CMB polarization experiment. The goals of CMB-S4 are to detect the signature of primordial gravitational waves in order to shed light on models of inflation, to search for previously undiscovered light relic particles in order to study the dark Universe, to map normal and dark matter in the cosmos separately and to explore the time-variable millimeter-wave sky [1]. The CMB-S4 survey in combination with external cluster surveys which are sensitive to different redshift ranges such as the Vera C. Rubin will provide detailed cluster data [1]. These data will be used to study the growth of cluster scale perturbations, to improve constraints on cosmological parameters and to test the alternative models or extensions of ΛCDM which can be used to clarify the origin of many of the tensions and non-standard signals referred in the present review. The CMB-S4 will also contribute to neutrino cosmology providing compelling sensitivity in the constraint of the effective number of relativistic species, N_{eff} and of the sum of the neutrino masses $\sum m_\nu$ (see Aghanim *et al.* 72, Di Valentino and Melchiorri 604, for cosmological constraints and Capozzi *et al.* 434, for three-neutrino analysis). This project has also the potential to constrain $\Delta N_{eff} \equiv N_{eff} - N_{eff}^{SM} \simeq 0.060$ at 95% confidence level [6]. Planck has provided a constraint $\Delta N_{eff} \simeq 0.126$ at 95% confidence level using temperature and polarization TT, TE, EE + lowE data [72]. The improved constraints on N_{eff} will enable us to test the scenarios with modifications of ΛCDM model in the light relic sector.

- **Gaia:** The Gaia satellite was launched at the end of 2013 [1446, 1484]. This European Space Agency (ESA) mission Gaia provides data that allow us to determine with high accuracy positions, parallaxes and proper motions for more than 1 billion sources. There have been two data releases GDR1 [382] and GDR2 [381] of Gaia results. Using quasars Lindegren *et al.* [1188] found that the GDR2 suffer from the parallax zero point (ZP) error. Riess *et al.* [1521] refer to this additional error as parallax offset because it is not a

[113] Planck was the third generation space mission which mapped the anisotropies of CMB.

single value but depends on the color or/and magnitude of the source and its position on the sky. Riess et al. [311] found that the parallax offset can be measured directly from the Cepheids, but with a reduced precision of the distance scale from GDR2. This reduction leads to a increased uncertainty of the determination of H_0 value by a factor of 2.5.

Recently the Gaia team presented the Gaia Early Data Release 3 (EDR3) (the full Gaia DR3 release is expected in 2022) [751] with improved parallaxes since GDR2. Using the EDR3 parallaxes and Cepheid PL relation the latest analysis of the SH0ES Team [1521] achieved a precision of 1.0% in the geometric calibration of Cepheid luminosities. The precision of the geometric calibration of Cepheids will approach 0.5% by Gaia DR4 [1521]. This higher precision will be sufficient to confirm the present H_0 tension.

- **James Webb Space Telescope:** The James Webb Space Telescope (JWST or 'Webb') [1551, 762] is a joint NASA-ESA-CSA (National Aeronautics and Space Administration -European Space Agency-Canadian Space Agency) large, cold (under $50\,K$), infrared optimized ($0.6 < \lambda < 28.3.0\,\mu m$), space telescope and its launch is currently planned for 31 October 2021. JWST is a scientific successor to HST and will extend its discoveries to higher redshifts. It is nearly twice as big as HST with $6.6\,m$ gold-plated primary mirror much larger than $2.4\,m$ of HST.

  The two main goals of this upcoming, next-generation telescope are to look much closer to the Big Bang and to investigate the light from the first stars and galaxies that formed in the Universe [see 762, for other goals]. The observational data of this mission will essentially enhance our understanding of the formation and evolution of galaxies, stars, and planetary systems. The JWST will detect galaxies out to a redshift of $z \geq 15$. It will probably be able to detect Pop III stars (see Subsection 11.3.7) in the high-redshift galaxies [1887, 1550] in a mass range $140 - 260\,M_{\odot}$ as pair-instability supernovae [1824, 857]. Various projects using the JWST observations will provide stronger nucleosynthesis constraints inside the first supernova [375] and constraints on the nature of dark matter [1606, 1225]. These constraints and other unprecedented information from JWST could potentially help address the lithium problem, explain small-scale curiosities and improve constraints on the age of Universe.

- **Simons Observatory:** The Simons Observatory (SO) [752, 54, 37] is a next generation CMB ground-based experiment. SO consists of one $6\,m$ Large Aperture Telescope (LAT) and three $0.42\,m$ Small Aperture Telescopes (SATs) at the Atacama Desert, Chile. It will provide more accurate measurements of the primary CMB temperature and polarization signals. The main targets of SO as described by Ade *et al.* [54] are: primordial perturbations, effective number of relativistic species, neutrino mass, deviations from ΛCDM, galaxy evolution (feedback efficiency and non-thermal pressure in massive halos) and reionization (measurement of duration). Also a goal of SO survey is to provide a catalog of 16.000 galaxy clusters and more than 20.000 extragalactic sources. The sky region from SO survey overlaps with many surveys such as LSST, DES, DESI and Euclid at different wavelengths [54]. This overlap is extremely beneficial as it will allow data cross correlation tests [for a detailed discussion, see 630]. Like CMB-S4, SO will provide improved constraints on the effective number of relativistic species N_{eff}, the sum of the neutrino masses $\sum m_{\nu}$ and the dark energy equation of state w_{DE} [see 54, for the forecast constraints on cosmological parameters]. Also the SO and CMB-S4 experiments will measure the primordial tensor-to-scalar ratio r to a target sensitivity of $\sigma_r \sim 0.002$ (for an $r = 0$ model). This will be an improvement by a factor of approximately 5 compared to Planck sensitivity. In addition the uncertainty of the determinations of H_0 from SO will be two and five times better than that inferred from Planck and local direct measurement

respectively. Therefore, the SO data will enable us to improve constraints on extensions of ΛCDM which alleviate its tensions and curiosities. In addition the improved quality of lensing data from SO as well as CMB-S4 will improve our understanding of the CMB anisotropy anomalies.

- **SPT-3G:** This is a third generation CMB experiment [278, 643]. It uses the third survey camera SPT-3G which was installed on the South Pole Telescope (SPT) in 2017. The SPT-3G with the 10-meter diameter telescope targets at least $1,500\,deg^2$ region of low-foreground sky in three spectral bands centered at 95, 150, and $220\,GHz$ with $\sim 16,000$ detectors (10 times more than its predecessor SPTpol Austermann *et al.* 190, Keisler *et al.* 1040).

  Its scientific goals aim to constrain the physics of the cosmic inflation, to explore the neutrino sector, and to constrain the relativistic energy density of the Universe [278, 643]. The SPT-3G survey in combination with the deep and wide optical survey DES, will provide detailed data on $\sim 200\,Mpc$ scales which may be used to test General Relativity. The SPT-3G will also provide stringent and improved constraints on the effective number of relativistic species, N_{eff} and on the sum of the neutrino masses $\sum m_\nu$ by synergy with Planck.

- **GAUSS:** This space mission concept combines the WL and galaxy clustering probes using three two-point correlation functions (3×2pt analysis) of gravitational lensing and galaxy positions: the cosmic shear, the galaxy clustering and galaxy-galaxy lensing. GAUSS aims to fully map the cosmic web up to redshift $z \sim 5$ and to provide a catalog with the spectroscopic redshifts and the shapes of 10 billions of galaxies [316] increased by a factor of approximately 10^3 compared to DESI which measures the spectra of 35 million galaxies and quasars [64]. The very large sky coverage and the high galaxy density provided by the GAUSS will facilitate the construction of the 3D matter power spectrum of all scales (large and small) in detail. The 3×2pt correlation functions in combination with 3D matter power spectrum will provide stronger constraints and break parameter degeneracies [316]. The constraining power of the GAUSS will be an order of magnitude larger than that of any currently planned projects such as Euclid and Vera C. Rubin Observatory.

- **Laser Interferometer Gravitational-Wave Observatory:** The Laser Interferometer Gravitational-Wave Observatory (LIGO) proposed by Abramovici *et al.* [39] is a large-scale experiment that uses ground-based laser interferometers with $L = L_x = L_y = 4\,km$ long orthogonal arms to detect GWs[114].

  There are two identical LIGO instruments, one in Hanford (LHO) and one in Livingston (LLO) separated by roughly $3000\,km$. The principle of operation of laser interferometers [781, 1317] is similar to that of a simple interferometer, such as that used by Michelson and Morley. Detection of GWs with strain amplitude $h \sim 10^{-21}$ by a ground detector with arms of length $L = 4\,km$ requires length change measurement [781]

$$\Delta L = \delta L_x - \delta L_y \sim hL \sim 4 \cdot 10^{-18}\,m \tag{180}$$

[114] Another class of GW detectors are the resonant mass antennas [1463, 1224, 76, 185] in the frequency range from $15\,Hz$ to few kHz. The principle of operation of mass resonance detectors is related to the periodic dimensional changes caused by the ripple effect of GWs on solid bodies. The Weber bar [1813] is a first generation resonant mass detector, the ALLEGRO [1256], NAUTILUS [116, 183, 184], EXPLORER [485], AURIGA [454], NIOBE [306, 875] are the second generation and Mario Schenberg [77], MiniGRAIL [1784] are the third generation.

For the period between 2002 and 2010, the two LIGO observatories were unable to detect GWs. The detectors were later replaced by much improved Advanced LIGO versions [2, 852]. The improved detectors that officially went into operation in 2015 have about ten times the sensitivity to detect GWs in the frequency range around $\sim 100\,Hz$ compared to the initial LIGO interferometers [2]. In addition Advanced LIGO extends the low frequency end from $40\,Hz$ down to $10\,Hz$. Much of the research and development work for LIGO/Advanced LIGO projects was based on the groundbreaking work of the GEO 600 detector [1830, 63] which is a $600\,m$ interferometer in Hanover, Germany.

On February 11, 2016 the LIGO Scientific Collaboration [12] and the Virgo Collaboration [790, 40] announced the first directly observed GWs from a signal detected on September 14, 2015 by the Advanced LIGO devices (the Virgo was not working at the time due to an upgrade). The detected signal was named GW150914 and its source was the merger of two stellar-mass BHs [13].

The Advanced Virgo [41] with $3\,km$ arm length interferometer contributes to the reliability of Advanced LIGO experimental device detections allowing for greater accuracy in locating the source in the sky (triangulation i.e 3-detector localization) [1644] and more accurate reconstruction of the signal waveform [for the relevant method, see 518, 1194, 519, 555, 783]. For example, in the case of the event GW170814 the three detectors improved the sky localization of the source, reducing the area of the 90% credible region from $1160\,deg^2$ using only the two LIGO detectors to $60\,deg^2$ using all three LIGO/Virgo detectors and reduced the luminosity distance uncertainty from $570^{+300}_{-230}\,Mpc$ to $540^{+130}_{-210}\,Mpc$ [15].

In 2019 the Advanced LIGO [2], the Advanced Virgo [41] and the Japanese successor of the Tama300 [142, 148], Kamioka Gravitational (KAGRA) wave detector [180, 1673, 92] (previously called LCGT Kuroda 1125) signed collaboration agreement to begin joint observation. The LIGO, Virgo and KAGRA collaboration will be probably complemented by other interferometers like the planned Indian LIGO by the Indian Initiative in Gravitational Wave Observations (IndIGO) consortium [1746]. In addition a future third-generation ground-based detector the Cosmic Explorer (CE) [14, 1506, 1505] is envisioned to begin operation in the 2030s in the USA. It will contribute to the GW Astronomy beyond LIGO. CE with ten times longer arms ($40\,km$) than Advanced LIGO's will amplify the amplitude of the observed signals [680, 457] and will significantly increase the sensitivity of the observations [1505, 1506].

Many events (~ 50 compact binary coalescences) were observed by the Advanced LIGO/Virgo interferometers during three observing run periods (O1, O2 and O3)[115]. The full three-detector network provided data which enabled the standard siren measurement of the Hubble constant H_0 (see Subsections 11.2.1). These data are not yet sufficiently constraining the Hubble constant but in the future they are expected to improve significantly.

- **Laser Interferometer Space Antenna:** The Laser Interferometer Space Antenna (LISA) [437, 120] is a large-scale space mission proposed by ESA, planned for launch in 2034. It will consist of three spacecrafts placed in an equilateral triangle with arms 2.5 million kilometers long which will be placed near the Earth in a heliocentric orbit. In order to pave the way for the LISA mission ESA launched LISA Pathfinder in 2015 and it was operational from 2016 to 2017 [167, 168]. The results from scientific research show

[115] For the first Gravitational Wave Transient Catalog (GWTC-1) during O1 and O2, see Abbott *et al.* [11], for second Gravitational Wave Transient Catalog (GWTC-2) from the first part of the third observing run (O3a), see Abbott *et al.* [19] and for third Gravitational Wave Transient Catalog (GWTC-3) from the second part of the third observing run (O3a), see Abbott *et al.* [22]

that LISA Pathfinder works exactly five times better than required, with a successful demonstration of the basic technologies for a large gravitational wave observatory.

LISA is designed to detect GWs in the frequency range from $0.1\,mHz$ to $10^{-1}\,Hz$ [119, 118] targeting very different source populations from ground-based detectors such as LIGO, Virgo and KAGRA which operate in the frequency range[116] from $10\,Hz$ to $10^{3}\,Hz$ [16].

There are many different sources of GWs [see 1587, 833, 405, for a review of GW physics]. LIGO and Virgo can detect the merger events of binaries with masses $\lesssim 100 M_{\odot}$ while LISA will be able to detect the merger of massive BHs ($10^{5} - 10^{7} M_{\odot}$) with higher signal-to-noise ratio (SNR) and thus to perform precision tests in the strong gravity regime of ΛCDM model. LISA will detect events lasting weeks, months or years allowing us to observe a much larger volume of the Universe. It may improve our understanding of the early Universe. In addition the LISA mission will be able to detect sources like primordial BHs ($\sim 10^{-12} M_{\odot}$) which correspond to the mHz frequency [227, 228, 409]. This possibility can help to test primordial BH dark matter scenario. Finally, LISA and the Big Bang Observer (BBO) [526, 853], which is a proposed LISA's successor will detect many other known or currently unknown exotic sources. Thus it will enable us to explore alternative gravity theories and to address the problems of the ΛCDM cosmology.

- **Taiji:** Taiji [1542, 922] meaning 'supreme ultimate' is a Chinese large-scale space mission, planned for launch in 2033. Like LISA, Taiji is a laser interferometric GW detector which will consist of three spacecraft placed in an equilateral triangle with arms 2 million kilometers long in orbit around the Sun. Taiji will detect GWs in the frequency range from $0.1\,mHz$ to $10^{-1}\,Hz$. Like LISA, Taiji can detect many possible GW sources such as a stochastic GW background generated in the early Universe and the merger of two super massive BHs.

  A potential LISA-Taiji network was explored by Ruan *et al.* [1543, 1544]. This network with a separation distance of about $0.7\,AU$ can accurately localize the sky position of a GW source and may completely identify the host galaxy.

- **TianQin:** TianQin [1209, 1276] is a Chinese large-scale space mission. It aims to launch a laser interferometric GW detector around 2035. Like other space-base observatories, TianQin observatory consist of three spacecrafts placed in an equilateral triangle with arms $\sim 0.1\,Gm$ long but in geocentric orbit with an orbital radius of about $10^{5}\,km$ [1209, 1276]. TianQin aims to detect GWs in the frequency range from $10^{-4}\,Hz$ to $1\,Hz$ (overlapping with that of LISA near $10^{-4}\,Hz$ and with that of DECIGO near $1\,Hz$). It will search for GW signals from various cosmological sources such as the inspiral of supermassive BBH [703], stellar-mass BBH [1198], the merger of massive BBHs [1797] and stochastic GW background originating from primordial BHs [593] and/or cosmic strings [1387]. As a precursor mission of TianQin, TianQin-1 experimental satellite has been launched on 20 December 2019. The results from scientific research shows that TianQin-1 satellite has exceeded all of its mission requirements.

- **Deci-hertz Interferometer Gravitational wave Observatory:** The DECi-hertz Interferometer Gravitational wave Observatory (DECIGO) is a Japanese large-scale space mission [1022, 1020, 1021, 1019] which was proposed by Seto *et al.* [1619] and is planned for launch in 2027. DECIGO consists of four clusters (with two of them at the same position) and each cluster consists of three spacecrafts placed in an equilateral triangle with $1000\,km$ arm lengths in heliocentric orbit [1024]. As a precursor mission of DECIGO,

[116] The frequency range from $0.1\,mHz$ to $10^{-1}\,Hz$ is unobservable by any proposed ground based detectors, due to seismic noise.

B-DECIGO (smaller version of DECIGO) will be launched before 2030 with 100 km arm lengths orbiting around the earth at 2000 km altitude above the surface of the earth [1589, 1023, 1024].

DECIGO is designed to detect GWs in the frequency range from 0.1 Hz to 10 Hz which is located in a gap between the frequency band of the LISA/Taiji and ground-based detectors such as advanced LIGO, advanced Virgo, and KAGRA. It aims to observe the primordial gravitational waves i.e. the beginning of the universe ($10^{-36} - 10^{-34}\,sec$ right after the birth of the Universe), the formation of giant black holes in the center of galaxies and the compact binaries, such as white dwarf binaries [691].

- **Einstein Telescope:** Einstein Telescope (ET) or Einstein Observatory is a European proposed underground laser interferometric GW detector [1485]. It will be located underground at a depth of about $100 - 300\,m$ in order to reduce the seismic noises. ET will consist of three nested detectors placed in an equilateral triangle, each in turn composed of two interferometers with arms 10 km long. Using two arms in each side of the triangle will enable the determination of the polarisation of GWs. As a third-generation observatory is targeting a sensitivity 10 times better than of current second-generation laser-interferometric detectors such as advanced LIGO, advanced Virgo, and KAGRA [885]. ET will reduce thermal noise compared to the first and second generations of GW detectors by operating the mirrors at cryogenic temperatures as low as 10 K [886]. It is planned to start observations in 2035 with two candidate sites: north of Lula in Sardinia (Italy) and in Meuse-Rhine Euroregion (the border area of Belgium, Germany, and the Netherlands) [117]. ET will detect GWs in the frequency range from $\sim 1\,Hz$ to $\sim 10\,kHz$. This will allow the detection of BNS up to a redshift of $z \sim 2$, stellar-mass BBH at $z \sim 15$, and intermediate-mass BBH ($10^2 - 10^4 M_\odot$) at $z \sim 5$. The observations of these standard sirens will be useful to calibrate the cosmic distance ladder and will improve the estimation of the Hubble constant.

Cosmology is entering an even more exciting era! The combination of the existing puzzling observational signals discussed in this review, along with the upcoming revolutionary improvement in the quality and quantity of data creates anticipation for exciting new effects and new physics discoveries in the coming two decades.

11.5 Appendix

In this appendix we present the list of used acronyms.

Table 23: List of used acronyms.

Acronym	Meaning	Acronym	Meaning
ACS	Advanced Camera for Surveys	KiDS	Kilo Degree Survey
ACT	Atacama Cosmology Telescope	LAT	Large Aperture Telescope
AEDGE	Atomic Experiments for Dark Matter and Gravity Exploration	LDE	Late Dark Energy

Continued on next

Table 23 – continued from previous

Acronym	Meaning	Acronym	Meaning
AGB	Asymptotic Giant Branch	LIGO	Laser Interferometer Gravitati Wave Observatory
AGN	Active Galactic Nucleus	LISA	Laser Interferometer Space Ante
AIC	Akaike Information Criterion	LMC	Large Magellanic Cloud
AO	Adaptive Optics	LOS	Line-Of-Sight
AvERA	Average Expansion Rate Approximation	LSS	Large Scale Structure
BAO	Baryon Accoustic Oscillations	LSST	Large Synoptic Survey Telescope
BBH	Binary Black Holes	LwMPT	Late $w - M$ Phantom Transitior
BBN	Big Bang Nucleosynthesis	MCP	Megamaser Cosmology Project
BBO	Big Bang Observer	MCT	Multi-Cycle Treasury
BH	Black Hole	MEDE	Modified Emergent Dark Energy
BIC	Bayesian Information Criterion	MGS	Main Galaxy Sample
BNS	Binary Neutron Stars	MM	Many Multiplet
BOSS	Baryon Oscillation Spectroscopic Survey	MST	Mass Sheet Transformation
BTFR	Baryonic Tully Fisher Relation	MW	Milky Way
CC	Cluster Counts	NANOGrav	North American Nanohertz Obs tory for Gravitational-waves
CCH	Cosmic CHronometric	NASA	National Aeronautics and Space ministration
CCHP	Carnegie–Chicago Hubble Program	NEDE	New Early Dark Energy
CDI	Cold Dark matter Isocurvature	NIR	Near InfraRed
CDM	Cold Dark Matter	NRAO	National Radio Astronomy Obs tory
CE	Cosmic Explorer	NS	Neutron Star
CFHTLenS	Canada-France-Hawaii Telescope Lensing	NVSS	NRAO VLA Sky Survey
CHP	Carnegie Hubble Program		
CL	Confidence Level	PEDE	Phenomenologically Emergent Energy
CMB	Cosmic Microwave Backgroun	PL	Period–Luminosity
COBE	Cosmic Background Explorer	PTAs	Pulsar Timing Arrays
COSMOGRAIL	COSmological MOnitoring of GRAvItational Lenses	QFT	Quantum Field Theory
CP	Cosmological Principle	QSO	Quasi-Stellar Object (quasar)
CPL	Chevallier - Polarski - Linder	ROSAT	ROentgen SATellite
CSA	Canadian Space Agency	RSD	Redshift Space Distortions
CSP	Carnegie Supernova Project	RVM	Running Vacuum Model
DE	Dark Energy	SATs	Small Aperture Telescopes
DEBs	Detached Eclipsing Binary stars		
DECIGO	DECi-hertz Interferometer Gravitational wave Observatory	SBF	Surface Brightness Fluctuations
DES	Dark Energy Survey	SDSS	Sloan Digital Sky Survey

Continued on next

Table 23 – continued from previous

Acronym	Meaning	Acronym	Meaning
DESI	Dark Energy Spectroscopic Instrument	SH0ES	Supernovae H_0 for the Equatio State
DIC	Deviance Information Criterion		
DM	Dark Matter	SKA	Square Kilometre Array
EBL	Extragalactic Background Light	SLACS	Sloan Lens ACS Survey
eBOSS	Extended Baryon Oscillation Spectroscopic Survey	SM	Standard Model
EDE	Early Dark Energy	SMBH	SuperMassive Black Hole
EDGES	Experiment to Detect the Global Epoch of Reionization Signature	SnIa	Supernova Type Ia
EDR	Early Data Release	SneII	Supernovae Type II
EDS	Early Dark Sector		
EFTofLSS	Effective Field Theory of Large-Scale Structure	SNR	Signal-to-Noise Ratio
EM	ElectroMagnetic	SO	Simons Observatory
EPTA	European Pulsar Timing Array	SOLME	Stanford Optically Levitated M sphere Experiment
eROSITA	extended ROentgen Survey with an Imaging Telescope Array	SPH	Smooth Particle Hydrodynamics
ESA	European Space Agency	SPT	South Pole Telescope
ET	Einstein Telescope	STRIDES	STRong-lensing Insights into Energy Survey
ETHOS	Effective THeory Of Structure formation	TBTF	Too Big To Fail
FJ	Faber–Jackson	TD	Time-Delay
FLRW	Friedmann-Lemaitre-Roberson-Walker	TDCOSMO	Time-Delay COSMOgraphy
FMOS	Fiber Multi-Object Spectrograph	TDE	Transitional Dark Energy
FP	Fundamental Plane	TFR	Tully-Fisher Relation
GAMA	Galaxy and Mass Assembly	TGSS	TIFR GMRT Sky Survey
GAUSS	Gravitation And the Universe from large Scale-Structures	TIFR	Tata Institute of Fundamental search
GDR	Gaia Data Release	TPCF	Two-Point Correlation Functions
GEDE	Generalised Emergent Dark Energy	TRGB	Tip of the Red Giant Branch
GEHR	Giant Extragalactic HII Region	tSZ	thermal Sunyaev-Zel'dovich
GGL	Galaxy-Galaxy Lensing	UV	Ultraviolet
GMRT	Giant Metrewave Radio Telescope	UVES	Ultraviolet and Visual Echelle S trograph
GP	Gaussian Process	VCDM	Vacuum Cold Dark Matter
GR	General Relativity	VHE	Very High Energy
GRB	Gamma-Ray Burst	VIKING	VISTA Kilo-Degree Infrared Gal
GW	Gravitational Waves	VIMOS	VIsible MultiObject Spectrograp
GWTC	Gravitational Wave Transient Catalog	VIPERS	VIMOS Public Extra-galactic shift Survey

Continued on next

Table 23 – continued from previous

Acronym	Meaning	Acronym	Meaning
HMF	Halo Mass Function	VISTA	Visible and Infrared Survey Teles for Astronomy
HSC	Subaru Hyper Suprime-Cam lensing survey	VLT	Very Large Telescope
HST	Hubble Space Telescope	VM	Vacuum Metamorphosis
H0LiCOW	H_0 Lenses in COSMOGRAIL's Wellspring	VSF	Violent Star Formation
ICM	IntraCluster Medium	VVDS	VIMOS-VLT Deep Survey
IDE	Interacting Dark Energy	WISE	Wide-field Infrared Survey Explo
IGM	InterGalactic Medium	WL	Weak Lensing
ISW	Integrated Sachs–Wolfe	WMAP	Wilkinson Microwave Anisot Probe
JWST	James Webb Space Telescope	WtG	Weighting the Giant
IndIGO	Indian Initiative in Gravitational wave Observations consortium	ZTF	Zwicky Transient Facility
IPTA	International Pulsar Timing Array	2dFGRS	2-degree Field Galaxy Redshift Su
JLA	Joint Light-curve Analysis	2dFlenS	2-degree Field Lensing Survey
KAGRA	Kamioka Gravitational	6dFGS	6-degree Field Galaxy Survey

Part IV
Properties of the Universe in Detail

12 Theory by Dodelson

12.1 The Standard Model and Beyond

Einstein's discovery of general relativity in the last century enabled us for the first time in history to come up with a compelling, testable theory of the universe. The realization that the universe is expanding and was once much hotter and denser allows us to modernize the deep age-old questions "Why are we here?" and "How did we get here?" The updated versions are now "How did the elements form?", "Why is the universe so smooth?", and "How did galaxies form from this smooth origin?" Remarkably, these questions and many like them have quantitative answers, answers that can be found only by combining our knowledge of fundamental physics with our understanding of the conditions in the early universe. Even more remarkable, these answers can be tested against astronomical observations.

This chapter describes the idea of an expanding universe, without using the equations of general relativity. The success of the Big Bang rests on three observational pillars: the Hubble diagram exhibiting expansion; light element abundances which are in accord with Big Bang nucleosynthesis; and the blackbody radiation left over from the first few hundred thousand years, the cosmic microwave background. After introducing these pieces of evidence, I move beyond the Standard Model embodied by the three pillars. Developments in the last two decades of the 20^{th} century - both theoretical and observational - point to

- the existence of dark matter and perhaps even dark energy
- the need to understand the evolution of perturbations around the zero order, smooth universe
- inflation, the generator of these perturbations

The emergent picture of the early universe is summarized in the time line of Figure 1.15.

12.1.1 The Expanding Universe

We have good evidence that the universe is expanding. This means that early in its history the distance between us and distant galaxies was smaller than it is today. It is convenient to describe this effect by introducing the scale factor a, whose present value is set to one. At earlier times a was smaller than it is today. We can picture space as a grid as in Figure 1.1 which expands uniformly as time evolves. Points on the grid maintain their coordinates, so the comoving distance between two points - which just measures the difference between coordinates remains constant. However, the physical distance is proportional to the scale factor, and the physical distance does evolve with time.

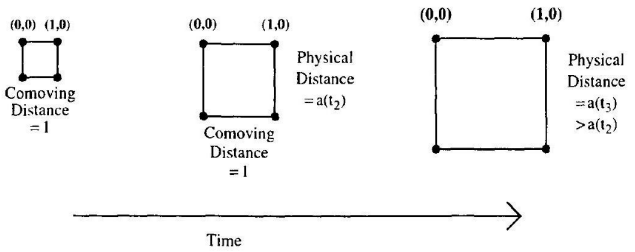


Figure 1.1. Expansion of the universe. The comoving distance between points on a hypothetical grid remains constant as the universe expands. The physical distance is proportional to the comoving distance times the scale factor, so it gets larger as time evolves.

In addition to the scale factor and its evolution, the smooth universe is characterized by one other parameter, its geometry. There are three possibilities: flat, open, or closed universes. These different possibilities are best understood by considering two freely traveling particles which start their journeys moving parallel to each other. A flat universe is Euclidean: the particles remain parallel as long as they travel freely. General relativity connects geometry to energy. Accordingly, a flat universe is one in which the energy density is equal to a critical value, which we will soon see is approximately 10^{-29} g cm^{-3}. If the density is higher than this value, then the universe is closed: gradually the initially parallel particles converge, just as all lines of constant longitude meet at the North and South Poles. The analogy of a closed universe to the surface of a sphere runs even deeper: both are said to have positive curvature, the former in three spatial dimensions and the latter in two. Finally, a low-density universe is open, so that the initially parallel paths diverge, as would two marbles rolling off a saddle.

To understand the history of the universe, we must determine the evolution of the scale factor a with cosmic time t. Again, general relativity provides the connection between this evolution and the energy in the universe. Figure 1.2 shows how the scale factor increases as the universe ages. Note that the dependence of a on t varies as the universe evolves. At early times, $a \propto t^{1/2}$ while at later times the dependence switches to $a \propto t^{2/3}$. How the scale factor varies with time is determined by the energy density in the universe. At early times, one form of energy, radiation, dominates, while at later times, nonrelativistic matter accounts for most of the energy density. In fact, one way to explore the energy content of the universe is to measure changes in the scale factor. We will see that, partly as a result of such exploration, we now believe that, very recently, a has stopped growing as $t^{2/3}$, a signal that a new form of energy has come to dominate the cosmological landscape.

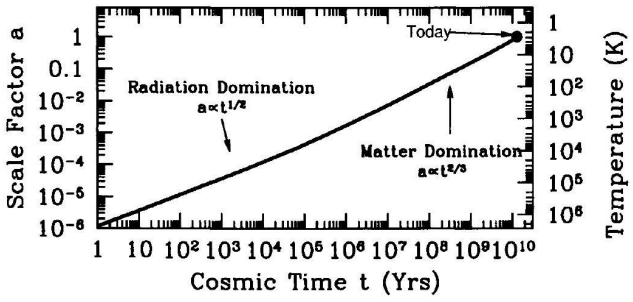


Figure 1.2. Evolution of the scale factor of the universe with cosmic time. When the universe was very young, radiation was the dominant component, and the scale factor increased as $t^{1/2}$. At later times, when matter came to dominate, this dependence switched to $t^{2/3}$. The right axis shows the corresponding temperature, today equal to 3 K.

To quantify the change in the scale factor and its relation to the energy, it is first useful to define the Hubble rate

$$H(t) := \frac{da/dt}{a}$$

which measures how rapidly the scale factor changes. For example, if the universe is flat and matter-dominated, so that $a \propto t^{2/3}$, then $H = (2/3)t^{-1}$. Thus a powerful test of this cosmology is to measure separately the Hubble rate today, H_0, and the age of the universe today. Here and throughout, subscript 0 denotes the value of a quantity today. In a flat, matter-dominated universe, the product $H_0 t_0$ should equal 2/3.

More generally, the evolution of the scale factor is determined by the Friedmann equation

$$H^2(t) = \frac{8\pi G}{3}\left[\rho(t) + \frac{\rho_{\rm cr} - \rho_0}{a^2(t)}\right]$$

where $\rho(t)$ is the energy density in the universe as a function of time with ρ_0 the present value. The critical density

$$\rho_{\rm cr} := \frac{3H_0^2}{8\pi G}$$

where G is Newton's constant. To use Einstein's equation, we must know how the energy density evolves with time. This turns out to be a complicated question because ρ in Eq. (1.2) is the sum of several different components, each of which scale differently with time. Consider first nonrelativistic matter. The energy of one such particle is equal to its rest mass energy, which remains constant with time. The energy density of many of these is therefore equal to the rest mass energy times the number density. When the scale factor was smaller, the densities were necessarily larger. Since number density is inversely proportional to volume, it should be proportional to a^{-3}. Therefore the energy density of matter scales as a^{-3}.

The photons which make up the cosmic microwave background (CMB) today have a well-measured temperature $T_0 = 2.725 \pm 0.002K$ (Mather et al., 1999). A photon with an energy $k_B T_0$ today has a wavelength $\hbar c/k_B T_0$. Early on, when the scale factor was smaller than it is today, this wavelength would have been correspondingly smaller. Since the energy of a photon is inversely proportional to its wavelength, the photon energy would have been larger than today by a factor of $1/a$. This argument applied to the thermal bath of photons implies that the temperature of the plasma as a function of time is

$$T(t) = T_0/a(t).$$

At early times, then, the temperature was higher than it is today, as indicated in Figure 1.2. The energy density of radiation, the product of number density times average energy per particle, therefore scales as a^{-4}.

Evidence from distant supernovae (Chapter 2; Riess et al., 1998; Perlmutter et al., 1999) suggests that there may well be energy, dark energy, besides ordinary matter and radiation. One possibility is that this new form of energy remains constant with time, i.e., acts as a cosmological constant, a possibility first introduced (and later abandoned) by Einstein. Cosmologists have explored other forms though, many of which behave very differently from the cosmological constant. We will see more of this in later chapters.

Equation (1.2) allows for the possibility that the universe is not flat: if it were flat, the sum of all the energy densities today would equal the critical density, and the last term in Eq. (1.2) would vanish. If the universe is not flat, the curvature energy scales as $1/a^2$. In most of this book we will work within the context of a flat universe. In such a universe, the evolution of perturbations is much easier to calculate than in open or closed universes. Further, there are several persuasive arguments, both theoretical and more recently observational, which strongly support the flatness of the universe. More on this in Chapters 2 and 8.

Figure 1.3 illustrates how the different terms in Eq. (1.2) vary with the scale factor. While today matter, and possibly a cosmological constant dominate the landscape, early on, because of the a^{-4} scaling, radiation was the dominant constituent of the universe.

Let's introduce some numbers. The expansion rate is a measure of how fast the universe is expanding, determined (Section 1.2) by measuring the velocities of distant galaxies and dividing by their distance from us. So the expansion is often

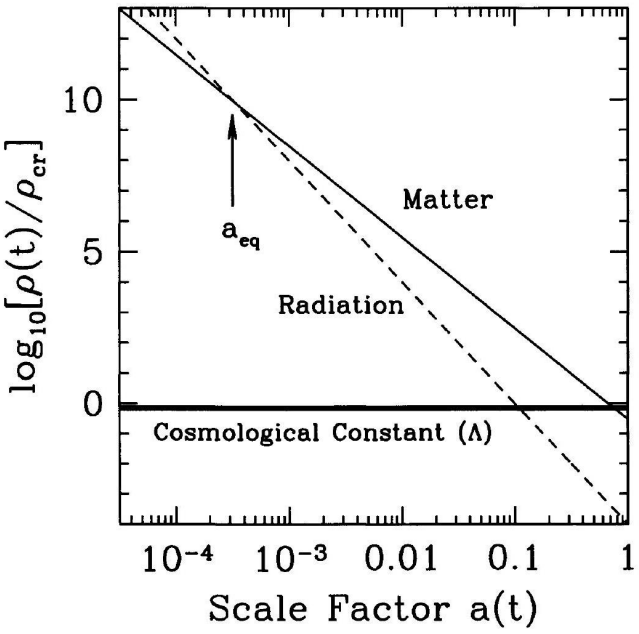


Figure 1.3. Energy density vs scale factor for different constituents of a flat universe. Shown are nonrelativistic matter, radiation, and a cosmological constant. All are in units of the critical density today. Even though matter and cosmological constant dominate today, at early times, the radiation density was largest. The epoch at which matter and radiation are equal is $a_{\rm eq}$.

written in units of velocity per distance. Present measures of the Hubble rate are parameterized by h defined via

$$\begin{aligned} H_0 &= 100h \text{ kmsec}^{-1}\text{Mpc}^{-1} \\ &= \frac{h}{0.98 \times 10^{10} \text{ years}} = 2.133 \times 10^{-33} h\text{eV}/\hbar \end{aligned}$$

where h has nothing to do with Planck's constant $\hbar$. The astronomical length scale of a megaparsec (Mpc) is equal to 3.0856×10^{24} cm. Current measurements set $h = 0.72 \pm 0.08$ (Freedman et al., 2001).

The predicted age for a flat, matter-dominated universe, $(2/3)H_0^{-1}$, is then of order 8 to 10 Gyr. The current best estimate for the age of the universe is 12.6Gyr, with a 95% confidence level lower limit of 10.4 Gyr (Krauss and Chaboyer, 2001), so this test suggests that a flat, matter-dominated universe is barely viable. You will show in Exercise 2 that the age of the universe with a cosmological constant is larger (for fixed h); in fact one of the original arguments in favor of this excess energy was to make the universe older.

Newton's constant in Eq. (1.3) is equal to 6.67×10^{-8} cm^3 g^{-1}sec^{-2}. This, together with Eq. (1.5), enables us to get a numerical value for the critical density:

$$\rho_{\rm cr} = 1.88h^2 \times 10^{-29} \text{ g cm}^{-3}.$$

An important ramification of the higher densities in the past is that the rates for particles to interact with each other, which scale as the density, were also much higher early on. Figure 1.4 shows some important rates as a function of the scale

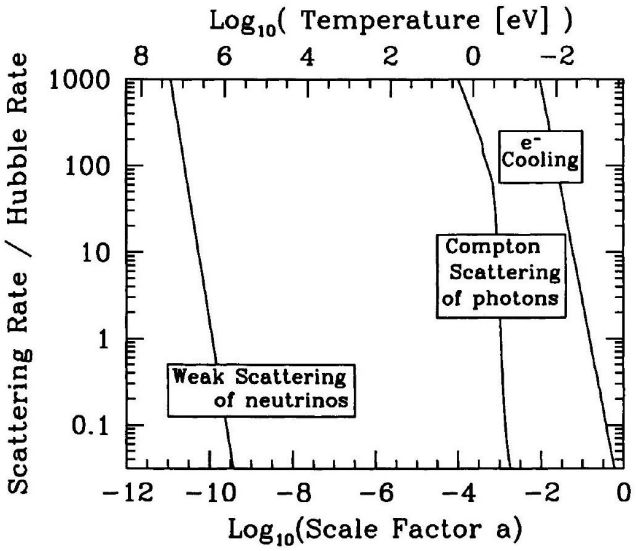


Figure 1.4. Rates as a function of the scale factor. When a given rate becomes smaller than the expansion rate H, that reaction falls out of equilibrium. Top scale gives (k_B times) the temperature of the universe, an indication of the typical kinetic energy per particle.

factor. For example, when the temperature of the universe was greater than several MeV/k_B, the rate for electrons and neutrinos to scatter was larger than the expansion rate. Thus, before the universe could double in size, a neutrino scattered many times off background electrons. All these scatterings brought the neutrinos into equilibrium with the rest of the cosmic plasma. This is but one example of a very general, profound fact: if a particle scatters with a rate greater than the expansion rate, that particle stays in equilibrium. Since rates were typically quite large, the early universe was a relatively simple environment: not only was it very smooth, but many of its constituents were in equilibrium. Chapter 2 explores some manifestations of the equilibrium conditions, while Chapter 3 touches on several cases where equilibrium could not be maintained because the reaction rates dropped beneath the expansion rate.

12.1.2 The Hubble Diagram

If the universe is expanding as depicted in Figure 1.1, then galaxies should be moving away from each other. We should therefore see galaxies receding from us. Recall that the wavelength of light or sound emitted from a receding object is stretched out so that the observed wavelength is larger than the emitted one. It is convenient to define this stretching factor as the redshift z:

$$1 + z := \frac{\lambda_{\text{obs}}}{\lambda_{\text{emit}}} = \frac{1}{a}.$$

For low redshifts, the standard Doppler formula applies and $z \simeq \frac{v}{c}$. So a measurement of the amount by which absorption and/or emission lines are redshifted is a direct measure of how fast the structures in which they reside are receding from us.

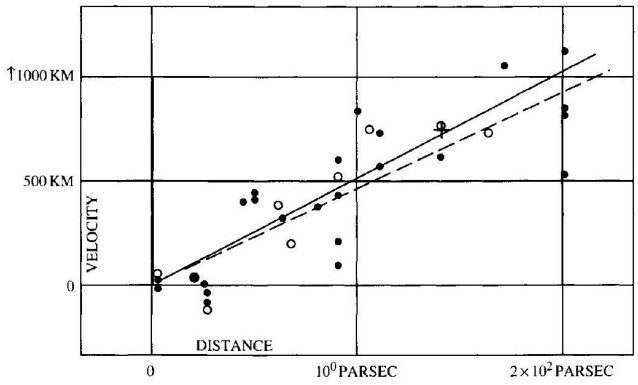


Figure 1.5. The original Hubble diagram (Hubble, 1929). Velocities of distant galaxies (units should be kmsec^{-1}) are plotted vs distance (units should be Mpc). Solid (dashed) line is the best fit to the filled (open) points which are corrected (uncorrected) for the sun's motion.

Hubble (1929) first found that distant galaxies are in fact receding from us. He also noticed the trend that the velocity increases with distance. This is exactly what we expect in an expanding universe, for the physical distance between two galaxies is $d = ax$ where x is the comoving distance. In the absence of any comoving motion ($\dot{x} = 0$, no peculiar velocity) the relative velocity $v = \dot{d}$ is therefore equal to $\dot{a}x = Hd$. Therefore, velocity should increase linearly with distance (at least at low redshift) with a slope given by H, the Hubble constant. Hubble's Hubble constant can be easily extracted from Figure 1.5. It is simply $H = 1000/2$ kmsec^{-1}Mpc^{-1}, almost a factor of 10 higher than current estimates. Also notice that Hubble's data went out to redshift $z = 1000$ kmsec$^{-1}/c \simeq 0.003$.

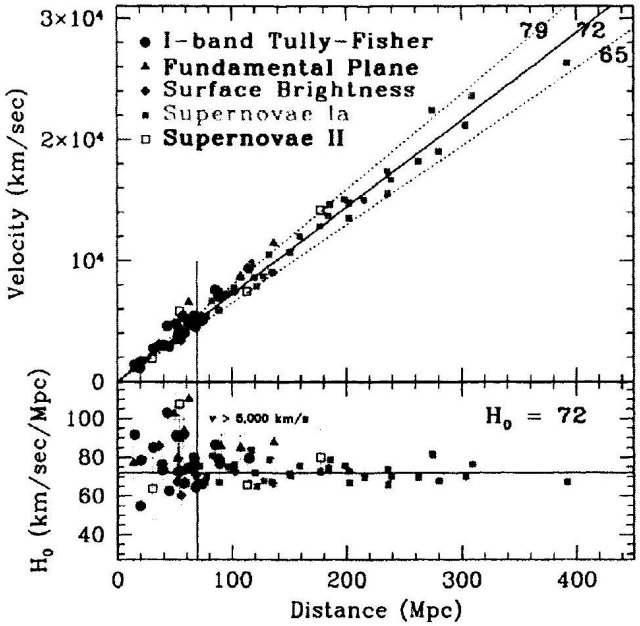


Figure 1.6. Hubble diagram from the Hubble Space Telescope Key Project (Freedman et al., 2001) using five different measures of distance. Bottom panel shows H_0 vs distance with the horizontal line equal to the best fit value of 72 kmsec^{-1}Mpc^{-1}.

The Hubble diagram is still the most direct evidence we have that the universe is expanding. Current incarnations use the same principle as the original: find the distance and the redshift of distant objects. Measuring redshifts is straightforward; the hard part is determining distances for objects of unknown intrinsic brightness. One of the most popular techniques is to try to find a standard candle, a class of objects which have the same intrinsic brightness. Any difference between the apparent brightness of two such objects then is a result of their different distances

from us. This method is typically generalized to find a correlation between an observable and intrinsic brightness. For example, Cepheid variables are stars for which intrinsic brightness is tightly related to period. The Hubble Space Telescope measured the periods of thousands of Cepheid variables in galaxies as far away as 20Mpc. With distances to these galaxies fixed, five different distance measures were used to go much further, as far away as 400Mpc. Figure 1.6 shows that all of these five indicators agree with one another and have converged on $H_0 = 72$ kmsec Mpc^{-1} with 10% errors.

As shown in Figure 1.6 the standard candle that can be seen at largest distances is a Type Ia supernova. Since they are so bright, supernovae can be used to extend the Hubble diagram out to very large redshifts (the current record is of order $z \simeq 1.7$), a regime where the simple Doppler law ceases to work. Figure 1.7 shows a recent Hubble diagram using very these very distant objects. In the next chapter, we will derive the correct expression for the distance (in this case the luminosity distance) as a function of redshift. For now, I simply point out that this expression depends on the energy content of the universe. The three curves in Figure 1.7 depict three different possibilities: flat matter dominated; open; and flat with a cosmological constant (Λ). The high-redshift data are now good enough to distinguish among these possibilities, strongly disfavoring the previously favored flat, matter-dominated universe. The current best fit is a universe with about 70% of the energy in the form of a cosmological constant, or some other form of dark energy. More on this in Chapter 2.

12.1.3 Big Bang Nucleosynthesis

When the universe was much hotter and denser, when the temperature of order an MeV/k_B, there were no neutral atoms or even bound nuclei. The vast amounts of radiation in such a hot environment ensured that any atom or nucleus produced would be immediately destroyed by a high energy photon. As the universe cooled well below the binding energies of typical nuclei, light elements began to form. Knowing the conditions of the early universe and the relevant nuclear cross-sections, we can calculate the expected primordial abundances of all the elements (Chapter 3).

Figure 1.8 shows the predictions of Big Bang Nucleosynthesis (BBN) for the light element abundances [1]. The boxes and arrows in Figure 1.8 show the current estimates for the light element abundances. These are consistent with the predictions, and this consistency test provides yet another ringing confirmation of the Big Bang. The measurements do even more though. The theoretical predictions, which we will explore in detail in Chapter 3, depend on the density of protons and neutrons at the time of nucleosynthesis. The combined proton plus neutron density

[1] Recall nuclear notation: The 4 in ^{4}He refers to the total number of nucleons (protons and neutrons). So ^{4}He has two neutrons and two protons, while ^{3}He has two protons and one neutron. See the box on page 63 for more details.

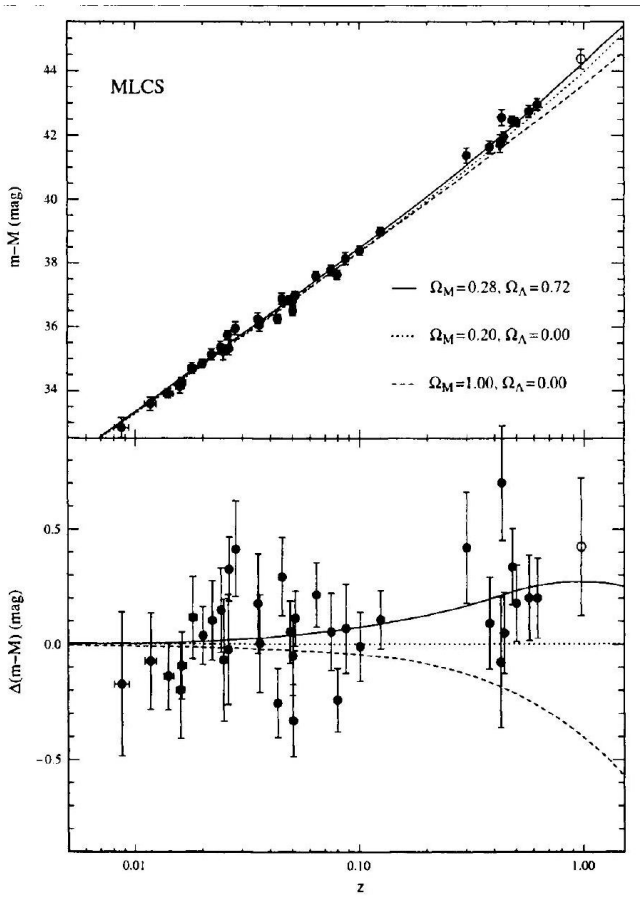


Figure 1.7. Hubble diagram from distant Type la supernovae. Top panel shows apparent magnitude (an indicator of the distance) vs redshift. Lines show the predictions for different energy contents in the universe, with Ω_M the ratio of energy density today in matter compared to the critical density and Ω_Λ the ratio of energy density in a cosmological constant to the critical density. Bottom panel plots the residuals, making it clear that the high-redshift supernovae favor a Λ-dominated universe over a matter-dominated one.

is called the baryon density since both protons and neutrons have baryon number one and these are the only baryons around at the time. Thus, BBN gives us a way of measuring the baryon density in the universe. Since we know how those densities scale as the universe evolves (they fall as a^{-3}), we can turn the measurements of light element abundances into measures of the baryon density today.

In particular, the measurement of primordial deuterium pins down the baryon density extremely accurately to only a few percent of the critical density. Ordinary matter (baryons) contributes at most 5% of the critical density. Since the total matter density today is almost certainly larger than this-direct estimates give values of order $20 - 30\%$ - nucleosynthesis provides a compelling argument for nonbaryonic dark matter.

The deuterium measurements (Burles and Tytler, 1998) are the new developments in the field. These measurements are so exciting because they explore the deuterium abundance at redshifts of order $3 - 4$, well before much processing could have altered the primordial abundances. Figure 1.9 shows one such detection. The basic idea is that light from distant QSOs is absorbed by intervening neutral hydrogen systems. The key absorption feature arises from transition from the $(n = 1)$

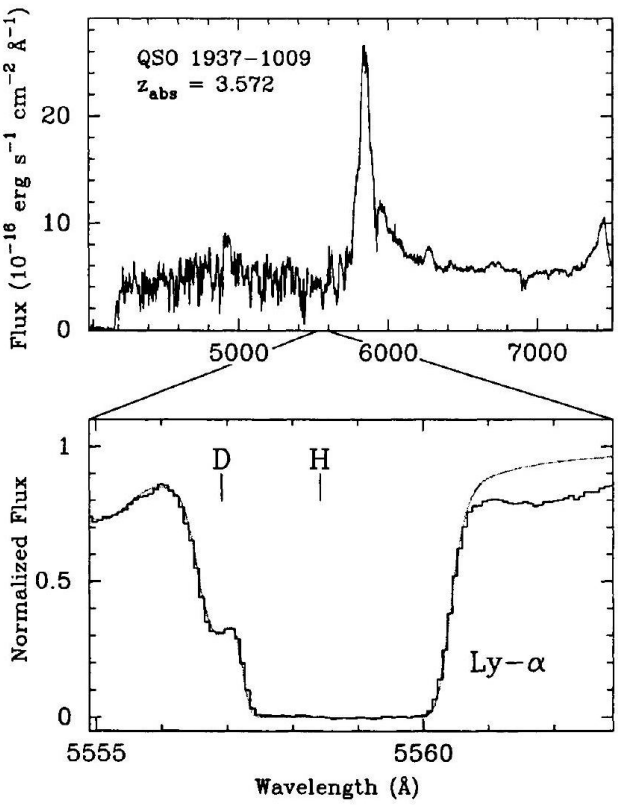


Figure 1.9. Spectrum from a distant QSO (Burles, Nollett, and Turner, 1999). Absorption of photons with rest wavelength 1216Å corresponding to the $n = 1$ to $n = 2$ state of hydrogen is redshifted up to $1216(1 + 3.572)$Å. Bottom panel provides details of the spectrum in this range, with the the presence of deuterium clearly evident.

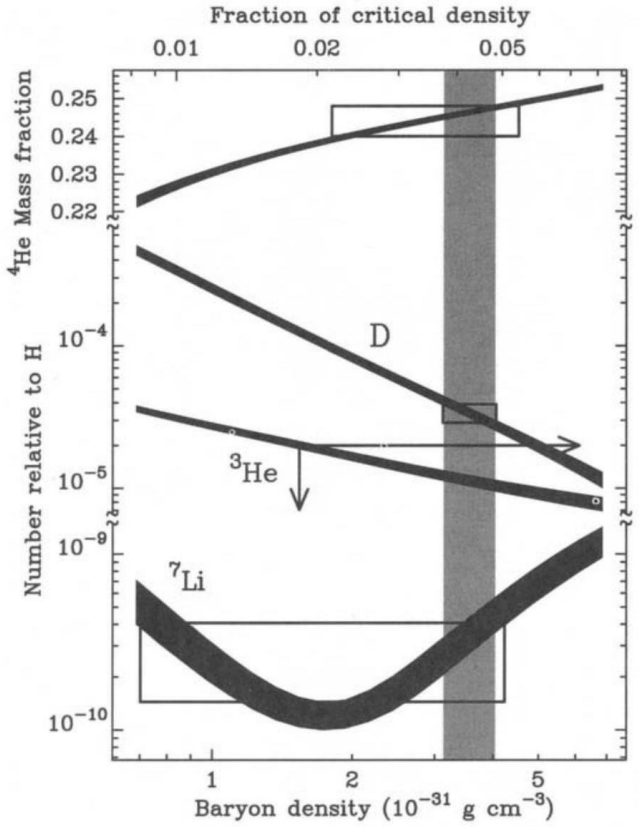


Figure 1.8. Constraint on the baryon density from Big Bang Nucleosynthesis (Burles, Nollett, and Turner, 1999). Predictions are shown for four light elements- ^{4}He, deuterium, ^{3}He,

and lithium - spanning a range of 10 orders of magnitude. The solid vertical band is fixed by measurements of primordial deuterium. The boxes are the observations; there is only an upper limit on the primordial abundance of ^{3}He.

ground state of hydrogen to the first excited state ($n = 2$), requiring a photon with wavelength $\lambda = 1215.7$Å. Since photons are absorbed when exciting hydrogen in this fashion, there is a trough in the spectrum at $\lambda = 1215.7$Å, redshifted by a factor of $1 + z$. The corresponding line from deuterium should be (i) shifted over by $0.33(1 + z)$Å (see Exercise 3) and (ii) much less damped since there is much less deuterium. Figure 1.9 shows just such a system; there are now half a dozen with detections precisely in the neighborhood shown in Figure 1.8. Note that the steep decline in deuterium as a function of baryon density helps here: even relatively large errors in D measurements translate into small errors on the baryon density. COSMIC MICROWAVE BACKGROUND SPECTRUM FROM COBE

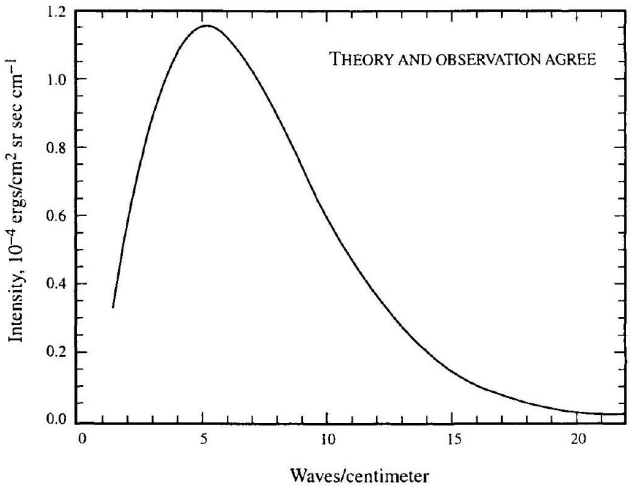


Figure 1.10. Intensity of cosmic microwave radiation as a function of wavenumber from Far InfraRed Absolute Spectrophotometer (FIRAS) (Mather et al., 1994), an instrument on the COBE satellite. Hidden in the theoretical blackbody curve are dozens of measured points, all of which have uncertainties smaller than the thickness of the curve!

12.1.4 The Cosmic Microwave Background

The CMB offers us a look at the universe when it was only 300,000 years old. The photons in the cosmic microwave background last scattered off electrons at redshift 1100 ; since then they have traveled freely through space. When we observe them today, they literally come from the earliest moments of time. They are therefore the most powerful probes of the early universe. We will spend an inordinate amount of time in this book working through the details of what happened before the epoch of last scattering and also developing the mathematics of the freestreaming process since then. A crucial fact about this history, though, is that the collisions with electrons before last scattering ensured that the photons were in equilibrium. That is, they should have a blackbody spectrum.

The specific intensity of a gas of photons with a blackbody spectrum is

$$I_\nu = \frac{4\pi\hbar\nu^3/c^2}{\exp\{2\pi\hbar\nu/k_BT\} - 1}$$

Figure 1.10 shows the remarkable agreement between this prediction (see Exercise 4) of Big Bang cosmology and the observations by the FIRAS instrument aboard the COBE spacecraft. We have been told [2] that detection of the 3 K background by Penzias and Wilson in the mid-1960s was sufficient evidence to decide the controversy in favor of the Big Bang over the Steady

State universe. Penzias and Wilson, though, measured the radiation at just one wavelength. If even their one-wavelength result was enough to tip the scales, the current data depicted in Figure 1.10 should send skeptics from the pages of physics journals to the far reaches of radical Internet chat groups.

The most important fact we learned from our first 25 years of surveying the CMIB was that the early universe was very smooth. No anisotropies were detected in the CMB. This period, while undoubtedly frustrating for observers searching for anisotropies, solidified the view of a smooth Big Bang. We are now moving on. We have discovered anisotropies in the CMB, indicating that the early universe was not completely smooth. There were small perturbations in the cosmic plasma. To understand these, we must go beyond the Standard Model.

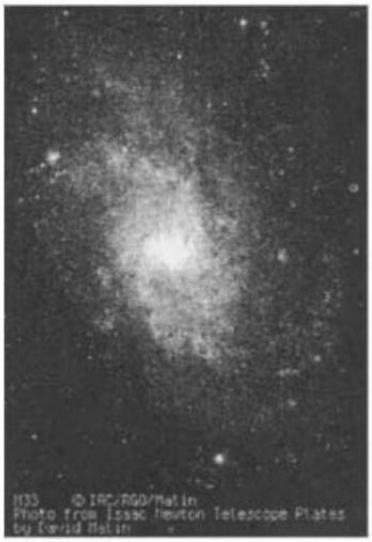
12.1.5 Beyond the Standard Model

While the three pillars put the Big Bang model on firm footing, other observations cry out for more details. I hinted above at one of these, the notion that there must be nonbaryonic matter in the universe. Dark matter is a familiar concept to astronomers; the first suggestion was put forth by Zwicky in 1933(!). Figure 1.11 illustrates the way dark matter can be found in galaxies, with the use of rotation curves probing the gravitational field. Indeed, a mismatch between the matter inferred from gravity and that we can see exists on almost all observable scales.

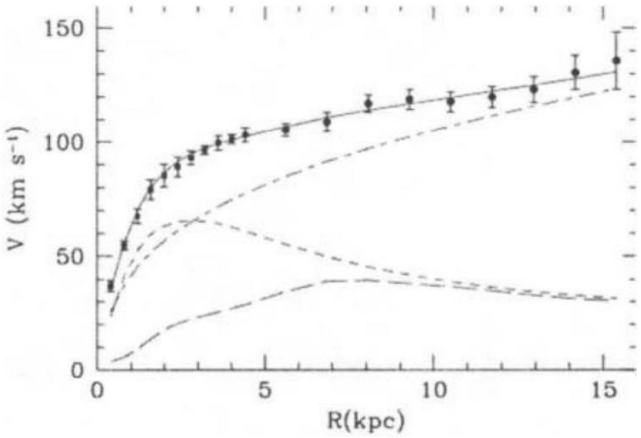
Because of the limits inferred from Big Bang nucleosynthesis, the dark matter, or at least an appreciable fraction of it, must be nonbaryonic. What is this new form of matter? And how did it form in the early universe? The most popular idea currently is that the dark matter consists of elementary particles produced in the earliest moments of the Big Bang. In Chapter 3, we will explore this possibility in detail, arguing that dark matter was likely produced when the temperature of the universe was of order hundreds of GeV/k_B. As we will see, the hypothesis that dark matter consists of fundamental relics from the early universe may soon be tested experimentally.

The last decades of the 20$^{\text{th}}$ century saw a number of large surveys of galaxies designed to measure structure in the universe. These culminated in two large surveys, the Sloan Digital Sky Survey and the Two Degree Field Galaxy (Figure 1.12) Redshift Survey, which between them will compile the redshifts of, and hence the distances to, a million galaxies. Galaxies in Figure 1.12 are clearly not distributed randomly: the universe has structure on large scales. To understand this structure, we must go beyond the Standard Model not only by including dark matter, but also by allowing for deviations from smoothness. We must develop the tools to study

[2] For a fascinating first-hand account of the history of the discovery of the CMB, see Chapter 1 of Partridge (1995).



(a)



(b)

Figure 1.11. (a) Image of spiral galaxy M33. The inner brightest region has a radius of several kpc. (b) Rotation curve for M33 (Corbelli and Salucci, 2000). Points with error bars come from the $21-$cm line of neutral hydrogen. Solid line is a model fitting the data. Different contributions to the total rotation curve are: dark matter halo (dot-dashed line), stellar disk (short dashed line), and gas (long dashed line). At large radii, dark matter dominates.

perturbations around the smooth background of the Standard Model. We will see in Chapters 4 and 5 that this is straightforward in theory, as long as the perturbations remain small.

The best ways to learn about the evolution of structure and to compare theory with observations are to look at anisotropies in the CMB and at how matter is distributed on large scales. To compare theory with observations, we must at first try to avoid scales dominated by nonlinearities. As an extreme example, we can never hope to understand cosmology by carefully examining rock formations on Earth. The intermediate steps - collapse of matter into a galaxy; molecular cooling; star formation; planetary formation; etc. - are much too complicated to allow comparison between linear theory and observations. While perturbations to the matter on small scales (less than about 10Mpc) have grown nonlinear, large-scale perturbations are still small. So they have been processed much less than the corresponding small-scale structure. Similarly, anisotropies in the CMB have remained small because the photons that make up the CMB do not clump.

Identifying large-scale structure and the CMB as the two most promising areas of study solves just one issue. Another very important challenge is to understand how to characterize

these distributions so that theory can be compared to experiment. It is one thing to look at a map and quite another to quantitative tests of cosmological models. To make such tests, it is often useful to take the Fourier transform of the distribution in question; as we will see, working in Fourier space makes it easier to separate large from small scales. The most important statistic in the cases of

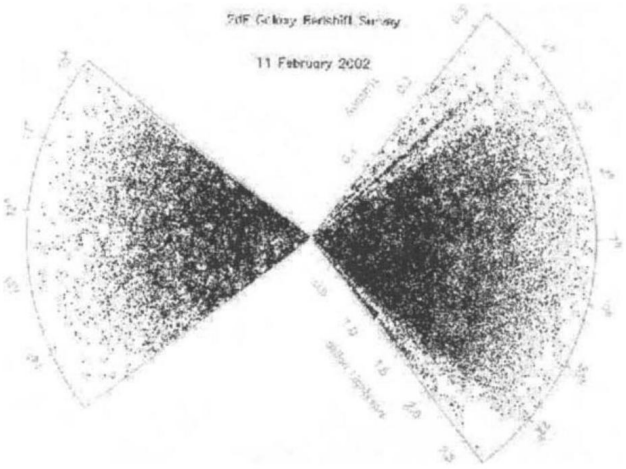


Figure 1.12. Distribution of galaxies in the Two Degree Field Galaxy Redshift Survey (2dF) (Colless et al., 2001). By the end of the survey, redshifts for 250,000 galaxies will have been obtained. As shown here, they probe structure in the universe out to $z = 0.3$, corresponding to distances up to 1000 h^{-1}Mpc away from us (we are located at the center). See color Plate 1.12.

both the CMB and large-scale structure is the two-point function, called the power spectrum in Fourier space. If the mean density of the galaxies is $\bar{n}$, then we can characterize the inhomogeneities with $\delta(\vec{x}) = (n(\vec{x}) - \bar{n})/\bar{n}$, or its Fourier transform $\tilde{\delta}(\vec{k})$. The power spectrum $P(k)$ is defined via

$$\left\langle \tilde{\delta}(\vec{k})\tilde{\delta}\left(\vec{k}'\right)\right\rangle = (2\pi)^3 P(k)\delta^3\left(\vec{k} - \vec{k}'\right)$$

Here the angular brackets denote an average over the whole distribution, and $\delta^3()$ is the Dirac delta function which constrains $\vec{k} = \vec{k}'$. The details aside, Eq. (1.9) indicates that the power spectrum is the spread, or the variance, in the distribution. If there are lots of very under- and overdense regions, the power spectrum will be large, whereas it is small if the distribution is smooth. Figure 1.13 shows the power spectrum of the galaxy distribution. Since the power spectrum has dimensions of k^{-3} or (length) ^{3}, Figure 1.13 shows the combination $k^3P(k)/2\pi^2$, a dimensionless number which is a good indication of the clumpiness on scale k.

The best measure of anisotropies in the CMB is also the two-point function of the temperature distribution. There is a subtle technical difference between the two power spectra which are used to measure the galaxy distribution and the CMB, though. The difference arises because the CMB temperature is a two-dimensional field, measured everywhere on the sky (i.e., with two angular coordinates). Instead of Fourier transforming the CMB temperature, then, one typically expands it in

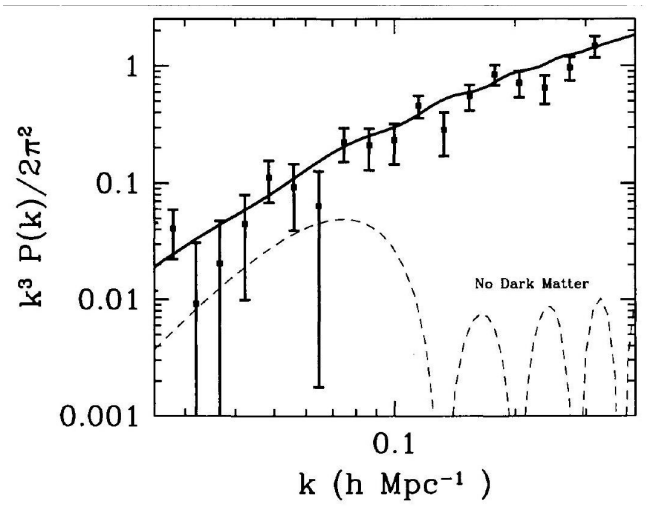


Figure 1.13. The variance $\Delta^2 := k^3P(k)/2\pi^2$ of the Fourier transform of the galaxy distribution as a function of scale. On large scales, the variance is smaller than unity, so the distribution is smooth. The solid line is the theoretical prediction from a model in which the universe contains dark matter, a cosmological constant, with perturbations generated by inflation. The dashed line is a theory with only baryons and no dark matter. Data come from the PSCz survey (Saunders et al., 2000) as analyzed by Hamilton and Tegmark (2001).

spherical harmonics, a basis more appropriate for a 2D field on the surface of a sphere. Therefore the two-point function of the CMB is a function of multipole moment l, not wave number k. Figure 1.14 shows the measurements of dozens of groups since 1992, when COBE first discovered large-angle (low l in the plot) anisotropies.

Figures 1.13 and 1.14 both have theoretical curves in them which appear to agree well with the data. The main goal of much of this book is to develop a firstprinciples understanding of these theoretical predictions. Indeed, understanding the development of structure in the universe has become a major goal of most cosmologists today. Note that this second aspect of cosmology beyond the Standard Model reinforces the first: i.e., observations of structure in the universe lead to the conclusion that there must be dark matter. In particular, the dashed curve in Figure 1.13 is the prediction of a model with baryons only, with no dark matter. The inhomogeneities expected in this model (when normalized to the CMB observations) are far too small. In Chapter 7, we will come to understand the reason why a baryon-only universe would be so smooth. For now, though, the message is clear: Dark matter is needed not only to explain rotation curves of galaxies but to explain

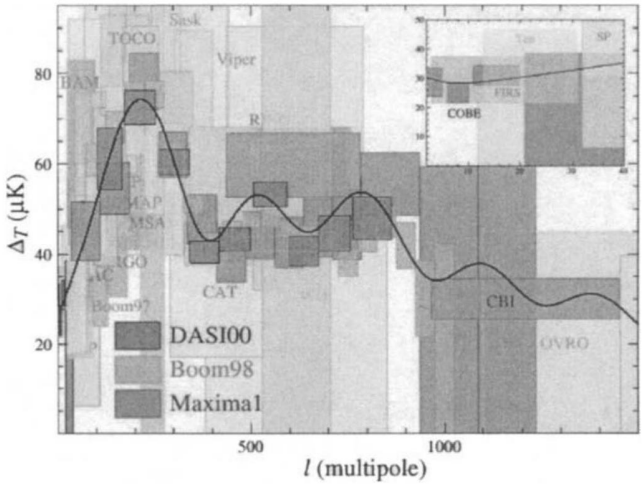


Figure 1.14. Anisotropies in the CMB predicted by the theory of inflation compared with observations. x-axis is multipole moment (e.g., $l = 1$ is the dipole, $l = 2$ the quadrupole) so

that large angular scales correspond to low l; y-axis is the root mean square anisotropy (the square root of the two-point function) as a function of scale. The characteristic signature of inflation is the series of peaks and troughs, a signature which has been verified by experiment. See color Plate 1.14.

structure in the universe at large!

While trying to understand the evolution of structure in the universe, we will be forced to confront the question of what generated the initial conditions, the primordial perturbations that were the seeds for this structure. This will lead us to a third important aspect of cosmology beyond the Standard Model: the theory of inflation. Chapter 6 introduces this fascinating proposal, that the universe expanded exponentially fast when it was but 10^{-35} sec old. Until recently, there was little evidence for inflation. It survived as a viable theory mainly because of its aesthetic appeal. The discoveries of the past several years have changed this. They have by and large confirmed some of the basic predictions of inflation. Most notably, this theory makes concrete predictions for the initial conditions, predictions that have observable consequences today. For me, the most profound and exciting discovery in cosmology has been the observation of anisotropies in the CMB, with a characteristic pattern predicted by inflation.

The theory encompassing all these Beyond the Standard Model ingredients-dark matter plus evolution of structure plus inflation - is called Cold Dark Matter, or CDM. The "Cold" part of this moniker comes from requiring the dark matter particles to be able to clump efficiently in the early universe. If they are hot instead, i.e., have large pressure, structure will not form at the appropriate levels.

SUGGESTED READING

There are many good textbooks covering the homogeneous Big Bang. I am most familiar with The Early Universe (Kolb and Turner), which has especially good discussions on nucleosynthesis and inflation. Peacock's Cosmological Physics is the most up-to-date and perhaps the broadest of the standard cosmology texts, with more of an emphasis on extragalactic astronomy than either The Early Universe or this book. A popular account which still captures the essentials of the homogeneous Big Bang (testifying to the success of the model: it hasn't changed that much in 25 years) is The First Three Minutes (Weinberg). More recently, three books of note are: The Whole Shebang (Ferris), The Little Book of the Big Bang (Hogan), and *A* Short History of the Universe (Silk).

A nice article summarizing the evidence for an expanding universe and some methods to quantify it is Freedman (1998). Two of the pioneers in the field of Big Bang nucleosynthesis, Schramm and Turner, wrote a very clear review article (1998) right before a tragic accident took the life of the first author. An excellent account of the evidence for dark matter in spiral galaxies is Vera Rubin's 1983 article in Scientific American.

I have not attempted to record the history of the discovery of the Big Bang. Three books I am familiar with which treat this history in detail are Blind Watchers of the Sky (Kolb), 3K: The Cosmic Microwave Background (Partridge), and Three Degrees Above Zero: Bell Labs in the Information Age (Bernstein). An article which sheds light on this history is Alpher and Herman (1988).

D-1.

Suppose (incorrectly) that H scales as temperature squared all the way back until the time when the temperature of the universe was $10^{19}\mathrm{GeV}/k_B$ (i.e., suppose the universe was radiation dominated all the way back to the Planck time). Also suppose that today the dark energy is in the form of a cosmological constant Λ, such that ρ_Λ today is equal to $0.7\rho_{\mathrm{cr}}$ and ρ_Λ remains constant throughout the history of the universe. What was $\rho_\Lambda/\left(3H^2/8\pi G\right)$ back then?

D-2.

Assume the universe today is flat with both matter and a cosmological constant, the latter with energy density which remains constant with time. Integrate Eq. (1.2) to find the present age of the universe. That is, rewrite Eq. (1.2) as

$$dt = H_0^{-1} \frac{da}{a} \left[\Omega_\Lambda + \frac{1 - \Omega_\Lambda}{a^3} \right]^{-1/2}$$

where Ω_Λ is the ratio of energy density in the cosmological constant to the critical density. Integrate from $a = 0$ (when $t = 0$) until today at $a = 1$ to get the age of the universe today. In both cases below the integral can be done analytically.

(a) First do the integral in the case when $\Omega_\Lambda = 0$.

(b) Now do the integral in the case when $\Omega_\Lambda = 0.7$. For fixed H_0, which universe is older?

D-3.

Using the fact that the reduced mass of the electron-nucleus in the D atom is larger than in hydrogen, and the fact that the Lyman $\alpha (n = 1 \to n = 2)$ transition in H has a wavelength 1215.67Å, find the wavelength of the photon emitted in the corresponding transition in D. Astronomers often define

$$v := c \frac{\Delta\lambda}{\lambda}$$

to characterize the splitting of two nearby lines. What is v for the H-D pair?

D-4.

Convert the specific intensity in Eq. (1.8) into an expression for what is plotted in Figure 1.10, the energy per square centimeter per steradian per second. Note that the x-axis is $1/\lambda$, the inverse wavelength of the photons. Show that the peak of a 2.73 K blackbody spectrum does lie at $1/\lambda = 5$ cm^{-1}.

12.2 The Smooth, Expanding Universe

Just as the early navigators of the great oceans required sophisticated tools to help them find their way, we will need modern technology to help us work through the ramifications of an expanding universe. In this chapter I introduce two of the necessary tools, general relativity and statistical mechanics. We will use them to derive some of the basic results laid down in Chapter 1: the expansion law of Eq. (1.2), the dependence of different components of energy density on the scale factor which governs expansion, the epoch of equality $a_{\rm eq}$ shown in Figure 1.3, and the luminosity distance needed to understand the implications of the supernovae diagram in Figure 1.7. Indeed, with general relativity and statistical mechanics, we can go a long way toward performing a cosmic inventory, identifying those components of the universe that dominate the energy budget at various epochs.

Implicit in this discussion will be the notion that the universe is smooth (none of the densities vary in space) and in equilibrium (the consequences of which will be explored in Section 2.3). In succeeding chapters, we will see that the deviations from equilibrium and smoothness are the source of much of the richness in the universe. Nonetheless, if only in order to understand the framework in which these deviations occur, a basic knowledge of the "zero order" universe is a must for any cosmologist.

In this chapter, I begin using units in which

$$\hbar = c = k_B = 1$$

Many papers employ these units, so it is important to get accustomed to them. Please work through Exercise 1 if you are uncomfortable with the idea of setting the speed of light to 1.

12.2.1 General Relativity

Most of cosmology can be learned with only a passing knowledge of general relativity. One must be familiar with the concept of a metric, understand geodesics, and be able to apply the Einstein equations to the Friedmann-Robertson-Walker (FRW) metric thereby relating the parameters in the metric to the density in the universe. Eq. (1.2) is the result of applying the Einstein equations to the zero order universe. We will derive it in this section. Chapters 4 and 5 apply them to the perturbed universe. With the experience we gain in this section, there will be nothing difficult about these subsequent chapters. The principles are identical; only the algebra will be a touch harder.

The Metric

Figure 1.1 from Chapter 1 highlights the fact that even if one knows the components of a vector, say the difference between two grid points there, the physical distance associated with this vector requires additional information. In the case of a smooth expanding universe, the scale factor connects the coordinate distance with the physical distance. More generally, the metric turns coordinate distance into physical distance and so will be an essential tool in our quest to make quantitative predictions in an expanding universe.

We are familiar with the metric for the Cartesian coordinate system which says that the square of the physical distance between two points separated by dx and dy in a 2D plane is $(dx)^2 + (dy)^2$. However, were we to use polar coordinates instead, the square of the physical distance would no longer be the sum of the square of the two coordinate differences. Rather, if the differences dr and $d\theta$ are small, the square of the distance between two points is $(dr)^2 + r^2(d\theta)^2 \neq (dr)^2 + (d\theta)^2$. This distance is invariant: an observer using Cartesian coordinates to find it would get the same result as one using polar coordinates. Thus another way of stating what a metric does is this: it turns observer-dependent coordinates into invariants. Mathematically, in the 2D plane, the invariant distance squared $dl^2 = \sum_{i,j=1,2} g_{ij} dx^i dx^j$. The metric g_{ij} in this 2D example is a 2×2 symmetric matrix. In Cartesian coordinates the metric is diagonal with each element equal to 1. In polar coordinates (taking $x^1 = r$ and $x^2 = \theta$) it is also diagonal with $g_{11} = 1$, but g_{22} which multiplies $(d\theta)^2$ is equal to r^2.

There is yet another way of thinking about a metric, using pictures. When handed a vector, we immediately think of a line with an arrow attached, the length of the line corresponding to the length of the vector and the arrow to its direction. In fact, this notion is rooted too firmly in Euclidean space. In actuality, the length of the vector depends on the metric. An intuitive way of understanding this is to consider the contour map in Figure 2.1. The number of lines crossed by a vector is a measure of the vertical distance traveled by a hiker. Vectors of the same apparent 2D length - corresponding to identical coordinate distances - can correspond to significantly different physical distances. Mathematically the surface of the Earth can be parametrized by two coordinates, say θ and ϕ. Then the metric is a very nontrivial function of θ and ϕ which accounts for all the elevation changes on the surface.

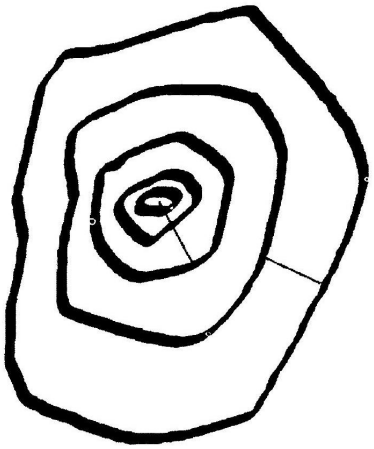


Figure 2.1. Contour map of a mountain. The closely spaced contours near the center correspond to rapid elevation gain. The two thin lines correspond to hikes of significantly different difficulty even though they appear to be of the same length. Similarly, the true length of a vector requires knowledge of the metric.

The great advantage of the metric is that it incorporates gravity. Instead of thinking of gravity as an external force and talking of particles moving in a gravitational field, we can include gravity in the metric and talk of particles moving freely in a distorted or curved space-time, one in which the metric cannot be converted everywhere into Euclidean form.

In four space-time dimensions the invariant includes time intervals as well, so

$$ds^2 = \sum_{\mu,\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu$$

where the indices μ and ν range from 0 to 3 (see the box on page 27), with the first one reserved for the time-like coordinate ($dx^0 = dt$) and the last three for spatial coordinates. Here I have explicitly written down the summation sign, but from now on we will use the convention that repeated indices are summed over. The metric $g_{\mu\nu}$ is necessarily symmetric, so in principle has four diagonal and six off-diagoral components. It provides the connection between values of the coordinates and the more physical measure of the interval ds^2 (sometimes called proper time). Special relativity is described by Minkowski space-time with the metric: $g_{\mu\nu} = \eta_{\mu\nu}$, with

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

What is the metric which describes the expanding universe? Let us return to the grid depicted in Figure 1.1. We said earlier that two grid points move away from each other, so that the distance between the two points is always proportional to the scale factor. If the comoving distance today is x_0, the physical distance between the two points at some earlier time t was $a(t)x_0$. At least in a flat (as opposed to open or closed) universe, the metric then is almost identical to the Minkowski metric, except that distance must be multiplied by the scale factor. This suggests that the metric in an expanding, flat universe is

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & a^2(t) & 0 & 0 \\ 0 & 0 & a^2(t) & 0 \\ 0 & 0 & 0 & a^2(t) \end{pmatrix}.$$

This is called the Friedmann-Robertson-Walker (FRW) metric.

As noted in Eq. (1.2), which we will shortly derive, the evolution of the scale factor depends on the density in the universe. When perturbations are introduced, the metric will become more complicated, and the perturbed part of the metric will be determined by the inhomogeneities in the matter and radiation.

Indices

In three dimensions, a vector $\vec{A}$ has three components, which we refer to as A^i, superscript i taking the values 1,2, or 3. The dot product of two vectors is then

$$\vec{A} \cdot \vec{B} = \sum_{i=1}^{3} A^i B^i := A^i B^i$$

where I have introduced the Einstein summation convention of not explicitly writing the $\sum$ sign when an index (in this case i) appears twice. Similarly, matrices can be written in component notation. For example, the product of two matrices **M** and **N** is

$$(\mathbf{MN})_{ij} = M_{ik} N_{kj}$$

again with implicit summation over k.

In relativity, two generalizations must be made. First, in relativity a vector has a fourth component, the time component. Since the spatial indices run from 1 to 3, it is conventional to use 0 for the time component. Greek letters are used to represent all four components, so $A^\mu = (A^0, A^i)$. The second, more subtle, feature of relativity is the distinction between upper and lower indices, the former associated with vectors and the latter with 1 -forms. One goes back and forth with the metric tensor, so

$$A_\mu = g_{\mu\nu} A^\nu \quad ; \quad A^\mu = g^{\mu\nu} A_\nu$$

A vector and a 1 -form can be contracted to produce an invariant, a scalar. For example, the statement that the four-momentum squared of a massless particle must vanish is

$$P^2 := P_\mu P^\mu = g_{\mu\nu} P^\mu P^\nu = 0.$$

This contraction is the equivalent of counting the contours crossed by a vector, as depicted in Figure 2.1.

Just as the metric can turn an upper index on a vector into a lower index, the metric can be used to raise and lower indices on tensors with an arbitrary number of indices. For example, raising the indices on the metric tensor itself leads to

$$g^{\mu\nu} = g^{\mu\kappa} g^{\nu\beta} g_{\alpha\beta}$$

If the index $\alpha = \nu$, then the first term on the right is equal to the term on the left, so if the combination of the last two terms on the right force α to be equal to ν, then the equation is satisfied. Therefore,

$$g^{\nu\beta} g_{\alpha\beta} = \delta^\nu_\alpha$$

where $\delta^\nu{}_\alpha$ is the Kronecker delta equal to zero unless $\nu = \alpha$ in which case it is 1.

The Geodesic Equation

In Minkowski space, particles travel in straight lines unless they are acted on by a force. Not surprisingly, the paths of particles in more general space-times are more complicated. The notion of a straight line gets generalized to a geodesic, the path followed by a particle in the absence of any forces. To express this in equations, we must generalize Newton's law with no forces. $d^2\vec{x}/dt^2 = 0$, to the expanding universe.

The machinery necessary to generalize $d^2\vec{x}/dt^2 = 0$ is perhaps best introduced by starting with a simple example: particle motion in a Euclidean 2D plane. In that case, the equations of motion in Cartesian coordinates $x^i = (x, y)$ are

$$\frac{d^2x^2}{dt^2} = 0$$

However, if we use polar coordinates $x'^2 = (r, \theta)$ instead, the equations for a free particle look significantly different. The fundamental difference between the two coordinate systems is that the basis vectors for polar coordinates $\hat{r}, \hat{\theta}$ vary in the plane. Therefore, $d^2\vec{x}'/dt^2 = 0$ does not imply that each coordinate r and θ satisfies $d^2 \cdot r'^i/dt^2 = 0$

To determine the equation satisfied by the polar coordinates, we can start from the Cartesian equation and then transform. In particular,

$$\frac{dx^i}{dt} = \frac{\partial x^i}{\partial x'^j}\frac{dx'^j}{dt}$$

$\partial x^i/\partial x'^j$ is called the transformation matrix going from one basis to another. In the case of Cartesian to polar coordinates in 2D, $x^1 = x'^1 \cos x'^2$ and $x^2 = x'^1 \sin x'^2$, so the transformation matrix is

$$\frac{\partial x^i}{\partial x'^3} = \begin{pmatrix} \cos x'^2 & -x'^1 \sin x'^2 \\ \sin x'^2 & x'^1 \cos x'^2 \end{pmatrix}$$

Therefore, the geodesic equation becomes

$$\frac{d}{dt}\left[\frac{dx'}{dt}\right] = \frac{d}{dt}\left[\frac{\partial x^i}{\partial x'}\frac{dx'}{dt}\right] = 0$$

The derivative with respect to time acts on both terms inside the brackets. If the transformation from the Cartesian basis to the new basis was linear, then the derivative acting on the transformation matrix would vanish, and the geodesic equation in the new basis would still be $d^2x'^i/dt^2 = 0$. In the case of polar coordinates, though, the transformation is not linear, and we need the fact that

$$\frac{d}{dt}\left(\frac{\partial x'}{\partial x'^j}\right) = \frac{\partial}{\partial x'^j}\left(\frac{dx'}{dt}\right)$$

$$= \frac{\partial^2 x^i}{\partial x'^j \partial x'^k}\frac{dx'^k}{dt}$$

where the first equality holds since derivatives commute and the second comes from inserting dx^i/dt from Eq. (2.12), changing dummy indices from $j \to k$. The geodesic equation in the new coordinates therefore becomes

$$\frac{d}{dt}\left[\frac{\partial x^i}{\partial x'^j}\frac{dx'^j}{dt}\right] = \frac{\partial x^i}{\partial x'^j}\frac{d^2x'^j}{dt^2} + \frac{\partial^2 x^i}{\partial x'^j \partial x'^k}\frac{dx'^k}{dt}\frac{dx'^j}{dt} = 0.$$

To get this in a more recognizable form, note that the term multiplying the second time derivative is the transformation matrix. If we multiply the equation by the inverse of this transformation matrix, then the second time derivative will stand alone, leaving

$$\frac{d^2x'^l}{dt^2} + \left[\left(\left\{\frac{\partial x}{\partial x'}\right\}^{-1}\right)_i^l \frac{\partial^2 x^i}{\partial x'^j \partial x'^k}\right] \frac{dx'^k}{dt}\frac{dx'^j}{dt} = 0$$

You can check that this rather cumbersome expression does indeed give the correct equations of motion in polar coordinates. More importantly, by keeping things general, we have derived the geodesic equation in a non-Cartesian basis.

It is convenient to define the Christoffel symbol, $\Gamma^l{}_{jk}$, to be the coefficient of the $\left(dx'^k/dt\right)\left(dx'^j/dt\right)$ term in Eq. (2.17). Note that by definition it is symmetric in its lower indices j and k. In a Cartesian coordinate system, the Christoffel symbol vanishes and the geodesic equation is simply $d^2x^i/dt^2 = 0$. But in general, the Christoffel symbol does not vanish; its presence describes geodesics in nontrivial coordinate systems. The reason why this generalized geodesic equation is so powerful is that in a nontrivial space-time such as the expanding universe it is not possible to find a fixed Cartesian coordinate system, so we need to know how particles travel in the more general case.

There are two small changes we need to make when importing the geodesic equation (2.17) into relativity. The first is trivial: allow the indices to range from 0 to 3 to include time and the three spatial dimensions. The second is also not surprising: since time is now one of our coordinates, it will not do to use it as the evolution parameter. Instead introduce a parameter λ which monotonically increases along the particle's path as in Figure 2.2. The geodesic equation then

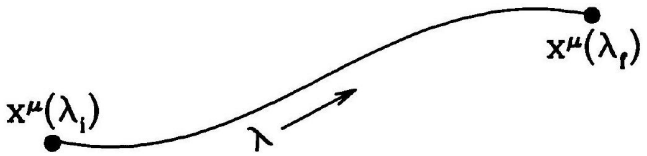


Figure 2.2. A particle's path is parametrized by λ, which monotonically increases from its

initial value λ_i to its final value λ_f. becomes

$$\frac{d^2x^\mu}{d\lambda^2} = -\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda}$$

We derived this equation transforming from a Cartesian basis, so that the Christoffel symbol is given by the term in square brackets in Eq. (2.17). It is almost always more convenient, however, to obtain the Christoffel symbol from the metric directly. A convenient formula expressing this dependence is

$$\Gamma^\mu_{\alpha,\beta} = \frac{g^{\mu\nu}}{2}\left[\frac{\partial g_{\alpha\nu}}{\partial x^\beta} + \frac{\partial g_{\beta\nu}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\nu}\right]$$

Note that the raised indices on $g^{\mu\nu}$ are important: $g^{\mu\nu}$ is the inverse of $g_{\mu\nu}$ (see the box on page 27). So $g^{\mu\nu}$ in the flat, FRW metric is identical to $g_{\mu\nu}$ except that its spatial elements are $1/a^2$ instead of a^2.

Using the general expression in Eq. (2.19) and the FRW metric in Eq. (2.4), we can derive the Christoffel symbol in an expanding, homogeneous universe. First we compute the components with upper index equal to zero, $\Gamma^0{}_{\alpha\beta}$. Since the metric is diagonal, the factor of $g^{0\nu}$ vanishes unless $\nu = 0$ in which case it is -1. Therefore,

$$\Gamma^0_{\alpha\beta} = \frac{-1}{2}\left[\frac{\partial g_{\alpha 0}}{\partial x^\beta} + \frac{\partial g_{\beta 0}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^0}\right]$$

The first two terms here reduce to derivatives of g_{00}. Since the FRW metric has constant g_{00}, these terms vanish, and we are left with

$$\Gamma^0_{\alpha\beta} = \frac{1}{2}\frac{\partial g_{\alpha\beta}}{\partial x^0}$$

The derivative is nonzero only if α and β are spatial indices, which will be identified with Roman letters i, j running from 1 to 3. Since $x^0 = t$, we have

$$\begin{aligned}\Gamma^0_{00} &= 0\\ \Gamma^0_{0i} &= \Gamma^0_{i0} = 0\\ \Gamma^0_{ij} &= \delta_{ij}\dot{a}a\end{aligned}$$

where overdots indicate derivatives with respect to time. [1] It is a straightforward, but useful, exercise to show that $\Gamma^i_{\alpha\beta}$ is nonzero only when one of its lower indices is zero and one is spatial, so that

$$\Gamma^i_{0j} = \Gamma^i_{j0} = \delta_{ij}\frac{\dot{a}}{a}$$

with all other $\Gamma^i_{\alpha\beta}$ zero.

[1] I will use this convention until Chapter 4. After that, overdots will denote derivatives with respect to conformal time. This has been a long, rather formal subsection, opening with the generalization of the geodesic equation to curved space-time and then proceeding with a calculation of the Christoffel symbol in the expanding universe described by the FRW metric. Before completing our main task and using the Einstein equations to derive Eq. (1.2), let's take a break and apply the geodesic equation to a single particle. In particular let's see how a particle's energy changes as the universe expands. We'll do the calculation here for a massless particle; an almost identical problem for a massive particle is relegated to

D-4.

Start with the four-dimensional energy-momentum vector $P^\alpha = (E, \vec{P})$, whose time component is the energy. We use this four-vector to define the parameter λin Eq. (2.18):

$$P^\alpha = \frac{dx^\alpha}{d\lambda}$$

This is an implicit definition of λ. Fortunately, one never needs to find λ explicitly, for it can be directly eliminated by noting that

$$\begin{aligned}\frac{d}{d\lambda} &= \frac{dx^0}{d\lambda}\frac{d}{dx^0}\\ &= E\frac{d}{dt}.\end{aligned}$$

The zeroth component of the geodesic equation (2.18) then becomes

$$E\frac{dE}{dt} = -\Gamma^0_{ij}P^iP^j$$

where the equality holds since only the spatial components of $\Gamma^0{}_{\alpha\beta}$ are nonzero. Inserting these components leads to a right-hand side equal to $-\delta_{ij}a\dot{a}P^iP^j$. A massless particle has energy-momentum [2] vector $(E, \vec{P})$ with zero magnitude:

$$g_{\mu\nu}P^\mu P^\nu = -E^2 + \delta_{ij}a^2 P^i P^j = 0$$

which enables us to write the right hand side of Eq. (2.26) as $-(\dot{a}/a)E^2$. Therefore, the geodesic equation yields

$$\frac{dE}{dt} + \frac{\dot{a}}{a}E = 0$$

the solution to which is

$$E \propto \frac{1}{a}.$$

This confirms our hand-waving argument in Chapter 1 that the energy of a massless particle should decrease as the universe expands since it is inversely proportional to its wavelength, which is being stretched along with the expansion. In Chapter 4 we will rederive this result in yet another way using the Boltzmann equation.

[2] Note that $\vec{P}$ measures motion on the comoving (nonexpanding) grid. The physical momentum which measures changes in physical distance is related to $\vec{P}$ by a factor of a. Hence the factor of a^2 in Eq. (2.27).

Einstein Equations

If you did a word search on the previous two subsections, you might be surprised to discover that the words "general relativity" never appeared. The concept of a metric and the realization that nontrivial metrics affect geodesics both exist completely independently of general relativity. The part of general relativity that is hidden above is that gravitation can be described by a metric, in our case by Eq. (2.4). There is a second aspect of general relativity, though: one which relates the metric to the matter and energy in the universe. This second part is contained in the Einstein equations, which relate the components of the Einstein tensor describing the geometry to the energy-momentum tensor describing the energy:

$$G_{\mu\nu} := R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\mathcal{R} = 8\pi G T_{\mu\nu}$$

Here $G_{\mu\nu}$ is the Einstein tensor; $R_{\mu\nu}$ is the Ricci tensor, which depends on the metric and its derivatives; $\mathcal{R}$, the Ricci scalar, is the contraction of the Ricci tensor ($\mathcal{R} := g^{\mu\nu}R_{\mu\nu}$) : G is Newton's constant; and $T_{\mu\nu}$ is the energy-momentum tensor. We will spend some time on the energy-momentum tensor in Section 2.3. For now, all we need to know is that it's a symmetric tensor describing the constituents of the universe. The left-hand side of Eq. (2.30) is a function of the metric, the right a function of the energy: the Einstein equations relate the two.

The Ricci tensor is most conveniently expressed in terms of the Christoffel symbol,

$$R_{\mu\nu} = \Gamma^\alpha_{\mu\nu,\alpha} - \Gamma^\alpha_{\mu\alpha,\nu} + \Gamma^\alpha_{\beta\alpha}\Gamma^\beta_{\mu\nu} - \Gamma^\alpha_{\beta\nu}\Gamma^\beta_{\mu\alpha}.$$

Here commas denote derivatives with respect to x. So, for example, $\Gamma^\alpha{}_{\mu\nu,\alpha} := \partial\Gamma^\alpha{}_{\mu\nu}/\partial x^\alpha$. Although this expression looks formidable, we have already done the hard work by computing the Christoffel symbol in an FRW universe. It turns out that there are only two sets of nonvanishing components of the Ricci tensor: one with $\mu = \nu = 0$ and the other with $\mu = \nu = i$.

Consider

$$R_{00} = \Gamma^\alpha_{00,\alpha} - \Gamma^\alpha_{0\alpha.0} + \Gamma^\alpha_{\beta\alpha}\Gamma^\beta{}_{00} - \Gamma^\alpha{}_{\beta 0}\Gamma^\beta{}_{0\alpha}.$$

Recall that the Christoffel symbol vanishes if its two lower indices are zero, so the first and third terms on the right vanish. Similarly, the indices α and β in the second and fourth terms must be spatial. We are left with

$$R_{00} = -\Gamma^{i}_{0i.0} - \Gamma^{i}{}_{j0}\Gamma^{j}{}_{0i}.$$

Using Eq. (2.23) leads directly to

$$\begin{aligned} R_{00} &= -\delta_{ii}\frac{\partial}{\partial t}\left(\frac{\dot{a}}{a}\right) - \left(\frac{\dot{a}}{a}\right)^2 \delta_{ij}\delta_{ij} \\ &= -3\left[\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2}\right] - 3\left(\frac{\dot{a}}{a}\right)^2 \\ &= -3\frac{\ddot{a}}{a} \end{aligned}$$

The factors of 3 on the second line arise since δ_{ii} means sum over all three spatial indices, counting one for each. I will leave the space-space component as an exercise; it is

$$R_{ij} = \delta_{ij}\left[2\dot{a}^2 + a\ddot{a}\right]$$

The next ingredient in the Einstein equations is the Ricci scalar, which we can now compute since

$$\begin{aligned} \mathcal{R} &:= g^{\mu\nu}R_{\mu\nu} \\ &= -R_{00} + \frac{1}{a^2}R_{ii} \end{aligned}$$

Again the sum over i leads to a factor of 3, so

$$\mathcal{R} = 6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right]$$

To understand the evolution of the scale factor in a homogeneous universe, we need consider only the time-time component of the Einstein equations:

$$R_{00} - \frac{1}{2}g_{00}\mathcal{R} = 8\pi G T_{00}$$

The terms on the left sum to $3\dot{a}^2/a^2$, and the time-time component of the energymomentum tensor is simply the energy density ρ. So we finally have

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho$$

To get this into the form of Eq. (1.2), recall that the left-hand side here is the square of the Hubble rate and that the critical density was defined as $\rho_{\rm cr} := 3H_0^2/8\pi G$. So, dividing both sides by H_0^2 leads to

$$\frac{H^2(t)}{H_0^2} = \frac{\rho}{\rho_{\rm cr}}$$

Here the energy density ρ counts the energy density from all species: matter, radiation, and the dark energy. In our derivation, we have assumed the universe is flat, so Eq. (2.40) does not contain a term corresponding to the curvature of the universe. I leave it as an exercise to derive the Einstein equation in an open universe.

12.2.2 Distances

We can anticipate that measuring distance in an expanding universe will be a tricky business. Referring back to the expanding grid of Figure 1.1, we immediately see two possible ways to measure distance, the comoving distance which remains fixed as the universe expands or the physical distance which grows simply because of the expansion. Frequently, neither of these two measures accurately describes the process of interest. For example light leaving a distant QSO at redshift 3 starts its journey towards us when the scale factor was only a quarter of its present value and ends it today when the universe has expanded by a factor of 4. Which distance do we use in that case to relate, say, the luminosity of the QSO to the flux we see?

The fundamental distance measure, from which all others may be calculated, is the distance on the comoving grid. If the universe is flat, as we will assume through most of this book, then computing distances on the comoving grid is easy: the distance between two points $\vec{x}_1$ and $\vec{x}_2$ is equal to $\left[(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - \ z_2)^2\right]^{1/2}$.

One very important comoving distance is the distance light could have traveled (in the absence of interactions) since $t = 0$. In a time dt, light travels a comoving distance $dx = dt/a$ (recall that we are setting $c = 1$), so the total comoving distance light could have traveled is

$$\eta := \int_0^t \frac{dt'}{a\left(t'\right)}$$

The reason this distance is so important is that no information could have propagated further (again on the comoving grid) than η since the beginning of time. Therefore, regions separated by distance greater than η are not causally connected. If they appear similar, we should be suspicious! We can think of η then as the comoving horizon. We can also think of η, which is monotonically increasing, as a time variable and call it the conformal time. Just like the time t, the temperature T, the redshift z, and the scale factor a, η can be used to discuss the evolution of the universe. In fact, for most purposes η is the most convenient time variable, so when we begin to study the evolution of perturbations, we will use it instead of t. In some simple cases, η can be expressed analytically in terms of a (Exercise 11). For example, in a matter-dominated universe, $\eta \propto a^{1/2}$, while $\eta \propto a$ in a radiation-dominated universe.

Another important comoving distance is that between a distant emitter and us. In that case, the comoving distance out to an object at scale factor a (or redshift $z = 1/a - 1$) is

$$\chi(a) = \int_{t(a)}^{t_0} \frac{dt'}{a\left(t'\right)} = \int_a^1 \frac{da'}{a'^2 H\left(a'\right)}.$$

Here I have changed the integration over t' to one over a', which brings in the additional factor of $da/dt = aH$ in the denominator. Typically we can see objects out to $z \lesssim 6$; at these late times radiation can be ignored (recall Figure 1.3). If the universe is purely matter dominated at such times, then $H \propto a^{-3/2}$ and we can do the integral in Eq. (2.42) analytically,

$$\chi^{\text{Flat.MD}}\left(a\right) = \frac{2}{H_0}\left[1 - a^{1/2}\right]$$

$$= \frac{2}{H_0}\left[1 - \frac{1}{\sqrt{1+z}}\right]$$

This comoving distance goes as z/H_0 for small z (verifying our hand-waving discussion of the small- z Hubble diagram in Section 1.2) and then asymptotes to $2/H_0$ as z gets very large.

A classic way to determine distances in astronomy is to measure the angle θ subtended by an object of known physical size l. The distance to that object (assuming the angle subtended is small) is then

$$d_A = \frac{l}{\theta}$$

The subscript $_A$ here denotes angular diameter distance. To compute the angular diameter distance in an expanding universe, we first note that the comoving size of the object is l/a. The comoving distance out to the object is given by Eq. (2.42), so the angle subtended is $\theta = (l/a)/\chi(a)$. Comparing with Eq. (2.44), we see that the angular diameter distance is

$$d_A^{\rm flat} = a\chi = \frac{\chi}{1+z}$$

Note that the angular diameter distance is equal to the comoving distance at low redshift, but actually decreases at very large redshift. At least in a flat universe, objects at large redshift appear larger than they would at intermediate redshift! The superscript here is a warning that this result holds only in a flat universe. In an open or closed universe, the curvature density is defined as $\Omega_k = 1 - \Omega_0$ where Ω_0 is the ratio of total to critical density today, including contributions from matter, radiation, and any other form of energy such as a cosmological constant. If the curvature is nonzero, the angular diameter distance generalizes to

$$d_A = \frac{a}{H_0\sqrt{|\Omega_k|}} \begin{cases} \sinh\left[\sqrt{\Omega_k} H_0 \chi\right] & \Omega_k > 0 \\ \sin\left[\sqrt{-\Omega_k} H_0 \chi\right] & \Omega_k < 0 \end{cases}$$

Note that both of these expressions reduce to the flat case in the limit that the curvature density Ω_k goes to zero. Figure 2.3 shows the angular diameter distance in a flat universe, both with and without a cosmological constant.

Another way of inferring distances in astronomy is to measure the flux from an object of known luminosity. Recall that (forgetting about expansion for the moment) the observed flux F a distance d from a source of known luminosity L is

$$F = \frac{L}{4\pi d^2}$$

since the total luminosity through a spherical shell with area $4\pi d^2$ is constant. How does this result generalize to an expanding universe? Again it is simplest to work

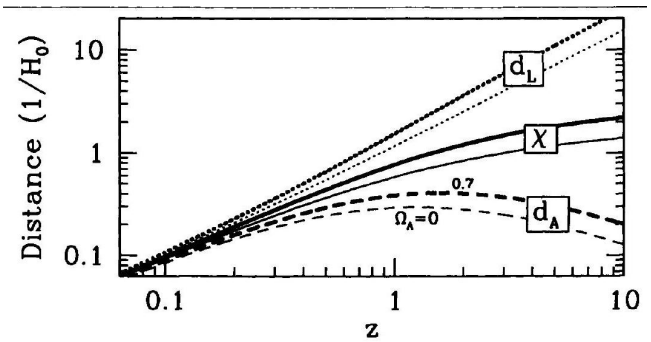


Figure 2.3. Three distance measures in a flat expanding universe. From top to bottom, the luminosity distance, the comoving distance, and the angular diameter distance. The pair of lines in each case is for a flat universe with matter only (light curves) and 70% cosmological constant Λ (heavy curves). In a Λ-dominated universe, distances out to fixed redshift are larger than in a matter-dominated universe.

on the comoving grid, this time with the source centered at the origin. The flux we observe is

$$F = \frac{L(\chi)}{4\pi\chi^2(a)}$$

where $L(\chi)$ is the luminosity through a (comoving) spherical shell with radius $\chi(a)$. To further simplify, let's assume that the photons are all emitted with the same energy. Then $L(\chi)$ is this energy multiplied by the number of photons passing through a (comoving) spherical shell per unit time. In a fixed time interval, photons travel farther on the comoving grid at early times than at late times since the associated physical distance at early times is smaller. Therefore, the number of photons crossing a shell in the fixed time interval will be smaller today than at emission, smaller by a factor of a. Similarly, the energy of the photons will be smaller today than at emission, because of expansion. Therefore, the energy per unit time passing through a comoving shell a distance $\chi(a)$ (i.e., our distance) from the source will be a factor of a^2 smaller than the luminosity at the source. The flux we observe therefore will be

$$F = \frac{La^2}{4\pi\chi^2(a)}$$

where L is the luminosity at the source. We can keep [3] Eq. (2.47) in an expanding universe as long as we define the luminosity distance

$$d_L := \frac{\chi}{a}$$

[3] Actually there is one more difference that needs to be accounted for: the observed luminosity is related to the emitted luminosity at a different wavelength. Here we have assumed a detector which counts all the photons. The luminosity distance is shown in Figure 2.3.

All three distances are larger in a universe with a cosmological constant than in one without. This follows from the fact that the energy density, and therefore the expansion rate, is smaller in a Λ-dominated universe. The universe was therefore expanding more slowly early on, and light had more time to travel from distant objects to us. These distant objects will therefore appear fainter to us than if the universe was dominated by matter only.

12.2.3 Evolution of Energy

Let us return to the energy-momentum tensor on the right-hand side of the Einstein equations. We will eventually include perturbations to $T^\mu{}_\nu$, but in the spirit of this chapter, first consider the case of a perfect isotropic fluid. Then,

$$T^\mu{}_\nu = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & \mathcal{P} & 0 & 0 \\ 0 & 0 & \mathcal{P} & 0 \\ 0 & 0 & 0 & \mathcal{P} \end{pmatrix}$$

where $\mathcal{P}$ is the pressure of the fluid.

How do the components of the energy-momentum tensor evolve with time? Consider first the case where there is no gravity and velocities are negligible. The pressure and energy density in that case evolve according to the continuity equation, $\partial\rho/\partial t = 0$, and the Euler equation, $\partial\mathcal{P}/\partial x^2 = 0$. This can be promoted to a 4 component conservation equation for the energy-momentum tensor: $\partial T^\mu{}_\nu/\partial x^\mu = 0$. In an expanding universe, however, the conservation criterion must be modified. Instead, conservation implies the vanishing of the covariant derivative:

$$T^{\mu}{}_{\nu;\mu} := \frac{\partial T^{\mu}{}_{\nu}}{\partial x^{\mu}} + \Gamma^{\mu}{}_{\alpha\mu}T^{\alpha}{}_{\nu} - \Gamma^{\alpha}{}_{\nu\mu}T^{\mu}{}_{\alpha}.$$

The vanishing of $T^{\mu}_{\nu;\mu}$ is four separate equations; let's consider the $\nu = 0$ component. This is

$$\frac{\partial T^{\mu}{}_{0}}{\partial x^{\mu}} + \Gamma^{\mu}_{\alpha\mu}T^{\alpha}_{0} - \Gamma^{\alpha}_{0\mu}T^{\mu}_{\alpha} = 0$$

Since we are assuming isotropy, $T^{i}{}_{0}$ vanishes, so the dummy indices μ in the first term and α in the second must be equal to zero:

$$\frac{-\partial\rho}{\partial t} - \Gamma^{\mu}{}_{0\mu}\rho - \Gamma^{\alpha}{}_{0\mu}T^{\mu}{}_{\alpha} = 0.$$

From Eq. (2.23), $\Gamma^{\alpha}_{0\mu}$ vanishes unless α, μ are spatial indices equal to each other, in which case it is $\dot{a}/a$. So, the conservation law in an expanding universe reads

$$\frac{\partial\rho}{\partial t} + \frac{\dot{a}}{a}[3\rho + 3\mathcal{P}] = 0$$

Rearranging terms, we have

$$a^{-3}\frac{\partial\left[\rho a^{3}\right]}{\partial t} = -3\frac{\dot{a}}{a}\mathcal{P}$$

The conservation law can be applied immediately to glean information about the scaling of both matter and radiation with the expansion. Matter has effectively zero pressure, so

$$\frac{\partial\left[\rho_{\mathrm{m}}a^{3}\right]}{\partial t} = 0$$

implying that the energy density of matter $\rho_{\mathrm{m}} \propto a^{-3}$. We anticipated this result in Chapter 1 based on the simple notion that the mass remains constant, while the number density scales as the inverse volume. The application to radiation also offers no surprises. Radiation has $\mathcal{P} = \rho/3$ (Exercise 14), so working from Eq. (2.55),

$$\begin{aligned}\frac{\partial\rho_{r}}{\partial t} + \frac{\dot{a}}{a}4\rho_{r} &= a^{-4}\frac{\partial\left[\rho_{r}a^{4}\right]}{\partial t}\\ &= 0.\end{aligned}$$

Therefore, the energy density of radiation $\rho_r \propto a^{-4}$, accounting for the decrease in energy per particle as the universe expands.

Through most of the early universe, reactions proceeded rapidly enough to keep particles in equilibrium, different species sharing a common temperature. We will often want to express the energy density and pressure in terms of this temperature. For this reason, and many others which will emerge over the next few chapters, it is convenient to introduce the occupation number, or distribution function, of a species. This counts the number of particles in a given region in phase space around position $\vec{x}$ and momentum $\vec{p}$[4] The energy of a species is then obtained by summing the energy over all of phase space elements: $\sum f(\vec{x},\vec{p})E(p)$ with $E(p) = \sqrt{p^2+m^2}$. How many phase space elements are there in a region of "volume" d^3xd^3p? By Heisenberg's principle, no particle can be localized into a region of phase space smaller than $(2\pi\hbar)^3$, so this is the size of a fundamental element. Therefore, the number of phase space elements in d^3xd^3p is $d^3xd^3p/(2\pi\hbar)^3$ (see Figure 2.4), and the energy density is

$$\rho_i = g_i\int\frac{d^3p}{(2\pi)^3}f_i(\vec{x},\vec{p})E(p)$$

where i labels different species, g_i is the degeneracy of the species (e.g., equal to 2 for the photon for its spin states), and I have gone back to $\hbar = 1$. In equilibrium at temperature T, bosons such as photons have Bose-Einstein distributions,

$$f_{\rm BE} = \frac{1}{e^{(E-\mu)/T} - 1}$$

and fermions such as electrons have Fermi-Dirac distributions,

[4] By p here I mean not the comoving momentum defined in Eq. (2.24), but rather the proper momentum which decreases with the expansion. See Exercise 15 for a discussion.

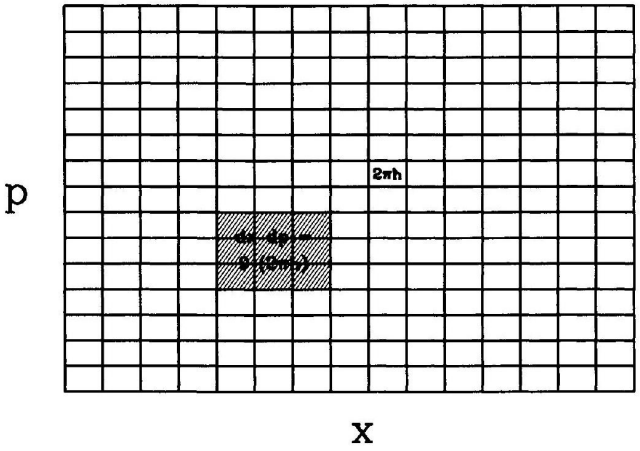


Figure 2.4. Phase space of position and momentum in one dimension. Volume of each cell is $2\pi\hbar$, the smallest region into which a particle can be confined because of Heisenberg's principle. Shaded region has infinitesmal volume $dxdp$. This covers nine cells. To count the appropriate number of cells, therefore, the phase space integral must be $\int dxdp/(2\pi\hbar)$.

$$f_{\rm FD} = \frac{1}{e^{(E-\mu)/T} + 1}$$

with μ the chemical potential. It should be noted that these distributions do not depend on position $\vec{x}$ or on the direction of the momentum $\hat{p}$, simply on the magnitude p. This is a feature of the zero-order, smooth universe. When we come to consider inhomogenities and anisotropies, we will see that the distribution functions have small perturbations around these zero order values, and the perturbations do depend on position and on the direction of propagation.

The pressure can be similarly expressed as an integral over the distribution function,

$$\mathcal{P}_i = g_i \int \frac{d^3p}{(2\pi)^3} f_i(\vec{x}, \vec{p}) \frac{p^2}{3E(p)}$$

For almost all particles at almost all times in the universe, the chemical potential is much smaller than the temperature. To a good approximation, then, the distribution function depends only on E/T and the pressure satisfies (Exercise 14)

$$\frac{\partial \mathcal{P}_i}{\partial T} = \frac{\rho_i + \mathcal{P}_i}{T}$$

This relation can be used to show that the entropy density in the universe scales as a^{-3}. To see this, let's rewrite Eq. (2.56) as

$$a^{-3} \frac{\partial \left[(\rho + \mathcal{P})a^3\right]}{\partial t} - \frac{\partial \mathcal{P}}{\partial t} = 0$$

The derivative of the pressure with respect to time can be written as $(dT/dt)(\partial\mathcal{P}/\partial T)$ so

$$a^{-3}\frac{\partial\left[(\rho+\mathcal{P})a^3\right]}{\partial t}-\frac{dT}{dt}\frac{\rho+\mathcal{P}}{T}=a^{-3}T\frac{\partial}{\partial t}\left[\frac{(\rho+\mathcal{P})a^3}{T}\right]$$
$$=0.$$

So the entropy density [5]

$$s:=\frac{\rho+\mathcal{P}}{T}$$

scales as a^{-3}. Although we have framed the argument in terms of a single species, this scaling holds for the total entropy including all species in equilibrium. In fact, even if two species have different temperatures, the sum of their entropy densities still scales as a^{-3}. We will make use of this fact shortly when computing the relative temperatures of neutrinos and photons in the universe.

12.2.4 Cosmic Inventory

Armed with an expression for the energy density of a given species (Eq. (2.59)), and a knowledge of how it evolves in time (Eq. (2.56)), we can now tackle quantitatively the question of how much energy is contributed by the different components of the universe.

Photons

The temperature of the CMB photons has been measured extraordinarily precisely by the FIRAS instrument aboard the COBE satellite, $T=2.725\pm0.002$ K (Mather et al., 1999). The energy density associated with this radiation is

$$\rho_\gamma=2\int\frac{d^3p}{(2\pi)^3}\frac{1}{e^{p/T}-1}p$$

The factor of 2 in front of Eq. (2.67) accounts for the two spin states of the photon. The energy of a given state is simply equal to p since the photon is massless. The chemical potential is zero; we expect this theoretically because early in the universe, photon number is not conserved (e.g., electrons and positrons can annihilate to produce photons). We also know it observationally because the spectrum of the CMB has been measured so accurately. The limits on a chemical potential are $\mu/T<9\times10^{-5}$ (Fixsen et al., 1996), so μ can be safely ignored. Since there is

[5] Technically, there is another term in the entropy density-proportional to the chemical potential-but, as mentioned above, this term is usually irrelevant in cosmology. Even with nonzero chemical potential, though, the entropy density scales as a^{-3}. no angular dependence in the integrand of Eq. (2.67). the angular integral yields a factor of 4π and we are left with a ome-dimensional integral. Define a dummy variable $x:=p/T$; then

$$\rho_\gamma=\frac{8\pi T^4}{(2\pi)^3}\int_0^\infty\frac{dxx^3}{e^x-1}$$

The integral can be expressed in terms of the Riemann ζ function; it is $6\zeta(4)=\pi^4/15$, so

$$\rho_\gamma=\frac{\pi^2}{15}T^4$$

Since we derived (Eq. (2.58)) that the energy density of radiation scales as a^{-4}, the temperature of the CMB must scale as a^{-1}.

It will be useful to have all energy densities in the same units. The simplest way to do this is to divide all energy densities by the critical density today. [6] Thus,

$$\frac{\rho_\gamma}{\rho_{\rm cr}} = \frac{\pi^2}{15}\left(\frac{2.725\ {\rm K}}{a}\right)^4 \frac{1}{8.098\times 10^{-11}h^2{\rm eV}^4}$$
$$= \frac{2.47\times 10^{-5}}{h^2a^4}$$

where to get the last line, it is useful to remember the conversion between kelvin and eV : 11605 K = 1eV. To reiterate an important point, the photon energy density in Eq. (2.70) depends on time via the scale factor, but has no spatial dependence. This is because we have used the zero-order distribution function, the Bose-Einstein function, for the photons. In fact there are small perturbations around this zeroorder distribution function. These do have a spatial dependence and correspond to the anisotropies in the CMB.

Baryons

Unlike the CMB, baryons [7] cannot be simply described as a gas with a temperature and zero chemical potential. Therefore, the baryon density must be measured directly, not via a temperature. There are now four established ways of measuring the baryon density, and these all seem to agree reasonably well (Fukugita, Hogan, and Peebles 1998). These are all measurements at different redshifts, and we know that the density scales as a^{-3}, so to facilitate comparison, one defines Ω_b via

$$\frac{\rho_b}{\rho_{\rm cr}} = \Omega_b a^{-3}$$

That is, Ω_b is the ratio of the baryon density to the critical density today.

[6] The critical density - just like the Hubble rate which defines it - changes with time. However, it is common to define $\rho_{\rm cr}$ to be a constant, the critical density today, and I will follow this convention.

'I refer to all the nuclei and electrons in the universe as baryons. This is technically incorrect (electrons are leptons), but nuclei are so much more massive than electrons that virtually all the mass is in the baryons. The simplest way is to observe the baryons today in galaxies. The greatest contribution to the density, though, comes not from stars in galaxies, but rather from gas in groups of galaxies. In these groups, Ω_b is about 0.02. The second way to count baryons is by looking at the spectra of distant quasars. The amount of light absorbed from these beacons is a measure of the intervening hydrogen, and hence the baryon density. These estimates (Rauch et al., 1997) suggest $\Omega_b h^{1.5} \simeq 0.02$ with a fairly large uncertainty. A third method is to infer the baryon density by careful scrutiny of the anisotropies in the universe. As we will see in Chapter 8, these depend on the baryon density. Preliminary results (Pryke et al., 2001; Netterfield et al., 2001) give $\Omega_b h^2 = 0.024^{+0.004}_{-0.003}$ from the CMB. Finally, we will see in Chapter 3 that the light element abundances are sensitive to the baryon density, and that estimates of these abundances pin down $\Omega_b h^2 = 0.0205 \pm 0.0018$.

Remarkably, then, these estimates of the baryon density with very different techniques all agree. [8] They all place the baryon density at roughly $2-5\%$ of the critical density. The total matter density in the universe is higher than this, so there must be matter in the universe that is nonbaryonic.

Matter

All of the methods of measuring the baryon density mentioned above involve the interaction of matter and radiation. For example, simply counting stars works at some level because we

roughly know how much mass is required to output the light from a typical star. There are, however, methods of measuring the mass of matter that do not rely on the way light and matter interact. These classically have involved measuring the gravitational field in a given system, thereby inferring information about the mass responsible for that field. Figure 2.5 shows the inferred mass-tolight ratios of many systems, ranging from galaxies to superclusters. Historically this ratio was measured on small scales first, suggesting that the density in the universe was far less than critical. As more large-scale data were obtained, the steady increase in the mass/light ratio led some cosmologists to speculate that eventually we would find that the density was critical. Bahcall and collaborators (Bahcall, Lubin, and Dorman 1995; Bahcall et al., 2000), however, have argued that mass-to-light ratios do not increase past $R \sim 1$Mpc; a leveling off occurs consistent with a matter density $\Omega_m \simeq 0.3$, where Ω_m is the ratio of the total matter density today to the critical density and

$$\rho_{\rm m} = \Omega_m \rho_{\rm cr} a^{-3}.$$

Recently a number of other techniques for inferring the matter density have emerged. We will see in Chapter 7 that the distribution of galaxies in the universe, in particular the power spectrum of this distribution, is very sensitive to $\Omega_m h$;

[8] Whether or not agreement holds is subject to debate. There have been claims (e.g., Persic and Salucci, 1992) that there is a missing baryon problem because the present-day abundance appears to be lower than that inferred from the light element abundances.

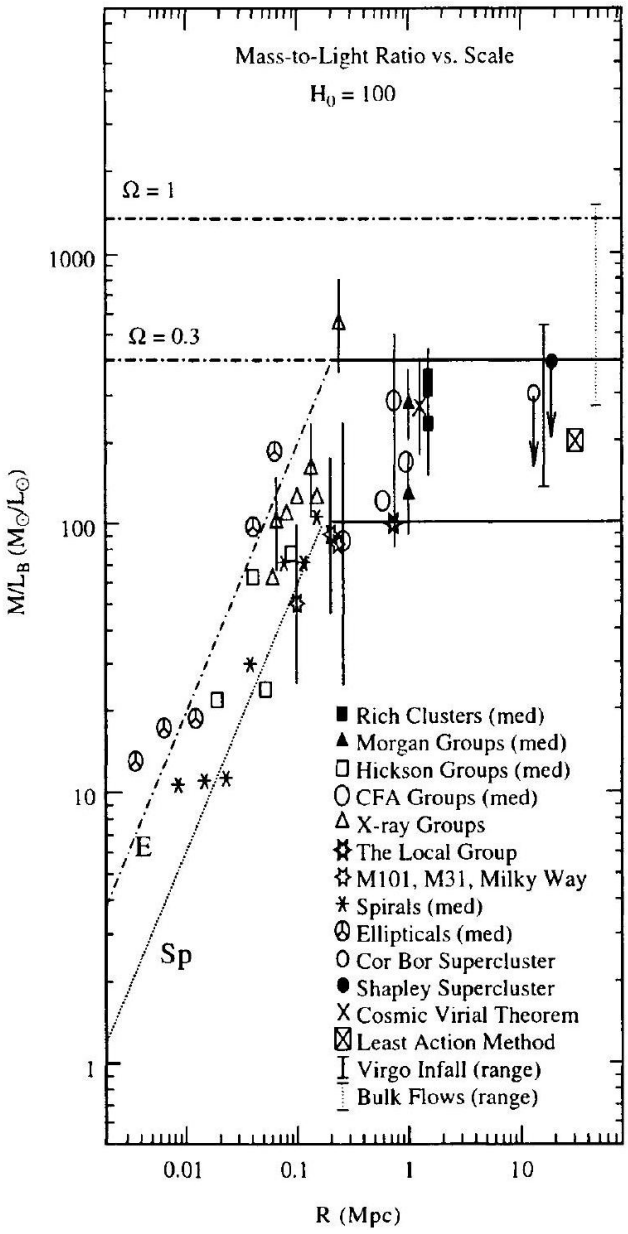


Figure 2.5. Mass vs. light ratio as a function of scale (Bahcall et al., 2000). On the largest scales, the ratio flattens so that $\Omega_m \simeq 0.3$.

virtually all galaxy surveys [9] have inferred $\Omega_m h \simeq 0.2$. Another cosmological probe

[9] To mention three examples, the Automated Plate Measuring (APM) Survey, to be discussed further in Chapter 9, has been analyzed by Efstathiou and Moody (2001); the Two Degree Field (2DF) by Percival et al. (2001), and early data from the Sloan Digital Sky Survey by Dodelson *et* al. (2001). These groups found $\Omega_m h = 0.14^{+0.24}_{-0.09}, 0.20 \pm 0.03$, and $0.14^{+0.11}_{-0.06}$, respectively. we will encounter in Chapter 9 is the cosmic velocity field (Strauss and Willick, 1995) and its relation to the observed galaxy distribution. These are related by the continuity equation, a relation sensitive to Ω_n. Again most of the measurements are clustered around $\Omega_m \simeq 0.3$.

Another way of measuring the total mass density is to pick out observations sensitive to Ω_b/Ω_m and use the apparent value of Ω_b to infer the matter density. For example, most of the baryonic mass in a galaxy cluster is in the form of hot gas. The ratio of the mass of gas in clusters of galaxies to the total mass can be measured either by looking for X-ray emission (White et al., 1993) or by looking at the electron-heated CMB in the direction of the cluster (Grego et al., 2001). If this ratio is characteristic of the universe as a whole - and it probably

is, because clusters are so large - then the cosmic baryon to matter ratio is around 20%. Since baryons make up only about 5% of the critical density, the total matter density is inferred to be about 0.25. Another way of inferring the baryon/matter ratio is by looking for features in the power spectrum of galaxies; if the baryon fraction is truly of order 20%, then there will be wiggles in the spectrum (again Chapter 7). There are tentative hints of these wiggles in the early data from the Two Degree Field (2DF) survey (Percival et al., 2001). These pin down $\Omega_b/\Omega_m = 0.15 \pm 0.07$, consistent with the cluster observations. Finally, the anisotropies in the CMB (Chapter 8) are sensitive to the matter density $\Omega_m h^2$. Recent determinations indicate $\Omega_m h^2 = 0.16 \pm 0.04$ (Pryke et al., 2001; Netterfield et al., 2001). Given the fact that current best estimates of the Hubble constant give $h = 0.72$, the CMB observations also are consistent with a matter density equal to 30% of the critical density.

We therefore have an enormous amount of evidence telling us that the baryon density is of order 5% the critical density, while the total matter density is some five times larger. Most of the matter in the universe must not be baryons. It must be some new form of matter: dark matter.

Neutrinos

The next component we need consider is the neutrino. Unlike photons and baryons, cosmic neutrinos have not been observed, so arguments about their contribution to the energy density are necessarily theoretical. However, these theoretical arguments are quite strong, based on very well-understood physics.

A basic understanding of the interaction rates of neutrinos enables us to argue that neutrinos were once kept in equilibrium with the rest of the cosmic plasma. Since they are fermions, their distribution was Fermi-Dirac with zero chemical potential. At late times, they lost contact with the plasma because their interactions are weak. Nonetheless, their distribution remained Fermi-Dirac, with their temperature simply falling as a^{-1}. The main task therefore is to relate the neutrino temperature to the photon temperature today. The tricky part of this is the annihilation of electrons and positrons when the cosmic temperature was of order the electron mass. Neutrinos lost contact with the cosmic plasma slightly before this annihilation, so they did not inherit any of the associated energy. The photons, which did, are therefore hotter than the neutrinos.

We can account for the annhilation of electrons and positrons by using the fact that the total entropy density s (Eq. (2.66)) scales as a^{-3}. Massless bosons contribute $2\pi^2 T^3/45$ to the entropy density for each spin state, while massless fermions contribute 7/8 this, and massive particles contribute negligibly (Exercise 17). Before annihilation, the fermions are electrons)2 spin states), positrons (2), neutrinos (3 generations and one spin state) and anti-neutrinos (3). The bosons are photons (2 spin states). So at a_1 before annihilation,

$$\begin{aligned} s\left(a_1\right) &= \frac{2\pi^2}{45} T_1^3 [2 + (7/8)(2+2+3+3)] \\ &= \frac{43\pi^2}{90} T_1^3 \end{aligned}$$

where T_1 is the common temperature at a_1. After annihilation, the electrons and positrons have gone away and the photon and neutrino temperatures are no longer identical: we must distinguish between them. Therefore, the entropy density is

$$s\left(a_2\right) = \frac{2\pi^2}{45}\left[2T_\gamma^3 + \frac{7}{8}6T_\nu^3\right].$$

Equating $s\left(a_1\right)a_1^3$ with $s\left(a_2\right)a_2^3$ leads to

$$\frac{43}{2}\left(a_1 T_1\right)^3 = 4\left[\left(\frac{T_\gamma}{T_\nu}\right)^3 + \frac{21}{8}\right]\left(T_\nu\left(a_2\right)a_2\right)^3.$$

But the neutrino temperature scales throughout as a^{-1}, so $a_1 T_1 = a_2 T_\nu\left(a_2\right)$. Therefore, the ratio of the two temperatures is

$$\frac{T_\nu}{T_\gamma} = \left(\frac{4}{11}\right)^{1/3}$$

We can now evaluate the energy density of neutrinos in the universe. Let's sum up what we know about the cosmic abundance of neutrinos

- One spin degree of freedom for neutrinos
- Neutrino has antiparticle
- Three generations of neutrinos
- Neutrinos are fermions $\rightarrow$ Fermi-Dirac distribution function
- Neutrino temperature is lower by a factor of $(4/11)^{1/3}$ since photons are heated by e^+e^-annihilation

The first three items on the list then imply that the degeneracy factor of neutrinos is equal to six. The fourth means we need to change the denominator in the integrand in Eq. (2.67) to $e^{p/T}+1$. The Fermi-Dirac integral is then smaller by a factor of 7/8. Finally, since the energy density of a massless particle scales as T^4, the last item implies that the neutrino energy density is smaller than the photon density by $(4/11)^{4/3}$. Putting all these factors together leads to

$$\rho_\nu = 3\frac{7}{8}\left(\frac{4}{11}\right)^{4/3}\rho_\gamma$$

Equivalently, if there were three species of massless neutrinos today, then their contribution to the energy density would be

$$\Omega_\nu := \left.\frac{\rho_\nu}{\rho_{\rm cr}}\right|_{\rm today} = \frac{1.68\times10^{-5}}{h^2} \quad m_\nu = 0.$$

In reality, all the neutrinos do not appear to be massless. Observations of neutrinos from both the sun (Bahcall, 1989) and from our atmosphere (Fukuda et al., 1998) strongly suggest that neutrinos of different flavors (generations) oscillate into each other. This can happen only if the neutrinos have mass. The atmospheric neutrino observations in particular imply that at least one neutrino has a mass larger than 0.05eV.[10] The energy density of a massive neutrino is

$$\rho_\nu = 2\int\frac{d^3p}{(2\pi)^3}\frac{\sqrt{p^2+m_\nu^2}}{e^{p/T_\nu}+1}$$

At high temperatures, this reduces to Eq. (2.77) (without the 3), so when considering neutrinos in the early universe, it is often sufficient to use Eq. (2.77). Indeed, we will do this shortly when we come to esimtate the epoch at which the energy density of matter equals that of radiation. At late times, the massive neutrino energy density is $m_\nu n_\nu$, with the neutrino number density equal to $3n_\gamma/11$ (Exercise 18). As can be seen from Figure 2.6, the transition takes place when $T_\nu \sim m_\nu$. Therefore,

$$\Omega_\nu = \frac{m_\nu}{94h^2\text{eV}} \quad m_\nu \neq 0$$

Those who trafficked in both astrophysics and particle physics (Gerstein and Zel'dovich, 1966; Marx and Szalay, 1972; Cowsik and McClelland, 1972) early on noted that the simple observation that the total density was not much greater than the critical density leads to constraints on neutrino mass, constraints much more stringent than those obtainable at accelerators. When the need for nonbaryonic dark matter first became evident, a number of cosmologists (e.g., Gunn et al., 1978) proposed neutrinos as the natural candidate. Subsequent studies (Bond, Efstathiou, and Silk, 1980; White, Frenk, and Davis 1983) of the structure of the universe with neutrinos as the dominant dark matter component looked significantly different from the actual universe. Nonetheless, the possibility that neutrinos might make up a fraction of the total density reemerged in the 1990s. We can then hope to detect a trace amount - corresponding to masses smaller than an eV - by observing its effect on large-scale structure.

10 The oscillation experiments are sensitive to mass differences, $m_2^2 - m_1^2$, so the actual constraint is that the mass squared difference is of order 10^{-3}. This could also be accomodated with two nearly degenerate masses with a small splitting.

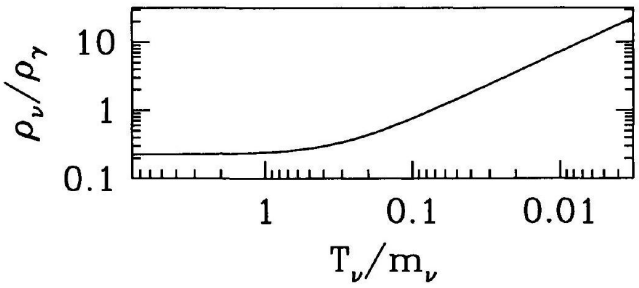


Figure 2.6. Energy density of one generation of massive neutrinos as compared with the density in the CMB. At high temperatures, the ratio is a fixed constant; at low temperatures, the neutrino behaves like nonrelativistic matter (scaling as a^{-3}) and so begins to dominate over the photon density (which scales as a^{-4}).

Dark Energy

There are two sets of evidence pointing toward the existence of something else, something beyond the radiation and matter itemized above. The first comes from a simple budgetary shortfall. The total energy density of the universe is very close to critical. We expect this theoretically (Chapter 6) and we observe it in the anisotropy pattern of the CMB (Chapter 8). Yet, the total matter density inferred from observations is only a third critical. The remaining two-thirds of the density in the universe must be in some smooth, unclustered form, dubbed [11] dark energy. The second set of evidence is more direct. Given the energy composition of the universe, one can compute a theoretical distance vs redshift diagram. This relation can then be tested observationally.

In 1998, two groups (Riess et al., 1998, Perlmutter et al., 1999) observing supernovae reported direct evidence for dark energy. The evidence is based on the difference between the luminosity distance in a universe dominated by dark matter and one dominated by dark energy. As Figure 2.3 indicates, the luminosity distance is larger for objects at high redshifts in a dark-energy-dominated universe. Therefore, objects of fixed intrinsic brightness will appear fainter if the universe is composed of dark energy.

[11] After the discovery, quite a bit of attention was focused on choosing an appropriate name. Cosmological constant, everyone's initial moniker, is too restrictive in that the energy density is constant at all times, and we do not yet know that this is true of the dark energy.

Variable cosmological constant fixes that problems but introduces an inherent contradiction ("variable" and "constant"?). Variable Lambda using the Greek letter reserved for the cosmological constant is too obscure. Quintessence is a good choice: it expresses the fact that, after cosmological photons, baryons, neutrinos, and dark matter, there is a fifth essence in the universe. It seems to me that dark energy has become a bit more popular, with quintessence referring to the subset of models in which the energy density can be associated with a time-dependent scalar field. More concretely, the luminosity distance of Eq. (2.50) can be used to find the apparent magnitude m of a source with absolute magnitude M. Magnitudes are related to fluxes via $m = -(5/2)\log(F)+$ constant. Since the flux scales as d_L^{-2}, the apparent magnitude $m = M + 5\log(d_L) +$ constant. The convention is that

$$m - M = 5\log\left(\frac{d_L}{10\text{pc}}\right) + K$$

where K is a correction for the shifting of the spectrum into or out of the wavelength range measured due to expansion.

The two groups measured the apparent magnitudes of dozens of Type Ia supernovae, which are known to be standard candles, i.e., they have nearly identical absolute magnitudes. Although they were able to place tight constraints on dark energy using the many supernovae that they detected, we can get a feel for the measurement by simply considering two of these. Consider then Supernova 1997ap, found at redshift $z = 0.83$ with apparent magnitude $m = 24.32$, and Supernova 1992P, found at low redshift $z = 0.026$ with apparent magnitude $m = 16.08$. Since the absolute magnitudes of these are the same, the difference in apparent magnitudes is due solely to the difference in luminosity distance:

$$24.32 - 16.08 = 5\log\left(d_L(z = 0.83)\right) - 5\log\left(d_L(z = 0.026)\right).$$

The nearby luminosity distance is independent of cosmology, simply equal to $z/H_0 = 0.026/H_0$. Therefore, the only unknown remaining in Eq. (2.82) is fxed by the observations to be

$$H_0 d_L(z = 0.83) = 1.16$$

In a flat, matter-dominated universe ($\Omega_m = 1$), the luminosity distance out to $z = 0.83$ is equal to $0.95H_0^{-1}$, whereas a universe with $\Omega_m = 0.3$ and a cosmological constant $\Omega_\Lambda = 0.7$ has a luminosity distance of 1.23H_0^{-1}. The apparent magnitude of this single distant supernova then suggests that dark energy pervades the universe.

Of course, the discussion of the previous paragraph does not account for uncertainties (typical uncertainties in the magnitudes are of order 0.2), nor does it do a careful fit to all known supernovae, allow for extinction by dust, or allow for the variation of the absolute magnitude correlated with the duration. The supernova groups did all of those things, and emerged with the constraints shown in Figure 2.7. The two free cosmological parameters are the matter density Ω_m and a cosmological constant Ω_Λ, something we now recognize as one possible form of dark energy, one in which the energy density is constant. Note that the "theorists' dream" universe - flat and matter-dominated ($\Omega_m = 1$) - is excluded with high confidence. Indeed, even a pure open universe with $\Omega_m = 0.3, \Omega_\Lambda = 0$ is strongly disfavored by the supernova data.

While highly popular, Figure 2.7 suffers from two drawbacks. It allows for too much freedom in one sector and too little in another. Most of the region in the figure is taken up by a universe with both dark energy and nonzero curvature (not flat).

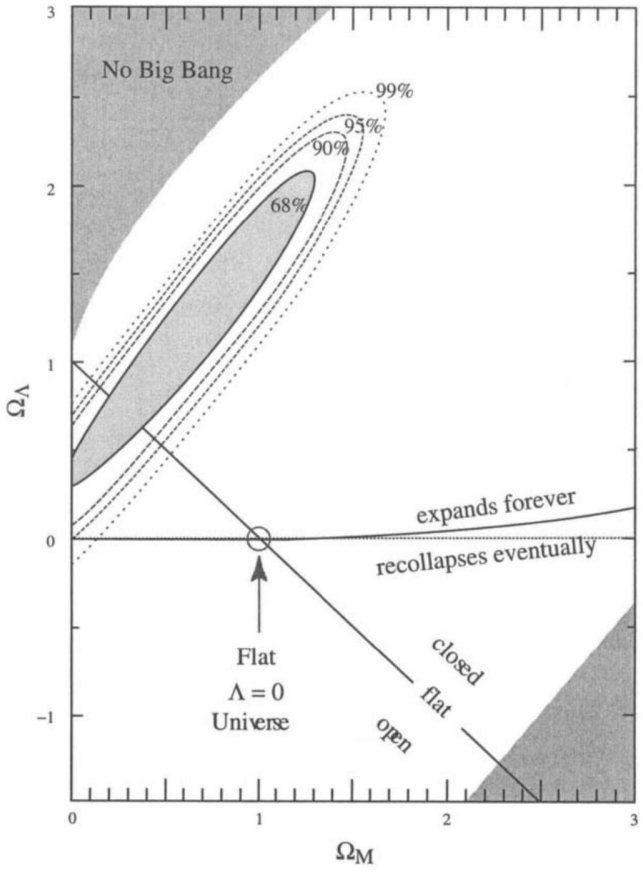


Figure 2.7. Constraints from Type la supernovae on the parameters $(\Omega_m, \Omega_\Lambda)$ (Perlmutter et al., 1999). Flat, matter-dominated universe, the dot with $\Omega_m = 1, \Omega_\Lambda = 0$, is ruled out with high confidence. The line extending from upper left to lower right corresponds to a flat universe.

Although one or the other of these has been argued for, seldom have cosmologists suggested that the universe contains both. Thus, except for the "flat" line and the $\Omega_\Lambda = 0$ line, most of the region in Figure 2.7 is, at least aesthetically, unappealing: the figure allows for too much freedom. On the other hand, the only form of dark energy budgeted for is the cosmological constant. To open up other possibilities,

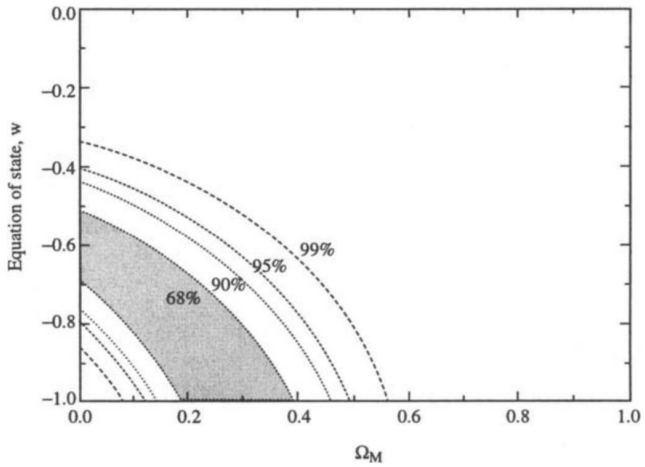


Figure 2.8. Constraints in a flat universe from Type la supernovae on the matter density (Ω_m) and equation of state of the dark energy (w) (Perlmutter et al., 1999). Cosmological constant corresponds to $w = -1$, matter to $w = 0$.

consider Eq. (2.55) as applied to the cosmological constant. The only way for this equation to be satisfied with constant energy density is if the pressure is equal to $-\rho$. One might imagine energy with a slightly different pressure and therefore energy evolution. Define

$$w := \frac{\mathcal{P}}{\rho}$$

A cosmological constant corresponds to $w = -1$, matter to $w = 0$, and radiation to $w = 1/3$. With this new freedom, let's see what the supernova data imply for the equation of state of the dark energy if we fix the universe to be flat. Figure 2.8 shows that values of w greater than ~ -0.5 are disfavored; a cosmological constant is consistent with the data, but it is by no means the only possibility. Equation (2.55) can be integrated to find the evolution of the dark energy,

$$\rho_{\rm de} \propto \exp\left\{-3\int^{a}\frac{da'}{a'}\left[1+w\left(a'\right)\right]\right\}$$

Note that - if w is constant - this expression agrees with our knowledge of the cases explicated above $w = 1/3, 0, -1$.

Epoch of Matter-Radiation Equality

The epoch at which the energy density in matter equals that in radiation is called matter-radiation equality. It has a special significance for the generation of large- scale structure and for the development of CMB anisotropies, because perturbations grow at different rates in the two different eras. It is therefore a useful exercise to calculate the epoch of matter-radiation equality. To do this, we need to compute the energy density of both matter and radiation, and then find the value of the scale factor at which they were equal.

Using Eqs. (2.70) and (2.78), we see that the total energy density in radiation is

$$\frac{\rho_r}{\rho_{\rm cr}} = \frac{4.15\times10^{-5}}{h^2a^4} := \frac{\Omega_r}{a^4}.$$

To calculate the epoch of matter-radiation equality, we equate Equations (2.86) and (2.72) to find

$$a_{\rm eq} = \frac{4.15\times10^{-5}}{\Omega_m h^2}.$$

A different way to express this epoch is in terms of redshift z; the redshift of equality is

$$1 + z_{\rm eq} = 2.4\times10^4\Omega_m h^2.$$

Note that - obviously - as the amount of matter in the universe today, $\Omega_m h^2$, goes up, the redshift of equality also goes up. For our purposes it will be very important that the redshift of equality is at least several times larger than the redshift when photons decouple from matter, $z_* \simeq 10^3$. Thus, we expect photons to decouple when the universe is already well into the matter-dominated era.

SUGGESTED READING

My favorite book on general relativity at this level is A First Course in General Relativity (Schutz), which gives many simple examples to introduce the seemingly profound ideas of general relativity. Also very good are Flat and Curved Spacetimes (Ellis and Williams) and

Essential Relativity (Rindler). Slightly more advanced is Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity (Weinberg), which also has a nice discussion of the early universe in Chapter 15. Two more advanced books are General Relativity (Wald) and the classic Gravitation (Misner, Thorne, and Wheeler). Some of the thermodynamics and statistical mechanics introduced in this chapter is presented in The Early Universe (Kolb and Turner). The distance formulae of Section 2.2 are covered in all standard texts. Neutrinos and their relation to cosmology are covered in the standard texts as well, but there are also several other good books focused solely on neutrinos and astrophysics, Neutrino Astrophysics by the pioneer of the field, Bahcall, and Massive Neutrinos in Physics and Astrophysics by Mohapatra and Pal.

A number of papers treat the topics in this chapter at an accessible level. An especially coherent review of all the different distance measures is given by Hogg (1999). Fukugita, Hogan, and Peebles (1999) do the baryon inventory outlined in Section 2.4.2. Since the supernova discoveries in the late 1990s, many popular articles have appeared attempting to explain the dark energy. The two seminal articles though - Perlmutter et al. (1999) and Riess et al. (1998) - are extremely clear and well worth reading.

D-1.

Convert the following quantities by inserting the appropriate factors of $c, \hbar$, and k_B:

- $T_0 = 2.725$ K $\rightarrow$ eV
- $\rho_\gamma = \pi^2 T_0^4/15 \rightarrow \mathrm{eV}^4$ and gcm^{-3}
- $1/H_0 \rightarrow$ cm
- $m_{\rm Pl} := 1.2 \times 10^{19}\mathrm{GeV} \rightarrow \mathrm{K}, \mathrm{cm}^{-1}, \mathrm{sec}^{-1}$

D-2.

Show that the geodesic equation gets the correct equations of motion for a particle traveling freely in two dimensions using polar coordinates. You can get the Christoffel symbols one of two ways (or both!) and then proceed to (b).

(a) Get the Christoffel symbol either directly from the term in brackets in Eq. (2.17) or from the 2D metric

$$g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}$$

using Eq. (2.19). Show that the only nonzero Christoffel symbols are

$$\Gamma^2_{12} = \Gamma^2_{21} = \frac{1}{r} \quad ; \quad \Gamma^1_{22} = -r$$

with 1,2 corresponding to r, θ. (b) Write down the two components of the geodesic equation using these Christoffel symbols. Show that these give the proper equations of motion for a particle traveling in a plane.

D-3.

The metric for a particle traveling in the presence of gravitational field is $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ where $h_{00} = -2\phi$ where ϕ is the Newtonian gravitational potential; $h_{i0} = 0$; and $h_{ij} = -2\phi\delta_{ij}$. Find the equation of motion for a massive particle traveling in this field.

(a) Show that $\Gamma^0_{00} = \partial\phi/\partial t$ and $\Gamma^i{}_{00} = \delta^{ij}\partial\phi/\partial x^j$.

(b) Show that the time component of the geodesic equation implies that energy $p^0 + m\phi$ is conserved.

(c) Show that the space components of the geodesic equation lead to $d^2x^i/dt^2 = -m\delta^{ij}\partial\phi/\partial x^j$ in agreement with Newtonian theory. Use the fact that the particle is nonrelativistic so $p^0 \gg p^i$.

D-4.

Find how the energy of a massive, nonrelativistic particle changes as the universe expands. Recall that in the massless case we used the fact that $g_{\mu\nu}P^\mu P^\nu = 0$. In this case, it is equal not to zero, but to $-m^2$.

D-5.

Fill in some of the blanks left in our derivation of the Einstein equations.

(a) Compute the Christoffel symbol $\Gamma^i_{\alpha\beta}$ for a flat FRW metric.

(b) Compute the spatial components of the Ricci tensor in a flat FRW universe, R_{ij}. Show that the space-time component, R_{i0}, vanishes.

D-6.

Show that the space-space component of the Einstein equations in a flat universe is

$$\frac{d^2a/dt^2}{a} + \frac{1}{2}\left(\frac{da/dt}{a}\right)^2 = -4\pi G\mathcal{P}$$

where $\mathcal{P}$ is the pressure, the T^i_i (no sum over i) component of the energymomentum tensor.

D-7.

Find and apply the metric, Christoffel symbols, and Ricci scalar for a particle trapped on the surface of a sphere with radius r.

(a) Using coordinates t, θ, ϕ, the metric is

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2\theta \end{pmatrix}.$$

Show that the only nonvanishing Christoffel symbols are $\Gamma^\theta_{\phi\phi}, \Gamma^\phi_{\phi\theta}$, and $\Gamma^\phi_{\theta\phi}$. Express these in terms of θ.

(b) Use these and the geodesic equation to find the equations of motion for the particle. (c) Find the Ricci tensor. Show that contraction of this tensor leads to

$$\mathcal{R} := g^{\mu\nu}R_{\mu\nu} = \frac{2}{r^2}.$$

D-8.

Apply the Einstein equations to the case of an open universe. The interval in an open universe is

$$ds^2 = -dt^2 + a^2(t)\left\{\frac{dr^2}{1+\Omega_k H_0^2 r^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right\}$$

where r, θ, ϕ are the standard 3D spherical coordinates, and Ω_k is the curvature density.

(a) First compute the Christoffel symbols. Show that the only nonzero ones are equal to

$$\Gamma^i_{0j} = H\delta^i_j \quad \Gamma^0_{ij} = g_{ij}H$$
$$\Gamma^i_{jk} = \frac{g^{il}}{2}\left[g_{lj,k} + g_{lk,j} - g_{jk,l}\right].$$

(b) Show that the components of the Ricci tensor are

$$R_{00} = -3\frac{\ddot{a}}{a}$$
$$R_{ij} = g_{ij}\left[\frac{\ddot{a}}{a} + 2H^2 - \frac{2\Omega_k H_0^2}{a^2}\right].$$

(c) From these, compute the Ricci scalar, and then derive the time-time component of Einstein equations.

D-9.

Show that the geodesic equation we derived in a flat universe implies that

$$\frac{d^2\vec{x}}{d\eta^2} = 0$$

where η is the conformal time.

D-10.

Assume that there is only matter and radiation in the universe (no cosmological constant) and that the universe is flat ($\rho_0 = \rho_{\rm cr}$). Integrate Eq. (1.2) to determine the times when the cosmic temperature was 0.1MeV and 1/4eV.

D-11.

Derive some simple expressions for the conformal time η as a function of a.

(a) Show that $\eta \propto a^{1/2}$ in a matter dominated universe and a in one dominated by radiation.
(b) Consider a universe with only matter and radiation, with equality at $a_{\rm eq}$. Show that

$$\eta = \frac{2}{\sqrt{\Omega_m H_0^2}}\left[\sqrt{a + a_{\rm eq}} - \sqrt{a_{\rm eq}}\right]$$

What is the conformal time today? At decoupling?

D-12.

Consider a galaxy of physical (visible) size 5kpc. What angle would this galaxy subtend if situated at redshift 0.1? Redshift 1? Do the calculation in a flat universe, first matter-dominated and then with 30% matter and 70% cosmological constant.

D-13.

How is the energy density of a gas of photons with a blackbody spectrum related to the specific intensity of the radiation? That is, what is the relation between ρ_γ and I_ν defined in Eq. (1.8)?

D-14.

(a) Compute the pressure of a relativistic species in equilibrium with temperature T. Show that $\mathcal{P} = \rho/3$ for both Fermi-Dirac and Bose-Einstein statistics.

(b) Suppose the distribution function depends only on E/T as it does in equilibrium. Find $d\mathcal{P}/dT$. A simple way to do this is to rewrite df/dT in the integral as $-(E/T)df/dE$ and then integrate Eq. (2.62) by parts.

D-15.

The general relativistic expression for the energy-momentum tensor in terms of the distribution functions is given by

$$T^{\mu}{}_{\nu}(\vec{x},t)|_{\text{species i}} = g_i \int \frac{dP_1 dP_2 dP_3}{(2\pi)^3} \left(-\det\left[g_{\alpha\beta}\right]\right)^{-1/2} \frac{P^{\mu}P_{\nu}}{P^0} f_i(\vec{x},\vec{p},t)$$

where P_μ was defined in Eq. (2.24), g_i is the number of spin states for species i, and $\det[g_{\mu\nu}]$ is the determinant of the 4D matrix $g_{\mu\nu}$. Eliminate the comoving momenta P_μ in favor of the magnitude of the proper momentum defined via

$$p^2 := g^{ij}P_iP_j$$

and the direction vector $\hat{p}$. Note that while the comoving momenta P_i remain constant as the universe expands, p falls off as a^{-1}. Show that the time-time component of Eq. (2.101) agrees with the expression for the energy density given in Eq. (2.59). Use the fact that $P^2 := g_{\mu\nu}P^\mu P^\nu = -m^2$ for a particle of mass m.

D-16.

Plot $m - M$ as a function of redshift for a flat, matter-dominated universe (this can be done analytically) and for a flat universe with $\Omega_\Lambda = 0.7, \Omega_m = 0.3$ (for this you need to evaluate numerically a 1D integral). Neglect the K correction. Compare with Figure 1.7.

D-17.

Consider the entropy density, s, defined in Eq. (2.66). For a massless particle, you showed in Exercise 14 that $\mathcal{P} = \rho/3$, so $s = 4\rho/3T$. Express s as a function of T for both bosons and fermions (assumed massless) in equilibrium with zero chemical potential. Show that the entropy density for a massive particle in equilibrium ($T \ll m; \mu = 0$) is exponentially small.

D-18.

Show that the number density of one generation of neutrinos and anti-neutrinos in the universe today is

$$n_\nu = \frac{3}{11} n_\gamma = 112 \text{ cm}^{-3}$$

For this calculation, you will also have to compute the photon number density; both can be expressed in terms of Riemann zeta functions (Eq. (C.27)). Using this result, verify Eq. (2.80)

D-19.

We computed the epoch of equality in the event that all three neutrinos are massless. Suppose instead that two are massless, but the third has mass $m = 0.1$eV. What is a_{eq} is this case?

12.3 Beyond Equilibrium

The very early universe was hot and dense. As a result, interactions among particles occurred much more frequently than they do today. As an example, a photon today can travel across the observable universe without deflection or capture, so it has a mean free path greater than 10^{28} cm. When the age of the universe was equal to 1 sec, though, the mean free path of a photon was about the size of an atom. Thus in the time it took the universe to expand by a factor of 2, a given photon interacted many, many times. These multiple interactions kept the consitituents in the universe in equilibrium in most cases. Nonetheless, there were times when reactions could not proceed rapidly enough to maintain equilibrium conditions. These times are -perhaps not coincidentally - of the utmost interest to cosmologists today.

Indeed, we will see in this chapter that out-of-equilibrium phenomena played a role in (i) the formation of the light elements during Big Bang nucleosynthesis; (ii) recombination of electrons and protons into neutral hydrogen when the temperature was of order 1/4eV; and quite possibly in (iii) production of dark matter in the early universe. Each of these three periods, for obvious reasons, is the subject of intense study by many different groups. Often these studies are carried out in parallel, without the link among the three mentioned. I think it is important to understand that all three phenomena are the result of nonequilibrium physics and that all three can be studied with the same formalism: the Boltzmann equation. Section 3.1 introduces the Boltzmann equation and some approximations to it that are common to all three processes. The remaining three sections of the chapter are simply applications of this general formula.

Beyond the intrinsic importance of these nonequilibrium phenomena, this chapter also serves as a bridge between the smooth, homogeneous universe described in Chapter 2 and the inhomogeneous perturbations we will explore in the rest of the book. One way to think of this transition is in terms of phase-space distributions f of Eqs. (2.60) and (2.61). Until now, we have assumed that the chemical potentials are zero and temperatures uniform. In this chapter, we will have to abandon the idea of trivial chemical potentials in order to track abundances of particles losing contact with the plasma. In succeeding chapters, we will move beyond uniformity and explore temperatures which depend on both position and direction of propagation.

12.3.1 Boltzmann Equation for Annihilation

The Boltzmann equation formalizes the statement that the rate of change in the abundance of a given particle is the difference between the rates for producing and eliminating that species. Suppose that we are interested in the number density n_1 of species 1. For simplicity, let's suppose that the only process affecting the abundance of this species is an annihilation with species 2 producing two particles, imaginatively called 3 and 4. Schematically, $1 + 2 \leftrightarrow 3 + 4$; i.e., particle 1 and particle 2 can annihilate producing particles 3 and 4, or the inverse process can produce 1 and 2. The Boltzmann equation for this system in an expanding universe is

$$\begin{aligned} a^{-3}\frac{d\left(n_1 a^3\right)}{dt} = & \int \frac{d^3p_1}{(2\pi)^3 2E_1}\int \frac{d^3p_2}{(2\pi)^3 2E_2}\int \frac{d^3p_3}{(2\pi)^3 2E_3}\int \frac{d^3p_4}{(2\pi)^3 2E_4} \\ & \times (2\pi)^4\delta^3\left(p_1+p_2-p_3-p_4\right)\delta\left(E_1+E_2-E_3-E_4\right)|\mathcal{M}|^2 \\ & \times \left\{f_3 f_4\left[1\pm f_1\right]\left[1\pm f_2\right]-f_1 f_2\left[1\pm f_3\right]\left[1\pm f_4\right]\right\} \end{aligned}$$

In the absence of interactions, the left-hand side of Eq. (3.1) says that the density times the scale factor cubed is conserved. This reflects the nature of the expanding universe: as the comoving grid expands, the volume of a region containing a fixed number of particles grows as a^3. Therefore, the physical number density of these particles falls off as a^{-3}. Interactions are included in the right-hand side of the Boltzmann equation. Let's consider the interaction

term starting from the last line and moving up. Putting aside the $1 \pm f$ terms on the last line, we see that the rate of producing species 1 is proportional to the occupation numbers of species 3 and 4, f_3 and f_4. Similarly the loss term is proportional to $f_1 f_2$. The $1 \pm f$ terms, with plus sign for bosons such as photons and minus sign for fermions such as electrons, represent the phenomena of Bose enhancement and Pauli blocking. If particles of type 1 already exist, a reaction producing more such particles is more likely to occur if 1 is a boson and less likely if a fermion. I have suppressed the momentum dependence of f, but of course all the occupation numbers depend on the corresponding momentum (e.g., $f_1 = f_1(p_1)$). Moving upward, the Dirac delta functions on the second line in Eq. (3.1) enforce energy and momentum conservation; the factors of 2π are the result of moving from discrete Kronecker δ 's to the continuous Dirac version. The energies here are related to the momenta via $E = \sqrt{p^2 + m^2}$. The amplitude on the second line $\mathcal{M}$ is determined from the fundamental physics in question. For example, if we were interested in the Compton scattering of electrons off of photons, $\mathcal{M}$ would be proportional to α, the fine structure constant. In almost all cases of interest, this amplitude is reversible, identical for $1 + 2 \to 3 + 4$ and $3 + 4 \to 1 + 2$. Indeed, reversibility has been assumed in Eq. (3.1). The last two lines of Eq. (3.1) depend on the momenta of the particles involved. To find the total number of interactions, we must sum over all momenta. The integrals on the first line do precisely that. As in Figure 2.4, the factors of $(2\pi)^3$ [really $(2\pi\hbar)^3$] represent the volume of one unit of phase space; we want to sum over all such units. Finally, the factors of $2E$ in the denominator arise because, relativistically, the phase space integrals should really be four-dimensional, over the three components of momentum and one of energy. However, these are constrained to lie on the 3 -sphere fixed by $E^2 = p^2 + m^2$. In equations,

$$\int d^3p \int_0^\infty dE \delta\left(E^2 - p^2 - m^2\right) = \int d^3p \int_0^\infty dE \frac{\delta\left(E - \sqrt{p^2 + m^2}\right)}{2E}$$

Performing the integral over E with the delta function yields the factor of $2E$.

Equation (3.1) is an integrodifferential equation for the phase space distributions. Further, in principle at least, it must be supplemented with similar equations for the other species. In practice, these formidable obstacles can be overcome for many practical cosmological applications. The first, most important realization is that scattering processes typically enforce kinetic equilibrium. That is, scattering takes place so rapidly that the distributions of the various species take on the generic Bose-Einstein/Fermi-Dirac forms (Eqs. (2.61) and (2.60)). This form condenses all of the uncertainty in the distribution into a single function of time μ. If annihilations were also in equilibrium, μ would be the chemical potential, and the sum of the chemical potentials in any reaction would have to balance. For example, the reaction $e^+ + e^- \leftrightarrow \gamma + \gamma$ would cause $\mu_{e^+} + \mu_{e^-} = 2\mu_\gamma$. In the out-of-equilibrium cases we will study, the system will not be in chemical equilibrium and we will have to solve a differential equation for μ. The great simplifying feature of kinetic equilibrium, though, is that this differential equation will be a single ordinary differential equation, as opposed to the very complicated form of Eq. (3.1).

We will typically be interested in systems at temperatures smaller than $E - \mu$. In this limit, the exponential in the Bose-Einstein or Fermi-Dirac distribution is large and dwarfs the \pm 1 in the denominator. Thus, another simplification emerges: we can ignore the complications of quantum statistics. The distributions become

$$f(E) \to e^{\mu/T} e^{-E/T}$$

and the Pauli blocking/Bose enhancement factors in the Boltzmann equation can be neglected.

Under these approximations, the last line of Eq. (3.1) becomes

$$f_3 f_4 \left[1 \pm f_1\right]\left[1 \pm f_2\right] - f_1 f_2 \left[1 \pm f_3\right]\left[1 \pm f_4\right] \\ \to e^{-(E_1+E_2)/T}\left\{e^{(\mu_3+\mu_4)/T} - e^{(\mu_1+\mu_2)/T}\right\}.$$

Here I have used energy conservation. $E_1 + E_2 = E_3 + E_4$. We will use the number densities themselves as the time-dependent functions to be solved for. instead of μ. The number density of species i is related to μ_i via Table 3.1. Reactions in This Chapter: $1 + 2 \leftrightarrow 3 + 4$

	1	2	3	4
Neutron-Proton Ratio	n	ν_e or e^+	p	e^-or $\bar{\nu}_e$
Recombination	e	p	H	γ
Dark Matter Production	X	X	l	l

$$n_i = g_i e^{\mu_i/T}\int \frac{d^3p}{(2\pi)^3} e^{-E_i/T}$$

where g_i is the degeneracy of the species, e.g., equal to 2 for the two spin states of the photon. It will also be useful to define the species-dependent equilibrium number density as

$$n_i^{(0)} := g_i \int \frac{d^3p}{(2\pi)^3} e^{-E_i/T} = \begin{cases} g_i \left(\frac{m_i T}{2\pi}\right)^{3/2} e^{-m_i/T} & m_i \gg T \\ g_i \frac{T^3}{\pi^2} & m_i \ll T \end{cases}.$$

With this defintion, $e^{\mu_i/T}$ can be rewritten as $n_i/n_i^{(0)}$, so the last line of Eq. (3.1) is equal to

$$e^{-(E_1+E_2)/T}\left\{\frac{n_3 n_4}{n_3^{(0)} n_4^{(0)}} - \frac{n_1 n_2}{n_1^{(0)} n_2^{(0)}}\right\}$$

With these approximations the Boltzmann equation now simplifies enormously. Define the thermally averaged cross section as

$$\langle\sigma v\rangle := \frac{1}{n_1^{(0)} n_2^{(0)}}\int \frac{d^3p_1}{(2\pi)^3 2E_1}\int \frac{d^3p_2}{(2\pi)^3 2E_2}\int \frac{d^3p_3}{(2\pi)^3 2E_3}\int \frac{d^3p_4}{(2\pi)^3 2E_4} e^{-(E_1+E_2)/T} \\ \times (2\pi)^4 \delta^3\left(p_1 + p_2 - p_3 - p_4\right)\delta\left(E_1 + E_2 - E_3 - E_4\right)|\mathcal{M}|^2$$

Then, the Boltzmann equation becomes

$$a^{-3}\frac{d\left(n_1 a^3\right)}{dt} = n_1^{(0)} n_2^{(0)}\langle\sigma v\rangle\left\{\frac{n_3 n_4}{n_3^{(0)} n_4^{(0)}} - \frac{n_1 n_2}{n_1^{(0)} n_2^{(0)}}\right\}$$

We thus have a simple ordinary differential equation for the number density. Although the details will vary from application to application (see Table 3.1), we will in the remainder of the chapter start from this equation when tracking abundances.

One qualitative note about Eq. (3.9). The left-hand side is of order n_1/t, or, since the typical cosmological time is $H^{-1}, n_1 H$. The right-hand side is of order $n_1 n_2\langle\sigma v\rangle$. Therefore, if the reaction rate $n_2\langle\sigma v\rangle$ is much larger than the expansion rate, then the terms on the right side will be much larger than the one on the left. The only way to maintain equality then is for the individual terms on the right to cancel. Thus, when reaction rates are large,

$$\frac{n_3 n_4}{n_3^{(0)} n_4^{(0)}} = \frac{n_1 n_2}{n_1^{(0)} n_2^{(0)}}$$

This equation, which follows virtually by inspection from the Boltzmann equation, goes under different names in different venues. The particle physics community, which first studied the production of heavy relics in the early universe, tends to call it chemical equilibrium. In the context of Big Bang nucleosynthesis, it is called nuclear statistical equilibrium (NSE), while students of recombination, the process of electrons and protons combining to form neutral hydrogen, use the terminology Saha equation.

12.3.2 Big Bang Nucleosynthesis

As the temperature of the universe cools to 1MeV, the cosmic plasma consists of:

- Relativistic particles in equilibrium: photons, electrons and positrons. These are kept in close contact with each other by electromagnetic interactions such as $e^+e^- \leftrightarrow \gamma\gamma$. Besides a small difference due to fermion/boson statistics, these all have the same abundances.
- Decoupled relativistic particles: neutrinos. At temperatures a little above 1 MeV, the rate for processes such as $\nu e \leftrightarrow \nu e$ which keep neutrinos coupled to the rest of the plasma drops beneath the expansion rate. Neutrinos therefore share the same temperature as the other relativistic particles, and hence are roughly as abundant, but they do not couple to them.
- Nonrelativistic particles: baryons. If there had been no asymmetry in the initial number of baryons and anti-baryons, then both would be completely depleted by 1MeV. However, such an asymmetry did exist: $(n_b - n_{\bar{b}})/s \sim 10^{-10}$ initially, [1] and this ratio remains constant throughout the expansion. By the time the temperature is of order 1MeV, all anti-baryons have annihilated away (Exercise 12) so

$$\eta_b := \frac{n_b}{n_\gamma} = 5.5 \times 10^{-10} \left(\frac{\Omega_b h^2}{0.020}\right)$$

There are thus many fewer baryons than relativistic particles when $T \sim$ MeV.

Our task in this section will be to determine how the baryons end up. Were the system to remain in equilibrium throughout, the final state would be dictated solely by energetics, and all baryons would relax to the nuclear state with the lowest energy per baryon, iron (Figure 3.1). However, nuclear reactions, which scale as the second - or higher - power of the density, are too slow to keep the system in equilibrium as the temperature drops. So, in principle, we need to solve the

[1] s is the entropy density which scales as a^{-3}, as we saw in Chapter 2.

equivalent of Eq. (3.9) for all the nuclei, i.e., a set of coupled differential equations. In practice, at least for a qualitative understanding of the result, we can make use of two simplifications that obviate the need to solve the full set of differential equations.

Lightning Introduction to Nuclear Physics

A single proton is a hydrogen nucleus, referred to as ^{1}H or simply p; a proton and a neutron make up deuterium, ^{2}H or D; one proton and two neutrons make tritium, ^{3}H or T. Nuclei with two protons are helium; these can have one neutron (^{3}He) or two (^{4}He). Thus unique elements have a fixed number of protons, and isotopes of a given element have differing numbers of neutrons. The total number of neutrons and protons in the nucleus, the atomic number, is a superscript before the name of the element.

The total mass of a nucleus with Z protons and $A - Z$ neutrons differs slightly from the mass of the individual protons and neutrons alone. This difference is called the binding energy, defined as

$$B := Zm_p + (A - Z)m_n - m$$

where m is the mass of the nucleus. For example, the mass of deuterium is 1875.62MeV while the sum of the neutron and proton masses is 1877.84MeV, so the binding energy of deuterium is 2.22MeV. Nuclear binding energies are typically in the MeV range, which explains why Big Bang nucleosynthesis occurs at temperatures a bit less than 1MeV even though nuclear masses are in the GeV range.

Neutrons and protons can interconvert via weak interactions:

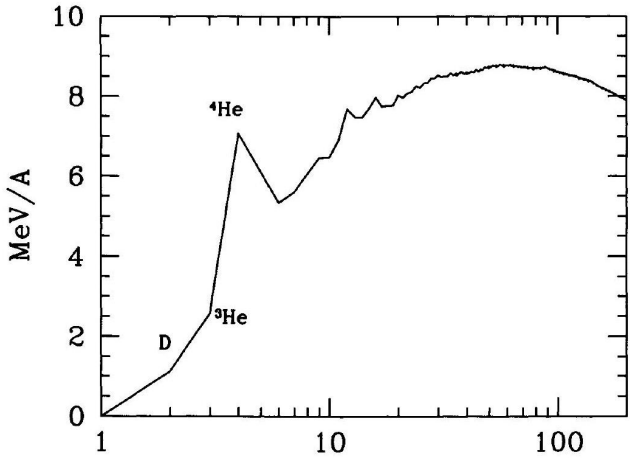
$$p + \bar{\nu} \leftrightarrow n + e^+ \quad ; \quad p + e^- \leftrightarrow n + \nu \quad ; \quad n \leftrightarrow p + e^- + \bar{\nu}$$

where all the reactions can proceed in either direction. The light elements are built up via electromagnetic interactions. For example, deuterium forms from $p + n \rightarrow \mathrm{D} + \gamma$. Then, $\mathrm{D} + \mathrm{D} \rightarrow n + {}^3\mathrm{He}$, after which ${}^3\mathrm{He} + \mathrm{D} \rightarrow p + {}^4\mathrm{He}$ produces ${}^4\mathrm{He}$.

The first simplification is that essentially no elements heavier than helium are produced at appreciable levels. [2] So the only nuclei that need to be traced are hydrogen and helium, and their isotopes: deuterium, tritium, and ${}^3\mathrm{He}$. The second simplification is that, even in the context of this reduced set of elements, the physics splits up neatly into two parts since above $T \simeq 0.1\mathrm{MeV}$, no light nuclei form: only free protons and neutrons exist. Therefore, we first solve for the neutron/proton ratio and then use this abundance as input for the synthesis of helium and isotopes such as deuterium.

[2] An exception is lithium, produced at a part in $10^9 - 10^{10}$, and this trace abundance may be observable today. See, e.g., Pinsonneault et al. (2001).

Binding Energy per Nucleon



A

Figure 3.1. Binding energy of nuclei as a function of mass number. Iron has the highest binding energy, but among the light elements, ${}^4\mathrm{He}$ is a crucial local maximum. Nucleosynthesis in the early universe essentially stops at ${}^4\mathrm{He}$ because of the lack of tightly bound isotopes at $A = 5 - 8$. In the high-density environment of stars, three ${}^4\mathrm{He}$ nuclei fuse to form ${}^{12}\mathrm{C}$, but the low baryon number precludes this process in the early universe.

Both of these simplifications - no heavy elements at all and only n/p above 0.01 MeV - rely on the physical fact that, at high temperatures, comparable to nuclear binding energies, any time a nucleus is produced in a reaction, it is destroyed by a high-energy photon. This

fact is reflected in the fundamental equilibrium equation (3.10). To see how, let's consider this equation applied to deuterium production, $n + p \leftrightarrow \mathrm{D} + \gamma$. Since photons have $n_\gamma = n_\gamma^{(0)}$, the equilibrium condition becomes

$$\frac{n_D}{n_n n_p} = \frac{n_D^{(0)}}{n_n^{(0)} n_p^{(0)}}$$

The integrals on the right, as given in Eq. (3.6), lead to

$$\frac{n_D}{n_n n_p} = \frac{3}{4}\left(\frac{2\pi m_D}{m_n m_p T}\right)^{3/2} e^{[m_n+m_p-m_D]/T},$$

the factor of 3/4 being due to the number of spin states)3 for D and 2 each for p and n). In the prefactor, m_D can be set to $2m_n = 2m_p$, but in the exponential the small difference between $m_n + m_p$ and m_D is important: indeed the argument of the exponential is by defintion equal to the binding energy of deuterium, $B_D = 2.22$ MeV. Therefore, as long as equilibrium holds,

$$\frac{n_D}{n_n n_p} = \frac{3}{4}\left(\frac{4\pi}{m_p T}\right)^{3/2} e^{B_D/T}$$

Both the neutron and proton density are proportional to the baryon density, so roughly,

$$\frac{n_D}{n_b} \sim \eta_b \left(\frac{T}{m_p}\right)^{3/2} e^{B_D/T}$$

As long as B_D/T is not too large, the prefactor dominates this expression. And the prefactor is very small because of the smallness of the baryon-to-photon ratio, Eq. (3.11).

The small baryon to photon ratio thus inhibits nuclei production until the temperature drops well beneath the nuclear binding energy. At temperatures above 0.1 MeV, then, virtually all baryons are in the form of neutrons and protons. Around this time, deuterium and helium are produced, but the reaction rates are by now too low to produce any heavier elements. We could have anticipated this by considering Figure 3.1. The lack of a stable isotope with mass number 5 implies that heavier elements cannot be produced via ${}^4\mathrm{He} + p \to X$. In stars, the triple alpha process ${}^4\mathrm{He} + {}^4\mathrm{He} + {}^4\mathrm{He} \to {}^{12}\mathrm{C}$ produces heavier elements, but in the early universe, densities are far too low to allow three nuclei to find one another on relevant time scales.

Neutron Abundance

We begin by solving for the neutron-proton ratio. Protons can be converted into neutrons via weak interactions, $p + e^- \to n + \nu_e$ for example. As we will see, reactions of this sort keep neutrons and protons in equilibrium until $T \sim$ MeV. Thereafter, one must solve the rate equation (3.9) to track the neutron abundance.

From Eq. (3.6), the proton/neutron equilibrium ratio in the nonrelativistic limit ($E = m + p^2/2m$) is

$$\frac{n_p^{(0)}}{n_n^{(0)}} = \frac{e^{-m_p/T}\int dp p^2 e^{-p^2/2m_p T}}{e^{-m_n/T}\int dp p^2 e^{-p^2/2m_n T}}$$

The integrals here are proportional to $m^{3/2}$, but the resulting ratio $(m_p/m_n)^{3/2}$ is sufficiently close to unity that we can neglect the mass difference. However, in the exponential the mass difference is very important, and we are left with

$$\frac{n_p^{(0)}}{n_n^{(0)}} = e^{\mathcal{Q}/T}$$

with $\mathcal{Q} := m_n - m_p = 1.293\text{MeV}$. Therefore, at high temperatures, there are as many neutrons as protons. As the temperature drops beneath 1MeV, the neutron fraction goes down. If weak interactions operated efficiently enough to maintain equilibrium indefinitely, then it would drop to zero. The main task of this section is to find out what happens in the real world where weak interactions are not so efficient.

It is convenient to define

$$X_n := \frac{n_n}{n_n + n_p}$$

that is, X_n is the ratio of neutrons to total nuclei. In equilibrium,

$$X_n \to X_{n,\text{EQ}} := \frac{1}{1 + \left(n_p^{(0)}/n_n^{(0)}\right)}$$

To track the evolution of X_n, let's start from Eq. (3.9), with 1 = neutron, 3 = proton, and $2, 4$ = leptons in complete equilibrium $\left(n_l = n_l^{(0)}\right)$. Then,

$$a^{-3}\frac{d\left(n_n a^3\right)}{dt} = n_l^{(0)}\langle\sigma v\rangle\left\{\frac{n_p n_n^{(0)}}{n_p^{(0)}} - n_n\right\}.$$

We have already determined the ratio $n_n^{(0)}/n_p^{(0)} = e^{-\mathcal{Q}/T}$ and we can identify $n_l^{(0)}\langle\sigma v\rangle$ as λ_{np}, the rate for neutron $\to$ proton conversion since it multiplies n_n in the loss term. Also if we rewrite n_n on the left as $(n_n + n_p)\,X_n$, then the total density times a^3 can be taken outside the derivative, leaving

$$\frac{dX_n}{dt} = \lambda_{np}\left\{(1 - X_n)\,e^{-\mathcal{Q}/T} - X_n\right\}$$

Equation (3.23) is a differential equation for X_n as a function of time, but it contains the temperature T and the reaction rate λ_{np}, both of which have complicated time dependences. It is simplest therefore to recast the equation using as the evolution variable

$$x := \frac{\mathcal{Q}}{T}.$$

The left-hand side of Eq. (3.23) then becomes $\dot{x}dX_n/dx$, so we need an expression for $dx/dt = -x\dot{T}/T$. Since $T \propto a^{-1}$,

$$\frac{1}{T}\frac{dT}{dt} = -H = -\sqrt{\frac{8\pi G\rho}{3}},$$

the second equality following from Eq. (2.39). Nucleosynthesis occurs in the radiation-dominated era, so the main contribution to the energy density ρ comes from relativistic particles. Recall from Chapter 2 that the contribution to the energy density from relativistic particles is

$$\rho = \frac{\pi^2}{30}T^4\left[\sum_{i=\text{ bosons}} g_i + \frac{7}{8}\sum_{i=\text{ fermions}} g_i\right] \quad (i \text{ relativistic })$$

$$:= g_*\frac{\pi^2}{30}T^4$$

The effective numbers of relativistic degrees of freedom, g_*, is a function of temperature. At temperatures of order 1MeV, the contributing species are: photons ($g_\gamma = 2$), neutrinos ($g_\nu = 6$), and electrons and positrons ($g_{e^+} = g_{e^-} = 2$). Adding up leads to $g_* \simeq 10.75$, roughly constant throughout the regime of interest. Then, Eq. (3.23) becomes

$$\frac{dX_n}{dx} = \frac{x\lambda_{np}}{H(x=1)} \left\{ e^{-x} - X_n \left(1 + e^{-x}\right) \right\}$$

with

$$H(x=1) = \sqrt{\frac{4\pi^3 G \mathcal{Q}^4}{45}} \times \sqrt{10.75} = 1.13\text{sec}^{-1}.$$

Finally, we need an expression for the neutron-proton conversion rate, λ_{np}. Under the approximations we are using, the rate is (Bernstein 1988 or Exercise 3)

$$\lambda_{np} = \frac{255}{\tau_n x^5} \left(12 + 6x + x^2\right)$$

with the neutron lifetime $\tau_n = 886.7$ sec. Thus, when $T = \mathcal{Q}$ (i.e., when $x = 1$), the conversion rate is 5.5sec^{-1}, somewhat larger than the expansion rate. As the temperature drops beneath 1MeV, though, the rate rapidly falls below the expansion rate, so conversions become inefficient.

We can now integrate Eq. (3.27) numerically to track the neutron abundance (Exercise 4). Figure 3.2 shows the results of this integration. Note that the result agrees extremely well at temperatures above ~ 0.1MeV with the exact solution which includes proper statistics, nonzero electron mass, and changing g_*. The neutron fraction X_n does indeed fall out of equlibrium once the temperature drops below 1MeV: it freezes out at 0.15 once the temperature drops below 0.5MeV.

At temperatures below 0.1MeV, two reactions we have not included yet become important: neutron decay ($n \to p + e^- + \bar{\nu}$) and deuterium production ($n + p \to \text{D} + \gamma$). Decays can be added trivially by adding in a factor of e^{-t/τ_n} to the results of Figure 3.2. By the time decays become important, electrons and positrons have annihilated, so g_* in Eq. (3.26) is 3.36 and the time-temperature relation is (Exercise 5):

$$t = 132\,\text{sec} \left(\frac{0.1\text{MeV}}{T}\right)^2$$

We will see shortly that production of deuterium, and other light elements, begins in earnest at $T \sim 0.07$MeV. By then, decays have depleted the neutron fraction by a factor of $\exp\left[-(132/886.7)(0.1/0.07)^2\right] = 0.74$. So the neutron abundance at the onset of nucleosynthesis is 0.15×0.74, or

$$X_n\left(T_{\text{nuc}}\right) = 0.11$$

We now turn to light element formation to understand the ramifications of this number.

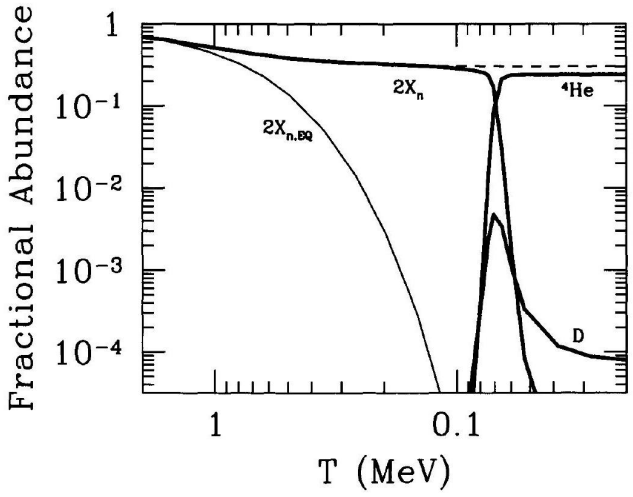


Figure 3.2. Evolution of light element abundances in the early universe. Heavy solid curves are results from Wagoner (1973) code; dashed curve is from integration of Eq. (3.27); light solid curve is twice the neutron equilibrium abundance. Note the good agreement of Eq. (3.27) and the exact result until the onset of neutron decay. Also note that the neutron abundance falls out of equilibrium at $T \sim$ MeV.

Light Element Abundances

A useful way to approximate light element production is that it occurs instantaneously at a temperature $T_{\rm nuc}$ when the energetics compensates for the small baryon to photon ratio. Let's consider deuterium production as an example, with Eq. (3.17) as our guide. The equilibrium deuterium abundance is of order the baryon abundance (i.e. if the universe stayed in equilibrium, all neutrons and protons would form deuterium) when Eq. (3.17) is of order unity, or

$$\ln\left(\eta_b\right) + \frac{3}{2}\ln\left(T_{\rm nuc}\,/m_p\right) \sim -\frac{B_D}{T_{\rm nuc}}.$$

Equation (3.32) suggests that deuterium production takes place at $T_{\rm nuc} \sim 0.07$ MeV, with a weak logarithmic dependence on η_b.

Since the binding energy of helium is larger than that of deuterium, the exponential factor $e^{B/T}$ favors helium over deuterium. Indeed, Figure 3.2 illustrates that helium is produced almost immediately after deuterium. Virtually all remaining neutrons at $T \sim T_{\rm nuc}$ then are processed into ^{4} He. Since two neutrons go into ^{4}He, the final ^{4}He abundance is equal to half the neutron abundance at $T_{\rm nuc}$. Often, results are quoted in terms of mass fraction; then,

$$X_4 := \frac{4n_4He}{n_b} = 2X_n\left(T_{\rm nuc}\right)$$

Figure 3.2 shows that this relation holds. Indeed, to find the final helium mass fraction, we need only take the neutron fraction at $T_{\rm nuc}$, Eq. (3.31), and multiply by 2, so the final helium mass fraction is 0.22. This rough estimate, obtained by solving a single differential equation, is in remarkable agreement with the exact solution, which can be fit via (Olive, 2000; Kolb and Turner, 1990)

$$Y_p = 0.2262 + 0.0135\ln\left(\eta_b/10^{-10}\right)$$

One important feature of this result is that it depends only logarithmically on the baryon fraction. We saw in Eq. (3.32) that $T_{\rm nuc}$ has this dependence. You might think that the

exponential sensitivity to $T_{\rm nuc}$ in the decay fraction would turn this into linear dependence. However, $T_{\rm nuc}$ is sufficiently early that only a small fraction of neutrons have decayed: the exponential in this regime is linear in the time. Therefore, the final helium abundance maintains only logarithmic dependence on the baryon density.

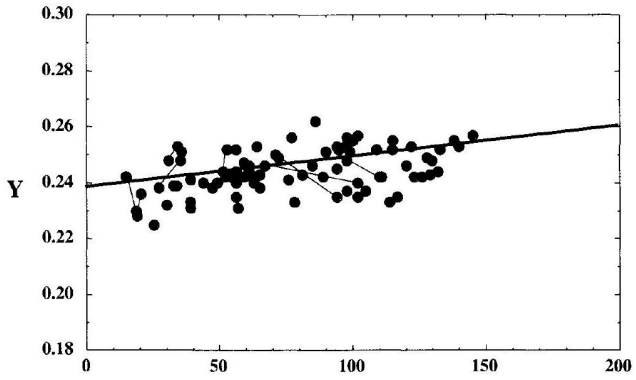


Figure 3.3. Helium abundance ($Y := Y_p$) as a function of oxygen/hydrogen ratio. Lower oxygen systems have undergone less processing, so the helium abundance in those systems is closer to primordial. Line, and extrapolation to $Y_p = 0.238$, from Olive (2000). Data from Pagel et al. (1992), Skillman and Kennicutt (1993), Skillman et al. (1994), and Izotov and Thuan (1998). Short lines connect the same region observed by different groups.

The prediction agrees well with the observations, as indicated in Figure 3.3. The best indication of the primordial helium abundance comes from the most unprocessed systems, typically identified by low metallicities. As Figure 3.3 indicates, the primordial abundance almost certainly lies between 0.22 and 0.25. Although there have been claims of discord in the past, the agreement remains one of the pillars of observational cosmology.

Figure 3.2 shows that not all of the deuterium gets processed into helium. A trace amount remains unburned, simply because the reaction which eliminates it ($\mathrm{D} + p \to {}^3\mathrm{He} + \gamma$) is not completely efficient. Figure 3.2 shows that after $T_{\rm nuc}$, deuterium is depleted via these reactions, eventually freezing out at a level of order $10^{-5} - 10^{-4}$. If the baryon density is low, then the reactions proceed more slowly, and the depletion is not as effective. Therefore, low baryon density inevitably results in more deuterium; the sensitivity is quite stark, as illustrated in Figure 1.8. As a result, deuterium is a powerful probe of the baryon density. Complementing this sensitivity is the possiblity of measuring deuterium in gas clouds as $z \sim 3$ by looking for absorbtion in the spectra of distant QSOs. For example, O'Meara et al. (2001) combine the measurements of primordial deuterium in four systems to obtain, $\mathrm{D/H} = 3.0 \pm 0.4 \times 10^{-5}$, corresponding to $\Omega_b h^2 = 0.0205 \pm 0.0018$.

12.3.3 Recombination

As the temperature drops to $\sim$ 1eV, photons remain tightly coupled to electrons via Compton scattering and electrons to protons via Coulomb scattering. It will come as no surprise that at these temperatures, there is very little neutral hydrogen. Energetics of course favors the production of neutral hydrogen with a binding energy of $\epsilon_0 = 13.6$eV, but the high photon/baryon ratio ensures that any hydrogen atom produced will be instantaneously ionized. This phenomenon is identical to the delay in the production of light nuclei we saw above, replayed on the atomic scale.

As long as the reaction ${}^3e^- + p \leftrightarrow \mathrm{H} + \gamma$ remains in equilibrium, the condition in Eq. (3.10) (with $1 = e, 2 = p, 3 = \mathrm{H}$) ensures that

$$\frac{n_e n_p}{n_H} = \frac{n_e^{(0)} n_p^{(0)}}{n_H^{(0)}}$$

We can go further here by recognizing that the neutrality of the universe ensures that $n_e = n_p$. Let's define the free electron fraction

$$X_e := \frac{n_e}{n_e + n_H} = \frac{n_p}{n_p + n_H}$$

the denominator equal to the total number of hydrogen nuclei. Carrying out the integrals on the right of Eq. (3.35) leads then to

$$\frac{X_e^2}{1 - X_e} = \frac{1}{n_e + n_H} \left[\left(\frac{m_e T}{2\pi} \right)^{3/2} e^{-[m_e + m_p - m_H]/T} \right]$$

where we have made the familiar approximation of neglecting the small mass difference of H and p in the prefactor. The argument of the exponential is $-\epsilon_0/T$.

[3] Here p stands for free protons and H for neutral hydrogen, i.e., a proton with an electron attached. Neglecting the relatively small number of helium atoms, the denominator $n_e + n_H$ (or $n_p + n_H$) is equal to the baryon density, $\eta_b n_\gamma \sim 10^{-9} T^3$. So when the temperature is of order ϵ_0, the right-hand side is of order $10^9 \, (m_e/T)^{3/2} \simeq 10^{15}$. In that case, Eq. (3.37) can be satisfied only if the denominator on the left is very small, that is if X_e is very close to 1: all hydrogen is ionized. Only when the temperature drops far below ϵ_0 does appreciable recombination take place. As X_e falls, the rate for recombination also falls, so that equilibrium becomes more difficult to maintain. Thus, in order to follow the free electron fraction accurately, we need to solve the Boltzmann equation, just as we did for the neutron-proton ratio.

In this case, Eq. (3.9) for the electron density becomes

$$\begin{aligned} a^{-3} \frac{d \left(n_e a^3 \right)}{dt} &= n_e^{(0)} n_p^{(0)} \langle \sigma v \rangle \left\{ \frac{n_H}{n_H^{(0)}} - \frac{n_e^2}{n_e^{(0)} n_p^{(0)}} \right\} \\ &= n_b \langle \sigma v \rangle \left\{ (1 - X_e) \left(\frac{m_e T}{2\pi} \right)^{3/2} e^{-\epsilon_0/T} - X_e^2 n_b \right\} \end{aligned}$$

where the last line follows since the ratio $n_e^{(0)} n_p^{(0)} / n_H^{(0)}$ is equal to the term in square brackets in Eq. (3.37). Meanwhile, since $n_b a^3$ is constant it can be passed through the derivative on the left after expressing n_e as $n_b X_e$, so that

$$\frac{dX_e}{dt} = \left\{ (1 - X_e) \, \beta - X_e^2 n_b \alpha^{(2)} \right\}$$

where the ionization rate is typically denoted

$$\beta := \langle \sigma v \rangle \left(\frac{m_e T}{2\pi} \right)^{3/2} e^{-\epsilon_0/T}$$

and the recombination rate

$$\alpha^{(2)} := \langle \sigma v \rangle$$

The recombination rate has superscript (2) because recombination to the ground ($n = 1$) state is not relevant. Ground-state recombinations lead to production of an ionizing photon, and this photon immediately ionizes a neutral atom. The net effect of such a recombination is

zero: no new neutral atoms are formed this way. The only way for recombination to proceed is via capture to one of the excited states of hydrogen; to a good approximation, this rate is

$$\alpha^{(2)} = 9.78\frac{\alpha^2}{m_e^2}\left(\frac{\epsilon_0}{T}\right)^{1/2}\ln\left(\frac{\epsilon_0}{T}\right)$$

The Saha approximation, Eq. (3.37), does a good job predicting the redshift of recombination, but fails as the electron fraction drops and the system goes out of equilibirium. Therefore, the detailed evolution of X_e must be obtained by a numerical integration of Eq. (3.39) (Exercise 8). Results from numerical integration of Eq. (3.39) are shown in Figure 3.4.

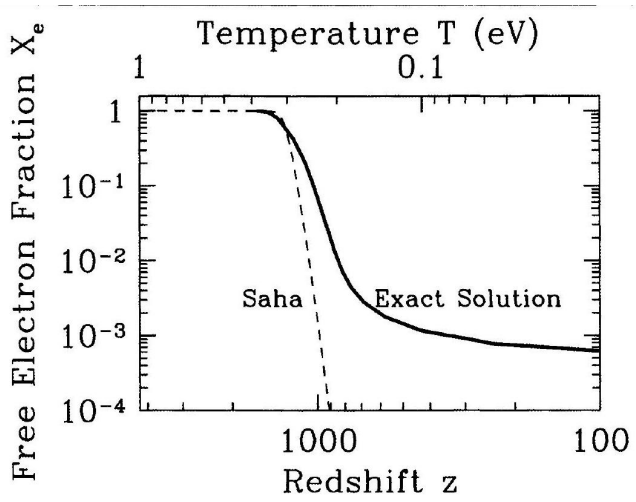


Figure 3.4. Free electron fraction as a function of redshift. Recombination takes place suddenly at $z \sim 1000$ corresponding to $T \sim 1/4$eV. The Saha approximation. Eq. (3.37), holds in equilibrium and correctly identifies the redshift of recombination, but not the detailed evolution of X_e. Here $\Omega_b = 0.06, \Omega_m = 1, h = 0.5$.

The computation of the neutron/proton ratio affects the abundance of light elements today. Similarly, the evolution of the free electron abundance has major ramifications for observational cosmology. Recombination at $z_* \sim 1000$ is directly tied to the decoupling of photons from matter. [4] This decoupling, in turn, directly affects the pattern of anisotropies in the CMB that we observe today.

Decoupling occurs roughly when the rate for photons to Compton scatter off electrons becomes smaller than the expansion rate. [5] The scattering rate is

$$n_e\sigma_T = X_e n_{\rm b}\sigma_T$$

where $\sigma_T = 0.665\times10^{-24}$ cm^2 is the Thomson cross section, and I continue to ignore helium, thereby assuming that the total number of hydrogen nuclei (free protons + hydrogen atoms) is equal to the total baryon number. Since the ratio of the

[4] Notice from Figure 1.4 that even though photons stop scattering off electrons at $z \sim 1000$, electrons do scatter many times off photons until much later. This is not a contradiction: there are many more photons than baryons. In any event, many cosmologists shy away from the word decoupling to describe what happens at $z \sim 1000$ for this reason.

[5] In Chapter 8 we will define a more precise measure of decoupling, making use of the visibility function, the probability that a photon last scattered at a given redshift. Using the visibility function, we will show that a CMB photon today most likely last scattered at a slightly higher redshift than inferred by the simple $n_e\sigma_T = H$ criterion. baryon density to the critical density is $m_p n_{\rm b}/\rho_{\rm cr} = \Omega_b a^{-3}$, n_b can be eliminated in Eq. (3.43) in favor of Ω_b:

$$n_e\sigma_T = 7.477 \times 10^{-30}\ \mathrm{cm}^{-1} X_e \Omega_b h^2 a^{-3}.$$

Dividing by the expansion rate leads to

$$\frac{n_e\sigma_T}{H} = 0.0692 a^{-3} X_e \Omega_b h \frac{H_0}{H}.$$

The ratio on the right depends on the Hubble rate, which is given in Eq. (1.2). From that equation or from Figure 1.3, we see that at early times, the main contribution comes from either matter or radiation, so $H/H_0 = \Omega_m^{1/2} a^{-3/2} \left[1 + a_{\mathrm{eq}}/a\right]^{1/2}$. Then,

$$\frac{n_e\sigma_T}{H} = 113 X_e \left(\frac{\Omega_b h^2}{0.02}\right) \left(\frac{.15}{\Omega_m h^2}\right)^{1/2} \left(\frac{1+z}{1000}\right)^{3/2} \left[1 + \frac{1+z}{3600}\frac{0.15}{\Omega_m h^2}\right]^{-1/2}.$$

Here I have normalized with "best fit" values for the baryon and matter densities. When the free electron fraction X_e drops below $\sim 10^{-2}$, photons decouple. From Figure 3.4, we see that X_e drops very quickly from unity to 10^{-3}. Therefore, decoupling takes place during recombination.

Let's forget all we just learned and ask what would happen if the universe remained ionzied throughout its history. In that hypothetical case, $X_e = 1$, and Eq. (3.46) can be trivially solved to find the redshift of decoupling. Setting the right hand side to 1 leads to

$$1 + z_{\mathrm{decouple}} = 43 \left(\frac{0.02}{\Omega_b h^2}\right)^{2/3} \left(\frac{\Omega_m h^2}{0.15}\right)^{1/3} \qquad \text{(no recombination)}.$$

Equation (3.47) tells us that even if the electrons remained ionized throughout the history of the universe, eventually the photons decoupled simply because expansion made it more difficult to find the increasingly dilute electrons. In theory, we do not expect the electrons to remain ionized throughout, so this calculation would appear academic. However, Eq. (3.47) is relevant for a more general reason. We do expect that at some late time. the electrons were reionized. We expect this because the universe we observe back to redshift $z \sim 6$ appears to be ionized. If the universe was reionized at very late times, much after the z_{decouple} of Eq. (3.47), there would not be a huge change in the CMIB anisotropy pattern. However, if the universe was reionized earlier than this redshift, multiple scattering of the photons would dramatically alter the primordial anisotropy pattern set up at $z \sim 1000$. Observations of the most distant quasars (Becker et al., 2001; Fan et al., 2002) suggest that reionization took place at $z \simeq 6$, so the alteration is expected to be slight.

12.3.4 Dark Matter

There is strong evidence for nonbaryonic dark matter in the universe, with $\Omega_{\mathrm{dm}} \simeq 0.3$. Perhaps the most plausible candidate for dark matter is a weakly interacting massive particle (WIMP), which was in close contact with the rest of the cosmic plasma at high temperatures, but then experienced freeze-out as the temperature dropped below its mass. Freeze-out is the inability of annihilations to keep the particle in equlibrium. Indeed, were it kept in equilibrium indefinitely, its abundance would be suppressed by $e^{-m/T}$: there would be no such particles in the observable universe. The purpose of this section, then, is to solve the Boltzmann equation for such a particle, determining the epoch of freeze-out and its relic abundance. The hope is that, by fixing its relic abundance so that $\Omega_{\mathrm{dm}} \simeq 0.3$, we will learn something about the fundamental properties of the particle, such as its mass and cross section. We then might use this knowledge to detect the particles in a laboratory.

In the generic WIMP scenario, two heavy particles X can annihilate producing two light (essentially massless) particles l. The light particles are assumed to be very tightly coupled

to the cosmic plasma, so they are in complete equilibrium (chemical as well as kinetic), with $n_l = n_l^{(0)}$. There is then only one unknown, n_X, the abundance of the heavy particle. We can use Eq. (3.9) to solve for this abundance:

$$a^{-3}\frac{d\left(n_X a^3\right)}{dt} = \langle\sigma v\rangle\left\{\left(n_X^{(0)}\right)^2 - n_X^2\right\}$$

To go further, recall that the temperature typically scales as a^{-1}, so if we multiply and divide the factor of $n_X a^3$ inside the parentheses on the left by T^3, we can remove $(aT)^3$ outside the derivative, leaving $T^3 d\left(n_X/T^3\right)/dt$. Let's define then

$$Y := \frac{n_X}{T^3}$$

The differential equation for Y becomes

$$\frac{dY}{dt} = T^3\langle\sigma v\rangle\left\{Y_{\rm EQ}^2 - Y^2\right\}$$

with $Y_{\rm EQ} := n_X^{(0)}/T^3$.

To go further, as in the neutron-proton case, it is convenient to introduce a new time variable,

$$x := m/T$$

where m, the mass of the heavy particle, sets a rough scale for the temperature during the region of interest. Very high temperature corresponds to $x \ll 1$, in which case reactions proceed rapidly so $Y \simeq Y_{\rm EQ}$. Since the X particles are relativistic at these epochs, the $m \ll T$ limit of Eq. (3.6) implies that $Y \simeq 1$. For high x, the equilibrium abundance $Y_{\rm EQ}$ becomes exponentially suppressed (e^{-x}). Ultimately, X particles will become so rare because of this suppression that they will not be able to find each other fast enough to maintain the equilibrium abundance. This is the onset of freeze-out. To change from t to x, we need the Jacobian $dx/dt = Hx$. Dark matter production typically occurs deep in the radiation era where the energy density scales as T^4, so $H = H(m)/x^2$. Then the evolution equation becomes

$$\frac{dY}{dx} = -\frac{\lambda}{x^2}\left\{Y^2 - Y_{\rm EQ}^2\right\}$$

where the ratio of the annihilation rate to the expansion rate is parameterized by

$$\lambda := \frac{m^3\langle\sigma v\rangle}{H(m)}.$$

In many theories λ is a constant, but in some, the thermally averaged cross section is temperature dependent; this leads to slight numerical changes in the following but unchanged qualitative solutions.

Equation (3.52) is a form of the Riccati equation, for which in general there are no analytic solutions. In this case, though, we can make use of our understanding of the freeze-out process to get an analytic expression for the final freeze-out abundance $Y_\infty := Y(x = \infty)$. Let's review this understanding in the context of Eq. (3.52). The left-hand side is of order Y (for $x \sim 1$) while the right is of order $Y^2\lambda$. We will see that λ is typically quite large, so as long as Y is not too small, the right-hand side must zero itself by setting $Y = Y_{\rm EQ}$. At late times, as $Y_{\rm EQ}$ drops precipitously, the terms on the right-hand side will no longer be much larger than the one on the left. In fact, well after freeze-out, Y will be much larger than $Y_{\rm EQ}$: the X particles will not be able to annihilate fast enough to maintain equilibrium. Thus at late times,

$$\frac{dY}{dx} \simeq -\frac{\lambda Y^2}{x^2} \quad (x \gg 1).$$

Integrate this analytically from the epoch of freeze-out x_f until very late times $x = \infty$ to get

$$\frac{1}{Y_\infty} - \frac{1}{Y_f} = \frac{\lambda}{x_f}.$$

Typically Y at freeze-out Y_f is significantly larger than Y_∞, so a simple analytic approximation is

$$Y_\infty \simeq \frac{x_f}{\lambda}.$$

This approximation is incomplete, in that it depends on the freeze-out temperature, which we have not determined. Although more precise determinations are possible (Exercise 10), a simple order-of-magnitude estimate for the dark matter problem is $x_f \sim 10$.

Figure 3.5 shows the numerical solution to Eq. (3.52) for several different values of λ. The abundances do track the equilibrium abundances until $m/T \sim 10$, after which they level off to a constant. The rough estimate $Y_\infty \sim 10/\lambda$ is seen to be a reasonable approximation for the relic abundance. Note that particles with larger cross sections (e.g. in the figure, $\lambda = 10^{10}$) freeze out later, and this later freeze-out carries along with it a lower relic abundance. Also note from the inset in Figure 3.5 that the distinction between Bose-Einstein, Fermi-Dirac, and Boltzmann statistics is important only at temperatures above the particle's mass. For temperatures relevant to the freeze-out process, our use of Boltzmann statistics is completely warranted.

There is one more piece of physics needed in order to determine the present-day abundance of these heavy particle relics. After freeze-out, the heavy particle density

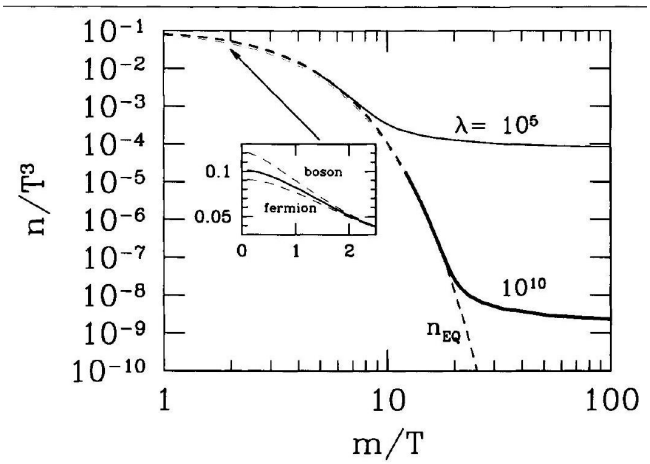


Figure 3.5. Abundance of heavy stable particle as the temperature drops beneath its mass. Dashed line is equilibrium abundance. Two different solid curves show heavy particle abundance for two different values of λ, the ratio of the annihilation rate to the Hubble rate. Inset shows that the difference between quantum statistics and Boltzmann statistics is important only at temperatures larger than the mass.

simply falls off as a^{-3}. So its energy density today is equal to $m\,(a_1/a_0)^3$ times its number density where a_1 corresponds to a time sufficiently late that Y has reached its aymptotic value, Y_∞. The number density at that time is $Y_\infty T_1^3$, so

$$\rho_X = m Y_\infty T_0^3 \left(\frac{a_1 T_1}{a_0 T_0}\right)^3 \simeq \frac{m Y_\infty T_0^3}{30}$$

The second equality here is nontrivial. You might expect that aT remains constant through the evolution of the universe, so that the ratio a_1T_1/a_0T_0 would be unity. It is not, for the same reason that the CMB and neutrinos have different temperatures. We saw in Chapter 2 that photons are heated by $e^{\pm}$annihilation, while neutrinos which have already decoupled are not. Similarly, as the universe expands, photons are heated by the annihilation of the zoo of particles with masses between 1MeV and 100GeV, so T does not fall simply as a^{-1}. You can show in Exercise 11 that as a result $(a_1T_1/a_0T_0)^3 \simeq 1/30$. Finally, to find the fraction of critical density today contributed by X, insert our expression for Y_∞ and divide by $\rho_{\rm cr}$:

$$\begin{aligned}\Omega_X &= \frac{x_f}{\lambda}\frac{mT_0^3}{30\rho_{\rm cr}} \\ &= \frac{H(m)x_fT_0^3}{30m^2\langle\sigma v\rangle\rho_{\rm cr}}.\end{aligned}$$

To find the present density of heavy particles, then, we need to compute the Hubble rate when the temperature was equal to the X mass, $H(m)$, for which we need the energy density when the temperature was equal to m. The energy density in the radiation era is given by Eq. (3.26) with g_* a function of temperature. Therefore,

$$\Omega_X = \left[\frac{4\pi^3 G g_*(m)}{45}\right]^{1/2}\frac{x_fT_0^3}{30\langle\sigma v\rangle\rho_{\rm cr}}$$

We see that Ω_X does not explicitly depend on the mass of the X particle. [6] So it is mainly the cross section which determines the relic abundance.

Let's now see what order of magnitude is needed to get dark matter today, i.e., to get $\Omega_X = \Omega_{\rm dm} \simeq 0.3$. At the temperatures of interest for dark matter production, $T \sim 100$GeV, $g_*(m)$ includes contributions from all the particles in the standard model (three generations of quarks and leptons, photons, gluons, weak bosons, and perhaps even the Higgs boson) and so is of order 100. Normalizing $g_*(m)$ and x_f by their nominal values leads to

$$\Omega_X = 0.3h^{-2}\left(\frac{x_f}{10}\right)\left(\frac{g_*(m)}{100}\right)^{1/2}\frac{10^{-39}\ {\rm cm}^2}{\langle\sigma v\rangle}.$$

The fact that this estimate is of order unity for cross sections of order 10^{-39} cm^2 is taken as a good sign: there are several theories which predict the existence of particles with cross sections this small.

Perhaps the most notable of these theories is supersymmetry, the theory which predicts that every particle has a partner with opposite statistics. For example, the supersymmetric partner of the spin zero Higgs boson is the spin- 1/2 Higgino (fermion). Initially it was hoped that the observed fermions could be the partners of the observed bosons, but this hope is not realized in nature. Instead, supersymmetry must be broken and all the supersymmetric partners of the known particles must be so massive that they have not yet been observed even in accelerators: they must have masses greater than 10 to 100GeV. Which of the supersymmetric partners is the best candidate for the dark matter today? The particle must be neutral since the evidence points to dark matter that is truly dark, i.e., does not interact much with the known particles and especially does not emit photons. The particle must also be stable: if it could decay to lighter particles, then decays would have kept it in equilibrium throughout the early universe and there would be none left today. The first of these criteria restricts the dark matter to be the partner of one of the neutral particles, such as the Higgs or the photon. [7] The second requires the particle to be the lightest (LSP for lightest supersymmetric partner) of these, for any heavier particles could decay into the lightest one plus some ordinary particles.

[6] There is a slight implicit dependence on mass in the freeze-out temperature x_f and in g_*, which is to evaluated when $T = m$.

7'he other neutral boson in the Standard Model is the Z vector boson. The lightest supersymmetric partner is likely a linear combination of the partners of all of these.

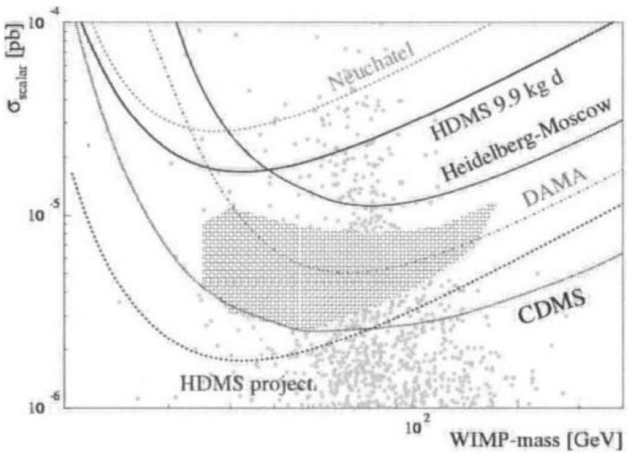


Figure 3.6. Constraints on supersymmetric dark matter (Baudis et al., 2000). Region above the solid curves is excluded, while filled region is reported detection by DAMA (Bernabei et al., 1999). Lowest curve (labled HDMS project) is only a projected limit based on one future experiment. The CDMS experiment (Abusaidi et al., 2000) appears to rule out the DAMA detection. Points scattered throughout correspond to different parameter choices in a class of supersymmetric models. Note the limits on the cross section are in units of picobarns)1 picobarn = 10^{-36} cm^2).

Not only would weakly interacting particles such as LSP's annihilate in the early universe, but if they were around today they would scatter off ordinary matter. Although it is difficult to detect these reactions because the rate is so low, a number of experiments have been performed searching for dark matter particles. Figure 3.6 shows the limits on the masses and cross sections of dark matter from these experiments (note that the scattering cross section, while related to the annihilation cross section, is not identically equal to it). Apart from a tantalizing detection from the DAMA experiment, so far we have only upper limits. However, as indicated in the figure, in the coming years the experiments are expected to pierce into the region predicted by supersymmetric theories.

The fact that a thermal relic with a cross section characteristic of the weak interaction gives the right dark matter abundance is called the WIMP miracle.

SUMMARY BEYOND EQUILIBRIUM

The light elements in the universe formed when the temperature of the cosmic plasma was of order 0.1MeV. Roughly a quarter of the mass of the baryons is in the form of ^{4}He, the remaining in the form of free protons with only trace amounts of deuterium, ^{3}He, and lithium. These elements remain ionized until the temperature of the universe drops well below the ionization energy of hydrogen. The epoch of recombination - at which time electrons and protons combine to form neutral hydrogen - is at redshift $z \sim 1000$ corresponding to a temperature $T \sim 0.25$eV. Before recombination, photons and electrons and protons are tightly coupled with one another because of Compton and Coulomb scattering. After this time, photons travel freely through the universe without interacting, so the photons in the CMB we observe today offer an excellent snapshot of the universe at $z \sim 1000$. The importance of this snapshot cannot be overstated.

The details of both nucleosynthesis and recombination are heavily influenced by the fact that the reactions involved eventually become too slow to keep up with the expansion rate.

This feature may also be responsible for the production of dark matter in the universe. We explored the popular scenario wherein a massive, neutral stable particle stops annihilating when the temperature drops significantly beneath its mass. The present-day abundance of such a particle can be deterimined in terms of its annihilation cross section, as in Eq. (3.60). Larger cross sections correspond to more efficient annhilation and therefore a lower abundance today. Roughly cross sections of order 10^{-40} are needed to get the dark matter abundance observed today. Such cross sections and the requisite stable, neutral particles emerge fairly natrually in extensions of the Standard Model of particle physics, such as supersymmetry, and may well be tested by accelerator experiments in the near future.

SUGGESTED READING

The Early Universe (Kolb and Turner) contains especially clear treatments of the heavy particle freeze-out problem and Big Bang nucleosynthesis, also based on the Boltzmann equation. Kinetic Theory in an Expanding Universe (Bernstein) offers some semianalytic solutions to these problems, as well as the requisite Boltzmannology.

Work on nucleosynthesis in the early universe dates back to Gamow and collaborators, summarized in Alpher, Follin, and Herman (1953). The first post-CMB papers were Peebles (1966) and Wagoner, Fowler, and Hoyle (1967); these got the basics right: 25% helium and roughly the correct amount of deuterium. Yang et al. (1984) helped many people of my generation understand that the baryonic density could be constrained with observations of the light elements. A nice review article on nucleosynthesis is Olive, Steigman, and Walker (2000).

As I tried to indicate in the text, the process of recombination is very rich; it involves some subtle physics. The original paper which worked through all the details was by Peebles (1968). Ma and Bertschinger (1996) however managed to describe the physics succintly in just one page in their Section 5.8. Seager, Sasselov, and Scott (1999) have presented a more accurate treatment (although, as they emphasize, Peebles' more intuitive work holds up remarkably well), including many small effects previously neglected.

Jungman, Kamionkowski, and Griest (1996) is a comprehensive review of all aspects of supersymmetric dark matter. Many papers have explored limits on supersymmetric dark matter candidates from cosmology and accelerators. To mention just several: Roszkowski (1991) showed that the Higgsino is likely not the lightest supersymmetric partner; Nath and Arnowitt (1992) and Kane et al. (1994) showed that the bino, the partner of the initial gauge eigenstate B, is the most likely both from the point of view of physics and cosmology - LSP; Ellis et al. (1997) combined accelerator constraints with those from cosmology to place a lower limit on the mass of the LSP; Edsjo and Gondolo (1997) included some subtle effects in the relic abundance calculation which affects the limits if the LSP is composed primarily of the partner of the Higgs boson; and Bottino et al. (2001) explored the consequences if the signal seen in the DAMA signal is due to dark matter particles.

D-1.

Compute the equilibrium number density (i.e., zero chemical potential) of a species with mass m and degeneracy $g = 2$ in the limits of large and small m/T. Take these limits for all three types of statistics: Boltzmann, BoseEinstein, and Fermi-Dirac. You will find Eqs. (C.26) and (C.27) helpful for the high- T Bose-Einstein and Fermi-Dirac limits.

D-2.

In the text, we treated $e^{\pm}$as relevant to the energy density at temperatures above m_e, but irrelevant afterwards (Eq. (3.30) and the discussion leading up to it). Track the $e^{\pm}$density

through annihilation assuming $n_{e^\pm} = n_{e^\pm}^{(0)}$. This equality holds during the BBN epoch because electromagnetic interactions (e.g., $e^+ + e^- \leftrightarrow \gamma + \gamma$) are so strong. When does the density fall to 1% of the photon energy density? If $\eta_b \simeq 6 \times 10^{-10}$, at what temperature do you expect $n_{e^-}$to depart from $n_{e^-}^{(0)}$?

D-3.

Compute the rate for neutron-to-proton conversion, λ_{np}. Show that it is equal to Eq. (3.29). There are two processes which contribute to $\lambda_{np} : n + \nu_e \to p + e^-$and $n + e^+ \to p + \bar{\nu}_e$. Assume that all particles can be described by Boltzmann statistics and neglect the mass of the electron. With these approximations the two rates are identical.

(a) Use Eq. (3.8) to write down the rate for $n + \nu_e \to p + e^-$. Perform the integrals over heavy particle momenta to get

$$\lambda_{np} = n_{\nu_e}^{(0)} \langle \sigma v \rangle = \frac{\pi}{4m^2} \int \frac{d^3 p_\nu}{(2\pi)^3 2 p_\nu} e^{-p_\nu / T} \times \int \frac{d^3 p_e}{(2\pi)^3 2 p_e} \delta \left(\mathcal{Q} + p_\nu - p_e \right) |\mathcal{M}|^2 .$$

(b) The amplitude squared is equal to $|\mathcal{M}|^2 = 32 G_F^2 \left(1 + 3 g_A^2\right) m_p^2 p_\nu p_e$, where g_A is the axial-vector coupling of the nucleon. The present best measurement of g_A is via the neutron lifetime, $\tau_n = \lambda_0 G_F^2 \left(1 + 3 g_A^2\right) m_e^5 / \left(2\pi^3\right)$, where the phase space integral

$$\lambda_0 := \int_1^{\mathcal{Q}/m_e} dx x \left(x - \mathcal{Q}/m_e\right)^2 \left(x^2 - 1\right)^{1/2} = 1.636$$

Carry out the integrals in Eq. (3.61) to get the rate, λ_{np} in terms of τ_n. Don't forget to multiply by 2 for the two different reactions.

D-4.

Solve the rate equation (3.27) numerically to determine the neutron fraction as a function of temperature. Ignore decays. There are (at least) two ways to perform this computation. The first is to treat it as a simple ordinary differential equation and solve numerically. The second is to proceed analytically and reduce the problem to an evaluation of a single numerical integral. This second method, which I'll lead you through here, is based on a numerical coincidence noted by Bernstein, Brown, and Feinberg (1988).

(a) Using standard differential equation techniques, show that a formal solution to Eq. (3.27) is

$$X_n(x) = \int_{x_i}^{x} dx' \frac{\lambda_{np}\left(x'\right) e^{-x'}}{x' H\left(x'\right)} e^{\mu(x') - \mu(x)}$$

where x_i is some initial, very high temperature, and

$$\mu(x) := \int_{x,}^{x} \frac{dx'}{x' H\left(x'\right)} \lambda_{np}\left(x'\right) \left[1 + e^{-x'}\right] .$$

(b) Use Eqs. (3.29) and (3.26) to compute the integrating factor μ analytically. Show that it is equal to

$$\mu = -\frac{255}{\tau_n \mathcal{Q}} \left[\frac{4\pi^3 G \mathcal{Q}^2 g_*}{45}\right]^{-1/2} \times \left[\left(\frac{4}{x^3} + \frac{3}{x^2} + \frac{1}{x}\right) + \left(\frac{4}{x^3} + \frac{1}{x^2}\right) e^{-x}\right]\Bigg|_{x_i}^{x}.$$

The simple form for μ is the result of numerical coincidence alone.

(c) With the results of part (b), do the single numerical integral in (a) numerically. Compare the asymptotic result at $x = \infty$ with the result in the text, $X_n(x = \infty) = 0.15$.

D-5.

Integrate the Friedmann equation (1.2) to verify the time-temperature relation in Eq. (3.30) in the epoch after $e^{\pm}$annihilation, but before matter domination.

D-6.

Determine η_b in terms of $\Omega_b h^2$. Show that it is given by Eq. (3.11).

D-7.

An important parameter for CMB anisotropies is the sound speed at decoupling. This is determined by the ratio of baryons to photons.

(a) Find

$$R := \frac{3\rho_b}{4\rho_\gamma}$$

as a function of a. Evaluate it at decoupling. Your answer should depend on $\Omega_b h^2$. (b) We will see in Chapter 8 that the sound speed of the combined photon/baryon fluid is

$$c_s = \sqrt{\frac{1}{3(1+R)}}.$$

Use your answer from (a) to plot the sound speed at decoupling as a function of $\Omega_b h^2$.

D-8.

Solve for the evolution of the free electron fraction. Do not compare your results with Figure 3.4 until you finish part (d). Throughout, take parameters $\Omega_m = 1, \Omega_b = 0.06, h = 0.5$.

(a) Use as an evolution variable $x := \epsilon_0/T$ instead of time in Eq. (3.39). Rewrite the equation in terms of x and the Hubble rate at $T = \epsilon_0$.

(b) Using the methods of Section 3.4, find the final freeze-out abundance of the free electron fraction, $X_e(x = \infty)$.

(c) Numerically integrate the equation from (a) from $x = 1$ down to $x = 1000$. What is the final frozen-out X_e?

(d) Peebles (1968) argued that even captures to excited states would not be important except for the fraction of times that the $n = 2$ state decays into two photons or expansion redshifts the Lyman alpha photon so that it cannot pump up a groundstate atom. Quantitatively, he multiplied the right-hand side of Eq. (3.39) by the correction factor,

$$C = \frac{\Lambda_\alpha + \Lambda_{2\gamma}}{\Lambda_\alpha + \Lambda_{2\gamma} + \beta^{(2)}}$$

where the two-photon decay rate is $\Lambda_{2\gamma} = 8.227\text{sec}^{-1}$; Lyman alpha production is $\beta^{(2)} = \beta e^{3\epsilon_0/4T}$; and

$$\Lambda_\alpha = \frac{H\left(3\epsilon_0\right)^3}{(8\pi)^2}$$

Do this and show how it changes your final answer. Now compare the freeze-out abundance with the result of (c) and the evolution with Figure 3.4.

D-9.

Find the redshift of decoupling as a function of Ω_b. If you do not have the evolution code of Exercise 8, use the Saha equation to determine X_e.

D-10.

Find an approximation to the freeze-out temperature of annihilating heavy particles by setting x_f such that $n^{(0)}\left(x_f\right)\langle\sigma v\rangle = H\left(x_f\right)$

D-11.

Typically the temperature of the cosmic plasma cools as a^{-1} with the expansion. However, when particles annihilate, they deposit energy into the plasma, thereby slowing the cooling. (Scherrer and Turner, 1986, showed that the annihilations do not actually heat the universe: T never increases, it simply decreases more slowly than a^{-1}.) Use the fact that the entropy density (Eq. (2.66)) scales as a^{-3} to compute the ratio of $(aT)^3$ at $T = 10\text{GeV}$ (roughly the time when WIMPs decouple) to its present value today.

D-12.

Suppose that there were no baryon asymmetry so that the number density of baryons exactly equaled that of anti-baryons. Determine the final relic density of (baryons+anti-baryons). At what temperature is this asymptotic value reached?

D-13.

There is a fundamental limitation on the annihilation cross section of a particle with mass m. Because of unitarity, $\langle\sigma v\rangle$ must be less than or equal to $1/\text{m}^2$, give or take a factor of order unity. Determine Ω_X for a particle which saturates this bound, i.e., for a particle with $\langle\sigma v\rangle = 1/m^2$. For what value of m is Ω_X equal to 1? (Keep x_f and g_* equal to the nominal values given in Eq. (3.60).) Note that if m is greater than this critical value, $\Omega_X > 1$, which is ruled out. This is a strong argument against stable particles (and therefore dark matter candidates) with masses above this critical value.

12.4 The Boltzmann Equations

We are interested in the anisotropies in the cosmic distribution of photons and inhomogeneities in the matter. Figure 4.1 shows why these are complicated to calculate. The photons are affected by gravity and by Compton scattering with free electrons. The electrons are tightly coupled to the protons. Both of these, of course, are also affected by gravity. The metric which determines the gravitational forces is influenced by all these components plus the neutrinos and the dark matter. Thus to solve for the photon and dark matter distributions, we need to simultaneously solve for all the other components.

There is a systematic way to account for all of these couplings. We write down a Boltzmann equation for each species in the universe. We have already encountered the Boltzmann equation in its integrated form in Chapter 3. There we were interested solely in the number density of the dark matter, the neutrons, and the free electrons. The number density is the integral over all momenta of the distribution function. Here we will be interested in more detailed information, not just the integrated number density, but the full distribution of photons, say, as a function of momentum. We then need a more primitive version of Eq. (3.1). Schematically, the unintegrated Boltzmann equation is

$$\frac{df}{dt} = C[f]$$

The right-hand side of the Boltzmann equation contains all possible collision terms. These terms in general are complicated functionals of the distribution functions of the various components. In the absence of collisions, the distribution function obeys $df/dt = 0$. This seemingly innocent equation says that the number of particles in a given element of phase space does not change with time. The catch is that the phase space elements themselves are moving in time in complicated ways due to the nontrivial metric. This catch makes the problem more difficult than it seems from Eq. (4.1). Nonetheless, we can still progress systematically by reexpressing the full derivative in terms of partial derivatives.

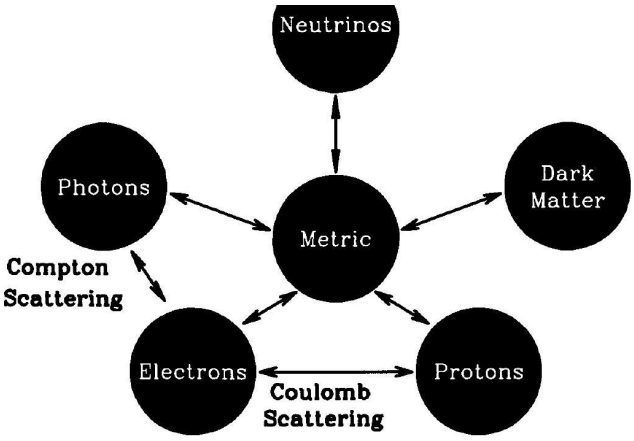


Figure 4.1. The ways in which the different components of the universe interact with each other. These connections are encoded in the coupled Boltzmann-Einstein equations.

In this chapter, we derive the Boltzmann equations for photons, eletrons, protons, dark matter, and massless neutrinos. This set of equations governs the evolution of perturbations in the universe.

12.4.1 The Boltzmann Equation for the Harmonic Oscillator

Before tackling the problem of interest - the Boltzmann equation for all species in an expanding universe - let us treat a much simpler example of the Boltzmann equation: the nonrelativistic harmonic oscillator. This simple example is very similar to the full general relativistic version we will encounter in the next section, but the algebra is much less cumbersome. So here the physics will be quite transparent. It will be useful to keep this example in mind when the algebra threatens to obscure the physics in the next section.

Consider a one-dimensional harmonic oscillator with energy

$$E = \frac{p^2}{2m} + \frac{1}{2}kx^2$$

The distribution function of the harmonic oscillator depends on time t, position x, and momentum p. Thus, the full time derivative in Eq. (4.1) can be rewritten as

$$\frac{df(t,x,p)}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial p}\frac{dp}{dt}.$$

Figure 4.2 illustrates the movement through phase space of a distribution of collisionless ($C = 0$) oscillators. The full time derivative df/dt vanishes since the number of particles in the bunch at t_1 equals that at t_2. What has changed is the location of the phase space elements $x(t)$ and $p(t)$ themselves. Alternatively, we can think of x and p as independent variables (not dependent on t) and take partial derivatives of f with respect to t, x, and p. All of these partial derivatives are nonzero, but the appropriate weighted sum of the three vanishes.

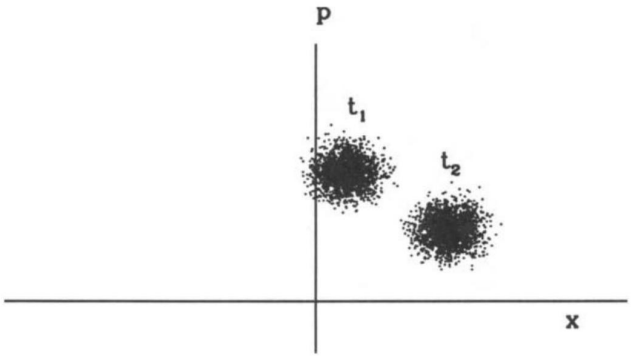


Figure 4.2. Distribution function for a set of collisionless harmonic oscillators. The initial distribution at t_1 moves in phase space by time t_2. The distribution function $f(t, x, p)$ remains constant as long as the evolution of $x(t)$ and $p(t)$ is accounted for.

To determine the coefficients dx/dt and dp/dt we must use the equations of motion. By the definition of momentum,

$$\frac{dx}{dt} := \frac{p}{m}$$

This equation will be generalized to a fully relativistic, three-dimensional version in the next section. Indeed we already got a preview of this when we defined $P^\mu := dx^\mu/d\lambda$ in Chapter 2. Newton's equation governing the motion of the oscillator is

$$\frac{dp}{dt} = -kx$$

The analogue of this familiar equation in the next section will be the geodesic equation of general relativity.

The collisionless Boltzmann equation for the harmonic oscillator is thus

$$\frac{\partial f}{\partial t} + \frac{p}{m}\frac{\partial f}{\partial x} - kx\frac{\partial f}{\partial p} = 0$$

The second term here governs how rapidly the oscillator moves in real space; the coefficient in front is just the velocity, p/m. The last term governs how quickly particles lose momentum.

In order to solve the Boltzmann equation, we need to know the initial conditions on the distribution function. Even without these, though, the Boltzmann equation offers some useful physics. Consider the equilibrium distribution, wherein $\partial f/\partial t = 0$. A general solution for the equilibrium distribution is

$$f(p,x) = f_{\rm EQ}(E)$$

that is, f is a function only of energy E. To see that this is indeed a solution, consider

$$\frac{p}{m}\frac{\partial f(E)}{\partial x} - kx\frac{\partial f(E)}{\partial p} = \frac{df}{dE}\left[\frac{p}{m}\frac{\partial E}{\partial x} - kx\frac{\partial E}{\partial p}\right] = 0.$$

So any function of the energy alone is an equilibrium distribution. Of course, in general, there will be interactions, or collisions. The only way for the full Boltzmann equation to be satisfied is if the collision terms also vanish. This will in general drive f to one of the familiar equilibrium distributions, e.g., $e^{-E/T}$ for the classical Maxwell-Boltzmann distribution.

12.4.2 The Collisionless Boltzmann Equation for Photons

Let us begin then by considering the left hand side of Eq. (4.1) for massless photons. First we must specify the form of the metric, accounting for perturbations around the smooth universe described by Eq. (2.4). Whereas the smooth universe is characterized by a single function, $a(t)$, which depends only on time and not on space, the perturbed universe requires two more functions, Ψ and Φ, both of which depend on both space and time. In terms of them, the metric can be written as

$$\begin{aligned} g_{00}(\vec{x},t) &= -1 - 2\Psi(\vec{x},t) \\ g_{0i}(\vec{x},t) &= 0 \\ g_{ij}(\vec{x},t) &= a^2\delta_{ij}(1 + 2\Phi(\vec{x},t)). \end{aligned}$$

In the absence of Ψ and Φ, Eq. (4.9) is simply the FRW metric of the zero-order homogeneous, flat cosmology. Similarly, in the absence of expansion ($a = 1$) this metric describes a weak gravitational field (

D-2.

3). The perturbations to the metric are Ψ, which corresponds to the Newtonian potential, and Φ, the perturbation to the spatial curvature. Since the perturbations in the universe are small at the times and scales of interest, we will treat these Ψ and Φ as small quantities, dropping all terms quadratic in them. There are two technical points about the metric in Eq. (4.9) which you don't need to worry about for most of this book, but which nonetheless are important to be aware of, if only to better understand the literature. First, one can break up perturbations into those behaving as scalars, vectors, and tensors under a transformation from one 3D coordinate system to another. Equation (4.9) contains only scalar perturbations. In principle, it is possible that the metric of our universe also has vector or tensor perturbations. If so, $g_{\mu\nu}$ would require other functions besides Ψ and Φ to fully describe all perturbations. For example, the off-diagonal elements become nonzero if there are vector perturbations. Indeed, there are many cosmological theories wherein there are both tensor and vector perturbations. For example, inflation tends to predict that there will be tensor perturbations, while models based on topological defects tend to produce large vector perturbations. For now we focus solely on the scalar perturbations; these are the only ones that couple to matter perturbations and are the most important that couple to photon perturbations as well.

The other feature of Eq. (4.9) worth noting is that its form corresponds to a choice of gauge. The simplest way to understand this gauge freedom is to think back to electricity and magnetism. There, the vector potential A_μ and its derivatives contain all possible information about the electric and magnetic fields. Since the physical $\vec{E}$ and $\vec{B}$ fields remain unchanged if a constant is added to A_μ, there is some residual freedom in choosing the potential. (For example, one often chooses $A_0 = 0$ or $\partial_\mu A^\mu = 0$.) In our case of perturbations to the metric, a similar

freedom exists. Even if only scalar perturbations are considered, there is still considerable freedom in the variables one chooses to describe the fluctuations. Although any physical results must be insensitive to the gauge choice, it is possible to use a gauge which looks quite different from Eq. (4.9) and still describes the same physics. For the record, the gauge in Eq. (4.9) is called the conformal Newtonian gauge. [1]

We want to reexpress the total derivative in Eq. (4.1) as a sum of partial derivatives. The distribution function depends on the space-time point $x^\mu = (t, \vec{x})$ and also on the momentum vector defined as

$$P^\mu := \frac{dx^\mu}{d\lambda}$$

where λ again parametrizes the particle's path, as in Eq. (2.18) (and again we will not need to specify λ explicitly). Thus, in principle, f is a function defined in an 8 -dimensional space. However, not all the components of the momentum vector are independent since the masslessness of the photon implies that

$$P^2 := g_{\mu\nu} P^\mu P^\nu = 0$$

[1] Historically, the initial ground-breaking work on the evolution of fluctuations was carried out in synchronous gauge (Peebles and Yu, 1970; Wilson and Silk, 1981; Peebles 1982; Bond and Szalay, 1983; Bond and Efstathiou, 1984). Recently, the physics of the anisotropies has been elucidated most clearly by using conformal Newtonian gauge (e.g., Hu and Sugiyama, 1995). Exercise 2 works out some of the relevant equations in synchronous gauge. So there are only three independent components of the momentum vector. Before we choose which three we will use, let us enforce the constraint of Eq. (4.11), using the metric of Eq. (4.9).

$$P^2 = 0 = -(1 + 2\Psi)\left(P^0\right)^2 + p^2 = 0$$

where I have defined

$$p^2 := g_{ij} P^i P^j$$

We can use the constraint equation then to eliminate the time component of P^μ:

$$P^0 = \frac{p}{\sqrt{1 + 2\Psi}} = p(1 - \Psi)$$

This last equality holds since we are doing first-order perturbation theory in the small quantity Ψ. With our sign convention, an overdense region has $\Psi < 0$. Therefore, in an overdense region, the term in parentheses on the right-hand side here is greater than one. Thus, Eq. (4.14) tells us photons lose energy-redshift-as they move out of a potential well.

Equation (4.14) is the generalization of the relativistic expression $E = pc$ to a perturbed Friedmann-Robertson-Walker metric. It allows us to eliminate P^0 whenever it occurs in favor of p, the generalized magnitude of the momentum. Recall that in the harmonic oscillator case, we did not include a term proportional to $\partial f/\partial E$ in Eq. (4.3). Here, too, we do not need to include a term proportional to $\partial f/\partial P^0$ when expanding the total time derivative. We need include only the dependence of f on the momentum: both the magnitude p and the angular direction. For the direction vector, we'll use the unit vector $\hat{p}^i = \hat{p}_i$, which by definition satisfies $\delta_{ij}\hat{p}^i\hat{p}^j = 1$.

We can now write Eq. (4.1) as

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x^i} \cdot \frac{dx^i}{dt} + \frac{\partial f}{\partial p}\frac{dp}{dt} + \frac{\partial f}{\partial \hat{p}^i} \cdot \frac{d\hat{p}^i}{dt}.$$

The easiest term in Eq. (4.15) is the last one since it does not contribute at first order in perturbation theory. To see this, first recall that the zero-order distribution function is simply the Bose-Einstein function which depends only on p, not on the direction $\hat{p}^i$. Therefore, $\partial f/\partial \hat{p}^i$ is nonzero only if we consider the perturbation to the zero-order f; i.e., it is a first-order term. But so is the term which multiplies it, $d\hat{p}^i/dt$, for the direction of a photon changes only in the presence of potentials Φ and Ψ. In the absence of these potentials, a photon moves in a straight line. Thus the last term is the product of two first-order terms, rendering it a second-order term. We can neglect it.

Next let us reexpress the second term on the right-hand side of Eq. (4.15) by recalling that (Eq. (4.10)) $P^i := dx^i/d\lambda$ and $P^0 := dt/d\lambda$. Therefore,

$$\frac{dx^i}{dt} = \frac{dx^i}{d\lambda}\frac{d\lambda}{dt}$$

$$= \frac{p^i}{p^0}.$$

We want to reexpress this ratio in terms of our favored variables p and $\hat{p}^i$. Equation (4.14) does this for P^0; let's do the same for the numerator P^i. The comoving momentum P^i is proportional to $\hat{p}^i$; call the proportionality constant C:

$$P^i := C\hat{p}^i$$

To determine the coefficient C, we can use Eq. (4.13):

$$\begin{aligned} p^2 &= g_{ij}\hat{p}^i\hat{p}^j C^2 \\ &= a^2(1+2\Phi)\delta_{ij}\hat{p}^i\hat{p}^j C^2 \\ &= a^2(1+2\Phi)C^2 \end{aligned}$$

where the last equality holds because the direction vector is a unit vector. Equation (4.18) tells us that $C = p(1-\Phi)/a$ so whenever we encounter P^i, we can always eliminate it in terms of $p, \hat{p}^i$ via

$$P^i = p\hat{p}^i\frac{1-\Phi}{a}$$

From Eqs. (4.16) and (4.19), we see that

$$\frac{dx^i}{dt} = \frac{\hat{p}^i}{a}(1+\Psi-\Phi)$$

An overdense region has $\Psi < 0$ and $\Phi > 0$, rendering the term in parentheses less than one. So, Eq. (4.20) says that a photon slows down (dx/dt becomes smaller) when traveling through an overdense region. This makes perfect sense: we expect the gravitational force of an overdense region to slow down even photons. Having said that, I now claim that we can neglect the potentials in Eq. (4.20). For, in the Boltzmann equation they multiply $\partial f/\partial x^i$ which is a first-order term. (Again, the zero-order distribution function does not depend on position.) So collecting terms up to this point, we have

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial f}{\partial x^i} + \frac{\partial f}{\partial p}\frac{dp}{dt}$$

The remaining term to be calculated is dp/dt. Alas, unlike the harmonic oscillator, here $dp/dt \neq -kx$. Rather we will need the geodesic equation from general relativity and more fortitude to compute dp/dt for photons in a perturbed FRW metric.

To begin, let us recall that the time component of the geodesic equation (2.18) can be written as

$$\frac{dP^0}{d\lambda} = -\Gamma^0_{\alpha\beta} P^\alpha P^\beta.$$

We can rewrite the derivative with respect to λ as a derivative with respect to time multiplied by $dt/d\lambda = P^0$. Also, we can use Eq. (4.14) to eliminate P^0 in terms of our favored variable p. Then the geodesic equation reduces to

$$\frac{d}{dt}[p(1-\Psi)] = -\Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p}(1+\Psi)$$

Expand out the time derivative to get

$$\frac{dp}{dt}(1-\Psi) = p\frac{d\Psi}{dt} - \Gamma^0_{\alpha\beta}\frac{P^\alpha p^\beta}{p}(1+\Psi).$$

Now we multiply both sides by $(1+\Psi)$; drop all terms quadratic in Ψ; and reexpress the total time derivative of Ψ in terms of partial derivatives so that

$$\frac{dp}{dt} = p\left\{\frac{\partial\Psi}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}\right\} - \Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p}(1+2\Psi).$$

In order to evaluate dp/dt then we need to evaluate the product $\Gamma^0_{\alpha\beta}P^\alpha P^\beta/p$. Recall that the Christoffel symbol is best written as a sum of derivatives of the metric (Eq. (2.19)). Here we are interested only in the $\Gamma^0_{\alpha\beta}$ component. It multiplies $P^\alpha P^\beta$, which is symmetric in α, β. Thus, the first two metric derivatives contribute equally, and we have

$$\Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p} = \frac{g^{0\nu}}{2}\left[2\frac{\partial g_{\nu\alpha}}{\partial x^\beta} - \frac{\partial g_{\alpha\beta}}{\partial x^\nu}\right]\frac{P^\alpha P^\beta}{p}$$

Now $g^{0\nu}$ is nonzero only when $\nu = 0$, in which case it is simply the inverse of g_{00}, so

$$\Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p} = \frac{-1+2\Psi}{2}\left[2\frac{\partial g_{0\alpha}}{\partial x^\beta} - \frac{\partial g_{\alpha\beta}}{\partial t}\right]\frac{P^\alpha P^\beta}{p}$$

Once again, $g_{0\alpha}$ in the first term in brackets is nonzero only when $\alpha = 0$, in which case its derivative is $-2\partial\Psi/\partial x^\beta$. The second term in brackets multiplied by the product of momenta is

$$\begin{aligned}-\frac{\partial g_{\alpha\beta}}{\partial t}\frac{P^\alpha P^\beta}{p} &= -\frac{\partial g_{00}}{\partial t}\frac{P^0 P^0}{p} - \frac{\partial g_{ij}}{\partial t}\frac{P^i P^j}{p} \\ &= 2\frac{\partial\Psi}{\partial t}p - a^2\delta_{ij}\left[2\frac{\partial\Phi}{\partial t} + 2H(1+2\Phi)\right]\frac{P^i P^j}{p}\end{aligned}$$

But, via Eq. (4.19), $\delta_{ij}P^iP^j = p^2(1-2\Phi)/a^2$, so we have

$$\begin{aligned}\Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p} = \frac{-1+2\Psi}{2}\Bigg[&-4\frac{\partial\Psi}{\partial x^\beta}P^\beta + 2p\frac{\partial\Psi}{\partial t} \\ &-p\left\{2\frac{\partial\Phi}{\partial t} + 2H(1+2\Phi)\right\}(1-2\Phi)\Bigg]\end{aligned}$$

The last line here simplifies since $(1+2\Phi)(1-2\Phi) \to 1$ at first order, and $1-2\Phi$ can be set to 1 when multiplying $\partial\Phi/\partial t$. Summing over the index β in the first term then leads to

$$\Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p} = \frac{-1+2\Psi}{2}\left[-4\left(\frac{\partial\Psi}{\partial t}p + \frac{\partial\Psi}{\partial x^i}\frac{p\hat{p}^i}{a}\right) + 2p\frac{\partial\Psi}{\partial t} - p\left\{2\frac{\partial\Phi}{\partial t} + 2H\right\}\right]$$
$$= \{-1+2\Psi\}\left[-\frac{\partial\Psi}{\partial t}p - 2\frac{\partial\Psi}{\partial x^i}\frac{p\hat{p}^i}{a} - p\left\{\frac{\partial\Phi}{\partial t} + H\right\}\right]$$

We can insert this into Eq. (4.25) to get

$$\frac{dp}{dt} = p\left\{\frac{\partial\Psi}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}\right\} - \frac{\partial\Psi}{\partial t}p - 2\frac{\partial\Psi}{\partial x^i}\frac{p\hat{p}^i}{a} - p\left\{\frac{\partial\Phi}{\partial t} + H\right\}.$$

Collecting terms, we finally have

$$\frac{1}{p}\frac{dp}{dt} = -H - \frac{\partial\Phi}{\partial t} - \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}$$

Equation (4.32) is what we were after. It describes the change in the photon momentum as it moves through a perturbed FRW universe. The first term accounts for the loss of momentum due to the Hubble expansion. To understand the significance of the next two terms in Eq. (4.32), we first need to remember that an overdense region has $\Phi > 0$ and $\Psi < 0$ with our sign conventions. Therefore, the second term says that a photon in a deepening gravitational well ($\partial\Phi/\partial t > 0$) loses energy. This is understandable: the deepening well makes it more difficult for the photon to emerge, thereby increasing the magnitude of the redshift. Finally, a photon traveling into a well ($\hat{p}^i\partial\Psi/\partial x^i < 0$) gains energy because it is being pulled toward the center. Conversely, as it leaves the well, it gets redshifted.

We are now in a position to write down the Boltzmann equation for photons. Using Eq. (4.32) in Eq. (4.21) leads to

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial f}{\partial x^i} - p\frac{\partial f}{\partial p}\left[H + \frac{\partial\Phi}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}\right]$$

This equation incorporates much of the physics with which we are already familiar, such as the fact that photons redshift in an expanding universe. It also leads directly to the equations governing anisotropies. Working through the terms on the right, the first two are familiar from standard hydrodynamics; when integrated, they lead to the continuity and Euler equations (Exercise 1). The third term dictates that photons lose energy in an expanding universe. We saw some of this in Chapter 2 when considering geodesics. Shortly, we will see how the Boltzmann formalism enforces this result. Finally, the last two encode the effects of under-/overdense regions on the photon distribution function. To go further we must now expand the photon distribution function f about its zero-order Bose-Einstein value. I will do this in a way that may seem odd at first. Let us write

$$f(\vec{x}, p, \hat{p}, t) = \left[\exp\left\{\frac{p}{T(t)[1+\Theta(\vec{x},\hat{p},t)]}\right\} - 1\right]^{-1}$$

Here the zero-order temperature T is a function of time only (i.e., scales as a^{-1}), not space. The perturbation to the distribution function is characterized by Θ, which could also be called $\delta T/T$. In the smooth zero-order universe, photons are distributed homogeneously, so T is independent of $\vec{x}$, and isotropically, so T is independent of the direction of propagation $\hat{p}$. Now that we want to describe perturbations about this smooth universe, we need to allow for inhomogeneities in the photon distribution (so Θ depends on $\vec{x}$) and anisotropies (so Θ depends on $\hat{p}$). There is one assumption built into Eq. (4.34). I have explicitly written down that Θ depends on $\vec{x}, \hat{p}$, and t. This assumes that it does not depend on the magnitude of the momentum p. We will soon see that this is a valid assumption, following directly from that

fact that the magnitude of the photon momentum is virtually unchanged during a Compton scatter. The perturbation Θ is small, so we can expand (again keeping only terms up to first order)

$$\begin{aligned} f &\simeq \frac{1}{e^{p/T}-1} + \left(\frac{\partial}{\partial T}\left[\exp\left\{\frac{p}{T}\right\}-1\right]^{-1}\right) T\Theta \\ &= f^{(0)} - p\frac{\partial f^{(0)}}{\partial p}\Theta. \end{aligned}$$

In the last line I have identified the zero-order distribution function as the BoseEinstein distribution with zero chemical potential,

$$f^{(0)} := \left[\exp\left\{\frac{p}{T}\right\}-1\right]^{-1}$$

and made use of the fact that for this function $T\partial f^{(0)}/\partial T = -p\partial f^{(0)}/\partial p$.

Zero-Order Equation

We can now set about systematically collecting the terms of similar order in Eq. (4.33). Let us start with the zero-order terms, those with no Φ, Ψ, or Θ. These lead immediately to

$$\left.\frac{df}{dt}\right|_{\text{zero order}} = \frac{\partial f^{(0)}}{\partial t} - Hp\frac{\partial f^{(0)}}{\partial p} = 0.$$

I have set df/dt here equal to zero, i.e., set the collision term on the right of Eq. (4.1) to zero. I could justify this by claiming that we are now looking only at the collisionless Boltzmann equation. But there is a much deeper justification. In fact, even when we come around to including collisions, we will see that there is no zero-order collision term. That is, the collision terms will be proportional to Θ and other perturbatively small quantities. There is a profound reason for this: the zero-order distribution function is set precisely by the requirement that the collision term vanishes. Another, perhaps more familiar, way of saying this is to point out that any collision term includes the rate for the given reaction and for its inverse. If the distribution functions are set to their equilibrium values, the rate for the reaction precisely cancels the rate for its inverse. If a given component is out of equilibrium, collisions will drive it toward its equilibrium distribution. This is the reason we expected a Bose-Einstein distribution in the first place. Its observation is convincing evidence that photons were at one point in the early universe tightly coupled to the electrons.

Returning to Eq. (4.37), we can rewrite the time derivative as

$$\frac{\partial f^{(0)}}{\partial t} = \frac{\partial f^{(0)}}{\partial T}\frac{dT}{dt} = -\frac{dT/dt}{T}p\frac{\partial f^{(0)}}{\partial p}$$

so that the zero-order equation becomes

$$\left[-\frac{dT/dt}{T} - \frac{da/dt}{a}\right]\frac{\partial f^{(0)}}{\partial p} = 0$$

Thus $dT/T = -da/a$ or

$$T \propto \frac{1}{a}$$

This is precisely what we expected from the heuristic argument about the photon's wavelength getting stretched as the universe expands (Section 1.1) and the more concrete argument of Section 2.1. It is reassuring to see this result emerge from the Boltzmann treatment.

First-Order Equation

We now return to Eq. (4.33) and extract the equation for the deviation of the photon temperature from its zero-order value, i.e., an equation for Θ. To do this, everywhere we encounter f in Eq. (4.33), we insert the expansion of Eq. (4.35):

$$\left.\frac{df}{dt}\right|_{\text{first order}} = -p\frac{\partial}{\partial t}\left[\frac{\partial f^{(0)}}{\partial p}\Theta\right] - p\frac{\hat{p}^i}{a}\frac{\partial\Theta}{\partial x^i}\frac{\partial f^{(0)}}{\partial p} + Hp\Theta\frac{\partial}{\partial p}\left[p\frac{\partial f^{(0)}}{\partial p}\right]$$
$$- p\frac{\partial f^{(0)}}{\partial p}\left[\frac{\partial\Phi}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}\right].$$

Consider the first term on the right-hand side here. The time derivative can be rewritten as a temperature derivative so

$$-p\frac{\partial}{\partial t}\left[\frac{\partial f^{(0)}}{\partial p}\Theta\right] = -p\frac{\partial f^{(0)}}{\partial p}\frac{\partial\Theta}{\partial t} - p\Theta\frac{dT}{dt}\frac{\partial^2 f^{(0)}}{\partial T\partial p}$$
$$= -p\frac{\partial f^{(0)}}{\partial p}\frac{\partial\Theta}{\partial t} + p\Theta\frac{dT/dt}{T}\frac{\partial}{\partial p}\left[p\frac{\partial f^{(0)}}{\partial p}\right]$$

The second line follows here since $\partial f^{(0)}/\partial T = -(p/T)\partial f^{(0)}/\partial p$. The second term on this second line cancels the third term on the right in Eq. (4.40), so we can finally write down the equation governing the perturbation θ:

$$\left.\frac{df}{dt}\right|_{\text{first order}} = -p\frac{\partial f^{(0)}}{\partial p}\left[\frac{\partial\Theta}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial\Theta}{\partial x^i} + \frac{\partial\Phi}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}\right]$$

The first two terms here account for "free streaming," which translates into anisotropies on increasingly small scales as the universe evolves. The last two account for the effect of gravity. Note that every time x appears it is multiplied by a, the scale factor. This must happen, for physical distances are ax.

12.4.3 Collision Terms: Compton Scattering

Our task in this section is to determine the influence Compton scattering has on the photon distribution function. The scattering process of interest is

$$e^-(\vec{q}) + \gamma(\vec{p}) \leftrightarrow e^-(\vec{q}') + \gamma(\vec{p}')$$

where I have explicitly indicated the momentum of each particle.

We are interested in the change of distribution of photons with momentum $\vec{p}$ (with magnitude p and direction $\hat{p}$). Therefore we must sum over all other momenta $(\vec{q}, \vec{q}', \vec{p}')$ which affect $f(\vec{p})$. Schematically, then, the collision term is

$$C[f(\vec{p})] = \sum_{\vec{q},\vec{q}',\vec{p}'} |\text{ Amplitude }|^2 \{f_e(\vec{q}')f(\vec{p}') - f_e(\vec{q})f(\vec{p})\}$$

The amplitude is reversible so it multiplies both the reaction and its inverse. The products of the electron distribution function f_e and the photon distribution function simply count the number of particles with the given momenta. I have neglected stimulated emission and Pauli blocking, which would lead to factors of $1 + f$ and $1 - f_e$ with the appropriate momenta. At first order this turns out to be a valid assumption. If one were to go to second order, though, stimulated emission would have to be included. Pauli blocking is never important after electron-positron annihilation because the occupation numbers f_e are very small (Exercise 4).

Unfortunately, the collision term becomes a bit messier than the schematic version when we put in all the factors of 2π to properly account for the sums over phase space. Explicitly, the collision term is [2]

$$C[f(\vec{p})] = \frac{1}{p}\int \frac{d^3q}{(2\pi)^3 2E_e(q)} \int \frac{d^3q'}{(2\pi)^3 2E_e\left(q'\right)} \int \frac{d^3p'}{(2\pi)^3 2E\left(p'\right)} |\mathcal{M}|^2 (2\pi)^4$$

[2] Most of the phase space factors here follow from our discussion in Section 3.1, the exception being the factor of $1/p$ in front. You may have wondered about one other feature of the Boltzmann equation presented in this chapter: I started at the outset taking df/dt; does not general relativity require us to take the derivative with respect to the affine parameter λ? Exercise 5 illustrates how these problems solve each other.

$$\begin{aligned} &\times \delta^3[\vec{p}+\vec{q}-\vec{p}'-\vec{q}']\,\delta\left[E(p)+E_e(q)-E\left(p'\right)-E_e\left(q'\right)\right] \\ &\quad \times \times \left\{f_e\left(\vec{q}'\right) f\left(\vec{p}'\right) - f_e(\vec{q}) f(\vec{p})\right\} \end{aligned}$$

Here the delta functions enforce energy momentum conservation. The energies at this order are the relativistic limit for photons and non-relativistic limit for electrons: $E(p) = p$ and $E_e(q) = m_e + q^2/\left(2m_e\right)$. Note the similarity between this collision term and the general one (Eq. (3.1)) we considered in Chapter 3. The only difference is that I have not integrated this collision term over all photon momenta $\vec{p}$, so there are only three momentum integrals. Again, this reflects our need to understand how photons traveling in different directions interact: we will see that the collision term depends on $\hat{p}$.

Since the kinetic energy of the electrons is very small at the epochs of interest compared with their rest energy, the factors of E_e in the denominator of Eq. (4.45) may be replaced with m_e. Then using the three-dimensional momentum delta function to do the $\vec{q}'$ integral, we have

$$\begin{aligned} C[f(\vec{p})] = \frac{\pi}{4m_e^2 p}\int \frac{d^3q}{(2\pi)^3} \int \frac{d^3p'}{(2\pi)^3 p'} \delta\left[p + \frac{q^2}{2m_e} - p' - \frac{\left(\vec{q}+\vec{p}-\vec{p}'\right)^2}{2m_e}\right] \\ \times |\mathcal{M}|^2 \left\{f_e\left(\vec{q}+\vec{p}-\vec{p}'\right) f\left(\vec{p}'\right) - f_e(\vec{q}) f(\vec{p})\right\} \end{aligned}$$

To go further, we need to understand the kinematics of nonrelativistic Compton scattering. The most important feature of this process for our purposes is that very little energy is transferred. In particular,

$$\begin{aligned} E_e(q) - E_e\left(\vec{q}+\vec{p}-\vec{p}'\right) &= \frac{q^2}{2m_e} - \frac{\left(\vec{q}+\vec{p}-\vec{p}'\right)^2}{2m_e} \\ &\simeq \frac{\left(\vec{p}'-\vec{p}\right)\cdot\vec{q}}{m_e}, \end{aligned}$$

where the last approximate equality holds since q is much larger than $\vec{p}$ and $\vec{p}'$. In nonrelativistic Compton scattering, $p' \simeq p$, scattering is nearly elastic. Therefore, $\vec{p}'-\vec{p}$ is of order p, of order the ambient temperature T. So the right-hand side of Eq. (4.47) is of order $Tq/m_e \sim Tv_{\rm b}$ where the baryonic velocity $v_{\rm b}$ is very small. The change in the electron energy due to Compton scattering is therefore of order $Tv_{\rm b}$. Since the typical kinetic energy of the electrons is also of order T, the fractional energy change in a single Compton collision is very small, of order $v_{\rm b}$. It makes sense, therefore, to expand the final electron kinetic energy $\left(\vec{q}+\vec{p}-\vec{p}'\right)^2/\left(2m_e\right)$ around its zero-order value of $q^2/\left(2m_e\right)$. The delta function can be expanded as

$$\delta\left[p+\frac{q^2}{2m_e}-p'-\frac{\left(\vec{q}+\vec{p}-\vec{p}'\right)^2}{2m_e}\right]\simeq\delta\left(p-p'\right)$$
$$+\left(E_e\left(q'\right)-E_e(q)\right)\left.\frac{\partial\delta\left(p+E_e(q)-p'-E_e\left(q'\right)\right)}{\partial E_e\left(q'\right)}\right|_{E_e(q)=E_e(q')}$$
$$=\delta\left(p-p'\right)+\frac{\left(\vec{p}-\vec{p}'\right)\cdot\vec{q}}{m_e}\frac{\partial\delta\left(p-p'\right)}{\partial p'}$$

where the second equality makes use of the fact that for a general function f of the sum of two variables, $\partial f(x-y)/\partial x=-\partial f(x-y)/\partial y$. This formal expansion appears ill-defined at present, but when integrating over momenta, the derivatives of delta functions will be handled by integrating by parts. With this expansion, and using the fact that $f_e\left(\vec{q}+\vec{p}-\vec{p}'\right)\simeq f_e(\vec{q})$, the collision term becomes

$$C[f(\vec{p})]=\frac{\pi}{4m_e^2p}\int d^3q\frac{f_e(\vec{q})}{(2\pi)^3}\int\frac{d^3p'}{(2\pi)^3p'}|\mathcal{M}|^2$$
$$\times\left\{\delta\left(p-p'\right)+\frac{\left(\vec{p}-\vec{p}'\right)\cdot\vec{q}}{m_e}\frac{\partial\delta\left(p-p'\right)}{\partial p'}\right\}\left\{f\left(\vec{p}'\right)-f(\vec{p})\right\}.$$

To proceed further, we need the amplitude for Compton scattering. This can be computed using Feynman rules as explicated for example in Bjorken and Drell (1965). We will take it to be constant:

$$|\mathcal{M}|^2=8\pi\sigma_Tm_\epsilon^2$$

where σ_T is the Thomson cross-section. This is wrong, and it is wrong for two reasons. First of all, the amplitude squared has an angular dependence $\propto(1+\cos^2[\hat{p}\cdot\hat{p}\prime])$. Ignoring this angular dependence, as I now propose to do, makes a small difference in the final collision term. It needs to be included in calculations which aspire to 1% accuracy. But it would simply distract us here, so let us ignore it for the present. The second reason a constant amplitude is wrong is a little more subtle and, when properly accounted for, opens up a whole new branch of CMB study. In particular, the amplitude squared has a polarization dependence $\left(\propto\left|\hat{\epsilon}\cdot\hat{\epsilon}'\right|^2\right.$, where $\hat{\epsilon}$ and $\hat{\epsilon}'$ are the polarizations of the incoming and outgoing photons) which I have implicitly summed over here. The dependence on polarization means that at a small level the CMB will be polarized due to Compton scattering (Bond and Efstathiou, 1984; Polnarev, 1985). It turns out that the information carried by the polarization spectrum is as valuable as that carried by the temperature spectrum (Seljak, 1997; Seljak and Zaldarriaga, 1997; Kamionkowski, Kosowsky, and Stebbins, 1997a,b). We will devote considerable time in Chapter 10 to understanding polarization. Even if we were not concerned with polarization, the temperature anisotropies are coupled to the polarization field, so an accurate determination of the former requires a treatment of the latter. Again, though, I will neglect this small effect here in the derivation of the collision term. It is straightforward to include both the effects of polarization and the angular dependence of Compton scattering using the same formalism we are now in the midst of. The algebra is simply a bit more tedious.

Once we have assumed that $|\mathcal{M}|^2$ is constant, we can multiply out the terms in brackets in Eq. (4.49) keeping only terms first order in energy transfer. Also, the $\vec{q}$ integral simply gives a factor of n_e (or $n_e\vec{v}_{\rm b}$ for the term which has a factor of $\vec{q}/m_e$). So

$$
\begin{aligned}
C[f(\vec{p})] =& \frac{2\pi^2 n_e \sigma_T}{p} \int \frac{d^3p'}{(2\pi)^3 p'} \left\{ \delta\left(p-p'\right) + \left(\vec{p}-\vec{p}'\right)\cdot\vec{v}_{\rm b} \frac{\partial \delta\left(p-p'\right)}{\partial p'} \right\} \\
& \times \left\{ f\left(\vec{p}'\right) - f(\vec{p}) - p' \frac{\partial f^{(0)}}{\partial p'} \Theta\left(\hat{p}'\right) + p \frac{\partial f^{(0)}}{\partial p} \Theta(\hat{p}) \right\} \\
=& \frac{n_e \sigma_T}{4\pi p} \int_0^\infty dp' p' \int d\Omega' \left[\delta\left(p-p'\right) \left(-p' \frac{\partial f^{(0)}}{\partial p'} \Theta\left(\hat{p}'\right) + p \frac{\partial f^{(0)}}{\partial p} \Theta(\hat{p}) \right) \right. \\
& \left. + \left(\vec{p}-\vec{p}'\right)\cdot\vec{v}_{\rm b} \frac{\partial \delta\left(p-p'\right)}{\partial p'} \left(f^{(0)}\left(p'\right) - f^{(0)}(p) \right) \right]
\end{aligned}
$$

where Ω' is the solid angle spanned by the unit vector $\hat{p}'$. On the first line, I have broken up the difference $f\left(\vec{p}'\right) - f(\vec{p})$ into a zero-order piece, [3] which doesn't contribute when multiplying $\delta\left(p-p'\right)$, and a first-order part which can be neglected when multiplying the velocity term.

There are only two terms in Eq. (4.51) which depend on $\hat{p}'$ and therefore which must be accounted for when integrating over solid angle Ω'. First, there is the perturbation to the distribution function, $\Theta\left(\hat{p}'\right)$. It is convenient at this stage to introduce the notation

$$
\Theta_0(\vec{x},t) := \frac{1}{4\pi} \int d\Omega' \Theta\left(\hat{p}', \vec{x}, t\right)
$$

So Θ_0 does not depend on the direction vector; it is an integral of the perturbation over all directions. In other words, it is the monopole part of the perturbation. Note that we cannot absorb this monopole into the definition of the zero order temperature since the latter is constant over all space. The perturbation Θ_0 therefore represents the deviation of the monopole at a given point in space from its average in all space. Later on we will generalize Eq. (4.52) to all other multipoles.

The second term in Eq. (4.51) which depends on $\hat{p}'$ is the explicit factor $\hat{p}'\cdot\vec{v}_{\rm b}$. This term integrates to zero since $\vec{v}_{\rm b}$ is a fixed vector. Thus, the integration over solid angle leaves

$$
\begin{aligned}
C[f(\vec{p})] = \frac{n_e \sigma_T}{p} \int_0^\infty dp' p' & \left[\delta\left(p-p'\right) \left(-p' \frac{\partial f^{(0)}}{\partial p'} \Theta_0 + p \frac{\partial f^{(0)}}{\partial p} \Theta(\hat{p}) \right) \right. \\
& \left. + \vec{p}\cdot\vec{v}_{\rm b} \frac{\partial \delta\left(p-p'\right)}{\partial p'} \left(f^{(0)}\left(p'\right) - f^{(0)}(p) \right) \right].
\end{aligned}
$$

Now the p' integral can be done: in the first line by trivially integrating over the delta function and in the second by integrating by parts. We are left with

[3] Note that we are expanding in two small quantities simultaneously, the small perturbations and the small energy transfer. Here, we are breaking up $f(\vec{p}) - f(\vec{p})$ into terms zero and first order in the small perturbations.

$$
C[f(\vec{p})] = -p \frac{\partial f^{(0)}}{\partial p} n_e \sigma_T \left[\Theta_0 - \Theta(\hat{p}) + \hat{p}\cdot\vec{v}_b \right]
$$

Already, we can anticipate the effect of Compton scattering on the photon distribution. In the absence of a bulk velocity for the electrons ($v_{\rm b} = 0$), the collision terms serve to drive Θ to Θ_0. That is, when Compton scattering is very efficient, only the monopole perturbation survives; all other moments are washed out (Figure 4.3). Intuitively, strong scattering means that the mean free path of a photon is very small. Therefore, photons arriving at a given point in space last scattered off very nearby electrons if Compton scattering is efficient. These nearby electrons most likely had a temperature very similar to the point of observation. Therefore,

photons from all directions have the same temperature. This is the characteristic signature of a monopole distribution: the temperature on the sky is uniform.

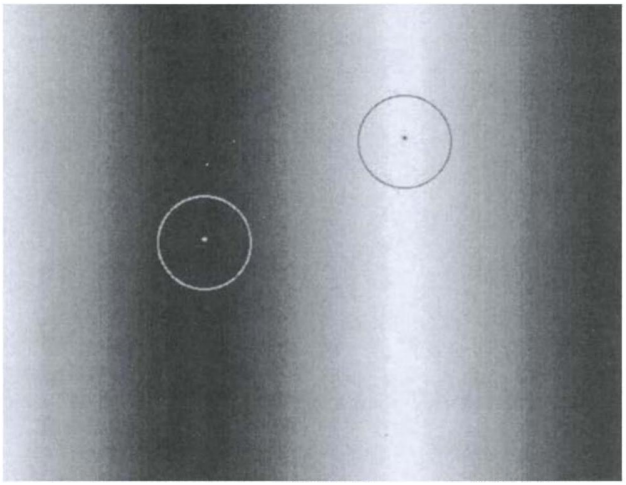


Figure 4.3. A plane wave perturbation in the matter and its effect on tightly coupled photons. Dark (white) regions represent hot (cold) spots in the electron temperature. If Compton scattering is very efficient then photons last scattered very near the point of observation. Circles denote last scattering surfaces for observation points indicated by stars. The temperature on these surfaces is very close to uniform, so the distribution is almost all monopole. Note though that different circles (corresponding to different observers) have different temperatures due to the perturbation. So the monopole varies in space.

The situation changes slightly if the electrons carry a bulk velocity. In that case, the photons will also have a dipole moment, fixed by the amplitude and direction of the electron velocity. Even in this case, though, all higher moments will vanish. Thus Compton scattering produces a photon distribution which is extremely simple to categorize: it has only a nonvanishing monopole and dipole. This is equivalent to saying that the photons behave like a fluid. Indeed, strong scattering, or tight coupling, produces a situation wherein the photons and electrons behave as a single fluid.

12.4.4 The Boltzmann Equation for Photons

We can now collect the left- and right-hand sides of the Boltzmann equations from the previous two sections. A few more definitions will complete the first goal of this chapter, a linear equation for the perturbation to the photon distribution. Equating Eqs. (4.42) and (4.54) leads to

$$\frac{\partial \Theta}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial \Theta}{\partial x^i} + \frac{\partial \Phi}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial \Psi}{\partial x^i} = n_e \sigma_T \left[\Theta_0 - \Theta + \hat{p} \cdot \vec{v}_{\rm b}\right]$$

At this point, it is convenient to reintroduce the conformal time η, defined in Eq. (2.41), as our time variable. In terms of the conformal time, the Boltzmann equation becomes

$$\dot{\Theta} + \hat{p}^i \frac{\partial \Theta}{\partial x^i} + \dot{\Phi} + \hat{p}^i \frac{\partial \Psi}{\partial x^i} = n_e \sigma_T a \left[\Theta_0 - \Theta + \hat{p} \cdot \vec{v}_{\rm b}\right]$$

Here, and from now on, overdots represent derivatives with respect to conformal time.

Equation (4.56) is a partial differential linear equation coupling Θ to other variables Φ, Ψ, and $\vec{v}_{\rm b}$ which also evolve linearly. If we Fourier transform all these variables, so that $\partial/\partial x^i \to k_i \, (:= k^i)$, the resulting Fourier amplitudes obey ordinary differential equations, which are much simpler to solve. In the case of small perturbations around a smooth universe, there is an added

benefit of Fourier transforming. Since the background is smooth, the only $\vec{x}$ dependence in Eq. (4.56) is hidden in the perturbation variables themselves. In general, an equation of the form

$$aA(\vec{x}) = bB(\vec{x})$$

gets transformed into

$$a\tilde{A}(\vec{k}) = b\tilde{B}(\vec{k})$$

That is, every Fourier mode evolves independently: $A\left(\vec{k}_1\right)$ can be evolved even if we know nothing of $A\left(\vec{k}_2\right)$. So the Fourier transform of Eq. (4.56) produces a set of ordinary differential equations for the Fourier modes, and this set of equations is uncoupled. Instead of solving an infinite number of coupled equations, we can solve for one k-mode at a time.

Note that this simplification arises because the perturbations are small (equivalently the equations are linear). In this case, the different Fourier modes all evolve independently. Perturbations to the CMB remain small at all cosmological epochs, so Fourier transforms are very useful. In contrast, perturbations to matter are more complicated. Initially they are small. and they remain small until relatively recently. The largest scales today are still in the linear regime, so Fourier transforming is certainly useful for the matter perturbations as well. However, to completely characterize the matter field today requires accounting for nonlinearities, and for this purpose, Fourier transforms lose much of their appeal. Different Fourier modes couple when nonlinear behavior becomes important, so the codes which follow matter perturbations all the way until today work in real space. Even these codes, however, start at $z \sim 20$ with the initial conditions set by linear evolution.

Our Fourier convention will be

$$\Theta(\vec{x}) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{x}} \tilde{\Theta}(\vec{k})$$

We will often characterize a mode by the magnitude of its wavevector [4] : $k = \sqrt{k^i k^i}$.

Before rewriting Eq. (4.56) in terms of Fourier modes, let us make two final definitions. First, define the cosine of the angle between the wavenumber $\vec{k}$ and the photon direction $\hat{p}$ to be

$$\mu := \frac{\vec{k}\cdot\hat{p}}{k}$$

From now on, μ will be the variable describing the direction of photon propagation. A good way to think of μ is to go back to Figure 4.3. The wavevector $\vec{k}$ is pointing in the direction in which the temperature is changing, so it is perpendicular to the gradient ($\vec{k}$ is horizontal in the figure). When $\mu = 1$ then the photon direction is aligned with $\vec{k}$, so the photon is traveling in the direction along which the temperature is changing. A photon traveling in a direction in which the temperature remains the same (vertically in the figure) has $\mu = 0$. We will typically assume that the velocity points in the same direction as $\vec{k}$ (this is equivalent to saying that the velocity is irrotational), so $\tilde{\vec{v}}_{\rm b}\cdot\hat{p} = \tilde{v}_{\rm b}\mu$. Next, we define the optical depth

$$\tau(\eta) := \int_{\eta}^{\eta_0} d\eta' n_e \sigma_T a$$

At late times, the free electron density is small, so $\tau \ll 1$, while at early times, it is very large. Note that I have defined the limits of integration in such a way that

$$\dot{\tau} := \frac{d\tau}{d\eta} = -n_e \sigma_T a$$

With these definitions, we are finally left with

$$\dot{\tilde{\Theta}} + ik\mu\tilde{\Theta} + \dot{\tilde{\Phi}} + ik\mu\tilde{\Psi} = -\dot{\tau}\left[\tilde{\Theta}_0 - \tilde{\Theta} + \mu\tilde{v}_{\rm b}\right].$$

[4] Note that k^i is a 3D vector in Eudclidean space so that $k_i = k^i$: you do not need a factor of g_{ij} to go back and forth. The same goes for the velocity v_b^i.

12.4.5 The Boltzmann Equation for Cold Dark Matter

We can apply the formalism developed in the previous sections to derive the Boltzmann equation for any other constituent in the universe. Of particular importance is the evolution of the dark matter. In almost all currently popular models of structure formation, dark matter plays an important role in structure formation and in determining the gravitational field in the universe.

It is perhaps simplest to derive the evolution equations for dark matter by imposing conservation of the energy-momentum tensor, as we did in Chapter 2 in the homogeneous case. Unlike the photons, the dark matter always behaves like a fluid so can always be described completely by $T_{\mu\nu}$. Nonetheless, here we will sacrifice simplicity and use the Boltzmann formalism to derive the dark matter equations. This will (i) reinforce the calculations of the previous sections and also (ii) pave the way for the electron/proton equations of the next section.

There are several ways in which the dark matter distribution differs from that of the photons. First, by definition, "dark" matter does not interact with any of the other constituents in the universe. Thus we need not deal with any collision terms. Second, cold dark matter, in contrast to the photons, is nonrelativistic. So we need to redo some of the kinematics which led to the left side of the Boltzmann equation. In particular, the constraint Eq. (4.11) now becomes

$$g_{\mu\nu}P^\mu P^\nu = -m^2$$

where m is the mass of the dark matter particle. It is also useful to define the energy as

$$E := \sqrt{p^2 + m^2}$$

where p is defined exactly as in Eq. (4.13): $p^2 = g_{ij}P^iP^j$. In the massless case, of course, Eq. (4.65) says that $E = p$, so E is superfluous. Here it will be convenient to let E replace p as one of the variables on which the distribution function depends (in addition to position $\vec{x}$, time t, and the direction vector $\hat{p}$). We can now derive the equivalent of equations (4.14) and (4.19) for the four-momentum of a massive particle:

$$P^\mu = \left[E(1-\Psi), p\hat{p}^i\frac{1-\Phi}{a}\right].$$

Only the time component is different from that of a massless particle, with E replacing p.

Using E as one of the dependent variables means that the total time derivative of the dark matter distribution function $f_{\rm dm}$ is

$$\frac{df_{\rm dm}}{dt} = \frac{\partial f_{\rm dm}}{\partial t} + \frac{\partial f_{\rm dm}}{\partial x^i}\frac{dx^i}{dt} + \frac{\partial f_{\rm dm}}{\partial E}\frac{dE}{dt} + \frac{\partial f_{\rm dm}}{\partial \hat{p}^i}\frac{d\hat{p}^i}{dt}.$$

Once again, the last term here vanishes since it is the product of two first-order terms. Because of the change in the constraint equation, the coefficients of the derivatives of the distribution function with respect to x^i and E are slightly different than they were in the

massless case. Working through the algebra, which is otherwise identical to the calculation presented in Section 4.2, leads to the collisionless Boltzmann equation for nonrelativistic matter:

$$\frac{\partial f_{\rm dm}}{\partial t} + \frac{\hat{p}^i}{a}\frac{p}{E}\frac{\partial f_{\rm dm}}{\partial x^i} - \frac{\partial f_{\rm dm}}{\partial E}\left[\frac{da/dt}{a}\frac{p^2}{E} + \frac{p^2}{E}\frac{\partial \Phi}{\partial t} + \frac{\hat{p}^i p}{a}\frac{\partial \Psi}{\partial x^i}\right] = 0$$

Equation (4.68) reduces to Eq. (4.33) in the massless limit as it must. The main difference between the two is the presence of factors of p/E, or velocity. For dark matter particles, these velocity factors suppress any free streaming, as we will shortly see.

In the massless case, to proceed further we used our knowledge of the distribution function. Namely, we knew that the zero-order distribution function was Bose-Einstein, and we perturbed around this zero-order solution. For cold dark matter particles, we do not need such detailed information about the zero-order distribution function. All we need to know is that these particles are very nonrelativistic. So we can neglect the thermal motion of the dark matter (Exercise 9). We cannot however neglect p/m completely, because the density perturbations themselves induce velocity flows in the dark matter via the continuity equation. These coherent flows give rise to $p/m \sim v$ terms, which must be retained. What we can do, however, in our linear treatment, is to neglect terms second-order in p/E.

Instead of assuming a form for $f_{\rm dm}$, we will take moments of Eq. (4.68). First, multiply both sides by the phase space volume $d^3p/(2\pi)^3$ and integrate. This leads to

$$\frac{\partial}{\partial t}\int \frac{d^3p}{(2\pi)^3} f_{\rm dm} + \frac{1}{a}\frac{\partial}{\partial x^i}\int \frac{d^3p}{(2\pi)^3} f_{\rm dm}\frac{p\hat{p}^i}{E} - \left[\frac{da/dt}{a} + \frac{\partial \Phi}{\partial t}\right]\int \frac{d^3p}{(2\pi)^3}\frac{\partial f_{\rm dm}}{\partial E}\frac{p^2}{E}$$
$$-\frac{1}{a}\frac{\partial \Psi}{\partial x^i}\int \frac{d^3p}{(2\pi)^3}\frac{\partial f_{\rm dm}}{\partial E}\hat{p}^i p = 0$$

Note that, since they are independent variables, the integral over p passes through the partial derivatives with respect to x^i and t. The last term here can be neglected since the integral over the direction vector is nonzero only for the perturbed part of $f_{\rm dm}$. Thus the integral is first order and it multiplies the first-order term $\partial\Psi/\partial x^i$. The rest of the terms are all relevant, though. To simplify, let us recall that the dark matter density is [5]

$$n_{\rm dm} = \int \frac{d^3p}{(2\pi)^3} f_{\rm dm}$$

while the velocity is defined as

$$v^i := \frac{1}{n_{\rm dm}}\int \frac{d^3p}{(2\pi)^3} f_{\rm dm}\frac{p\hat{p}^i}{E}$$

[5] Here I have incorporated the spin degeneracy $g_{\rm dm}$ into the phase space distribution $f_{\rm dm}$. We implicitly did the same thing in the last section for the electrons. The first two terms in Eq. (4.69), then. can be simply expressed in terms of the velocity and the density. The third term is a bit more subtle: to relate it to the density, we need to integrate by parts. Since $dE/dp = p/E$, the integrand can be reexpressed as $p\partial f_{\rm din}/\partial p$. Thus, the integral becomes

$$\begin{aligned}\int \frac{d^3p}{(2\pi)^3} p\frac{\partial f_{\rm dm}}{\partial p} &= \frac{4\pi}{(2\pi)^3}\int_0^\infty dp p^3\frac{\partial f_{\rm dm}}{\partial p} \\ &= -3\frac{4\pi}{(2\pi)^3}\int_0^\infty dp p^2 f_{\rm dm} \\ &= -3n_{\rm dn}.\end{aligned}$$

So the zeroth moment of the Boltzmann equation leads to the cosmological generalization of the continuity equation:

$$\frac{\partial n_{\rm dm}}{\partial t} + \frac{1}{a}\frac{\partial\left(n_{\rm dm}v^i\right)}{\partial x^i} + 3\left[\frac{da/dt}{a} + \frac{\partial\Phi}{\partial t}\right] n_{\rm dm} = 0.$$

The first two terms here are the standard continuity equation from fluid mechanics. The last term arises due to the FRW metric and its perturbations.

To go further, we can collect zero-order and first-order terms in Eq. (4.73). The velocity is first order as is Φ, so the only zero-order terms are

$$\frac{\partial n^{(0)}_{\rm dm}}{\partial t} + 3\frac{da/dt}{a} n^{(0)}_{\rm dm} = 0$$

where $n^{(0)}_{\rm dm}$ is the zero-order, homogeneous part of the density. Equivalently, we have

$$\frac{d\left(n^{(0)}_{\rm dm}a^3\right)}{dt} = 0 \Longrightarrow n^{(0)}_{\rm dm} \propto a^{-3}$$

a relation we anticipated back in Chapter 1 as an obvious ramification of the expansion. We also proved this scaling in Chapter 2 by using the conservation of the energy momentum tensor.

Now let us extract the first-order part of Eq. (4.73). All factors of $n_{\rm dm}$ multiplying the first-order quantities v and Φ may be set to $n^{(0)}_{\rm dm}$. Everywhere else, we need to expand $n_{\rm dm}$ out to include a first-order perturbation. In particular, we will set

$$n_{\rm dm} = n^{(0)}_{\rm dm}[1 + \delta(\vec{x}, t)].$$

which defines the first-order piece as $n^{(0)}_{\rm dm}\delta$. Since the energy density of matter is equal to mass times n, δ is also the fractional overdensity, $\delta\rho/\rho$, of the dark matter. After dividing by $n^{(0)}_{\rm dm}$, the first-order equation is therefore

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\frac{\partial v^i}{\partial x^i} + 3\frac{\partial\Phi}{\partial t} = 0.$$

As it stands, we have introduced two new perturbation variables for the dark matter, the density perturbation δ and the velocity $\vec{v}$. Equation (4.77) is only one equation, though, for these two variables. We need another. To get it. we return to the unintegrated Boltzmann equation (4.68). We have just taken its zeroth moment: to extract a second equation, let us take its first moment. In particular, multiply Eq. (4.68) by $d^3p(p/E)\hat{p}^j/(2\pi)^3$ and then integrate. The first moment equation is then

$$\begin{aligned} 0 = &\frac{\partial}{\partial t}\int\frac{d^3p}{(2\pi)^3} f_{\rm dm}\frac{p\hat{p}^j}{E} + \frac{1}{a}\frac{\partial}{\partial x^i}\int\frac{d^3p}{(2\pi)^3} f_{\rm dm}\frac{p^2\hat{p}^i\hat{p}^j}{E^2} \\ &- \left[\frac{da/dt}{a} + \frac{\partial\Phi}{\partial t}\right]\int\frac{d^3p}{(2\pi)^3}\frac{\partial f_{\rm dm}}{\partial E}\frac{p^3\hat{p}^j}{E^2} - \frac{1}{a}\frac{\partial\Psi}{\partial x^i}\int\frac{d^3p}{(2\pi)^3}\frac{\partial f_{\rm dm}}{\partial E}\frac{\hat{p}^i\hat{p}^j p^2}{E} \end{aligned}$$

The first two terms are straightforward: the first is the time derivative of $n_{\rm dm}v^i$ while the second can be safely neglected since it is of order $\langle(p/E)^2\rangle$. The last sets of terms must be handled more carefully, though, because of the partial derivatives. Since $(p/E)\partial/\partial E = \partial/\partial p$ the third term is actually of order p/E while the last is independent of velocity. Let us do the integration by parts explicitly in the third term. The integral is:

$$\int \frac{d^3p}{(2\pi)^3}\frac{\partial f_{\rm dm}}{\partial p}\frac{p^2\hat{p}^j}{E} = \int \frac{d\Omega\hat{p}^j}{(2\pi)^3}\int_0^\infty dp\frac{p^4}{E}\frac{\partial f_{\rm dm}}{\partial p}$$
$$= -\int \frac{d\Omega\hat{p}^j}{(2\pi)^3}\int_0^\infty dp f_{\rm dm}\left(\frac{4p^3}{E}-\frac{p^5}{E^3}\right)$$

The p^5/E^3 term is completely negligible, so the only relevant contribution to the integral comes from the $-4p^3/E$ term: its integral is $-4n_{\rm dm}v^j$. The same steps carry through for the last term in Eq. (4.78); the one additional fact we need is that

$$\int d\Omega\hat{p}^i\hat{p}^j = \delta^{ij}\frac{4\pi}{3}$$

So the first moment of the Boltzmann equation is

$$\frac{\partial\left(n_{\rm dm}v^j\right)}{\partial t} + 4\frac{da/dt}{a}n_{\rm dm}v^j + \frac{n_{\rm dm}}{a}\frac{\partial\Psi}{\partial x^j} = 0.$$

This equation has no zero-order parts, since the velocity is a first-order quantity. Therefore, we need extract only the first-order terms, which allows us to set $n_{\rm dm}\to n_{\rm dm}^{(0)}$ everywhere. Using the time dependence we found in Eq. (4.75) we arrive at

$$\frac{\partial v^j}{\partial t} + \frac{da/dt}{a}v^j + \frac{1}{a}\frac{\partial\Psi}{\partial x^j} = 0$$

Equations (4.77) and (4.82) are the two equations governing the evolution of the density and the velocity of the cold, dark matter. The momentum conservation equation (4.82) does not have the standard $(\vec{v}\cdot\nabla)\vec{v}$ term, since any term with two factors of v is manifestly second order. An interesting feature of the two equations is generic to this process of integrating the Boltzmann equations to get the fluid equations. Note that the equation for the density depends on the next highest moment, the velocity. This is general: the integrated Boltzmann equation for the lth moment depends on the $l+1$ moment. In principle, then, this process of integrating leads to an infinite heirarchy of equations for the moments of the distribution function. Indeed, we will see that this is one way of solving the Boltzmann equation for the photons, Eq. (4.63), which we have not yet integrated over. One might expect, then, that the velocity equation would depend on the next highest moment, the quadrupole, of the dark matter distribution. Why doesn't it? The answer lies in our assumption that the dark matter is cold. We have explicitly dropped all terms of order $(p/E)^2$ and higher. These terms correspond to the higher moments of the distribution, but since we are dealing with cold, dark matter they are irrelevant. Thus, the set of two equations, (4.77) and (4.82), are a closed set of equations for the cold, dark matter distribution. [6] If we were interested in dark matter particles with much smaller masses, such as massive neutrinos, we would need to keep these higher moments.

Let us finally rewrite Eqs. (4.77) and (4.82) in terms of conformal time η and the Fourier transforms. The density equation becomes

$$\dot{\tilde{\delta}} + ik\tilde{v} + 3\dot{\tilde{\Phi}} = 0$$

where I have assumed that the velocity is irrotational so $\tilde{v}^i = (k^i/k)\,\tilde{v}$. The velocity equation is

$$\dot{\tilde{v}} + \frac{\dot{a}}{a}\tilde{v} + ik\tilde{\Psi} = 0$$

12.4.6 The Boltzmann Equation for Baryons

The final components of the universe which require a set of Boltzmann equations are the electrons and protons. These components are often grouped together and called baryons, nomenclature which is obviously ridiculous (electrons are leptons, not baryons) but nonetheless common.

Electrons and protons are coupled by Coulomb scattering ($e + p \to e + p$). The Coulomb scattering rate is much larger than the expansion rate at all epochs of interest (Exercise 12). This tight coupling forces the electron and proton overdensities to a common value:

$$\frac{\rho_e - \rho_e^{(0)}}{\rho_e^{(0)}} = \frac{\rho_p - \rho_p^{(0)}}{\rho_p^{(0)}} := \delta_b$$

where we bow to common usage with the subscript b. Similarly the velocities of the two species are forced to a common value,

[6] Of course, we still need equations for the gravitational potentials Φ and Ψ. These come from Einstein's equations, as does the zero-order equation for a.

$$\vec{v}_e = \vec{v}_p := \vec{v}_\mathrm{b}$$

We need to derive equations then for δ_b and v_b. The starting point will be the unintegrated equations for electrons and protons:

$$\frac{df_e(\vec{x}, \vec{q}, t)}{dt} = \langle c_{ep} \rangle_{QQ'q'} + \langle c_{e\gamma} \rangle_{pp'q'}$$
$$\frac{df_p(\vec{x}, \vec{Q}, t)}{dt} = \langle c_{ep} \rangle_{qq'Q'} \, .$$

The notation here is more compact, and therefore more deceiving, than that in previous sections. We will need this compactness in what follows, so let's walk through it slowly. First, notice that initial and final momenta for the photon are $\vec{p}$ and $\vec{p}'$; for electron $\vec{q}$ and $\vec{q}'$; and the proton has been assigned $\vec{Q}$ and $\vec{Q}'$. Consider the Compton collision term in the equation for the electron distribution function. I have defined the unintegrated part of the collision term as

$$c_{e\gamma} := (2\pi)^4 \delta^4 \left(p + q - p' - q'\right) \frac{|\mathcal{M}|^2}{8E(p)E\left(p'\right) E_e(q) E_e\left(q'\right)} \left\{ f_e\left(q'\right) f_\gamma\left(p'\right) - f_e(q) f_\gamma(p) \right\}$$

and the angular brackets denote integration over all momenta in the subscripts:

$$\langle (\ldots) \rangle_{pp'q'} := \int \frac{d^3p}{(2\pi)^3} \int \frac{d^3p'}{(2\pi)^3} \int \frac{d^3q'}{(2\pi)^3} (\ldots)$$

The Coulomb collision term is similar, the main difference being the amplitude for the two processes.

In principle, Eq. (4.88) should contain a term accounting for scattering of protons off photons. In practice, though, the cross section for this process is much smaller than for Compton scattering off electrons (in each case the cross section is inversely proportional to the mass squared). So the interactions of the combined electron-proton fluid with the photons is driven by Compton scattering of electrons, and the proton-photon process can be ignored. Also, in principle, we should include ionization and recombination terms in Eqs. (4.87) and (4.88). These however would merely distract us here, so we treat all electrons as ionized.

With this notation defined, we can now proceed and derive equations for δ_b and v_b. First, multiply both sides of Eq. (4.87) by the phase space volume $d^3q/(2\pi)^3$ and integrate. The left-hand side then becomes identical to the left-hand side we derived for dark matter in Eq. (4.73). So we can immediately write

$$\frac{\partial n_e}{\partial t} + \frac{1}{a}\frac{\partial \left(n_e v_{\rm b}^i\right)}{\partial x^i} + 3\left[\frac{da/dt}{a} + \frac{\partial \Phi}{\partial t}\right] n_e = \left\langle c_{ep}\right\rangle_{QQ'q'q} + \left\langle c_{e\gamma}\right\rangle_{pp'q'q}$$

Both terms on the right vanish. The mathematical way to see this is to realize that the integration measure in the first term on the right, e.g., is completely symmetric under the interchange of $Q \leftrightarrow Q'$ and $q \leftrightarrow q'$. Because of the factors of the distribution function, the integrand $-c_{ep}$ - is antisymmetric under this interchange. So the full integral vanishes. More intuitively, the processes we are considering conserve electron number so they certainly cannot contribute to dn/dt. That is, the integral over $f_e\left(q'\right) f_p\left(Q'\right)$ counts the total number of electrons that are produced in Coulomb scattering. But this is obviously equal to the integral over $f_e(q) f_p(Q)$, which counts the number of electrons lost in Coulomb scattering. More generally, any time we multiply an unintegrated collision term by a conserved quantity and then integrate we will get zero.

The perturbed version of Eq. (4.91) equation is therefore identical to Eq. (4.77). Switching to Fourier space and using conformal time leads to

$$\dot{\tilde{\delta}}_b + ik\tilde{v}_{\rm b} + 3\dot{\tilde{\Phi}} = 0$$

The second equation for the baryons is obtained by taking the first moments of both Eqs. (4.87) and (4.88) and adding them together. We did something similar for the dark matter; there we first multiplied by $\vec{p}/E$ and then integrated over all momenta. Here we will take the moments by first multiplying the unintegrated equations by $\vec{q}$ (and $\vec{Q}$ for the protons) instead of by $\vec{q}/E$. Therefore, our results from the dark matter case carry over as long as we multiply them by a factor of m. The left-hand side of the integrated electron equation, for example, will look exactly like the left-hand side of Eq. (4.81) except it will be multiplied by m_e. The proton equation will be multiplied by m_p. Since the proton mass is so much larger than the electron mass, the sum of the two left-hand sides will be dominated by the protons. So, following Eq. (4.81), we have

$$\begin{aligned} m_p\frac{\partial \left(n_{\rm b} v_{\rm b}^j\right)}{\partial t} + 4\frac{da/dt}{a} m_p n_{\rm b} v_{\rm b}^j + \frac{m_p n_{\rm b}}{a}\frac{\partial \Psi}{\partial x^i} \\ = \left\langle c_{ep}\left(q^j + Q^j\right)\right\rangle_{QQ'q'q} + \left\langle c_{e\gamma} q^j\right\rangle_{pp'q'q}. \end{aligned}$$

The right-hand side here is the sum of that from both the electron and proton equations. Both equations have the Coulomb term, so it is weighted by $\vec{q}$ (by which we multiplied the electron equation) $+\vec{Q}$ (from the proton equation). Only the electron equation has the Compton term, so there is only the factor of $\vec{q}$ there. Once again we can use a conservation law, this time conservation of momentum, to argue that the integral of $c_{ep}(\vec{q}+\vec{Q})$ over all momenta vanishes. So dividing both sides by ${}^7\rho_b = m_p n_{\rm b}^{(0)}$, we are left with

$$\frac{\partial v_{\rm b}^j}{\partial t} + \frac{da/dt}{a} v_{\rm b}^j + \frac{1}{a}\frac{\partial \Psi}{\partial x^j} = \frac{1}{\rho_b}\left\langle c_{e\gamma} q^j\right\rangle_{pp'q'q}$$

Here I have used the by now familiar $n_{\rm b}^{(0)} \propto a^{-3}$ scaling to eliminate the $n_{\rm b}^{(0)}$ time derivative and three of the four factors of the da/dt term on the left.

[7] Note that convention here, which I will stick with, that ρ_b is the zero order baryon energy density. The total baryon density is therefore $\rho_b\left(1+\delta_{\rm b}\right)$. The final step is to evaluate the average momentum $\vec{q}$ in Compton scattering. As before, we can use the conservation of total momentum $\vec{q}+\vec{p}$ to argue that

$$\left\langle c_{e\gamma}\vec{q}\right\rangle_{pp'q'q} = -\left\langle c_{e\gamma}\vec{p}\right\rangle_{pp'q'q}$$

Now switch to Fourier space and multiply both sides of Eq. (4.94) by $\hat{k}^j$. Since $\hat{k}\cdot\vec{p} = p\mu$, the right-hand side of Eq. (4.94) becomes $-\langle c_{e\gamma} p\mu\rangle_{pp'q'q}/\rho_b$. We have already computed $\langle c_{e\gamma}\rangle_{p'q'q}$ in Eq. (4.54). We need simply multiply this by $p\mu$ and integrate over all $\vec{p}$ to find the right-hand side of Eq. (4.94):

$$\begin{aligned} -\frac{\langle c_{e\gamma} p\mu\rangle_{pp'q'q}}{\rho_b} &= \frac{n_e\sigma_T}{\rho_b}\int\frac{d^3p}{(2\pi)^3}p^2\frac{\partial f^{(0)}}{\partial p}\mu\left[\tilde{\Theta}_0 - \tilde{\Theta}(\mu) + \tilde{v}_{\rm b}\mu\right] \\ &= \frac{n_e\sigma_T}{\rho_b}\int_0^\infty\frac{dp}{2\pi^2}p^4\frac{\partial f^{(0)}}{\partial p}\int_{-1}^{1}\frac{d\mu}{2}\mu\left[\tilde{\Theta}_0 - \tilde{\Theta}(\mu) + \tilde{v}_{\rm b}\mu\right]. \end{aligned}$$

The integral over p can be done by integrating by parts: it is $-4\rho_\gamma$. The μ-integration over the first and third terms is straightforward (first term vanishes and second gives $v_{\rm b}/3$). The second term is the first moment of the perturbation Θ. Recall that the zeroth moment was defined as Θ_0. It makes sense therefore to define the first moment as

$$\Theta_1 := i\int_{-1}^{1}\frac{d\mu}{2}\mu\Theta(\mu)$$

where the factor of i is a convention and the definition holds in either real or Fourier space.

We now have an expression for the collision term which can be inserted into Eq. (4.94), and after switching to conformal time, we have:

$$\dot{\tilde{v}}_{\rm b} + \frac{\dot{a}}{a}\tilde{v}_{\rm b} + ik\tilde{\Psi} = \dot{\tau}\frac{4\rho_\gamma}{3\rho_b}\left[3i\tilde{\Theta}_1 + \tilde{v}_{\rm b}\right].$$

Why is there a factor of ρ_b in the denominator? That is, since photons scatter primarily off electrons, why does the total baryon density (which is dominated by protons) appear in this velocity equation? Physically, it arises from the fact that moving electrons is difficult because they are tightly coupled to protons via Coulomb scattering. If the proton was infinitely heavy, so $\rho_b \to \infty$, Compton scattering would not change the electron velocity at all; it would not have any impact on the combined proton-electron fluid. We derived Eq. (4.98) by setting $n_e = n_p = n_b$, but it turns out to be valid even if there is an appreciable amount of neutral hydrogen, so that $n_e \neq n_b$. Indeed after recombination, most protons are bound in neutral hydrogen atoms. And even before recombination, a small fraction are in helium atoms or ions. You might be tempted therefore to replace ρ_b in the denominator of Eq. (4.98) by the density of free protons. In fact, though, even neutral hydrogen and helium are tightly coupled to electrons and protons (see Exercise 12), so all baryons should be included. Equation (4.98) quite generally governs the evolution of the baryon velocity.

SUGGESTED READING

In the 1960s a national magazine ran a cartoon showing dozens of businessmen and -women walking the streets of Manhattan looking very important and serious. Thought bubbles over each head revealed their true focus: each was imagining a raucus sex scene. In at least some ways, the Boltzmann equation plays a similar role for physicists and astronomers: no one ever talks about it, but everyone is always thinking about it.

Two excellent astronomy textbooks which do make abundant use of the Boltzmann equation - either explicitly or implicitly - are Radiative Processes in Astrophysics (Rybicki and Lightman) and Galactic Dynamics (Binney and Tremaine). In the context of cosmology, in addition to the

books mentioned in Chapter 1, The Large Scale Structure of the Universe (Peebles), written by the field's pioneer, uses the Boltzmann equation extensively, working in synchronous gauge. If you struggled through Section 4.3, you will be amused (angered?) to see §92 of Peebles' book, where he takes much less space to derive terms due to Compton scattering.

A number of papers deriving the Boltzmann equation for cosmological perturbations are well worth reading. There is the path-breaking work by Lifshitz (1946), Peebles and Yu (1970), and Bond and Szalay (1983). A nice review was written by Efstathiou (1990). The treatment of Compton scattering presented here is based on Dodelson and Jubas (1995). If you were to read just one paper in this area, I would recommend Ma and Bertschinger (1995), which skips many of the steps presented here but has all the relevant formulae and the added virtue of equations in both conformal Newtonian and synchronous gauges. For derivation of the polarization terms in the Boltzmann equations, see Kosowsky (1996). The first paper to present the Boltzmann equation for tensors was Crittenden et al. (1993).

We will not spend too much time in this book on different gauges or on the decomposition of perturbations into scalar, vector, and tensor parts. Two excellent review articles which discuss both of these topics in detail are Mukhanov, Feldman, and Brandenberger (1992) and Kodama and Sasaki (1984). Both of these are also very good on the subjects of the next two chapters, the perturbed Einstein equations and inflation.

D-1.

Derive the fluid equations for the collisionless, one-dimensional harmonic oscillator by taking the moments of Eq. (4.6). The relevant quantities are the number density and the velocity defined as integrals over the distribution function:

$$n := \int_{-\infty}^{\infty} \frac{dp}{2\pi} f \quad ; \quad v := \frac{1}{n} \int_{-\infty}^{\infty} \frac{dp}{2\pi} \frac{p}{m} f$$

D-2.

The metric in a synchronous gauge is

$$g_{00}(\vec{x}, t) = -1$$

$$g_{0i}(\vec{x}, t) = 0$$
$$g_{ij}(\vec{x}, t) = a^2 \left[\delta_{ij} + h_{ij}\right],$$

with perturbations

$$h_{ij} = \begin{pmatrix} -2\tilde{\eta} & 0 & 0 \\ 0 & -2\tilde{\eta} & 0 \\ 0 & 0 & h + 4\tilde{\eta} \end{pmatrix}$$

where $\tilde{\eta}$ has nothing to do with conformal time. Here I have chosen the wavevector $\vec{k}$ to lie in the $\hat{z}$ direction. Derive the equivalent of Eq. (4.63) in synchronous gauge:

$$\dot{\Theta} + ik\mu\Theta - \frac{\mu^2 \dot{h}}{2} - \mathcal{P}_2(\mu)\dot{\tilde{\eta}} = -\dot{\tau}\left[\Theta_0 - \Theta + \mu v\right].$$

D-3.

Start from the zero-order unintegrated Boltzmann equation (4.37). Integrate this equation over all momenta to show that the number density falls off as a^3. In the course of this, you will have justified the left-hand side of Eq. (3.1).

D-4.

Show that the Pauli blocking factor $1 - f_e$ can be set to 1 for all epochs of interest. First find f_e as a function of temperature and number density using the results/approximations of Section 3.1 (i.e. assume that $T_e \ll m_e$). Then, show that as long as the temperature is much less than m_e, f_e is much less than 1.

D-5.

Suppose we started this chapter by writing

$$\frac{df}{d\lambda} = C'$$

Change from this form to the one in Eq. (4.1) (with df/dt) on the left. How is the collision term here, C' related to C in Eq. (4.1)? Argue that the first-order perturbations in the factor relating the two collision terms can be dropped since the collision terms themselves are first-order.

D-6.

Derive Eq. (4.68), the unintegrated Boltzmann equation for a massive particle.

D-7.

Account for the angular dependence of Compton scattering. Start from Eq. (4.49) but instead of assuming the amplitude is constant, take

$$|\mathcal{M}|^2 = 6\pi\sigma_T m_e^2 \left(1 + \cos^2\left[\hat{p} \cdot \hat{p}'\right]\right).$$

Show that correctly accounting for the angular dependence introduces the factor of $(1/2)\mathcal{P}_2(\mu)\Theta_2$ presented in Eq. (4.100).

D-8.

Show that the temperature of nonrelativistic matter scales as a^{-2} in the absence of interactions. Start from the zero-order part of Eq. (4.68) and assume a form $f_{\rm dm} \propto e^{-E/T} = e^{-p^2/2mT}$. Note that this argument does not apply to electrons and protons: as long as they are coupled to the photons, their temperature scales as a^{-1}.

D-9.

In Exercise 8, you showed that a thermal distribution of nonrelativistic particles which do not interact has a temperature which scales as a^{-2}, as opposed to that of relativistic particles which we have seen scales as a^{-1}. So $T_{\rm dm} \propto T^2$. Fix the normalization by requiring $T_{\rm dm} = T$ when each is equal to the dark matter mass. Estimate the typical thermal velocity of a dark matter particle with mass equal to 100GeV when the photon temperature is 1eV.

D-10.

The purpose of this problem is to derive the results of Section 2.3 using the Boltzmann equation. Multiply the zero-order part of Eq. (4.68) by $d^3pE(p)/(2\pi)^3$ and integrate. Show that the resulting equation is identical to Eq. (2.55).

D-11.

Consider the effect of a massive neutrino on the evolution equations. (a) Start from the Boltzmann equation for a massive particle (4.68). Turn it into an equation for $\mathcal{N}$, the perturbation to the massive neutrino distribution function. Use the fact that to first order the neutrino distribution function is

$$f_\nu = f_\nu^{(0)} + \frac{\partial f_\nu^{(0)}}{\partial T_\nu} T_\nu \mathcal{N}$$

where $f_\nu^{(0)} = \left[e^{p/T_\nu} + 1\right]^{-1}$. Express the final equation in Fourier space using conformal time as the evolution variable.

(b) Recent experiments measuring the atmospheric neutrino flux suggest that the mass of the tau neutrino is 0.07eV, far larger than either the electron or muon neutrino. Find the contribution of a $0.07 - \text{eV}$ neutrino to the energy density today. You may assume it is nonrelativistic.

(c) Consider the following two scenarios. Each has energy density equal to the critical density divided up between only two components: a cold, dark matter particle and a neutrino. The neutrino in each case has the standard abundance and temperature. The only difference between the two scenarios is in one the neutrino is massless while in the other it has a mass of 0.07eV. Plot the energy density as a function of scale factor in each of the these scenarios. Note that they should agree very early on (in each case there is only a relativistic neutrino early on) and very late. The only difference comes in the middle.

D-12.

Show that ordinary matter is tightly coupled during the relevant epochs in the early universe.

(a) Compute the ratio of the Coulomb scattering rate to the Hubble rate. You may assume that all electrons and protons are ionized.

(b) Show that the rate for neutral hydrogen to scatter off ionized protons is always much larger than the expansion rate even when the ionization fraction is on the order of 10^{-1}.

D-13.

Consider tensor perturbations to the metric. These do not perturb $g_{00}(= -1)$ or $g_{0i}(= 0)$. However, the spatial part of the metric is now

$$g_{ij} = a^2 \begin{pmatrix} 1 + h_+ & h_\times & 0 \\ h_\times & 1 - h_+ & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Derive the equation for the photon distribution function in the presence of tensor perturbations. Unlike scalar perturbations. tensor perturbations induce an azimuthal dependence in Θ_l, so decompose the anisotropy due to tensors into

$$\Theta^T(k, \mu, \phi) = \Theta_+^T(k, \mu)\left(1 - \mu^2\right)\cos(2\phi) + \Theta_\times^T(k, \mu)\left(1 - \mu^2\right)\sin(2\phi).$$

Show that both the $+$ and the $\times$ component satisfy

$$\frac{d\Theta_i^T}{d\eta} + ik\mu\Theta_i^T + \frac{1}{2}\frac{dh_i}{d\eta} = \dot{\tau}\left[\Theta_i^T - \frac{1}{10}\Theta_{i,0}^T - \frac{1}{7}\Theta_{i,2}^T - \frac{3}{70}\Theta_{i,4}^T\right]$$

where i stands for either $\times$ or $+$, and the moments are defined as were the scalar moments, in Eq. (4.99).

12.5 Einstein Equations

The previous chapter set up the formalism to describe how perturbations in the gravitational field affect particle distributions. This formalism led to the set of equations (4.100) -(4.107). We need to supplement these equations with an account of how the perturbations to the particle distributions affect the gravitational field. For this, we need the Einstein equations of general relativity. The calculation detailed in this chapter expands the Einstein equations perturbatively around the zeroorder homogeneous solution. Far from being subtle or complex as one might expect from general relativity's reputation, this calculation is completely straightforward, although a bit long. Still, working through it is a "must-do-once" exercise, so the steps are presented in some detail.

12.5.1 The Perturbed Ricci Tensor and Scalar

The fundamental equation of general relativity (2.30) is a 4D tensor equation, so in principle it represents 16 separate equations. However, since both sides of the equation are symmetric tensors, only 10 of these are distinct. We are interested though in only two of the 10, since the metric we are focusing on has only two independent functions, Φ and Ψ.

Evaluating the left-hand side of the Einstein equation requires three pre-steps:

- Compute the Christoffel symbols, $\Gamma^{\mu}{}_{\alpha\beta}$, for the perturbed metric of Eq. (4.9).
- From these, form the Ricci tensor, $R_{\mu\nu}$, using Eq. (2.31).
- Contract the Ricci tensor to form the Ricci scalar, $\mathcal{R} := g^{\mu\nu}R_{\mu\nu}$.

Note that, unfortunately, even if we are interested in only several components of the Einstein equations, we need to compute all the elements of the Ricci tensor. For all the components of the Einstein tensor $G_{\mu\nu} := R_{\mu\nu} - g_{\mu\nu}\mathcal{R}/2$ depend on the Ricci scalar, which depends on all elements of $R_{\mu\nu}$.

Christoffel Symbols

We have already computed the zero-order Christoffel symbols in Eqs. (2.22) and (2.23). Now we need to look at the first-order terms, those that are linear in Φ and/or Ψ. First let us consider $\Gamma^0_{\mu\nu}$, which by definition is

$$\Gamma^0_{\mu\nu} = \frac{1}{2}g^{0\alpha}\left[g_{\alpha\mu,\nu} + g_{\alpha\nu,\mu} - g_{\mu\nu,\alpha}\right]$$

where again $_{,\alpha}$ means the derivative with respect to x^α. The only nonzero component of $g^{0\alpha}$ is the time component, [1] which is the inverse of $g_{00} = -1 - 2\Psi$. So, to first-order in the perturbations, $g^{00} = -1 + 2\Psi$, and

$$\Gamma^0_{\mu\nu} = \frac{-1+2\Psi}{2}\left[g_{0\mu,\nu} + g_{0\nu,\mu} - g_{\mu\nu,0}\right].$$

Take each component in turn: first the one with $\mu = \nu = 0$. Each of the terms in square brackets is identical, so the brackets give $g_{00,0} = -2\Psi_{,0}$. Since we are interested only in first-order terms the factor of 2Ψ out in front can be dropped and we are left with

$$\Gamma^0_{00} = \Psi_{,0}.$$

The next possibility is that one of the indices μ or ν is spatial and the other time. It doesn't matter which, since the Christoffel symbol is symmetric in its lower indices. In this case, only one of terms in brackets in Eq. (5.2) is nonzero, $g_{00,i} = -2\Psi_{,i}$. Once again since this is first-order, we can drop the factor of 2Ψ in front, leading to

$$\Gamma^0_{0i} = \Gamma^0_{i0} = \Psi_{,i} = ik_i\Psi$$

The final equality here moves to Fourier space, where we will stay from now on. Recall our convention of not using s for Fourier transformed variables: Ψ on the far right is really $\tilde{\Psi}$.

Finally, if both lower indices in Eq. (5.2) are spatial, the first two terms in brackets vanish since $g_{0i} = 0$ and only the last term survives, leaving

$$\Gamma^0_{ij} = \frac{1-2\Psi}{2}\frac{\partial}{\partial t}\left[\delta_{ij}a^2(1+2\Phi)\right]$$

There is a zero-order term here, the one we computed in Eq. (2.22), and three first-order terms:

$$\Gamma^0_{ij} = \delta_{ij}a^2\left[H + 2H(\Phi-\Psi) + \Phi_{,0}\right]$$

with $H = (da/dt)/a$.

Computing the Christoffel symbols, $\Gamma^i_{\mu\nu}$, will be left as an exercise. They are

$$\Gamma^i_{00} = \frac{ik^i}{a^2}\Psi$$

[1] We will do the calculation with $x^0 = t$, not conformal time. Therefore, $\Psi_{,0}$ for example means derivative with respect to time. Since our convention is now $\dot{\Psi} := \partial\Psi/\partial\eta$, $\Psi_{,0} = \dot{\Psi}/a$.

$$\Gamma^i_{j0} = \Gamma^i{}_{jj} = \delta_{ij}\left(H + \Phi_{,0}\right)$$
$$\Gamma^i{}_{jk} = i\Phi\left[\delta_{ij}k_k + \delta_{ik}k_j - \delta_{jk}k_i\right].$$

Note that the only nonvanishing zero-order component is $\Gamma^i{}_{j0}$, in agreement with Eq. (2.23). Also remember that both δ_{ij} and the 3 -vector k_i live in Euclidean space, so we can freely interchange their upper and lower indices.

Ricci Tensor

The Ricci tensor is most easily expressed in terms of the Christoffel symbols, as in Eq. (2.31). First, consider the time-time component:

$$R_{00} = \Gamma^\alpha_{00,\alpha} - \Gamma^\alpha_{0\alpha,0} + \Gamma^\alpha_{\beta\alpha}\Gamma^\beta_{00} - \Gamma^\alpha_{\beta 0}\Gamma^\beta_{0\alpha}.$$

All of these terms contribute at first-order. One simplification comes from considering the $\alpha = 0$ part of all these terms. The first and second terms are equal and opposite to each other as are the last two. So the sum over the index α contributes only when α is spatial. Let's consider each of the terms one by one.

- The first is

$$\Gamma^i_{00,i} = \frac{-k^2}{a^2}\Psi$$

using the first of equations (5.7).

- The second term in Eq. (5.8) is

$$-\Gamma^i_{0i,0} = -3\left(\frac{d^2a/dt^2}{a} - H^2 + \Phi,00\right)$$

using the second of equations (5.7). The factor of 3 in front comes from the implicit sum in δ_{ii}.

- The next term is $\Gamma^i{}_{i\beta}\Gamma^\beta{}_{00}$. Note that $\Gamma^\beta{}_{00}$ is first order no matter what β is, so we need keep only the zero-order part of $\Gamma^i_{i\beta}$. However, the last of equations (5.7) shows that $\Gamma^i{}_{i\beta}$ is first-order unless $\beta = 0$. So to first-order,

$$\begin{aligned}\Gamma^i{}_{i\beta}\Gamma^\beta{}_{00} &= \Gamma^i{}_{i0}\Gamma^0{}_{00}\\ &= 3H\Psi_{,0}.\end{aligned}$$

- Finally the last term is $-\Gamma^i{}_{\beta 0}\Gamma^\beta{}_{0i}$. In this case, if $\beta = 0$ both Γ 's are first-order, so their product is second-order and can be neglected. Therefore, only spatial β need be considered, leading to

$$\begin{aligned}-\Gamma^i{}_{30}\Gamma^\beta{}_{0i} &= -\Gamma^i{}_{j0}\Gamma^j{}_{0i}\\ &= -3\left(H^2 + 2H\Phi_{,0}\right).\end{aligned}$$

Collecting these four sets of terms gives

$$R_{00} = -3\frac{d^2a/dt^2}{a} - \frac{k^2}{a^2}\Psi - 3\Phi_{.00} + 3H\left(\Psi_{.0} - 2\Phi_{.0}\right).$$

Note that the zero-order term agrees with Eq. (2.34).
The space-space part of the Ricci tensor is left as an exercise. It is

$$\begin{aligned}R_{ij} =&\delta_{ij}\left[\left(2a^2H^2 + a\frac{d^2a}{dt^2}\right)(1 + 2\Phi - 2\Psi)\right.\\ &\left.+a^2H\left(6\Phi_{,0} - \Psi_{,0}\right) + a^2\Phi_{,00} + k^2\Phi\right] + k_ik_j(\Phi + \Psi).\end{aligned}$$

We can now contract the indices on the Ricci tensor and find the Ricci scalar:

$$\begin{aligned}\mathcal{R} :=&g^{\mu\nu}R_{\mu\nu} = g^{00}R_{00} + g^{ij}R_{ij}\\ =&[-1 + 2\Psi]\left[-3\frac{d^2a/dt^2}{a} - \frac{k^2}{a^2}\Psi - 3\Phi_{.00} + 3H\left(\Psi_{,0} - 2\Phi_{,0}\right)\right]\\ &+\left[\frac{1 - 2\Phi}{a^2}\right]\left[3\left\{\left(2a^2H^2 + a\frac{d^2a}{dt^2}\right)(1 + 2\Phi - 2\Psi)\right.\right.\\ &\left.\left.+a^2H\left(6\Phi_{,0} - \Psi_{,0}\right) + a^2\Phi_{.00} + k^2\Phi\right\} + k^2(\Phi + \Psi)\right].\end{aligned}$$

First let us check the zero-order part of $\mathcal{R}$. Combining terms, we find that it is $6\left(H^2 + \frac{d^2a/dt^2}{a}\right)$, in agreement with Eq. (2.37). To get the first-order part, $\delta\mathcal{R}$, we go through the by-now-familiar routine of multiplying terms, keeping only those first-order in Φ and Ψ. This gives

$$\begin{aligned}\delta\mathcal{R} =& -6\Psi\frac{d^2a/dt^2}{a} + \frac{k^2}{a^2}\Psi + 3\Phi_{,00} - 3H\left(\Psi_{,0} - 2\Phi_{,0}\right)\\ & - 6\Psi\left(2H^2 + \frac{d^2a/dt^2}{a}\right) + 3H\left(6\Phi_{.0} - \Psi_{,0}\right)\\ & + 3\Phi_{,00} + 4\frac{k^2\Phi}{a^2} + \frac{k^2\Psi}{a^2},\end{aligned}$$

where the first line contains the terms from R_{00} (the second line in Eq. (5.15)) and the last two from R_{ij} (the last two lines in Eq. (5.15)). Combining these leads to

$$\begin{aligned}\delta\mathcal{R} =& -12\Psi\left(H^2 + \frac{d^2a/dt^2}{a}\right) + \frac{2k^2}{a^2}\Psi + 6\Phi_{,00}\\ & - 6H\left(\Psi_{,0} - 4\Phi_{,0}\right) + 4\frac{k^2\Phi}{a^2}.\end{aligned}$$

12.5.2 Two Components of the Einstein Equations

We can now derive the evolution equations for Φ and Ψ, the perturbations to the Friedmann-Robertson-Walker metric. There is some freedom here because the Einstein equations

$$G^\mu_\nu = 8\pi G T^\mu_\nu$$

have 10 components and we need only two. All of the other eight components will either be zero at first-order or be redundant. [2]

The first component we will use is the time-time component. Thus we need to evaluate

$$\begin{aligned}G^0_0 &= g^{00}\left[R_{00} - \frac{1}{2}g_{00}\mathcal{R}\right]\\ &= (-1+2\Psi)R_{00} - \frac{\mathcal{R}}{2}\end{aligned}$$

Here one of the indices has been raised by multiplying G_{00} by g^{00} (recall that g^{0i} vanish). This turns out to simplify the energy-momentum tensor (see Exercise 3) which sources the Einstein tensor. Also note that the second line follows from the first since $g^{00}g_{00} = 1$. We have computed the time-time component of the Ricci tensor (Eq. (5.13)) and the perturbed Ricci scalar (Eq. (5.17)), so the first-order part of the time-time component of the Einstein tensor is

$$\begin{aligned}\delta G^0_0 =& -6\Psi\frac{d^2a/dt^2}{a} + \frac{k^2}{a^2}\Psi + 3\Phi_{,00} - 3H\left(\Psi_{,0} - 2\Phi_{,0}\right)\\ & + 6\Psi\left(H^2 + \frac{d^2a/dt^2}{a}\right) - \frac{k^2}{a^2}\Psi - 3\Phi_{,00}\\ & + 3H\left(\Psi_{.0} - 4\Phi_{,0}\right) - 2\frac{k^2\Phi}{a^2}\end{aligned}$$

Combining terms leads to

$$\delta G^0_0 = -6H\Phi_{.0} + 6\Psi H^2 - 2\frac{k^2\Phi}{a^2}$$

Einstein's equation equates G^0_0 with $8\pi G T^0_0$ where $T_{\mu\nu}$ is the energymomentum tensor. To complete our derivation of the first evolution equation for Φ and Ψ, therefore, we need to compute the first-order part of the source term, T^0_0. Recall from Section 2.3 that $-T^0_0$ is the energy density of all the particles

[2] This is true for scalar perturbations. When we come to conside tensor perturbations, some of the other components will be useful. in the universe, and that the contribution from each species is an integral over the distribution function (Eq. (2.59)),

$$T_0^0(\vec{x},t) = -\sum_{\text{all species } i} g_i \int \frac{d^3p}{(2\pi)^3} E_i(p) f_i(\vec{p},\vec{x},t).$$

Recall also that g_i is the spin degeneracy of the species (has nothing to do with the metric); $E_i = \sqrt{p^2+m_i^2}$ is the energy of a particle with momentum p and mass m_i; and f_i is the distribution function. In Section 2.3, we considered the zero-order distributions of the smooth universe. To get the first-order part of the energy-momentum tensor, we must naturally consider the first-order part of the distribution functions, i.e. the perturbation variables we defined in Chapter 4 for the photons, neutrinos, dark matter, and baryons. This is easiest for the dark matter and baryons. For we defined the right-hand side as $-\rho_i\,(1+\delta_i)$ where i labels either dark matter or baryons. For photons, a little more care is required. Using the definition of Θ in Eq. (4.35), we have

$$T_0^0 = -2\int \frac{d^3p}{(2\pi)^3} p\left[f^{(0)} - p\frac{\partial f^{(0)}}{\partial p}\Theta\right] \qquad \text{(photons)}.$$

The first term here is just the zero-order photon energy density, ρ_γ. To reduce the second term, we first do the angular integral, which picks out the monopole Θ_0 from Θ. Then, we do the radial integral by parts. This changes the sign and introduces a factor of 4 since $\partial p^4/\partial p = 4p^3$, leading to

$$T_0^0 = -\rho_\gamma\,[1+4\Theta_0] \qquad \text{(photons)}.$$

The factor of 4 here is obvious in retrospect. The perturbation variable Θ is the fractional temperature change, while the energy momentum tensor is interested in the perturbed energy density, $\delta\rho$. We should have expected that since $\rho \propto T^4$, $\delta\rho/\rho = 4\delta T/T$. In any event, it falls out of the algebra. I harp on it only to warn those who turn to the literature that authors are virtually split between those who define Θ as $\delta\rho/\rho$ and those who opt for the convention we use here. Finally, note that the first-order contribution from massless neutrinos is identical in form,

$$T_0^0 = -\rho_\nu\,[1+4\mathcal{N}_0] \qquad \text{(neutrinos)}\ .$$

In principle, we should also include a term for the perturbation to the dark energy. In practice, though, most models of the dark energy predict that (i) it should be smooth and (ii) it should be important only very recently. Both of these features are inherent in the cosmological constant model for example. There are some models which deviate from one or both of these conditions, but for the most part we are justified in neglecting the dark energy as a source of perturbations to the metric. Returning to Einstein's equation, we equate Eq. (5.21) with $8\pi G$ times the firstorder part of the time-time component of the energy-momentum tensor. Dividing both sides by 2 leads to

$$-3H\Phi_{,0} + 3\Psi H^2 - \frac{k^2\Phi}{a^2} = -4\pi G\left[\rho_{\rm dm}\delta + \rho_b\delta_{\rm b} + 4\rho_\gamma\Theta_0 + 4\rho_\nu\mathcal{N}_0\right]$$

It is again useful to write the equation in terms of conformal time. This introduces an extra factor of $1/a$ every time a time derivative appears, so

$$k^2\Phi + 3\frac{\dot{a}}{a}\left(\dot{\Phi} - \Psi\frac{\dot{a}}{a}\right) = 4\pi G a^2\left[\rho_{\rm dm}\delta + \rho_b\delta_{\rm b} + 4\rho_\gamma\Theta_0 + 4\rho_\nu\mathcal{N}_0\right]$$

This is our first evolution equation for Φ and Ψ. In the limit of no expansion ($a =$ constant), Eq. (5.27) reduces to the ordinary Poisson equation for gravity (in Fourier space). The left-hand side is $-\nabla^2\Phi$ while the right-hand side is $4\pi G\delta\rho$. The terms proportional to $\dot{a}$ account for expansion and are typically important for modes with wavelengths ($\sim a/k$) comparable to, or larger than, the Hubble radius, H^{-1}. We need this general relativistic expression when we consider the evolution of parturbations, because almost all modes of interest today used to have wavelengths larger than the Hubble radius. More on this in Chapter 6.

We now obtain a second evolution equation for Φ and Ψ. Since we have already dealt with the time-time component of the Einstein tensor, let's focus on the spatial part of G^μ_ν,

$$G^i{}_j = g^{ik}\left[R_{kj} - \frac{g_{kj}}{2}\mathcal{R}\right] = \frac{\delta^{ik}(1-2\Phi)}{a^2}R_{kj} - \frac{\delta_{ij}}{2}\mathcal{R}$$

From Eq. (5.14), we see that most of the terms in R_{kj} are proportional to δ_{kj}. When contracted with δ^{ik} this will lead to a host of terms proportional to δ_{ij}, in addition to the last term here, the one proportional to $\mathcal{R}$. Therefore,

$$G^i{}_j = A\delta_{ij} + \frac{k_ik_j(\Phi+\Psi)}{a^2}$$

where A has close to a dozen terms which we would rather not write down. Since all of these terms are proportional to δ_{ij} they all contribute to the trace of $G^i{}_j$. To avoid dealing with these terms, consider the longitudinal, traceless part of $G^i{}_j$, which can be extracted by contracting $G^i{}_j$ with $\hat{k}_i\hat{k}^j - (1/3)\delta^j_i$, a projection operator. That is, it picks out the piece which is longitudinal, traceless and only that part (Exercise 4). This projection operator kills all terms proportional to δ_{ij}, leaving only

$$\left(\hat{k}_i\hat{k}^j - (1/3)\delta^j_i\right)G^i{}_j = \left(\hat{k}_i\hat{k}^j - (1/3)\delta^j_i\right)\left(\frac{k^ik_j(\Phi+\Psi)}{a^2}\right) = \frac{2}{3a^2}k^2(\Phi+\Psi).$$

This is to be equayed with the longitudinal, traceless part of the energymomentum tensor, extracted in the same fashion:

$$\left(\hat{k}_i\hat{k}^j - (1/3)\delta^j_i\right)T^i{}_j = \sum_{\text{all species i}} g_i\int\frac{d^3p}{(2\pi)^3}\frac{p^2\mu^2 - (1/3)p^2}{E_i(p)}f_i(\vec{p}).$$

We can immediately recognize the combination $\mu^2 - 1/3$ as proportional to the second Legendre polynomial, more precisely equal to $(2/3)\mathcal{P}_2(\mu)$. Therefore, the integral picks out the quadrupole part of the distribution. Of course the zero-order part of the distribution function has no quadrupole, so the source term is first order, proportional to Θ_2, which is nonzero only for neutrinos and photons. The integral in Eq. (5.31) for photons is

$$\begin{aligned}-2\int\frac{dpp^2}{2\pi^2}p^2\frac{\partial f^{(0)}}{\partial p}\int_{-1}^{1}\frac{d\mu}{2}\frac{2\mathcal{P}_2(\mu)}{3}\Theta(\mu) &= 2\frac{2\Theta_2}{3}\int\frac{dpp^2}{2\pi^2}p^2\frac{\partial f^{(0)}}{\partial p}\\ &= -\frac{8\rho^{(0)}\Theta_2}{3}\end{aligned}$$

where the first equality follows from the definition of the quadrupole and the second from an integration by parts. This component of the energy-momentum tensor is called the anisotropic stress. Nonrelativistic particles, such as baryons and dark matter, do not contribute anisotropic stress.

For the second Einstein equation, we therefore equate Eq. (5.30) with $8\pi G$ times the photon and neutrino anisotropic stresses:

$$k^2(\Phi + \Psi) = -32\pi G a^2 \left[\rho_\gamma \Theta_2 + \rho_\nu \mathcal{N}_2\right].$$

That is, the two gravitational potentials are equal and opposite unless the photons or neutrinos have appreciable quadrupole moments. In practice, the photons' quadrupole contributes little to this sum, because it is very small during the time when it has appreciable energy density. [Recall the argument after Eq. (4.54).] Only the collisionless neutrino has an appreciable quadrupole moment early on when radiation dominates the universe.

12.5.3 Tensor Perturbations

Until now, we have focused almost exclusively on scalar perturbations to the homogeneous FRW universe. Formally, this means that the perturbations $\Phi(\vec{x}, t)$ and $\Psi(\vec{x}, t)$ transform as scalars as $\vec{x} \to \vec{x}'$; i.e., they remain unchanged under a spatial coordinate transformation. This focus is reasonable: as we have seen, scalar perturbations to the metric are sourced by density fluctuations and vice versa. For the most part, the density fluctuations that led to the structure of the universe are our primary interest.

Nonetheless, many theories of structure formation produce, in addition to scalar fluctuations, tensor perturbations to the metric. These are potentially detectable because they produce observable distortions in the CMB, especially on large scales. Sprinkled throughout the book, therefore, are exercises (with hints) relating to tensor perturbations. The tools needed to study these are precisely those we crafted when studying scalar perturbations. For the most part, therefore, I regard the evolution of tensor perturbations as one rather large homework problem, one which introduces no new physics.

One question which naturally arises when working out these exercises, though, is why consider scalar and tensor perturbations separately? To answer this question (and to alleviate the homework load) this section derives Einstein's equations for tensor perturbations. We will see that scalar and tensor perturbations decouple; that is, they evolve completely independently. So the presence of tensor perturbations does not affect the scalars and vise versa. Contrast this with Φ and Ψ. We have just shown that they are quite tightly coupled to each other. It is impossible to learn about Φ without also solving for Ψ. The decoupling of scalars and tensors is a manifestation of the decomposition theorem. Needless to say, it is much more instructive to work out an example of this theorem than to prove it abstractly. Incidentally, as you would expect, the same theorem can be applied to vector perturbations. These too are produced by some early-universe models (but not as ubiquitously as tensors) and can be treated completely independently.

Tensor perturbations can be characterized by a metric with $g_{00} = -1$, zero space-time components $g_{0i} = 0$, and spatial elements

$$g_{ij} = a^2 \begin{pmatrix} 1 + h_+ & h_\times & 0 \\ h_\times & 1 - h_+ & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

That is, the perturbations to the metric are described by two functions, $h_+$and $h_\times$, assumed small. For definiteness, I have chosen the perturbations to be in the $x - y$ plane. This corresponds to an implicit choice of axes; in particular, it corresponds to choosing the z-axis to be in the direction of the wavevector, $\vec{k}$. More generally, h_+ and $h_\times$are two components of a divergenceless, traceless, symmetric tensor. If this perturbation tensor is written as $\mathcal{H}_{ij}$, divergenceless means that $k^i \mathcal{H}_{ij} = k^j \mathcal{H}_{ij} = 0$. This is clearly satisfied by Eq. (5.34) since there are no components in the $\hat{k} = \hat{z}$ direction. Tracelessness is also satisfied since the sum of the perturbations along the diagonal vanishes.

Once the metric in Eq. (5.34) has been written down, we can blast away and derive the Einstein equations. Once again the derivation proceeds in three steps: (i) Christoffel symbols, (ii) Ricci tensor, and (iii) Ricci scalar.

Christoffel Symbols for Tensor Perturbations

First consider $\Gamma^0_{\alpha\beta}$. The metric we are considering in Eq. (5.34) has constant g_{00} and vanishing g_{0i}. Recall that the Christoffel symbol is a sum of derivatives of the metric. The only terms that will be nonzero are those which involve derivatives of the spatial part of the metric, $g_{ij,\alpha}$. Therefore, we can immediately argue that

$$\Gamma^0{}_{00} = \Gamma^0_{i0} = 0$$

The term with two lower spatial indices is

$$\begin{aligned}\Gamma^0_{ij} &= -\frac{g^{00}}{2} g_{ij,0} \\ &= \frac{1}{2} g_{ij,0}.\end{aligned}$$

Let's use the notation mentioned earlier: the 3D matrix $\mathcal{H}_{ij}$ contains the perturbations, which in this basis (with $\hat{k}$ in the $\hat{z}$ direction) is equal to

$$\mathcal{H}_{ij} = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

so that $g_{ij} = a^2 (\delta_{ij} + \mathcal{H}_{ij})$. Therefore,

$$g_{ij,0} = 2H g_{ij} + a^2 \mathcal{H}_{ij,0}$$

where the Hubble rate H is not to be confused with tensor perturbations $\mathcal{H}$. The first nonzero Christoffel symbol is therefore

$$\Gamma^0_{ij} = H g_{ij} + \frac{a^2 \mathcal{H}_{ij,0}}{2}$$

When both lower indices on Γ are 0, the Christoffel symbol vanishes. The two remaining components are $\Gamma^i{}_{0j}$ and $\Gamma^i{}_{jk}$. The former is

$$\Gamma^i_{0j} = \frac{g^{ik}}{2} g_{jk,0}$$

The time derivative of g_{jk} acts on both the scale factor and on the perturbations $h_{+,x}$, as in Eq. (5.38), so

$$\Gamma^i{}_{0j} = \frac{g^{ik}}{2} \left[2H g_{jk} + a^2 \mathcal{H}_{jk,0} \right]$$

But $g^{ik} g_{jk} = \delta_{ij}$, so the first term here is simply $\delta_{ij} H$. To get the second, we can set $g^{jk} = \delta_{jk}/a^2$ (i.e., neglect first-order terms) since it multiplies the first-order $\mathcal{H}$. So,

$$\Gamma^i_{0j} = H \delta_{ij} + \frac{1}{2} \mathcal{H}_{ij,0},$$

where I have used the fact that $\mathcal{H}_{ij}$ is symmetric.

The last Christoffel symbol we need is Γ^i_{jk}. In Exercise 7 you will show that

$$\Gamma^i{}_{jk} = \frac{i}{2} \left[k_k \mathcal{H}_{ij} + k_j \mathcal{H}_{ik} - k_i \mathcal{H}_{jk} \right]$$

Ricci Tensor for Tensor Perturbations

Following the same steps as in the scalar perturbation case, we now combine these Christoffel symbols to form the Ricci tensor. First we compute the time-time component of the Ricci tensor:

$$R_{00} = \Gamma^{\alpha}_{00,\alpha} - \Gamma^{\alpha}_{0\alpha,0} + \Gamma^{\alpha}_{\beta\alpha}\Gamma^{\beta}_{00} - \Gamma^{\alpha}_{\beta 0}\Gamma^{\beta}{}_{0\alpha}.$$

We have shown that the Christoffel symbol vanishes for tensor perturbations when the two lower indices are time-time. Therefore, the first and third terms here are zero. Using the same argument, the indices α and β in the second and fourth terms must be spatial, so

$$R_{00} = -\Gamma^{i}_{0i,0} - \Gamma^{i}{}_{j0}\Gamma^{j}{}_{0i}.$$

Using Eq. (5.42) for Γ^{i}_{j0} which is the only element appearing, we find that

$$\begin{aligned} R_{00} = &- 3\frac{dH}{dt} - \frac{1}{2}\mathcal{H}_{ii,00} \\ &- \left(H\delta_{ij} + \frac{1}{2}\mathcal{H}_{ij,0}\right)\left(H\delta_{ij} + \frac{1}{2}\mathcal{H}_{ij,0}\right). \end{aligned}$$

On the first line, the trace $\mathcal{H}_{ii}$ vanishes since h_{+}appears in the metric with opposite signs along the diagonal. Expanding the second line out to first-order leads to a similar cancellation: $\mathcal{H}_{ij}$ is multiplied by δ_{ij}, so there are no first-order terms. The zero-order terms combine to form

$$R_{00} = -3\frac{d^2a/dt^2}{a}$$

an equation in which we are by now quite confident since this is the third time we have derived it (see equations (2.34) and (5.13)). Of course the big news here is not that we have correctly derived the zero-order term, but rather that tensor perturbations do not appear at first-order in R_{00}. Looking ahead, we will soon see that the Ricci scalar also has no tensor contribution (even though R_{ij} does). Therefore, we can anticipate that the time-time component of Einstein's equations contains no tensor perturbations. This is important for it tells us that density perturbations - which form the right hand side of the time-time component as shown in Eq. (5.26) - do not induce any tensor perturbations. We are beginning therefore to get a glimmer of the decomposition theorem. Density perturbations and scalar perturbations to the metric are coupled; indeed their names are often used as synonyms. Tensor perturbations, however, are decoupled from these and evolve on their own.

The spatial components of the Ricci tensor do depend on the tensor perturbation variables. We now turn to

$$R_{ij} = \Gamma^{\alpha}_{ij,\alpha} - \Gamma^{\alpha}_{i\alpha,j} + \Gamma^{\alpha}_{\alpha\beta}\Gamma^{\beta}_{ij} - \Gamma^{\alpha}_{\beta j}\Gamma^{\beta}_{i\alpha}.$$

Let's consider the first two terms together. Expanding out leads to

$$\Gamma^{\alpha}_{ij,\alpha} - \Gamma^{\alpha}_{i\alpha,j} = \Gamma^{0}_{ij.0} + \Gamma^{k}_{ij,k} - \Gamma^{k}_{ik,j}$$

since $\alpha = 0$ does not contribute in $\Gamma^{\alpha}_{i\alpha.j}$ because of Eq. (5.35). The hardest (i.e., longest) term here is the first, which involves multiple time derivatives. Let's postpone its calculation by recalling that $\Gamma^{0}{}_{ij} = g_{ij.0}/2$ so that the first term can be written in shorthand as $g_{ij,00}/2$. The last term in Eq. (5.49) vanishes since $\Gamma^{k}_{ik} = 0$ for tensor perturbations. Combining the other terms then leads to

$$\Gamma^{\alpha}_{ij,\alpha} - \Gamma^{\alpha}_{i\alpha,j} = \frac{g_{ij,00}}{2} + \frac{1}{2}\left[-k_ik_k\mathcal{H}_{jk} - k_jk_k\mathcal{H}_{ik} + k^2\mathcal{H}_{ji}\right]$$

Recall that we chose $\vec{k}$ to be along the z-axis. Therefore, the indices i and k in the first term in brackets must be equal to 3. But these multiply $\mathcal{H}_{jk} = \mathcal{H}_{j3} = 0$ so this term and its cousin $k_j k_k \mathcal{H}_{ik}$ must both vanish. Therefore,

$$\Gamma^{\alpha}{}_{ij,\alpha} - \Gamma^{\alpha}{}_{i\alpha,j} = \frac{g_{ij.00}}{2} + \frac{k^2}{2}\mathcal{H}_{ji}$$

The third term in Eq. (5.48), $\Gamma^{\alpha}_{\alpha\beta}\Gamma^{\beta}{}_{ij}$, is nonzero only when the index α is spatial, so

$$\Gamma^{\alpha}_{\alpha\beta}\Gamma^{\beta}{}_{ij} = \Gamma^{k}_{k0}\Gamma^{0}_{ij} + \Gamma^{k}_{kl}\Gamma^{l}_{ij}.$$

But each of the Christoffel symbols in the second term here are first-order, so their product vanishes. In the first term, the sum over k makes the first-order terms go away, so Γ^{k}_{k0} is purely zero-order, $3H$. Therefore,

$$\Gamma^{\alpha}{}_{\alpha\beta}\Gamma^{\beta}{}_{ij} = \frac{3}{2}H g_{ij,0}$$

The final term in Eq. (5.48) will be left as an exercise; it is

$$\Gamma^{\alpha}{}_{\beta j}\Gamma^{\beta}_{i\alpha} = 2H^2 g_{ij} + 2a^2 H\mathcal{H}_{ij,0}.$$

We can now combine all four terms in Eq. (5.48) to get

$$\begin{aligned} R_{ij} = \frac{g_{ij,00}}{2} + \frac{k^2}{2}\mathcal{H}_{ji} + \frac{3}{2}H g_{ij,0} \\ - 2H^2 g_{ij} - 2a^2 H\mathcal{H}_{ij,0}. \end{aligned}$$

We now need to expand out the time derivatives of the metric. Using Eq. (5.38), one finds

$$g_{ij,00} = 2g_{ij}\left(\frac{d^2a/dt^2}{a} + H^2\right) + 4a^2 H\mathcal{H}_{ij,0} + a^2\mathcal{H}_{ij,00}.$$

Therefore the Ricci tensor is

$$\begin{aligned} R_{ij} = g_{ij}\left(\frac{d^2a/dt^2}{a} + 2H^2\right) + \frac{3}{2}a^2 H\mathcal{H}_{ij,0} \\ +a^2\frac{\mathcal{H}_{ij,00}}{2} + \frac{k^2}{2}\mathcal{H}_{ij} \end{aligned}$$

Again we see that we have successfully recaptured the zero-order part of the Ricci tensor. Remarkably, we will see that the first-order parts - when used in Einstein's equations – do not couple to the scalar perturbations.

First, though, we must compute the Ricci scalar:

$$\mathcal{R} = g^{00}R_{00} + g^{ij}R_{ij}$$

The time-time product is all zero-order, so we can neglect it when considering the first-order piece $\delta\mathcal{R}$. The space-space contraction has two types of terms. First, there are the terms in Eq. (5.57) proportional to the metric g_{ij}. But $g^{ij}g_{ij} = 3$, so there are no first-order terms here. All the other terms in Eq. (5.57) are proportional to $\mathcal{H}_{ij}$, so when contracting them we can set g^{ij} to its zero-order value, δ_{ij}/a^2. This corresponds to taking the trace of the first-order terms in Eq. (5.57). Since all first-order terms are proportional to $\mathcal{H}_{ij}$, the trace vanishes. Therefore, tensor perturbations do not affect (at first order) the Ricci scalar.

Einstein Equations for Tensor Perturbations

Now let's read off the perturbations to the Einstein tensor induced by tensor modes. Since the Ricci scalar is unperturbed by tensor perturbations, the first-order Einstein tensor is simply

$$\delta G^i_j = \delta R^i_j$$

To get R^i_j, we contract $g^{ik}R_{kj}$, using the Ricci tensor we computed in Eq. (5.57). The first term, proportional to the contraction of $g^{ik}g_{kj} = \delta^i_j$, has no first-order piece; the remaining terms are explicitly first-order in $\mathcal{H}$, so we can set $g^{ik} = \delta^{ik}/a^2$, leading to

$$\delta G^i{}_j = \delta^{ik}\left[\frac{3}{2}H\mathcal{H}_{kj,0} + \frac{\mathcal{H}_{kj,00}}{2} + \frac{k^2}{2a^2}\mathcal{H}_{kj}\right]$$

We can now derive a set of equations governing the evolution of the tensor variables, $h_+$and $h_\times$.

To derive an equation for h_+, let us consider the difference between the $_1$ and $^2_2$ components of the Einstein tensor. The Einstein tensor in Eq. (5.60) is proportional to $\mathcal{H}_{ij}$ and its derivatives. Since $\mathcal{H}_{11} = -\mathcal{H}_{22} = h_+, \delta G^1{}_1$ is equal and opposite to $\delta G^2{}_2$. Therefore,

$$\delta G^1{}_1 - \delta G^2{}_2 = 3Hh_{+,0} + h_{+,00} + \frac{k^2 h_+}{a^2}.$$

Change to conformal time so that $h_{+,0} = \dot{h}_+/a$ and $h_{+,00} = \ddot{h}_+/a^2 - (\dot{a}/a^3)\,\dot{h}_+$. Then,

$$a^2\left[\delta G^1_1 - \delta G^2_2\right] = \ddot{h}_+ + 2\frac{\dot{a}}{a}\dot{h}_+ + k^2 h_+$$

The right-hand side of this component of Einstein's equations is zero (Exercise 8), and $h_\times$obeys the same equation (Exercise 9), so the tensor modes are governed by

$$\ddot{h}_\alpha + 2\frac{\dot{a}}{a}\dot{h}_\alpha + k^2 h_\alpha = 0$$

where $\alpha = +, \times$. Equation (5.63) is a wave equation, and the corresponding solutions are called gravity waves. For example, if we neglect the expansion of the universe so that the damping term in Eq. (5.63) vanishes, we immediately see that the two solutions are $h_\alpha \propto e^{\pm ik\eta}$. In real space, then the perturbation to the metric is of the form

$$h_\alpha(\vec{x},\eta) = \int d^3k e^{i\vec{k}\cdot\vec{x}}\left[Ae^{ik\eta} + Be^{-ik\eta}\right] \qquad \text{(no expansion)}$$

The two modes here corresponds to waves traveling in the $\pm\hat{z}$ direction at the speed of light.

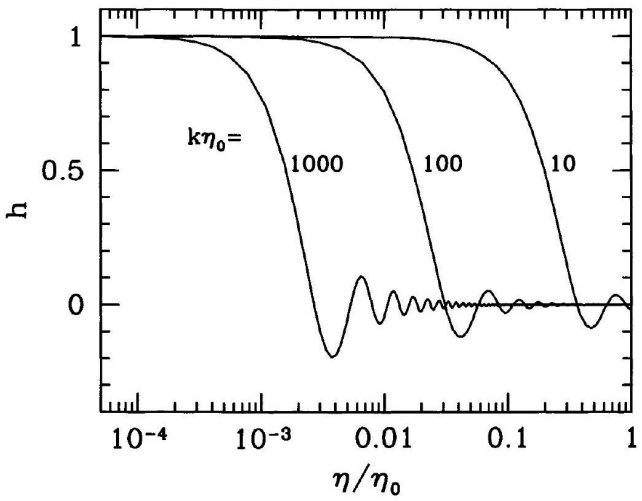


Figure 5.1. Evolution of gravity waves as a function of conformal time. Three different modes are shown, labeled by their wave numbers. Smaller scale modes decay earlier.

Equation (5.63) is a generalization of the wave equation to an expanding universe. Exercise 12 illustrates that if the universe is purely radiation or matter dominated, exact analytic solutions can be obtained. These are oscillatory, like the simple ones in Eq. (5.64), but also damp out. Figure 5.1 shows the evolution of h_α for three different wavelength modes. The large-scale mode (with $k\eta_0 = 10$) remains constant at early times when its wavelength is larger than the horizon $k\eta < 1$. Once its wavelength becomes comparable to the horizon, the amplitude begins to die off, oscillating several times until the present. The small-scale mode $k\eta_0 = 1000$ shown in Figure 5.1 also begins to decay when its wavelength becomes comparable to the horizon. Its entry into the horizon occurs much earlier, though, so the decay is much more efficient. By today, the amplitude is extremely small.

An important point about the effect of gravity waves on the CMB anisotropy spectrum can be gleaned from Figure 5.1. Because small-scale modes decay earlier than large-scale modes, at decoupling (at $\eta/\eta_0 \simeq 0.02$) only modes with $k\eta_0$ less than about 100 persist. All smaller scale modes can be neglected. Therefore, anisotropies on small angular scales will not be affected by gravity waves. Only the large-scale anisotropies are impacted by gravity waves.

12.5.4 The Decomposition Theorem

The decomposition theorem states that perturbations to the metric can be divided up into three types: scalar, vector, and tensor. Each of these types of perturbations evolves independently. That is, if some physical process in the early universe sets up tensor perturbations, these do not induce scalar perturbations. Conversely, to determine the evolution of scalar perturbations, we will not have to worry about possible vector or tensor perturbations.

Now that we have computed the contributions to the Einstein tensor $G_{\mu\nu}$ from scalars and tensors, we can demonstrate the decomposition of these two types of perturbations. To do this, remember that we obtained the scalar equations by considering the two components of Einstein's tensor:

$$G^0_0 \quad ; \quad \left(\hat{k}_i\hat{k}_j - (1/3)\delta_{ij}\right) G^i_j.$$

Inserting these components into Einstein's equations led to equations (5.27) and (5.33). If we can show that tensor perturbations do not contribute to these two components, then we will have convinced ourselves of at least part of the decomposition theorem, namely that the equations governing scalar equations are not affected by tensors.

It is easy to see that tensor perturbations do not contribute to $G^0{}_0$. For $G^0{}_0$ depends on R_{00} and $\mathcal{R}$. But we have seen that both of these do not depend on h_+ or $h_\times$.

Now let's show that $\left(\hat{k}_i\hat{k}_j - \delta_{ij}/3\right) G^i{}_j$ also does not pick up a contribution from tensor perturbations. Multiply Eq. (5.60) by the projection operator:

$$\left(\hat{k}_i\hat{k}_j - (1/3)\delta_{ij}\right) \delta G^i{}_j = (\delta_{i3}\delta_{j3} - (1/3)\delta_{ij}) \times \left[\frac{3}{2}H\mathcal{H}_{ij,0} + \frac{\mathcal{H}_{ij,00}}{2} + \frac{k^2}{2a^2}\mathcal{H}_{ij}\right]$$

where the equality holds since we have chosen $\hat{k}$ to lie in the $\hat{z}$ direction. The terms in which indices i and j are set to 3 vanish since $\mathcal{H}_{33} = 0$. The only remaining terms are those proportional to δ_{ij}. But the Kronecker delta instructs us to take the trace of $\mathcal{H}$. This too vanishes. The scalar equations we derived in the previous section are therefore unchanged by the presence of tensor modes. This is a manifestation of the decomposition theorem.

12.5.5 from Gauge to Gauge

Let's go back to scalar perturbations. Until now, we have characterized these with Ψ and Φ in the form of Eq. (4.9). This corresponds to a choice of gauge or a choice of a coordinate system with which to study the space-time. If we changed the coordinate system we use, we would get a metric of a different form, i.e., a different gauge. Although we will work almost exclusively in the gauge corresponding to Eq. (4.9), the conformal Newtonian gauge, historically many other gauges have been used, and for different parts of the "cosmological perturbation" problem, different gauges have their advantages. Indeed, we will see in Section 6.5.3 that people who work on the theory of inflation sometimes prefer a gauge with spatially flat slicing (g_{ij} unperturbed), since the equations for the perturbations generated by inflation simplify considerably. Also, the code currently used by most people to compute anisotropies and inhomogeneities in the universe uses synchronous gauge, partly because the equations are better behaved numerically in that gauge. So the ability to move back and forth between different gauges is useful, and I want to spend a few pages describing how to do this.

Most generally, scalar perturbations to the metric can be written down as

$$\begin{aligned} g_{00} &= -(1+2A) \\ g_{0i} &= -aB_{,i} \\ g_{ij} &= a^2\left(\delta_{ij}[1+2\psi] - 2E_{,ij}\right). \end{aligned}$$

There are four functions which characterize scalar perturbations to the metric: A, B, ψ, and E. They all depend on space and time, and they are all scalars. For example, the g_{0i} components are the derivatives of a scalar function, not an independent vector function with its own orientation. In conformal Newtonian gauge, $A = \Psi$ and $\psi = \Phi$, while $B = E = 0$.

How do we transform from one gauge to another? The invariant distance of Eq. (2.2) does not change if different coordinates $\tilde{x}$ are used instead of x. Therefore,

$$\tilde{g}_{\alpha\beta}(\tilde{x})d\tilde{x}^\alpha d\tilde{x}^\beta = g_{\mu\nu}(x)dx^\mu dx^\nu,$$

where I have used a different set of dummy indices on both sides to make the upcoming few lines clearer. One of the differentials on the left-hand side can be written as $d\tilde{x}^\alpha = (\partial\tilde{x}^\alpha/\partial x^\mu)\,dx^\mu$ and similarly with the other differential, so equating coeffients of $dx^\mu dx^\nu$ leads to

$$\tilde{g}_{\alpha\beta}(\tilde{x})\frac{\partial\tilde{x}^\alpha}{\partial x^\mu}\frac{\partial\tilde{x}^\beta}{\partial x^\nu} = g_{\mu\nu}(x).$$

This equation is what we are after: a prescription for how the metric changes under a coordinate transformation.

The most general coordinate transformation is generated by

$$\begin{aligned} t \to \tilde{t} &= t + \xi^0(t,\vec{x}) \\ x^i \to \tilde{x}^i &= x^i + \delta^{ij}\xi_{,j}(t,\vec{x}), \end{aligned}$$

where we take ξ^0 and ξ to be small perturbations of the same order as the variables characterizing the perturbations. Let's examine how the metric changes under such a transformation. I'll work out one component explicitly and leave the rest as an exercise. Consider the oo component of Eq. (5.69):

$$\tilde{g}_{\alpha\beta}(\tilde{x})\frac{\partial\tilde{x}^\alpha}{\partial t}\frac{\partial\tilde{x}^\beta}{\partial t} = -[1+2A].$$

I claim that the only term that contributes to the left-hand side is the one with $\alpha = \beta = 0$. Consider for example $\alpha = 0$ and $\beta = i$. The off-diagonal component of the metric $\tilde{g}_{0i}$ is

proportional to $\tilde{B}_{,i}$ a first-order perturbation. But $\partial\tilde{x}^i/\partial t$ is proportional to the first-order variable ξ, so the product is second-order and can be neglected. A similar argument holds for the $\alpha = i; \beta = j$ terms. Therefore, the left-hand side is simply

$$-[1+2\tilde{A}]\left(\frac{\partial\tilde{t}}{\partial t}\right)^2 = -[1+2\tilde{A}]\left(1+\frac{\partial\xi^0}{\partial t}\right)^2$$
$$\simeq -1-2\tilde{A}-2\frac{\partial\xi^0}{\partial t}.$$

Equating this with g_{00} leads to

$$-2\tilde{A}-2\frac{\partial\xi^0}{\partial t} = -2A$$

so under the coordinate transformation specified by Eq. (5.70)

$$A \rightarrow \tilde{A} = A - \frac{1}{a}\dot{\xi}^0.$$

In a similar vein, the other components of the metric transform into

$$\tilde{\psi} = \psi - H\xi^0$$

$$\tilde{B} = B - \frac{\xi^0}{a} + \dot{\xi}$$
$$\tilde{E} = E + \xi.$$

One technical point: Eqs. (5.74) and (5.75) describe how the components of the metric tensor transform under general coordinate changes. These equations, which have become standard, are a bit misleading, though, because each of the individual functions, A for example, transforms as a scalar, i.e., does not change under a spatial coordinate transformation. Here we have transformed the metric and accomodated the resulting changes in new definitions of A, B, ψ, and E. This is not the same thing as seeing how A by itself changes under a transformation.

To sum up, then, there are four functions which characterize scalar perturbations, but these can be manipulated with two other functions which characterize coordinate transformations. For example, starting with a metric in which $E \neq 0$, it is trivial to make a transformation to eliminate E: simply choose $\xi = -E$, and the resulting $\tilde{E} = 0$. Thus, there are really only $4 - 2 = 2$ functions which matter. Indeed, this is the reason that we had only Φ and Ψ in conformal Newtonian gauge. More generally, one might hope to construct two gauge invariant variables, those which remain unchanged under a general coordinate transformation. Bardeen (1980) first identified two such variables:

$$\Phi_A := A + \frac{1}{a}\frac{\partial}{\partial\eta}[a(\dot{E}-B)]$$
$$\Phi_H := -\psi + aH(B-\dot{E}).$$

In conformal Newtonian gauge, in which $E = B = 0, \Phi_A = \Psi$ and $\Phi_H = -\Phi$. These invariants are very useful: if equations simplify in a particular gauge, then one can do calculations in that gauge, form the gauge-invariant variables, and then turn these into the perturbations in any other gauge. We will do precisely this in Section 6.5.3. In other words, Φ_A and Φ_H are useful shortcuts or recipes for transforming from one gauge to another.

Under a general coordinate transformation, the components of the energymomentum tensor $T_{\mu\nu}$ also change. In exact analogy with the metric tensor,

$$\tilde{T}_{\mu\nu}(\tilde{x}) = \frac{\partial x^\alpha}{\partial \tilde{x}^\mu}\frac{\partial x^\beta}{\partial \tilde{x}^\nu}T_{\alpha\beta}(x)$$

Again, though, Bardeen found combinations of the components of $T_{\mu\nu}$ which remain invariant and therefore facilitate mapping from one gauge to another. In particular, in Fourier space

$$v := ikB + \frac{\hat{k}^i T_i^0}{(\rho + \mathcal{P})a}$$

remains invariant under a coordinate transformation. In conformal Newtonian gauge, for matter, v is indeed equal to the v we defined in Chapter 4. For radiation, $v = -3i\Theta_{r,1}$, i.e., proportional to the dipole, again in conformal Newtonian gauge. A second invariant is the generalized perturbation to the energy density,

$$\epsilon_m := -1 - \frac{T_0^0}{\rho} + \frac{3H}{k^2\rho}k^i T_i^0$$

For matter, in conformal Newtonian gauge, $\epsilon_m = \delta + (3aHv/k)$; i.e., it reduces to the ordinary overdensity δ on scales smaller than the horizon. For radiation, $\epsilon_m = 4\Theta_{r,0} - 12i\Theta_{r,1}aH/k$, again reducing to the standard overdensity on small scales.

12.6 Initial Conditions

In order to understand structure in the universe, we have derived the equations governing perturbations around a smooth background. Before we start solving these equations, we need to know the initial conditions. This quest for initial conditions will lead to an entirely new realm of physics, the theory of inflation. Inflation was introduced (Guth, 1981; Linde 1982; Albrecht and Steinhardt, 1982) partly to explain how regions which could not have been in causal contact with each other have the same temperature. It was soon realized (Starobinsky, 1982; Guth and Pi, 1982; Hawking, 1982; Bardeen, Steinhardt, and Turner, 1983; Brandenberger, Kahn, and Press, 1983; Guth and Pi, 1985) that the very mechanism that explains the uniformity of the temperature in the universe can also account for the origin of perturbations in the universe. Therefore, in order to produce a set of initial conditions, we will need to detour into the world of inflation. One warning: we are not sure that inflation is the mechanism that generated the initial perturbations. It is very difficult to test a theory based on energy scales well beyond the reach of accelerators. Nonetheless, it is by far the most plausible explanation. Indeed, one of the current problems in cosmology is that there is really no viable alternative to inflation. Also, the next generation of CMB and large-scale structure observations will put inflation to some stringent tests.

12.6.1 The Einstein-Boltzmann Equations at Early Times

Theory

Chapters 4 and 5 contain nine first-order differential equations for the nine perturbation variables we need to track. In principle, we need initial conditions for $\therefore$ 'l of these variables. In practice, though, a combination of arguments will relate many of these variables to each other, and we need only determine the initial conditions for one of these. This section determines the way all variables depend on the gravitational potential Φ at early times; the remaining sections work out the initial conditions for Φ.

Let us consider first the Boltzmann equations (4.100)-(4.107) at very early times. In particular, we want to consider times so early that for any k-mode of interest, $k\eta \ll 1$. This

inequality immediately leads to several important simplifications. Consider the terms $\dot{\Theta}$ and $ik\mu\Theta$ in Eq. (4.100). The first term is of order Θ/η, while the second is of order $k\Theta$. Therefore, the first is larger than the second by a factor of order $1/(k\eta)$, which, by assumption, is much greater than 1. In a similar way, we can argue that all terms in the Boltzmann equations multiplied by k can be neglected at early times. Physically, this means that, at early times, all perturbations of interest have wavelengths ($\sim k^{-1}$) much larger than the distance over which causal physics operates. A hypothetical observer then who sees only photons from within his causal horizon will see a uniform sky. Thus higher multipoles $(\Theta_1, \Theta_2, \ldots)$ are much smaller than the monopole, Θ_0. Therefore, the perturbations to the photon and neutrino temperatures evolve according to

$$\dot{\Theta}_0 + \dot{\Phi} = 0$$
$$\dot{\mathcal{N}}_0 + \dot{\Phi} = 0.$$

The same principles can be applied to the matter distributions. The overdensity equations reduce to

$$\dot{\delta} = -3\dot{\Phi}$$
$$\dot{\delta}_{\rm b} = -3\dot{\Phi}.$$

The velocities are comparable to the first moments of the radiation distributions, so they are smaller than the overdensities by a factor of order $k\eta$ and may be set to zero initially. In fact, the baryon velocity is not only comparable to the photon first moment, Θ_1: it is equal to it by virtue of the strength of Compton scattering. That is, the largeness of $\dot{\tau}$ in Eq. (4.106) ensures that $v_{\rm b} = -3i\Theta_1$. We will use this later when reexamining the Boltzmann equations closer to decoupling. For now, we are interested in times so early that the only relevant fact is that higher moments are all negligibly small.

Now let us turn to the Einstein equations at early times. First consider Eq. (5.27). The first term there contains a factor of k^2 so may be neglected. Also the two matter terms on the right are negligible at early times since radiation dominates. Therefore, we have

$$3\frac{\dot{a}}{a}\left(\dot{\Phi} - \frac{\dot{a}}{a}\Psi\right) = 16\pi G a^2 \left(\rho_\gamma \Theta_0 + \rho_\nu \mathcal{N}_0\right)$$

But since radiation dominates, $a \propto \eta$ (recall Eq. (2.100) and the discussion immediately afterward) so $\dot{a}/a = 1/\eta$. Therefore,

$$\begin{aligned}\frac{\dot{\Phi}}{\eta} - \frac{\Psi}{\eta^2} &= \frac{16\pi G\rho a^2}{3}\left(\frac{\rho_\gamma}{\rho}\Theta_0 + \frac{\rho_\nu}{\rho}\mathcal{N}_0\right) \\ &= \frac{2}{\eta^2}\left(\frac{\rho_\gamma}{\rho}\Theta_0 + \frac{\rho_\nu}{\rho}\mathcal{N}_0\right)\end{aligned}$$

where the last equality follows by virtue of the zero-order Einstein equation. To simplify further, we can define the ratio of neutrino energy density to the total radiation density as

$$f_\nu := \frac{\rho_\nu}{\rho_\gamma + \rho_\nu}$$

Then multiplying Eq. (6.4) by η^2 leads to

$$\dot{\Phi}\eta - \Psi = 2\left(\left[1 - f_\nu\right]\Theta_0 + f_\nu \mathcal{N}_0\right)$$

Recall that Eq. (6.1) relates the derivative of the monopoles to the derivative of the potential. We can therefore eliminate both monopoles from Eq. (6.6) by differentiating both right- and left-hand sides. Then,

$$\ddot{\Phi}\eta + \dot{\Phi} - \dot{\Psi} = -2\dot{\Phi}$$

where the right-hand side follows since both $\dot{\Theta}_0$ and $\dot{\mathcal{N}}_0$ are equal to $-\dot{\Phi}$ for these large scale modes.

So far we have used only one Einstein equation. The second, Eq. (5.33), describes how the higher moments of the photon and neutrino distributions cause $\Psi + \Phi$ to be nonzero. Let us here neglect these higher order moments, which cause the sum of the gravitational potentials to be slightly nonzero. [1] Under this approximation, we can eliminate Ψ everywhere by simply setting it to $-\Phi$. Then,

$$\ddot{\Phi}\eta + 4\dot{\Phi} = 0$$

Setting $\Phi = \eta^p$ leads to the algebraic equation

$$p(p-1) + 4p = 0$$

which allows two solutions: $p = 0, -3$. The $p = -3$ mode is the decaying mode. If it is excited very early on, it will quickly die out and have no impact on the universe. The $p = 0$ mode, on the other hand, does not decay if excited. It is the mode we are interested in. If some mechanism can be found which excites this mode, this mechanism may well be responsible for the perturbations in the universe.

Focusing therefore on only the $p = 0$ mode, we see that Eq. (6.6) relates the gravitational potential to the neutrino and photon overdensities:

$$\Phi = 2\left(\left[1 - f_\nu\right]\Theta_0 + f_\nu \mathcal{N}_0\right)$$

Both Θ_0 and $\mathcal{N}_0$ are also constant in time. In most models of structure formation, they are equal since whatever causes the perturbations tends not to distinguish between photons and neutrinos. Therefore, we will set

$$\Theta_0\left(k, \eta_i\right) = \mathcal{N}_0\left(k, \eta_i\right)$$

which leads to

$$\Phi\left(k, \eta_i\right) = 2\Theta_0\left(k, \eta_i\right)$$

where I have explicitly written the k-dependence of all these variables and the fact that we are setting up the initial conditions at some early time η_i.

[1] See Exercise 2 for a careful accounting of the effect of the neutrino quadrupole; the photon quadrupole is kept minuscule by Compton scattering, so it really does not contribute to Eq. (5.33). The initial conditions for matter, both δ and $\delta_{\rm b}$, depend upon the nature of the primordial perturbations. Combining the first of equations (6.1) and (6.2) leads to

$$\delta = 3\Theta_0 + \text{ constant}$$

for the dark matter overdensity, with an identical equation for the baryon overdensity. Primordial perturbations are often divided into those for which the constant in Eq. (6.13) is zero (adiabatic perturbations) and those for which the constant is nonzero (isocurvature perturbations). Adiabatic perturbations have a constant matter-to-radiation ratio everywhere since

$$\frac{n_{\rm dm}}{n_\gamma} = \frac{n_{\rm dmm}^{(0)}}{n_\gamma^{(0)}}\left[\frac{1+\delta}{1+3\Theta_0}\right].$$

The prefactor, the ratio of zero-order number densities, is a constant in both space and time. For the ratio of matter to radiation number density to be uniform, therefore, the combination inside the brackets which linearizes to $1 + \delta - 3\Theta_0$ must be independent of space. So the perturbations must sum to zero,

$$\delta = 3\Theta_0$$

for adiabatic perturbations. By similar arguments for the baryons, $\delta_b = 3\Theta_0$. There are models based on isocurvature perturbations, but these have not been very successful to date; we will focus on adiabatic initial conditions.

For the most part, velocities and dipole moments are negligibly small in the very early universe. However, we will encounter situations where we need to know the initial conditions for these as well. You will show in Exercise 3 that the appropriate initial conditions are

$$\begin{aligned}\Theta_1 = &\mathcal{N}_1 = \frac{iv_b}{3} = \frac{iv}{3} \\ &= -\frac{k\Phi}{6aH}.\end{aligned}$$

12.6.2 The Horizon

Theory

If this book were a novel or a biography, a better title for this section might be Midlife Crisis. The main character would have attended a good high school, studied hard, and gone on to a solid university. There he fell in love with an exciting, but sensible, woman; upon graduating, he set up some interviews, and got a good job downtown. He married his college girlfriend, and after several years in the city, they moved to the suburbs and had three kids. Our hero contributed to the community and was recognized all over town as a solid citizen. He was moving up fast in his company and there was talk about a political position. Just when he was about to declare his candidacy, he began to have doubts. "What have I been doing with my life? What is really important? Were all those years of study and work simply a 'track'? Did I take this path just because everyone else was moving in the same direction? Where is the innovation and the signature that my life is mine?" And worse, he has a secret, an underlying feeling that everything he has built is based on a fallacy.

OK, maybe it wouldn't be a bestseller, but it does serve as a useful metaphor for our study of perturbations in the universe. Until now, we have done everything in a systematic, proper way. We reviewed the standard Big Bang cosmology. We expanded about this zero-order smooth universe, getting evolution equations for the perturbations to the particle distributions and to the gravitational fields. We realized that these coupled differential equations needed initial conditions so in the last section we set those up. However, now we must ask, What caused those initial perturbations? It is one thing to say that $\Phi = 2\Theta_0$ initially. It is quite another to explain what caused Φ to be nonzero in the first place.

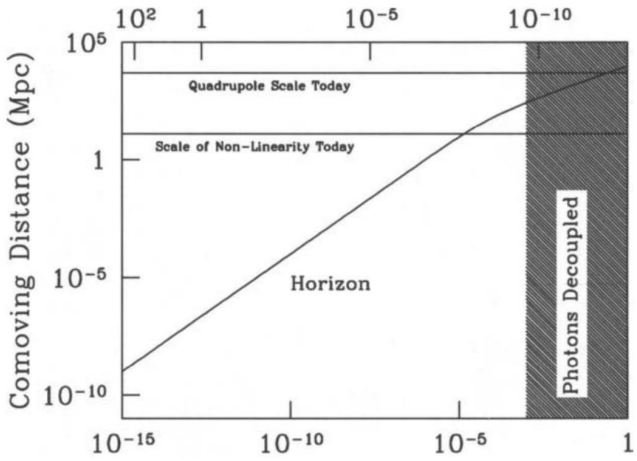
And it is worse than that. To understand why let us recall the physical meaning of the conformal time η: it is the maximum comoving distance traveled by light since the beginning of the universe. Equivalently, objects separated by comoving distances larger than η today were not ever in causal contact: there is simply no way information could have propagated over distances larger than η. For this reason, η is called the comoving horizon.

With this in mind, we can now revisit the condition used in the previous section that $k\eta \ll 1$. The wavenumber k is roughly equal to the inverse of the wavelength of the mode in question (give or take a factor of 2π). Therefore $k\eta$ is the ratio of the comoving horizon to the comoving wavelength of the perturbation. If this ratio is much smaller than 1, then the mode in question

has a wavelength so large that no causal physics could possibly have affected it. A picture worth remembering is shown in Figure 6.1. The horizon grows as the scale factor increases. On the other hand, comoving wavelengths remain constant. All modes of cosmological interest therefore had wavelengths much larger than the horizon early on. Eventually these cosmological modes enter the horizon; after that, causal physics begins to operate on them.

The truly disturbing feature of this realization is most apparent when considering the microwave background today. On all scales observed the CMB is very close to isotropic. How can this be? The largest scales observed have entered the horizon just recently, long after decoupling. (An example is the scale corresponding to the quadrupole moment of the CMB, shown in Figure 6.1.) Before decoupling, the wavelengths of these modes are so large that no causal physics could force deviations from smooothness to go away. After decoupling, the photons do not interact at all; they simply freestream. So even though it is technically possible that photons reaching us today from opposite directions had a chance to communicate with each other and equilibrate to the same temperature, practically this could not have happened. Why then is the CMB temperature so uniform? This is a profound problem that we have glossed over by simply assuming that the temperature is uniform and that perturbations about the zero-order temperature are small.

T(GeV)



a

Figure 6.1. The comoving horizon as a function of the scale factor. Also shown are two comoving wavelengths, which remain constant with time. Early in the history of the universe, both of these modes - as well as all other modes of cosmological interest — had wavelengths much larger than the horizon. The CMB comes from the last scattering surface at $a \simeq 10^{-3}$. At that time, the largest scales (e.g., the one labeled "quadrupole") were still outside the horizon. The horizon problem asks how regions separated by distances larger than the horizon at the last scattering surface can have the same temperature.

A more intuitive picture of the horizon problem is shown in Figure 6.2. At any given time, the region within the cone is causally connected to us (at the center). Photons that we observe today from the last scattering surface were well outside our horizon when they were first emitted. The most disturbing aspect of this is the observation of large-angle isotropy, an indication that photons apparently separated by many horizons at the last scattering surface nonetheless shared the same temperature (to a part in 10^5).

12.6.3 Inflation

Theory

This section describes a beautiful solution to the horizon problem outlined in the previous section. First, we explore a logical way out of the previous argument by realizing that an early epoch of rapid expansion solves the horizon problem. Then we consider the Einstein equations to tell us what type of energy is needed in order to produce this rapid expansion, showing that negative pressure is required.

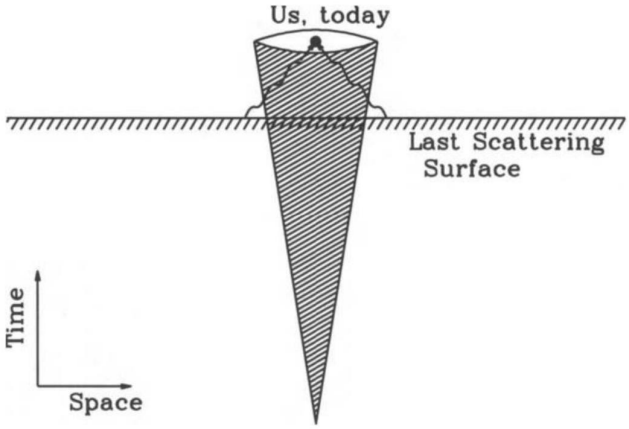


Figure 6.2. The horizon problem. The region inside the cone at any time is causally connected to us (at the center). Photons emitted from the last scattering surface (at redshift ~ 1000) started outside of this region. Therefore, at the last scattering surface, they were not in causal contact with us and certainly not with each other. Yet their temperatures are almost identical.

Finally, we consider a scalar field theory and show that negative pressure is easy to accommodate in such a theory.

Two comments about the field theory implementation. First, field theory has a reputation as a difficult subject. It is, but the part we will need for inflation is decidedly simple. Indeed, almost all we will need to know about field theory we've already used in the previous chapter on general relativity. The second point is that there is no known scalar field which can drive inflation. (A skeptic might point out that there is no known fundamental scalar field at all!) Therefore, it may well be true that the idea of inflation is correct but it is driven by something other than a scalar field. Having said that, there are a number of reasons to work with scalar fields, as we will do whenever we need to specify the source of inflation. Almost all fundamental particle physics theories contain scalar fields. In fact, historically it was particle physicists studying high-energy extensions of the Standard Model (in particular Grand Unified Theories) who proposed the idea of inflation driven by a scalar field as a natural byproduct of some of these extensions. Indeed, almost all current work on inflation is based on a scalar field (or sometimes two). The alternative from a particle physics point of view is to use a vector field (such as the electromagnetic potential) or a set of fermions (similar to the way condensates induce superconductivity) to drive inflation. Neither of these choices works very well, but they both complicate things severely, so we will stick to a scalar field.

A Solution to the Horizon Problem

To motivate a solution to the horizon problem, let me rewrite the comoving horizon as

$$\eta = \int_0^a \frac{da'}{a'} \frac{1}{a' H\left(a'\right)}$$

The comoving horizon then is the logarithmic integral of the comoving Hubble radius, $1/aH$. The Hubble radius is the distance over which particles can travel in the course of one expansion time, i.e., roughly the time in which the scale factor doubles. So the Hubble radius is another way of measuring whether particles are causally connected with each other: if they are separated by distances larger than the Hubble radius, then they cannot currently communicate. There is a subtle distinction between the comoving horizon η and the comoving Hubble radius $(aH)^{-1}$. If particles are separated by distances greater than η, they never could have communicated with one another; if they are separated by distances greater than $(aH)^{-1}$, they cannot talk to each other now. It is therefore possible that η could be much larger than $(aH)^{-1}$ now, so that particles cannot communicate today but were in causal contact early on. This might happen if the comoving Hubble radius early on was much larger than it is now so that η got most of its contribution from early times. This could happen, but it does not happen during matter- or radiationdominated epochs. In those cases, the comoving Hubble radius increases with time, so typically we expect the largest contribution to η to come from the most recent times. Indeed, this is precisely what Figure 6.1 indicates.

Look again at Figure 6.1. On top of the figure I have drawn an axis which depicts the temperature of the cosmic plasma for the given value of the scale factor. We know quite a bit about physics going up to the limits on the plot, several hundred GeV. Beneath these energies, the standard model of particle physics works very well. Beyond those energies, although we have ideas, there is no experimental reason to prefer one theory over another. Since the energy content of the universe determines $a(t)$, when you mentally extrapolate the horizon in Figure 6.1 back to $a = 0$, or equivalently to infinitely high temperatures, you are really making an assumption. You are assuming that nothing strange happened early on, in particular that the universe was always radiation dominated at early times. If this were true, then it does indeed follow that the comoving horizon received a negligible contribution from the very early universe, that photons can travel only very small distances in the first fraction of a second after the Big Bang.

This suggests a solution to the horizon problem: perhaps early on, the universe was not dominated by either matter or radiation. Perhaps, for at least a brief time, the comoving Hubble radius decreased. Then, we would have the situation depicted in Figure 6.3. The comoving Hubble radius would decrease dramatically during this epoch. In that case, the comoving horizon would get most of its contribution not from recent times, but rather from primordial epochs before the rapid expansion of the grid. Particles separated by many Hubble radii today, for example those outside the circle in the bottom panel of the figure, were in causal contact - were inside the Hubble circle in the top panel - before this epoch of rapid expansion.

How must the scale factor evolve in order to solve the horizon problem? We can first answer this question qualitatively. If the comoving Hubble radius is to decrease, then aH must increase. That is,

$$\frac{d}{dt}\left[a\frac{da/dt}{a}\right] = \frac{d^2a}{dt^2} > 0$$

So to solve the horizon problem, the universe must go through a period in which it is accelerating, expanding ever more rapidly. This is the origin of the term inflation. To understand the epoch of inflation more quantitatively, let me give away the punchline that most inflationary models typically operate at energy scales of order 10^{15}GeV or larger. How big was the comoving Hubble radius when the temperature was 10^{15}GeV? We can get an order of magnitude estimate by ignoring the relatively brief epoch of recent matter domination and assuming that the universe has been radiation dominated since the end of inflation (you can correct this assumption in Exercise 6). Then H scales as a^{-2} so $a_0H_0/a_eH_e = a_e$ where a_e is the scale factor at the end of inflation. If a_e corresponds to a time at which the temperature was 10^{15}GeV, then

$a_e \simeq T_0/10^{15}\text{GeV} \simeq 10^{-28}$. So the comoving Hubble radius at the end of inflation was 28 orders of magnitude smaller than it is today. For inflation to work, the comoving Hubble radius at the onset of inflation had to be larger than the largest scales observable today, i.e., larger than the current comoving Hubble radius. So during inflation, the comoving Hubble radius had to decrease by some 28 orders of magnitude.

The most common way to arrange this is to construct a model wherein H is constant during inflation. In that case, since $da/a = Hdt$, the scale factor evolves as

$$a(t) = a_e e^{H(t-t_e)} \quad t < t_e$$

where t_e is the time at the end of inflation. The decrease in the comoving Hubble radius $(aH)^{-1}$ is now due solely to the exponential increase in the scale factor. For the scale factor to increase by a factor of 10^{28}, the argument of the exponential must be of order $\ln(10^{28}) \sim 64$ (but remember the corrections in Exercise 6), so inflation can solve the horizon problem if the universe expands exponentially for more than 60 e-folds.

Thus, consider Figure 6.4, which shows the comoving Hubble radius as a function of the scale factor. The right side of this plot is virtually identical to Figure 6.1, which tells us that the comoving scales of interest to us were much larger than the Hubble radius in the standard cosmology. The left-hand side of the plot though shows that an inflationary epoch reduces the comoving Hubble radius dramatically. This makes sense: since the scale factor is inflating very rapidly, it becomes increasingly difficult for photons to move along the comoving grid (which is itself expanding with a). Before inflation started, the comoving Hubble radius was very large, larger than any scale of cosmological interest today, so all such scales were well within the horizon.

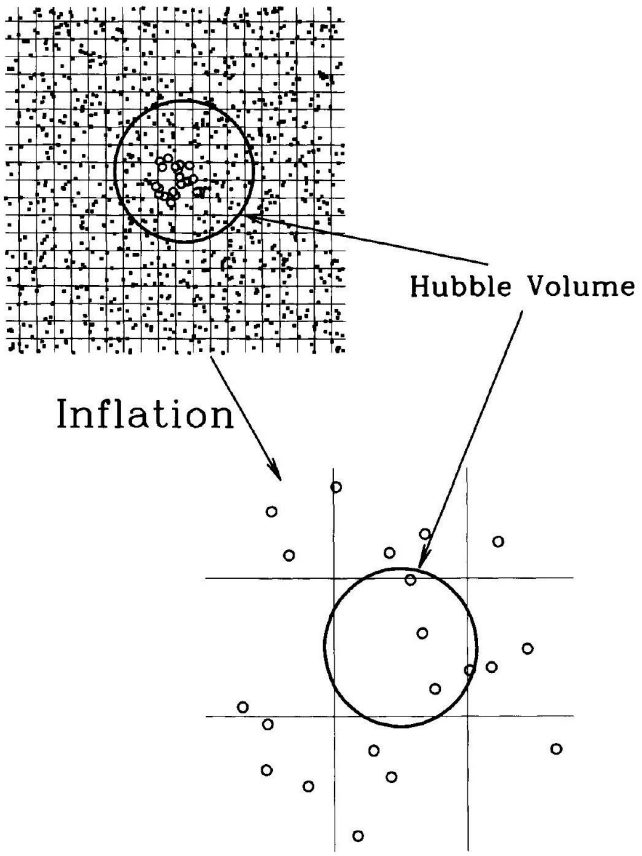


Figure 6.3. Particles on the comoving grid before (top) and after (bottom) inflation. Open circles are the same particles on top and bottom. Before inflation, the comoving Hubble radius was quite large, encompassing dozens of cells on the grid. After inflation, the comoving Hubble

radius has shrunk to just one cell. (In this caricature, the scale factor has grown by a factor of order 7 ; during inflation the scale factor increases by greater than e^{60}.) The shrinkage of the comoving Hubble radius means that particles which were initially in causal contact with one another (within the large circle at top) can now no longer communicate. Note that the physical Hubble radius, depicted by large circles on the top and bottom grids, remains roughly constant during inflation.

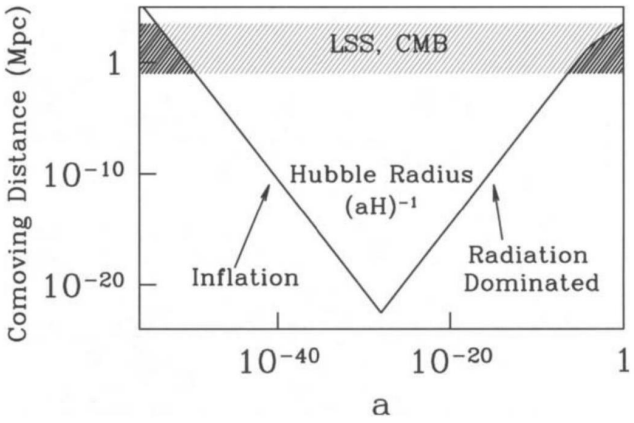


Figure 6.4. The comoving Hubble radius as a function of scale factor. Scales of cosmological interest (shaded band) were larger than the Hubble radius until $a \sim 10^{-5}$. Dark shaded regions show when these scales were smaller than the Hubble radius, and therefore susceptible to microphysical processes. Very early on, before inflation operated, all scales of interest were smaller than the Hubble radius and therefore susceptible to microphysical processing. Similarly, at very late times, scales of cosmological interest came back within the Hubble radius.

Note the symmetry in Figure 6.4. Scales just entering the horizon todayroughly 60 e-folds after the end of inflation - left the horizon 60 e-folds before the end of inflation. The amplitude of the perturbations on these scales remained constant as long as they were super-horizon. So, when we measure them today, we are actually seeing them as they were when they first left the horizon during the inflationary era (modulo whatever processing has taken place since they reentered the horizon, processing we will study in great detail in Chapters 7 and 8). To explain the structure in the universe today, then, it is clearly important to understand the generation of perturbations during inflation.

We have until now discussed inflation in comoving coordinates. But it is also profitable to think of the exponential expansion in physical coordinates. The idea that the horizon blows up early on is depicted in Figure 6.5. The physical size (a times the comoving size) of a causally connected region blows up exponentially quickly during inflation. So regions that we observe to be astronomical today were actually microscopically small before inflation, and they were in causal contact with each other.

The total comoving horizon ceases to be an effective time parameter after inflation because it becomes large very early on, and then changes very little as the universe expands during the matter- and radiation-dominated eras. A simple way to rectify this is to subtract off its primordial part $\eta_{\rm prim}$, and redefine η as

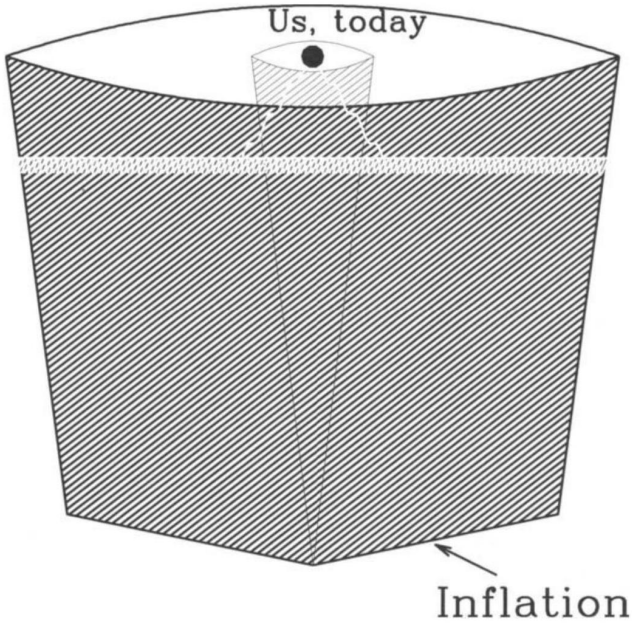


Figure 6.5. Inflationary solution to the horizon problem. Larger cone shows the true horizon in an inflationary model; smaller inner cone shows the horizon without inflation. During inflation, the physical horizon blows up very rapidly. All scales in the shaded region were once in causal contact so it is not surprising that the temperature is uniform.

$$\eta = \int_{t_e}^{t} \frac{dt'}{a\left(t'\right)}$$

so that the total comoving horizon is $\eta_{\rm prim} + \eta$. This is the convention we will follow; note that this means that during inflation, η is negative, but always monotonically increasing. A scale leaves the horizon in the sense of Figure 6.4 when $k|\eta|$ becomes less than 1, and returns at late times when $k\eta$ becomes larger than 1.

To sum up, inflation - an epoch in which the universe accelerates – solves the horizon problem. During the accelerated expansion the physical Hubble radius remains fixed, so particles initially in causal contact with one another can no longer communicate. Regions which are separated by vast distances today were actually in causal contact before and during inflation. At that time, these regions were given the necessary initial conditions, the smoothness we observe today, but also, as we will soon see, the small perturbations about smoothness that eventually grew into galaxies and other structure in the universe.

Negative Pressure

We have shown that an accelerating universe can solve the horizon problem. Since general relativity ties the expansion of the universe to the energy in it, we now need to ask what type of energy can produce acceleration. We can get an immediate answer if we appeal to the time-time and space-space components of the zero-order Einstein equations. They are (Eqs. (2.39) and (2.93))

$$\left(\frac{da/dt}{a}\right)^2 = \frac{8\pi G}{3}\rho$$

$$\frac{d^2a/dt^2}{a} + \frac{1}{2}\left(\frac{da/dt}{a}\right)^2 = -4\pi G\mathcal{P}$$

Multiplying the first of these by one-half and then subtracting one from the other eliminates the first derivative of a, leaving

$$\frac{d^2a/dt^2}{a} = -\frac{4\pi G}{3}(\rho + 3\mathcal{P}).$$

Acceleration is defined to mean that d^2a/dt^2 is positive. For this to happen, the terms in parentheses on the right must be negative. So inflation requires

$$\mathcal{P} < -\frac{\rho}{3}$$

Since the energy density is always positive, the pressure must be negative. This result is perhaps not surprising: we saw back in Chapter 2 that the accelerated expansion which cause supernovae to appear very faint can be caused only by dark energy with negative pressure. Inflation was apparently driven by the a similar form of energy, one with $\mathcal{P} < 0$. To reiterate what we emphasized in Chapter 2, negative pressure is not something with which we have any familiarity. Nonrelativistic matter has small positive pressure proportional to temperature divided by mass, while a relativistic gas has $\mathcal{P} = +\rho/3$, again positive. So whatever it is that drives inflation is not ordinary matter or radiation.

Implementation with a Scalar Field

We have become familiar with the fields $\Psi(\vec{x}, t)$ and $\Phi(\vec{x}, t)$, deriving equations for them which govern their evolution (the Einstein equations) and the evolution of particles which are affected by them. These two fields are parts of the multicomponent field, the metric $g_{\mu\nu}$. The metric is one of the fundamental fields in physics, but there are others. Every elementary particle - the electron, neutrino, quarks, photon, etc. - is associated with its own field. It would be wonderful if one of these fields, the electromagnetic potential associated with photons say, could serve as the source for an inflationary model. Unfortunately we do not yet have such a concrete model. Instead, I will discuss inflation in terms of a generic scalar field (not a fermion like the quarks and leptons or a vector like the electromagnetic field). The simplest version of the standard model does indeed have within it one such field, the Higgs field. But, again unfortunately, we know too much about the Higgs of the standard model. Its interactions and properties are constrained enough for us to know that it cannot serve as the source for inflation. So we will drop any pretensions of connecting the generic scalar field which drives inflation to known physics. Making this connection is left as a homework problem for a future Nobel laureate.

We want to know if a scalar field - which I will call $\phi(\vec{x}, t)$, not to be confused with the metric perturbation $\Phi(\vec{x}, t)$ - can have negative $\rho + 3\mathcal{P}$. So our first task is to write down the energy-momentum tensor for ϕ. This is

$$T^\alpha_\beta = g^{\alpha\nu}\frac{\partial\phi}{\partial x^\nu}\frac{\partial\phi}{\partial x^\beta} - g^\alpha_\beta\left[\frac{1}{2}g^{\mu\nu}\frac{\partial\phi}{\partial x^\mu}\frac{\partial\phi}{\partial x^\nu} + V(\phi)\right].$$

Here $V(\phi)$ is the potential for the field. For example a free field with mass m has a potential $V(\phi) = m^2\phi^2/2$. A warning about signs: if you delve into the literature you will invariably find different signs than those in Eq. (6.24). These are dictated by the choice of metric. Although our metric signature $(-,+,+,+)$ is probably most common in the context of cosmology, it is probably not as common in particle physics. Beware. We will assume that ϕ is mostly homogeneous, consisting of a zero-order part, $\phi^{(0)}(t)$, and a first-order perturbation, $\delta\phi(\vec{x}, t)$. In this section we will derive information about the zero-order homogeneous part, its energy density and pressure, and its time evolution. Later we will consider its perturbations, $\delta\phi$, and how they are generated.

For the homogeneous part of the field, only time derivatives of ϕ are relevant so the indices α and β in the first term in Eq. (6.24) and μ, ν in the second must be equal to zero. The energy-momentum tensor then reduces to

$$T_{\beta}^{(0)\alpha} = -g^{\alpha}{}_{0} g^{0}_{\beta} \left(\frac{d\phi^{(0)}}{dt} \right)^2 + g^{\alpha}{}_{\beta} \left[\frac{1}{2} \left(\frac{d\phi^{(0)}}{dt} \right)^2 - V\left(\phi^{(0)}\right) \right].$$

The time-time component of T^0_0 is equal to $-\rho$, so the energy density is

$$\rho = \frac{1}{2} \left(\frac{d\phi^{(0)}}{dt} \right)^2 + V\left(\phi^{(0)}\right)$$

The first term here is the kinetic energy density of the field, the second its potential energy density. A homogeneous scalar field therefore has much the same dynamics as a single particle moving in a potential [think of $\phi^{(0)}(t)$ as the position of the particle $x(t)$]. In fact this analogy dominates even the language used to describe inflation. The pressure for the homogeneous field is $\mathcal{P} = T_i^{(0)i}$ (no sum over spatial index i), so

$$\mathcal{P} = \frac{1}{2} \left(\frac{d\phi^{(0)}}{dt} \right)^2 - V\left(\phi^{(0)}\right)$$

A field configuration with negative pressure is therefore one with more potential energy than kinetic. An example is shown in Figure 6.6, in which a field is trapped in a false vacuum, i.e., a local, but not the global, minimum of the potential.

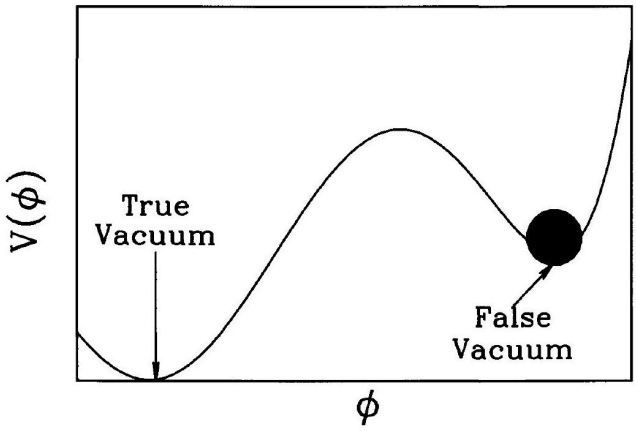


Figure 6.6. A scalar field trapped in a false vacuum. Since it is trapped, it has little kinetic energy. The potential energy is nonzero, however, so the pressure is negative. The global minimum of the potential is called the true vacuum, since a homogeneous field sitting at the global minimum of the potential is in the ground state of the system.

There is something important to notice about a field trapped in a false vacuum. Since $\phi^{(0)}$ is constant, its energy density, which is all potential, remains constant with time. Constant energy density is much different than anything with which we are familiar. The densities of both matter and radiation, for example, fall off very rapidly as the universe expands. Therefore, even if the universe initially contains a mixture of matter, radiation, and false vacuum energy, it will quickly become dominated by the vacuum energy. For a trapped field, it is trivial to determine the evolution of the scale factor. Since the energy density is constant, Einstein's equation for the evolution of a is

$$\frac{da/dt}{a} = \sqrt{\frac{8\pi G\rho}{3}} = \text{ constant}$$

We immediately see that a field trapped in a false vacuum produces exponential expansion as in Eq. (6.19), with $H \propto \rho^{1/2}$ constant. The primordial comoving horizon, that generated before the end of inflation, is then obtained by integrating the inverse of Eq. (6.19) over time,

$$\eta_{\rm prim} = \frac{1}{H_e a_e}\left(e^{H(t_e - t_b)} - 1\right)$$

where t_b is the beginning of inflation. So if the field is trapped for at least 60 e-foldings ($H\,(t_e - t_b) > 60$), the horizon problem is solved. Guth's (1981) initial formulation of inflation used a scalar field trapped in a false minimum of the potential, but it was quickly realized (Guth and Weinberg, 1983; Hawking, Moss, and Stewart, 1982) that such a scenario is not viable. The only way for the field to evolve to its true minimum is similar to the way an alpha particle migrates out of the potential barrier in a nucleus: it tunnels quantum mechanically. Thus, initially small localized regions tunnel from the false to the true vacuum. These bubbles of the true vacuum state must coalesce in order for the universe as a whole to move to the true vacuum state. Careful calculations showed that these bubbles would never coalesce, that the regions of false vacuum would expand rapidly and remain, so that the true vacuum state of the universe would never be attained.

To avoid the problem of the universe never reaching its true vacuum state, subsequent models of inflation (Linde, 1982; Albrecht and Steinhardt, 1982) made use of a scalar field slowly rolling toward its true ground state. The energy density of such a field is also very close to constant (if the potential is not too steep) so it quickly comes to dominate. To determine the evolution of $\phi^{(0)}$ in general when the field is not trapped, we return to the Einstein equations as given in Eq. (6.21). Consider the first of these. If the dominant component in the universe is ϕ, then the energy density on the right-hand side becomes $\left(d\phi^{(0)}/dt\right)^2/2 + V$. Differentiating this first equation therefore leads to

$$2\frac{da/dt}{a}\left[\frac{d^2a/dt^2}{a} - \left(\frac{da/dt}{a}\right)^2\right] = \frac{8\pi G}{3}\left[\left(\frac{d\phi^{(0)}}{dt}\right)\left(\frac{d^2\phi^{(0)}}{dt^2}\right) + V'\frac{d\phi^{(0)}}{dt}\right]$$

where V' is defined as the derivative of V with respect to the field $\phi^{(0)}$. We can replace the first term in brackets on the left by $-4\pi G(\rho/3 + \mathcal{P})$ as in Eq. (6.22). Similarly the second term on the left is $8\pi G\rho/3$. The left-hand side therefore becomes

$$\frac{da/dt}{a}8\pi G[-(\rho/3) - \mathcal{P} - 2\rho/3] = -8\pi G H\left(\frac{d\phi^{(0)}}{dt}\right)^2$$

Equating this to the right side of Eq. (6.30) leads to the evolution equation for a homogeneous scalar field in an expanding universe,

$$\frac{d^2\phi^{(0)}}{dt^2} + 3H\frac{d\phi^{(0)}}{dt} + V' = 0.$$

A more useful form for us will be with the conformal time η as the time variable; then it is straightforward (Exercise 8) to show that

$$\ddot{\phi}^{(0)} + 2aH\dot{\phi}^{(0)} + a^2V' = 0$$

where overdots still denote derivatives with respect to conformal time η.

Most models of inflation are slow roll models, in which the zero-order field, and hence the Hubble rate, vary slowly. Therefore, a simple relation between the conformal time η and the expansion rate holds. In particular, during inflation

$$\eta := \int_{a_e}^{a} \frac{da}{Ha^2}$$
$$\simeq \frac{1}{H}\int_{a_e}^{a} \frac{da}{a^2}$$
$$\simeq \frac{-1}{aH}$$

where the rough equality on the second line holds because H is nearly constant, and the one on the third because the scale factor at the end of inflation is much larger than in the middle ($a_e \gg a$). To quantify slow roll, cosmologists typically define two variables which vanish in the limit that ϕ remains constant. First, define

$$\epsilon := \frac{d}{dt}\left(\frac{1}{H}\right) = \frac{-\dot{H}}{aH^2}$$

Since H is always decreasing, ϵ is always positive. During inflation, it is typically small, whereas it is equal to 2 during a radiation era. In fact, one definition of an inflationary epoch is one in which $\epsilon < 1$. A complementary variable which also quantifies how slowly the field is rolling is:

$$\delta := \frac{1}{H}\frac{d^2\phi^{(0)}/dt^2}{d\phi^{(0)}/dt} = \frac{-1}{aH\dot{\phi}^{(0)}}\left[aH\dot{\phi}^{(0)} - \ddot{\phi}^{(0)}\right]$$
$$= \frac{-1}{aH\dot{\phi}^{(0)}}\left[3aH\dot{\phi}^{(0)} + a^2V'\right]$$

Here the paucity of Greek letters becomes a hindrance. The second slow-roll parameter is more conventionally defined as η, but we obviously cannot follow that convention as η is our conformal time. (Early universe cosmologists use τ for conformal time, freeing up η, but we do not have that luxury since we need τ for optical depth.) My choice of δ is also fairly common, but we need to bear in mind that this has nothing to do with the overdensities introduced in Chapter 4. The second line here follows from Eq. (6.33). Again, in most models δ is small. We will see in Section 6.6 that some unique features of inflation, deviations from the simplest possible spectrum and the production of gravity waves, are proportional to ϵ and δ. If these features are one day measured, they will not only be unique signatures of inflation but also allow us to learn something about the physics driving inflation.

12.6.4 Gravity Wave Production

Theory

Inflation does more than solve the horizon problem. The power of inflation is its ability to correlate scales that would otherwise be disconnected. The zero-order scheme outlined in the previous section ensures that the universe will be uniform on all scales of interest today. There are perturbations about this zero-order scheme, though, and these perturbations - produced early on when the scales are causally connected - persist long after inflation has terminated.

We are most interested in scalar perturbations to the metric since these couple to the density of matter and radiation and ultimately are responsible for most of the inhomogeneities and anisotropies in the universe. In Section 6.5 we will study these in detail. In addition to scalar perturbations, though, inflation also generates tensor fluctuations in the gravitational metric, so-called gravity waves. As we saw in Chapter 5, these are not coupled to the density and so

are not responsible for the large-scale structure of the universe, but they do induce fluctuations in the CMB. In fact, these fluctuations turn out to be a unique signature of inflation and offer the best window on the physics driving inflation, so they are clearly worthy of our study. I choose to study the production of tensor perturbations first before scalar perturbations for a subtle technical reason. Tensor perturbations to the metric are not coupled to any of the other perturbation variables, [2] so when we consider them, we will be looking at the fluctuations in a single field. Scalar perturbations to the metric couple to energy density fluctutations. The coupled fields fluctuate together and this coupling requires a bit of work to understand. This work, while important, is not the main point: the most important idea is that quantum mechanical fluctuations during inflation are responsible for the variations around the smooth background that so fascinate us. This idea is best introduced in the much simpler context of a single field, so we start with tensor perturbations.

During inflation, the universe consists primarily of a uniform scalar field and a uniform background metric. Against this background, the fields fluctuate quantum mechanically. At any given time, the average fluctuation is zero, because there are regions in which the field is slightly larger than its average value and regions in which it is smaller. The average of the square of the fluctuations (the variance), however, is not zero. Our goal is to compute this variance and see how it evolves as inflation progresses. Looking ahead, once we know this variance, we can draw from a distribution with this variance to set the initial [3] conditions.

Quantizing the Harmonic Oscillator

In order to compute the quantum fluctuations in the metric, we need to quantize the field. The way to do this, in the case of both tensor and scalar perturbations, is to rewrite the problem so that it looks like a simple harmonic oscillator. Once that is done, we will appeal to our knowledge of this simple system. Therefore, let's first record some basic facts about the quantization of the harmonic oscillator.

- A simple harmonic oscillator with unit mass and frequency ω is governed by the equation

$$\frac{d^2x}{dt^2} + \omega^2 x = 0$$

- Upon quantization, x becomes a quantum operator

$$\hat{x} = v(\omega, t)\hat{a} + v^*(\omega, t)\hat{a}^\dagger$$

[2] This is not quite true. The quadrupole moments act as sources for tensor perturbations, but these vanish if a scalar field drives inflation. See Exercise 10

[3] Initial here means those when the modes of interest are still far outside the horizon. This is well before any processing can take place, but well after inflation has generated them. where $\hat{a}$ is a quantum operator which acts on the state of the system, and v is a solution to Eq. (6.37), $v \propto e^{-i\omega t}$.

- In particular, $\hat{a}$ annihilates the vacuum state $|0\rangle$, in which there are no particles. It also satisfies the commutation relation

$$\left[\hat{a}, \hat{a}^\dagger\right] := \hat{a}\hat{a}^\dagger - \hat{a}^\dagger\hat{a} = 1.$$

Other commutators vanish: $[\hat{a}, \hat{a}] = \left[\hat{a}^\dagger, \hat{a}^\dagger\right] = 0$. It is straightforward to show (Exercise 9) that these commutation relations are equivalent to the (perhaps more familiar) relations between $\hat{x}$ and its momentum $\hat{p}$:

$$[\hat{x}, \hat{p}] = i,$$

as long as v is normalized via

$$v(\omega, t) = \frac{e^{-i\omega t}}{\sqrt{2\omega}}$$

These facts enable us to compute the quantum fluctuations of the operator $\hat{x}$ in the ground state $|0\rangle$:

$$\begin{aligned}\left\langle |\hat{x}|^2 \right\rangle &:= \left\langle 0 \left| \hat{x}^\dagger \hat{x} \right| 0 \right\rangle \\ &= \left\langle 0 \left| \left(v^* \hat{a}^\dagger + v\hat{a}\right)\left(v\hat{a} + v^*\hat{a}^\dagger\right) \right| 0 \right\rangle .\end{aligned}$$

Since $\hat{a}|0\rangle = 0$, the first term in the second set of parentheses vanishes. Similarly, $\langle 0|\hat{a}^\dagger = (a|0\rangle)^\dagger = 0$, so we are left with

$$\begin{aligned}\left\langle |\hat{x}|^2 \right\rangle &= |v(\omega, t)|^2 \left\langle 0 \left| \hat{a}\hat{a}^\dagger \right| 0 \right\rangle \\ &= |v(\omega, t)|^2 \left\langle 0 \left| \left[\hat{a}, \hat{a}^\dagger\right] + \hat{a}^\dagger\hat{a} \right| 0 \right\rangle .\end{aligned}$$

The second term again vanishes since $\hat{a}$ annihilates the vacuum, while the first is unity, so the variance in $\hat{x}$ is

$$\left\langle |\hat{x}|^2 \right\rangle = |v(\omega, t)|^2,$$

in this case $1/2\omega$. This is (almost) all we need to know about quantum fluctuations in order to compute the fluctuations in the early universe generated by inflation.

Tensor Perturbations

Recall that tensor perturbations to the metric are described by two functions h_+ and $h_\times$, each of which obeys Eq. (5.63),

$$\ddot{h} + 2\frac{\dot{a}}{a}\dot{h} + k^2 h = 0.$$

We would like to massage this equation into the form of a harmonic oscillator, so that h can be easily quantized. To do this, define [4]

$$\tilde{h} := \frac{ah}{\sqrt{16\pi G}}$$

Derivatives of h with respect to conformal time can be rewritten as

$$\frac{\dot{h}}{\sqrt{16\pi G}} = \frac{\dot{\tilde{h}}}{a} - \frac{\dot{a}}{a^2}\tilde{h}$$

and

$$\frac{\ddot{h}}{\sqrt{16\pi G}} = \frac{\ddot{\tilde{h}}}{a} - 2\frac{\dot{a}}{a^2}\dot{\tilde{h}} - \frac{\ddot{a}}{a^2}\tilde{h} + 2\frac{(\dot{a})^2}{a^3}\tilde{h}$$

Inserting these into Eq. (6.45), and multiplying by $\sqrt{16\pi G}$, gives

$$\frac{\ddot{\tilde{h}}}{a} - 2\frac{\dot{a}}{a^2}\dot{\tilde{h}} - \frac{\ddot{a}}{a^2}\tilde{h} + 2\frac{(\dot{a})^2}{a^3}\tilde{h} + 2\frac{\dot{a}}{a}\left(\frac{\dot{\tilde{h}}}{a} - \frac{\dot{a}}{a^2}\tilde{h}\right) + k^2\frac{\tilde{h}}{a}$$
$$= \frac{1}{a}\left[\ddot{\tilde{h}} + \left(k^2 - \frac{\ddot{a}}{a}\right)\tilde{h}\right] = 0.$$

This is precisely the form we know how to use. It has no damping term ($\alpha\dot{\tilde{h}}$) so we can immediately write down an expression for the quantum operator

$$\hat{\tilde{h}}(\vec{k},\eta) = v(k,\eta)\hat{a}_{\vec{k}} + v^*(k,\eta)a^\dagger_{\vec{k}},$$

where the coefficients of the creation and annihilation operators satisfy the equation

$$\ddot{v} + \left(k^2 - \frac{\ddot{a}}{a}\right)v = 0$$

We will shortly solve Eq. (6.51), but first let's see how the eventual solution determines the power spectrum of the fluctuations of the tensor perturbations.

[4] Regarding the factor of $\sqrt{16\pi G}$ here, the only way I know of deriving this is to write down the action for the fields $h_{+,\times}$. The kinetic term is then multiplied by a factor of $1/32\pi G$. A canonical scalar field has prefactor equal to a half. So the additional $16\pi G$ must be absorbed into a redefinition of the field. The hard part of this is writing down the action to second order in perturbation variables. We have seen that even first-order perturbations are cumbersome to track. On the other hand, by dimensional analysis - the fact that $h(\vec{x})$ is dimensionless while a canonical scalar field has dimensions equal to mass - we could have guessed that the factor of $m_{\rm Pl} = G^{-1/2}$ is required. Note that this prefactor does not affect the equation for $\tilde{h}$; it simply provides the normalization that becomes important when trying to determine the amplitude of the gravity-wave spectrum. Using our harmonic oscillator analogy, we can write the variance of perturbations in the $\tilde{h}$ field as

$$\left\langle \hat{\tilde{h}}^\dagger(\vec{k},\eta)\hat{\tilde{h}}\left(\vec{k}',\eta\right)\right\rangle = |v(\vec{k},\eta)|^2(2\pi)^3\delta^3\left(\vec{k}-\vec{k}'\right)$$

There is one difference between this expression and the analogous expression for the one-dimensional harmonic oscillator in Eq. (6.44). A quantum field is defined in all space, so it can be considered as a collection, an infinite collection, of oscillators, each at a different spatial position (or, in Fourier space, at different values of $\vec{k}$). The quantum fluctuations in each of these oscillators are independent (as long as the equations are linear) so $\hat{\tilde{h}}(\vec{k})$ is completely uncorrelated with $\hat{\tilde{h}}\left(\overrightarrow{k'}\right)$ if $\vec{k} \neq \vec{k}'$. The Dirac delta function in Eq. (6.52) enforces this independence; the $(2\pi)^3$ allows for the fact that we have moved to the continuum limit. Recalling that $\tilde{h} = ah/\sqrt{16\pi G}$, we see that

$$\left\langle \hat{h}^\dagger(\vec{k},\eta)\hat{h}\left(\vec{k}',\eta\right)\right\rangle = \frac{16\pi G}{a^2}|v(k,\eta)|^2(2\pi)^3\delta^3\left(\vec{k}-\vec{k}'\right)$$
$$:= (2\pi)^3 P_h(k)\delta^3\left(\vec{k}-\vec{k}'\right)$$

where the second line defines the power spectrum of the primordial perturbations to the metric. Conventions for the power spectrum abound in the literature; the one I've chosen in Eq. (6.53) is not the most popular in the early universe community. Often a factor of k^{-3} is added so that the power spectrum is dimensionless. I prefer to omit this factor to be consistent with the large scale structure community which likes its power spectra to have dimensions of k^{-3}. In any event, with this definition,

$$P_h(k) = 16\pi G \frac{|v(k,\eta)|^2}{a^2}$$

We have now reduced the problem of determining the spectrum of tensor perturbations produced during inflation to one of solving a second-order differential equation for $v(k,\eta)$, Eq. (6.51). To solve this equation, we first need to evaluate $\ddot{a}/a$ during inflation. Recall that overdots denote derivative with respect to conformal time, so $\dot{a} = a^2 H \simeq -a/\eta$ by virtue of Eq. (6.34). Therefore, the second derivative of a in Eq. (6.51) is

$$\begin{aligned}\frac{\ddot{a}}{a} &\simeq -\frac{1}{a}\frac{d}{d\eta}\left(\frac{a}{\eta}\right) \\ &\simeq \frac{2}{\eta^2}\end{aligned}$$

So the equation for v is

$$\ddot{v} + \left(k^2 - \frac{2}{\eta^2}\right) v = 0$$

The initial conditions necessary to solve this equation come from considering v at very early times before inflation has done most of its work. At that time, $-\eta$ is large, of order $\eta_{\rm prim}$, so the k^2 term dominates, and the equation reduces precisely to that of the simple harmonic oscillator. In that case, we know (Eq. (6.41)) that the properly normalized solution is $e^{-ik\eta}/\sqrt{2k}$. This knowledge enables us to choose (Exercise 11) the proper solution to Eq. (6.56),

$$v = \frac{e^{-ik\eta}}{\sqrt{2k}}\left[1 - \frac{i}{k\eta}\right]$$

This obviously goes into the correct solution when the mode is well within the horizon ($k|\eta| \gg 1$). Even if you don't work through Exercise 11 (which arrives at the relatively simple solution of Eq. (6.57) in a rather tortured way), you should at least check that the v given here is indeed a solution to Eq. (6.56).

After inflation has worked for many e-folds $k|\eta|$ becomes very small. Now that v has been normalized, we can determine the amplitude of v, and hence the variance of the super-horizon gravitational wave amplitude, by taking the small argument limit of Eq. (6.57):

$$\lim_{-k\eta\to 0} v(k,\eta) = \frac{e^{-ik\eta}}{\sqrt{2k}}\frac{-i}{k\eta}$$

Figure 6.7 shows the evolution of $h \propto v/a$ during inflation. At early times h falls as $1/a$ as inflation reduces the amplitude of the modes. Once $-k\eta$ becomes smaller than unity, the mode leaves the horizon, after which h remains constant.

The primordial power spectrum for tensor modes, which scales as $|v|^2/a^2$, is therefore constant in time after inflation has stretched the mode to be larger than the horizon. This constant determines the initial conditions for the gravity waves, those with which to start off $h_{+,x}$ at early times (where in this context "early" means well after inflation has ended but before decoupling). Equations (6.54) and (6.58) show that this constant is

$$\begin{aligned}P_h(k) &= \frac{16\pi G}{a^2}\frac{1}{2k^3\eta^2} \\ &= \frac{8\pi G H^2}{k^3}.\end{aligned}$$

The second line here follows from Eq. (6.34). We have assumed that H is constant in deriving this result; more generally, H is to be evaluated at the time when the mode of interest

leaves the horizon. This is our final expression for the primordial power spectrum of gravity waves. Detection of these waves would, quite remarkably, measure the Hubble rate during inflation. Since potential energy usually dominates kinetic energy in inflationary models, a measure of H would be tantamount to measuring the potential V, again quite remarkable in view of the likelihood that inflation was generated by physics at energy scales above 10^{15}GeV, 12 orders of magnitude beyond the capacity of present-day accelerators. There is no guarantee that gravity waves produced during inflation will be detectable. Indeed, since $H^2 \propto \rho/m_{\rm Pl}^2$, the power spectrum is proportional to $\rho/m_{\rm Pl}^4$, the energy density at the time

Tensor Modes (Unnormalized)

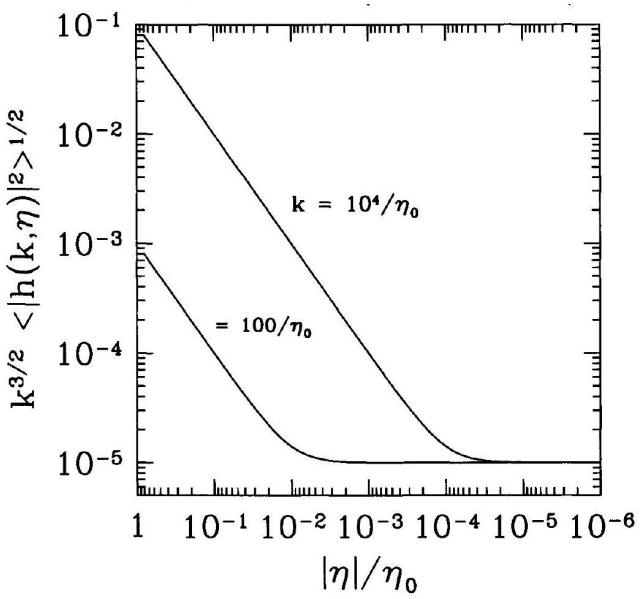


Figure 6.7. The root mean square fluctuations in the tensor field during inflation for two different k-modes. Time evolves from left to right: η is negative but gets closer to zero during inflation. Once a mode "leaves the horizon" ($\eta \sim -1/k$), its RMS amplitude remains constant. Note that after a mode has left the horizon, its RMS amplitude times $k^{3/2}$ is the same for all modes. This is called a scale-free spectrum, strictly true only if the Hubble rate is constant when the scale of interest leaves the horizon (the choice here).

of inflation in units of the Planck mass. If inflation takes place at scales sufficiently smaller than the Planck scale, then primordial gravity waves will not be detected. Later in the book, we will develop the machinery necessary to answer the question, How small can the gravity wave component be and still be detected?

Two final technical points are in order regarding Eq. (6.59). Although I have not emphasized this feature of the spectrum of these primordial perturbations, the fluctuations in h are Gaussian, just as are the quantum-mechanical fluctuations of the simple harmonic oscillator. Gaussianity is a fairly robust prediction of inflation; as such, many studies have been undertaken searching for signs of primordial nonGaussianity in CMB and large-scale structure data, signs that would jeopardize the inflationary picture. Although there have been some hints, none have held up under greater scrutiny, so this prediction of inflation too appears to be verified. Second, Eq. (6.59) is the power spectrum for $h_+$and $h_\times$separately; these are uncorrelated, so the power spectrum for all modes must be multiplied by a factor of 2.

12.6.5 Scalar Perturbations

Theory

The goal of this chapter is to find the perturbation spectrum of Ψ (or Φ; we assume throughout that they are equal in magnitude) emerging from inflation. With that spectrum, we can

use the relations in Section 6.1 to determine the spectrum of the other perturbation variables. Finding the spectrum for Ψ, however, turns out to be complicated, more so than was the tensor case considered earlier. The primary complication is the presence of perturbations in the scalar field driving inflation, perturbations which are coupled to Ψ.

To deal with this problem, we will first ignore it: in Section 6.5.1, we compute the spectrum of perturbations in the scalar field ϕ generated during inflation, neglecting Ψ. This turns out to be relatively simple to do, since it is virtually identical to the tensor calculation we went through above. Why are we justified in neglecting Ψ and how do the perturbations get transferred from ϕ to Ψ? The next two subsections take turns answering this question from two different points of view. First, Section 6.5.2 argues that - in a sense to be defined there - until a mode moves far outside the horizon, Ψ is indeed negligibly small. Once it is far outside the horizon, this no longer holds, but we will find that a linear combination of Ψ and $\delta\phi$ (the perturbations to the scalar field driving inflation) is conserved. This will allow us to convert the initial spectrum for $\delta\phi$ into a final spectrum for Ψ. The second way of justifying the neglect of perturbations to the metric is to switch gauges and work in a gauge in which the spatial part of the metric is unperturbed, a so-called spatially flat slicing. In such a gauge, the calculation of Section 6.5.1 is exact; the only question remaining is how to convert back to conformal Newtonian gauge to move on with the rest of the book. In Section 6.5.3, we identify a gauge-invariant variable, one which does not change upon a gauge transformation, which is proportional to $\delta\phi$ in a spatially flat slicing. It is then a simple matter to determine this variable in conformal Newtonian gauge, thereby linking Ψ in conformal Newtonian gauge to $\delta\phi$ in spatially flat slicing. Note that the two solutions to the coupling problem, as worked out in Sections 6.5.2 and 6.5.3, are simply alternative approaches to the same problem. If you are comfortable with gauge transformations, Section 6.5.3 is probably a more elegant approach; the more brute-force approach of Section 6.5.2 gives the same answer though and requires less formalism and background.

Scalar Field Perturbations around a Smooth Background

Let's decompose the scalar field into a zero-order homogeneous part and a perturbation,

$$\phi(\vec{x},t) = \phi^{(0)}(t) + \delta\phi(\vec{x},t),$$

and find an equation governing $\delta\phi$ in the presence of a smoothly expanding universe, i.e., with metric $g_{00} = -1; g_{ij} = \delta_{ij}a^2$. Consider the conservation of the energymomentum tensor,

$$T^{\mu}_{\nu;\mu} = \frac{\partial T^{\mu}_{\nu}}{\partial x^{\mu}} + \Gamma^{\mu}_{\alpha\mu}T^{\alpha}_{\nu} - \Gamma^{\alpha}_{\nu\mu}T^{\mu}_{\alpha} = 0$$

The $\nu = 0$ component of this equation, expanded out to first order, gives the desired equation for $\delta\phi$. Since we are assuming a smooth metric, the only first-order pieces are perturbations in the energy-momentum tensor. All the Γ 's are either zero-order $\left(\Gamma^{0}_{ij} = \delta_{ij}a^2H\right.$ and $\left.\Gamma^{i}_{0j} = \Gamma^{i}_{j0} = \delta_{ij}H\right)$ or zero (the rest of the components), as we found in Eqs. (2.22) and (2.23). So, writing the perturbed part of the energymomentum tensor as δT^{μ}_{ν} and considering the $\nu = 0$ component of the perturbed conservation equation leads to

$$0 = \frac{\partial \delta T^{0}_{0}}{\partial t} + ik_i\delta T^{i}_{0} + 3H\delta T^{0}_{0} - H\delta T^{i}_{i}$$

It remains to determine the perturbations to the energy-momentum tensor in terms of the perturbations to the scalar field.

First let's compute δT^{i}_{0}. Since the time-space components of the scalar metric are zero, the second set of terms in Eq. (6.24), those with prefactor g^{α}_{β}, vanish. Therefore,

$$T^{i}_{0} = g^{i\nu}\phi_{,\nu}\phi_{,0}$$

where I have returned to using.ν to denote the derivative with respect to x^ν. Since $g^{i\nu} = a^{-2}\delta_{i\nu}$, the index ν must be equal to i. Recall that the zero-order field $\phi^{(0)}$ is homogeneous, so $\phi^{(0)}_{,i} = 0$. The space-time component of the energy-momentum tensor therefore has no zero-order piece. To extract the first-order piece, we can set $\phi_{,i}$ to $\delta\phi_{,i} = ik_i\delta\phi$. Then, setting all other factors to their zero-order values leads to

$$\delta T^i_0 = \frac{ik_i}{a^3}\dot{\phi}^{(0)}\delta\phi$$

The additional factor of a enters the denominator here because $\phi^{(0)}_{,0} = \dot{\phi}^{(0)}/a$ (recall that ' is derivative with respect to conformal time).

The time-time component of the energy-momentum tensor is a little more difficult:

$$T^0_0 = g^{00}\left(\phi_{,0}\right)^2 - \frac{1}{2}g^{\alpha 3}\phi_{.\alpha}\phi_{,\beta} - V$$

Setting $\phi = \phi^{(0)} + \delta\phi$ leads to

$$T^0_0 = \frac{-1}{2}\left(\phi^{(0)}_{,0} + \delta\phi_{.0}\right)^2 - \frac{1}{2a^2}\delta\phi_{,i}\delta\phi_{.i} - V\left(\phi^{(0)} + \delta\phi\right).$$

The spatial derivatives come in pairs, and pairs of first-order variables ($\delta\phi_{.i}$) lcad to second-order terms. These may therefore be neglected. The potential may be expanded as a zero-order term, $V\left(\phi^{(0)}\right)$ plus a first-order correction, $V'\delta\phi$, so the first-order correction to the energy-momentum tensor is

$$\begin{aligned}\delta T^0{}_0 &= -\phi^{(0)}_{.0}\delta\phi, 0 - V'\delta\phi \\ &= \frac{-\dot{\phi}^{(0)}\dot{\delta\phi}}{a^2} - V'\delta\phi.\end{aligned}$$

Similarly, you can show that the space-space component is

$$\delta T^i{}_j = \delta_{ij}\left(\frac{\dot{\phi}^{(0)}\dot{\delta\phi}}{a^2} - V'\delta\phi\right)$$

Therefore, the conservation equation (6.62) becomes

$$\left(\frac{1}{a}\frac{\partial}{\partial\eta} + 3H\right)\left(\frac{-\dot{\phi}^{(0)}\dot{\delta\phi}}{a^2} - V'\delta\phi\right) - \frac{k^2}{a^3}\dot{\phi}^{(0)}\delta\phi - 3H\left(\frac{\dot{\phi}^{(0)}\dot{\delta\phi}}{a^2} - V'\delta\phi\right) = 0.$$

Carrying out the time derivatives (the only subtle one is $\partial V'/\partial\eta = V''\dot{\phi}^{(0)}$), multiplying by a^3, and collecting terms leads to

$$-\dot{\phi}^{(0)}\ddot{\delta\phi} + \dot{\delta\phi}\left(-\ddot{\phi}^{(0)} - 4aH\dot{\phi}^{(0)} - a^2V'\right) + \delta\phi\left(-a^2V''\dot{\phi}^{(0)} - k^2\dot{\phi}^{(0)}\right) = 0$$

The V'' term here is typically small, proportional to the slow-roll variables ϵ and δ (Exercise 14), so it can be neglected. The coefficient of $\dot{\delta\phi}$, the first set of parentheses, is equal to $-2aH\dot{\phi}^{(0)}$ using the zero-order equation (6.33), so after dividing by $-\dot{\phi}^{(0)}$, we are left with

$$\ddot{\delta\phi} + 2aH\dot{\delta\phi} + k^2\delta\phi = 0$$

This equation for perturbations to $\delta\phi$ is identical to Eq. (6.45) for tensor perturbations to the metric. Thus we can trivially copy our result from Section 6.4.2 and write immediately that the power spectrum of fluctuations in $\delta\phi$ is equal to

$$P_{\delta\phi} = \frac{H^2}{2k^3}$$

Compare this with Eq. (6.57). It is identical apart from a factor of $16\pi G$. Recall that we inserted this factor (with a bit of hand-waving: see the footnote on page 158) in the tensor case to turn the dimensionless h into a field with dimensions of mass. To get the result for $\delta\phi$ which is already a scalar field with the proper dimensions, we simply remove this factor.

Super-Horizon Perturbations

Until now, we have neglected the metric perturbations. When the wavelength of the perturbation is of order the horizon or smaller. this approximation is valid, as we will shortly see. In the process of seeing this, we will also find that, by the end of inflation, the metric perturbation has become important. So, although the inflationinduced perturbations start out all- " $\delta\phi$," they end up as a linear combination of Ψ and $\delta\phi$ or more generally as a linear combination of Ψ and perturbations to the energy-momentum tensor. The trick is to find the linear combination which is conserved outside the horizon. The value of this conserved linear combination is determined by $\delta\phi$ at horizon crossing; we can then evaluate it after inflation solely in terms of Ψ. The resulting equation will be of the form $\Psi \propto \delta\phi$ with the left-hand side the post-inflation metric perturbation and the right the scalar field perturbation produced during inflation (the power spectrum for which we have calculated above). We can then finally relate P_Ψ (and the spectra of all other perturbation variables using the results of Section 6.1) to the $P_{\delta\phi}$ of Eq. (6.72).

Let's begin by rewriting the equation for conservation of energy, this time in the presence of the metric perturbation. It is straightforward to show that Eq. (6.62) gets generalized to

$$\frac{\partial \delta T^0_0}{\partial t} + ik_i \delta T^i_0 + 3H\delta T^0_0 - H\delta T^i_i = -3(\mathcal{P}+\rho)\frac{\partial \Psi}{\partial t}$$

where $\mathcal{P}$ and ρ are the zero-order pressure and energy density. Were we correct to neglect Ψ in the last section? We were, as long as the right-hand side is significantly smaller than the individual terms on the left. Taking the first term on the left as an example, we require

$$\Psi \ll \frac{\delta T^0_0}{\mathcal{P}+\rho}$$

A simple way to see that this inequality holds is to use the Einstein equations we derived in Chapter 5. The most convenient for these purposes is the time-time (Eq. (5.21)) component:

$$k^2\Psi + 3aH(\dot{\Psi} + aH\Psi) = 4\pi G a^2 \delta T^0_0$$

Here I have simply copied the results from Chapter 5, replacing Φ with $-\Psi$. The left-hand side here is of order $k^2\Psi \sim a^2H^2\Psi$ for modes which are crossing the horizon. Therefore,

$$\begin{aligned}\Psi &\sim G\frac{\delta T^0_0}{H^2} \sim \frac{\delta T^0_0}{\rho} \\ &= \frac{\mathcal{P}+\rho}{\rho}\left(\frac{\delta T^0_0}{\mathcal{P}+\rho}\right).\end{aligned}$$

The left-hand side must be much less than the term in parentheses; equivalently, the prefactor $(\mathcal{P}+\rho)/\rho$ must be small. In fact, during inflation, the pressure is almost equal to minus the energy density, so this prefactor is very small. In terms of the slow-roll parameters, it is equal to $2\epsilon/3$. So, at least in slow-roll models of inflation, we are justified in neglecting metric perturbations when computing the spectrum of $\delta\phi$.

The above argument holds only for modes that have not yet passed outside the horizon. Super-horizon modes, on the other hand, require more careful treatment. Indeed, it is inevitable that the inequality of Eq. (6.74) will break down sometime before the end of inflation. To see this, recall that after inflation, when the universe is dominated by radiation, $\delta T^0{}_0 = -4\rho_r\Theta_0$ and $\mathcal{P} + \rho = 4\rho_r/3$. Therefore, after inflation, the right-hand side of Eq. (6.74) is $-3\Theta_0$. According to Eq. (6.12), $\Psi = -2\Theta_0$ right after inflation, so it is certainly not true that the inequality of Eq. (6.74) is satisfied for all times. At some point before inflation ends, perturbations to Ψ must grow in importance relative to those in the energy-momentum tensor.

One way to deal with the coupling between the metric perturbations and those to the energy density is to define

$$\zeta := -\frac{ik_i\delta T^0_i H}{k^2(\rho + \mathcal{P})} - \Psi$$

For sub-horizon modes and those which have just left the horizon, Ψ is negligible; $P + \rho = \left(\dot{\phi}^{(0)}/a\right)^2$ from Eqs. (6.26) and (6.27); and Eq. (6.64) fixes the numerator of the first term in ζ. We are left with

$$\zeta = -aH\delta\phi/\dot{\phi}^{(0)}$$

around the time of horizon crossing. After inflation ends, $ik_i\delta T^0{}_i = 4ak\rho_r\Theta_1$, proportional to the dipole of the radiation. Since the pressure of radiation is equal to a third of the energy density,

$$\begin{aligned}\zeta &= -\frac{3aH\Theta_1}{k} - \Psi \\ &= -\frac{3}{2}\Psi \quad \text{(post inflation)}.\end{aligned}$$

The second equality follows from the initial conditions relating the dipole to the potential (Eq. (6.16)).

The variable ζ is so important because it is conserved when the perturbation moves outside the horizon (Figure 6.8). We will show that ζ is conserved shortly, but first let's appreciate the importance of this conservation. Since we know that, after inflation, $\zeta = -3\Psi/2$, we can immediately relate Ψ coming out of inflation to the $\delta\phi$ at horizon crossing,

$$\Psi|_{\text{post inflation}} = \frac{2}{3}aH\frac{\delta\phi}{\dot{\phi}^{(0)}}\bigg|_{\text{horizon crossing}}.$$

Equivalently, the post-inflation power spectrum of Ψ is simply related to the horizon-crossing spectrum of $\delta\phi$:

$$P_\Psi = \frac{4}{9}\left(\frac{aH}{\dot{\phi}^{(0)}}\right)^2 P_{\delta\phi}\bigg|_{aH=k}$$

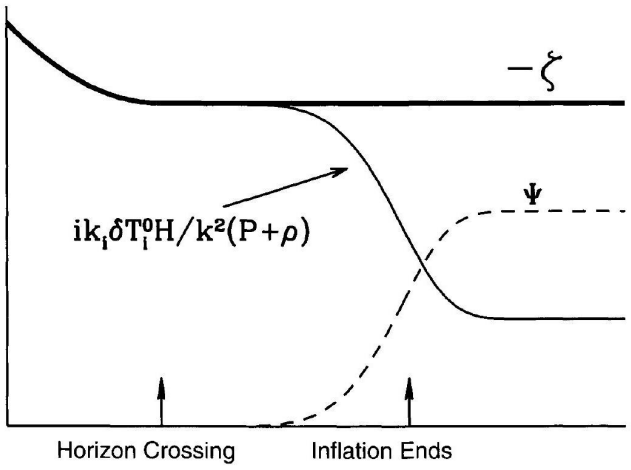


Figure 6.8. Cartoon view of the evolution of scalar, adiabatic perturbations during inflation in conformal Newtonian gauge. When a mode is sub-horizon, quantum-mechanical fluctuations are set up in the scalar field driving inflation $\left(ik_i\delta T^0_i H/k^2(\rho+\mathcal{P}) = aH\delta\phi/\dot{\phi}^{(0)}\right)$. Scalar perturbations to the metric are negligible at this time. Once the mode leaves the horizon, the linear combination $\zeta = -ik_i\delta T^0{}_i H/k^2(\rho+\mathcal{P}) - \Psi$ is conserved. Well after inflation has ended, the metric perturbation has grown in importance, but the linear combination ζ remains unchanged.

$$= \frac{2}{9k^3}\left(\frac{aH^2}{\dot{\phi}^{(0)}}\right)^2\Bigg|_{aH=k}$$

the second line following from Eq. (6.72). Another way to express the power spectrum of scalar perturbations is to eliminate $\dot{\phi}^{(0)}$ in favor of the slow-roll parameter ϵ. You will show (Exercise 12) that $\left(aH/\dot{\phi}^{(0)}\right)^2 = 4\pi G/\epsilon$, so

$$P_\Psi = P_\Phi(k) = \frac{8\pi G}{9k^3}\frac{H^2}{\epsilon}\Bigg|_{aH=k}.$$

The first equality here follows from our ubiquitous assumption that anisotropic strsses are small, so that $\Psi = -\Phi$. Comparing to Eq. (6.59), we see that the ratio of scalar to tensor modes is of order $1/\epsilon$; that is, we expect scalar modes to dominate. Finally, another way of writing the scalar power spectrum is to eliminate ϵ in favor of the potential and its derivative, using the result of Exercise 14,

$$P_\Phi(k) = \frac{128\pi^2 G^2}{9k^3}\left(\frac{H^2V^2}{V'^2}\right)\Bigg|_{aH=k}.$$

It remains to prove that ζ is conserved on super-horizon scales. To see this, let's turn to the conservation equation, Eq. (6.73). On large scales, $k_i\delta T^i{}_0$ is proportional to k^2 and so can be ignored, leaving

$$\frac{\partial\delta T^0_0}{\partial t} + 3H\delta T^0_0 - H\delta T^i_i = -3(\mathcal{P}+\rho)\frac{\partial\Psi}{\partial t}.$$

On large scales, you will show (Exercise 13) that the energy-momentum tensor satisfies

$$\frac{ik_i\delta T^0{}_i H}{k^2} = \frac{\delta T^0_0}{3}$$

Therefore, on large scales

$$\zeta = -\Psi - \frac{1}{3}\frac{\delta T^0_0}{\rho + \mathcal{P}}$$

Eliminating Ψ in favor of ζ in the conservation equation leads to

$$\frac{\partial \delta T^0 0}{\partial t} + 3H\delta T^0{}_0 - H\delta T^i{}_i = 3(\mathcal{P}+\rho)\frac{\partial \zeta}{\partial t} + (\rho+\mathcal{P})\frac{\partial}{\partial t}\left[\frac{\delta T^0{}_0}{\rho+\mathcal{P}}\right]$$

The partial derivative on the right acting on $\delta T^0{}_0$ cancels the first term on the left, leaving

$$\delta T^0_0\left[3H + \frac{1}{\rho+\mathcal{P}}\left(\frac{d\rho}{dt} + \frac{d\mathcal{P}}{dt}\right)\right] - H\delta T^i_i = 3(\mathcal{P}+\rho)\frac{\partial\zeta}{\partial t}$$

Recall from Eq. (2.55) that $d\rho/dt = -3H(\rho+\mathcal{P})$, so the first term in brackets cancels the second, and

$$\frac{\partial\zeta}{\partial t} = \frac{-1}{3(\rho+\mathcal{P})^2}\left[H(\rho+\mathcal{P})\delta T^i{}_i - \delta T^0_0\frac{d\mathcal{P}}{dt}\right].$$

I claim that the two terms in brackets on the right cancel for the class of perturbations we are considering. To see this, first rewrite $H(\rho+\mathcal{P})$ as $-(1/3)d\rho/dt$. Thus, the terms in brackets are proportional to

$$\frac{\delta T^i_i}{3} + \frac{d\mathcal{P}}{d\rho}\delta T^0_0 = \delta\mathcal{P} - \frac{d\mathcal{P}}{d\rho}\delta\rho$$

since $-\delta T^0{}_0$ is the perturbation to the energy density, while $\delta T^i{}_i/3$ is the perturbation to the pressure. If we know the background pressure-energy density relation $d\mathcal{P}/d\rho$, then given an overdensity $\delta\rho$, we expect the pressure perturbation to be proportional to the overdensity with coefficient $d\mathcal{P}/d\rho$. Indeed, this is the characteristic feature of adiabatic perturbations, precisely those set up during inflation. Thus, ζ is indeed conserved on large scales.

Spatially Flat Slicing

The treatment of the previous subsection is complete, but it is not the most elegant way to understand scalar perturbations in inflation. A much simpler way is to move back and forth between different gauges, making use along the way of the concept of a gauge-invariant variable, one which does not change under these transformations. Here I outline this method, leaving some of the more detailed calculations as problems.

We saw earlier that one of the major complications in conformal Newtonian gauge was that perturbations to the scalar field $\delta\phi$ are coupled to the potential Ψ. It would obviously be nice to transform to a gauge in which these perturbations decoupled. Consider a gauge with spatially flat slicing, with the spatial part of the metric $g_{ij} = \delta_{ij}a^2$. In this gauge the line element is

$$ds^2 = -(1+2A)dt^2 - 2aB_{,i}dx^i dt + \delta_{ij}a^2 dx^i dx^j$$

i.e., there are two functions A and B characterizing the perturbations. In this case, the equation for $\delta\phi$ is given exactly (Exercise 16) by Eq. (6.71): the perturbations in the scalar field do not couple to those in the gravitational metric. Therefore, without having to neglect any couplings, we can identify the power spectrum for $\delta\phi$ as given by Eq. (6.72)

The next step is to identify a gauge-invariant variable, one which remains the same when transforming from one gauge to the next. Bardeen (1980) identified several such variables, two characterizing scalar perturbations to the metric and two characterizing perturbations to the

matter. Of course any linear combination of these is still gauge invariant. We would like to identify the combination that is proportional to $\delta\phi$ in the gauge with spatially flat slicing. In this gauge, Bardeen's velocity (Eq. (5.78)) is

$$v = ikB - \frac{ik\dot{\phi}^{(0)}\delta\phi}{(\rho+\mathcal{P})a^2} \quad \text{(spatially flat slicing)}$$

where I have evaluated δT^0_i with Eq. (6.64). Thus, we can create a gauge-invariant variable proportional to $\delta\phi$ in a spatially flat slicing if we subtract off the kB term. Bardeen's Φ_H (Eq. (5.76)) is simply equal to aHB, so the combination

$$\zeta := -\Phi_H - \frac{iaH}{k}v$$

is gauge invariant and in spatially flat slicing is equal to

$$\zeta = -\frac{aH}{\dot{\phi}^{(0)}}\delta\phi \quad \text{(spatially flat slicing)}$$

We can immediately relate the power in ζ to the power in $\delta\phi$,

$$P_\zeta = \left(\frac{aH}{\dot{\phi}(0)}\right)^2 P_{\delta\phi}$$

We know $P_{\delta\phi}$ from Eq. (6.72) and the prefactor is $4\pi G/\epsilon$, so

$$P_\zeta = \left.\frac{2\pi G H^2}{\epsilon k^3}\right|_{aH=k}.$$

Equation (6.96) is very useful, for it expresses the power spectrum of a gauge invariant quantity. Although we computed it in a gauge of the form in Eq. (6.91), once we have this answer, we can compute ζ in any gauge and then relate the power in the perturbation variables of that gauge to P_ζ.

Throughout this book, we have been working in conformal Newtonian gauge. In this gauge, $\Phi_H = -\Phi$, so ζ as defined in Eq. (6.93) is indeed given by Eq. (6.77). We argued in Section 6.5.2 that in conformal Newtonian gauge, after inflation, $\zeta = 3\Phi/2$, so $P_\Phi = 4P_\zeta/9$, or using Eq. (6.96),

$$P_\Phi = \left.\frac{8\pi G H^2}{9\epsilon k^3}\right|_{aH=k}$$

in exact agreement with our earlier calculation finalized in Eq. (6.82).

This is the end of the calculation, but not quite the end of the story. Bardeen and others have argued that Φ_H has a nice geometrical interpretation, one shared by ζ in certain gauges. In particular, the curvature of the three-dimensional space at fixed time is equal to $4k^2\Phi_H/a^2$. Therefore, perturbations in Φ_H represent curvature perturbations: even though the zero-order space is flat, perturbations induce a curvature which varies from place to place. In conformal Newtonian gauge or in a spatially flat slicing this interpretation would seem irrelevant to perturbations in ζ, since ζ is a combination of both Φ_H and the velocity. However, if one moves to a comoving gauge, one in which the velocities vanish, then ζ is equal to Φ_H. In comoving gauges, then, it is clear that a perturbation to ζ is a curvature perturbation, and indeed the scalar perturbations generated during inflation are often called curvature perturbations.

Other topics about initial conditions

12.7 Inhomogeneities

Having set up the system of equations to be solved and the initial conditions for the perturbations, we can now calculate the inhomogeneities and anisotropies in the universe. In this first solutions chapter, we start with the perturbations to the dark matter. In principle these are coupled to all other perturbations. In practice, though, perturbations to the dark matter depend very little on the details of the radiation perturbations. Dark matter, by definition, is affected by radiation only indirectly, through the gravitational potentials. At late times, when the universe is dominated by matter, these potentials are independent of the radiation. At early times, while it is true that the potentials are determined by the radiation, it is also true that the radiation perturbations are relatively simple, so that all moments beyond the monopole and dipole can be neglected. The converse is not true, as we will see in the next chapter: To treat the anisotropies properly we will need to know how the matter perturbations behave.

The ultimate goal of this exercise is to compare theory with observations. We will solve for the evolution of each Fourier mode, $\delta(k, \eta)$. Given this solution, and the initial power spectrum generated by inflation, we can construct the power spectrum of matter today. At least on large scales, this is the most important observable. On small scales, comparison with observation today is more difficult: one must worry about nonlinearities and gas dynamics when comparing with the galaxy distribution. Nonetheless, even on small scales, the linear power spectrum, which we compute in this chapter, is often the starting point for any quantitative statement about the distribution of matter.

12.7.1 Prelude

Gravitational instability is a powerful idea, easy to understand, and most likely responsible for the structure in our universe. As time evolves, matter accumulates in initially overdense regions. It doesn't matter how small the initial overdensity was (e.g., in typical cosmological scenarios, the overdensity was of order 1 part in 10^5); eventually enough matter will be attracted to the region to form structure.

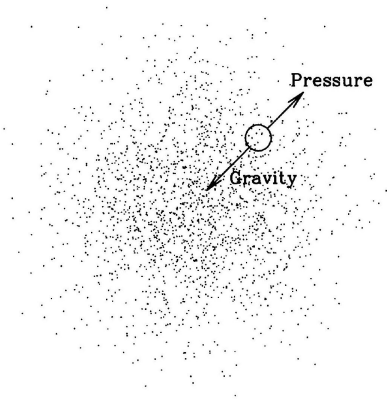


Figure 7.1. Gravitational instability. Mass near an overdense region is attracted to the center by gravity but repelled by pressure. If the region is dense enough, gravity wins and the overdensity grows with time.

The $F = ma$ of gravitational instability is the equation governing overdensities δ. Schematically, it reads

$$\ddot{\delta} + [\text{ Pressure } - \text{ Gravity }]\delta = 0$$

These basic forces, depicted in Figure 7.1, act in opposite directions. Gravity acts to increase overdensities, grabbing more matter into the region. Since there are more particles in an overdense region, random thermal motion causes a net loss of mass in an overdense region. Therefore, if pressure is strong, inhomogeneities do not grow. As indicated by the cartoon equation (7.1), if pressure is low, δ grows exponentially; if it is large, δ oscillates with time.

We will see many manifestations of the simple form of gravitational instability depicted in Eq. (7.1). Different ambient cosmological conditions alter the growth rate. For example, in a matter-dominated universe, δ grows only as a power of time, not exponentially, whereas in a radiation-dominated universe, the growth is but logarithmic. We will treat super-horizon versions of this equation as well as the more familiar sub-horizon version. When going though the math, though, it is useful to bear in mind the dueling concepts of gravity and pressure.

Three Stages of Evolution

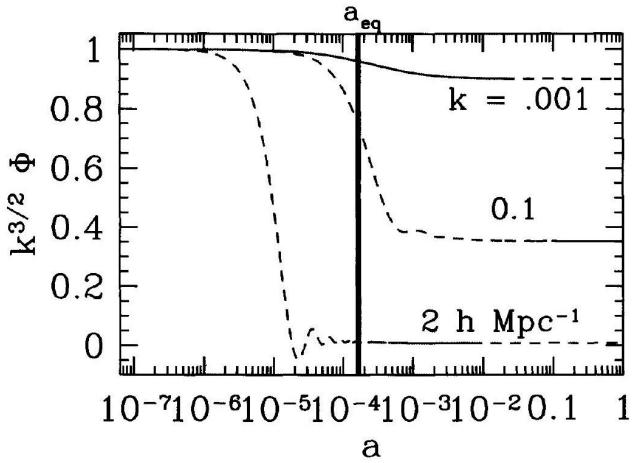


Figure 7.2. The linear evolution of the gravitional potential Φ. Dashed line denotes that the mode has entered the horizon. Evolution through the shaded region is described by the transfer function. The potential is unnormalized, but the relative normalization of the three modes is as it would be for scale-invariant perturbations. Here baryons have been neglected, $\Omega_m = 1$, and $h = 0.5$.

The evolution of cosmological perturbations breaks up naturally into three stages. To see this, let's cheat and look at the solutions for several different modes. Figure 7.2 shows the gravitational potential as a function of scale factor for long-, medium-, and short-wavelength modes. Early on, all of the modes are outside the horizon ($k\eta \ll 1$) and the potential is constant. At intermediate times (shaded in the figure), two things happen: the wavelengths fall within the horizon and the universe evolves from radiation domination ($a \ll a_{\rm eq}$) to matter domination ($a \gg a_{\rm eq}$). Without getting into the details, we see that the order of these epochs ($a_{\rm eq}$ and the epoch of horizon crossing) greatly affects the potential. The large-scale mode, which enters the horizon well after $a_{\rm eq}$, evolves much differently than the small-scale mode, which enters the horizon before equality. Finally, at late times, all the modes evolve identically again, in this case (where $\Omega_m = 1$) remaining constant. We are able to observe the distribution of matter predominantly at late epochs, in the third stage of evolution, when all modes are evolving identically. If we wish to relate the potential during these times to the primordial potential set up during inflation, and we do, we can write schematically

$$\Phi(\vec{k}, a) = \Phi_{\rm p}(\vec{k}) \times \{\text{Transfer Function } (k)\} \times \{\text{ Growth Function } (a)\}$$

where Φ_p is the primordial value of the potential, set during inflation. The transfer function describes the evolution of perturbations through the epochs of horizon cossing and radiation/matter transition (the shaded region in Figure 7.2), while the growth factor describes the

wavelength-independent growth at late times. This schematic equation is indeed roughly how the growth factor and the transfer function are defined, with two caveats, both due to convention. Notice from Figure 7.2 that even the largest wavelength perturbations decline slightly as the universe passes through the epoch of equality. This decline is conventionally removed so that the transfer function on large scales is equal to 1. Therefore, the transfer function is defined as

$$T(k) := \frac{\Phi\left(k, a_{\rm late}\right)}{\Phi_{\text{Large-Scale}}\left(k, a_{\rm late}\right)}$$

where $a_{\rm late}$ denotes an epoch well after the transfer function regime and the LargeScale solution is the primordial Φ decreased by a small amount. We will derive in Section 7.2 that - neglecting anisotropic stresses - this factor is equal to (9/10). The second caveat concerns the growth function. The ratio of the potential to its value right after the transfer function regime is defined to be

$$\frac{\Phi(a)}{\Phi\left(a_{\rm late}\right)} := \frac{D_1(a)}{a} \quad (a > a_{\rm late})$$

where D_1 is called the growth function. In the flat, matter-dominated case depicted in Figure 7.2, then, the potential is constant so $D_1(a) = a$. With these conventions, we have

$$\Phi(\vec{k}, a) = \frac{9}{10}\Phi_{\rm p}(\vec{k})T(k)\frac{D_1(a)}{a} \quad (a > a_{\rm late})\,.$$

The easiest way to probe the potential is to measure the matter distribution. Figure 7.3 shows the evolution of the matter overdensity for three different modes. Notice that at late times - when the potential is constant and all the modes are within the horizon - the overdensity grows with the scale factor ($\delta \propto a$). This explains the seemingly odd nomenclature above (Why is it called a growth function if the potential remains constant?): D_1 describes the growth of the matter perturbations at late times. This growth is completely consistent with our intuition that as time evolves, overdense regions attract more and more matter, thereby becoming more overdense.

We can now express the power spectrum of the matter distribution in terms of the primordial power spectrum generated during inflation, the transfer function, and the growth function. The simplest way to relate the matter overdensity to the

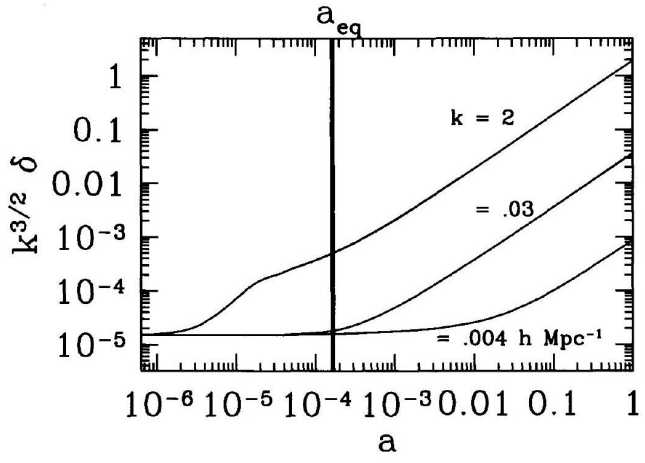


Figure 7.3. The evolution of perturbations to the dark matter in the same model as plotted in Figure 7.2. Amplitude starts to grow upon horizon entry (different times for the three different modes shown here). Well after $a_{\rm eq}$, all sub-horizon modes evolve identically, scaling as

the growth factor. In the case plotted, a flat, matter dominated universe, the growth factor is simply equal to a.

potential at late times is to use Poisson's equation (the large- k, no-radiation limit of Eq. (5.81))

$$\Phi = \frac{4\pi G \rho_{\rm m} a^2 \delta}{k^2} \quad (a > a_{\rm late})$$

The background density of matter is $\rho_{\rm m} = \Omega_m \rho_{\rm cr}/a^3$, and $4\pi G \rho_{\rm cr} = (3/2)H_0^2$, so

$$\delta(\vec{k}, a) = \frac{k^2 \Phi(\vec{k}, a) a}{(3/2)\Omega_m H_0^2} \quad (a > a_{\rm late}\,).$$

This, together with Eq. (7.5), allows us to relate the overdensity today to the primordial potential

$$\delta(\vec{k}, a) = \frac{3}{5}\frac{k^2}{\Omega_m H_0^2}\Phi_{\rm p}(\vec{k}) T(k) D_1(a) \quad (a > a_{\rm late}\,)$$

Equation (7.8) holds regardless of how the initial perturbation Φ_p was generated. In the context of inflation, $\Phi_{\rm p}(\vec{k})$ is drawn from a Gaussian distribution with mean zero and variance (Eq. (6.100)) $P_\Phi = (50\pi^2/9k^3)\,(k/H_0)^{n-1}\,\delta_H^2\,(\Omega_m/D_1(a=1))^2$. So the power spectrum of matter at late times is

$$P(k, a) = 2\pi^2 \delta_H^2 \frac{k^n}{H_0^{n+3}} T^2(k) \left(\frac{D_1(a)}{D_1(a=1)}\right)^2 \quad (a > a_{\rm late}\,).$$

The power spectrum has dimensions of (length $)^3$. If we want to express the power as a dimensionless function, then, we must multiply by k^3. More precisely, one often associates $d^3k P(k)/(2\pi)^3$ with the excess power in a bin of width dk centered at k. After integrating over all orientations of $\vec{k}$, this becomes $(dk/k)\Delta^2(k)$, with

$$\Delta^2(k) := \frac{k^3 P(k)}{2\pi^2}$$

Small Δ then corresponds to small inhomogeneities, while large Δ indicates nonlinear perturbations. Note that, with our conventions, a Harrison-Zel'dovich-Peebles spectrum today has $\Delta^2 = \delta_H^2$ on a horizon-sized scale ($k = H_0$).

Figure 7.4 shows the power spectrum today for two different models. Note that in both of the models $P \propto k$ on large scales, where the transfer function is unity. This behavior is apparent from Eq. (7.9) and corresponds to the simplest inflationary model, wherein $n = 1$. On small scales the power spectrum turns over. To understand this, look back at Figure 7.2. The small-scale mode there ($k = 2h$ Mpc^{-1}) enters the horizon well before matter/radiation equality. During the radiation epoch the potential decays, so the transfer function is much smaller than unity. The effect of this on matter perturbations can be seen in Figure 7.3, where the growth of δ is retarded starting at $a \simeq 10^{-5}$ after the mode has entered the horizon and ending at $a \simeq 10^{-4}$ when the universe becomes matter dominated. Modes that enter the horizon even earlier undergo more suppression. Thus, the power spectrum is a decreasing function of k on small scales.

This leads to the realization that there will be a turnover in the power spectrum at a scale corresponding to the one which enters the horizon at matter/radiation equality. The power of this realization is apparent in Figure 7.4, which shows two different models: one corresponding to a flat, matter-dominated universe today (often called standard Cold Dark Matter or SCDM) and the other a universe with a cosmological constant today (Lambda Cold Dark Matter or Λ

CDM). The major difference between the two models is that sCDM has more matter ($\Omega_m = 1$) and hence an earlier $a_{\rm eq}$. An earlier $a_{\rm eq}$ means only the very small scales enter the horizon during the radiation-dominated epoch, and therefore the turnover occurs on smaller scales. Finally, another important scale to keep in mind is the scale above which nonlinearities cannot be ignored. This is roughly set by $\Delta(k_{\rm nl}) \simeq 1$, which corresponds to $k_{\rm nl} \simeq 0.2h{\rm Mpc}^{-1}$ in most models. The power spectra shown in Figure 7.4 are the linear power spectra today. On scales smaller than $k_{\rm nl}$, one cannot blindly compare the spectra from Figure 7.4 with the matter distribution today.

Method

What are the evolution equations for the dark matter overdensity? In principle, these are the full set of Boltzmann equations derived in Chapter 4 and the pair of Einstein equations from Chapter 5. In practice, though, the full set of equations is not needed. To understand why, recall that early on (before recombination at

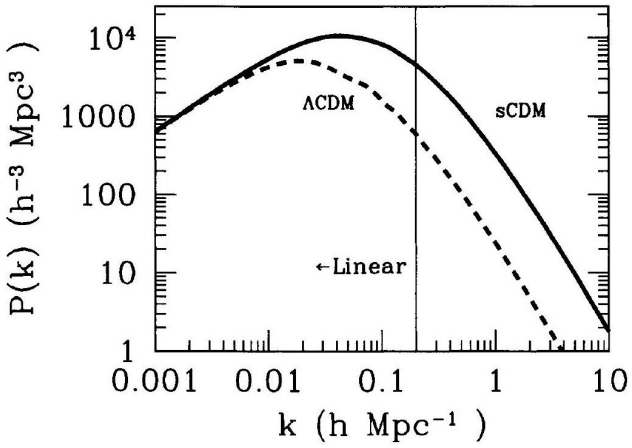


Figure 7.4. The power spectrum in two Cold Dark Matter models, with (ΛCDM) and without (sCDM) a cosmological constant. The spectra have been normalized to agree on large scales. The spectrum in the cosmological constant model turns over on larger scales because of a later $a_{\rm eq}$. Scales to the left of the vertical line are still evolving linearly.

$a = a_*$), the photon distribution can be characterized by only two moments, the monopole Θ_0 and the dipole Θ_1. All other moments are suppressed because the photons are tightly coupled to the electron/proton gas. After decoupling this ceases to be true, and to completely characterize the photon distribution we will need to follow high moments. However, for the purposes of the matter distribution, what the photons are doing after a_* is irrelevant. For, by that time, which is typically well into the matter era, the potential is dominated by the dark matter itself. To sum up then, we can neglect all photon moments except for the monopole and dipole when we are considering the evolution of the matter distribution.

Neglecting the higher radiation moments, the four relevant Boltzmann equations (Section 4.7) become

$$\dot{\Theta}_{r.0} + k\Theta_{r.1} = -\dot{\Phi}$$
$$\dot{\Theta}_{r.1} - \frac{k}{3}\Theta_{r.0} = \frac{-k}{3}\Phi$$
$$\dot{\delta} + ikv = -3\dot{\Phi}$$
$$\dot{v} + \frac{\dot{a}}{a}v = ik\Phi$$

Even with the assumption that only the monopole and dipole are retained, getting from Eq. (4.100) to Eqs. (7.11) and (7.12) requires some explanation and work. First, the explanation: The subscript $_r$ here refers to radiation, both neutrinos and photons. Both species contribute to the gravitational potential (which is our interest in this chapter) and both start out with the same initial conditions. It is not quite as obvious that both follow the same evolution equations (the $\dot{\tau}$ terms can be neglected in Eq. (4.100)) or that these evolution equations are the ones given in (7.11) and (7.12). But it is true, at least in the limit of small baryon density, and again only for the purposes of following the matter evolution. You can work out the details in Exercise 1, and we will explore the full photon evolution equation in the next chapter.

To close the set of equations for the dark matter density, we need an equation for the gravitational potential Φ. You may have noticed that in Eqs. (7.11)-(7.14), I set $\Psi \to -\Phi$, an approximation valid in the limit that there are no quadrupole moments (Eq. (5.33)). Since some of the Einstein equations are redundant, we have several choices for one last equation relating Φ to the radiation and matter overdensities. We can use the time-time component, Eq. (5.27),

$$k^2\Phi + 3\frac{\dot{a}}{a}\left(\dot{\Phi} + \frac{\dot{a}}{a}\Phi\right) = 4\pi G a^2\left[\rho_{\rm dm}\delta + 4\rho_r\Theta_{r,0}\right]$$

Here, again I have set Ψ to $-\Phi$, neglected the baryons, [1] and merged the neutrino and photon contributions to the potential. The alternative is to use the algebraic (no time derivatives) equation (5.81):

$$k^2\Phi = 4\pi G a^2\left[\rho_{\rm dm}\delta + 4\rho_r\Theta_{r,0} + \frac{3aH}{k}\left(i\rho_{\rm dm}v + 4\rho_r\Theta_{r,1}\right)\right].$$

Both of these equations will be useful to us at various times, although only one is necessary to close the set of equations for the five variables $\delta, v, \Theta_{r,0}, \Theta_{r,1}$, and Φ.

At this stage, the simplest thing to do is solve the set of five coupled equations numerically (Exercise 2). If Eq. (7.15) is used, there are no numerical difficulties, and with very little work, you can have a code which computes the transfer function (in the absence of baryons) in less than a second.

Analytic solutions for the dark matter density are harder to come by. I know of no analytic solution valid on all scales at all times. To make progress, we will have to take some limits which reduce the full set of five equations to a more managable two or three. The cost is that these limits will be valid only for certain scales at certain times. Patching these analytic solutions together to obtain a reasonable transfer function is as much art as science.

As a guide to this analytic work which will occupy us much of the rest of this chapter, consider Figure 7.5. The solid curve is the comoving horizon (conformal time), which increases with time, equal to about $30h^{-1}$Mpc at the epoch of

[1] This is a fairly good approximation since in most models, the baryon density is much smaller than the dark matter density. We will explore the effects of baryons in Section 7.6.

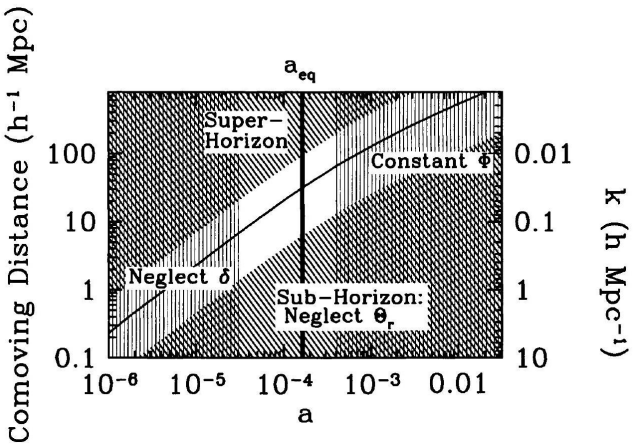


Figure 7.5. Physics of the transfer function. Hatched regions show where analytic expressions exist. The gaps in the center show that no analytic solutions exist to capture the full evolution of intermediate scale modes. The curve monotonically increasing from bottom left to top right is the comoving horizon.

equality. [2] A given comoving scale remains constant with time. Take for example, a comoving distance of $10h^{-1}$Mpc, corresponding to wavenumber $k = 0.1h\text{Mpc}^{-1}$. At early times ($a < 10^{-5}$) this distance is larger than the horizon, so $k\eta \ll 1$. We can then drop all terms proportional to k in the evolution equations. In Section 7.2.1, we will derive an exact solution for the potential in this super-horizon limit. Unfortunately, Figure 7.5 indicates that, for the mode in question, this superhorizon solution is valid only until $a \simeq 10^{-5}$. At much later times ($a > 10^{-3}$) the mode is well within the horizon and the radiation perturbations have become irrelevant (since the universe is matter dominated). We will see in Section 7.3.2 that, under these conditions, another analytic solution can be found. The difficulty is matching the super-horizon solution to the sub-horizon solution.

The problem of matching the super-horizon solution to the sub-horizon solution can be solved for very large scale $\left(k < 0.01h\text{Mpc}^{-1}\right)$ and very small scale ($k > 0.5h$ Mpc^{-1}) modes. In the large-scale case, we will see in Section 7.2.2 that once the universe becomes matter dominated, $\Phi =$ constant is a solution to the evolution

[2]This is model dependent; the plot shows SCDM, with $h = 0.5$. equations even as the mode crosses the horizon. This fact serves as a bridge between the super- and sub-horizon solutions, both of which have constant Φ in the matterdominated regime. In the small-scale case, we can neglect matter perturbations as the mode crosses the horizon, since these modes cross the horizon when the universe is deep in the radiation era. Then, once the mode is sufficiently within the horizon, radiation perturbations decay away, and we can match on to the sub-horizon, noradiation perturbation solution of Section 7.3.2.

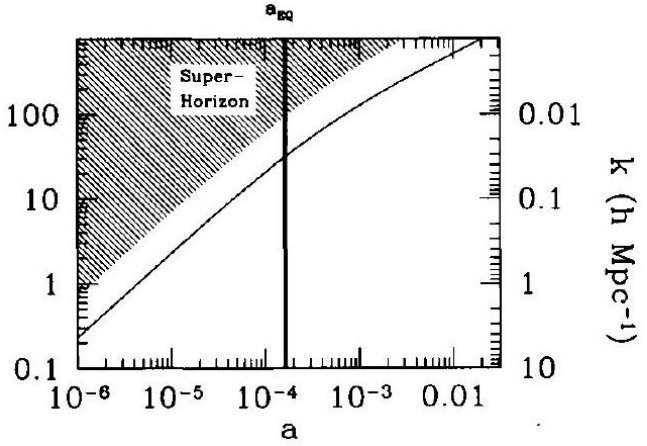
With analytic expressions on both large and small scales, we can obtain a good fit to the transfer function by splining the two solutions together. We will see in Section 7.4 that this works, primarily because the transfer function is so smooth, monotonically decreasing from unity on large scales.

12.7.2 Large Scales

On very large scales, we can get analytic solutions for the potential first through the matter-radiation transition and then through horizon crossing. We start with the super-horizon solution valid through the matter-radiation transition. The results of Section 7.2.1 will be that the potential drops by a factor of 9/10 as the universe goes from radiation to matter domination.

Super-horizon Solution

For modes that are far outside the horizon, $k\eta \ll 1$ and we can drop all terms in the evolution equations dependent on k. From Eqs. (7.11) and (7.13), we see that, in this limit, the velocities (v and $\Theta_{r,1}$) decouple from the evolution equations. This immediately reduces the number of equations to solve from five to three. For the third equation, we notice that Eq. (7.16) has terms inversely propor-



tional to k. These will be difficult to deal with, so let us choose Eq. (7.15) instead. We are left with

$$\begin{aligned}
\dot{\Theta}_{r,0} &= -\dot{\Phi} \\
\dot{\delta} &= -3\dot{\Phi} \\
3\frac{\dot{a}}{a}\left(\dot{\Phi} + \frac{\dot{a}}{a}\Phi\right) &= 4\pi G a^2 \left[\rho_{\rm dm}\delta + 4\rho_r\Theta_{r,0}\right].
\end{aligned}$$

We can go a step further by realizing that the first two equations require $\delta - 3\Theta_{r,0}$ to be constant. Further, we know that this constant is zero (these are the initial conditions). So let us use the dark matter equation (7.18) and the Einstein equation with $\Theta_{r,0}$ set to $\delta/3$. The Einstein equation is then

$$3\frac{\dot{a}}{a}\left(\dot{\Phi} + \frac{\dot{a}}{a}\Phi\right) = 4\pi G a^2 \rho_{\rm dm}\delta\left[1 + \frac{4}{3y}\right]$$

Here I have introduced

$$y := \frac{a}{a_{\rm eq}} = \frac{\rho_{\rm dm}}{\rho_r}$$

which we will use as an evolution variable instead of η or a. Again I emphasize that we are ignoring baryons, so $a_{\rm eq}$ is determined solely by $\rho_{\rm dm}$; in the real world, the numerator in the last term in Eq. (7.21) would be $\rho_{\rm m}$, accounting for all matter including baryons.

Equations (7.18) and (7.20) are two first-order equations for the two variables δ and Φ. The stategy will be to turn these two first-order equations into one secondorder equation and then solve. First, though, let us rewrite the equations in terms of the new variable y. The derivative with respect to y is related to that with respect to η via the Jacobian,

$$\begin{aligned}
\frac{d}{d\eta} &= \frac{dy}{d\eta}\frac{d}{dy} \\
&= aHy\frac{d}{dy},
\end{aligned}$$

where the second line follows from the definition of y and the fact that $\dot{a} = a^2 H$. In terms of y then, the Einstein equation becomes

$$\begin{aligned} y\Phi' + \Phi &= \frac{y}{2(y+1)}\delta\left[1 + \frac{4}{3y}\right] \\ &= \frac{3y+4}{6(y+1)}\delta \end{aligned}$$

where prime denotes derivatives with respect to y and the right side of the first line follows since $8\pi G\rho_{\rm dm}/3 = (8\pi G\rho/3)y/(y+1) = H^2 y/(y+1)$.

In general, to turn two first-order equations into one second-order equation, the trick is to differentiate one of them. Here, to simplify the algebra, we first rewrite Eq. (7.23) as an expression for δ; then differentiate with respect to y; and finally set δ' to $-3\Phi'$ thanks to the dark matter equation (7.18). This leads to

$$-3\Phi' = \frac{d}{dy}\left\{\frac{6(y+1)}{3y+4}\left[y\Phi' + \Phi\right]\right\}$$

Carrying out the derivative is tedious but straightforward. We are left with

$$\Phi'' + \frac{21y^2 + 54y + 32}{2y(y+1)(3y+4)}\Phi' + \frac{\Phi}{y(y+1)(3y+4)} = 0.$$

Remarkably, Kodama and Sasaki (1984) found an analytic solution to Eq. (7.25). They introduced a new variable

$$u := \frac{y^3}{\sqrt{1+y}}\Phi$$

In terms of this variable, you will show (Exercise 4) that Eq. (7.25) becomes

$$u'' + u'\left[\frac{-2}{y} + \frac{3/2}{1+y} - \frac{3}{3y+4}\right] = 0$$

That is, there is no term proportional to u. Instead of a second-order equation for Φ, then, we have a first-order equation for u'. Fortunately, this first-order equation is integrable. Starting from

$$\frac{du'}{u'} = dy\left[\frac{2}{y} - \frac{3/2}{1+y} + \frac{3}{3y+4}\right]$$

we can integrate to get

$$\ln(u') = \text{ constant } + 2\ln(y) - (3/2)\ln(1+y) + \ln(3y+4)$$

Then exponentiating gives

$$u' = A\frac{y^2(3y+4)}{(1+y)^{3/2}}$$

where A is a constant to be determined.

We are one integral away from an analytic expression for the gravitational potential. Remembering the definition of u (Eq. (7.26)), we can integrate Eq. (7.30) to obtain

$$\frac{y^3}{\sqrt{1+y}}\Phi = A\int_0^y dy' \frac{y'^2(3y'+4)}{(1+y')^{3/2}}$$

Note that there should be another constant, $u(0)$, here. However, since $y^3\Phi \to 0$ early on, this constant is vanishes. By similar logic, we can determine the constant A even before performing the integral. For small y, the integrand becomes $4y'^2$, so for small y, Eq. (7.31) becomes $\Phi = 4A/3$. Therefore, $A = 3\Phi(0)/4$. The integral can be done analytically (Exercise 4 again) leaving

$$\Phi = \frac{\Phi(0)}{10}\frac{1}{y^3}\left[16\sqrt{1+y} + 9y^3 + 2y^2 - 8y - 16\right]$$

Equation (7.32) is our final expression for the potential on super-horizon scales. Although it is not obvious, at small y this expression sets $\Phi = \Phi(0)$, a constant. This must be so, since we chose the two constants of integration with precisely this condition. At large y, once the universe has become matter-dominated, the y^3 term in the brackets dominates, so $\Phi \to (9/10)\Phi(0)$. This is precisely the result we were after: the potential on even the largest scales drops by 10% as the universe passes through the epoch of equality.

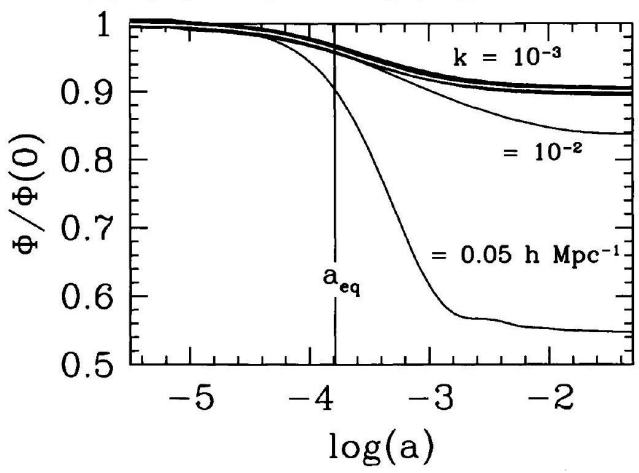


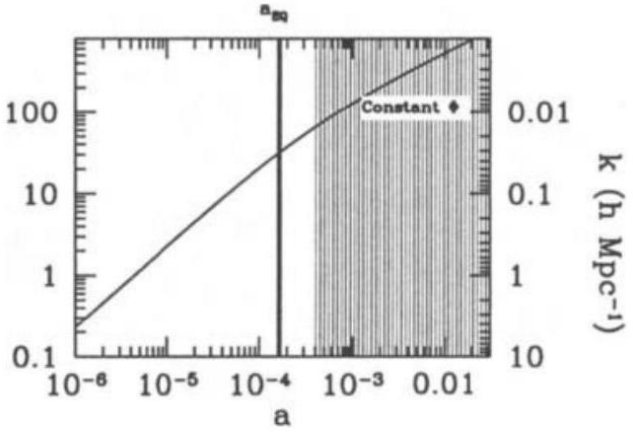
Figure 7.6. Super-horizon evolution of the potential in a CDM model with no baryons, $h = 0.5$ and $\Omega_m = 1$. Thick solid line shows the analytic result of Eq. (7.32), valid only on large scales. White curve within is for the mode $k = 0.001h\text{Mpc}^{-1}$. Two other smaller scale modes are shown.

Let us compare this analytic result, valid only when modes are super-horizon, with the numerical results. Figure 7.6 shows that the solution works perfectly on the largest scales and even tolerably well (better than 10%) for scales as small as $k = 0.01$ hMpc^{-1}. This is slightly better than we had anticipated from a crude estimate of where the super-horizon solution is valid (Figure 7.5) and will be important for us later on when spline together the large and small scales solutions. A feature of the analytic solution which may be surprising to you is that, although it is true that the (large scale) potentials are constant in both the matter and radiation epochs, the transition between the pure matter and pure radiation eras is quite long. For example, and this is an important example for the purposes of the CMB as we will see in the next chapter, the potentials, even for the largest scale modes, are still decaying as late as $a \simeq 10^{-3}$, significantly after $a_{\rm eq}$. In models with less matter, $a_{\rm eq}$ is pushed even closer to 10^{-3} so the decay of the potentials becomes even more apparent at the time of recombination.

Through Horizon Crossing

One interesting feature of Figure 7.6 which you should take note of is that large-scale potential (the numerical solution) becomes constant at very late times ($a \gtrsim 10^{-2}$). For $k = 10^{-3}h$ Mpc^{-1}, the mode enters the horizon at $\eta \sim k^{-1} = 1000h^{-1}$Mpc which corresponds to

$a \sim 0.03$ in the flat, matter-dominated universe depicted in Figure 7.6. The potential remains constant as the mode crosses the horizon.



This result is valid as long as the universe is matter dominated. We now set out to prove it.

We are interested then in our set of five equations in the limit that radiation is not important. The potential depends only on the matter inhomogeneities, so we can neglect the two radiation equations, (7.11) and (7.12). In addition to the two matter equations, we now keep the second of Einstein's equations (7.16). This is an algebraic equation, meaning that we could in principle eliminate Φ in the two matter equations and be left with a system of two first-order differential equations. These two first-order equations in general have two solutions. Instead of solving them directly, though, we can cheat using our knowledge of the initial conditions. Here is the idea: we just learned that, deep in the matter epoch, super-horizon potentials are constant. Therefore, the initial conditions for our problem are that the potential is constant ($\dot{\Phi} = 0$). If we can show that constant Φ is one of the two general solutions to the set of matter-dominated equations, then we don't care what the other solution is. For, the initial conditions ensure that the constant Φ solution will be the solution.

We want to see, then, if the set of equations

$$\begin{aligned} \dot{\delta} + ikv &= 0 \\ \dot{v} + aHv &= ik\Phi \\ k^2\Phi &= \frac{3}{2}a^2H^2\left[\delta + \frac{3aHiv}{k}\right] \end{aligned}$$

admits a solution with Φ a constant in time. We can use the algebraic equation (7.35) to eliminate δ from the other two equations. In the matter dominated era, $H \propto a^{-3/2}$, so $d(aH)/d\eta = -a^2H^2/2$. Replacing δ in Eq. (7.33) with Φ and v therefore leads to

$$\frac{2k^2\dot{\Phi}}{3a^2H^2} + \frac{2k^2\Phi}{3aH} - \frac{3aHiv}{k} + \frac{3a^2H^2iv}{2k} + ikv = 0$$

We now have two first-order equations for Φ and v. The strategy is to turn these two equations into one second-order equation for Φ. First eliminate $\dot{v}$ from Eq. (7.36) by using the velocity equation. This leaves

$$\frac{2k^2\dot{\Phi}}{3a^2H^2} + \left[\frac{iv}{k} + \frac{2\Phi}{3aH}\right]\left(\frac{9a^2H^2}{2} + k^2\right) = 0$$

If the second-order equation is of the form $\alpha\ddot{\Phi} + \beta\dot{\Phi} = 0$, that is, if it has no terms proportional to Φ, then $\Phi =$ constant is a solution to the equations. So we differentiate Eq. (7.37) with respect to η but consider only the terms proportional to Φ, dropping all terms proportional to derivatives of Φ. Using the fact that $(d/d\eta)(aH)^{-1} = 1/2$, we see that the remaining terms are

$$\left[\frac{i\dot{v}}{k}+\frac{\Phi}{3}\right]\left(\frac{9a^2H^2}{2}+k^2\right)+\left[\frac{iv}{k}+\frac{2\Phi}{3aH}\right]\frac{d}{d\eta}\frac{9a^2H^2}{2}$$
$$=-\left[\frac{iaHv}{k}+\frac{2\Phi}{3}\right]\left(9a^2H^2+k^2\right)$$

where I have eliminated $\dot{v}$ by using the velocity equation again. But Eq. (7.37) tells us that the term in square brackets on the right here is proportional to $\dot{\Phi}$. So there are no terms in the second-order equation proportional to Φ. Constant potentials are therefore a solution in the matter-dominated era. Since the initial conditions pick out this mode, constant potential is the solution in the matter-dominated era.

Potentials remain constant as long as the universe is matter dominated. At much later times ($a > 1/10$), it is conceivable that the universe becomes dominated by some other form of energy - dark energy for example - or, less likely, by curvature. If so, then the potentials will decay. This decay is described by the growth function, though (Section 7.5), and does not affect the transfer function. The main result of this section is that the transfer function as defined in Eq. (7.3) is very close to unity on all scales that enter the horizon after the universe becomes matter dominated. That is, it is unity for all $k \ll a_{\rm eq}H\,(a_{\rm eq})$, the inverse comoving Hubble radius at equality. You will show in Exercise 5 that the relevant scale is

$$k_{\rm eq} = 0.073{\rm Mpc}^{-1}\Omega_m h^2$$

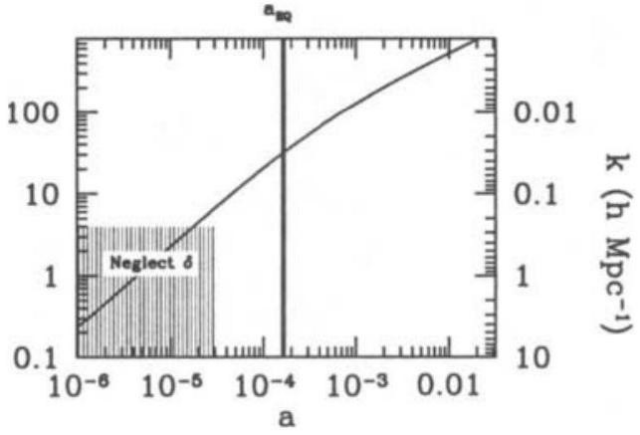
In the limit in which we are working, where baryons and anisotropic stresses are neglected, the transfer function depends only on $k/k_{\rm eq}$. To get a feel for when the large-scale approximations of this section are valid, look back at Figure 7.6, plotted for the standard CDM model with $\Omega_m = 1$ and $h = 0.5$. The transfer function for the curve labeled 10^{-2} is 7% lower (0.84/0.9) than unity. For that mode, $k/k_{\rm eq} = 0.01/(0.073h) = 0.27$. So if we are interested in 10% accuracy in the transfer function, then we can use the large-scale approximation for $k \lesssim k_{\rm eq}/3$.

12.7.3 Small Scales

We were able to solve for the evolution of large-scale perturbations in the previous section because the modes crossed the horizon well after the epoch of equality. Therefore, the problem neatly divided into (i) super-horizon modes passing through the epoch of equality and then (ii) modes in the matter-dominated era which cross the horizon. The converse is true for the small-scale modes considered in this section. They cross the horizon when the universe is deep in the radiation era. So the problem divides neatly into (i) modes in the radiation era crossing the horizon and then (ii) sub-horizon modes passing through the epoch of equality. Step (i) we treat in Section 7.3.1, step (ii) in Section 7.3.2. Notice that we are unable to treat analytically modes which enter the horizon around the epoch of equality.

Horizon Crossing

When the universe is radiation dominated, the potential is determined by perturbations to the radiation. The dark matter perturbations - the ones we are interested in in this chapter are influenced by the potential, but do not themselves influence the potential. So the situation is as depicted in Figure 7.7. To solve for matter perturbations in this epoch, therefore, is a twostep problem. First, we must solve the



coupled equations for $\Theta_{r,0}, \Theta_{r,1}$, and Φ. Then we solve the equation for matter evolution using the potential as an external driving force.

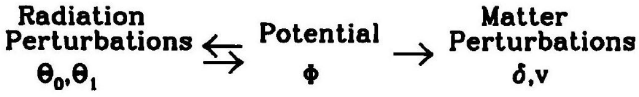


Figure 7.7. Coupling of perturbations in the radiation era. Radiation perturbations and the gravitational potential affect each other. Matter perturbations do not affect the potential but are driven by it.

To solve for the potential in the radiation dominated era, we choose Eq. (7.16). Dropping the matter source terms, we have

$$\Phi = \frac{6a^2H^2}{k^2}\left[\Theta_{r,0} + \frac{3aH}{k}\Theta_{r,1}\right]$$

since $H^2 = 8\pi G\rho_r/3$ in the radiation era. Also in the radiation era, $aH = 1/\eta$. Armed with this fact, we can use Einstein's equation (7.40) to eliminate $\Theta_{r,0}$ from the two radiation equations, (7.11) and (7.12). These become

$$-\frac{3}{k\eta}\dot{\Theta}_{r,1} + k\Theta_{r,1}\left[1 + \frac{3}{k^2\eta^2}\right] = -\dot{\Phi}\left[1 + \frac{k^2\eta^2}{6}\right] - \Phi\frac{k^2\eta}{3}$$

$$\dot{\Theta}_{r,1} + \frac{1}{\eta}\Theta_{r,1} = \frac{-k}{3}\Phi\left[1 - \frac{k^2\eta^2}{6}\right]$$

We can turn these two first order-equations for Φ and $\Theta_{r,1}$ into one second-order equation for the potential. Use Eq. (7.42) to eliminate $\dot{\Theta}_{r,1}$ from the first equation, which then becomes

$$\dot{\Phi} + \frac{1}{\eta}\Phi = \frac{-6}{k\eta^2}\Theta_{r,1}$$

We now have an expression for $\Theta_{r,1}$ solely in terms of the potential and its first derivative. To arrive at a second-order equation for Φ, we differentiate. When we do, we will encounter terms proportional to $\Theta_{r,1}$ and its derivative. Each of these can be eliminated with Eq. (7.42) and Eq. (7.43). The resulting second-order equation is

$$\ddot{\Phi} + \frac{4}{\eta}\dot{\Phi} + \frac{k^2}{3}\Phi = 0$$

To determine the behavior of the potential in the radiation-dominated era, we must solve Eq. (7.44) subject to the initial conditions that Φ is constant. It can be solved analytically by defining $u := \Phi\eta$. Then Eq. (7.44) becomes

$$\ddot{u} + \frac{2}{\eta}\dot{u} + \left(\frac{k^2}{3} - \frac{2}{\eta^2}\right) u = 0.$$

This is the spherical Bessel equation of order 1 (see Eq. (C.13)) with solutions $j_1(k\eta/\sqrt{3})$ - the spherical Bessel function - and $n_1(k\eta/\sqrt{3})$ - the spherical Neumann function. The latter blows up as η gets very small, so we discard it on the basis of the initial conditions. The spherical Bessel function of order 1 can be expressed in terms of trigonometric functions (Eq. (C.14)), so

$$\Phi = 3\Phi_{\rm p}\left(\frac{\sin(k\eta/\sqrt{3}) - (k\eta/\sqrt{3})\cos(k\eta/\sqrt{3})}{(k\eta/\sqrt{3})^3}\right)$$

where $\Phi_{\rm p}$ is the primordial value of Φ. The factor of 3 in front here arises because the $\eta \to 0$ limit of the expression in parentheses is 1/3.

Equation (7.46) tells us that, as soon as a mode enters the horizon during the radiation-dominated era, its potential starts to decay. After decaying, the potential oscillates, as depicted in Figure 7.8. Qualitatively, we could have anticipated as much. From the qualitative discussion surrounding Eq. (7.1), we expected that when the pressure is large, as it is when radiation dominates, perturbations will oscillate in time. If perturbations to the dominant component (here radiation) do not grow, then the potential in an expanding universe will begin to decay simply due to the dilution of the zero-order density. This is evident in Eq. (7.40) which (neglecting the dipole well within the horizon) says that $\Phi \sim \Theta_0/\eta^2$. Since Θ_0 oscillates with fixed amplitude, the potential also ocillates, but with an amplitude decreasing as η^{-2}. Indeed, this is precisely the large $k\eta$ limit of Eq. (7.46). The decay and oscillation of the potential is shown in Figure 7.8, with both the analytic

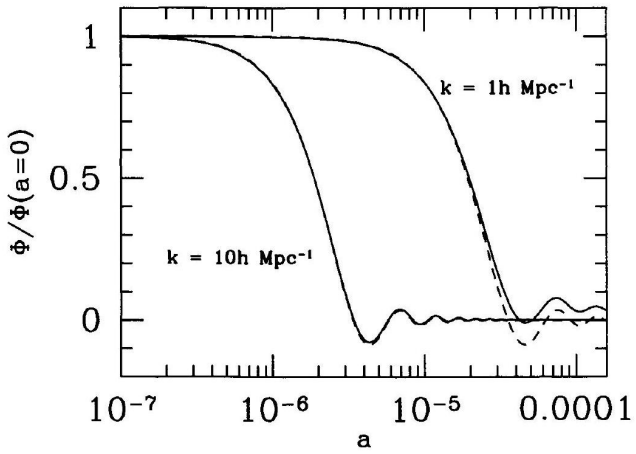


Figure 7.8. Evolution of the potential in the radiation-dominated era. For two small scale modes which enter the horizon well before equality, the exact (solid curve) solution is shown along with the approximate analytic solution (dashed curve) of Eq. (7.46).

expression of Eq. (7.46) and the numerical solution including matter perturbations. Note that the approximate description - in which the effect of matter on the potential is neglected - is valid only deep in the radiation era. The analytic solution for the $k = 1\,h$ Mpc^{-1} mode already begins to depart from the exact solution at $a \simeq 3 \times 10^{-5}$, well before equality (here, in the sCDM model I have taken for illustrative purposes, at $a \simeq 2 \times 10^{-4}$).

Armed with knowledge of the potential in the radiation dominated era, we can now determine the evolution of the matter perturbations, the second half of Figure 7.7. To do this, we turn the two matter evolution equations- (7.13) and (7.14) into one second-order equation with

the potentials serving as an external source. Differentiate Eq. (7.13) and use Eq. (7.14) to eliminate $\dot{v}$:

$$\ddot{\delta} + ik\left(-\frac{\dot{a}}{a}v + ik\Phi\right) = -3\ddot{\Phi}.$$

Now we can use Eq. (7.13) to eliminate v, leading to

$$\ddot{\delta} + \frac{1}{\eta}\dot{\delta} = S(k,\eta)$$

where the source term is

$$S(k,\eta) = -3\ddot{\Phi} + k^2\Phi - \frac{3}{\eta}\dot{\Phi}$$

The two solutions to the homogeneous equation ($S = 0$) associated with Eq. (7.48) are $\delta =$ constant and $\delta = \ln(a)$ (or, equivalently in the radiationdominated era, $\ln[\eta]$). In general, the solution to a second-order equation is a linear combination of the two homogeneous solutions and a particular solution. In the absence of a revelation about the particular solution, one can construct it from scratch from the two homogeneous solutions (call them s_1 and s_2) and the source terms. It is the integral of the source term weighted by the Green's function $\left[s_1(\eta)s_2\left(\eta'\right) - s_1\left(\eta'\right)s_2(\eta)\right] / \left[\dot{s}_1\left(\eta'\right)s_2\left(\eta'\right) - s_1\left(\eta'\right)\dot{s}_2\left(\eta'\right)\right]$. So here, we have

$$\delta(k,\eta) = C_1 + C_2\ln(\eta) - \int_0^\eta d\eta' S\left(k,\eta'\right)\eta'\left(\ln\left[k\eta'\right] - \ln[k\eta]\right).$$

At very early times the integral is small, so our initial conditions (δ constant) dictate that the coefficient of $\ln(\eta), C_2$, vanishes and $C_1 = \delta(k,\eta = 0) = 3\Phi_{\rm p}/2$. Now let us consider the integral in Eq. (7.50). The source function decays to zero along with the potential as the mode enters the horizon. Thus, the dominant contribution to the integral comes from the epochs during which $k\eta$ is of order 1. The integral over $S\left(\eta'\right)\ln\left(k\eta'\right)$ therefore will just asymptote to some constant, while the integral over $S\left(\eta'\right)\ln(k\eta)$ will lead to a term proportional to $\ln(k\eta)$ with the constant of proportionality being just that, a constant. Thus, we expect that after the mode has entered into the horizon,

$$\delta(k,\eta) = A\Phi_{\rm p}\ln(Bk\eta)$$

i.e., a constant ($A\Phi_{\rm p}\ln[B]$) plus a logarithmic growing mode ($A\Phi_{\rm p}\ln[k\eta]$).

We can determine the constants A and B in Eq. (7.51) by referring to the relevant parts of Eq. (7.50). The constant term, $A\Phi_{\rm p}\ln(B)$, is equal to C_1 plus the integral over $\ln\left(\eta'\right)$, or

$$A\Phi_{\rm p}\ln(B) = \frac{3}{2}\Phi_{\rm p} - \int_0^\infty d\eta' S\left(k,\eta'\right)\eta'\ln\left(k\eta'\right)$$

while the coefficient of the $\ln(k\eta)$ term is set by the remaining integral

$$A\Phi_{\rm p} = \int_0^\infty d\eta' S\left(k,\eta'\right)\eta'$$

Note that in both integrals here, I have set the upper limit to infinity in accord with our expectation that the integrals asymptote to some constant value at large η. Using the expression for the source term, Eq. (7.49), and our analytic approximations to the potential, Eq. (7.46),

we can evaluate the integrals here and determine A and B. I find $A = 9.0$ and $B = 0.62$. Hu and Sugiyama (1996), who introduced this method for following the dark matter evolution at early times, found that integrating the exact potentials (instead of the approximate ones of Eq. (7.46)) leads to slightly different values, $A = 9.6$ and $B = 0.44$.

Figure 7.9 shows the exact solution for δ in the radiation era along with the approximation of Eq. (7.51). Setting aside the details for a moment, we see that

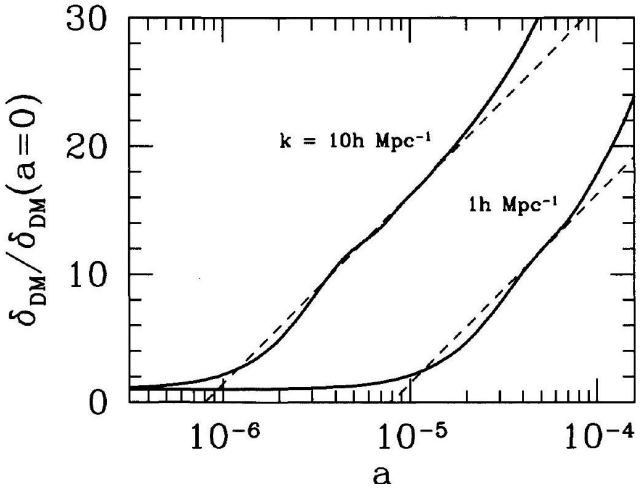
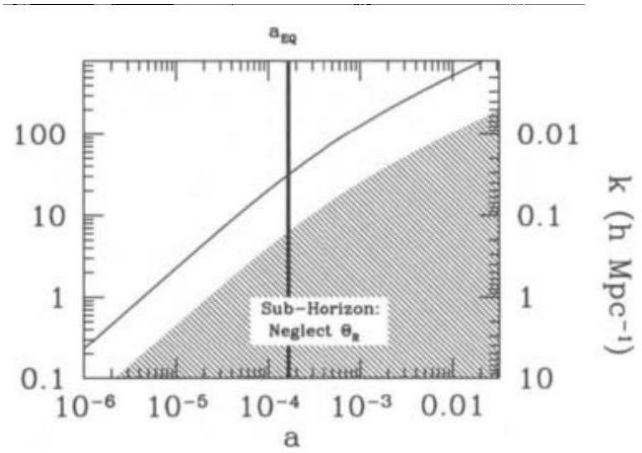


Figure 7.9. Matter perturbations in the radiation-dominated era. The two scales shown here both enter the horizon in the radiation era and lock onto the logarithmically growing mode after some oscillations. Heavy solid curves are the exact solutions, light dashed curves the logarithmic mode of Eq. (7.51). The perturbations have been artificially normalized by their values at early times: inflation actually predicts a larger initial amplitude (by a factor of $10^{3/2}$) for the larger scale mode.

matter perturbations do indeed grow even during the radiation era. The growth is not as prominent as during the matter era (when the constant potentials derived in Section 7.2imply$\delta \propto a$) due to the pressure of the radiation, but it still exists. For both scales shown in Figure 7.9 the perturbations do indeed settle into the logarithmic growing mode once they enter the horizon. As the universe gets closer to matter domination, though, the pressure of the radiation becomes less important, and the perturbations begin to grow faster. Indeed, you might be worried that our approximation for the $k = 1h\mathrm{Mpc}^{-1}$ mode is not very useful. Fortunately, we will be using these solutions only to set the initial conditions for growth in the subhorizon epoch (next subsection), so the approximation need be valid only for a very limited range of times. As long as we choose the matching epoch appropriately, the logarithmic approximation will be extremely good.

Sub-horizon Evolution

We saw in the last subsection that radiation pressure causes the gravitational potentials to decay as modes enter the horizon during the radiation era. Although I did not focus on the radiation perturbations themselves (we will do this in the next chapter), you might expect that the pressure suppresses any growth in $\Theta_{r,0}$. This is correct, and it is in sharp contrast to the matter perturbations which, we



just saw, grow logarithmically. Although initially the potential is determined by the radiation (since the universe is radiation dominated), eventually the growth in the matter perturbations more than offsets the fact that there is more radiation than matter. That is, eventually $\rho_{\rm dm}\delta$ becomes larger than $\rho_r\Theta_{r,0}$ even if $\rho_{\rm dm}$ is smaller than ρ_r. Once this happens, the gravitational potential and the dark matter perturbations evolve together and do not care what happens to the radiation. In this subsection, we want to solve the set of equations governing the matter perturbations and the potential and then match on to the logarithmic solution (7.51) set up during the epoch in which the potential decays.

Once again our starting point is the set of equations governing dark matter evolution, (7.13) and (7.14), and the algebraic equation for the gravitational potential (7.16). And, once again, we want to reduce this set of three equations (two of which are first-order differential equations) to one second-order equation. We will want to follow the sub-horizon dark matter perturbations through the epoch of equality, so it proves convenient again to use y (Eq. (7.21)) - the ratio of the scale factor to its value at equality - as the evolution variable. In terms of y, the three equations become

$$\begin{aligned}\delta' + \frac{ikv}{aHy} &= -3\Phi' \\ v' + \frac{v}{y} &= \frac{ik\Phi}{aHy} \\ k^2\Phi &= \frac{3y}{2(y+1)}a^2H^2\delta.\end{aligned}$$

Several comments are in order about this version of our fundamental equations. First, notice that the time derivatives in the first two equations have been replaced by derivatives with respect to y (indicated by primes), and this transformation leads to the factors of $\dot{y} = aHy$ in the denominators of the unprimed terms. Second, the gravitational potential is now expressed solely in terms of δ: there is no dependence on radiation perturbations because of our arguments above that these are subdominant, and there is no aHv/k dependence because the perturbations are well within the horizon and $aH/k \ll 1$. Finally, the coefficient of the δ source term is $4\pi G\rho_{\rm dm}a^2 \to (3/2)a^2H^2y/(y+1)$ since we are interested in times early enough that any curvature or dark energy is negligible.

We now go through the familiar routine of turning Eqs. (7.54) and (7.55) into a second-order equation for δ: differentiate the first of these to get

$$\delta'' - \frac{ik(2+3y)v}{2aHy^2(1+y)} = -3\Phi'' + \frac{k^2\Phi}{a^2H^2y^2}$$

where v' has been eliminated using the velocity equation. Also I have used the fact that $d(1/aHy)/dy = -(1+y)^{-1}(2aHy)^{-1}$. The first term on the right is much smaller than the

second, since the latter is multiplied by $(k/aH)^2$, and we are focusing on sub-horizon modes. Using Eq. (7.56), we recognize this second term as $3\delta/[2y(y+1)]$. We can rewrite the velocity on the left using Eq. (7.54) but neglecting the potential which on sub-horizon scales is much smaller than δ. Thus, the combination $ikv/(aHy)$ can be simply replaced by $-\delta'$ leaving

$$\delta'' + \frac{2+3y}{2y(y+1)}\delta' - \frac{3}{2y(y+1)}\delta = 0$$

This is the Meszaros equation governing the evolution of sub-horizon cold, dark matter perturbations once radiation perturbations have become negligible.

To understand the growth of dark matter perturbations, we need to obtain the two independent solutions to the Meszaros equations and then match on to the logarithmic mode established in the previous subsection. To solve this differential equation, we can use our knowledge of the solution deep in the matter era. We know that sub-horizon perturbations in the matter era grow with the scale factor, so one of the solutions to Eq. (7.58) is a polynomial in y of order 1. Therefore, for one mode at least, δ'' vanishes. Therefore, the equation governing this first mode, the growing mode,[3] is $D_1'/D_1 = 3/(2+3y)$, the solution to which is

$$D_1(y) = y + 2/3$$

To find the second solution, notice that the Meszaros equation tells us that $u := \delta/D_1$ satisfies

$$(1+3y/2)u'' + \frac{u'}{y(y+1)}\left[(21/4)y^2 + 3y + 1\right] = 0$$

Since there is no term proportional to u, Eq. (7.60) is actually a first-order equation [4] for u'. We can therefore integrate to obtain a solution for u' and then integrate again to get the second Meszaros solution. The first integral gives

[3] D_1 is the growth function mentioned in Section 7.1. Note though that in this section we are assuming that only matter, and not curvature or dark energy, dominates the landscape. Therefore, our expression for the growth function will be valid only when $a \lesssim 0.1$. For the generalization to later times, see Section 7.5.

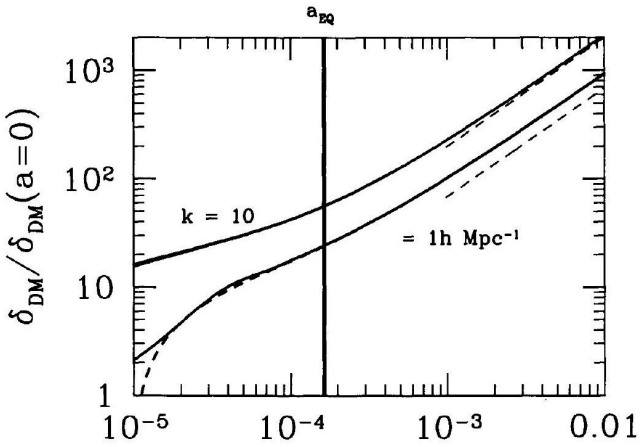
[4] Indeed this is a general trick for obtaining the second solution to a differential equation once the first is known. We will use it again later on to obtain the growth factor.

$$u' \propto (y+2/3)^{-2} y^{-1} (y+1)^{-1/2}$$

Integrating again leads to the second Meszaros solution

$$D_2(y) = D_1(y) \ln\left[\frac{\sqrt{1+y}+1}{\sqrt{1+y}-1}\right] - 2\sqrt{1+y}$$

At late times ($y \gg 1$), the growing solution D_1 scales as y while the decaying mode D_2 falls off as $y^{-3/2}$.



a

Figure 7.10. Evolution of small-scale, sub-horizon, dark matter perturbations. Solid curves are exact solutions; dashed curves (almost imperceptible because the goodness of fit in the $10\,h$ Mpc^{-1} case) the Meszaros solution with coefficients given by the matching condition, Eq. (7.64). The dashed straight lines at $a > 10^{-3}$ are the asymptotic solution of Eq. (7.67).

The general solution to the Meszaros equation is therefore

$$\delta(k, y) = C_1 D_1(y) + C_2 D_2(y) \quad y \gg y_H$$

where y_H is the scale factor when the mode enters the horizon divided by the scale factor at equality (Exercise 6). To determine the constants C_1 and C_2 we can match on to the logarithmic solution of Eq. (7.51). That solution is valid within the horizon but before equality: $y_H \ll y \ll 1$. So we can hope to arrive at a reasonable approximation for the evolution of dark matter perturbations only for those modes that enter the horizon before equality. For those modes, we match the two solutions and their first derivatives

$$A\Phi_{\rm p} \ln\left(By_m/y_H\right) = C_1 D_1\left(y_m\right) + C_2 D_2\left(y_m\right)$$

$$\frac{A\Phi_{\rm p}}{y_m} = C_1 D_1'\left(y_m\right) + C_2 D_2'\left(y_m\right)$$

where the matching epoch y_m must satisfy $y_H \ll y_m \ll 1$. Note that I have replaced the argument of the log in Eq. (7.51) — $k\eta$ — with y/y_H, valid as long as the matching epoch is deep in the radiation era. Figure 7.10 shows the evolution of two modes along with the analytic solutions to the Meszaros equation with coefficients set by the matching conditions laid out in Eq. (7.64). Not suprisingly, for larger scale modes than the ones shown the approximation breaks down.

12.7.4 Numerical Results and Fits

In Section 7.2 and Section 7.3, we derived analytic solutions following the dark matter perturbations deep into the matter era. Here, we assimilate these results and spline them together to form the transfer function. Also, I will present a wellknown fitting function for the transfer function.

First, we need to transform our expression ((7.63) along with Eqs. (7.64)) for the small-scale matter density into an expression for the transfer function. The transfer function is determined by the behavior of δ well after equality when the decaying mode has long since vanished. We can extract an even simpler form for δ in this $a \gg a_{\rm eq}$ limit. The key constant in that case is

C_1, the coefficient of the growing mode. Multiplying the first matching condition in Eq. (7.64) by D_2' and the second by D_2 and then subtracting leads to

$$C_1 = \frac{D_2'(y_m)\, A \ln(By_m/y_H) - D_2(y_m)\,(A/y_m)}{D_1(y_m)\, D_2'(y_m) - D_1'(y_m)\, D_2(y_m)} \Phi_{\rm p}$$

The denominator $D_1 D_2' - D_1' D_2 = -(4/9) y_m^{-1} (y_m + 1)^{-1/2}$, which is approximately equal to $-4/9y_m$ since $y_m \ll 1$. Similarly for small y_m, $D_2 \to (2/3)\ln(4/y) - 2$ and $D_2' \to -2/3y$. Therefore,

$$C_1 \to \frac{-9A\Phi_{\rm p}}{4} \left[\frac{-2}{3} \ln(By_m/y_H) - (2/3)\ln(4/y_m) + 2 \right]$$

which fortuitously does not depend on y_m. Therefore, at late times we have an approximate solution for the small-scale dark matter perturbations

$$\delta(\vec{k}, a) = \frac{3A\Phi_{\rm p}(\vec{k})}{2} \ln\left[\frac{4Be^{-3} a_{\rm eq}}{a_H} \right] D_1(a) \quad a \gg a_{\rm eq}$$

On very small scales, the argument of the log simplifies because $a_{\rm eq}/a_H = \sqrt{2}k/k_{\rm eq}$ (Exercise 6). To turn Eq. (7.67) into a transfer function, we need to remember how δ is related to $\Phi_{\rm p}$. Comparing Eq. (7.8) with Eq. (7.67) leads to an analytic expression for the transfer function on small scales:

$$T(k) = \frac{5A\Omega_m H_0^2}{2k^2 a_{\rm eq}} \ln\left[\frac{4Be^{-3}\sqrt{2}k}{k_{\rm eq}} \right] \quad k \gg k_{\rm eq}$$

Recall that the wavenumber entering the horizon at equality is defined as $k_{\rm eq} := a_{\rm eq} H(a_{\rm eq}) = \sqrt{2} H_0 a_{\rm eq}^{-1/2}$, so the prefactor is also a function of $k/k_{\rm eq}$ only. Then, plugging in numbers leads to

$$T(k) = \frac{12 k_{\rm eq}^2}{k^2} \ln\left[\frac{k}{8k_{\rm eq}} \right] \quad k \gg k_{\rm eq}.$$

Figure 7.11 shows the power spectrum for a standard CDM model ($n = 1$; $h = 0.5$; but no baryons) matching the large-scale transfer function ($T = 1$) with the small-scale transfer function of Eq. (7.69). Also shown is the exact solution

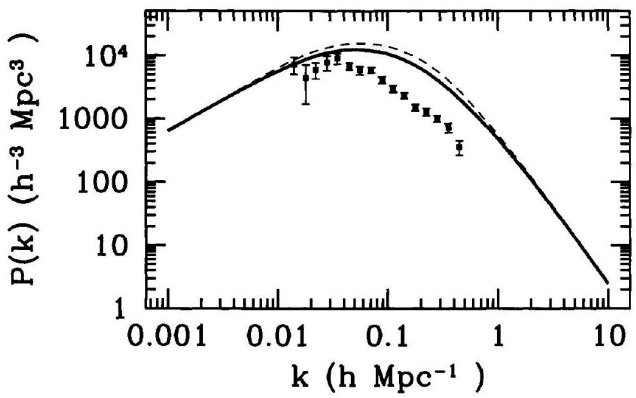


Figure 7.11. The power spectrum in a standard CDM model with a Harrison-Zel'dovich-Peebles spectrum. The thick solid curve uses the BBKS transfer function; the dashed curve interpolates between the analytic transfer function on large scales (equal to 1) and small scales

(Eq. (7.68)). The data points are a compilation (and interpretation) by Peacock and Dodds (1994).

(again in the no-baryon limit), or equivalently, the fitting form of Bardeen, Bond, Kaiser, and Szalay (1986, BBKS),

$$T\left(x := k/k_{\rm eq}\right) = \frac{\ln[1+0.171x]}{(0.171x)}\left[1+0.284x+(1.18x)^2+(0.399x)^3 +(0.490x)^4\right]^{-0.25}.$$

Note that the BBKS form agrees very well with the analytic solution on small scales; i.e., both aymptote to $\ln(k)/k^2$ with the same coefficients. Since wavenumbers are measured in units of $h{\rm Mpc}^{-1}$, the ratio $k/k_{\rm eq}$ depends on $\Omega_m h$. So defining $\Gamma := \Omega_m h$, the BBKS transfer function can also be written as $T\left(q := k/\Gamma h{\rm Mpc}^{-1}\right) = \frac{\ln[1+2.34q]}{(2.34q)}\left[1+3.89q+(16.2q)^2+(5.47q)^3+(6.71q)^4\right]^{-0.25}$.

Several final comments are in order. First, our analytic work has enabled us to understand the origin of the asymptotic, small-scale behavior of the power spectrum. Had there been no logarithmic growth in the radiation era, the modes which entered very early on would have experienced no growth from horizon entry until the epoch of equality. Their amplitude relative to large-scale modes would then have been suppressed by a factor of order $\left(k_{\rm eq}/k\right)^2$. The logarithmic growing mode in the radiation era somewhat ameliorates this suppression. Second, although our analytic expression and its BBKS counterpart are good approximations, it is important to be aware of some small effects which affect the transfer function in the real world. We have assumed no anisotropic stresses ($\Phi = -\Psi$). Dropping this assumption changes the factor of 9/10 by which the potential drops for large-scale modes to 0.86, resulting in a corresponding rise in the small-scale transfer function. Including a realistic amount of baryons leads to even more severe small-scale changes. We will address these in Section 7.6. Third, all of our work in this section has been on the transfer function, i.e., on the evolution of perturbations early on when the only components of the universe were matter and radiation. At very late times, the growth function depends on other hypothetical components, the most likely of which is dark energy. Finally, the theoretical power spectrum in Figure 7.11 has been normalized by fixing δ_H in Eq. (7.9) using the observations of CMB anisotropies on large scales (more on this in Chapter 8). We see that (i) the large-scale normalization is roughly correct and that (ii) the shape of the standard CDM power spectrum is wrong. The sCDM power spectrum turns over on relatively small scales, in distinct disagreement with the data. The universe as we observe it appears to have a smaller $k_{\rm eq}$ than sCDM. This observation motivates consideration of variations of sCDM; we will consider these in Section 7.6.

12.7.5 Growth Function

At late times ($z \lesssim 10$) all modes of interest have entered the horizon. You might think then, that the $y \gg 1$ limit of the Meszaros equation, which describes subhorizon modes in the matter era, would apply. This is true if $\Omega_m = 1$. If the energy budget of the universe has another item at late times - either dark energy or curvature - then we must retrace the steps which led to the Meszaros equation. Before doing this, I want to point out that, no matter what constitutes the energy budget today, all modes will experience the same growth factor. We saw this in the previous section, where the Meszaros equation was independent of k. And we will soon see it again, when we generalize the Meszaros equation to account for other forms of energy. This uniform growth is a direct result of the fact that cold, dark matter has zero pressure. Therefore, once a mode enters the horizon, there is no way for pressure to smooth out the inhomogeneities and all modes evolve identically. We want to derive an evolution equation analagous to the Meszaros equation, but allowing for the possibility of energy other than matter or radiation. We can take the $y \gg 1$ limit of Eqs. (7.54)-(7.56), but we must rethink the coefficient of the source

term in the Poisson equation. Since radiation can be ignored, the coefficient multiplying δ in Eq. (7.56) is now $4\pi G\rho_{\rm dm} = (3/2)H_0^2\Omega_m a^{-3}$. Also when differentiating Eq. (7.54) previously, we set $(1/aHy)' = -(1+y)^{-1}(2aHy)^{-1}$; here we need to account for other contributions to H' so Eq. (7.57) becomes

$$\delta'' + ikv\left(\frac{d(aHy)^{-1}}{dy} - \frac{1}{aHy^2}\right) = \frac{3\Omega_m H_0^2}{2y^3a^2H^2a_{\rm eq}}\delta$$

Replacing the velocity term using the continuity equation as before leads to

$$\frac{d^2\delta}{da^2} + \left(\frac{d\ln(H)}{da} + \frac{3}{a}\right)\frac{d\delta}{da} - \frac{3\Omega_m H_0^2}{2a^5H^2}\delta = 0$$

Here I have divided by $a_{\rm eq}^2$ and we will now use a as the variable instead of y. In this large y limit, all factors of $a_{\rm eq}$ disappear.

There are two solutions to Eq. (7.73). One solution is $\delta \propto H$. It is easy to check this if all the energy is nonrelativistic matter, so that the solution is proportional to $a^{-3/2}$. Then all three terms scale as $a^{-7/2}$; the coefficient of the first is $15/4$, the second $-9/4$, and the last $3/2$. The sum of these does indeed vanish. In Exercise 7, you will be asked to show that $\delta \propto H$ is a solution if there are other components of energy in the universe. This solution is pretty, but it is not the one we want since almost all current models of the universe have a nonincreasing Hubble rate. The modes we are interested in - those that remain long after horizon crossing-are the growing modes. So we are interested in the other solution of Eq. (7.73).

To obtain the growing mode, we try a solution of the form $u = \delta/H$. The evolution equation for u then becomes

$$\frac{d^2u}{da^2} + 3\left[\frac{d\ln(H)}{da} + \frac{1}{a}\right]\frac{du}{da} = 0$$

This first-order equation for u' can be integrated to obtain

$$\frac{du}{da} \propto (aH)^{-3}.$$

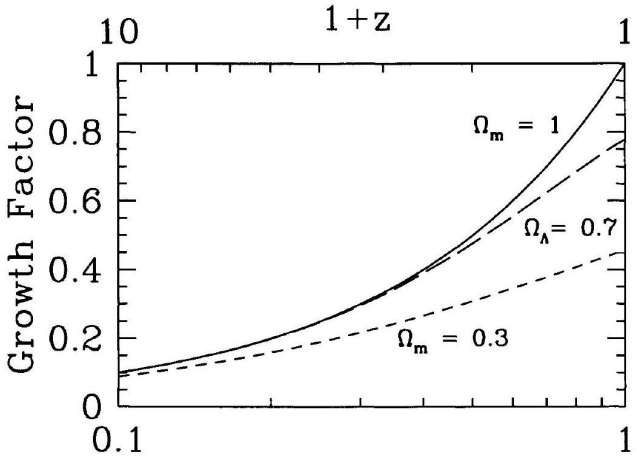
Integrating again and remembering that the second solution, the growth factor, is uH leads to an expression for the growth factor

$$D_1(a) \propto H(a)\int^a \frac{da'}{\left(a'H\left(a'\right)\right)^3}.$$

I have glossed over the proportionality constant. This is fixed by the definition of Eq. (7.4), which says that, early on when matter still dominates (say at $z \simeq 10$), D_1 should be equal to a. At those times, $H = H_0\Omega_m^{1/2}a^{-3/2}$ so the growth factor is

$$D_1(a) = \frac{5\Omega_m}{2}\frac{H(a)}{H_0}\int_0^a \frac{da'}{\left(a'H\left(a'\right)/H_0\right)^3}.$$

The growth factor in an open universe without dark energy can be computed analytically (see Exercise 8).



a

Figure 7.12. The growth factor in three cosmologies. Top two curves are for flat universes without and with a cosmological constant. Bottom curve is for an open universe.

Figure 7.12 shows the growth factor for three different cosmologies. As mentioned above, if the universe is flat and matter dominated, the growth factor is simply equal to the scale factor. In both open and dark energy cosmologies, though, growth is suppressed at late times. This leads to an important qualitative conclusion: structure in an open or dark energy universe developed much earlier than in a flat, matter-dominated universe. There has been relatively little evolution at recent times if the universe is open or dark energy-dominated. Therefore, whatever structure is observed today was likely in place at much earlier times. We will see some quantitative implications of this in Section 9.5.

12.7.6 Beyond Cold Dark Matter

There is more to the universe than just cold dark matter. Although CDM is the main component in most cosmological models, so that the transfer function we derived earlier is a good approximation to reality, there are trace amounts of other stuff. To be completely accurate we need to account for this other stuff. Here I

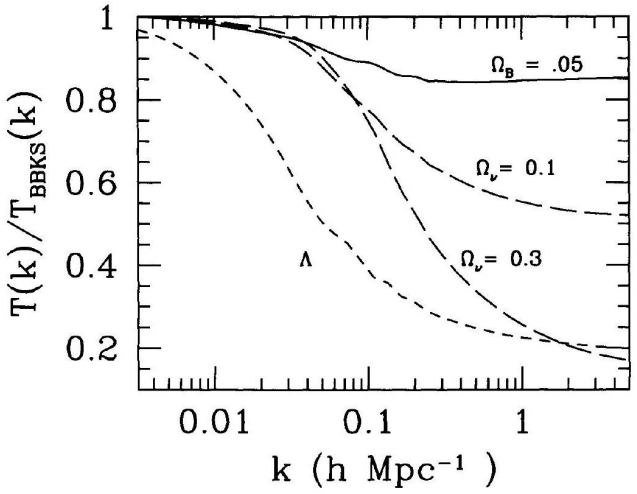


Figure 7.13. The ratio of the transfer function to the BBKS transfer function (Eq. (7.70)) which describes dark-matter-only (no baryons) perturbations. Top curve (and all other curves as well) has 5% baryons. Two middle curves show different values for a massive neutrino. Bottom curve has a cosmological constant $\Omega_\Lambda = 0.7$.

focus on three additional components. First, we consider the effect of the baryons, which constitute roughly 10% of the total matter in most models, on the transfer function. Then, we entertain the possibility that neutrinos have mass and examine the resultant effect on the transfer function. Finally, dark energy - one model for which is the cosmological constant - is considered.

Figure 7.13 shows the transfer functions accounting for these components. A realistic baryon fraction suppresses the transfer function on small scales. A massive neutrino does the same, with the nature and amplitude of the suppression depending on the neutrino mass. Dark energy, here in the form of a cosmological constant, moves the epoch of equality to later times, thereby reducing $k_{\rm eq}$. The break in the transfer function therefore comes on much larger scales than in the standard CDM model, in apparent agreement with the data exhibited in Figure 7.11.

Baryons

Baryons account for about 4% of the total energy density in the universe. As such, their effect on the matter power spectrum is small. A careful examination of Figure 7.13 reveals two signatures of a nonzero baryon density. The first is that the power spectrum is suppressed on small scales. This is not surprising: at early times, before decoupling, baryons are tightly coupled to photons. Therefore, just as radiation perturbations decay when entering the horizon, so too do baryon overdensities. After decoupling, baryons are released from the relatively smooth radiation field and fall into the gravitational potentials set up by the dark matter. The depth of these wells is smaller than we estimated in Section 7.3, though, because only a fraction $\Omega_{\rm cdm}/\Omega_m$ of the total matter was involved in the collapse.

The second effect of baryons is less noticable in Figure 7.13 and indeed may never get measured in real life either. Nonetheless, it is extremely important if only because it hints at a fundamental feature of the radiation field. In all the curves in Figure 7.13, except the $\Omega_\nu = 0.3$ case, you can see small oscillations in the transfer function centered around $k \simeq 0.1h{\rm Mpc}^{-1}$. These are not numerical artifacts. Rather, they are manifestations of the oscillations that the combined baryon/photon fluid experience before decoupling. We got a glimpse of these in Section 7.3.2 (e.g., Figure 7.8) when we considered the potential in the radiation-dominated era. Just as the potential oscillates in this era, the baryon/photon fluid also oscillates. It is the traces of these oscillations that are imprinted on the matter transfer function. They are barely (if at all) detectable because baryons are such a small fraction of the total matter. In the baryon-only model plotted in Figure 1.13, the oscillations were much more noticeable. [5] And these oscillations are also prominent in the spectrum of the radiation perturbations, as we will see in the next chapter.

Massive Neutrinos

Neutrinos are known to exist, and the standard Big Bang model gives a definite prediction for how many there are in the universe (Eq. (2.77)). Massive neutrinos may play an important role in structure formation. Conversely, an accurate measurement of the power spectrum may enable us to infer neutrino masses. For orientation, recall the difference between massless (Eq. (2.78)) and massive (Eq. (2.80)) neutrino energy densities. The best bet from experiments is that the most massive neutrino has a mass of order 0.05eV, therefore contributing $\Omega_\nu \simeq 10^{-3}$. Even this trace amount might eventually be detectable if the power spectrum can be measured accurately enough. There is also the possibility that one or more neutrinos has a larger mass (see the footnote on Page 46). Current upper limits from structure formation hover around 2eV (Elgaroy et al., 2002).

The reason why even a small admixture of massive neutrinos affects the power spectrum is that, especially if they are light, neutrinos can move fast (they are not cold dark matter)

and stream out of high-density regions. Perturbations on scales smaller than the free-streaming scale are therefore suppressed. Indeed, a long time ago, cosmologists considered the possibility that all the dark matter in the universe was in the form of neutrinos. If this were so, then there would be no power on small scales and structure would have to form from the "top down."

We can estimate the scale on which perturbations are damped by computing the comoving distance a massive neutrino can travel in one Hubble time at equality.

[5] Incidentally, we now also understand why the power in Figure 1.13 is so low in the baryon-only universe: there is no dark matter which can cluster before recombination. This calculation is trivial, however, if the neutrino mass is in the eV range. For then, the average velocity, T_ν/m_ν, is of order unity at equality. So neutrinos can freestream out of horizon-scale perturbations at equality. This leads to a suppression in power on all scales smaller than $k_{\rm eq}$.

Figure 7.13 shows this suppression. Note, though, that the effect is a little subtle. A lighter neutrino can free-stream out of larger scales, so the suppression begins at lower k for the $\Omega_\nu = 0.1$ mass than for the $\Omega_\nu = 0.3$ case. On the other hand, the more massive neutrino constitutes more of the total density so it suppresses small-scale power more than does the lighter neutrino.

Dark Energy

Cosmologists have recently accumulated tantalizing evidence for dark energy in the universe above and beyond the dark matter that we have spent so much time on in this chapter. If dark energy exists, how does it affect the matter perturbations?

The first effect of dark energy is indirect. Since theoretical prejudice and evidence both indicate that the universe is flat, $\Omega_{\rm de} \simeq 0.6 - 0.7$ implies that the matter density, Ω_m, is less than 1. This has a huge impact on the power spectrum, because we have seen that the power spectrum turns over at $k_{\rm eq}$, which is proportional to Ω_m. So dark energy leads to a turnover in the power spectrum on a scale much larger than predicted in standard CDM. In fact, as we saw in Figure 7.11, this is one of the pieces of evidence for dark energy. The turnover in the power spectrum does not appear on the scale predicted by standard CDM.

The second effect is again related to the smaller matter density in most models of dark energy. As a result of the Poisson equation (7.7), overdensities are inversely proportional to Ω_m for a fixed potential. Therefore, the amplitude of the power spectrum increases as the matter decreases, or equivalently in a flat universe as the dark energy content goes up. With a few caveats to be discussed in Chapter 8, large-angle CMB anisotropies fix the potential on large scales. When normalizing to these large-angle results, therefore, the power spectrum for a model with dark energy is normalized higher than one without.

The third effect of the dark energy on the density inhomogeneities is more direct and more model dependent. At late times, amplification of perturbations is controlled by the growth factor of Eq. (7.77). The evolution of the Hubble rate depends on the model of dark energy, so different models of dark energy predict different growth factors. If we parameterize the dark energy by its equation of state (2.84), then the Hubble rate in a flat universe evolves as

$$\frac{H(z)}{H_0} = \left[\frac{\Omega_m}{a^3} + \frac{\Omega_{\rm de}}{a^{3[1+w]}}\right]^{1/2}$$

at late times. Using this time dependence, it is straightforward to perform the integral in Eq. (7.77) and find the growth factor for a given equation of state (see Exercise 11). To sum up, dark energy affects the power spectrum by changing $k_{\rm eq}$ and the normalization (this depends only on $\Omega_{\rm de}$) and by changing the growth factor at late times (depends on both $\Omega_{\rm de}$ and w). Careful observations of the matter spectrum therefore may enable us to learn about dark energy.

SUGGESTED READING

Once again The Large Scale Structure of the Universe (Peebles) is a useful reference. Since it was written before the implications of cold dark matter and inflation were explored, though, it does not contain a transfer function or power spectrum such as the ones we have derived (although Peebles himself was instrumental in computing these things several years after the book was published). A more up-to-date book, which is particularly strong on large-scale structure is Structure Formation in the Universe (Padmanabhan).

The first papers to work out the CDM transfer function are particularly instructive to read, not least because they also focus on some of the physical implications of the hierarchical theories. See Blumenthal et al. (1984) and Peebles (1982). The most important recent paper is Seljak and Zaldarriaga (1996), not so much because it contains a concise description of the set of coupled equations to be solved (although it does that), but because it makes available CMBFAST, a code which computes transfer functions and CMB anisotropy spectra. It is currently available at http://physics.nyu.edu/matiasz/CMBFAST/cmbfast.html. The treatment in this chapter follows most closely the small scale analytic solution of Hu and Sugiyama (1996), a paper which is extremely rich and well worth reading. A more recent paper by Eisenstein and Hu (1998) employs the analytic small-scale solution to derive accurate fitting formulae that move beyond those presented by Bardeen, Bond, Kaiser, and Szalay (1986, BBKS).

12.8 Anisotropies

The primordial perturbations set up during inflation manifest themselves in the radiation as well as in the matter distribution. By understanding the evolution of the photon perturbations, we can make predictions about the expected anisotropy spectrum today. This evolution is again completely determined by the EinsteinBoltzmann equation we derived in Chapters 4 and 5, and one way to go would be to simply stick all the relevant equations in those chapters on a computer and solve them numerically. Historically, this is a pretty good caricature of what happened. Long before we developed deep insight into the physics of anisotropies, various groups had codes which determined the expected spectra from different models. Only much later did we come to understand both qualitatively and quantitatively why the spectra look like they do. [1] In this chapter, I will mangle the history and simply explain what we have learned about the physics of anisotropies.

Perturbations to the photons evolved completely differently before and after the epoch of recombination at $z \simeq 1100$. Before recombination, the photons were tightly coupled to the electrons and protons; all together they can be described as a single fluid (dubbed the "baryon-photon" fluid). After recombination, photons free-streamed from the "surface of last scattering" to us today. After an overview which qualitatively explains the anisotropy spectrum, Sections 8.2-8.4 work through the physics of the baryon-photon fluid before recombination. Then Sections 8.58.6 treat the post-recombination era, culminating in the predicted spectrum of anisotropies today. Finally Section 8.7 discusses how these spectra vary when the cosmological parameters change.

[1] Understanding the anisotropies actually helped make the codes much more efficient. The prime example of this is the popular code CMBFAST (Seljak and Zaldarriaga, 1996) which is based in part on the analytic solution presented in this chapter.

12.8.1 Overview

Let's begin as we did in the last chapter, by cheating and looking at the answers first. Figure 8.1 shows the evolution of the perturbations to the photons. Four Fourier modes corresponding to perturbations on four different scales are shown. Qualitatively, the most important feature of

Figure 8.1 is that perturbations to the photons do not grow appreciably with time. This stands in stark contrast to the matter perturbations, which do grow. And this contrast is something we should have expected: the pressure of the photons is so large that it can withstand the tendency toward collapse. This means that the small perturbations set up during inflation stay small; they remain linear all the way up to the present.

Before going further and examining the evolution of the different modes in more detail, a technical note: I have plotted not simply the perturbation to the photons but rather the combination $k^{3/2}(\Theta_0 + \Psi)$. The $k^{3/2}$ factor balances the fact that the amplitude of the perturbations (in a simple inflationary model) scales as $k^{-3/2}$. I have added the gravitational potential Ψ because the photons we see today had to travel out of the potentials they were in at the time of recombination. As they emerged from these potential wells, their wavelengths were stretched (if the region was overdense and $\Psi < 0$), thereby decreasing their energy. Thus, the temperature we see today is actually Θ_0 at recombination plus Ψ.

The large-scale mode in Figure 8.1 evolves hardly at all. This is not surprising: no causal physics can affect perturbations with wavelengths larger than the horizon, so a super-horizon mode should exhibit little evolution. This means that when we observe large-scale anisotropies - which are sensitive to modes with wavelengths larger than the horizon at recombination - we are observing perturbations in their most pristine form, as they were set down at very early times, presumably during inflation.

Figure 8.1 shows that the smaller scale modes evolve in a more complicated way than the super-horizon modes. Consider the curve labeled "First Peak." As the mode enters the horizon, the perturbation begins to grow until it reaches an apparent maximum at the time of recombination. If we observe anisotropies on scales corresponding to this mode, we would expect to see large fluctuations. Hence the label: the anisotropy spectrum will have a peak at the angular scales corresponding to the mode which has just reached its peak at recombination.

The mode in Figure 8.1 which enters the horizon slightly earlier peaks earlier and then turns over so that its amplitude at recombination is zero. By recombination, it has undergone half of an oscillation; so we see our first clear signal of the acoustic oscillations due to the pressure of the relativistic photons. The phase of this mode is such that, at recombination its amplitude is zero. Therefore, when we observe anistropies today corresponding to these scales, we expect very small fluctuations. There will be a trough in the anisotropy spectrum on these angular scales.

And on it goes. The curve labeled "Second Peak" entered the horizon even earlier and has gone through one full oscillation by recombination. As such, this mode will have large fluctuations and lead to a second peak in the anisotropy spectrum. You

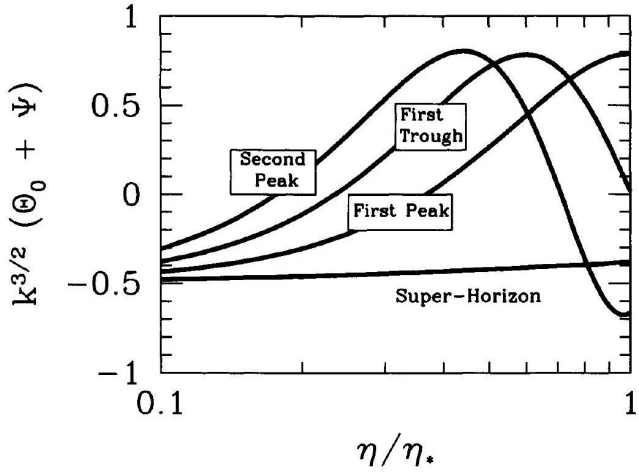


Figure 8.1. Evolution of photon perturbations of four different modes before recombination at η_*. Normalization is arbitrary, but the relative normalization of the 4 curves is appropriate for

perturbations with a Harrison-Zel'dovich-Peebles ($n = 1$) spectrum. Model is standard CDM with $h = 0.5, \Omega_m = 1$, and $\Omega_b = 0.06$. Starting from the bottom left and moving upward, the wavenumbers for the modes are $k = (7 \times 10^{-4}, 0.022, 0.034, 0.045)\,\mathrm{Mpc}^{-1}$ or $(8, 260, 400, 540)/\eta_0$

might expect that there will be a never-ending series of peaks and troughs in the anisotropy spectrum corresponding to modes that entered the horizon earlier and earlier. And you would be right: this is exactly what happens.

We can see this more clearly by looking at the spectrum of perturbations at one time, the time of recombination. Figure 8.2 shows this spectrum for two different models, one with a very low baryon content. We do indeed see this pattern of peaks and troughs. There are two more quantitative features of these oscillations that are important. First, note that - at least in the higher baryon model - the heights of the peaks seem to alternate: the odd peaks seem higher than the even peaks. Second, and this is clearest in the low baryon model, perturbations on small scales $k\eta_0 \gtrsim 500$ are damped.

To understand the first of these features, we can write down a cartoon version of the equation governing perturbations. Very roughly, this equation is

$$\ddot{\Theta}_0 + k^2 c_s^2 \Theta_0 = F$$

where F is a driving force due to gravity and c_s is the sound speed of the combined baryon-photon fluid. This is the equation of a forced harmonic oscillator (see box on page 220). Qualitatively, it predicts the oscillations we have seen above. But it also

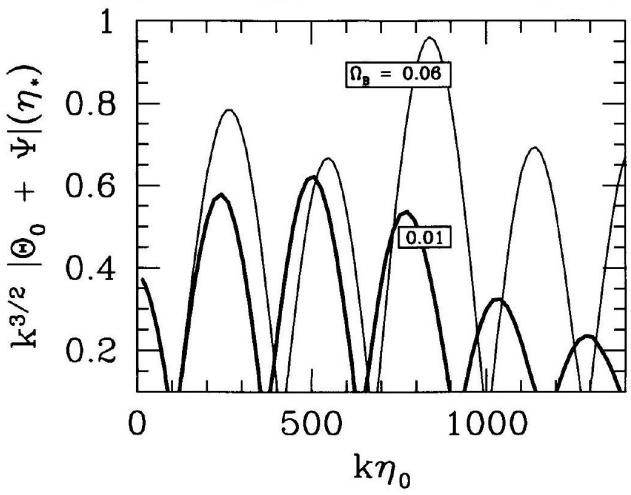


Figure 8.2. Perturbations to the photon distribution at recombination in two models. The larger damping length of the low- Ω_b model is clearly evident in the suppression of perturbations for modes with $k > 500/\eta_0$.

explains something about the heights of the peaks. As we add more baryons to the universe, the sound speed goes down (baryons are heavy so they reduce the speed). Thus the frequency of the oscillations goes down. The peaks at $n\pi/\omega$ are shifted to larger k (you really should read that box!), and the spacing between peaks gets correspondingly larger. Further, as the frequency goes down, the disparity between the heights of the odd and even peaks gets larger. We clearly see both of these features in Figure 8.2. Another way of understanding the alternating peak heights is to note that the perturbations for the first peak mode have been growing since they entered the horizon. By decreasing the pressure (or equivalently increasing the importance of gravity) these modes will grow even more. The second peak mode on the other hand, corresponds to an underdensity of photons in the potential wells. Decreasing the pressure makes it harder for photons to escape the well and therefore reduces the magnitude of the perturbation (makes it less underdense). Consider a simple harmonic oscillator with mass m and force constant k. In

addition to the restoring force, the oscillator is acted on by an external force F_0. Thus the full force is $F_0 - kx$ where x is the oscillator's position. The equation of motion is

$$\ddot{x} + \frac{k}{m}x = \frac{F_0}{m}$$

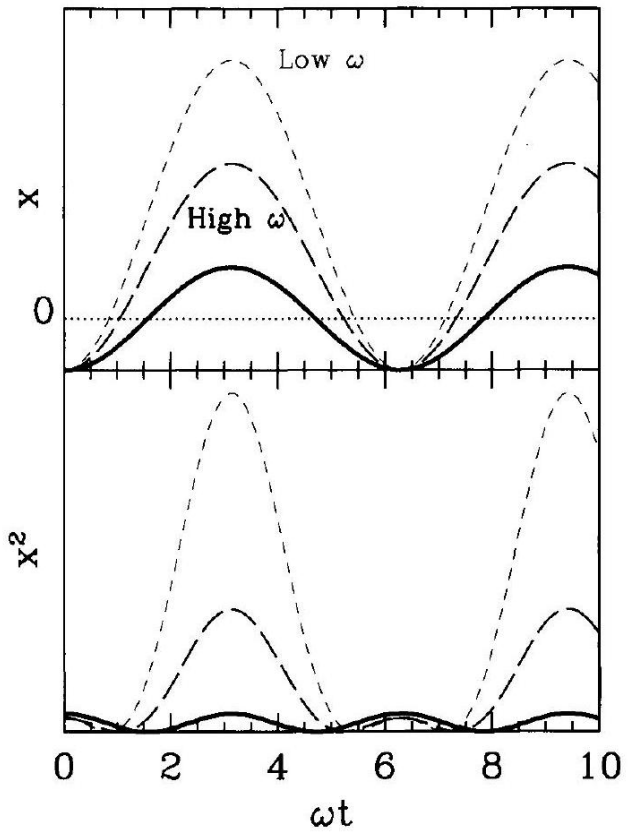
The term on the right-hand side-representing the external force-is driving the oscillator to large values of x. The restoring force on the other hand tries to keep the oscillator as close to the origin as possible. The solution therefore will be that oscillations will be set up around a new zero point, at positive x.

The solution to Eq. (8.2) is the sum of the general solution to the homogeneous equation (with the right-hand side set to zero) and a particular solution. The general solution has two modes, best expressed as a sine and cosine with arguments ωt, with the frequency ω defined as $\omega := \sqrt{\frac{k}{m}}$. A particular solution to Eq. (8.2) is constant $x = F_0/m\omega^2$, so the full solution is the sum of the sine and cosine modes plus this constant. Let us assume that the oscillator is initially at rest. Then, since $\dot{x}(0)$ is proportional to the coefficient of the sine mode, this coefficient must vanish, leaving

$$x = A\cos(\omega t) + \frac{F_0}{m\omega^2}.$$

This solution is shown in the figure at right. The solid line is the unforced solution: oscil-

Forced Harmonic Oscillator



lations about the origin. The dashed curves are the forced solutions for two different choices of frequencies. In both cases, the oscillations are not around $x = 0$ as they would be if the system was unforced. Once an external force is introduced, the zero point of the oscillations shifts in the direction of the force. Two curves are drawn to show that this shift is more dramatic for lower frequencies. The bottom panel shows the square of the oscillator position as a function of time. All three oscillators experience a series of peaks at $t = n\pi/\omega$ corresponding to the minima/maxima of the cosine mode. (Note that if only the sine mode was present these peaks would be at $t = (2n + 1)\pi/\omega$.) The heights of these peaks are identical in the case of the

unforced oscillator and equal to the height at $t = 0$. In the forced case, though, the height of the odd peaks - those at $t = (2n + 1)\pi/\omega$ - is greater than that of the even peaks. The effect is most dramatic for low frequencies. If the frequency is low, the force has a greater effect, producing the greater zero-point offset, and hence the greater odd/even disparity. The other feature of this example is that the even peaks correspond to negative positions of the oscillator: points at which it is farthest from where the force wants it to go.

Photon Diffusion

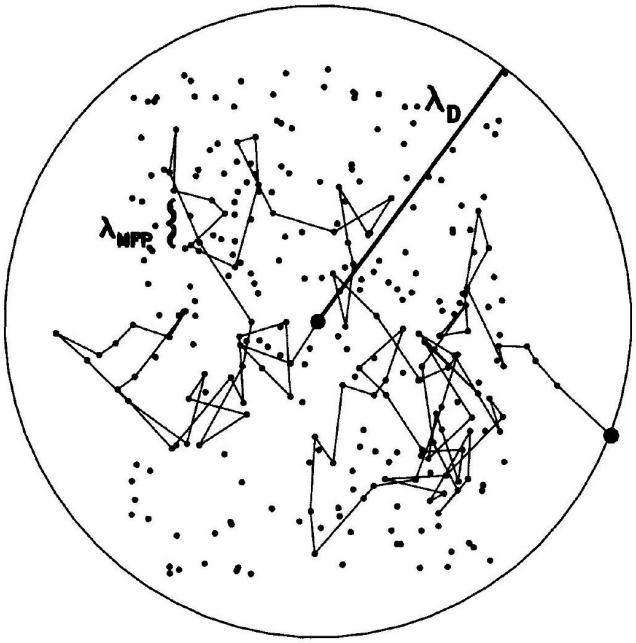


Figure 8.3. Photon diffusion through the electron gas. Electrons are denoted as points. Shown is a typical photon path as it scatters off electrons. The mean free path is λ_{MFP}. After a Hubble time, the photon has scattered many times, so that it has moved a distance of order λ_D.

To understand the damping evident in Figure 8.2, we need to remember that the approximation of the photons and electrons and baryons moving together as a single fluid is just that, an approximation. It is valid only if the scattering rate of photons off of electrons is infinite. Of course this condition is not met: photons travel a finite distance in between scatters. Consider Figure 8.3, which depicts the path of a single photon as it scatters off a sea of electrons. It travels a mean distance $\lambda_{\rm MFP}$ in between each scatter. In our case this distance is $(n_e\sigma_T)^{-1}$. If the density of electrons is very large, then the mean free path is correspondingly small. In the course of a Hubble time, H^{-1}, a photon scatters of order $n_e\sigma_T H^{-1}$ times (simply the product of the rate and the time). As depicted in Figure 8.3, each scatter contributes to the random walk of the photon. We know that the total distance traveled in the course of a random walk is the mean free path times the square root of the total number of steps. Therefore, a cosmological photon moves a mean distance

$$\begin{aligned}\lambda_D &\sim \lambda_{\rm MFP}\sqrt{n_e\sigma_T H^{-1}} \\ &= \frac{1}{\sqrt{n_e\sigma_T H}}\end{aligned}$$

in a Hubble time. Any perturbation on scales smaller than λ_D can be expected to be washed out. In Fourier space this will correspond to damping of all high k modes. Note that this crude estimate gets the Ω_b dependence right. Models with small baryon density have a larger λ_D (since n_e is proportional to Ω_b when the universe is ionized). Therefore, the damping sets in at larger scales, or smaller k. This is precisely what we saw in Figure 8.2.

The final step is to relate the perturbations at recombination, as depicted in Figure 8.2, to the anisotropies we observe today. The math of this is a little complicated, but the physics is perfectly straightforward. Consider one Fourier mode, a plane-wave perturbation. Figure 8.4 shows the temperature variations for one mode at recombination. Photons from hot and cold spots separated by a typical (comoving) distance k^{-1} travel to us coming from an angular separation $\theta \simeq k^{-1}/(\eta_0 - \eta_*)$ where $\eta_0 - \eta_*$ is the (comoving) distance between us and the surface of last scattering. [2] If we decompose the temperature field into multipole moments, then an angular scale θ roughly corresponds to $1/l$. So, using the fact that $\eta_* \ll \eta_0$, we project inhomogeneities on scales k onto anisotropies on angular scales $l \simeq k\eta_0$.

There is one final caveat to this picture of free-streaming. We have been implicitly assuming that nothing happens to the photons on their journey from the last scattering surface to Earth. In fact, if the universe was flat and matter dominated through this whole time, then gravitational potentials remain constant, and this assumption is correct. However, recombination takes places not too much later than the epoch of equality, so the remnant radiation density means potentials are not exactly constant right after recombination. Also, at late times, dark energy does not behave like matter and leads to potential decay. You can imagine other disruptions to matter domination. All of these so-called integrated Sachs-Wolfe effects produce new perturbations to the photons, leading to changes typically of order 10%.

And that's it; we now understand how primordial perturbations are processed to form the present-day anisotropy spectrum. Let's work through it again quantitatively.

12.8.2 Large-Scale Anisotropies

Theory

To find the large-scale solution for the photon perturbation, we make use of the super-horizon equation, (7.17). This immediately tells us that $\Theta_0 = -\Phi$ plus a

[2] This is true only in a flat universe. In an open universe, the distance to the last scattering surface is larger, so the same physical scale is projected onto a smaller angular scale.

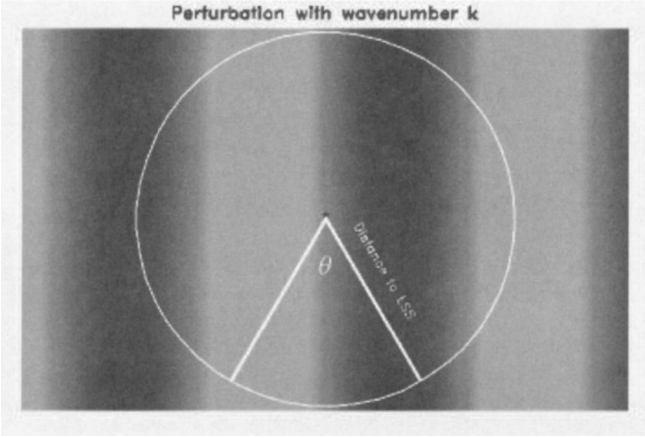


Figure 8.4. Free-streaming. Perturbations in the temperature at recombination from one plave wave with wavenumber k. Hot and cold spots are shaded light and dark. After recombination, photons from the hot and cold spots travel freely to us, here denoted by the star at the center. This mode contributes anisotropy on a scale $\theta \sim k^{-1}/$ (Distance to last scattering surface).

constant. The initial conditions are such that $\Theta_0 = \Phi/2$, so the constant is $3\Phi_{\mathbf{p}}/2$, where $\Phi_{\rm p}$ is the primordial potential set up during inflation. We have an exact expression for the large-scale evolution of Φ, Eq. (7.32). If recombination takes place long after the epoch of

equality, then we can take the $y \gg 1$ limit of this expression, $\Phi \rightarrow 9\Phi_{\rm p}/10$. Therefore, at recombination, large-scale photon perturbations satisfy

$$\begin{aligned}\Theta_0\left(k,\eta_*\right) &= -\Phi\left(k,\eta_*\right) + \frac{3\Phi_{\rm p}(k)}{2} \\ &= \frac{2\Phi\left(k,\eta_*\right)}{3}.\end{aligned}$$

The observed anisotropy is $\Theta_0 + \Psi$, which to a good approximation is $\Theta_0 - \Phi$ since $\Psi \simeq -\Phi$. Therefore,

$$\left(\Theta_0 + \Psi\right)\left(k,\eta_*\right) = \frac{1}{3}\Psi\left(k,\eta_*\right)$$

Another useful way of expressing the large-scale perturbations at recombination is in terms of the density field. The initial conditions derived in Chapter 6 were that $\delta = 3\Phi/2$. Integrating the large-scale evolution equation, $\dot{\delta} = -3\dot{\Phi}$, leads to

$$\begin{aligned}\delta\left(\eta_*\right) &= \frac{3}{2}\Phi_{\rm p} - 3\left[\Phi\left(\eta_*\right) - \Phi_{\rm p}\right] \\ &= 2\Phi\left(\eta_*\right)\end{aligned}$$

So the observed anisotropy expressed in terms of the dark matter overdensity is

$$\left(\Theta_0 + \Psi\right)\left(k,\eta_*\right) = -\frac{1}{6}\delta\left(\eta_*\right)$$

Equations (8.6) and (8.8) will be useful to us when we compute the largescale anisotropy spectrum. However, even now, they contain a fascinating piece of information. From the Fourier transform of Eq. (8.8), we see that the observed anisotropy of an overdense region will be negative. This is such a surprising result that it is worth repeating. For large-scale perturbations, overdense regions do indeed contain hotter photons at recombination than do underdense regions: i.e., $\Theta_0 > 0$ when $\Psi < 0$. However, to get to us today, these photons must travel out of their potential wells. In so doing they lose energy, and this energy loss more than compensates for the fact that the photons were initially hotter than average: i.e., $\Theta_0 + \Psi$ is negative when $\Psi < 0$. To sum up, when we observe large-scale hot spots on the sky today, we are actually observing regions that were underdense at the time of recombination.

The other important feature of Eq. (8.8) is the coefficient 1/6. It enables us to relate " $\delta T/T$" (the left-hand side) to " $\delta\rho/\rho$" (the right). Very roughly speaking, an anisotropy of order 10^{-5} corresponds to an overdensity of 6×10^{-5}. One of the important questions which must be addressed by the picture of gravitational instability is whether the observed anisotropy is consistent with the overdensities needed to form structure by today. This factor of 6 is a huge help. In almost all models of structure formation other than inflation, this factor of 6 is replaced by a number much closer to unity (see Exercise 1 for a specific example). Therefore, they struggle with the fact that the observed level of anisotropy is too small to account for the clustering of matter in the universe. Equivalently, when normalized to large-angle anisotropies, the matter power spectrum is too small. [3]

12.8.3 Acoustic Oscillations

Tightly Coupled Limit of the Boltzmann Equations

When all electrons were ionized, before η_*, the mean free path for a photon was much smaller than the horizon of the universe. Compton scattering caused the electron-proton fluid

to be tightly coupled with the photons. We now proceed to explore this regime quantitatively using the Boltzmann equations.

The tightly coupled limit corresponds to the scattering rate being much larger than the expansion rate: $\tau \gg 1$, where τ is the optical depth defined in Eq. (4.61). I

[3] This realization has been most important in theories in which structure is generated by cosmic strings. Several papers which pointed out the problem in the aftermath of the COBE detection include: Albrecht and Stebbins (1992), Perivolaropoulos and Vachaspati (1994), and Pen and Spergel (1995). want to argue that in the $\tau \gg 1$ limit, the only nonnegligible moments, Θ_l, are the monopole ($l = 0$) and the dipole ($l = 1$). All others are suppressed. In this sense, photons behave just like a fluid, which can be described with only two variables: the density ρ and the velocity $\vec{v}$. In order to show this, let's go back to the Boltzmann equation (4.100) for photons. We want to turn this differential equation for $\Theta(\eta, \mu)$ into an infinite set of coupled equations for $\Theta_l(\eta)$. The advantage is that - as we will see - the higher moments are small and so can be neglected. The strategy is to multiply by $\mathcal{P}_l(\mu)$ and then integrate over μ. Using Eq. (4.99), the Boltzmann equation for $l > 2$ becomes

$$\dot{\Theta}_l + \frac{k}{(-i)^{l+1}} \int_{-1}^{1} \frac{d\mu}{2} \mu \mathcal{P}_l(\mu) \Theta(\mu) = \dot{\tau} \Theta_l.$$

Note that all other terms (e.g., $-\dot{\Phi}$) have simple μ dependence (scale as μ^0 or μ^1) so all $l > 2$ moments vanish for them. To do the integral, we make use of the recurrence relation for Legendre polynomials, Eq. (C.3), to get

$$\dot{\Theta}_l - \frac{kl}{2l+1} \Theta_{l-1} + \frac{k(l+1)}{2l+1} \Theta_{l+1} = \dot{\tau} \Theta_l.$$

Let us consider the order of magnitude of the terms in Eq. (8.10). The first term on the left is of order Θ_l/η which is much smaller than the term on the right which is of order $\tau\Theta_l/\eta$. Neglecting the Θ_{l+1} term for the moment, this tells us that in the tightly coupled regime

$$\Theta_l \sim \frac{k\eta}{2\tau} \Theta_{l-1}$$

For horizon size modes $k\eta \sim 1$, this means that $\Theta_l \ll \Theta_{l-1}$. (By the way, this is justification for throwing out the Θ_{l+1} term in making our estimate.) This estimate is valid for all modes higher than the dipole, so all such modes are very small compared to the monopole and dipole.

Before making use of this fact and deriving the tightly coupled equations in the limit in which only the monopole and dipole are nonzero (the fluid approximation), I want to explain why higher moments are damped in a tightly coupled environment. Indeed this observation is extremely important not only in cosmology but in all settings in which the fluid approximation is used. To understand the fluid approximation, consider one plane-wave perturbation as depicted in Figure 8.5. An observer sitting at the center of the perturbation sees photons arriving from a distance of order the mean free path, η/τ. A wavelength of order the horizon η is much larger than this distance, so the photons arriving at the observer all have the same temperature. There is very little anisotropy. You might think that a perturbation with a very small wavelength (with $k\eta \sim \tau$) would lead to anisotropy. In fact, though, such a mode has a wavelength much smaller than the damping scale. So all perturbations on such small scales are smoothed out, again leading to no anisotropy. The bottom line is there is essentially no anisotropy beyond the monopole and the dipole in the tightly coupled regime. Tightly Coupled Limit

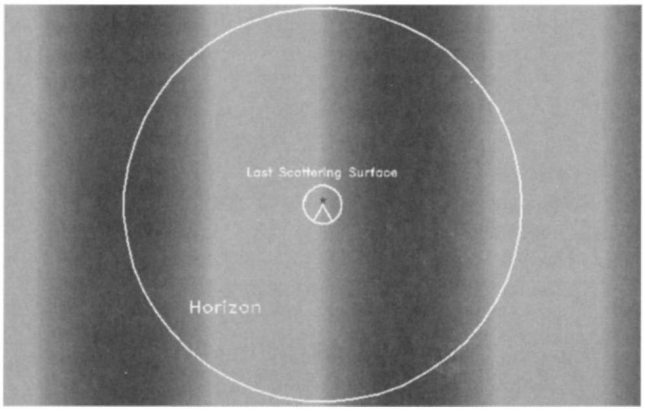


Figure 8.5. Anisotropies in the tightly coupled era. Perturbations on the scale of the horizon cannot be observed by an observer [denoted by the star here], for the photons observed come from the last scattering surface a distance η/τ away. This last scattering surface is so close that photons arriving from all angles have virtually identical temperatures.

Armed with this knowledge, we can now turn to the equations for the first two moments, which - after disposing of Θ_2 - read:

$$\dot{\Theta}_0 + k\Theta_1 = -\dot{\Phi}$$
$$\dot{\Theta}_1 - \frac{k\Theta_0}{3} = \frac{k\Psi}{3} + \dot{\tau}\left[\Theta_1 - \frac{iv_\mathrm{b}}{3}\right]$$

These follow by mutiplying Eq. (4.100) by $\mathcal{P}_0(\mu)$ and $\mathcal{P}_1(\mu)$ and integrating over μ. They are supplemented by the equations for the electron-baryon fluid, Eqs. (4.105) and (4.106). Let us first rewrite the velocity equation, (4.106), as

$$v_\mathrm{b} = -3i\Theta_1 + \frac{R}{\dot{\tau}}\left[\dot{v}_\mathrm{b} + \frac{\dot{a}}{a}v_\mathrm{b} + ik\Psi\right]$$

The second term on the right here is much smaller than the first since it is suppressed by a relative factor of order τ^{-1}. Thus, to lowest order, $v_\mathrm{b} = -3i\Theta_1$. A systematic way to expand, then, is to use this lowest order expression everywhere in the second term, leading to

$$v_\mathrm{b} \simeq -3i\Theta_1 + \frac{R}{\dot{\tau}}\left[-3i\dot{\Theta}_1 - 3i\frac{\dot{a}}{a}\Theta_1 + ik\Psi\right].$$

Now let us insert this expression into Eq. (8.13), eliminating v_b. After rearranging terms, we find

$$\dot{\Theta}_1 + \frac{\dot{a}}{a}\frac{R}{1+R}\Theta_1 - \frac{k}{3[1+R]}\Theta_0 = \frac{k\Psi}{3}.$$

We now have two first-order coupled equations for the first two photon moments, Eqs. (8.12) and (8.16). We can turn these into one second-order equation by differentiating Eq. (8.12) and using Eq. (8.16) to eliminate $\dot{\Theta}_1$:

$$\ddot{\Theta}_0 + k\left[\frac{k\Psi}{3} - \frac{\dot{a}}{a}\frac{R}{1+R}\Theta_1 + \frac{k}{3[1+R]}\Theta_0\right] = -\ddot{\Phi}$$

Finally, we use Eq. (8.12) to eliminate Θ_1 here. This leaves

$$\ddot{\Theta}_0 + \frac{\dot{a}}{a}\frac{R}{1+R}\dot{\Theta}_0 + k^2c_s^2\Theta_0 = -\frac{k^2}{3}\Psi - \frac{\dot{a}}{a}\frac{R}{1+R}\dot{\Phi} - \ddot{\Phi} := F(k,\eta)$$

where I have defined the forcing function on the right as F and the sound speed of the fluid as

$$c_s := \sqrt{\frac{1}{3(1+R)}}$$

The sound speed depends on the baryon density in the universe. In the absence of baryons, it has the standard value for a relativistic fluid, $c_s = 1/\sqrt{3}$. The presence of baryons, though, makes the fluid heavier, thereby lowering the sound speed. We will see shortly that the fluid oscillates in both space and time, with a period which is determined by the sound speed, and hence by the baryon density. Note that Eq. (8.18) is the "grown-up" version of Eq. (8.1); it differs only through the $\dot{\Theta}_0$ damping [4] term (see Exercise 2). The presence of this term does not change any of the qualitiative conlcusions we reached in Section 8.1. Finally, note that Φ enters on the right in a very similar way as Θ_0 does on the left. An alternate version of Eq. (8.18) takes adavantage of this:

$$\left\{\frac{d^2}{d\eta^2} + \frac{\dot{R}}{1+R}\frac{d}{d\eta} + k^2 c_s^2\right\}[\Theta_0 + \Phi] = \frac{k^2}{3}\left[\frac{1}{1+R}\Phi - \Psi\right]$$

Tightly Coupled Solutions

The equation we have derived governing acoustic oscillations of the photon-baryon fluid, (8.20), is a second-order ordinary differential equation. To solve it, we will again (as in Section 7.3.1) use Green's method to find the full solution. First we find the two solutions to the homogeneous equation. Then we use these to construct the particular solution.

In prnciple, to obtain the homogeneous solutions, we must solve the damped, harmonic oscillator equation, (8.20) with the right-hand side equal to zero. In practice, the damping term is of order $R(\Theta_0 + \Phi)/\eta^2$ while the pressure term is much

"This "damping" term is not to be confused with the damping of perturbations on small scales treated in the next section. They are completely different effects. larger, of order $k^2 c_s^2(\Theta_0 + \Phi)$ (at least it's larger when modes are within the horizon or when R is small). Physically we expect pressure to induce oscillations in the photon temperature; the time scale for these oscillations is much shorter than the damping introduced by the expansion of the universe. To a first approximation, then, let us neglect the damping term and simply obtain the oscillating solutions. [5] In this limit, the two homogeneous solutions are

$$S_1(k,\eta) = \sin[k r_s(\eta)] \quad ; \quad S_2(k,\eta) = \cos[k r_s(\eta)]$$

where I have defined the sound horizon as

$$r_s(\eta) := \int_0^\eta d\eta' c_s(\eta')$$

Since c_s is the sound speed, the sound horizon is the comoving distance traveled by a sound wave by time η.

The tightly coupled solution for the photon temperature can be constructed from the homogeneous solutions of Eq. (8.21):

$$\Theta_0(\eta) + \Phi(\eta) = C_1 S_1(\eta) + C_2 S_2(\eta) + \frac{k^2}{3}\int_0^\eta d\eta' [\Phi(\eta') - \Psi(\eta')] \frac{S_1(\eta') S_2(\eta) - S_1(\eta) S_2(\eta')}{S_1(\eta')\dot{S}_2(\eta') - \dot{S}_1(\eta') S_2(\eta')}$$

Here again, I have dropped all occurences of R except in the arguments of the rapidly varying sines and cosines. That is, the argument of S_1, for example, is still taken to be kr_s with its nonzero value of R. We can fix the constants C_1 and C_2 in Eq. (8.23) by appealing to the initial conditions, when both Θ_0 and Φ are constants. The coefficient of the sine term therefore, C_1, must vanish, and $C_2 = \Theta_0(0) + \Phi(0)$. The denominator in the integrand reduces to $-kc_s(\eta') \rightarrow -k/\sqrt{3}$ in the limit in which we are working. Finally, the difference of the products in the numerator of the integrand is simply $-\sin[k(r_s - r_s')]$, so

$$\Theta_0(\eta) + \Phi(\eta) = [\Theta_0(0) + \Phi(0)]\cos(kr_s) + \frac{k}{\sqrt{3}}\int_0^\eta d\eta' [\Phi(\eta') - \Psi(\eta')]\sin[k(r_s(\eta) - r_s(\eta'))]$$

Equation (8.24) is an expression for the anisotropy in the tightly coupled limit, first derived by Hu and Sugiyama in 1995. If you are not impressed with this solution since it still involves an integral over the gravitational potentials, I urge you to reconsider. First, look at Figure 8.6, which compares the solution of Eq. (8.24) with exact results obtained by integrating the full set of coupled Einstein-Boltzmann equations. The approximate solution gets the peak locations dead on, and it does fairly well with the heights as well. The later peaks - those at $k\eta_0 > 500$-are

[5] You can rectify this by applying the WKB approximation in

D-5.

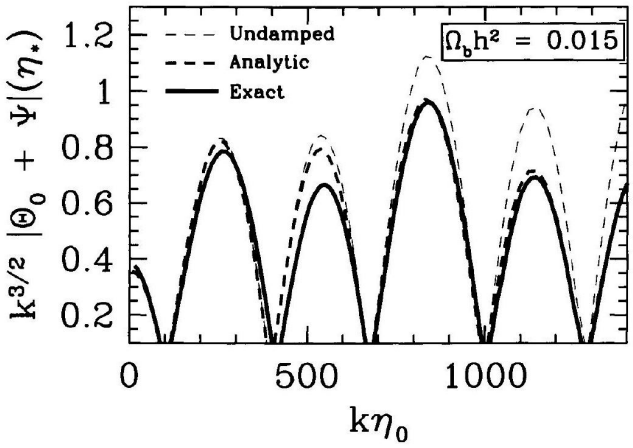


Figure 8.6. The monopole at recombination in a standard CDM model. The exact solution is the heavily weighted solid line. The light dashed line is the undamped solution of Section 8.3, Eq. (8.24); the heavier curve in the middle accounts for damping using the treatment of Section 8.4.

clearly overestimated by our solution, but we will shortly rectify this when we include damping due to diffusion in the next section. A second reason to respect the approximate solution is that it divides the problem neatly into first (i) a calculation of the external gravitional potentials generated by the dark matter and then (ii) the effect of these potentials on the anisotropies. Third, the solution clearly illustrates that the cosine mode is the one excited by inflationary models. This is important, because it is very hard to imagine this mode excited by any other mechanism. If causality is respected, then there should be no perturbations with $k\eta \ll 1$ early on. We know that inflation evades this constraint by changing the true horizon; it is tempting to say that if this mode is observed, we are seeing evidence for inflation. Fourth, we now have a more accurate expression for the frequency of oscillations and therefore for the locations of the acoustic peaks. In the limit that the first term in Eq. (8.24) dominates, the peaks should appear at the extrema of $\cos(kr_s)$, e.g., at

$$k_{\rm p} = n\pi/r_s \quad n = 1, 2, \ldots$$

And the final reason Eq. (8.24) is impressive is that the full set of EinsteinBoltzmann equations involve literally thousands of coupled variables (e.g., the Θ_l 's). Reducing those thousands of differential equations to just one is a huge leap in knowledge. In addition to the monopole, the photon distribution has a nonnegligible dipole at recombination. Using Eq. (8.12), we can obtain an analytic solution for the dipole by differentiating Eq. (8.24):

$$\Theta_1(\eta) = \frac{1}{\sqrt{3}}\left[\Theta_0(0) + \Phi(0)\right]\sin\left(kr_s\right) - \frac{k}{3}\int_0^\eta d\eta'\left[\Phi\left(\eta'\right) - \Psi\left(\eta'\right)\right]\cos\left[k\left(r_s(\eta) - r_s\left(\eta'\right)\right)\right].$$

The first term is completely out of phase with the monopole ($\sin\left(kr_s\right)$ versus $\cos\left(kr_s\right)$). Figure 8.7 shows that this feature remains even after accounting for the integral term. This mismatch of phase will have important implications for the final anisotropy spectrum.

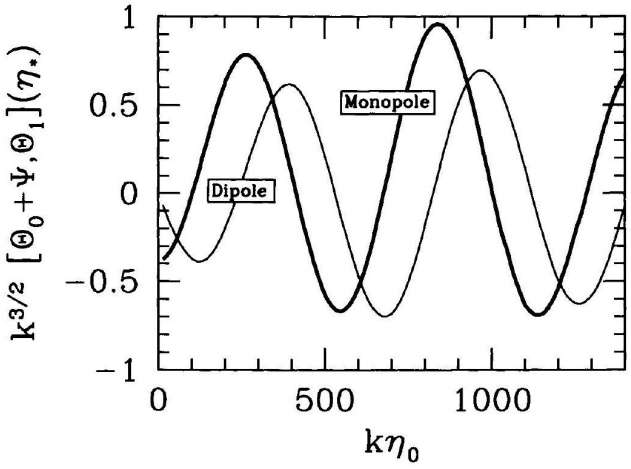


Figure 8.7. The monopole and dipole at recombination in a standard CDM model. The dipole vanishes for the longest wavelength modes that have not entered the horizon by recombination. It is completely out of phase with the monopole.

12.8.4 Diffusion Damping

Figure 8.6 makes it clear that we must account for diffusion to get accurate CMB spectra. To analyze diffusion quantitatively, we must return to the equations for the moments of the photon distribution, Eqs. (8.12), (8.13) and (8.10). Until now, we have neglected Θ_2 and all higher moments. Diffusion is characterized by a small but nonnegligible quadrupole. We must therefore supplement the set of equations we wrote down in the last section with an equation for the quadrupole, Θ_2. Our task is somewhat simplified by the fact that we will be interested in phenomena occurring only on small scales. On these scales, recall from Chapter 7 (e.g., Figure 7.8) that the potentials are very small because of radiation pressure, so we can drop Φ and Ψ everywhere. Also, we will see that diffusion manifests itself in the moments by making each successive moment proportional to a higher power of $1/\dot{\tau}$. Thus we will need to keep only the $l = 2$ mode; all higher ones can be neglected. With these approximations, we have

$$\dot{\Theta}_0 + k\Theta_1 = 0$$
$$\dot{\Theta}_1 + k\left(\frac{2}{3}\Theta_2 - \frac{1}{3}\Theta_0\right) = \dot{\tau}\left(\Theta_1 - \frac{iv_\mathrm{b}}{3}\right)$$
$$\dot{\Theta}_2 - \frac{2k}{5}\Theta_1 = \frac{9}{10}\dot{\tau}\Theta_2.$$

These three equations need to be supplemented by an equation for v_b. This is best expressed as a slight rewriting of Eq. (8.14):

$$3i\Theta_1 + v_\mathrm{b} = \frac{R}{\dot{\tau}}\left[\dot{v}_\mathrm{b} + \frac{\dot{a}}{a}v_\mathrm{b}\right],$$

where again I have dropped the gravitational potential.

To solve this set of equations, we appeal to the high-frequency nature of damping. Let us write the time dependence of the velocity as

$$v_\mathrm{b} \propto e^{i\int \omega d\eta}$$

and similarly for all other variables. We already know that $\omega \simeq kc_s$ in the tightly coupled limit. Now we are searching for damping, an imaginary part to ω. Since damping occurs on small scales, or high frequencies,

$$\dot{v}_\mathrm{b} = i\omega v_\mathrm{b} \gg \frac{\dot{a}}{a}v_\mathrm{b}$$

$\dot{a}/a$ is of order η^{-1} while ω is of order k. So we can drop the second term on the right in Eq. (8.30) and the velocity equation then becomes

$$v_\mathrm{b} = -3i\Theta_1\left[1 - \frac{i\omega R}{\dot{\tau}}\right]^{-1}$$
$$\simeq -3i\Theta_1\left[1 + \frac{i\omega R}{\dot{\tau}} - \left(\frac{\omega R}{\dot{\tau}}\right)^2\right]$$

where I have expanded out to $\dot{\tau}^{-2}$ because $v_\mathrm{b} + 3i\Theta_1$ is multiplied by $\dot{\tau}$ in Eq. (8.28). The equation for the second moment of the photon field, (8.29), can be reduced similarly. First we can drop the $\dot{\Theta}_2$ term since it is much smaller than $\dot{\tau}\Theta_2$. This leaves simply

$$\Theta_2 = -\frac{4k}{9\dot{\tau}}\Theta_1$$

which shows that our approximation scheme is controlled: higher moments are suppressed by additional powers of $k/\dot{\tau}$. The equation for the zeroth moment becomes

$$i\omega\Theta_0 = -k\Theta_1$$

Inserting all of these into Eq. (8.28) gives the dispersion relation

$$i\omega - \frac{8k^2}{27\dot{\tau}} + \left(k^2/3i\omega\right) = \dot{\tau}\left(1 - \left[1 + \frac{i\omega R}{\dot{\tau}} - \left(\frac{\omega R}{\dot{\tau}}\right)^2\right]\right).$$

Collecting terms we get

$$\omega^2(1+R) - \frac{k^2}{3} + \frac{i\omega}{\dot{\tau}}\left[\omega^2 R^2 + \frac{8k^2}{27}\right] = 0$$

The first two terms on the left, the leading ones in the expansion of $1/\dot{\tau}$, recover the result of the previous section, that the frequency is the wavenumber times the speed of sound. We can write the frequency as this zero-order piece plus a firstorder correction, $\delta\omega$. Then, inserting the zero-order part into the terms inversely proportional to $\dot{\tau}$ leads to

$$\delta\omega = -\frac{ik^2}{2(1+R)\dot{\tau}}\left[c_s^2 R^2 + \frac{8}{27}\right]$$

Therefore, the time dependence of the perturbations is

$$\Theta_0, \Theta_1 \sim \exp\left\{ik\int d\eta c_s\right\}\exp\left\{-\frac{k^2}{k_D^2}\right\}$$

where the damping wavenumber is defined via

$$\frac{1}{k_D^2(\eta)} := \int_0^\eta \frac{d\eta'}{6(1+R)n_e\sigma_T a\,(\eta')}\left[\frac{R^2}{(1+R)} + \frac{8}{9}\right]$$

Putting aside factors of order unity, this equation says that $1/k_D \sim [\eta/n_e\sigma_T a]^{1/2}$, which agrees with our heuristic estimate at the beginning of this chapter.

As a first estimate of the damping scale, we can work in the prerecombination limit, in which all electrons (except those in helium) are free. In Chapter 3 we estimated the optical depth in this limit, but ignored helium. The mass fraction of helium is usually denoted as Y_p and is approximately 0.24. Since each helium nucleus contains four nucleons, the ratio of helium to the total number of nuclei is $Y_p/4$. Each of these absorbs two electrons (one for each proton), so when counting the number of free electrons before hydrogen recombination, we must multiply our estimate of Eq. (3.46) by $1 - Y_p/2$. Using the fact that $H_0 = 3.33 \times 10^{-4}$ hMpc^{-1}, we have, in the prerecombination limit,

$$n_e\sigma_T a = 2.3 \times 10^{-5}\mathrm{Mpc}^{-1}\Omega_b h^2 a^{-2}\left(1 - \frac{Y_p}{2}\right)$$

Using this, you can show (Exercise 8) that an approximation for the damping scale is

$$k_D^{-2} = 3.1 \times 10^6 \mathrm{Mpc}^2 a^{5/2} f_D\,(a/a_{\rm eq})\left(\Omega_b h^2\right)^{-1}\left(1 - \frac{Y_p}{2}\right)^{-1}\left(\Omega_m h^2\right)^{-1/2}$$

where f_D, defined in Eq. (8.88), goes to 1 as $a/a_{\rm eq}$ gets large.

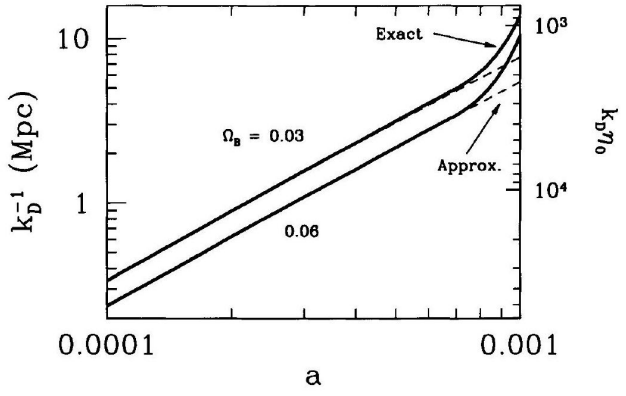


Figure 8.8. Damping scale as a function of the scale factor for two different values of Ω_b (with $h = 0.5$). Heavy curves (exact) numerically integrate over the standard recombination

history, while light curves use the approximation of Eq. (8.42) which assumes electrons remain ionized. Right axis shows the equivalent $k_D\eta_0$; damping occurs on angular scales $l > k_D\eta_0$.

Figure 8.8 shows the evolution of the damping scale before recombination. Neglecting recombination is a good approximation at early times but, as expected, leads to quantitative errors right near η_*, when using Eq. (8.41) for the free electron density does not accurately account for the electrons swept up into neutral hydrogen. In the absence of recombination, k_D scales as $\Omega_b^{1/2}$. Note from the late time behavior in Figure 8.8 that the messy details of recombination change this simple scaling: k_D for the $\Omega_b = 0.06$ case is less than $2^{0.5}$ as big as the $\Omega_b = 0.03$ case.

Figure 8.8 requires one final comment. The damping of anisotropies due to photon diffusion is sometimes referred to as being caused by the "finite thickness of the last scattering surface." That is, it is argued that if recombination took place instantaneously at η_*, then there would be no damping. Figure 8.8 shows that this is patently false. Even if recombination had occurred in this way, the universe before recombination would not have been inifinitely optically thick. Photons would still stream a reasonable distance and hence damp anisotropies. In the examples shown, the damping scale would have been smaller (larger l) by less than a factor of 2 if recombination had occurred instantaneously. On the other hand, we will see in the next section that the anisotropies today are determined by integrating over the visibility function, essentially a filter centered at the epoch of recombination but broadened due to the finite thickness of the last scattering surface. When incorporating the effects of damping (Seljak, 1994; Hu and Sugiyama, 1995), one must account for this finite thickness by integrating the damping function e^{-k^2/k_D^2} weighted by the visibility function. Thus the finite thickness of the last scattering surface has both qualitative and quantitative effects on the final anisotropy spectrum.

12.8.5 Inhomogeneities to Anisotropies

We now have a good handle on the perturbations to the photons at recombination. It is time to transform this understanding into predictions for the anisotropy spectrum today. First, we will solve for the moments Θ_l today in the next subsection. Then we will spend a bit of time relating these moments to the observables. Thus the main purpose of the following subsections is to derive Eq. (8.56), which relates the moments today to the monopole and dipole at recombination, and Eq. (8.68), which expresses the CMB power spectrum in terms of the Fourier moments today.

Free Streaming

We want to derive a formal solution for the photon moments today $\Theta_l(\eta_0)$ in terms of the monopole and dipole at recombination. A formal solution can be obtained by returning to Eq. (4.100). Subtracting $\dot\tau\Theta$ from both sides leads to

$$\dot\Theta + (ik\mu - \dot\tau)\Theta = e^{-ik\mu\eta+\tau}\frac{d}{d\eta}\left[\Theta e^{ik\mu\eta-\tau}\right] = \tilde S$$

where the source function is defined as

$$\tilde S := -\dot\Phi - ik\mu\Psi - \dot\tau\left[\Theta_0 + \mu v_{\rm b} - \frac{1}{2}\mathcal{P}_2(\mu)\Pi\right].$$

Hold your curiosity about the in the definition. Multplying both sides of Eq. (8.43) by the exponential and then integrating over η leads directly to

$$\Theta\left(\eta_0\right) = \Theta\left(\eta_{\rm init}\right) e^{ik\mu(\eta_{\rm init}-\eta_0)} e^{-\tau(\eta_{\rm init})} + \int_{\eta_{\rm init}}^{\eta_0} d\eta \tilde{S}(\eta) e^{ik\mu(\eta-\eta_0)-\tau(\eta)}$$

where I have used the fact that $\tau\left(\eta=\eta_0\right)=0$ since τ is defined as the scattering rate integrated from η up to η_0. We also know that, if the initial time $\eta_{\rm init}$ is early enough, then the optical depth $\tau\left(\eta_{\rm init}\right)$ will be extremely large. Therefore, the first term on the right side of Eq. (8.45) vanishes. This corresponds to the fact that any initial anisotropy is completely erased by Compton scattering. By the same reasoning, we can set the lower limit on the integral to zero: any contribution to the integrand from $\eta < \eta_{\rm init}$ is completely negligible. Thus, the solution for the perturbations is

$$\Theta\left(k,\mu,\eta_0\right) = \int_0^{\eta_0} d\eta \tilde{S}(k,\mu,\eta) e^{ik\mu(\eta-\eta_0)-\tau(\eta)}.$$

Equation (8.46) looks simple, but of course all of the complication is hidden in the source function $\tilde{S}$. Notice that $\tilde{S}$ depends somewhat on the angle μ. If it did not depend on μ, we could immediately turn Eq. (8.46) into an equation for each of the Θ_l 's. For, we could multiply each side by the Legendre polynomial $\mathcal{P}_l(\mu)$ and then integrate over all μ. By Eq. (4.99), the left side would give $(-i)^l\Theta_l$ and the right would contain the integral

$$\int_{-1}^{1} \frac{d\mu}{2} \mathcal{P}_l(\mu) e^{ik\mu(\eta-\eta_0)} = \frac{1}{(-i)^l} j_l\left[k\left(\eta-\eta_0\right)\right]$$

where j_l is the spherical Bessel function. This approach looks so promising that we should pursue it to its end, again forgetting for the moment that $\tilde{S}$ really does have some μ dependence. The expression for Θ_l would be

$$\Theta_l\left(k,\eta_0\right) = (-1)^l \int_0^{\eta_0} d\eta \tilde{S}(k,\eta) e^{-\tau(\eta)} j_l\left[k\left(\eta-\eta_0\right)\right]$$

What about the μ dependence in $\tilde{S}$? We can account for this by noting that $\tilde{S}$ multiplies the exponential $e^{ik\mu(\eta-\eta_0)}$ in Eq. (8.46). Thus, everywhere we encounter a factor of μ in $\tilde{S}$ we can replace it with a time derivative:

$$\mu \to \frac{1}{ik}\frac{d}{d\eta}$$

Let me demonstrate this explicitly with the $-ik\mu\Psi$ term in $\tilde{S}$. The integral is

$$\begin{aligned} -ik\int_0^{\eta_0} d\eta \mu\Psi e^{ik\mu(\eta-\eta_0)-\tau(\eta)} &= -\int_0^{\eta_0} d\eta \Psi e^{-\tau(\eta)} \frac{d}{d\eta} e^{ik\mu(\eta-\eta_0)} \\ &= \int_0^{\eta_0} d\eta e^{ik\mu(\eta-\eta_0)} \frac{d}{d\eta}\left[\Psi e^{-\tau(\eta)}\right] \end{aligned}$$

where the last line follows by integration by parts. Note that the surface terms can be dropped: at $\eta=0$ they are damped by the $e^{-\tau(0)}$ factor. The terms at $\eta=\eta_0$ are not small, but they are irrelevant since they have no angular dependence. They alter the monopole, an alteration which we cannot detect. Thus, accounting for the integration by parts changes the

substitution rule of Eq. (8.49) by a minus sign, with the understanding that the derivative does not act on the oscillating part of the exponential, $e^{ik\mu(\eta-\eta_0)}$. The solution in Eq. (8.48) therefore becomes

$$\Theta_l\left(k,\eta_0\right) = \int_0^{\eta_0} d\eta S(k,\eta) j_l\left[k\left(\eta_0-\eta\right)\right]$$

with the source function now defined as

$$\begin{aligned} S(k,\eta) := e^{-\tau}&\left[-\dot{\Phi}-\dot{\tau}\left(\Theta_0+\frac{1}{4}\Pi\right)\right] \\ &+\frac{d}{d\eta}\left[e^{-\tau}\left(\Psi-\frac{iv_{\rm b}\dot{\tau}}{k}\right)\right]-\frac{3}{4k^2}\frac{d^2}{d\eta^2}\left[e^{-\tau}\dot{\tau}\Pi\right]. \end{aligned}$$

In Eq. (8.51), I have also used the property of spherical Bessel functions: $j_l(x) = (-1)^l j_l(-x)$. At this stage, it is useful to introduce the visibility function

$$g(\eta) := -\dot{\tau}e^{-\tau}.$$

The visibility function has some interesting properties. The integral $\int_0^{\eta_0} d\eta g(\eta) = 1$, so we can think of it as a probability density. It is the probability that a photon last scattered at η. In the standard recombination, since τ is so large early on, this probability is essentially zero for η earlier than the time of recombination. It also declines rapidly after recombination, because the prefactor $-\dot{\tau}$, which is the scattering rate, is quite small. Figure 8.9 shows the visibility function for two values of the baryon density.

The source function in Eq. (8.52) can now be expressed in terms of the visibility function. If we drop the polarization tensor Π in the source since it is very small, then the source function becomes

$$\begin{aligned} S(k,\eta) \simeq\ & g(\eta)\left[\Theta_0(k,\eta)+\Psi(k,\eta)\right] \\ &+\frac{d}{d\eta}\left(\frac{iv_{\rm b}(k,\eta)g(\eta)}{k}\right) \\ &+e^{-\tau}[\dot{\Psi}(k,\eta)-\dot{\Phi}(k,\eta)]. \end{aligned}$$

We can take our analytic solution one step further by performing the time integral in Eq. (8.51). The source term proportional to $v_{\rm b}$ is best treated by integrating by parts. Then,

$$\begin{aligned} \Theta_l\left(k,\eta_0\right) =& \int_0^{\eta_0} d\eta g(\eta)\left[\Theta_0(k,\eta)+\Psi(k,\eta)\right] j_l\left[k\left(\eta_0-\eta\right)\right] \\ &-\int_0^{\eta_0} d\eta g(\eta)\frac{iv_{\rm b}(k,\eta)}{k}\frac{d}{d\eta}j_l\left[k\left(\eta_0-\eta\right)\right] \end{aligned}$$

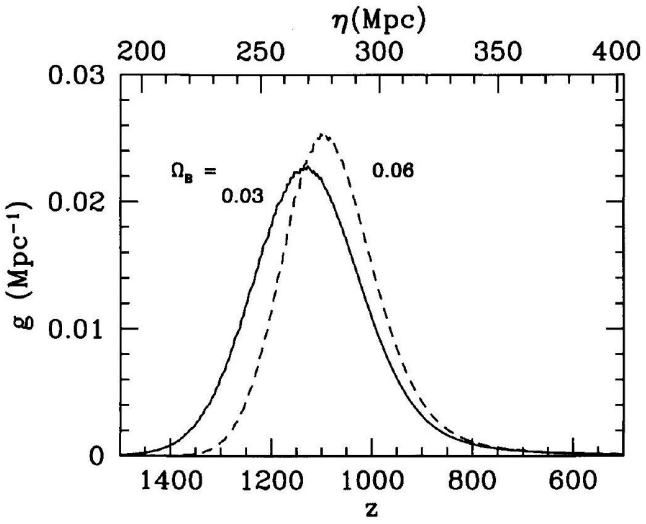
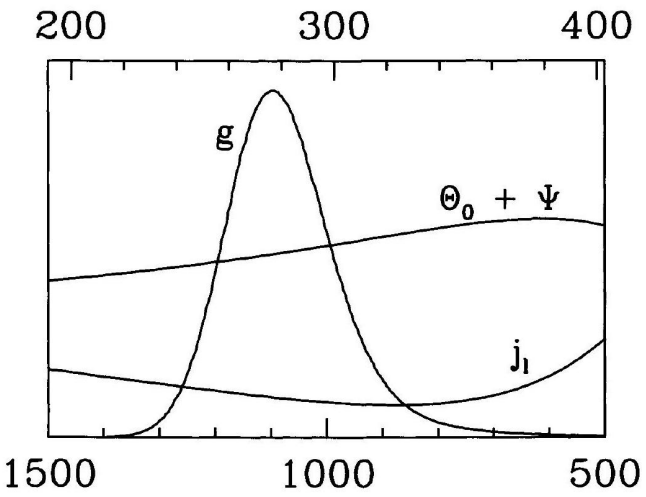


Figure 8.9. The visibility function. Most electrons last scatter at around $z \simeq 1100$ with little dependence on the baryon density. Note that the integral of g over conformal time is 1. Here $h = 0.5$.

$$+\int_0^{\eta_0} d\eta e^{-\tau}[\dot{\Psi}(k,\eta) - \dot{\Phi}(k,\eta)]j_l\left[k\left(\eta_0 - \eta\right)\right].$$

There are two types of terms in Eq. (8.55). First, there are those wherein the integral is weighted by $e^{-\tau}$. These contribute as long as $\tau \lesssim 1$, that is, at all times after recombination. Note that these are also proportional to derivatives of the potentials. If the potentials are constant after recombination, these terms vanish. In many theories, as we saw in Chapter 7, this is precisely what happens: the universe is purely matter dominated after recombination and in such an environment, the potentials generally remain constant. The corrections due to changing potentials are therefore important to get things right quantitatively, but do not affect the qualitative structure of the anisotropy spectrum. Rather, the dominant terms in Eq. (8.55) are the second types of terms, the ones with integrals weighted by the visibility function.

Since the visibility function is so sharply peaked, the integrals in the first two terms become very simple. To see why, consider Figure 8.10 which shows the three parts of the integrand of the first term (the monopole) in Eq. (8.55). Since the visibility function changes rapidly compared with the other two functions, we can evaluate those other functions at the peak of the visibility function, i.e., at $\eta = \eta_*$, and remove them from the integral. But then, the integral is simply $\int d\eta g(\eta) = 1$. Thus, we are left with η(Mpc)



Z

Figure 8.10. The three components of the integrand in the monopole term of Eq. (8.55). The visibility function is sharply peaked, so it changes rapidly compared with the monopole $\Theta_0 + \Psi$ and the Bessel function $j_l\left(k\left[\eta - \eta_0\right]\right)$. Figure is for $l = 100, k = 0.013h\mathrm{Mpc}^{-1}$.

$$\begin{aligned}\Theta_l\left(k,\eta_0\right) \simeq & \left[\Theta_0\left(k,\eta_*\right) + \Psi\left(k,\eta_*\right)\right] j_l\left[k\left(\eta_0 - \eta_*\right)\right] \\ & + 3\Theta_1\left(k,\eta_*\right)\left(j_{l-1}\left[k\left(\eta_0 - \eta_*\right)\right] - \frac{(l+1) j_l\left[k\left(\eta_0 - \eta_*\right)\right]}{k\left(\eta_0 - \eta_*\right)}\right) \\ & + \int_0^{\eta_0} d\eta e^{-\tau}[\dot{\Psi}(k,\eta) - \dot{\Phi}(k,\eta)] j_l\left[k\left(\eta_0 - \eta\right)\right]\end{aligned}$$

Here I have used the spherical Bessel function identity of Eq. (C.18) to rewrite the Bessel function derivative in the velocity term and also the fact that $v_\mathrm{b} \simeq -3i\Theta_1$ at η_*. On scales much smaller than the one shown in Figure 8.10, $\Theta_0 + \Psi$ changes more rapidly because of the rapid change in the damping scale around recombination. However, this effect can be incorporated by modifying the damping function from

$$e^{-k^2/k_D(\eta_*)^2} \rightarrow \int d\eta g(\eta) e^{-k^2/k_D(\eta)^2}$$

Equation (8.56) is the basis for semianalytic calculations (Seljak, 1994; Hu and Sugiyama, 1995) of C_l spectra which agree with the exact (numerical) solutions to within 10%. From Eq. (8.56), we see that, to solve for the anisotropies today, we must know the monopole (Θ_0), dipole (Θ_1), and potential (Ψ) at the time of recombination. Further, there will be small but noticable corrections if the potentials are time dependent. These corrections, encoded in the last line of Eq. (8.56), are often called integrated Sachs-Wolfe (ISW) terms.

The monopole term - the first in Eq. (8.56) - is precisely what we expected from the rough arguments of Section 8.1. In particular, the spherical Bessel function, $j_l\left[k\left(\eta_0 - \eta_*\right)\right]$, determines how much anisotropy on an angular scale l^{-1} is contributed by a plane wave with wavenumber k. On very small angular scales,

$$\lim_{l\rightarrow\infty} j_l(x) = \frac{1}{l}\left(\frac{x}{l}\right)^{l-1/2}$$

That is, $j_l(x)$ is extremely small for large l when $x < l$. In our case, this means that $\Theta_l\left(k,\eta_0\right)$ is very close to zero for $l > k\eta_0$. This makes sense physically. Returning to Figure 8.4, we see that very small angular scales will see little anisotropy from a perturbation with a large wavelength. The converse is also true: angular scales larger than $1/\left(k\eta_0\right)$ get little contribution from such a perturbation. To sum up, a perturbation with wavenumber k contributes predominantly on angular scales of order $l \sim k\eta_0$. One last comment about the monopole term: the final anisotropy today depends on not just Θ_0, but rather $\Theta_0 + \Psi$, again something we anticipated since photons must climb out of their potential wells to reach us today.

The C_l 's

How is the observed anisotropy pattern today related to the rather abstract $\Theta_l\left(k,\eta_0\right)$? To answer this question, we must first describe the way in which the temperature field is characterized today and then relate this characterization to Θ_l.

Recall that in Eq. (4.34), we wrote the temperature field in the universe as

$$T(\vec{x},\hat{p},\eta) = T(\eta)[1 + \Theta(\vec{x},\hat{p},\eta)]$$

Although this field is defined at every point in space and time, we can observe it only here (at $\vec{x}_0$) and now (at η_0) [6] Our only handle on the anisotropies is their dependence on the direction of the incoming photons, $\hat{p}$. So all the richness we observe comes from the changes in the temperature as the direction vector $\hat{p}$ changes. Observers typically makes maps, wherein the temperature is reported at a number of incoming directions, or "spots on the sky." These spots are labeled not by the $\hat{p}_x, \hat{p}_y, \hat{p}_z$ components of $\hat{p}$, but rather by polar coordinates θ, ϕ. However, it

[6] We do make small excursions from this point in space-time. For example, satellites are not located on Earth and anisotropy measurements have been made over the past 30 years. These are completely insignificant on scales over which the temperature is varying, which are of order the Hubble time (or distance). is a simple matter to move back and forth between the 3D unit vector $\hat{p}$ and polar coordinates. [7] I'll stick with $\hat{p}$ in the ensuing derivation.

We now expand the field in terms of spherical harmonics. That is, we write

$$\Theta(\vec{x}, \hat{p}, \eta) = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} a_{lm}(\vec{x}, \eta) Y_{lm}(\hat{p}).$$

The subscripts l, m are conjugate to the real space unit vector $\hat{p}$, just as the variable $\vec{k}$ is conjugate to the Fourier transform variable $\vec{x}$. We are all familiar with Fourier transforms, so it is useful to think of the expansion in terms of spherical harmonics as a kind of generalized Fourier transform. Whereas the complete set of eigenfunctions for the Fourier transform are $e^{i\vec{k}\cdot\vec{x}}$, here the complete set of eigenfunctions for expansion on the surface of a sphere are $Y_{lm}(\hat{p})$. All of the information contained in the temperature field T is also contained in the space-time dependent amplitudes a_{lm}. As an example of this, consider an experiment which maps the full sky with an angular resolution of 7°. The full sky has 4π radians $^2 \simeq 41,000$ degrees ^{2}, so there are 840 pixels with area of $(7^\circ)^2$. Thus, such an experiment would have 840 independent pieces of information. Were we to characterize this information with a_{lm} 's instead of temperatures in pixels, there would be some $l_{\max}$ above which there is no information. One way to determine this $l_{\max}$ is to set the total number of recoverable a_{lm} 's as $\sum_{l=0}^{l_{\max}} (2l+1) = (l_{\max}+1)^2 = 840$. So the information could be equally well characterized by specifying all the a_{lm} 's up to $l_{\max} = 28$. Incidentally, this is a fairly good caricature of the COBE experiment (Smoot et al., 1992; Bennett et al., 1996). They presented temperature data over many more pixels, but many of these pixels were overlapping. So, the independent information was contained in multipoles up to $l \sim 30$. Experiments currently under way or well along in the planning stage are capable of measuring the moments all the way up to $l \sim 10^4$.

We want to relate the observables, the a_{lm} 's, to the Θ_l we have been dealing with. To do this, we can use the orthogonality property of the spherical harmonics. The Y_{lm} 's are normalized via Eq. (C.11),

$$\int d\Omega Y_{lm}(\hat{p}) Y^*_{l'm'}(\hat{p}) = \delta_{ll'}\delta_{mm'}$$

where Ω is the solid angle spanned by $\hat{p}$. Therefore the expansion of Θ in terms of spherical harmonics, Eq. (8.60), can be inverted by multiplying both sides by $Y^*_{lm}(\hat{p})$ and integrating:

$$a_{lm}(\vec{x}, \eta) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{x}} \int d\Omega Y^*_{lm}(\hat{p}) \Theta(\vec{k}, \hat{p}, \eta)$$

Here I have written the right-hand side in terms of the Fourier transform ($\Theta(\vec{k})$ instead of $\Theta(\vec{x})$), since that is the quantity for which we obtained solutions.

[7] $\hat{p}_z = \cos\theta, \hat{p}_x = \sin\theta\cos\phi$, and $\hat{p}_y = \sin\theta\sin\phi$.

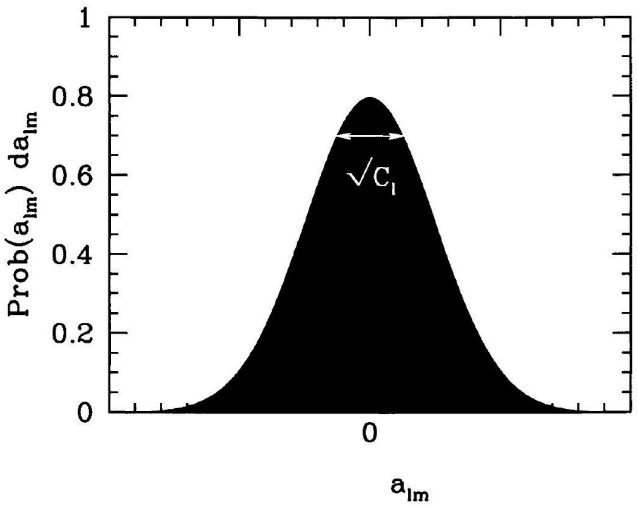


Figure 8.11. The distribution from which the a_{lm} 's are drawn. The distribution has expectation equal to zero and a width of $C_l^{1/2}$.

As with the density perturbations, we cannot make predictions about any particular a_{lm}, just about the distribution from which they are drawn, a distribution which traces its origin to the quantum fluctuations first laid down during inflation. Figure 8.11 illustrates this distribution. The mean value of all the a_{lm} 's is zero, but they will have some nonzero variance. The variance of the a_{lm} 's is called C_l. Thus,

$$\langle a_{lm} \rangle = 0 \quad ; \quad \langle a_{lm} a^*_{l'm'} \rangle = \delta_{ll'} \delta_{mm'} C_l$$

It is very important to note that, for a given l, each a_{lm} has the same variance. For $l = 100$, say, all $201 a_{100,m}$'s are drawn from the same distribution. When we measure these 201 coefficients, we are sampling the distribution. This much information will give us a good handle on the underlying variance of the distribution. On the other hand, if we measure the five components of the quadrupole ($l = 2$), we do not get very much information about the underlying variance, C_2. Thus, there is a fundamental uncertainty in the knowledge we may get about the C_l 's. This uncertainty, which is most pronounced at low l, is called cosmic variance. Quantitatively, the uncertainty scales simply as the inverse of the square root of the number of possible samples, or

$$\left(\frac{\Delta C_l}{C_l} \right)_{\text{cosmic variance}} = \sqrt{\frac{2}{2l+1}}$$

We can now obtain an expression for C_l in terms of $\Theta_l(k)$. First we square a_{lm} in Eq. (8.62) and take the expectation value of the distribution. For this we need $\left\langle \Theta(\vec{k}, \hat{p}) \Theta^* \left(\vec{k}', \hat{p}' \right) \right\rangle$, where from now on we will keep the η dependence implicit. This expectation value is complicated because it depends on two separate phenomena: (i) the initial amplitude and phase of the perturbation is chosen during inflation from a Gaussian distribution and (ii) the evolution we have studied in this chapter turns this initial perturbation into anisotropies, i.e. produces the dependence on $\hat{p}$. To simplify then, it makes sense to separate these two phenomena and write the photon distribution as $\delta \times (\Theta/\delta)$, where the dark matter overdensity δ does not depend on any direction vector. The ratio Θ/δ is precisely what we have solved for in the last two chapters: given the initial amplitude of a mode, we have learned how to evolve forward in time. The ratio does not depend on the initial amplitude, so it can be removed from the averaging over the distribution. Therefore,

$$\left\langle \Theta(\vec{k},\hat{p})\Theta\left(\vec{k}',\hat{p}'\right)\right\rangle = \left\langle \delta(\vec{k})\delta^*\left(\vec{k}'\right)\right\rangle \frac{\Theta(\vec{k},\hat{p})}{\delta(\vec{k})}\frac{\Theta^*\left(\vec{k}',\hat{p}\right)}{\delta^*\left(\vec{k}'\right)}$$
$$= (2\pi)^3\delta^3\left(\vec{k}-\vec{k}'\right)P(k)\frac{\Theta(k,\hat{k}\cdot\hat{p})}{\delta(k)}\frac{\Theta^*\left(k,\hat{k}\cdot\hat{p}'\right)}{\delta^*(k)},$$

where the second equality uses the definition of the matter power spectrum $P(k)$, but also contains a subtlety in the ratio Θ/δ. This ratio, which is determined solely by the evolution of both δ and Θ, depends only on the magnitude of $\vec{k}$ and the dot product $\hat{k}\cdot\hat{p}$. Two modes with the same k and $\hat{k}\cdot\hat{p}$ evolve identically even though their initial amplitudes and phases are different.

After squaring Eq. (8.62), we see that the anisotropy spectrum is

$$C_l = \int \frac{d^3k}{(2\pi)^3}P(k)\int d\Omega Y^*_{lm}(\hat{p})\frac{\Theta(k,\hat{k}\cdot\hat{p})}{\delta(k)}\int d\Omega' Y_{lm}\left(\hat{p}'\right)\frac{\Theta^*\left(k,\hat{k}\cdot\hat{p}'\right)}{\delta^*(k)}.$$

Now we can expand $\Theta\left(k,\hat{k}\cdot\hat{p}'\right)$ and $\Theta\left(k,\hat{k}\cdot\hat{p}'\right)$ in spherical harmonics using the inverse of Eq. (4.99), $\Theta(k,\hat{k}\cdot\hat{p}) = \sum_l (-i)^l(2l+1)\mathcal{P}_l(\hat{k}\cdot\hat{p})\Theta_l(k)$. This leaves

$$C_l = \int \frac{d^3k}{(2\pi)^3}P(k)\sum_{l'l''}(-i)^{l'}(i)^{l''}\left(2l'+1\right)\left(2l''+1\right)\frac{\Theta_{l'}(k)\Theta^*_{l''}(k)}{|\delta(k)|^2}$$
$$\times\int d\Omega\mathcal{P}_{l'}(\hat{k}\cdot\hat{p})Y^*_{lm}(\hat{p})\int d\Omega'\mathcal{P}_{l''}\left(\hat{k}\cdot\hat{p}'\right)Y_{lm}\left(\hat{p}'\right).$$

The two angular integrals here (Exercise 9) are identical. They are nonzero only if $l' = l$ and $l'' = l$, in which case they are equal to $4\pi Y_{lm}(\hat{k})/(2l+1)$ (or the complex conjugate). The angular part of the d^3k integral then becomes an integral over $|Y_{lm}|^2$, which is just equal to 1, leaving

$$C_l = \frac{2}{\pi}\int_0^\infty dk k^2 P(k)\left|\frac{\Theta_l(k)}{\delta(k)}\right|^2$$

For a given l, then, the variance of a_{lm}, C_l, is an integral over all Fourier modes of the variance of $\Theta_l(\vec{k})$. We can now use Eqs. (8.56) and (8.68) to plot the anisotropy spectrum today.

12.8.6 The Anisotropy Spectrum Today

Sachs-Wolfe Effect

Large-angle anisotropies are not affected by any microphysics: at the time of recombination, the perturbations responsible for these anisotropies were on scales far larger than could be connected via causal processes. On these largest of scales, only the monopole contributes to the anisotropy; this is the first term in Eq. (8.56). So the large-angle anisotropy is determined by $\Theta_0 + \Psi$ evaluated at recombination. The large-scale solution we found in Eq. (8.6) was that this combination is equal to $\Psi(\eta_*)/3$. In most cosmological models, recombination occurs far enough after matter/radiation equality that we can approximate the potential back then to be equal to the potential today modulo the growth factor, so

$$\Theta_0(\eta_*) + \Psi(\eta_*) \simeq \frac{1}{3D_1(a=1)}\Psi(\eta_0) = -\frac{1}{3D_1(a=1)}\Phi(\eta_0)$$

The last equality holds here because at very late times, there are no appreciable anisotropic stresses, and $\Phi = -\Psi$.

We may use Eq. (7.7) to express the potential Φ today in terms of the dark matter distribution, so that

$$\Theta_0(\eta_*) + \Psi(\eta_*) \simeq -\frac{\Omega_m H_0^2}{2k^2 D_1(a=1)}\delta(\eta_0)$$

This gives us what we need: an expression for the sum of $\Theta_0 + \Psi$ at recombination that we can plug into the monopole term in Eq. (8.56). To get the anisotropy spectrum today, we then integrate as in Eq. (8.68), leaving

$$C_l^{\rm SW} \simeq \frac{\Omega_m^2 H_0^4}{2\pi D_1^2(a=1)} \int_0^\infty \frac{dk}{k^2} j_l^2\left[k(\eta_0 - \eta_*)\right] P(k)$$

where the superscript denotes Sachs-Wolfe, in honor of the first people to compute the large-angle anisotropy (Sachs and Wolfe, 1967). The power spectrum is given by Eq. (7.9) with the transfer function set to 1 (since we're considering very large scales). Therefore,

$$C_l^{\rm SW} \simeq \pi H_0^{1-n}\left(\frac{\Omega_m}{D_1(a=1)}\right)^2 \delta_H^2 \int_0^\infty \frac{dk}{k^{2-n}} j_l^2\left[k(\eta_0 - \eta_*)\right]$$

The large-scale anisotropies in Eq. (8.72) can be computed analytically. First, we will use the fact that $\eta_* \ll \eta_0$ and define the dummy variable $x := k\eta_0$. Then the spectrum can be rewritten as

$$C_l^{\rm SW} \simeq \pi(\eta_0 H_0)^{1-n}\left(\frac{\Omega_m}{D_1(a=1)}\right)^2 \delta_H^2 \int_0^\infty \frac{dx}{x^{2-n}} j_l^2(x)$$

The integral over the spherical Bessel functions can be analytically expressed (Eq. (C.17) from Gradshteyn and Ryzhik, 6.574.2) in terms of gamma functions, leaving

$$C_l^{\rm SW} \simeq 2^{n-4}\pi^2(\eta_0 H_0)^{1-n}\left(\frac{\Omega_m}{D_1(a=1)}\right)^2 \delta_H^2 \frac{\Gamma\left(l + \frac{n}{2} - \frac{1}{2}\right)}{\Gamma\left(l + \frac{5}{2} - \frac{n}{2}\right)} \frac{\Gamma(3-n)}{\Gamma^2\left(2 - \frac{n}{2}\right)}.$$

If the spectrum is Harrison-Zel'dovich-Peebles, $n = 1$, then the first ratio of the gamma functions $\Gamma(l)/\Gamma(l+2)$ is equal to $[l(l+1)]^{-1}$ using Eq. (C.24). The remaining ratio of gamma functions $\Gamma(2)/\Gamma^2(3/2) = 4/\pi$ using Eq. (C.25), so

$$l(l+1)C_l^{\rm SW} = \frac{\pi}{2}\left(\frac{\Omega_m}{D_1(a=1)}\right)^2 \delta_H^2$$

a constant. Indeed, this is the reason why workers in the field typically plot $l(l+1)C_l$: at low l, where the Sachs-Wolfe approximation is a good one, we expect a plateau.

Figure 8.12 shows the COBE measurements of the large-angular-scale anisotropies along with the Boltzmann solutions of three CDM models. Note that, even for $n = 1$, the true spectrum is not completely flat as suggested by Eq. (8.75). The dipole at recombination (neglected in Eq. (8.74)) contributes slightly. The integrated Sachs-Wolfe effect also is not completely negligible, especially in the Λ model, wherein the potential starts to decay once the

universe becomes Λ dominated at late times. For an $n = 1$ spectrum, the best fit values of δ_H from COBE are

$$\delta_H = 1.9 \times 10^{-5} \quad \Omega_m = 1$$
$$\delta_H = 4.6 \times 10^{-5} \quad \Omega_m = 0.3; \Omega_\Lambda = 0.7.$$

Also shown in Figure 8.12 is a tilted model, one in which the primordial spectral index n is not equal to 1. In such models, the anisotropy should scale as l^{n-1} compared with the Harrison-Zel'dovich-Peebles $n = 1$ spectrum. You can see this scaling from Eq. (8.74) or more directly from the integral in Eq. (8.73). The integrand peaks at $x \sim l$, so very roughly every appearence of x there can be replaced by l. The generalization of the integrand from x^{-1} to x^{n-2} therefore leads to a change in the spectrum that scales as l^{n-1}. As indicated in Figure 8.12, the COBE data have the greatest weight at $l \sim 10$, but cover a range of l spanning an order of magnitude. Extreme values of tilt are therefore ruled out by COBE. To get much better constraints on the tilt, though, measurements spanning a larger range of l are necessary.

Small Scales

The small-scale anisotropy spectrum depends not only on the monopole, but also on the dipole and the integrated Sachs-Wolfe effect. Figure 8.13 shows all these contributions to the spectrum. Let's consider each in turn.

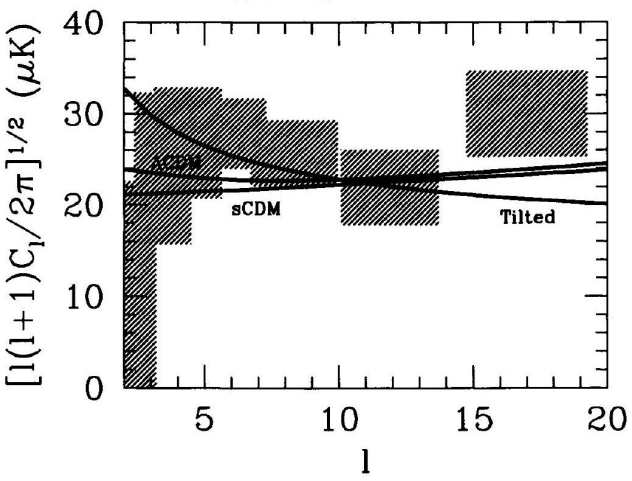


Figure 8.12. Large-scale anisotropies. Hatched boxes show measurements by COBE satellite (Bennett et al., 1996). Curves show the spectra for standard CDM and ACDM (both with $n = 1$). The tilted model is identical to standard CDM, except $n = 0.5$. The late time integrated Sachs-Wolfe effect enhances anisotropy on the largest scales in ΛCDM. Note that here, and in subsequent C_l figures, the root mean square anisotropy is plotted, proportional to $C_l^{1/2}$.C_l is dimensionless so the units of μK come from multiplying by the present background temperature, $T = 2.73$ K.

The monopole at recombination $(\Theta_0 + \Psi)(k, \eta_*)$ free-streams to us today, creating anisotropies on angular scales $l \sim k\eta_0$. This is what we expected back in Figure 8.4, showed to be true in Eq. (8.56), and can now see directly in the top panel of Figure 8.13. There are two interesting features of the quantitative aspect of the free-streaming process. First, note that the "zeroes" in the monopole spectrum, here at 400,650, and 970, are smoothed out because many modes contribute to anisotropy on a given angular scale. If only the $k = 400/\eta_0$ modes contributed to the anisotropy at $l = 400$, then C_{400} would really be zero. But many nonzero modes, with wavenumber greater than $400/\eta_0$, contribute. These change the zero to a trough in the C_l spectrum.

The second feature of free-streaming worth noticing is that our initial estimate tof the peak positions is not exactly right. Inhomogeneity on scale k does not show up as anisotropy precisely on angular scale $l = k\eta_0$. Rather, There is a noticable shift in the top panel, suggesting that a given k-mode contributes to slightly smaller l than we anticipated. This shift arises from the spherical Bessel function in Eq. (8.56). As indicated in Figure 8.14, the peak in the Bessel function comes not when $l = k\eta_0$, but rather at slightly smaller values of l. In addition, our initial

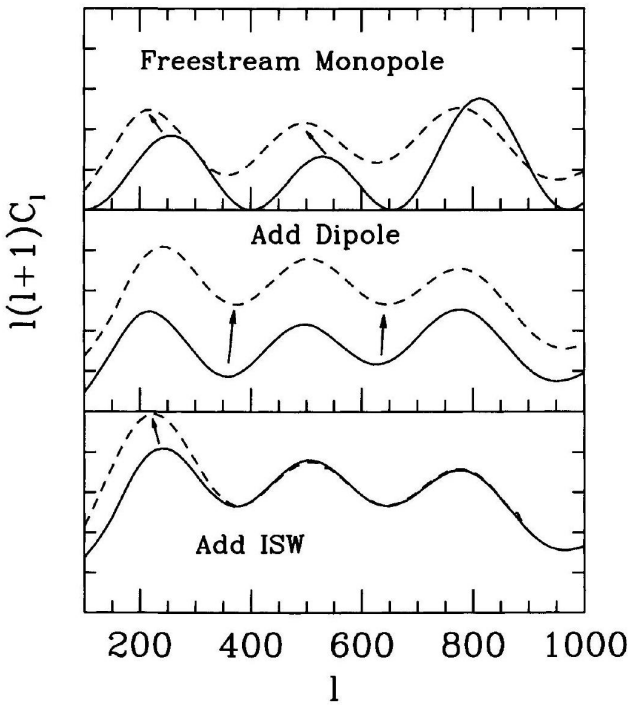


Figure 8.13. Small-scale anisotropy. Top panel: The monopole at recombination $(\Theta_0 + \Psi)\,(k = l/\eta_0, \eta_*)$ contains most of the structure of the final anisotropy spectrum. When free-streamed via the integral in Eq. (8.56), the spectrum shifts slightly to lower l. Middle panel: Accounting for the dipole raises the anisotropy spectrum. Since the dipole is out of phase with the monopole, the troughs become less pronounced. Bottom panel: The integrated Sachs-Wolfe effect enhances the anisotropy on scales comparable to the horizon. In this case, the potential changes near recombination since the universe is not purely matter dominated then. Thus the first peak gets most of the excess power. Throughout, $h = 0.5, \Omega_b = 0.06, \Omega_m = 1$.

estimate for the location of the peaks in k-space, Eq. (8.25), is also slightly high. For example, the expected position of the first peak, $\pi\eta_0/r_s$, for the model depicted in Figure 8.13 is a little over 280. The first peak in the monopole in k-space, however, appears at $k\eta_0 \sim 260$. These two effects - fixed k projects to slightly smaller l and peaks on slightly larger scales than expected from Eq. (8.25) - serve to move the predicted positions of the peaks to lower l. A better approximation for the first

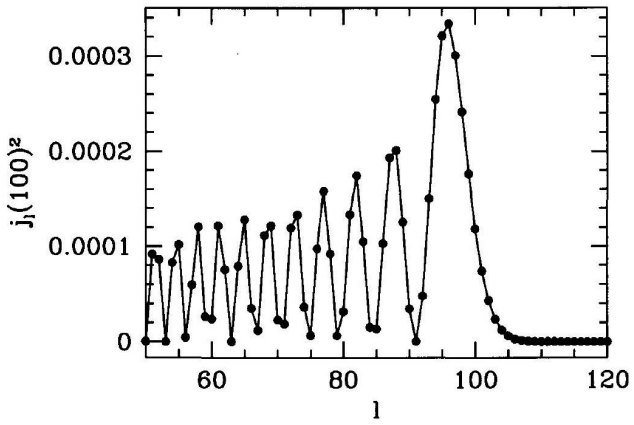


Figure 8.14. The spherical Bessel function, $j_l(100)$. Note that the peak occurs at $l \simeq 90$, slightly smaller than the argument.

peak position is $l_p \simeq 0.75\pi\eta_0/r_s$.

The dipole at recombination is smaller than the monopole and out of phase with it. The middle panel in Figure 8.13 shows that adding in the dipole raises the overall anisotropy level, but particularly fills in the troughs. Without the dipole (in this model) the ratio of the height of the first peak (at $l \sim 200$) to the height of the first trough (at $l \sim 400$) is about 2.5 : 1; the dipole lowers this ratio to 1.5 : 1. This is a direct manifestation of the dipole and monopole being out of phase with one another. That is, at the places where the monopole contributes least to the anisotropies, at its troughs, the dipole contributes the most. One other comment about the relation between the monopole and the dipole: they add incoherently. By incoherently, I mean that the cross term of Θ_l from the monopole multiplied by Θ_l from the dipole vanishes when integrating over all k-modes to get the C_l 's. This can be seen mathematically from the properties of the spherical Bessel function (Exercise 12). Incoherence implies that the dipole is not as important in the power spectrum as one might naively think. If the amplitude of the dipole is 30% of that of the monopole at recombination, the dipole's contribution to the C_l 's is only 10% $(1^2 + 0.3^2 = 1.1)$

The integrated Sachs-Wolfe effect is also important if the potential changes after recombination. To see which scales are affected by the ISW effect, consider the integral in Eq. (8.56). Suppose the potential changes at time η_c, with all subhorizon scales ($k\eta_c > 1$) being affected. The Bessel function peaks at $l \sim k\,(\eta_0 - \eta_c)$; so all angular scales $l > (\eta_0 - \eta_c)\,/\eta_c$ are affected. The largest effect is typically at the horizon.

The best, and most prevalent, example of the ISW effect is that due to residual radiation at recombination. If the universe were purely matter dominated, there would be no such effect. But, the transition to pure matter domination is not abrupt, and even for $a_{\rm eq} \sim 10^{-4}$, an ISW effect occurs right after recombination. This early ISW effect is particularly important because it adds coherently with the monopole. To see this, integrate the last term in Eq. (8.56) by parts. Then, the dominant contribution comes from $\eta \simeq \eta_*$, so the Bessel function can be evaluated there, leaving the trivial integral which gives

$$\Theta_l\,(k,\eta_0)^{\text{early ISW}} = [\Psi\,(k,\eta_0) - \Psi\,(k,\eta_*) - \Phi\,(k,\eta_0) + \Phi\,(k,\eta_*)]\,j_l\,[k\,(\eta_0 - \eta_*)]\,.$$

This adds exactly in phase with the monopole (which is proportional to the same Bessel function) so even though the magnitude of the effect on Θ_l is much smaller than is the dipole, the effect on the anisotropy spectrum is disproportionate. A 30% dipole leads to a 10% shift in the C_l 's, while a 5% ISW effect leads to the same 10% shift in the C_l 's. The bottom panel shows that the large scales, those with $l \sim \eta_0/\eta_*$, get a big boost from this early ISW effect.

12.8.7 Cosmological Parameters

The anisotropy spectrum depends on cosmological parameters. This fundamental realization initially caused great consternation ("We will never be able to measure any one parameter because there is too much degeneracy"). As more quantitative studies were carried out, the pendulum swung to the other side ("We will be able to disentangle the degeneracies and measure cosmological parameters to percent accuracy"). More recently, the community has settled into a state of cautious optimism. Indeed, just a decade after the initial discovery of large scale anisotropies by COBE, there were a host of experiments which together seemed to pin down one parameter (the total energy density) by measuring the location of the first peak. Several of these had measured the subsequent two peaks, allowing an inference of the baryon density, the parameter which most affects the heights and locations of these peaks.

We now have developed the theoretical tools needed to participate in the parameter determination discussion. In this section, we apply these tools to understand how the anisotropy spectrum varies as cosmological parameters vary.

One very important decision that must be made is which parameters will be allowed to vary. I will consider eight parameters:

- Curvature density, $\Omega_k := 1 - \Omega_m - \Omega_\Lambda$
- Normalization, C_{10}
- Primordial tilt, n
- Tensor modes, r (for a precise definition, see Exercise 18)
- Reionization, parametrized by τ back to recombination
- Baryon density, $\Omega_b h^2$
- Matter density, $\Omega_m h^2$
- Cosmological constant energy density, Ω_Λ There are two aspects of this list worth stressing. The first is that obviously it does not include all possible cosmological parameters. Some favorites missing are a neutrino mass (I will set all masses to zero in the following), the equation of state for dark energy w (will be fixed at -1 corresponding to a cosmological constant), and tensor tilt n_T (fixed at zero). The second important point is that I have deliberately chosen very specific combinations of these parameters, e.g., $\Omega_b h^2$, not Ω_b and h separately. While there is good reason for this (e.g., the alternating peaks effect depends on $\Omega_b h^2$), it also is a source of confusion. A common complaint is that, within the context of a flat universe (the first parameter, the curvature density, equal to zero), why should both the cosmological constant and the matter density be allowed to vary? Mustn't their sum equal 1? It is true that $\Omega_m + \Omega_\Lambda$ must equal 1 in a flat universe. But that does not preclude us from varying both $\Omega_m h^2$ and Ω_Λ, since h can change while the sum of the two densities is 1.

To harp on this point, consider two analysts. Analyst A works in the context of a flat universe and uses $\Omega_m h^2$ and Ω_Λ as her two free parameters. Analyst B also assumes the universe is flat, but takes h and Ω_Λ as his two parameters. When A raises Ω_Λ, the matter density ($\Omega_m h^2$) is kept fixed, so the epoch of equality is kept fixed. However, when analyst B raises his Ω_Λ, to keep the universe flat, he must lower Ω_m. He is therefore also lowering the matter density (since h is kept fixed), thereby moving $a_{\rm eq}$ closer to today. That change in $a_{\rm eq}$ will lead to an enhanced ISW effect, and therefore a larger first peak. Analyst A, who had the

foresight to separate out this effect by choosing $\Omega_m h^2$ as one of her parameters, sees no such enhancement. And, indeed the enhancement is caused only indirectly by Ω_Λ: rather it is the direct result of a smaller $\Omega_m h^2$.

Let's now consider the effect of each parameter in turn.

Curvature

If the universe is not flat, then the simple picture of Figure 8.4 is no longer accurate since the geodesics of massless particles are such that photons starting out parallel to each other slowly diverge. Consider the implication of this divergence for anisotropies. Suppose the identical pattern of inhomogeneities was in place at recombination in both a flat and open universe. As shown in Figure 8.15, the physical scale with maximal anisotropy (the first peak) gets projected onto a much smaller angular scale in an open universe. The peaks should therefore be shifted to higher l. As shown in Figure 8.16, this is precisely what happens.

The magnitude of this effect is determined by the comoving angular diameter distance to the last scattering surface, in a flat universe simply equal to $\eta_0 - \eta_*$, and in a universe with curvature given by Eq. (2.46) out to z_*. Figure 8.17 shows this distance as a function of the curvature density with all other parameters held fixed. The angular diameter distance scales as $(1 - \Omega_k)^{-0.45}$, so that it is a factor of 1.7 larger in an open $\Omega_k = 0.7$ universe than in a flat universe. Notice from Figure 8.16 that this is precisely the factor by which the first peak shifts from one model to

Freestreaming in on open unlyerse

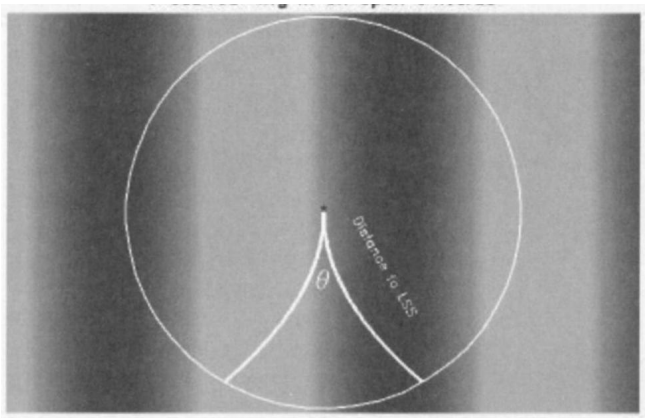


Figure 8.15. Photon trajectories in an open universe diverge. Perturbations at last scattering turn up on smaller scales in an open universe than they do in a flat universe.

the other. Of all the parameters under consideration, curvature by far causes the largest shift in the location of the peaks.

Figure 8.16 also shows data circa 2002. There is a clear rise up to a first peak at $l \sim 200$ and an equally clear fall past this first peak. When the data first started coming in (around 1998), a skeptic could plausibly claim that no one data set spanned the whole peak, and it is difficult to combine data sets. Within a year or two, though, this objection vanished as larger data sets such as TOCO (Miller et al., 1999), Boomerang (de Bernardis et al., 2000), and Maxima (Hanany et al., 2000) all contained enough information by themselves to rule out an open universe. The DASI detection (Halverson et al., 2002), together with the reanalyzed Boomerang and Maxima data, cemented the case for a flat universe.

Of course, a truly flat universe is only one point in parameter space, the point at which the sum of the energy densities exactly equals the critical density, and no data will ever rule out all values except for this one point. Rather, the data now suggest that the total density is equal to the critical density with an error of about 5%. The classic open universe once favored by astronomers had 30% of the critical density, and so is ruled out with very high confidence.

Degenerate Parameters

Figure 8.18 shows the results of varying four parameters. Before considering each in turn, it is important to state the obvious. All of these parameters change the spectrum in very similar ways. The shape of the spectrum varies hardly at all; rather, these parameters simply move the spectrum up and down.

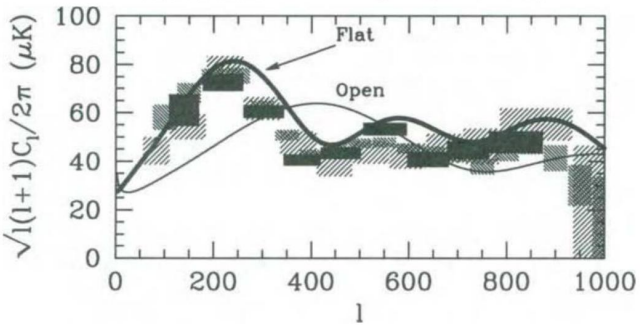


Figure 8.16. The anisotropy spectrum in flat versus open universe. Also shown are data from three small-scale experiments: DASI (darkest; Halverson et al., 2002), Boomerang (medium; Netterfield et al., 2002), and Maxima (lightest; Lee et al., 2001). The pattern of peaks and troughs persists in the open universe but is shifted to smaller scales. The data clearly favor the flat case. Both curves have identical parameters $n = 1, \Omega_m h^2 = 0.15, \Omega_b h^2 = 0.02$ with no reionization, tensors, or cosmological constant. Open curve has $\Omega_k = 1 - \Omega_m = 0.7$; flat has the same parameters except $\Omega_k = 0$.

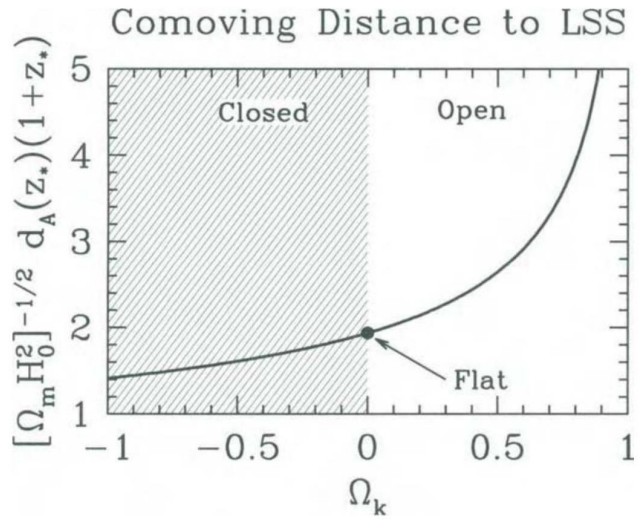


Figure 8.17. Comoving angular diameter distance back to the last scattering surface at $z_* \simeq 1100$ as a function of curvature. The distance is larger in an open universe than in a closed universe.

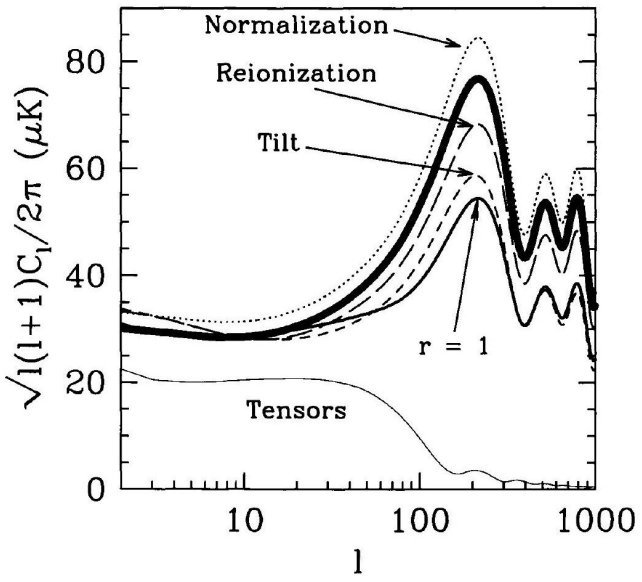


Figure 8.18. Changes in the anisotropy spectrum as C_{10}, τ, r, and n vary. The base model (thick curve) is a flat universe with no reionization or tensors, $n = 1, \Omega_m h^2 = 0.16$, $\Omega_b h^2 = 0.021$, and $\Omega_\Lambda = 0.7$. The thin curves vary one parameter each. Reionization corresponds to letting the optical depth back to the last scattering surface equal 0.2 instead of zero; tilt has a primordial spectum with $n = 0.8; r = 1$ has an equal contribution of scalars and tensors to the quadrupole; and normalization has $C_{10}10\%$ higher than the base model. The curve labeled tensors is the contribution to the anisotropy from tensors only. Only the $r = 1$ curve includes this contribution; all others assume no anisotropy from tensors.

Normalization. The parameter C_{10} trivially moves the spectrum up or down. Note that, of the four parameters varied in Figure 8.18, it is the only one which can raise the amplitude of the spectrum.

Tilt. We have already considered the large-angle effects of a tilted ($n \neq 1$) spectrum. If $n < 1$, then the small-scale anisotropies are smaller than in the $n = 1$ model. Figure 8.18 shows that, as smaller and smaller scales are probed, the effect becomes more pronounced. So of the four parameters considered here, tilt has the most distinctive shape - it is not a simple up-down shift - and perhaps will be most easily extracted. Quantitatively, the spectrum scales as

$$\frac{C_l(n)}{C_l(n=1)} \simeq \left(\frac{l}{l_{\rm pivot}}\right)^{n-1}$$

where here $l_{\rm pivot} = 10$ since we are fixing C_{10}. Accounting for the fact that $\sqrt{C_l}$ is plotted in Figure 8.18, we see from the point at $l = 1000$ that this scaling works extremely well.

Reionization. The universe was almost certainly reionized at late times. We see this in the absorption spectra of high-redshift quasars, where no evidence is seen of a uniform background of neutral hydrogen until we go back as least as far as $z \sim 6$ (Becker et al., 2001; Fan et al., 2002). Reionization brings the CMB back in contact with electrons. If enough scattering takes place, that is, if the optical depth back to the last scattering surface is high enough, isotropy is restored; equivalently, primordial anisotropies are washed out.

There are several ways to see the effect of reionization quantitatively. One is to imagine a photon traveling in our direction with temperature $T[1 + \Theta]$, where T is the background temperature and Θ is the perturbation for which we have solved. If these photons hit a region with optical depth τ, only a fraction $e^{-\tau}$ will escape and continue on their way to us. In addition to these, we will also get a fraction $1 - e^{-\tau}$ from the ionized region. All of these have the equilibrated temperature, T. So the temperature we see today is

$$T[1+\Theta]e^{-\tau} + T\left(1-e^{-\tau}\right) = T\left[1+\Theta e^{-\tau}\right]$$

Subtracting from this the mean temperature T tells us that the fractional anisotropy will be Θ, the primordial one set up at $z \simeq 1100$, multiplied by $e^{-\tau}$. Of course this argument can affect only those scales within the horizon at the time of reionization,

so multipoles l larger than $\eta_0/\eta_{\rm reion}$ will be suppressed by $e^{-\tau}$; small l will be

unaffected. This is seen in Figure 8.18, where the reionization curve falls on top of the base model on large scales but is uniformly suppressed on small scales.

Tensors. We saw in Chapter 5 that once they enter the horizon, the amplitude of gravitational waves dies away. Therefore, gravity waves affect the anisotropy spectrum only on scales larger than the horizon at recombination. Typically, this translates into angular scales $l < 100$. Indeed the tensors curve in Figure 8.18 shows that tensors die out after $l > 100$. We can observe only the sum of anisotropies due to tensors and scalars. So if tensor perturbations were produced during inflation, and if the total (scalar plus tensor) anisotropy spectrum is fit to the large-scale (COBE) data, then the small-scale scalar amplitude is smaller than it would otherwise be. Therefore, on scales $l > 100$ where only scalars remain, the anisotropy spectrum is identical to the base model in Figure 8.18, but with a lower amplitude.

Distinct Imprints

The final variations we will consider are changes in the baryon density $\Omega_b h^2$, the matter density $\Omega_m h^2$, and the cosmological constant. As can be seen from Figure 8.19, these changes lead to richer variations in the anisotropy spectrum; as such they are somewhat harder to understand (but easier to extract from the data!) than the parameters in the previous subsection.

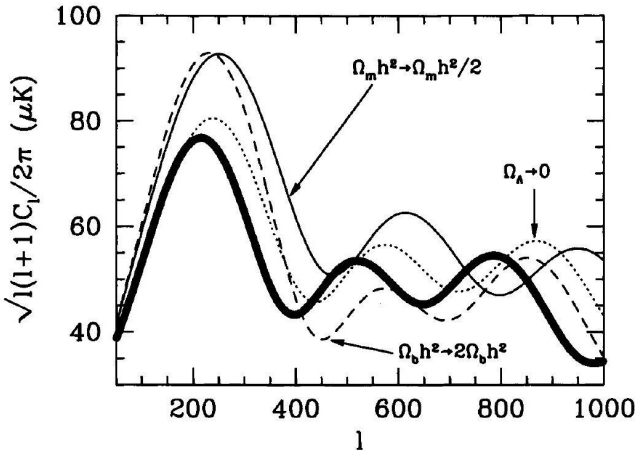


Figure 8.19. Changes in the anisotropy spectrum as baryon density, matter density, and cosmological constant vary. Same base model as Figure 8.18.

Each of these parameters induces a small shift in the locations of the peaks and troughs in the spectrum. To understand these shifts, it is important to recall that since inhomogenities on scales k show up at $l = k\eta_0$ in a flat universe, the peaks in a flat universe will show up at $l_p \simeq k_{\rm p}\eta_0 \simeq n\pi\eta_0/r_s\,(\eta_*)$ (Eq. (8.25), but also see the discussion on page 247 that argues that the actual value of l_p is $\sim 25\%$ lower). Figure 8.20 shows this ratio as a function of matter and baryon density. It is more sensitive to the matter density, so the peak spacing increases as the matter density goes down. But there is also a little sensitivity to the baryon density. With the densities fixed, introducing a cosmological constant does not change the sound horizon, but it does slightly affect η_0, so the peaks shift in that cases as well.

Baryon density. In addition to the lateral shift in the spectrum due to the change in the sound horizon, changes in the baryon density affect the heights of the peaks as well. We have

already touched on the ways in which the anisotropy spectrum depends on the baryon density. The foremost, clearly visible in Figure 8.19, is that odd peaks (first and third in the figure) are higher than the even peaks when the baryon density is large. This is a direct ramification of the lower frequency of oscillations due to the massive baryons. This change is virtually unique, making the baryon density one of the easiest parameters to extract from the CMB. Observations as of 2001 (e.g., Pryke et al., 2001) pin down $\Omega_b h^2 = 0.022 \pm 0.04$, and this constraint will undoubtedly get tighter with data from the Map and Planck satellites. The second change due to $\Omega_b h^2$ is that an increased baryon density reduces the diffusion length. Therefore, a larger baryon density means damping moves to smaller angular scales, so the anisotropy spectrum on scales $l > 1000$ is larger in a high- $\Omega_b h^2$ model.

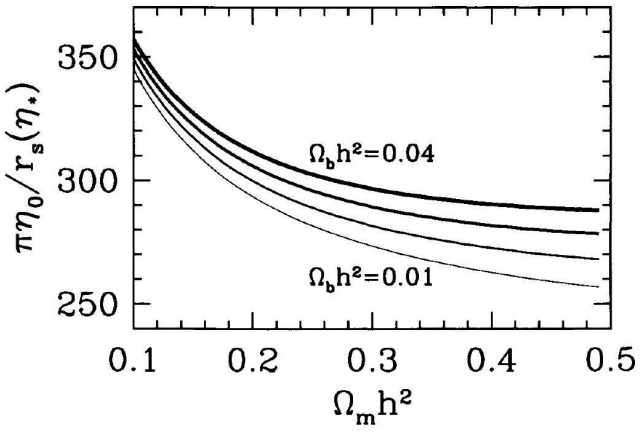


Figure 8.20. The inverse sound horizon at recombination. In a flat universe, the spacing between the acoustic peaks in the CMB is equal to $\pi\eta_0/r_s(\eta_*)$.

Cosmological constant. The cosmological constant is a late-time phenomenon. It was not around at recombination, and therefore could not have affected perturbations then. Therefore, the only possible effects of a cosmological constant are on freestreaming and on the largest angular scales just entering the horizon at recent times. The change due to free-streaming is evident in Figure 8.19. The spectrum is shifted to smaller angular scales if there is no cosmological constant. You will show in Exercise 13 that this small shift can be readily explained by comparing the conformal times in a Λ universe and a matter-dominated universe. Figure 8.19 also shows that the anisotropy spectrum is slightly lower on small scales in a Λ universe. This is a direct result of the large-angle normalization. In a Λ universe, there is a late-time ISW effect, which enhances the anisotropies on large angles. If we normalize on these scales, then the small-scale anisotropy gets correspondingly smaller.

Matter density. If the matter density is low, the epoch of equality occurred closer to recombination, so that the radiation density must be accounted for in computing the inhomogeneities at recombination. In particular, the decaying potential due to the inability of the radiation to cluster provides a strong driving force for the oscillations. Therefore, $\Theta_0(\eta_*)$ is larger than in a purely matter-dominated universe. Further, after recombination, since the potential is not constant, the ISW effect also contributes significantly to the final anisotropy spectrum. Therefore, the small-scale anisotropies increase if the matter density is low. This effect too has apparently been detected, with measurements (Pryke et al., 2001) implying $\Omega_m h^2 = 0.16 \pm 0.04$.

SUGGESTED READING

The large-scale Sachs-Wolfe effect was first predicted by Sachs and Wolfe (1967), just several years after the discovery of the CMB. Several groups initiated the study of anisotropies in the tightly coupled limit: Doroshkevich, Zel'dovich, and Sunyaev (1978), Atrio-Barandela and Doroshkevich (1994), and Jorgenson et al. (1995). The approach was perfected by Seljak (1994)

and Hu and Sugiyama (1995), the latter of which is the basis for the semianalytic treatment of this chapter. Again, CMBFAST described in Seljak and Zaldarriaga (1996) is a crucial tool for fast, accurate numerical work. Diffusion damping is sometimes called Silk damping because of the Silk (1968) paper recognizing its importance. Two other papers of interest are Zaldarriaga and Harari (1995) which discusses the effect of polarization on the damping scale (see Exercise 7) and Hu and White (1997a) which, among other things, gives fits to the damping scale valid for a wide range of parameters.

The question of how the anisotropy spectrum depends on cosmological parameters has been explored in literally hundreds of papers over the past decade. I remember Dick Bond, one of the pioneers in the field, giving a talk in 1992 at a conference about the early COBE data waving his hands through an invisible multidimensional parameter space, explaining that our goal now was to navigate through this space. Among the most important realizations were the dependence on curvature (Kamionkowski, Spergel, and Sugiyama 1994), the degeneracy of the height of the first peak (Bond et al., 1994), and breaking of this degeneracy by smaller scale information (Jungman et al., 1995). More recently, Hu et al. (2000) is a good reference.

I have given short shrift (or no shrift) to some important parameters. The effect of dark energy on the CMB has now been well studied: first by Coble, Dodelson, and Frieman (1997) and then more generally by Caldwell, Dave, and Steinhardt (1998). Massive neutrinos affect the anisotropy spectrum at the $5-10\%$ level (Ma and Bertschinger, 1995 and Dodelson, Gates and Stebbins, 1996). The anisotropies due to tensors became a hot topic after the COBE discovery. For a semianalytic treatment and references to the dozens of papers relating the tensor anisotropy to parameters in the potential, see Turner, White and Lidsey (1993). Although the effect of reionization on the primary anisotropies generated before recombination is well understood, a hot topic now is secondary anisotropies, those generated after reionization. These will likely be probed by the next generation of experiments.

3K: The Cosmic Microwave Background (Partridge) is a good introduction to some of the experimental issues I have neglected in this book. The COBE discovery paper is Smoot et al. (1992) with the 4-year observations presented in Bennett et al. (1996). There were many good analyses papers written on the COBE data; I've relied on Bunn and White (1997), which is especially good for using COBE to normalize the matter power spectrum, and Tegmark (1997), from which the points in Figure 8.12 are taken. The two satellite experiments are Map (http://map.gsfc.nasa.gov) and Planck (http://sci.esa.int/planck/). Map was launched in 2001, and Planck is scheduled to be launched in 2007.

12.9 Cosmological perturbations by J. Lesgorgues

Abstract

We present a self-contained summary of the theory of linear cosmological perturbations. We emphasize the effect of the six parameters of the minimal cosmological model, first, on the spectrum of Cosmic Microwave Background temperature anisotropies, and second, on the linear matter power spectrum. We briefly review at the end the possible impact of a few non-minimal dark matter and dark energy models.

Purposes

Cosmology is progressing by leaps and bounds thanks to a spectacular amount of observational data. It provides crucial clues for particle physics, and more generally for high energy physics. A key role is played by Cosmic Microwave Background (CMB) and Large Scale Structure (LSS) data. Their interpretation relies on the *theory of cosmological perturbations*.

The aim of this course is to introduce this theory, staying at the linear level, and hopefully with a simple, self-contained and original approach. It is targeted at PhD students in

astrophysics and cosmology, as well as researchers in particle physics willing to follow current developments in cosmology. We discuss in particular the effect of the parameters of the minimal cosmological model on the spectrum CMB temperature anisotropies, and on the matter power spectrum. We briefly review at the end the possible impact of non-minimal dark matter and dark energy models.

These notes are based on a cycle of five lectures of 90 minutes. Since this is not much for introducing linear cosmological perturbation theory, the lectures remained at a qualitative level, with more emphasis on physical ideas than on the algebra. These written notes contain roughly the same amount of information as the lectures. After reading these few pages, the reader may wish to improve his/her level of understanding, while sticking to the same structure and same notations. For this purpose, we suggest to read chapters 2, 5 and 6 of Ref. [1]: they contain a similar presentation, still simple and qualitative, but more detailed. Optionally, the reader can also grab there a description of neutrino effects, which are completely neglected here.

For a quantitative description of perturbation theory, gauge transformations, CMB physics and linear or non-linear structure formation, the reader should refer to the specialized literature. In the bibliography, we list a very incomplete selection of outstanding references.

Sections 12.9.2 and 12.9.3 of these notes are illustrated with several figures that represent the qualitative evolution of cosmological perturbations. It would have been easy to produce them with a Boltzmann code, as we did in Ref. [1]. We deliberately chose instead to include only scanned hand-drawings. The first reason is that drawings offer the possibility to exaggerate some effects for pedagogical purposes: they are often more illuminating than exact numerical plots. The second reason is that they better render the conviviality of the blackboard presentation at TASI.

12.9.1 Linear cosmological perturbations

Classification

We decompose the metric and stress-energy tensor of the universe into spatial averages and linear perturbations,

$$g_{\mu\nu}(t,\vec{x}) = \bar{g}_{\mu\nu}(t) + \delta g_{\mu\nu}(t,\vec{x}) \ , \tag{181}$$

$$T_{\mu\nu}(t,\vec{x}) = \bar{T}_{\mu\nu}(t) + \delta T_{\mu\nu}(t,\vec{x}) \ , \tag{182}$$

where $\bar{g}_{\mu\nu}$ stands for the metric of the homogeneous and isotropic Friedmann–Lemaître (FL) model. Being symmetric, the two perturbed tensors contain ten degrees of freedom each, describing different aspects of gravity. Bardeen showed in 1980 that they can be decomposed on the basis of scalars, vectors and tensors under spatial rotations (spatial rotations play a special role because they leave the FL background invariant). These three sectors are decoupled at first order in perturbation theory.

In the vacuum, scalar and vector perturbations vanish, while tensor perturbations can propagate if they have been excited: they account for gravitational waves, the only "real" (propagating) gravitational degrees of freedom in General Relativity (GR). In the presence of matter, scalars represent the response of the metric to an irrotational distribution of matter, and generalize Newton's theory of gravitation. Vectors represent the response of the metric to vorticity, and describe phenomena with no equivalent in Newton's theory, called "gravito-magnetism".

In minimal cosmological models, the vorticity of the various matter components decays with time, and vectors can be neglected. Tensors may play a small role in CMB anisotropies, that we will mention briefly in Sec. 12.9.2. They can be studied separately, since they decouple from the scalar sector at first order in perturbations. Hence this course will be essentially focused on scalar perturbations.

The four scalar components of both the metric and stress-energy perturbed tensors are contained in:

1. the (00) term,
2. the trace of the (ij) matrix,
3. the irrotational part of the $(0i)$ vector,
4. the traceless longitudinal part of the (ij) tensor.

For the perturbed metric $\delta g_{\mu\nu}$, these components correspond (in the same order) to:

1. the generalized gravitational potential ψ,
2. the local distortion ϕ of the average scale factor $a(t)$: the "local scale factor" is given by $(1-\phi)a$,
3. the potential b such that $\delta g_{0i} = \partial_i b$,
4. the potential μ of the metric shear: $\delta g_{ij} = (\partial_i\partial_j - \frac{1}{3}\delta_{ij}\Delta)\mu$.

For the perturbed stress-energy tensor $\delta T_{\mu\nu}$, these components (still in the same order) represent:

1. the energy density perturbations $\delta\rho$ (we will usually refer to the relative perturbations $\delta \equiv \delta\rho/\bar{\rho}$),
2. the pressure perturbations δp,
3. the potential v of the irrotational component of the flux of energy, $\delta T^0_i = \partial_i v$ (v is sometimes called the velocity potential, since in the case of a fluid it is related to the bulk velocity),
4. the potential s of the shear stress or anisotropic stress: $\delta T^i_j = (\partial_i\partial_j - \frac{1}{3}\delta_{ij}\Delta)s$.

It is equivalent to use as a variable the "velocity potential" v or the "velocity divergence" θ defined as

$$\partial_i \delta T^0_i \equiv (\bar{\rho} + \bar{p})\theta = \Delta v \ . \tag{183}$$

Similarly, we can use the function σ instead of s, with the definition

$$(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\Delta)\delta T^i_j \equiv (\bar{\rho} + \bar{p})\sigma = \Delta(\Delta s) \ . \tag{184}$$

The function σ is usually called the anisotropic stress, although the true anisotropic stress is the component of δT^i_j deriving from the potential s. The factors $(\bar{\rho} + \bar{p})$ in the two previous equations are introduced in the definitions in order to obtain simple equations (some authors use alternative notations without these factors or with different ones). To summarize, we see that we can manipulate four degrees of freedom representing the scalar perturbations of matter fields, that can be chosen to be the density fluctuation, pressure perturbation, velocity divergence and anisotropic stress: $\{\delta, \delta p, \theta, \sigma\}$.

Gauges

In an idealised FL universe, there is only one time slicing compatible with the assumption of homogeneity. Instead, in a perturbed universe, there is an infinity of time slicings compatible with perturbation theory (i.e. such that on each slice, all quantities remain close to their average value).

The perturbation of any quantity in a given point is the difference between the true and the average quantity in this point. For instance, for the total energy density ρ,

$$\delta\rho(t,\vec{x}) = \rho(t,\vec{x}) - \bar{\rho}(t) \ . \tag{185}$$

While $\rho(t,\vec{x})$ is a locally, unambiguously defined quantity, $\bar{\rho}(t)$ depends on the choice of equal-time hypersurface going through the point $(t,\vec{x})$. With a different choice, $\rho(t,\vec{x})$ would be compared to the average performed on a different sheet, that would take a different value. Hence $\delta\rho(t,\vec{x})$ also depends on the choice of time slicing.

A gauge is a choice of time slicing. Gauge transformations are induced by coordinate transformations $x_\mu \mapsto x_\mu + \epsilon_\mu$ mapping the points of one time slicing to those of another time slicing. All coordinate transformations do not induce a valid gauge transformation: the condition that perturbations must still be linear after the transformation restricts ϵ_μ to be small in every point.

A naive study of the equations of motion of perturbed quantities would be plagued by the freedom to change the gauge without changing physical results: some solutions of the full equations would be "gauge modes" with no observable consequences. To deal with this issue, one can adopt one of two point of views:

- one can derive gauge-invariant quantities (i.e., non-trivial integro-differential combinations of the metric and stress-energy tensor components left invariant by a gauge transformation), and gauge-invariant equations of motions for these quantities. Note that there are four scalar degrees of freedom in $\delta g_{\mu\nu}$ and two scalar degrees of freedom in the four-vector field ϵ_μ inducing gauge transformations: namely, ϵ_0 and the potential e such that $\epsilon_i = \partial_i e$. Hence we can use gauge transformations to cancel two scalar degrees of freedom, and build up two independent gauge-invariant scalars. One way to define them is through the two Bardeen potentials $\Phi_{\rm A}$ and Φ_H, defined by Bardeen (1980) as two integro-differential combinations of ψ, ϕ, b, μ.

- one can fix the gauge, i.e. introduce a condition such that the time slicing is unique. Then the number of independent solutions to the equations will be the same as in the gauge-invariant formalism. In any case, one can show that truly observable quantities are always independent of the gauge. Obtaining them after solving equations in the gauge-invariant formalism or in one particular gauge should not make any difference in practice.

A convenient gauge choice for a pedagogical introduction to CMB and matter scalar perturbations is the so-called Newtonian gauge or longitudinal gauge, in which one imposes that non-diagonal scalar perturbations of the metric vanish: $b = \mu = 0$. This prescription can be showed to fix a unique time slicing. In this gauge, adopting units such that $c = 1$ and using proper time t, the line element reads:

$$ds^2 = -(1+2\psi)dt^2 + (1-2\phi)a^2 d\vec{l}^{\,2} \ , \tag{186}$$

where $d\vec{l}^{\,2}$ stands for the cartesian measure $(dx^2 + dy^2 + dz^2)$ for a flat FL model, or for $(1 \pm kr^2)^{-1}dr^2 + r^2(\sin\theta^2 d\theta^2 + d\varphi^2)$ for an open/closed FL model in spherical coordinates. We are still free to redefine time (by definition, a time redefinition leaves the time slicing invariant). Lots of results in cosmological perturbation theory look simpler when using conformal time η,

defined up to a constant by $d\eta = dt/a$. In this course we fix the constant in such way that $\eta \longrightarrow 0$ at the vicinity of the initial singularity, when $\rho \longrightarrow \infty$ (this prescription would not work if we were studying cosmological inflation, but in this course, we are not). Conformal time is convenient because photons traveling in a flat unperturbed FL universe along geodesics crossing the origin of the system of coordinates obey to $dr = d\eta$ (this comes from $a\,dl = dt$, i.e. from $ds = 0$ with the restriction $d\theta = d\phi = 0$). Hence conformal time is a measure of time based on the comoving distance travelled by a photon[117], and the comoving distance to a given object coincides with its "look-back conformal time". In this course we will use dots for derivatives with respect to proper time and primes for derivatives with respect to conformal time. The Hubble parameter (or expansion rate parameter) reads

$$H = \frac{\dot{a}}{a} = \frac{a'}{a^2} \ , \tag{187}$$

the Hubble radius is $R_H = 1/H$, and the condition that a Fourier mode of physical wavelength λ crosses the Hubble radius is

$$\lambda = R_H \quad \Leftrightarrow \quad \frac{2\pi}{k}a = \frac{1}{H} \quad \Leftrightarrow \quad k \sim aH = \frac{a'}{a} \ . \tag{188}$$

One advantage of the Newtonian gauge is that the gauge-invariant Bardeen potentials $\{\Phi_{\rm A}, \Phi_H\}$ reduce in this gauge to the metric perturbations $\{\phi, \psi\}$: the evolution of the Newtonian metric perturbations informs us directly on that of two gauge-invariant quantities. Other interesting properties of $\{\phi, \psi\}$ appear when writing the Einstein equations in the Newtonian gauge. The full Einstein equations linearized at first order in perturbations feature four equations relating scalar degrees of freedom. One of them, associated to the traceless longitudinal part of $\delta G^i_j = 8\pi G T^i_j$, gives (assuming a flat FL background)

$$\frac{2}{3}(k/a)^2(\phi - \psi) = 8\pi G \sum_x (\bar{\rho}_x + \bar{p}_x)\sigma_x \ , \tag{189}$$

where the index x runs over all the species contributing to the total stress-energy tensor. This means that when the universe contains only shearless components with $\sigma_x = 0$ (as would be the case in the presence of perfect fluids), the two metric perturbations are equal. Next, the Einstein equation $\delta G^0_0 = 8\pi G T^0_0$ gives (still assuming a flat FL background)

$$2a^{-2}\left[k^2\phi + 3\frac{a'}{a}\left(\phi' + \frac{a'}{a}\psi\right)\right] = -8\pi G \sum_x \bar{\rho}_x \delta_x \ . \tag{190}$$

The term on the right-hand side involves the total energy perturbation $\delta\rho_{\rm tot} = \sum_x \bar{\rho}_x \delta_x$. In the short scale (more precisely, sub-Hubble) limit, the term containing k^2 dominates the other terms in the square brackets, and we recover the Poisson equation

$$-\frac{k^2}{a^2}\phi = 4\pi G\, \delta\rho_{\rm tot} \ , \tag{191}$$

where the factor $-k^2/a^2$ represents the Fourier transform of the physical Laplace operator in an expanding universe. Note that in the Poisson equation one may have expected to see the generalized gravitational potential ψ instead of ϕ: however, in the sub-Hubble limit, the shear of individual components is usually either null or negligible, so that $\phi = \psi$.

[117] In a universe with non-zero spatial curvature, this remains true, provided that the comoving distance is defined not like r, but like $\chi \equiv \int (1 \pm kr^2)^{-1/2} dr$.

Equations of motion

In the minimal cosmological scenario, the universe features several species with spatial fluctuations, described with different equations because of their distinct properties: cold dark matter (CDM) is non-relativistic and collisionless, neutrinos are ultra-relativistic and collisionless at the times of interest, baryons are non-relativistic and smoothly interpolating from a strongly coupled to decoupled regime, and finally photons are ultra-relativistic and interpolating between the same two regimes.

The equation of conservation of the total stress-energy tensor, $D_\mu T^\mu_\nu = 0$ (deriving from Bianchi identities), yields two scalar and two vector equations. The scalar ones are the conservation of energy equation and the Euler equation.

For an single component experiencing no interaction with other species (other than gravitational), the equation $D_\mu T^\mu_\nu = 0$ applies to the individual stress-energy tensor: it gives one continuity and one Euler equation for that component. The evolution of single component experiencing interactions is also given by the continuity and Euler equation, but with an extra source term accounting for stress-energy injection/leak caused by the interaction.

We have seen that the perturbations of each component x can be described by four variables $\{\delta_x, \delta p_x, \theta_x, \sigma_x\}$. Hence, in general, two equations of motion are not sufficient for closing the system. However:

- for a perfect fluid, microscopic interactions impose local thermodynamical equilibrium. The pressure is then isotropic[118], with $\sigma_x = 0$. In addition, pressure perturbations obey to $\delta p_x = c_{\rm a}^2 \delta\rho_x$, where $c_{\rm a}$ is the adiabatic sound speed inferred from the equation of state of the fluid. If σ_x vanishes and δp_x is a function of $\delta\rho_x$, perturbations in the fluid are described by only two independent functions $\delta_x \equiv \delta\rho_x/\bar{\rho}_x$ and θ_x. If the collision term also vanishes or is specified, the two equations of motion (continuity and Euler) are sufficient for closing the system and computing the evolution of perturbations.

- for a decoupled or weakly interacting species, there are no such simplifications concerning the anisotropic stress and pressure perturbation. Hence, in general, the two equations inferred from stress-energy conservation are not sufficient. For such species, one has to use the more general Boltzmann equation, giving the evolution of each phase-space distribution:
$$\frac{d}{d\eta} f_x = \sum_y C_{xy}[f_x, f_y] \ , \tag{192}$$
where the sum holds over the species y coupled with x. Each phase-space distribution can be decomposed into a background and perturbation part:
$$f_x(\eta, \vec{x}, \vec{p}) = \bar{f}_x(\eta, |\vec{p}|) + \delta f_x(\eta, \vec{x}, \vec{p}) \ , \tag{193}$$
where $\vec{p}$ stands for momentum, $|\vec{p}|$ for its modulus, and the background part does not depend on the direction of $\vec{p}$ by assumption of isotropy.

- fully decoupled CDM is a particular case of a collisionless species with negligible velocity dispersion (the word "cold" refers precisely to this last assumption). Hence, it behaves in the same way as a pressureless perfect fluid, although in reality it has no interactions and should not be called a fluid. Since the velocity dispersion is negligible, in a given

[118] An anisotropic shear stress $\sigma \neq 0$ reflects the fact that in each given point, particles travel with different velocities (due to some intrinsic velocity dispersion and/or a superposition of several flows in phase space), leading to anisotropic pressure. This contradicts the assumption of a perfect fluid, in which local interactions result in a unique bulk velocity (after coarse-graining over microscopic scales), and erase anisotropic pressure.

point, all particles share the same velocity, imposed by gravitational flows (while for non-cold collisionless species, the velocity would get two contributions, one from gravitational flows, and one from the phase-space distribution function). Hence the anisotropic pressure vanishes (see the previous footnote). The pressure perturbation δp_x is also related to the velocity dispersion in a given point, anf can be neglected with respect to the density perturbation in the CDM case. Hence CDM is formally equivalent to a perfect fluid with no anisotropic stress and no pressure perturbation (or in other words, with a sound speed $c_\mathrm{s} = 0$). In that case, the two equations of motion inferred from $D_\mu T^\mu_\nu = 0$ are sufficient, like for a fluid.

Gravitational interactions between species are accounted by the presence of metric perturbations in each equation of motion (more specifically, terms in $k^2\psi$ accounting for gravitational forces, and terms in ϕ' accounting for dilation effects, i.e. for local distortions of the scale factor with respect to a). Hence, in order to close the full system of equations, we still need two independent relations, to be chosen among the four scalar Einstein equations: they provide the value of $k^2\psi$ and ϕ' at each time, as a function of all matter fields.

Initial conditions

We wish to study the evolution of matter perturbations, starting from some early time at which all Fourier modes of interest (those which are observable in the CMB spectrum and in the matter power spectrum on linear or mildly non-scales) are still outside the Hubble radius. Indeed, super-Hubble modes experience a trivial evolution, unaffected by small scale interactions (Thomson or Coulomb scattering, usual gravitational force $\vec{\nabla}\phi$, etc.) Hence, the perturbations evaluated at some arbitrary time but on super-Hubble scales reflect directly the mechanism responsible for the formation of perturbations in the very early universe. In the standard cosmological model, these initial conditions can be inferred from inflation.

Typically, a good time for setting initial conditions is when the redshift $z = a_0/a - 1$ is of the order of 10^5: at this time, all comoving scales that are observable in the CMB and linear matter power spectrum still verify $k \ll aH$.

It is crucial to understand that, as long as the background cosmology is assumed to be of the FL type, the perturbed stress-energy momentum tensor δT^ν_μ must be diagonal on super-Hubble scales. Indeed, the background tensor $\bar{T}^\nu_\mu$ is diagonal, and of the form: $\mathrm{diag}(-\bar{\rho}, \bar{p}, \bar{p}, \bar{p})$. This can be showed to be the most general assumption compatible with homogeneity and isotropy. Let us assume that we Taylor-expand δT^ν_μ in powers of the variable $(k\eta)$. For any power-law scale factor, aH is given by $1/\eta$ times a factor of order one. Hence the limit $k\eta \ll 1$ represents precisely the super-Hubble limit. In the Taylor expansion, the zero mode should share the same properties as the background solution, and be diagonal. Higher order terms account for contributions to δT^ν_μ growing with time, and possibly becoming important around the time of Hubble crossing.

The total scalar perturbations $\delta\rho$ and δp are the only one preserving the diagonal form of δT^ν_μ: we conclude that they are the only one that do not vanish at order zero in the $(k\eta)$ expansion. Using stress-energy conservation equations, one can show that the part of δT^ν_μ associated with the velocity divergence is of order one in $(k\eta)$, while the part associated to the total anisotropic stress is of order two.

Suppose that the universe contains initially N uncoupled perfect fluids[119], with N known sound speeds $c_x^2 = \delta p_x/\delta\rho_x$. There are $2N$ independent initial conditions, corresponding to

[119]The discussion presented in this section could be generalized to N coupled species, not all of them being perfect fluids: the conclusions would not change qualitatively, and we restrict here to N uncoupled fluid for simplicity.

possible initial values of each δ_x and each δ'_x. Importantly, in the $2N$-dimensional basis of IC's, one basis vector is very special, as we shall see below.

Before studying perturbations, one should have solved for the background evolution: all background quantities should be known, including for instance the density $\bar{\rho}_x(\eta)$ and pressure $\bar{p}_x(\eta)$ of each species x. Now, let us assume that the real universe is perturbed initially by a single degree of freedom (one may say, by a single initial time shifting function). This is the case in single-field inflationary cosmology: during inflation, there is a single clock (the inflaton), and perturbations arise from a single time shifting function (the inflaton perturbation).

As long as we are dealing with super-Hubble modes, we can neglect microscopic interactions and say that the evolution in each point (in fact, in each Hubble patch) is still given by homogeneous cosmology, taking this shift function $\delta\eta(\vec{x})$ into account:

$$\begin{aligned} \forall x, \qquad \rho_x(\eta,\vec{x}) &= \bar{\rho}_x\left(\eta+\delta\eta(\vec{x})\right) = \bar{\rho}_x(\eta) + \bar{\rho}'_x(\eta)\,\delta\eta(\vec{x}) \\ p_x(\eta,\vec{x}) &= \bar{p}_x\left(\eta+\delta\eta(\vec{x})\right) = \bar{p}_x(\eta) + \bar{p}'_x(\eta)\,\delta\eta(\vec{x}) \end{aligned} \tag{194}$$

where in the last equalities, terms of order two or higher in $\delta\eta$ have been neglected. The above ansatz restricts a lot the choice of possible initial conditions. Indeed, for each uncoupled species, the energy conservation law for the background gives

$$\bar{\rho}'_x = -3\frac{a'}{a}(\bar{\rho}_x + \bar{p}_x) \ . \tag{195}$$

Combining this equation with the above ansatz, we get

$$\forall x, \qquad \frac{\delta\rho_x}{\bar{\rho}_x+\bar{p}_x} = -3\frac{a'}{a}\frac{\delta\rho_x}{\bar{\rho}'_x} = -3\frac{a'}{a}\,\delta\eta(\vec{x}) \ . \tag{196}$$

Since the last term is independent of the index x, we see that for each pair of fluids,

$$\forall x, y, \qquad \frac{\delta\rho_x}{\bar{\rho}_x+\bar{p}_x} = \frac{\delta\rho_y}{\bar{\rho}_y+\bar{p}_y} \ . \tag{197}$$

Note that this famous relation is a consequence of Eqs. (194). It shows that in presence of such initial conditions, everything is fixed up to a single function of $\vec{x}$. Let us take the example of a universe containing only photons, baryons, cold dark matter and neutrinos. We can use the fact that for non-relativistic species $\bar{p}_x \ll \bar{\rho}_x$, while for ultra-relativistic ones $\bar{p}_x = \bar{\rho}_x/3$. Hence, if one function is known — for instance, $\delta_\gamma(\vec{x})$ at initial time — the others can be derived from

$$\delta_{\rm b} = \delta_{\rm cdm} = \frac{3}{4}\delta_\nu = \frac{3}{4}\delta_\gamma \ . \tag{198}$$

It is a simple exercise to prove that the ansatz of Eqs. (194) implies additional important relations:

- first,

$$\forall x, \qquad \delta p_x(\eta,\vec{x}) = \frac{\bar{p}'_x(\eta)}{\bar{\rho}'_x(\eta)}\,\delta\rho_x(\eta,\vec{x}) \ , \tag{199}$$

  i.e., as long as we are dealing with super-Hubble modes and Eqs. (194) are satisfied, each component features an adiabatic sound speed independent of x, obeying to $c_x^2(\eta) \equiv \bar{p}'_x(\eta)/\bar{\rho}'_x(\eta)$,

- second,

$$\delta p_{\rm tot}(\eta,\vec{x}) = c_{\rm s}^2(\eta)\,\delta\rho_{\rm tot}(\eta,\vec{x}) \ , \tag{200}$$

  where the squared total sound speed $c_{\rm s}^2$ can be easily expressed as a weighted average over the squared adiabatic sound speed of individual components. Hence, the total matter content resulting from the sum of all components also features an adiabatic sound speed.

For the reason expressed in the last item, such initial conditions are usually called *adiabatic initial conditions*. The word adiabatic refers to thermodynamics, but it is actually more helpful to think of adiabatic initial conditions as resulting from the assumption of a single inhomogeneous degree of freedom in the early universe: then Eqs. (194) stands for the true definition of adiabatic initial conditions, and Eqs. (197, 198, 199, 200) for its consequences.

More general initial conditions not obeying to Eqs. (194) can be expanded on different bases. A famous basis is formed by (i) the adiabatic mode; (ii) $(N-1)$ non-decaying isocurvature modes, getting their name from the property that for each of them, the total density (and spatial curvature) perturbations vanish asymptotically in the super-Hubble limit, while two species have opposite density perturbations compensating each other; (iii) N decaying modes that are irrelevant for most purposes.

It is clear from the previous discussion that non-adiabatic initial conditions should only be considered when assuming that primordial perturbations are generated by more than one degree of freedom (for instance, two inflaton fields, or one inflaton and one axion, etc.) The assumption of several degrees of freedom is necessary but not sufficient. The primordial universe may contain a mixture of adiabatic and isocurvature modes until a time at which all species are brought in thermal equilibrium. At that time, if we further assume that all chemical potentials vanish, the perturbations of each species can be inferred from those of temperature, $T(\eta,\vec{x}) = \bar{T}(\eta) + \delta T(\eta,\vec{x})$. Then, temperature plays precisely the role of a single time-shifting function. Any non-adiabatic initial condition is washed out and becomes irrelevant. Hence, isocurvature modes can be observable — for instance in the CMB — only under additional assumptions. For instance, one species carrying isocurvature perturbations might remain out of equilibrium at all times, or might feature a chemical potential with spatial fluctuations. There exist a few non-minimal, but still reasonable scenarii featuring isocurvature perturbations (with axions, curvatons etc.). Ultimately, the presence of isocurvature modes is to be tested with observations. Current CMB observations put strong limits on the amplitude of such modes, and prefer purely adiabatic initial conditions. We will restrict to this case in the rest of this course.

For adiabatic initial conditions, we found the relation (197) holding between density fluctuations. Also, if we do not consider the case of species with a non-zero anisotropic stress, we can assume that $\phi = \psi$ at initial time. By plugging Eqs. (198), (200) and $\phi = \psi$ into the Einstein equations, one is led to a second-order differential equation for ψ only. During the radiation dominated era, one can show that this equation has two solutions, one constant in time, and one decaying. The (00) Einstein equation then shows that the constant solution is related to density fluctuations through

$$-2\psi = -2\phi = \delta_{\rm tot} \simeq \delta_\gamma = \text{constant} \ . \tag{201}$$

Hence, in the Newtonian gauge and for adiabatic initial conditions, super-Hubble metric fluctuations and density fluctuations are static. In fact, on super-Hubble scales, there can only be some time evolution when the universe changes of total equation of state (or equivalently, of expansion law). This is the case at the time of equality between radiation and matter. During matter domination and on super-Hubble scales, they freeze out again. Then, the relation $-2\psi = -2\phi = \delta_{\rm tot}$ and Eq. (198) are still satisfied, but now $\delta_{\rm tot} \simeq \delta_{\rm b} = \delta_{\rm cdm} = \frac{3}{4}\delta_\gamma$. This detail will become important when studying the Sachs-Wolfe effect in section 12.9.2.

Power spectra and transfer functions

The theory of cosmological perturbations is a stochastic theory: the fluctuations of a given quantity in a given point, $A(\eta,\vec{x})$, obey to a distribution of probability. As long as we stick to linear perturbation theory, there is a "linear transport of probability" from one time to another

time. Let the probability of A in a given point at time η_1 be $\mathcal{P}_1(A)$. In the same point and at η_2, the linear evolution would have transformed each value A into αA, where α is the linear growth (or decrease) factor of A between η_1 and η_2. So the probability of A at time η_2 is given by $\mathcal{P}_2(A) = \mathcal{P}_1(A/\alpha)$. This "linear transport of probability" implies that the linear evolution respects the shape of the probability distribution, and rescales all its statistical moments of order n by α^n. In particular, if the initial probability is Gaussian, the statistics will remain Gaussian at any later time, and the evolution of the system can be formulated as an evolution of the root mean square of all quantities. In summary, as long as we assume linear perturbations with Gaussian initial conditions, our goal is to solve for the evolution of the root mean squares of the fluctuations.

The equal-time 2-point correlation function of any quantity A in real space is given by

$$\langle A(\eta,\vec{x})A(\eta,\vec{x}')\rangle \equiv \xi(\eta,\vec{x},\vec{x}') \overset{SHI}{=} \xi(\eta,|\vec{x}'-\vec{x}|) \ , \tag{202}$$

where the last equality holds as a consequence of Statistical Homogeneity and Isotropy (SHI, assumed to hold in a perturbed FL universe). Indeed, the correlation function should be invariant under spatial translations and rotations.

We can go to Fourier space and use the same letter to denote the Fourier transform of A. If A is real, $A(\eta,-\vec{k}) = A^*(\eta,\vec{k})$. It is easy to show that the previous equation (using the assumption of SHI) implies that the two-point correlation function in Fourier space reads

$$\langle A(\eta,\vec{k})A^*(\eta,\vec{k}')\rangle \overset{SHI}{=} \delta_D(\vec{k}'-\vec{k})P_{\rm A}(k) \ , \tag{203}$$

where δ_D is the Dirac distribution. Here, statistical homogeneity is responsible for the fact that the two-point correlation vanishes for $\vec{k} \neq \vec{k}'$, and statistical isotropy for the fact that $P_{\rm A}$ only depends on the modulus $k = |\vec{k}|$. The function $P_{\rm A}(k)$ is usually called the power spectrum of A. Some authors prefer to refer to the "dimensionless power spectrum", defined as

$$\mathcal{P}_{\rm A}(k) \equiv \frac{k^3}{2\pi^2}P_{\rm A}(k) \ . \tag{204}$$

The reason is that typical expressions for the average of various quantitites in real space consist in the convolution of the power spectrum with some window function $f(k)$, like in

$$\int \frac{d^3\vec{k}}{(2\pi)^3}P_{\rm A}(k)f(k) = \int \frac{4\pi k^2 dk}{(2\pi)^3}P_{\rm A}(k)f(k) = \int d\log k \ \mathcal{P}_{\rm A}(k)f(k) \ . \tag{205}$$

Hence the dimensionless spectrum $\mathcal{P}_{\rm A}(k)$ stands for the weight of each logarithmic interval in the integral. The term "scale-invariant spectrum" refers to $\mathcal{P}_{\rm A}(k)$ being independent of k, i.e. $P_{\rm A}(k) \propto k^{-3}$.

We know that for adiabatic initial conditions, all perturbations are related to each other through Eqs. (198, 201). Hence, with Gaussian adiabatic initial conditions, specifying the primordial power spectrum of one quantity is sufficient for knowing everything about the system. For instance, if we assume that the power spectrum of the metric perturbation ψ is a given function $P_\psi(k)$, then we infer from Eqs. (198, 201) that the photon and baryon primordial spectra are given by $P_\gamma(k) = 4P_\psi(k)$ and $P_{\rm b}(k) = \frac{9}{16}P_\gamma(k) = \frac{9}{4}P_\psi(k)$.

By convention, the primordial spectrum is usually given for the variable $\mathcal{R}$, which represents the spatial curvature perturbation on one initial comoving hypersurface (i.e. an hypersurface orthogonal to the energy flux of the total cosmic fluid in each point). The advantage of using this quantity is that it is conserved on super-Hubble scales for adiabatic initial conditions (while ϕ and ψ get rescaled when the equation of state of the universe changes, e.g. at radiation-matter equality). In the Newtonian gauge, $\mathcal{R}$ relates to ψ and to the total density perturbation through

$$\mathcal{R} = \psi - \frac{1}{3}\frac{\delta\rho_{\rm tot}}{\bar{\rho}_{\rm tot}+\bar{p}_{\rm tot}} \ . \tag{206}$$

The power spectrum of a given quantity at some arbitrary time can be decomposed into two parts, one accounting for initial conditions, and one accounting for linear evolution with time:

$$\langle A(\eta,\vec{k})A^*(\eta,\vec{k'})\rangle = \delta_D(\vec{k'}-\vec{k})\left[\frac{A(\eta,\vec{k})}{\mathcal{R}(\vec{k})}\right]^2 P_{\mathcal{R}}(k)\ . \tag{207}$$

In the above equation, there is no time argument for $\mathcal{R}(\vec{k})$ since this quantity is conserved on super-Hubble scales. We just assume that $\mathcal{R}$ is evaluated at a time such that $k \ll aH$ for all modes of interest. In a FL universe, the equations of motion of all perturbations respect isotropy, and do not depend on the direction of the wavector $\vec{k}$. Hence the ratio between brackets in the last equation is a function of k, not $\vec{k}$. This function accounts for the linear evolution, independently of initial conditions. It is called the "transfer function" of A. In this course, we will denote transfer functions with the same letter as the perturbations themselves, but with the modulus of k as an argument, i.e.

$$A(\eta,k) \equiv \frac{A(\eta,\vec{k})}{\mathcal{R}(\vec{k})}\ . \tag{208}$$

In summary of this section, solving for cosmological perturbations (in a model with adiabatic Gaussian IC's) amounts in

- postulating a primordial spectrum, or calculating it within the framework of a model, for instance of inflation;
- solving the equations of motion of all perturbations, with quantities normalized initially to $\mathcal{R}(\vec{k}) = 1$, in such a way to obtain all transfer functions $A(\eta,k)$ and to infer the evolution of the various root mean squares.

12.9.2 CMB temperature anisotropies by Lesgourgues

From now on, we will assume for simplicity that the universe is flat.

Photon scattering rate

CMB physics consists is the study of electron, baryon and photon perturbations on cosmological scales, taking into account their gravitational coupling with collisionless species such as decoupled neutrinos and CDM. Electrons and baryons carry opposite electric charges and are coupled to each other through very efficient Coulomb scattering processes. Electrons and photons are coupled through Thomson scattering, which is the limit of Compton scattering when electrons are non-relativistic and photons carry a smaller energy than the rest mass of the electron. Then, the scattering process results mainly in a deflection of the photon, with a negligible transfer of energy between the two particles. The leading interaction between baryons and photons is the gravitational one.

In units such that $c = 1$, the Thomson scattering rate (with respect to conformal time) is given by $\Gamma = \sigma_T a n_e x_e$, where σ_T is the Thomson scattering rate, a is the scale factor, n_e is the total electron number density (scaling like a^{-3} due to dilution), and x_e is the ionized electron fraction. The product an_e scales like a^{-2}. The ionized fraction is close to one at high energy. Then, at the time of recombination between electrons and nuclei (around $z \sim 1080$), which falls at the beginning of matter domination, n_e drops abruptly to very small values. This causes Thomson scattering to become suddenly very inefficient, and photons to decouple from electrons.

Hence photon decoupling is the story of Thomson scattering becoming inefficient, while Coulomb scattering remains very strong. For that reason, one can describe electrons and baryons as a single tightly-coupled fluid. People often refer only to baryons for simplicity. At early time, the full system of (electrons)-baryons-photons is also tightly coupled, but later on, it splits progressively into two collisonless species, (electrons)-baryons on the one hand, photons on the other hand.

The thermodynamical description of recombination is very technical, due to the different energy level of atoms (in particular, of hydrogen). In order to understand CMB anisotropies, we only need to describe the main results of recombination studies at a very qualitative level.

- The free electron fraction $x_e(\eta)$ starts from one at high redshift. It decreases sharply at the recombination time (corresponding to $z \sim 1080$ or $T \sim 0.3$ eV), and freezes at a very small value (due to departure from thermal equilibrium). Around $z \sim 10$, star formation causes a reionization of the universe, and x_e goes up again to one.

- The Thomson scattering rate $\Gamma = \sigma_T a n_e x_e$ evolves like $a^{-2} x_e$. Before recombination, $\Gamma \gg \frac{a'}{a}$, and the universe is opaque. The sudden drop of x_e at recombination renders the universe transparent: $\Gamma \ll \frac{a'}{a}$. Due to the dilution factor coming from n_e, Γ remains much smaller than $\frac{a'}{a}$ even during reionization, and the universe keeps being transparent (this is why despite of reionization, most photons emitted at recombination do not interact anymore, and allow us to observe anisotropies on the last scattering surface).

- The optical depth $\tau(\eta) \equiv \int_{\eta}^{\eta_0} d\eta \Gamma(\eta)$ represents the opacity of the universe at a given time, when seen from today (when $\eta = \eta_0$). It tends to infinity when $\eta \longrightarrow 0$, falls below one at recombination and stabilize at a value of the order of 0.1 between recombination and reionization. After reionization it decreases smoothly and reaches zero today by definition.

- The visibility function $g(\eta) \equiv -\tau' e^{-\tau}$ gives the probability that a CMB photon seen today experienced its last scattering at time η. It starts from negligible values at high redshift (suppressed by the $e^{-\tau}$ factor). It has a narrow spike around the time of recombination, and then it falls again to negligible values due to the smallness of τ' between recombination and reionization. It develops a second smaller and wider spike around reionization. This function shows that most CMB photons did not interact between the last scattering surface and today, while a minority rescattered at reionization. The width of the recombination spike gives an indication on the thickness of the last scattering "surface".

- The diffusion length $\lambda_{\rm d}(\eta)$ is an important quantity for understanding the damping of temperature anisotropies on small scales. At any given time, the comoving mean free-path (mfp) of photons is given by $r_{\rm mfp} = 1/\Gamma(\eta)$ (still in units where $c = 1$). If the photons experience a random walk analogous to Brownian motion in a gas, the comoving distance over which they travel between time $\eta_{\rm ini}$ and η can be approximated by

$$r_{\rm d} \sim \left[\int_{\eta_{\rm ini}}^{\eta_0} d\eta \; \Gamma \; r_{\rm mfp}^2 \right]^{1/2} = \left[\int_{\eta_{\rm ini}}^{\eta_0} d\eta \; \Gamma^{-1} \right]^{1/2} . \qquad (209)$$

  This physical diffusion length of photons is given in this approximation by $\lambda_{\rm d} \simeq a r_{\rm d}$. It grows quickly from very small to very large scales (comparable to the Hubble scale) around the time of recombination.

Sketch of connection of Γ and τ by Craps (? need to learn it better?)

Idea is to consider a logarithm.
(???)

$$\Gamma(\eta)(1-\Gamma(\eta+\Delta\eta)\Delta\eta)(1-\Gamma(\eta+2\eta\Delta\eta)\Delta\eta)\cdots =$$

$$\prod_{n=1}^{\frac{\eta_0-\eta}{\Delta\eta}}(1-\Gamma(\eta+\Delta\eta)\Delta\eta)$$

take log:

$$\ln\prod_{n=1}^{\frac{\eta_0-\eta}{\Delta\eta}}(1-\Gamma(\eta+\Delta\eta)\Delta\eta)=\sum_{n=1}^{\frac{\eta_0-\eta}{\Delta\eta}}\ln(1-\Gamma(\eta+\Delta\eta)\Delta\eta)=-\int_{\eta}^{\eta_0}d\tilde{\eta}\Gamma(\tilde{\eta})=-\tau$$

(???)

$$\frac{dP}{d\eta}=\Gamma e^{-\eta}=-\dot{\eta}e^{-\eta}=g(\eta)$$

Thus

$$\tau=\int\Gamma(\tilde{\eta})d\tilde{\eta}$$

Γ is not σ (?)

(more explanation and illustrations are needed!)

Boltzmann equation

Since photons decouple from baryons near the recombination time, we cannot describe them with fluid equations, and need to solve the Boltzmann equation

$$\frac{d}{d\eta}f=C\left[f,f_e\right] \tag{210}$$

at order one in perturbations. The right-hand side stands for the photon-electron coupling due to Thomson scattering. As explained in the previous section, electrons and baryons are so tightly coupled that it makes no difference to think of this term as a photon-electron or photon-baryon coupling term. Solving this equation is involved because the photon phase-space distribution f involves many arguments, $f(\eta,\vec{x},\vec{p})$. Fortunately one can reduce the dimensionality of the problem. First, we notice that as long as photons are in thermal equilibrium with electrons (and hence with baryons), they are entirely described in any point by the local value of the equilibrium temperature $T(\eta,\vec{x})$. The phase space distribution is then of the Bose-Einstein form:

$$f(\eta,\vec{x},\vec{p})=\frac{1}{e^{\frac{p}{T(\eta,\vec{x})}}-1}\ . \tag{211}$$

It can be expanded into a background part and a first-order perturbation, $f=\bar{f}+\delta f$, with:

$$\bar{f}(\eta,p)=\frac{1}{e^{\frac{p}{\bar{T}(\eta)}}-1}\ ,\qquad \delta f(\eta,\vec{x},\vec{p})=\frac{d\bar{f}}{d\log p}\ \frac{\delta T(\eta,\vec{x})}{\bar{T}(\eta)}\ . \tag{212}$$

We see that in the tightly-coupled regime we could replace the variable $f(\eta,\vec{x},\vec{p})$ by the lower-dimensional variable $\Theta(\eta,\vec{x})\equiv[\delta T(\eta,\vec{x})/\bar{T}(\eta)]$. The Boltzmann equation leads to an equation of motion for $\Theta(\eta,\vec{x})$.

At later times, when photons decouple, the shape of the Bose-Eisntein distribution of photons is preserved. This can be infered from the geodesic equation. This equation tells how the individual momentum p of photons evolve when they are decoupled and they travel in the perturbed universe. In the Newtonian gauge, it reads

$$\frac{d(a\,p)}{d\eta} = -a\,p\,\phi' - a\,\epsilon\,\hat{n}\cdot\vec{\nabla}\psi \ . \tag{213}$$

The left-hand side represents the time evolution of the product (ap) for a photon of momentum p traveling over a geodesic. In a perfectly homogeneous universe, the photon would only experience the average cosmological redshifting, $p \propto a^{-1}$, and the product (ap) would be conserved. Due to presence of perturbations in the universe, photons experience gravitational interactions and the product (ap) varies. The first term $-a\,p\,\phi'$ accounts for dilation, i.e. for the fact that locally, the expansion of the universe is a bit more advanced or delayed than the average (remember that $a(1+\phi)$ can be seen as the "local scale factor"). This means that the redshifting of the photon is also a bit more advanced or delayed locally. The second term accounts for the gravitational blueshifting of photons falling in gravitational potential wells (or redshifting of those leaving potential wells). The energy $\epsilon \equiv \sqrt{p^2+m^2}$ can be simply replaced by p for massless photons. In that case we can rewrite the geodesic equation as

$$\frac{d\ln(a\,p)}{d\eta} = -\phi' - \hat{n}\cdot\vec{\nabla}\psi \ . \tag{214}$$

The fact that p disappeared from the right-hand side is crucial: it shows that photons of different momenta traveling in the perturbed universe along a given geodesic all experience the same relative momentum variation. The consequence is that there can be no distortions of the Bose-Einstein shape of f. However, photons traveling along different geodesics and in different directions experience different redshifting. Hence the distribution function acquires a dependence on one extra argument, the direction $\hat{n}$ of propagation ($\hat{n} \equiv \vec{p}/p$):

$$f(\eta,\vec{x},\vec{p}) = \frac{1}{e^{\frac{p}{T(\eta,\vec{x},\hat{n})}} - 1} \ . \tag{215}$$

We can perform the same decomposition in terms of background and perturbations as in Eq. (212). The difference is that $\Theta = \frac{\delta T}{T}$ is now a function of $(\eta,\vec{x},\hat{n})$. The Boltzmann equation can now be used to derive an equation of motion for $\Theta(\eta,\vec{x},\hat{n})$.

If we work in Fourier space, we can derive the equation of motion for the function $\Theta(\eta,\vec{k},\hat{n})$. Because of the statistical isotropy of the FL universe, this equation does not depend explicitly on $\vec{k}$ nor $\hat{n}$, since there is no preferred direction: it depends only on the direction of propagation relatively to the considered wavenumber, i.e. on the product $(\vec{k}\cdot\hat{n})$. Hence the equation of motion can be written in terms of k and of the angle θ such that $(\vec{k}\cdot\hat{n}) = k\cos\theta$. The initial conditions for Θ do depend on the wavevector $\vec{k}$ (since each mode gets random initial conditions), but they also depend only on θ rather than $\hat{n}$, for a reason that will become clear in the paragraphs below. Hence we can entirely eliminate $\hat{n}$ from the problem, and solve the equation of motion for $\Theta(\eta,\vec{k},\theta)$.

Finally, we can expand the temperature anisotropy with respect to θ using a Legendre transformation:

$$\Theta(\eta,\vec{k},\theta) = \sum_l (-i)^l(2l+1)\Theta_l(\eta,\vec{k})P_l(\cos\theta) \ . \tag{216}$$

Here the Θ_l's are the temperature anisotropy multipoles, and P_l the Legendre polynomials. It can be shown that the monopole Θ_0 is related to the photon density fluctuation δ_γ in a given point, the dipole Θ_1 to its velocity divergence θ_γ, and the quadrupole Θ_2 to its anisotropic stress

σ_γ. The Boltzmann equation can be written as an infinite hierarchy of equations of motion for the coupled multipoles Θ_l.

Actually, the equation of motion for Θ takes a striking form when written in real space, before Fourier and Legendre expansions:

$$\Theta' + \hat{n}\cdot\vec{\nabla}\Theta - \phi' + \hat{n}\cdot\vec{\nabla}\psi = -\Gamma\left(\Theta - \Theta_0 - \hat{n}\cdot\vec{v}_e\right) . \tag{217}$$

We recall that Γ is the conformal Thomson scattering rate, Θ_0 the temperature monopole (i.e. the average of $\Theta(\eta,\vec{x},\hat{n})$ over all directions $\hat{n}$), and $\vec{v}_e$ the bulk velocity of electrons, equal to that of baryons due to tight Coulomb interactions. We recall that the variable $\theta_{\rm b}$ is defined as the divergence of $\vec{v}_{\rm b} = \vec{v}_e$.

This equation is illuminating, since it shows that at early times, when photons, electrons and baryons form a tightly-coupled fluid, the fact that Γ is huge forces Θ to evolve in such way that the parenthesis on the right-hand side vanishes. When this is the case, Θ can only have two non-zero component: a monopole Θ_0, and a dipole equal to $\hat{n}\cdot\vec{v}_{\rm b}$. The condition on the dipole can be written equivalently as $\theta_\gamma = \theta_{\rm b}$.

This should remind us of the discussion of Sec. 12.9.1, when we noticed that a perfect fluid can be described in terms of δ_x, $c_{\rm s}^2$ and θ_x only, with a vanishing anisotropic stress σ_x. Here, we see concretely how this conclusion emerges from the Boltzmann equation in the tightly-coupled regime: Θ_2 (related to σ_γ) and all higher multipoles vanish; the photon perturbations can be described in terms of only two independent variables Θ_0 and Θ_1 (or δ_γ and θ_γ), like in a perfect fluid.

The fact that in the tightly-coupled limit the system is driven towards $\theta_\gamma = \theta_{\rm b}$ simply shows that the interaction imposes a common bulk velocity to photons, baryons and electrons, as it should be the case in any tightly-coupled fluid (note that we are referring to bulk velocities, not to individual velocities of particles, which are still equal to c for interacting photons).

In the tightly-coupled regime, the dipole component of $\Theta(\eta,\vec{x},\hat{n})$ is given by $\hat{n}\cdot\vec{v}_{\rm b}$, with $\theta_{\rm b} \equiv \vec{\nabla}\cdot\vec{v}_{\rm b}$. Hence in Fourier space this component reads $i(\hat{n}\cdot\vec{k})k^{-2}\theta_{\rm b} = i(\cos\theta)k^{-1}\theta_{\rm b}$. This justifies the fact that initial conditions for Θ in Fourier space depend on θ, but not on the two degrees of freedom of $\hat{n}$. As mentioned above, this property is preserved by the isotropic equations of motion, so that at all times we can study Θ as a function of the arguments $(\eta,\vec{k},\theta)$ instead of $(\eta,\vec{k},\hat{n})$.

Temperature anisotropy in a given direction

The map of temperature anisotropies that we observe today ($\eta = \eta_0$) in our location of the universe ($\vec{x} = \vec{o}$ with a proper choice of origin) when looking in a direction $\hat{n}$ is represented mathematically by

$$\frac{\delta T}{\bar{T}}(\hat{n}) = \Theta(\eta,\vec{o},-\hat{n}) \tag{218}$$

(since in a direction $\hat{n}$, we see photons traveling towards $-\hat{n}$). Our scope is now to relate this quantity to perturbations on the point of the last scattering surface seen in the same direction $\hat{n}$. This can be done by integrating the Boltzmann equation along the corresponding line-of-sight.

A good starting point consists in computing the total derivative of the product $e^\tau(\Theta+\psi)$ along the trajectory of photons between the last scattering surface and the observer. The reason for choosing this product rather than just Θ will become clear in a few lines: the derivative of this term will be easy to simplify, using the Boltzmann equation.

The total derivative of an arbitrary function $\mathcal{F}$ of $(\eta,\vec{x},\hat{n})$ along the trajectory of photons going in a direction $\hat{n}$ reads

$$\frac{d}{d\eta}\mathcal{F}(\eta,\vec{x},\hat{n}) = \mathcal{F}' + \frac{dx_i}{d\eta}\frac{\partial\mathcal{F}}{\partial x_i} + \frac{dn_i}{d\eta}\frac{\partial\mathcal{F}}{\partial n_i} . \tag{219}$$

If $\mathcal{F}$ is of order one in perturbations, the first two terms on the right-hand side are also of order one. Instead the last term is of order two, since $\frac{dn_i}{d\eta}$ is of order one (this is clear from the fact that in an unperturbed universe, photons would travel in straight line with $\frac{dn_i}{d\eta} = 0$). Hence we can drop this term in first-order perturbation theory, and for $\frac{dx_i}{d\eta}$ we only need to keep the zero-th order contribution, that can be computed assuming a homogeneous universe:

$$\frac{dx_i}{d\eta} = \hat{n} \ . \tag{220}$$

The previous relation is just telling that photons are traveling in the direction $\hat{n}$ and at the velocity of light: hence, in units where $c = 1$, $d\vec{x}^2 = d\eta^2$. In summary, the total derivative of $\mathcal{F}$ is given at first order by

$$\frac{d}{d\eta}\mathcal{F}(\eta, \vec{x}, \hat{n}) = \mathcal{F}' + \hat{n} \cdot \vec{\nabla}\mathcal{F} \ . \tag{221}$$

Let us now replace the generic function $\mathcal{F}$ by

$$\mathcal{F}(\eta, \vec{x}, \hat{n}) = e^{-\tau(\eta)} \left(\Theta(\eta, \vec{x}, \hat{n}) + \psi(\eta, \vec{x})\right) \ . \tag{222}$$

The total derivative of this function reads

$$\frac{d}{d\eta}\left[e^{-\tau}(\Theta + \psi)\right] = e^{-\tau}\left(\Theta' + \psi' + \hat{n} \cdot \vec{\nabla}(\Theta + \psi)\right) - \tau' e^{-\tau}(\Theta + \psi) \ . \tag{223}$$

We now use the linearized Boltzmann equation (217) and the fact that $\tau' = -\Gamma$ to write the result as

$$\frac{d}{d\eta}\left[e^{-\tau}(\Theta + \psi)\right] = -e^{-\tau}\tau'(\Theta_0 + \psi + \hat{n} \cdot \vec{v}_{\rm b}) + e^{-\tau}(\phi' + \psi') \ . \tag{224}$$

Finally, using the definition of the visibility function g given in Sec. 12.9.2, we get

$$\frac{d}{d\eta}\left[e^{-\tau}(\Theta + \psi)\right] = g\left(\Theta_0 + \psi + \hat{n} \cdot \vec{v}_{\rm b}\right) + e^{-\tau}(\phi' + \psi') \ . \tag{225}$$

We can integrate this relation along the line of sight, i.e. along a straight line seen by the observer in a given direction $-\hat{n}$ (since the photons go in the direction $\hat{n}$), starting from an early time *before* recombination (such that $e^{-\tau(\eta_{\rm ini})} \simeq 0$) until the present time at which photons reach the observer (with by definition $e^{-\tau(\eta_0)} = 1$). The result reads

$$(\Theta + \psi)|_{\rm obs} = \int_{\eta_{\rm ini}}^{\eta_0} d\eta \left[g\left(\Theta_0 + \psi + \hat{n} \cdot \vec{v}_{\rm b}\right) + e^{-\tau}(\phi' + \psi')\right] \ , \tag{226}$$

where the notation $|_{\rm obs}$ means "evaluated at the observer location, along this line of sight", i.e. at the coordinate $(\eta_0, \vec{o}, \hat{n})$.

We can gain further intuition from this equation if we use the instantaneous decoupling approximation, in which all photons are assumed to decouple precisely at the time $\eta_{\rm dec}$. In this limit, we can replace the visibility function g by the Dirac function $\delta_D(\eta - \eta_{\rm dec})$, and $e^{-\tau}$ by the Heaviside function $H(\eta - \eta_{\rm dec})$. Note that the approximation $g(\eta) = \delta_D(\eta - \eta_{\rm dec})$ is correctly normalized, since the definition of g implies $\int d\eta\, g(\eta) = 1$. In the instantaneous decoupling limit, Eq. (226) reads:

$$(\Theta + \psi)|_{\rm obs} = \left(\Theta_0 + \psi + \hat{n} \cdot \vec{v}_{\rm b})\right)|_{\rm dec} + \int_{\eta_{\rm dec}}^{\eta_0} d\eta\, (\phi' + \psi') \ , \tag{227}$$

where the notation $|_{\rm dec}$ means "evaluated on the last scattering surface, along this line of sight", i.e. at the coordinate $(\eta_{\rm dec}, -r_{\rm dec}\hat{n}, \hat{n})$ (here $r_{\rm dec}$ is the comoving radius of the last scattering surface). Let us now give the interpretation of each term in this crucial equation.

First, $\Theta|_{\rm obs}$ is the temperature anisotropy measured by the observer in the direction $-\hat{n}$, while $\Theta_0|_{\rm dec}$ is the temperature anisotropy in the point of the last scattering surface seen in the same direction. If only these two terms where present, this relation would simply tell us that that the temperature anisotropy seen today in a given direction is equal to the intrinsic anisotropy in the point where the observed photons last scattered.

Second, the term $\hat{n} \cdot \vec{v}_{\rm b}|_{\rm dec}$ stands for the correction to this temperature coming from by the usual Doppler. Indeed, this correction is caused by the velocity of the baryon-photon fluid (that we assumed to be tightly coupled until $\eta_{\rm dec}$) projected along the line of sight.

Next, we expect a correction from gravitational effects. The redshifting and blueshifting of the photons traveling along gravitational potential fluctuations should affect the observed temperature anisotropy relatively to the intrinsic one. It turns out that if the gravitational potential was constant in time (but, of course, not in space), this effect would be conservative, and would only depend on $\psi|_{\rm obs} - \psi|_{\rm dec}$. This explains the second term on the left-hand and right-hand sides. But if ψ varies in time, the effect is not conservative anymore: intuitively, the amount of blueshifting and redshifting experienced by photons traveling across a potential well do not compensate each other if the gravitational well gets deeper between the time at which the photon enters and leaves the well. This explains the addition term $\int d\eta\, \psi$. A similar effect is caused by dilation effects along the line-of-sight, and contributes like $\int d\eta\, \phi$.

Finally, we can drop the second term on the left-hand side, because this term represents only a tiny *isotropic* correction to the observed anisotropies. It is impossible to measure it with the CMB map only, because it is formally equivalent to a redefinition of the average temperature $\bar{T}$, but only by a tiny amount of the order of $10^{-5}\bar{T}$.

Let us write once more our result, dropping this unobservable correction, and grouping the terms in a particular way:

$$\Theta|_{\rm obs} = \underbrace{(\Theta_0 + \psi)|_{\rm dec}}_{\rm SW} + \underbrace{\hat{n} \cdot \vec{v}_{\rm b}|_{\rm dec}}_{\rm Doppler} + \underbrace{\int_{\eta_{\rm dec}}^{\eta_0} d\eta\, (\phi' + \psi')}_{\rm ISW} \quad . \tag{228}$$

The first term is conventionally called the Sachs-Wolfe (SW) term, and includes the intrinsic temperature term Θ_0 and the "gravitational Doppler shift" term ψ at one point on the last scattering surface. The second term is the conventional Doppler term. The last term is called the Integrated Sachs-Wolfe (ISW) term and contains all non-conservative gravitational effects occurring in a universe with non-static metric fluctuations.

We can gain further insight on the Sachs-Wolfe term. We will try to find a simpler expression for $(\Theta_0 + \psi)|_{\rm dec}$, that applies at least for describing large angular patterns on CMB maps, i.e. maps smoothed over small scales. For instance, this expression would describe very well the map of the COBE satellite, which had limited angular resolution. In more precise terms, we wish to calculate the contribution to the Sachs-Wolfe term of large wavelengths, which are bigger than the Hubble radius at the time of recombination.

Let us first focus on the term $\Theta_0|_{\rm dec}$. We have seen that on super-Hubble scales, temperature anisotropies only have a monopole and a dipole component, related respectively to δ_γ and θ_γ. We can be more precise now. We know from thermodynamics that the local value of the photon density is proportional to the temperature to the power four. Taking the derivative of $\log \rho_\gamma = \log T^4$, we get $\delta_\gamma = 4\,\delta T/\bar{T}$, where on the right-hand side the temperature anisotropy is averaged over all directions $\hat{n}$: hence $\delta_\gamma = 4\Theta_0$.

Let us now focus on the term $\psi|_{\rm dec}$. We have seen in Sec. 12.9.1 that on super-Hubble scales and for adiabatic initial conditions, $\delta_\gamma = \frac{4}{3}\delta_{\rm b}$. Also, we know that decoupling takes place at the beginning of matter domination, and we mentioned at the very end of section 12.9.1 that for super-Hubble scales and during matter domination, one has $-2\phi = -2\psi = \delta_{\rm tot} = \delta_{\rm b} = \delta_{\rm cdm}$. Hence, $-2\psi = \frac{3}{4}\delta_\gamma$. Putting all these equalities together, we conclude that on the last scattering surface,

$$\Theta_0 + \psi = \frac{1}{4}\delta_\gamma + \psi = (-\frac{1}{4}\,2\,\frac{4}{3} + 1)\,\psi = (-\frac{2}{3} + 1)\,\psi = \frac{1}{3}\psi \ . \tag{229}$$

Moreover, on super-Hubble scales, we can neglect the Doppler term, which can be shown to be important only on sub-Hubble scales. We will see later that the integrated Sachs-Wolfe term plays a role on large scales, but only a small role, because ϕ and ψ are static during most of the evolution after decoupling. Hence we get an approximation for CMB anisotropies smoothed over small scales (that was first derived by Sachs and Wolfe in 1967):

$$\Theta|_{\rm obs,\,large\,scales} \simeq \frac{1}{3}\psi|_{\rm dec} = -\frac{1}{8}\delta_\gamma|_{\rm dec} \ . \tag{230}$$

In this calculation, we have seen that the term ψ wins over the term Θ_0, leading to a minus sign in front of δ_γ in the above relation. This means that an over-density on the last scattering surface ($\delta_\gamma > 0$), corresponding to a potential well ($\psi < 0$), leads to a cold spot in the observed map ($\Theta < 0$). Conversely, a hot spot corresponds to an under-density. Hence, due to the "gravitational Doppler shift" effect accounted by the term ψ (often called the Sachs-Wolfe effect), the patterns that we observe on CMB map are inverted with respect to intrinsic fluctuations on the last scattering surface.

Equation (228) (and its large-scale approximation (230)) are important for pedagogical purposes, but they have no practical application: indeed, what we wish to calculate and to compare to observations is a theoretical prediction for the statistical properties of CMB anisotropies. Hence, we need to compute at least the CMB two-point correlation function.

Spectrum of temperature anisotropies

Definition. The map of CMB temperature anisotropies can be expanded in spherical harmonics:

$$\frac{\delta T}{\bar{T}}(\hat{n}) = \Theta(\eta_0, \vec{o}, -\hat{n}) = \sum_{lm} a_{lm} Y_{lm}(\hat{n}) \ . \tag{231}$$

Using the Legendre expansion of Θ introduced in Eq. (216), and some basic relations between Legendre polynomials and spherical harmonics, it is easy to express a_{lm} as a function of Θ_l:

$$a_{lm} = (-i)^l \int \frac{d^3\vec{k}}{2\pi^2} Y_{lm}(\hat{k}) \Theta_l(\eta_0, \vec{k}) \ , \tag{232}$$

where we recall that hats denote unit vectors: $\hat{k} \equiv \vec{k}/k$. In linear perturbation theory and assuming Gaussian initial conditions, both Θ_l and a_{lm} are Gaussian random variables. Using the orthogonality relation of spherical harmonics and the definitions given in Sec. 12.9.1, we can infer the two-point correlation function of the a_{lm}'s as a function of the power spectrum of Θ_l, or even better, of the primordial curvature power spectrum:

$$\langle a_{lm} a^*_{l'm'} \rangle = \delta^K_{ll'} \delta^K_{mm'} \left[\frac{1}{2\pi^2} \int \frac{dk}{k} \Theta_l^2(\eta_0, k) \mathcal{P}_{\mathcal{R}}(k) \right] \ . \tag{233}$$

Here $\delta^K_{ll'}$ represents the Kronecker symbol. The fact that $\langle a_{lm} a^*_{l'm'} \rangle$ vanishes for $l \neq l'$ or $m \neq m'$ comes out of the algebra, but physically, it is a consequence of the homogeneity of the universe

— just like the fact that all power spectra in Fourier space are proportional to $\delta_D(\vec{k}' - \vec{k})$. Similarily, the fact that the quantity between brackets is a function of l but not of m is a consequence of isotropy — like the fact that in Fourier space, power spectra are functions of k but not $\vec{k}$.

The quantity between brackets is usually denoted C_l, and is called the power spectrum of temperature anisotropies in harmonic space, or the temperature harmonic power spectrum:

$$C_l \equiv \frac{1}{2\pi^2} \int \frac{dk}{k} \Theta_l^2(\eta_0, k) \mathcal{P}_{\mathcal{R}}(k) \ . \tag{234}$$

In a universe with linear and Gaussian perturbations, the C_l's encode all the information concerning the cosmological model describing our universe that is contained in the CMB temperature map.

It is worth coming back on the meaning of the averaging symbols $\langle ... \rangle$ in Eq. (233). Since the theory of cosmological perturbation is stochastic, the a_{lm}'s should be seen as random numbers, and the average is meant over many realizations of the theory. In a sense, "a given realization" means "a given universe", and the average holds over many universes, all obeying to the same cosmological model, which is encoded in the spectrum C_l.

However, we observe CMB anisotropies in our universe, i.e. in only one realization. CMB maps allow us to measure a definite value of each a_{lm}. Hence the squares $|a_{lm}|^2$ are not expected to be equal to C_l, even if we postulated the right model: there should be some scattering around C_l. This scattering limits the possibility to find the best theory matching the observations. However we can reduce considerably the scattering by noticing that for any fixed l, the statistical distribution of $|a_{lm}|^2$ is independent of m (as a consequence of isotropy, and as expressed by Eq. (234)). Hence, for an ideal full-sky CMB experiment, the best estimator of the underlying C_l's is the average between all observed coefficients $|a_{lm}|^2$ with fixed l,

$$C_l^{obs} \equiv \frac{1}{2l+1} \sum_{-l \leq m \leq l} |a_{lm}^{obs}|^2 \ . \tag{235}$$

In a typical universe, this quantity should be closer to the underlying C_l than a single $|a_{lm}^{obs}|^2$. The way to see this mathematically is to consider again the theoretical (stochastic) a_{lm}'s, and to define

$$\hat{C}_l \equiv \frac{1}{2l+1} \sum_{-l \leq m \leq l} |a_{lm}|^2 \ . \tag{236}$$

By performing averages in the same sense as in equation (233), one can easily show that for Gaussian a_{lm}'s,

$$\langle \hat{C}_l \rangle = C_l \qquad \text{and} \qquad \langle (\hat{C}_l - C_l)^2 \rangle = \frac{2}{2l+1} C_l^2 \ . \tag{237}$$

The first equality shows that the observed C_l's defined in Eq. (235) are unbiased estimators of the true underlying C_l's. The second equality gives the typical scattering between the theory and the observations. Since for larger l we perform an average over more values of m, the relative scattering decreases. This is the case with true data points, which are distributed qualitatively like in Fig. 138. This scattering is called cosmic variance. It can be seen as a theoretical error: because of cosmic variance, we cannot reconstruct the underlying model with infinite precision, even if we have infinitely precise observations. Cosmic variance is large for small l's, meaning that the shape of the true underlying C_l's will always be poorly known at low l.

Line-of-sight integral in Fourier space. According to equation (234), for a given primordial spectrum, the shape of the CMB spectrum C_l depends on the square of the transfer function

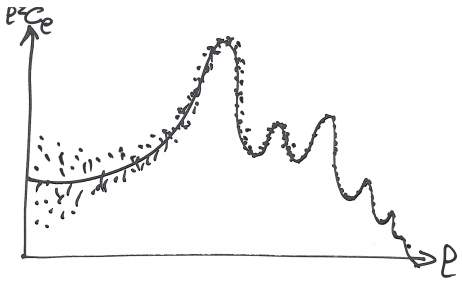


Figure 138: Observed values of the temperature harmonic spectrum, C_l^{obs}, are expected to be scattered around the true underlying C_l's. The scattering, called cosmic variance, decreases with increasing l.

$\Theta_l(\eta_0, k)$. We would like to understand this shape at least qualitatively. In real space, we did learn a lot on the behavior of $\Theta(\eta, \vec{x}, \hat{n})$ by using the line-of-sight integral approach presented in section 12.9.2. A similar approach can be worked out in Fourier and harmonic space, i.e. for the variable $\Theta_l(\eta, k)$. We do not present here the intermediate steps, which can be found in Ref. [2]. The final result shows many similarity with its real space counterpart, Eq. (226). It can be decomposed into:

$$\begin{aligned} \Theta_l(\eta_0, k) &= \int_{\eta_{\rm ini}}^{\eta_0} d\eta \; S_T(\eta, k) \; j_l(k(\eta_0 - \eta)) \; , \\ S_T(\eta, k) &\equiv \underbrace{g \, (\Theta_0 + \psi)}_{\rm SW} + \underbrace{\left(g \, k^{-2} \theta_{\rm b}\right)'}_{\rm Doppler} + \underbrace{e^{-\tau}(\phi' + \psi')}_{\rm ISW} \; . \end{aligned} \tag{238}$$

We see that $\Theta_l(\eta_0, k)$ is given by the convolution of spherical Bessel functions $j_l(x)$ with a function $S_T(\eta, k)$, called the temperature source function, which contains the usual three terms: Sachs-Wolfe, Doppler, and Integrated Sachs-Wolfe. Like in the previous section, we can use the instantaneous decoupling approximation, integrate the Doppler term by part, and write $\Theta_l(\eta_0, k)$ as:

$$\begin{aligned} \Theta_l(\eta_0, k) \simeq & \left[\Theta_0(\eta_{\rm dec}, k) + \psi(\eta_{\rm dec}, k)\right] j_l(k(\eta_0 - \eta_{\rm dec})) \\ & + k^{-1} \theta_{\rm b}(\eta_{\rm dec}, k) \; j_l'(k(\eta_0 - \eta_{\rm dec})) \\ & + \int_{\eta_{\rm dec}}^{\eta_0} d\eta \; \left[\phi'(\eta, k) + \psi'(\eta, k)\right] j_l(k(\eta_0 - \eta)) \end{aligned} \tag{239}$$

(note that in the second line, i.e. in the Doppler term, the prime stands for the derivative of the function $j_l(x)$ with respect to its argument, not with respect to conformal time). This approximate result can be plugged into Eq. (234) to obtain the final spectrum C_l. We see that each C_l can be decomposed into six terms: the power spectrum $C_l^{\rm SW}$ of the SW term, coming from the first line of Eq. (239) squared, that of the Doppler term, coming from the second line of Eq.(239) squared, that of the Integrated Sachs-Wolfe term, coming from the third line of Eq. (239) squared, and finally the three cross-spectra involving each pair of terms.

For large values of l, the spherical Bessel functions $j_l(x)$ and $j_l'(x)$ are very peaked near $x \simeq l$. Hence, for the Sachs-Wolfe and Doppler contributions to the spectrum C_l, the integral over k in Eq. (234) will pick up mainly modes with $k(\eta_0 - \eta_{\rm dec}) \simeq l$. This shows that the SW contribution to C_l is given by the product of the primordial spectrum with the squared transfer function $(\Theta_0 + \psi)$ at a given value of η and k:

$$C_l^{\rm SW} \sim \left[\Theta_0(\eta_{\rm dec}, k) + \psi(\eta_{\rm dec}, k)\right]^2 \mathcal{P}_{\mathcal{R}}(k) \; , \qquad k \simeq \frac{l}{(\eta_0 - \eta_{\rm dec})} \tag{240}$$

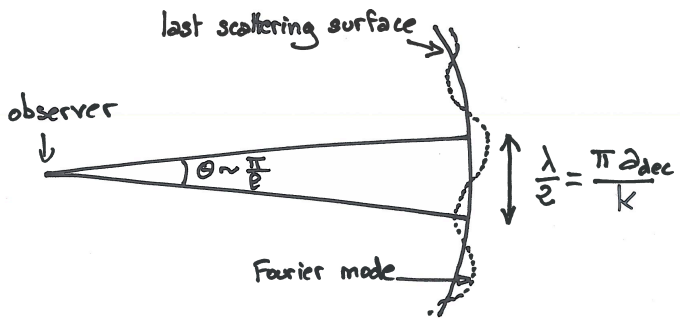


Figure 139: A multipole l refers to structures seen on the last scattering surface under an angle $\theta = \pi/l$. These structures are seeded by Fourier modes with a half wavelength $\frac{\lambda}{2} = \frac{\pi a_{\rm dec}}{k}$.

(for simplicity, we did not write numerical factors and powers of l or k in front of this expression). In other words, $C_l^{\rm SW}$ depends on the power spectrum of the perturbation $(\Theta_0 + \psi)$, evaluated at the time of decoupling, and for wavenumbers in the vicinity of $k = \frac{l}{(\eta_0 - \eta_{\rm dec})}$,

$$C_l^{\rm SW} \sim \langle |\Theta_0 + \psi|^2 \rangle_{(\eta,k) \simeq (\eta_{\rm dec},\, l/(\eta_0 - \eta_{\rm dec}))} \ . \tag{241}$$

We reached this result with mathematical arguments, but it has a very simple geometrical interpretation, illustrated in Fig. 139. The spectrum C_l encodes the correlation between structures on CMB maps seen under an angle $\theta = \pi/l$. This angle subtends a given physical scale on the last scattering surface, namely $\theta \times d_{\rm a}(z_{\rm dec})$, where $d_{\rm a}(z)$ is the angular diameter distance to objects of redshift z. Since the Sachs-Wolfe term $(\Theta_0 + \psi)$ contributes to the temperature map only at the time $\eta \simeq \eta_{\rm dec}$, the spectrum $C_l^{\rm SW}$ should depend on the power spectrum of the Sachs-Wolfe term $\langle |\Theta_0 + \psi|^2 \rangle$ at that time, and for a wavenumber such that

$$\frac{\lambda}{2} = \frac{\pi a(\eta_{\rm dec})}{k} = \theta \, d_{\rm a}(z_{\rm dec}) \ . \tag{242}$$

The reason for which λ has been divided by two is that for spherical harmonics, $\theta = \pi/l$ is the angle between a maximum and a minimum, while for a Fourier mode the distance between a maximum and a minimum is one half of the wavelength, $\frac{\lambda}{2} = \frac{\pi a}{k}$. Eq. (242) leads to

$$\frac{a(\eta_{\rm dec})}{k} = \frac{d_{\rm a}(z_{\rm dec})}{l} \ . \tag{243}$$

We recall that in a flat FL universe the angular diameter distance is given by

$$d_{\rm a}(z) = a(t(z)) \int_{t(z)}^{t_0} \frac{dt}{a} \ , \tag{244}$$

$t(z)$ being the proper time at which an object seen today with a redshift z emitted light. The conformal time $\eta(z)$ is defined similarly. In terms of conformal time,

$$d_{\rm a}(z) = a(\eta(z)) \int_{\eta(z)}^{\eta_0} d\eta = a(\eta(z)) \ [\eta_0 - \eta(z)] \ , \tag{245}$$

and for the case of a point located on the last scattering surface,

$$d_{\rm a}(z_{\rm dec}) = a(\eta_{\rm dec}) \ (\eta_0 - \eta_{\rm dec}) \ . \tag{246}$$

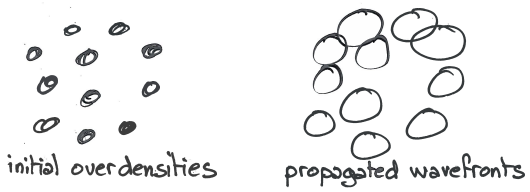


Figure 140: Initial over-densities in the early universe propagate in the form of wavefronts. The distance travelled by any wavefront at a given time is given by the sound horizon in the photon-baryon fluid. Here, we represent initial over-densities as spherical patterns at a given scale, while in the real universe primordial over-densities result from a superposition of structures on all scales.

Hence, Eq. (243) can be written as

$$\frac{1}{k} = \frac{(\eta_0 - \eta_{\rm dec})}{l} \ , \tag{247}$$

which is the same relation between k and l as in Eqs. (240, 241). In those equations, we implicitly performed a small-angle approximation. For large angles (small l's), it is inaccurate to say that a given angle/mutipole corresponds to a single Fourier mode on the last-scattering surface, and it is important to keep the spherical Bessel function of Eq. (238) and the integral over k of Eq.(234). In summary, Eqs. (240, 241) represent the instantaneous decoupling *and* small-angle limit of the true power spectrum $C_l^{\rm SW}$.

Using the same two limits, a similar discussion can be carried for the Doppler and Integrated Sachs-Wolfe power spectra. The Doppler term depends on the power spectrum of the baryon velocity divergence evaluated roughly at the same time and scale,

$$C_l^{\rm Doppler} \sim \langle |\theta_{\rm b}|^2 \rangle_{(\eta,k) \simeq (\eta_{\rm dec},\, l/(\eta_0 - \eta_{\rm dec}))} \ , \tag{248}$$

while the ISW term can be written approximately in terms of the integral

$$C_l^{\rm ISW} \sim \int_{\eta_{\rm dec}}^{\eta_0} d\eta \, (\eta_0 - \eta) \, \langle |\phi' + \psi'|^2 \rangle_{(\eta,k) \simeq (\eta,\, l/(\eta_0 - \eta))} \ . \tag{249}$$

Again, for simplicity, we did not write numerical factors and powers of l and k in front of these expressions.

In the next sections, we will infer the shape of the full spectrum C_l from that of the three power spectra appearing in Eqs. (241, 248, 249):

$$\begin{array}{lcll} \text{SW} & : & \langle |\Theta_0 + \psi|^2 \rangle & \text{at } (\eta, k) \simeq (\eta_{\rm dec},\ l/(\eta_0 - \eta_{\rm dec})) \ , \\ \text{Doppler} & : & \langle |\theta_{\rm b}|^2 \rangle & \text{at } (\eta, k) \simeq (\eta_{\rm dec},\ l/(\eta_0 - \eta_{\rm dec})) \ , \\ \text{ISW} & : & \langle |\phi' + \psi'|^2 \rangle & \text{for all } (\eta, k) \simeq (\eta,\ l/(\eta_0 - \eta)) \ . \end{array}$$

Before entering into details, we can make a guess. Even if the primordial spectrum $\mathcal{P}_{\mathcal{R}}(k)$ is smooth, the temperature spectrum C_l should contain some structure. Indeed, as illustrated by Fig. 140, any primordial over-density is expected to propagate. In Fig. 140, we represented naively the primordial over-densities with little dots, giving rise to spherical wavefronts at later time. In the real universe, wavefront patterns are less visible, because primordial over-densities result from a superposition of structures on all scales. However, there is always a characteristic scale in this problem: namely, the distance by which a wavefront travels between some time

in the primordial universe and the time of photon decoupling. This distance, called the sound horizon at decoupling $d_{\rm s}(\eta_{\rm dec})$, obeys to

$$d_{\rm s} \equiv a \int_{t_{\rm ini}}^{t} \frac{c_{\rm s} dt}{a} = a \int_{\eta_{\rm ini}}^{\eta} c_{\rm s} d\eta \ , \tag{250}$$

where $c_{\rm s}$ is the sound speed in the photon-baryon fluid (in units of the speed of light). Two points on the last scattering surface separated by this distance should be partially correlated, since density waves have propagated from one point to the other. Hence, in angular space, the two-point correlation function of CMB anisotropies should exhibit a characteristic feature for angular scales corresponding to the sound horizon at decoupling, $\theta \sim d_{\rm s}(\eta_{\rm dec})/d_{\rm a}(z_{\rm dec})$. Similarly, the harmonic power spectrum C_l should exhibit a feature at the corresponding scale, $l \sim \pi/\theta \sim \pi\, d_{\rm a}(z_{\rm dec})/d_{\rm s}(\eta_{\rm dec})$, and also for all the harmonics of this scale. We will get a confirmation of this in the next section.

Acoustic oscillations

As long as electrons, baryons and photons are tightly coupled, they form an effective single fluid in which density waves propagate at the sound speed

$$c_{\rm s}^2 = \frac{\delta p_\gamma + \delta p_{\rm b}}{\delta \rho_\gamma + \delta \rho_{\rm b}} \tag{251}$$

(the density and pressure of electrons is always negligible with respect to that of photons). The density fluctuation δ_x of each species $x = \gamma, b$ can be inferred from the local value of the equilibrium temperature. The fact that $\rho_{\rm b} \propto T^3$ and $\rho_\gamma \propto T^4$ implies $\delta_\gamma = \frac{4}{3}\delta_{\rm b}$, and tight coupling imposes $\theta_\gamma = \theta_{\rm b}$, as we already saw in Sec. 12.9.2. We can simplify the expression of the sound speed, using also the fact that $|\delta p_{\rm b}| \ll |\delta p_\gamma|$. The result reads

$$c_{\rm s}^2 = \frac{1}{3(1+R)} \ , \quad R \equiv \frac{4\bar{\rho}_{\rm b}}{3\bar{\rho}_\gamma} \propto a \ . \tag{252}$$

It is possible to derive a simple equation of motion for the photon temperature fluctuation $\Theta_0(\eta, \vec{x})$ in the tightly-coupled regime:

$$\Theta_0'' + \frac{R'}{1+R}\Theta_0' + k^2 c_{\rm s}^2 \Theta_0 = -\frac{k^2}{3}\psi + \frac{R'}{1+R}\phi' + \phi'' \ . \tag{253}$$

This equation follows from the combination of the continuity and Euler equations for photons and baryons. Given that R is proportional to the scale factor, we could replace R' by $(a'/a)R = aHR$. The second term on the left-hand side is a damping term, increasing with the contribution of baryons to the total energy of the fluid. The third term accounts for pressure forces in the effective fluid. The first term on the right-hand side accounts for the gravitational force, and the last two terms for dilation effects.

This equation would be that of a simple harmonic oscillator if R was a constant (no friction term, constant sound speed) and in absence of gravitational source terms. Then, the solution would be of the form

$$\Theta_0 = \Theta_{\rm ini} \cos(k c_{\rm s} \eta + \varphi) \ , \tag{254}$$

with two constants of integration $(\Theta_{\rm ini}, \varphi)$. We know that for adiabatic initial conditions and in the Newtonian gauge, photon density/temperature fluctuations should be constant in the super-Hubble limit, $k\eta \ll 1$: this fixes the phase to $\varphi = 0$. In the opposite limit, this solution corresponds to the propagation of acoustic oscillations. Actually, the limit between the constant

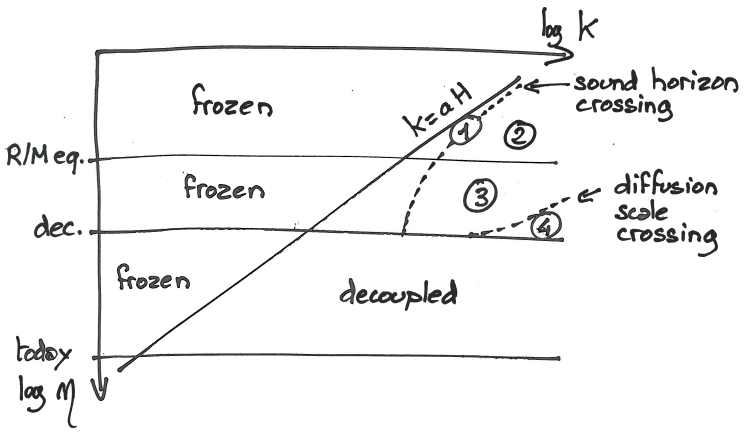


Figure 141: Different regions in (k, η) space, corresponding to qualitatively different behaviors for photon (and baryon) perturbations.

and oscillatory regime is not set by the value of $k\eta$, but by that of $kc_{\rm s}\eta$. In fact, the condition $kc_{\rm s}\eta \ll 1$ is equivalent to $\lambda \gg d_{\rm s}$, where λ is a physical wavelength ($\lambda = 2\pi a/k$), and $d_{\rm s}$ is the physical sound horizon, given in the case of a constant sound speed by:

$$d_{\rm s} = a \int_{\eta_{\rm ini}}^{\eta} c_{\rm s} d\eta \simeq a c_{\rm s} \eta \tag{255}$$

(assuming $\eta \gg \eta_{\rm ini}$). Hence, the phase $kc_{\rm s}\eta$ of the cosine stands for the ratio $2\pi d_{\rm s}/\lambda$. Modes start oscillating when their wavelength becomes smaller than the sound horizon, and later on, the number of oscillations is given by the ratio between these two scales.

In reality, R grows with time (and crosses one roughly around the time of decoupling). In addition, the gravitational source terms in Eq. (253) can play a role in some regimes. Let us describe qualitatively the evolution of Θ_0 in the different regions in (k, η) space shown in Fig. 141. In this figure, the horizontal axis corresponds to wavenumbers k (large wavelengths are on the left), and the vertical axis to conformal time, flowing from top to bottom. The super-Hubble and sub-Hubble regions are separated by the solid diagonal line corresponding to $k = aH$ (equivalent to $k\eta = 1$ during radiation domination). From top to bottom, the horizontal lines correspond to the time of equality between radiation and matter, to the time of photon decoupling, and to the time today. The upper dashed line separates wavelengths bigger/smaller than the sound horizon in the baryon-photon fluid before decoupling (after decoupling, this notion does not make sense anymore). As we have just seen, this limit corresponds to $\lambda = d_{\rm s}$, or (up to a factor 2π) to $k = (a/d_{\rm s})$. At early times, $c_{\rm s} = 1/\sqrt{3}$, and this condition reads $k\eta = \sqrt{3}$. Just before decoupling, R becomes large, $c_{\rm s}$ goes to zero, and the comoving sound horizon $(d_{\rm s}/a)$ becomes asymptotically constant, explaining the shape of the upper dashed line. Finally, the lower dashed line separates wavelengths bigger/smaller than the diffusion length defined in section 12.9.2: this line corresponds to $k = 1/r_{\rm d}$, where $r_{\rm d}$ is given in first approximation by Eq. (209).

The evolution of Θ_0 in the super-Hubble region is trivial: as long as $k\eta \ll 1$ — and *a fortiori* $kc_{\rm s}\eta \ll 1$ — the fluctuation Θ_0 is frozen, and remains approximately equal to its initial value.

The region marked with a ① in the figure corresponds to modes that are crossing the sound horizon before decoupling. This is precisely the region in which gravitational source terms are important. They shift the zero point of oscillations, and boost their amplitude (due to gravitational forces and dilution effects). This happens during a limited amount of time, because the metric fluctuations quickly decay inside the sound horizon during radiation

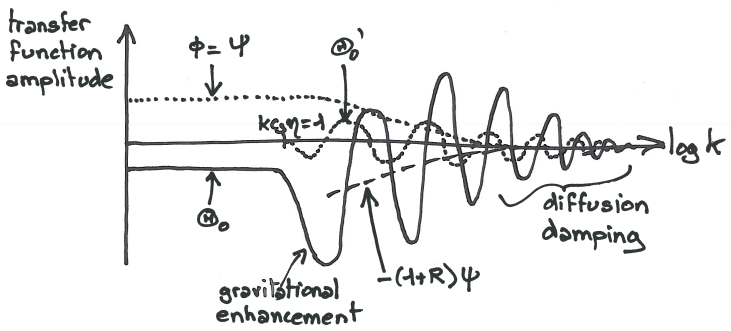


Figure 142: Transfer functions at the time of decoupling.

domination, making the gravitational source terms negligible. An approximation for the zero-point of oscillations can be found by setting Θ_0'' and Θ_0' to zero in equation (253), and by keeping only the first gravitational term:

$$\Theta_0^{equilibrium} = -\frac{1}{3c_{\rm s}^2}\psi = -(1+R)\psi \ . \tag{256}$$

Since the gravitational potential is non-zero on super-Hubble scales (we have seen that for adiabatic initial conditions $-2\psi = \delta_{\rm tot}$), the equilibrium point is shifted away from zero on those scales. It reaches asymptotically zero on sub-sound-horizon scales.

Region ② corresponds to wavelengths smaller than the sound horizon during radiation domination. In this regime, the metric fluctuations have decayed, so the source term in Eq. (253) can be neglected. The friction term can also be neglected, because during radiation domination, $R \ll 1$. Finally the effective mass $k^2c_{\rm s}^2$ is constant in time because $R \ll 1$ implies $c_{\rm s}^2 = 1/3$. Hence we are in the simple case discussed before, and the solution is proportional to $\cos(kc_{\rm s}\eta)$, corresponding to stationary oscillations, symmetric around $\Theta_0 = 0$.

Region ③ refers to wavelengths smaller than the sound horizon during the intermediate stage between the time of equality and that of photon decoupling. In this region, the metric perturbations have decayed, but R cannot be neglected (baryons and photons contribute to the total energy density with the same order of magnitude). Hence the oscillator equation has a non-negligible friction term (increasing with time), and a time-varying effective mass (decreasing with time). The solution of the equation corresponds to damped oscillations. Physically, this damping is caused by the increasing inertia and decreasing pressure of the baryon-photon fluid when the energy density of non-relativistic baryons takes over.

Region ④ refers to modes with smaller wavelength than the diffusion scale in the photon-baryon fluid. We have defined this scale in section 12.9.2. At early times, in the tightly-coupled limit, the mean free path of particles in the fluid is negligible, and cosmologically interesting scales are all well above the diffusion length. At the approach of decoupling, the diffusion length suddenly increases, and encompasses most of sub-sound-horizon wavelengths. In this regime, the oscillator equaion (253) does not apply anymore, because we cannot describe baryons and photons in terms of a perfect fluid. Perturbations are then strongly damped, since diffusion tends to average out any small-scale fluctuation.

After describing these different regions, we are ready for understanding the qualitative behavior of the various relevant transfer functions, evaluated at the time of decoupling. Figure 142 shows the transfer functions $\Theta_0(\eta_{\rm dec}, k)$ (solid line), $\psi(\eta_{\rm dec}, k)$ (upper dotted line) and $\Theta_0'(\eta_{\rm dec}, k)$ (middle dotted line) as a function of $\log k$. For simplicity, we neglect the role of the anisotropic stress generated by neutrinos (and to a lesser extent by photons near decoupling time). Hence we can assume $\phi = \psi$ at all times. For scales above the sound horizon ($kc_{\rm s}\eta \ll 1$),

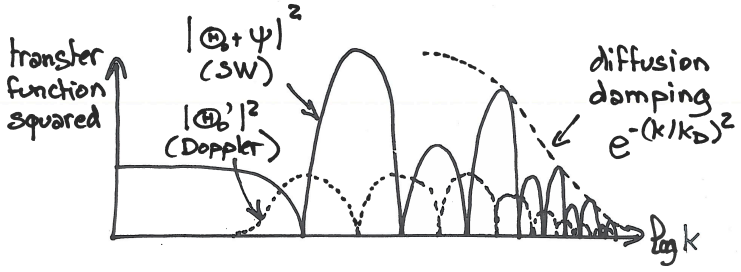


Figure 143: Squared transfer functions at the time of decoupling.

the transfer functions are constant: indeed we know that, on those scales and in the Newtonian gauge, density and metric fluctuations are frozen. Adiabatic initial conditions impose $-2\psi = \delta_{\rm tot} \simeq \delta_{\rm b} \simeq \frac{3}{4}\delta_\gamma \simeq 3\Theta_0$. The opposite sign of Θ_0 and ψ reflects the fact that an overdensity corresponds to a temperature excess and a gravitational potential well (and vice-versa). We remember that transfer functions are all normalized to $\mathcal{R}(\vec{k}) = 1$ (see section 12.9.1): this corresponds to a negative δ_γ (and Θ_0) and to a positive ψ, like in the figure.

Because of the decay of metric fluctuations inside the Hubble radius, the ψ curve smoothly decreases and tends towards zero in the small wavelength limit. The behavior of Θ_0 is more complicated. Modes which are just crossing the sound horizon near $\eta = \eta_{\rm dec}$ are experiencing the boost caused by gravitational source terms in Eq. (253): this explains the first bump in the solid line. For smaller wavelengths, we see oscillatory patterns corresponding to acoustic oscillations. Smaller wavelengths crossed the sound horizon earlier, and had more time for oscillating before decoupling. The maxima observed at the time of decoupling correspond to modes that could experience 0.5, 1, 1.5, 2,..., periods of oscillations before that time. The zero point of oscillations follows $-(1+R)\psi$, represented on the figure with a dashed line: this zero point reaches zero well inside the sound horizon. The amplitude of the oscillations is maximal for the first oscillatory pattern, i.e. for modes that crossed the sound horizon very recently. The second oscillatory pattern is reduced by the fact that those modes stayed for a longer time inside the sound horizon during the matter dominated regime, and experienced more damping due to baryons. The third and higher oscillatory patterns are reduced even more by diffusion damping just before photon decoupling. Temperature fluctuations on very small wavelengths are completely suppressed by photon diffusion.

Figure 142 also shows qualitatively the behavior of the time derivative $\Theta_0'(\eta_{\rm dec}, k)$, which exhibits oscillations that are out of phase with respect to those of $\Theta_0(\eta_{\rm dec}, k)$. This will be important in a few paragraphs, when discussing the Doppler effect.

Now that we understand qualitatively the behavior of the metric and photon transfer functions, we can go back to the decomposition of the CMB temperature spectrum C_l in three terms (Sachs-Wolfe, Doppler and integrated Sachs-Wolfe) discussed in section 12.9.2.

Sachs-Wolfe contribution. We have seen that the Sachs-Wolfe contribution to C_l is approximately given by the power spectrum of the combination $(\Theta_0 + \psi)$ at $\eta = \eta_{\rm dec}$, with a correspondence between k and l given by Eq. (247). We know that this power spectrum is given by the product of the primordial spectrum $\mathcal{P}_\mathcal{R}(k)$ (that we can choose to be scale-invariant in first approximation) by the square of the transfer function $(\Theta_0 + \psi)^2$. The qualitative behavior of the latter has no more secrets for us. We can pick up the solid and upper dotted lines in figure 142, add them up, and square the result. The result is shown in figure 143 as a function of $\log k$. For modes with $kc_{\rm s}^2\eta \ll 1$, we see a flat plateau: there, our previous calculation of the Sachs-Wolfe effect would apply (this part of the curve is actually called the Sachs-Wolfe plateau). We then observe a series of peaks. Due to the shift of the zero-point of oscillations given by $-(1+R)\psi$ for Θ_0, and hence by $-R\psi$ for the Sachs-Wolfe term $(\Theta_0 + \psi)$, there is an

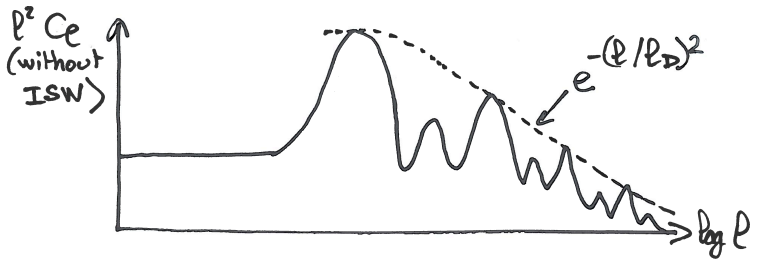


Figure 144: Contribution to the C_l's from the SW and Doppler terms.

asymmetry between the first few odd and even peaks, with odd peaks being enhanced. Moreover the overall amplitude of the peaks is suppressed in the large k limit by diffusion damping. A bit of algebra would show us that the envelope of the peaks is given in first approximation by the function $\exp[-(k/k_{\rm d})^2]$, where $k_{\rm d}$ is the diffusion wavenumber, related to the diffusion comoving scale of Eq. (209) by $k_{\rm d} r_{\rm d} = 1$.

Doppler contribution. Next, we know that the C_l's receive a second contribution from the Doppler effect, related to the power spectrum of $\theta_{\rm b}$, which is equal to θ_γ until baryon and photons decouple from each other. It turns out that the photon velocity divergence θ_γ is itself related to the time derivative of the temperature fluctuation Θ_0. At a very qualitative level, we can infer the Doppler contribution from the shape of the transfer function $\Theta_0'(\eta_{\rm dec}, k)$. This contribution is null for scales above the sound horizon, since in this regime there are no oscillations and no significant dynamics in the fluid. On smaller scales, the Doppler contribution has oscillatory patterns, that are out of phase with respect to those of the Sachs-Wolfe term. The Doppler contribution is represented schematically as a dotted line in figure 143.

We can now sum up the Sachs-Wolfe and Doppler contributions. Also, we can transpose our results for power spectra as a function of k in terms of C_l's as a function of l. We have already seen in section 12.9.2 that there is a mapping betwwen the two, at least in the small-angle and instantaneous decoupling approximation, with a correspondence between values of k and l given by Eq. (247). The result is shows in figure 144. Note that the vertical axis stands for $l^2 C_l$, for the following reason. If the primordial spectrum was scale invariant ($\mathcal{P}_{\mathcal{R}}(k) =$ constant) and the transfer functions were flat (as it is the case for large wavelengths/small l's), the quantity $l(l+1)C_l$ would also be flat and independent of l. This would follow from Eq. (240) if we had been more carefull in keeping all the factors. It is convenient to plot $l^2 C_l$ or $l(l+1)C_l$ instead of C_l, in order to display a roughly constant curve, just modulated by acoustic oscillations and diffusion damping.

Diffusion damping effect. In Fig. 144, we can identify all the features mentioned before: the flat Sachs-Wolfe plateau, the series of oscillations with enhanced odd peaks, and the exponentially decaying envelope of the peaks for large l. The envelope is now given by $\exp[-(l/l_{\rm d})^2]$, where $l_{\rm d}$ is the diffusion multipole, related to the diffusion angle by $l_{\rm d} = \pi/\theta_{\rm d}$. The diffusion angle is related to the diffusion scale $r_{\rm d}$ by the usual angular diameter distance relation,

$$\theta_{\rm d}\, d_{\rm a}(\eta_{\rm dec}) = \lambda_{\rm d} = a(\eta_{\rm dec}) r_{\rm d}(\eta_{\rm dec}). \tag{257}$$

At this point, we are almost done with the qualitative description of the CMB temperature spectrum C_l. We only missed the integrated Sachs-Wolfe contribution, and the effect of reionization.

Integrated Sachs-Wolfe contribution. We have seen that the integrated Sachs-Wolfe effect is given by an integral over $(\psi' + \phi')$ between photon decoupling and today. In fact, $(\psi' + \phi')$ remains vanishingly small in a large part of the space (k, η). In Figure 145, we

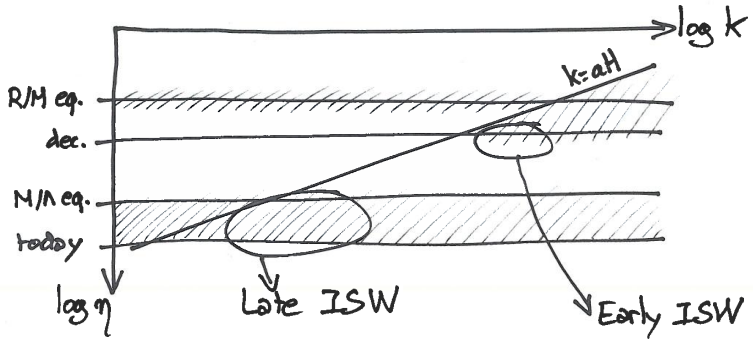


Figure 145: Regions in (k, η) space where metric fluctuations are expected to vary with time (giving rise potentially to an ISW effect after decoupling).

hatched all the regions in (k, η) space where metric fluctuations are expected to vary with time. Let us discuss these different regions.

Before photon decoupling, we know that metric fluctuations decay inside the sound horizon. Instead, in the Newtonian gauge, they remain frozen outside the Hubble radius, excepted near times at which the equation of state of the universe changes: namely, at the time of equality between radiation and matter.

Let us now discuss the variation of metric perturbations after photon decoupling (this is the relevant epoch for the ISW effect). Deep inside the matter dominated regime, one can show that metric fluctuations are static, even inside the Hubble radius (at least within linear perturbation theory). We will justify this result in section 12.9.3. Hence in figure 145 there are no hatches during matter domination on whatever scales. Note however that at the beginning of matter domination, it takes some time for sub-sound-horizon metric fluctuations to freeze around a constant value: hence the hatches continue below the line corresponding to the time of equality, and extend till the line corresponding to photon decoupling.

Later on, during the Λ (or dark energy) dominated regime, the equation of state of the universe changes again, so metric fluctuations vary on all scales, like at the time of equality. A simple calculation based on Einstein equations would show that metric fluctuations are damped during this stage.

In summary, contribution to the integrated Sachs-Wolfe effect can only come from two regions:

- just after photon decoupling, on sub-sound-horizon scales, and
- during Λ domination, on all scales.

These two distinct contributions are usually called the Early Integrated Sachs-Wolfe (EISW) and Late Integrated Sachs-Wolfe (LISW) effects. One can show that both effects decrease with wavelength, for geometrical reasons. Hence the EISW effect is maximal for scales crossing the sound horizon just at the time of photon decoupling, while the LISW effect is maximal for the largest observable scales today. In multipole space, this means that the EISW effect contributes mainly to the scale of the first peak, i.e. to $l \sim 200$, while the LISW contributes mainly to the smallest multipoles $l = 2, 3, 4$, etc. The two ISW contributions are drawn on figure 146. The EISW effect enhances the first peak, while the LISW effect tilts the Sachs-Wolfe plateau even if the primordial spectrum is exactly scale invariant.

Reionization effect. The last effect that we omitted to describe is that of reionization. We have seen in section 12.9.2 that at small redshift ($z \simeq 10$), the reionization of the universe produces a small secondary bump in the visibility function, corresponding physically to a small

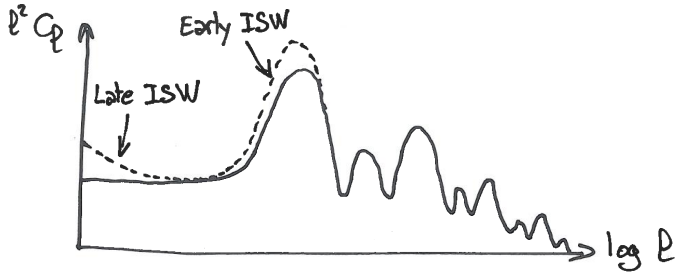


Figure 146: ISW contribution to the temperature spectrum.

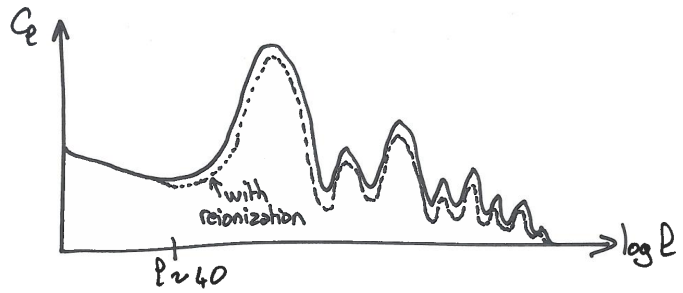


Figure 147: Impact of reionization on the temperature spectrum.

probability for CMB photons to rescatter at late times. This rescattering will tend to smooth out any temperature anisotropy pattern. Hence, reionization lowers the overall amplitude of the C_l', but only a small amount (by approximately 15%). Note that the suppression of power is not uniform over the whole multipole range: smoothing effects cannot reach the largest observable scales (corresponding to the smallest values of l). Hence the effect of reionization is step-like shaped, and saturates for l of the order of 40 or so (as illustrated on figure 147).

Parameter dependence of the temperature spectrum

We summarized in the last section the various physical effect contributing to the shape of the CMB temperature spectrum C_l. We will now recapitulate the effect of the various cosmological parameters on the C_l's, within the framework of the minimal ΛCDM model.

This model assumes zero spatial curvature, and a power-law primordial spectrum of scalar perturbations:

$$\mathcal{P}_\mathcal{R}(k) = A_\mathrm{s}(k/k_*)^{n_\mathrm{s}-1} \ , \tag{258}$$

where k_* is an arbitrary fixed pivot scale, A_s is the spectrum amplitude at this scale, and n_s is called the scalar tilt (the exponent is chosen to be $n_\mathrm{s}-1$ rather than just n_s for historical reasons; with such notations, a scale-invariant spectrum corresponds to $n_\mathrm{s}=1$).

We recall that the Hubble parameter today, H_0, can be expressed in terms of a dimensionless reduced Hubble parameter h:

$$H_0 \equiv 100\,h\,\mathrm{km.s^{-1}.Mpc^{-1}} \tag{259}$$

The physical energy density of a given component x today can be expressed in terms of a dimensionless parameter $\omega_x \equiv \Omega_x h^2$:

$$\rho_x^0 = \Omega_x \rho_\mathrm{crit}^0 = \Omega_x \frac{3H_0^2}{8\pi G} = \beta\,\omega_x \tag{260}$$

where $\beta = \frac{3(H_0/h)^2}{8\pi G}$ is a fixed number, with the dimension of an energy per volume.

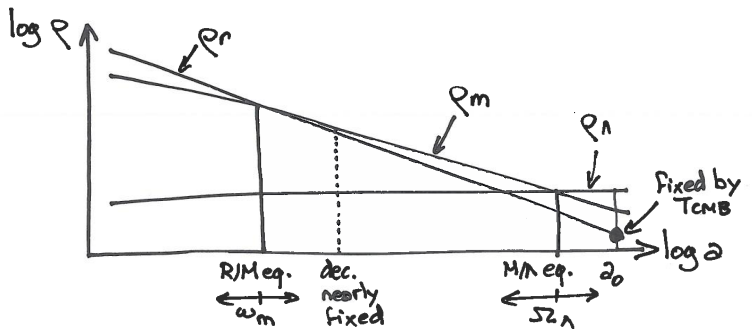


Figure 148: In the minimal ΛCDM model, evolution of the background density of radiation, matter and cosmological constant as a function of the scale factor.

The six free parameters of the minimal ΛCDM model can be chosen to be

$$\{A_{\rm s}, n_{\rm s}, \omega_{\rm b}, \omega_{\rm m}, \Omega_\Lambda, \tau_{\rm reio}\}, \tag{261}$$

where $\tau_{\rm reio}$ is the optical depth to recombination, which is non-zero because of reionization in the recent universe. At first order, this is the only parameter that one needs to introduce for describing reionization. For a very accurate description, one should introduce other parameters specifying the full reionization history, but these extra parameters are very difficult to probe experimentally. Hence, they are usually not specified.

The above parameter basis specifies the baryon density $\omega_{\rm b}$, the total non-relativistic (baryons + CDM) matter density $\omega_{\rm m}$, and the fractional density of cosmological constant Ω_Λ. The photon density is implicitly assumed to match the measured value of the CMB temperature ($T = 2.725$ K implies $\omega_\gamma \sim 2.10^{-5}$). Neutrinos are assumed to be still relativistic today, with an abundance relative to photons given by the prediction of the standard neutrino decoupling model[120]. Finally, we did not include the parameter H_0 (or h) in the parameter basis (261). Given that we are assuming a flat universe, h can be inferred from other parameters in (261): $h = \sqrt{\omega_{\rm m}/(1-\Omega_\Lambda)}$.

The parameters basis (261) is just one particular choice. Many other bases would be valid, for instance, $\{A_{\rm s}, n_{\rm s}, \omega_{\rm b}, \omega_{\rm cdm}, H_0, \tau_{\rm reio}\}$). The choice of (261) is dictated by purely pedagogical considerations: we will show that the shape of the CMB spectrum can easily be related to these parameters.

We summarize in Fig. 148 the evolution of background densities in the minimal ΛCDM model. The normalization of the radiation density ρ_r is fixed by the CMB temperature and by the standard neutrino decoupling model. The normalization of the matter density $\rho_{\rm m}$ and of the cosmological constant density ρ_Λ is given respectively by the parameters $\omega_{\rm m}$ and Ω_Λ. The scale factor at radiation/matter equality is given by

$$\rho_r = \rho_{\rm m} \quad \Longrightarrow \quad \omega_r \left(\frac{a_0}{a_{\rm eq}}\right)^4 = \omega_{\rm m} \left(\frac{a_0}{a_{\rm eq}}\right)^3 \quad \Longrightarrow \quad \frac{a_{\rm eq}}{a_0} = \frac{\omega_r}{\omega_{\rm m}} \, . \tag{262}$$

Since ω_r is considered as fixed, the redshift of equality is controlled by $\omega_{\rm m}$ only. The scale factor at matter/Λ equality is given by

$$\rho_{\rm m} = \rho_\Lambda \quad \Longrightarrow \quad \omega_{\rm m} \left(\frac{a_0}{a_\Lambda}\right)^3 = \Omega_\Lambda h^2 \quad \Longrightarrow \quad \frac{a_\Lambda}{a_0} = \left(\frac{1-\Omega_\Lambda}{\Omega_\Lambda}\right)^{1/3} \, , \tag{263}$$

[120] In this course, for simplicity, we do not discuss the effect of neutrinos; this effect is far from being negligible, but since we consider the abundance of neutrinos as fixed, and their mass as irrelevant, we can explain the effect of other free parameters without taking neutrinos into account. Neutrino effects are described in Ref. [1].

so it is controlled by Ω_Λ only. The scale factor at photon decoupling has a small (logarithmic) dependence on $\omega_{\rm b}$, that we can neglect — we will consider that decoupling takes place at a fixed temperature, and hence a fixed scale factor and redshift.

Among the various physical effects described in the previous section, we can identify eight independent leading effect. We summarize them in Table ??, and show by which parameters they are governed.

Independent leading effects controlling the shape of the CMB temperature power spectrum C_l in the minimal ΛCDM model.

	Effect	Relevant quantity	Parameter
(C1)	**Peak scale**	$\theta_{\rm peak} = \dfrac{\pi}{l_{\rm peak}} \sim \dfrac{d_{\rm s}\|_{\rm dec}}{d_{\rm a}\|_{\rm dec}}$	$\leftarrow \omega_{\rm m}, \omega_{\rm b}$ $\leftarrow \Omega_\Lambda, \omega_{\rm m}$
(C2)	**Odd/even peak amplitude ratio**	$R\|_{\rm dec}$	$\omega_{\rm b}$
(C3)	**Overall peak amplitude**	$\frac{a_{\rm dec}}{a_0}$	$\omega_{\rm m}$
(C4)	**Damping enveloppe**	$\theta_{\rm d} = \dfrac{\pi}{l_{\rm d}} = \dfrac{a_{\rm dec} r_{\rm d}\|_{\rm dec}}{d_{\rm a}\|_{\rm dec}}$	$\leftarrow \omega_{\rm m}, \omega_{\rm b}$ $\leftarrow \Omega_\Lambda, \omega_{\rm m}$
(C5)	**Global amplitude**	$\mathcal{P}_\mathcal{R}(k_*)$	$A_{\rm s}$
(C6)	**Global tilt**	$\frac{d\log\mathcal{P}_\mathcal{R}}{d\log k}$	$n_{\rm s}$
(C7)	**Additional plateau tilting (LISW)**	$\frac{a_\Lambda}{a_0}$	Ω_Λ
(C8)	**Amplitude for $l \geq 40$ only**	$\tau_{\rm reio}$	$\tau_{\rm reio}$

Below, we give more details on these effects.

(C1) Since all peaks in multipole space correspond to the harmonics of a single correlation length in real space, the scale of the peak is controlled (in good approximation) by a single number $l_{\rm peak}$. It depends on the ratio of the sound horizon at decoupling by the angular diameter distance to decoupling. The first quantity depends on the evolution prior to decoupling, and in particular on the expansion history and sound speed. Hence it depends on $\omega_{\rm m}$ (governing the time of equality) and $\omega_{\rm b}$ (governing $c_{\rm s}^2$ as a function of a). The second quantity depends on the expansion and geometry of the universe after decoupling, i.e. on Ω_Λ and H_0, or in our parameter basis Ω_Λ and $\omega_{\rm m}$.

(C2) When studying the SW contribution to the C_l's, we have seen that the asymmetry between the amplitude of odd and even peaks depends on the shift of the zero-point of acoustic oscillations by a term $-R\psi$. The value of the ratio R at decoupling is governed in our parameter basis by $\omega_{\rm b}$.

(C3) A shift in the time of radiation/matter equality affects the amplitude of all peaks for two reasons: it controls the duration of the intermediate stage between equality and decoupling, during which acoustic oscillations are damped by baryonic effects, and the EISW. Both effects go in the same direction. If equality takes place earlier, there is less time for damping (hence all peaks are increased). The metric fluctuations are also less stabilized at decoupling, and the EISW is larger (hence the first peak is increased even more).

(C4) Diffusion damping near the time of recombination controls the envelope of the peaks (the function $\exp[-(l/l_{\rm d})^2]$ suppresses them, starting essentially from the third one). It depends on the ratio of the damping scale at decoupling by the angular diameter distance to decoupling. The first quantity depends on the Thomson scattering rate prior to decoupling, and in particular on $\omega_{\rm m}$ (governing the value of conformal time at equality) and $\omega_{\rm b}$ (governing the ionization fraction as a function of a). The second quantity depends on the expansion and geometry of the universe after decoupling, i.e. on Ω_Λ and H_0, or in our parameter basis Ω_Λ and $\omega_{\rm m}$. The parameter dependence of effects (C1) and (C4) could be thought to be similar: in fact, it is not, because the sound horizon and the diffusion scale depend on very different combinations of $\omega_{\rm m}$ and $\omega_{\rm b}$.

(C5) The global amplitude of the C_l's depends trivially on that of the primordial spectrum, fixed by $A_{\rm s}$.

(C6) The global slope of the C_l's depends trivially on the tilt of the primordial spectrum, fixed by $n_{\rm s}$.

(C7) The LISW effect tilts the Sachs-Wolfe plateau (on top of the effect of $n_{\rm s}$). The plateau is more lifted at small l's if Λ domination is longer, i.e. if metric fluctuations decay during a larger amount of time. Hence this effect is enhanced by large values of Ω_Λ.

(C8) The C_l amplitude is suppressed if reionization takes place early, i.e. if $\tau_{\rm reio}$ is large, but without affecting the largest scales (small l's).

We see that in the framework of the minimal ΛCDM model, six parameters control eight distinct physical effects with different impacts on the C_l's. This suggests that an accurate enough measurement of the temperature spectrum is sufficient for fixing the six parameters of the cosmological model describing our universe, at least if the data is consistent with ΛCDM. This conclusion is roughly correct, but must be refined with some words of caution. Indeed, we remember that for small l's, cosmic variance is large, so that the average C_l's of the "true model" describing our universe cannot be measured precisely, even in the case of an ideal experiment. However, two of the previous effects **(C1)** - **(C8)** can only affect the smallest multipoles:

- effect **(C7)** affects only the Sachs-Wolfe plateau,
- a combination of effects **(C5)** and **(C8)**, corresponding to the product $e^{-2\tau_{\rm reio}} A_{\rm s}$, controls the global amplitude for $l \gg 40$, but a variation of both $\tau_{\rm reio}$ and $A_{\rm s}$ with the previous product being kept fixed would only affect the smallest multipoles.

Hence, the measurement of the CMB temperature spectrum is sufficient for constraining the six parameters of the ΛCDM model, but with a relatively large error bar for Ω_Λ and for a particular combination of $\tau_{\rm reio}$ and $A_{\rm s}$. In the next section, we will say a few words on the measurement of CMB polarization, which allows to better constrain reionization: polarization data allow to remove the degeneracy between $\tau_{\rm reio}$ and $A_{\rm s}$. Instead, the error bar on Ω_Λ can only be improved by combining CMB data with other cosmological probes (e.g. supernovae luminosity or large scale structure data).

A quick word on polarization

The Bolztmann equation presented in Eq. (217) was a bit over-simplified. We did as if the only degree of freedom describing photons with a blackbody spectrum was their temperature. In fact, photons are described by more degrees of freedom, called the Stokes parameters, involving also their polarization. In Eq. (217), the Thomson scattering rate was integrated over polarization parameters, but in reality some polarized correction terms are present.

Well before decoupling, photons remain unpolarized on average. Indeed, we have seen in section 12.9.2 that, throughout the tight-coupling regime and in the frame comoving with the fluid (i.e. such that $\theta_{\rm b} = \theta_\gamma = 0$), the photon temperature is isotropic is every point. This isotropy (resulting from frequent interactions) implies that photons acquire no net polarization patterns when they scatter over electrons.

Instead, at the approach of decoupling, when Thomson scattering becomes inefficient, the photon temperature is no longer isotropic in the frame comoving with the electrons. This means that a given electron will scatter simultaneously some hotter photons coming from one direction, and some colder photons coming from another direction. What is important for polarization is the quadrupolar component of the the temperature distribution in each point. This component starts from zero and grows at the approach of decoupling. When it becomes significant, photon scattering leads to a net linear polarization. Hence, today, CMB photons have a different polarization amplitude and orientation in each direction of the sky.

The map of CMB temperature anisotropies is a scalar map: it can be represented with a one-dimensional color code. The map of CMB polarization can be represented with sticks of different size and orientation in different points of the map. Roughly speaking, the size and orientation of the sticks can be related to the magnitude and orientation of the quadrupole anisotropy in each point of the last scattering surface. Since these quadrupolar patterns reflect variations of temperature in the region of the last scattering surface, it is clear that there is a non-zero correlation between temperature and polarization maps. Still, both maps contain some independent information.

In general, a vector field can be decomposed into a gradient and a curl component, like the electric and magnetic fields. Similarly, CMB polarization maps can be expanded in two scalar maps called, by analogy, E-type and B-type polarization maps.

One can show that primordial scalar perturbations can produce both temperature anisotropies and E-type polarization anisotropies. If the early universe only features Gaussian scalar perturbations on cosmological scales, all the information contained in CMB maps is encoded in the temperature power spectrum C_l^{TT}, the E-type polarization power spectrum C_l^{EE}, and the cross-correlation power spectrum C_l^{TE}. The later really contains additional independent information, because polarization patterns are only partially correlated with temperature patterns (in mathematical terms, $[C_l^{TE}]^2 \leq C_l^{TT} C_l^{EE}$).

We will mention in the next section the possibility that a significant amount of tensor perturbations are produced in the primordial universe. Such perturbations can also generate B-type polarization. If this is the case, all the information is encoded in C_l^{TT}, C_l^{EE}, C_l^{TE} and C_l^{BB}. We did not include in this list the cross-correlation spectra C_l^{TB} and C_l^{EB}, because the parity symmetry imposes that they should vanish (at the level of primary anisotropies).

Mapping the CMB polarization is interesting because it contains additional information with respect to temperature anisotropies.

First, reionization has a very distinct effect on the polarization spectrum. The small fraction of photons rescattered around the time of reionization acquire additional polarization that will show up in the form of a peak in the small-l branch of C_l^{EE} (see figure 149) and C_l^{TE}. We have seen in the previous section that with temperature data only, cosmic variance limits our ability to measure $\tau_{\rm reio}$, and *a fortiori* to constrain additional parameters describing the reionization history. The effect of reionization in the small-l branch of C_l^{EE} and C_l^{TE} is so big that despite of cosmic variance, $\tau_{\rm reio}$ can be well measured using polarization data.

So the effect of reionization, described before as **(C8)**, is very different for temperature and polarization. Instead, the other effects **(C1)** - **(C7)** have a qualitatively similar impact on C_l^{EE} and C_l^{TE}. Still, measuring polarization is interesting, because temperature anisotropies feature a particular combination of the SW, Doppler and ISW effects, while polarization provides information on the quadrupole Θ_2 on the last scattering surface, which is correlated with

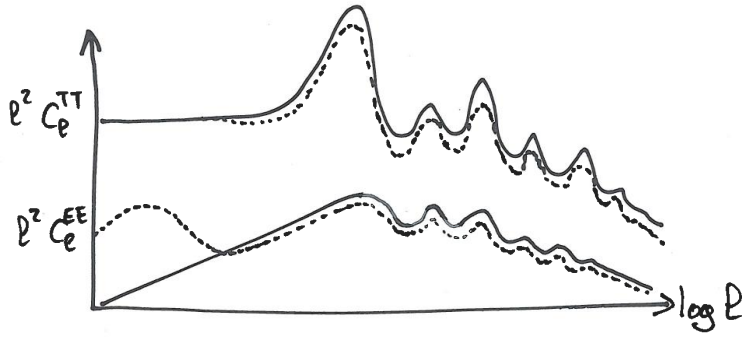


Figure 149: Typical shape of the temperature and E-type polarization power spectra in a ΛCDM model without (solid lines) or with (dotted lines) reionization of the universe at small redshift.

the SW and Doppler terms, but probing a different combination of them. Hence the role of polarization measurements is to remove some of the parameter degeneracies appearing in the analysis of temperature data.

Note that in the full Boltzmann equation, the equation of motion of temperature and polarization degrees of freedom are coupled. In the last sections, we neglected such a coupling. In fact, the impact of polarization on the evolution of temperature anisotropies is very small. Hence our qualitative discussion of the various effects affecting CMB temperature remains valid.

A quick word on tensors

We have seen in section 12.9.1 that tensor modes are related to the traceless divergenceless components of the metric and stress-energy tensor, obeying to

$$\sum_i \delta T_{ii} = 0 \qquad \text{and} \qquad \forall j, \quad \sum_i \partial_i \delta T_{ij} = 0 \ . \tag{264}$$

Since these relations impose four constraints on the six degrees of freedom of the 3×3 symmetric matrix δT_{ij} (or δg_{ij}), there are two tensor degrees of freedom in $g_{\mu\nu}$ and $T_{\mu\nu}$.

The two degrees of freedom in δg_{ij} are the two degrees of polarization of gravitational waves, that can propagate even in the vacuum. In the presence of matter with a non-diagonal stress tensor δT_{ij}, gravitational waves can be seeded by the tensor components δT_{ij}.

If tensor perturbations are sufficiently large at recombination or later, they can generate CMB temperature and polarization anisotropies: one can show that they interact with CMB photons, and produce effects similar to the SW and ISW ones. As mentioned in the previous section, gravitational waves can even seed B-type polarization.

The stress-energy tensor is diagonal for perfect fluids, and negligible for pressureless components like CDM. Hence, during radiation and matter domination, tensor perturbations can only be seeded by decoupled neutrinos and/or decoupled photons. However, neutrino and photon tensor modes are far too small for generating a detectable amount of CMB anisotropies. However, in the early universe, other mechanisms may produce gravitational waves on scales sufficiently large to be observable in the CMB. The most famous one is related to inflation. During inflation, quantum fluctuations of the metric can excite primordial tensor perturbations at a level that will produce a detectable signature in CMB maps. The amplitude of this signal is directly proportional to the energy scale of inflation. If this scale is large enough, tensors can contribute to C_l^{TT}, but only on small l's, because gravitational waves decay quickly inside the Hubble radius. Hence, one way to detect tensors would be through a small distortion of the

C_l^{TT} spectrum shape for $l \leq 100$ (i.e. for scales that are equal or larger than the Hubble radius at the time of decoupling).

However, if the tensor signal is small with respect to the error bars associated to cosmic variance, it will remain undetectable. In that case, further sensitivity could be obtained by measuring C_l^{BB}: in absence of tensors, C_l^{BB} would vanish, so even if the tensor amplitude is very small it can still dominate the C_l^{BB} signal. Unfortunately, this is true only up to some extent, because secondary anisotropies (in particular, those generated by weak lensing) produce a non-zero C_l^{BB} that could mask the primary tensor anisotropy spectrum.

It is very challenging for CMB experiments to reach the sensitivity level required for detecting a C_l^{BB} signal, even for that generated by weak lensing. Current limits on the tensor primordial spectrum (and on the energy scale of inflation) mainly come from the observation of C_l^{TT} at small l by WMAP. The sensitivity of Planck to tensors will also mainly come from temperature. Future experiments dedicated to CMB polarization will improve the sensitivity to B-type polarization and obtain more precise bounds, until they reach the theoretical limit set by the lensing contamination.

12.9.3 Matter power spectrum by Lesgourgues

Definition

The total energy perturbation in the universe can be expanded as

$$\delta\rho_{\rm tot} = \delta\rho_\gamma + \delta\rho_{\rm b} + \delta_{\rm cdm} + \delta\rho_\nu \,(+\,\delta\rho_{\rm de} + ...) \tag{265}$$

where $\delta\rho_{\rm de}$ refers to possible Dark Energy (DE) perturbations, and the three dots for extra relics. In the minimal ΛCDM model, only the first four components are present.

Many Large-Scale Structure (LSS) observables are related to the power spectrum of $\delta\rho_{\rm tot}$, at different wavenumbers and redshifts. This is the case of the galaxy and of the halo correlation function, of the cluster mass function, of CMB lensing, of the cosmic shear spectrum, etc.

All these observations probe the power spectrum during matter or Λ/DE domination, when photons are subdominant and $\delta\rho_\gamma$ can be neglected. If neutrinos are still ultra-relativistic today, $\delta\rho_\nu$ can also be neglected (in this course, we do not have time to discuss the impact of small neutrino masses, for which we refer the reader to Ref. [1]). If the acceleration of the universe is caused by a cosmological constant, there is no term $\delta\rho_{de}$. More generally, most Dark Energy models would predict a negligible amount of DE perturbations. In summary, in a wide category of cosmological scenarii including ΛCDM, we can use $\delta\rho_{\rm tot} \simeq \delta_{\rm m} \equiv \delta\rho_{\rm b} + \delta_{\rm cdm}$. Hence, in the context of LSS observations, it is customary to refer to the power spectrum of the non-relativistic matter fluctuation $\delta_{\rm m}$, defined as

$$\delta_{\rm m} = \frac{\delta\rho_{\rm m}}{\bar\rho_{\rm m}} = \frac{\delta\rho_{\rm b} + \delta\rho_{\rm cdm}}{\bar\rho_{\rm b} + \bar\rho_{\rm cdm}} \,, \tag{266}$$

which is indistinguishable from the ratio $\delta\rho_{\rm tot}/\bar\rho_{\rm m}$. Only in models with large dark energy perturbations or modifications of Einstein gravity, the two quantities might be different, and special care about the definition of the matter power spectrum is needed.

The matter power spectrum $P(z,k)$ of $\delta_{\rm m}$ is defined like in section 12.9.1:

$$\langle \delta_{\rm m}(z,\vec{k}) \delta^*_{\rm m}(z,\vec{k}') \rangle = \delta_D(\vec{k} - \vec{k}') \; P(z,k) \,. \tag{267}$$

Here we used the redshift as a time variable, but we could have indifferently used proper or conformal time. We have seen in section 12.9.1 that for Gaussian initial conditions and as long as perturbations are linear, the power spectrum at a given time can be written as the product

of the primordial spectrum by the square of the relevant transfer function, in our case $\delta_{\rm m}(z,k)$. Sticking to the same conventions as in section 12.9.1 and assuming a power-law primordial spectrum like in section 12.9.2, this gives

$$P(z,k) = \frac{2\pi^2}{k^3} A_{\rm s} \left(\frac{k}{k_*}\right)^{n_{\rm s}-1} \delta_{\rm m}^2(z,k) \ . \tag{268}$$

Hence, by studying qualitatively the evolution of the transfer function $\delta_{\rm m}(z,k)$, we will get some insight on the cosmological information encoded in the matter power spectrum.

Transfer function evolution

CDM dominated universe. In order to simplify the presentation, let us first assume that we leave in a ΛCDM universe with a negligible amount of baryons: $\Omega_{\rm b} \ll \Omega_{\rm cdm}$ and $\delta_{\rm m} \simeq \delta_{\rm cdm}$. In section 12.9.2, we wrote the master equation governing the evolution of photon perturbations during the tightly-coupled regime. Similarly, by combining the continuity and Euler equation of CDM perturbations, we can write here a master equation for $\delta_{\rm cdm}$, actually valid in all regimes:

$$\delta''_{\rm cdm} + \frac{a'}{a}\delta'_{\rm cdm} = -k^2\psi + 3\phi'' + 3\frac{a'}{a}\phi' \ . \tag{269}$$

In an expanding universe, the clustering rate depends on the expansion rate: expansion increases distances, weakens gravitational forces, and slows down clustering processes. In the above equation, this is accounted by the second term, often called the Hubble friction term. On the right-hand side, the first term represents gravitational forces, and the last two terms account for dilation effects.

On super-Hubble scales, we have seen that in the Newtonian gauge, adiabatic ICs predict constant density fluctuations $\delta_{\rm cdm}$. To be precise, ϕ and $\delta_{\rm cdm}$ vary on super-Hubble scales only when the total equation of state of the universe changes (i.e. around the time of radiation/matter equality, and during Λ domination). Instead, they remain constant on those scale during the radiation and matter dominated regime.

Inside the Hubble radius, we can neglect dilation terms, and replace the gravitational potential term:

$$\delta''_{\rm cdm} + \frac{a'}{a}\delta'_{\rm cdm} - \frac{3}{2}\left(\frac{a'}{a}\right)^2 \Omega_{\rm cdm}(a)\ \delta_{\rm cdm} = 0 \ , \tag{270}$$

where $\Omega_{\rm cdm}(a)$ is the fraction of the critical density coming from CDM at a given value of time (or of the scale factor). This equation is often called the Mészáros equation. It can be obtained by combining Eq. (269) with the (00) component of the Einstein equation (or its Poisson limit) and the Friedmann equation. The careful reader might have noticed something suspicious: the gravitational force term $k^2\psi$ has been eliminated in favor of $\delta_{\rm cdm}$, while the first Einstein equation (or its Poisson limit) involves the total density fluctuation. Hence, shouldn't Eq. (270) feature also δ_γ, $\delta_{\rm b}$ and δ_ν? Actually, it turns out that the above equation is a really good approximation for the CDM equation of evolution in all regimes, under our assumption $\Omega_{\rm b} \ll \Omega_{\rm cdm}$. It applies even during radiation domination, when photon fluctuations are potentially large. In this course, we do not have enough time for justifying this point, but we refer the reader to Weinberg[3, 4] for a detailed explanation.

During radiation domination, the Friedmann equation gives $a \propto \eta$, and $\Omega_{\rm cdm}$ is much smaller than one. Hence we can neglect the last term in the Mészáros equation, and find that the two solutions are $\delta_{\rm cdm} =$ constant and $\delta_{\rm cdm} \propto \log\eta$. Hence CDM fluctuations grow logarithmically. During matter domination, $a \propto \eta^2$ and $\Omega_{\rm cdm} \simeq 1$, so the solutions are $\delta_{\rm cdm} \propto \eta^2$ and $\delta_{\rm cdm} \propto \eta^{-3}$. Then, CDM fluctuations grow quadratically with η. (These are only asymptotic solutions, but the Mészáros equation can actually be solved analytically at all times). During Λ domination,

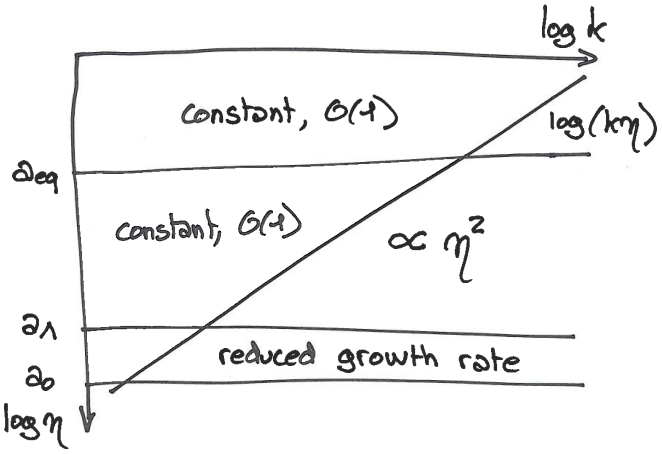


Figure 150: Qualitative evolution of the transfer function $\delta_{\rm cdm}(\eta, k)$ (normalized as usual to $\mathcal{R}(k) = 1$ at initial time) in different regimes: during radiation domination, matter domination, Λ domination, and on super/sub-Hubble scales.

the function $a(\eta)$ is more complicated, and $\Omega_{\rm cdm}$ decreases. With a bit of work, one can show that $\delta_{\rm cdm}$ grows at a smaller rate than during matter domination (i.e. slowlier than η^2), and that the reduction of the growth rate does not depend on k.

In summary, during radiation domination, $\delta_{\rm cdm}(\eta, k)$ is constant on super-Hubble scales and grows logarithmically on sub-Hubble scales. A more precise calculation would show that up to a numerical factor of order one, $\delta_{\rm cdm}$ is given on sub-Hubble scales by $\delta_{\rm cdm}(\eta, k) = \log(k\eta)$. During matter domination, $\delta_{\rm cdm}(\eta, k)$ is still constant on super-Hubble scales, and grows like η^2 on sub-Hubble scales. Finally, during Λ domination, it grows more slowly. These different behaviors are reported in Fig. 150.

This simple discussion is sufficient for understanding the shape of the matter power spectrum at different time. Like in a cartoon, Fig. 151 shows this shape at four different times: at some initial time when all relevant modes are super-Hubble; at radiation/matter equality; at matter/Λ equality; and today. Let us comment these plots. We first need to define the comoving wavenumbers corresponding to wavelengths crossing the Hubble radius at the time of radiation/matter equality, of matter/Λ equality, and today:

$$k_{\rm eq} = a_{\rm eq} H_{\rm eq} \ , \qquad k_\Lambda = a_\Lambda H_\Lambda \ , \qquad k_0 = a_0 H_0 \ . \tag{271}$$

We can now review the evolution of $P(k)$ with respect to time, following the same steps as in Fig. 151.

1. At initial time, if we assume a scale-invariant spectrum with $n_{\rm s} = 1$, we see from Eq. (268) that $P(k) \propto k^{-3}$, with an amplitude given by $A_{\rm s}$.

2. During radiation domination, modes grow logarithmically inside the Hubble radius, like $\log(k\eta)$. At equality, super-Hubble modes ($k_{\rm eq}\eta \ll 1$) are still shaped like at initial time, while sub-Hubble modes ($k_{\rm eq}\eta \gg 1$) have been enhanced by a factor $[\delta_{\rm cdm}(\eta_{\rm eq}, k)/\delta_{\rm cdm}(\eta_{\rm ini}, k)]^2 \simeq [\log(k\eta_{\rm eq})]^2$. The two asymptots of $P(k)$ are then given by k^{-3} for $k \ll k_{\rm eq}$ and $k^{-3}[\log(k)]^2$ for $k \gg k_{\rm eq}$.

3. At the end of matter domination, when $a = a_\Lambda$, modes still outside the Hubble radius keep being shaped like at initial time. This concerns all modes with $k \ll k_\Lambda$. Modes $k \gg k_{\rm eq}$ have been amplified during matter domination by a factor $[\delta_{\rm cdm}(\eta_\Lambda, k)/\delta_{\rm cdm}(\eta_{\rm eq}, k)]^2 \simeq$

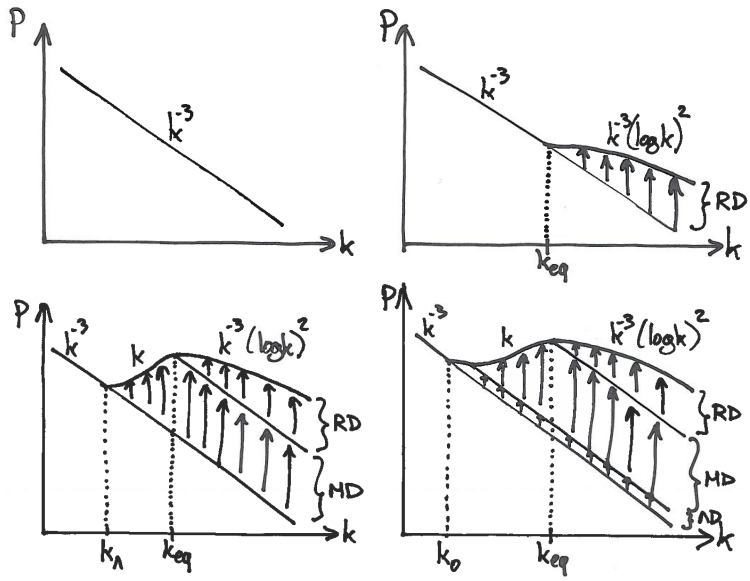


Figure 151: Shape of the matter power spectrum $P(k)$ (log-log scale) at four different times: *(upper left)* when initial conditions are imposed (and all wavenumbers are super-Hubble); *(upper right)* at radiation/matter equality (arrows show the logarithmic growth during radiation domination); *(lower left)* at matter/Λ equality (lower set of arrows show the growth during matter domination); *(lower right)* today (lower set of arrows show the growth during Λ domination).

$(\eta_\Lambda/\eta_{\rm eq})^4$. This factor does not depend on k and preserves the shape of the power spectrum on those scales. Finally, intermediate modes entering the Hubble scale during matter domination have been amplified by $(\eta_\Lambda/\eta_*)^4$, where η_* is their time of Hubble crossing, given approximately by $\eta_* = 1/k$. Hence they have been amplified by $(k\eta_\Lambda)^4$. Putting all these informations together, we see that the spectrum has three branches, scaling respectively like:

- $P(k) \propto k^{-3}$ for $k < k_\Lambda$,
- $P(k) \propto k^{-3}k^4 = k$ for $k_\Lambda < k < k_{\rm eq}$,
- $P(k) \propto k^{-3}(\log k)^2$ for $k > k_{\rm eq}$.

4. During Λ domination, $\delta_{\rm cdm}(k,\eta)$ grows more slowly than η^2, but it still grows at the same rate for all sub-Hubble modes. So the shape of the power spectrum today is unaltered by this stage, and given by:

   - $P(k) \propto k^{-3}$ for $k < k_0$,
   - $P(k) \propto k^{-3}k^4 = k$ for $k_\Lambda < k < k_{\rm eq}$,
   - $P(k) \propto k^{-3}(\log k)^2$ for $k > k_{\rm eq}$.

   We do not enter into details for the small range of modes obeying $k_0 < k < k_\Lambda$.

All this discussion was carried under the assumption of a scale-invariant primordial spectrum. If $n_{\rm s} \neq 1$, the above shape should simply be rescaled by $k^{n_{\rm s}-1}$, and the three branches of the power spectrum are given by:

- $P(k) \propto k^{n_{\rm s}-4}$ for $k < k_0$,
- $P(k) \propto k^{n_{\rm s}}$ for $k_\Lambda < k < k_{\rm eq}$,
- $P(k) \propto k^{n_{\rm s}-4}(\log k)^2$ for $k > k_{\rm eq}$.

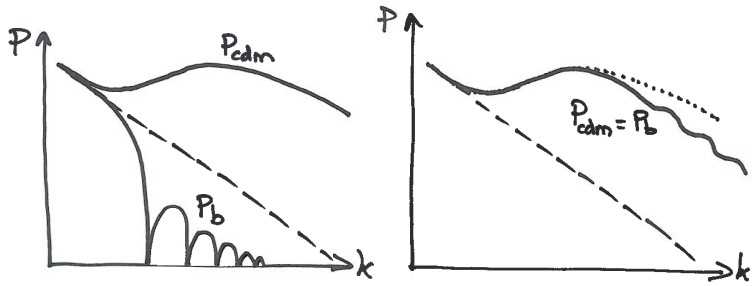


Figure 152: *(Left)* Shape of the baryon and CDM power spectra just before the baryon drag time, compared to the shape of the primordial spectrum (dashed line). *(Right)* Shape of the matter power spectrum well after the baryon drag time, when baryons and CDM perturbations have reached gravitational equilibrium, compared to the shape of the primordial spectrum (dashed lines) and to the spectrum of a CDM-dominated universe (dotted line).

This closes the presentation of the shape of $P(z,k)$ in the limit $\Omega_{\rm b} \ll \Omega_{\rm cdm}$. What remains to be seen is the impact of a non-negligible baryon fraction on the power spectrum.

Baryon corrections. We must define a new important time in the evolution of the universe: the baryon drag time. The photon decoupling time was defined through the maximum of the photon visibility function $g(\eta)$. But there are many more photons than baryons in the universe. Hence, for some amount of time after photon decoupling, baryons keep tracking photon perturbations: in other words, the baryon population decouples later than the photon population (seen as a whole).

Until the baryon drag time, we know that $\delta_{\rm b} = \frac{3}{4}\delta_\gamma$. After that time, baryons do not experience significant Thomson scattering anymore. They only feel gravity, and collapse in gravitational potential wells.

We know qualitatively the behavior of the baryon transfer function $\delta_{\rm b}(\eta,k)$ just before baryon drag, since in the CMB section we have studied the behavior of the photon transfer function. We know that $\delta_{\rm b} = \frac{3}{4}\delta_\gamma$ is constant on super-sound-horizon scales, experiences stationary oscillations on sub-sound-horizon scales during radiation domination, and finally damped oscillations on sub-sound-horizon scales during matter domination. Moreover, we expect that at any time — and particularly at the approach of decoupling — the relation $\delta_{\rm b} = \frac{3}{4}\delta_\gamma$ breaks on very small wavelengths comparable to the mean free path of baryons in the imperfect baryon-photon fluid, but here we will not discuss such small scales.

The behavior of the CDM transfer function $\delta_{\rm cdm}(\eta,k)$ before baryon drag is still given by the Mészáros equation: deep inside the sound horizon, baryons are much less clustered than CDM, so that CDM is almost self-gravitating.

In the left plot of Fig. 152, we sketch the qualitative behavior of the individual power spectrum of baryons and CDM just before baryon drag ($P_{\rm b}(k)$ and $P_{\rm cdm}(k)$ are defined as in Eq. (268), with $\delta_{\rm m}$ replaced either by $\delta_{\rm b}$ or $\delta_{\rm cdm}$).

After baryon drag, baryons collapse in potential gravitational wells. Since $\delta_{\rm b}$ grows, CDM will start to feel the gravity of baryons. Finally, because CDM and baryons are two collisionless species feeling the same gravitational forces, they will equilibrate with $\delta_{\rm b} = \delta_{\rm cdm}$. The gravitational potential is then related to this common value by the (00) Einstein equation (or its Poisson sub-Hubble limit). This is not true however on very small scales, for which the baryon pressure cannot be neglected, but here we do not discuss such small scales.

Hence, after a quick relaxation period, the two power spectra of baryons and CDM are equal to each other, $P_{\rm b}(k) = P_{\rm cdm}(k) = P(k)$. In order to relate this common power spectrum to the individual power spectrum of baryons and CDM before baryon drag, one must follow a

Independent leading effects controlling the shape of the matter power spectrum $P(k)$ in the minimal ΛCDM model.

	Effect	Relevant quantity	Parameter
(P1)	**scale of the maximum**	$k_{\rm eq}$	$\omega_{\rm m}, \Omega_\Lambda$
(P2)	**slope for $k \gg k_{\rm eq}$ and BAO amplitude**	$\Omega_{\rm b}/\Omega_{\rm cdm}$	$\omega_{\rm b}, \omega_{\rm m}$
(P3)	**BAO scale**	$r_{\rm s}(\eta_{\rm drag})$	$\omega_{\rm b}, \omega_{\rm m}$
(P4)	**Global amplitude**	**amplitude of primordial spectrum and duration of ΛD**	$A_{\rm s}, \Omega_\Lambda$
(P5)	**Global tilt**	**tilt of primordial spectrum**	$n_{\rm s}$

matching process described in details in Einsteinstein and Hu [661]. Intuitively, the matching depends on the relative weight of baryons and CDM, i.e. on the ratio $\Omega_{\rm b}/\Omega_{\rm cdm}$. In the limit $\Omega_{\rm b} \ll \Omega_{\rm cdm}$, what we wrote at the beginning of this section applies. In the limit $\Omega_{\rm b} \gg \Omega_{\rm cdm}$, the power spectrum is very suppressed with respect to the CDM-dominated case, with a much more negative slope on average, and some large oscillations corresponding to photon-baryon acoustic waves before decoupling. Finally, if $\Omega_{\rm b}$ is a bit smaller than $\Omega_{\rm cdm}$ but not negligible (as it is the case in our universe), the power spectrum departs slightly from a pure CDM one, with a smooth step-like suppression plus small oscillatory patterns, which are the smoking gun in the recent universe of baryon-photon acoustic oscillations happening before photon decoupling. This is illustrated on the right plot of Fig. 152. These oscillations are called Baryon Acoustic Oscillations (BAO).

LSS observations can only probe the power spectrum on scales much smaller than the radius of the observable universe, i.e. than the current Hubble radius. For this reason, the first of the three branches described above is unobservable[121]. We can only measure the second and third branches, behaving respectively like $P(k \ll k_{\rm eq}) \propto k^{n_{\rm s}}$ and, in first approximation, $P(k \gg k_{\rm eq}) \propto [k^{n_{\rm s}-4}(\log k)^2]$, plus the titling and superimposed oscillations coming from baryons.

Parameter dependence

The discussion presented in the previous section allows us to understand which effects and which parameters impact the shape of the matter power spectrum in the minimal (flat) ΛCDM model, parametrized by

$$\{A_{\rm s}, n_{\rm s}, \omega_{\rm b}, \omega_{\rm m}, \Omega_\Lambda, \tau_{\rm reio}\} \tag{272}$$

(see Sec. 12.9.2 for details on this parametrization). We can already notice that the reionization optical depth is relevant for the CMB spectrum but not for $P(k)$, since it only impacts the scattering rate of photons in the recent universe. Other effects are described in Table **??**. Below, we give more details on these effects.

[121] The first unobservable branch is actually gauge-dependent: it behaves like $k^{n_{\rm s}-4}$ in the Newtonian gauge, but not in other gauges. Note that truly observable quantities are always gauge-invariant. We have been a bit uncareful when saying that LSS data probes the power spectrum of $\delta_{\rm m}$. In fact, different LSS observations probe different quantities, each of them being gauge-invariant. However, well inside the Hubble region, all these quantities coincide with each other and with the spectrum of $\delta_{\rm m}$ computed in an arbitrary gauge (up to small corrections that may become important in the future, but not with current experimental sensitivities).

(P1) The time of equality determines the scale $k_{\rm eq}$ of the power spectrum peak. More precisely, if this scale is expressed in units of $h\,{\rm Mpc}^{-1}$, which is the usual convention, then one can show that the scale of the maximum depends on both $z_{\rm eq}$ and $\Omega_{\rm m}$ (i.e. on our parameter basis on $\omega_{\rm m}$ and $\Omega_\Lambda = 1 - \Omega_{\rm m}$).

(P2) The baryon abundance (relative to CDM) is crucial at the matching time: a high baryon abundance leads to more suppression of the power spectrum for $k \geq k_{\rm eq}$, and to more pronounced BAOs.

(P3) We have seen that the scale of acoustic oscillations is set by the sound horizon at a given time. Since photons decouple at $\eta_{\rm dec}$, the scale of oscillations on the last scattering surface is set by $d_s(\eta_{\rm dec})$, corresponding to the comoving scale $r_s(\eta_{\rm dec}) = d_s(\eta_{\rm dec})/a(\eta_{\rm dec})$. Similarly, since baryons decoupled at $\eta_{\rm drag}$, the scale of BAOs depends on $d_s(\eta_{\rm drag})$ at baryon drag, corresponding to the comoving scale $r_s(\eta_{\rm drag}) = d_s(\eta_{\rm drag})/a(\eta_{\rm drag})$. As explained in Sec. 12.9.2, the sound horizon at a given time depends on $\omega_{\rm b}$ and $\omega_{\rm m}$. In the case of $d_s(\eta_{\rm drag})$ the dependence on $\omega_{\rm b}$ is even stronger because the time of baryon drag itself depends strongly on the baryon abundance.

(P4) The global amplitude depends of course on the primordial spectrum amplitude, i.e. on $A_{\rm s}$. Also, we have seen that during Λ domination, the growth rate of fluctuations is reduced with respect to matter domination, but is independent of k for all sub-Hubble scales. Hence, Ω_Λ also affects the global amplitude of the power spectrum.

(P5) The global tilt depends of course on the primordial spectrum tilt, i.e. on $n_{\rm s}$.

This discussion shows that in principle, a precise measurement of the matter power spectrum today (or at a given redhsift) would allow to measure independently $\omega_{\rm b}$, $\omega_{\rm m}$, $n_{\rm s}$, $A_{\rm s}$ and Ω_Λ (assuming a flat ΛCDM universe). In practice, given the limited precision of current data sets, some of the effects described above are degenerate with each other, and matter power spectrum observations are mainly useful in combination with CMB observations, since they bring independent and complementary information.

Future experiments will use all the discriminating power of the matter power spectrum. They will perform accurate measurements of $P(k,z)$ at different redshifts. This is crucial for at least two reasons:

- First, the effect of Ω_Λ in **(P5)** is redshift dependent: at different redshifts, fluctuations have spent more or less time during the Λ dominated regime, and the power spectrum amplitude has been more or less affected. Hence, the comparison of the amplitude at different redshifts allows to measure Ω_Λ, or more generally to test the compatibility of the data with a cosmological constant (rather than some dynamical Dark Energy model).

- Second, BAOs appear at a fixed comoving wavenumber $k_{\rm BAO}$ in Fourier space (related to $r_s(\eta_{\rm drag})$), but if we measure it in different LSS datasets at different redshifts (corresponding to different shells in real space), this scale will be seen under different angles[122]. By comparing the BAO angle at different redshifts, one can reconstruct the angular diameter distance, and therefore the expansion history at different redshifts. This is another way of measuring Ω_Λ or testing the Λ model versus DE models.

[122]Here we are assuming that the BAO scale is measured in each redshift shell transversally, i.e. in the direction orthogonal to the line of sight. Real experiments probe the BAO scale both transversally and longitudinally, so the situation is a bit more subtle than in this simplified discussion.

Effects of dark matter on CMB and matter spectrum

The existence of Dark Matter (DM) was first postulated by astronomers in order to account for the velocity of stars in the Milky Way (Oort, 1932), and that of galaxies in the Coma cluster (Zwicky, 1933). Later on, it has been firmly established that if general relativity applies to galactic scales and above, some form of DM halo must be present in most galaxies and clusters.

Most of the dark matter is cold. If DM is made of collisionless and stable particles, the only DM property that can be probed by cosmological perturbations is its velocity distribution.

At the level of homogeneous cosmology, the DM phase-space distribution f only depends on momentum p. The number of particles in a given momentum bin is given by $dp\,p^2 f(p)$. If $p^2 f(p)$ is very peaked near $p = 0$ or the DM mass is very large, DM particles have a negligible velocity dispersion. In this case, they can be represented effectively as a pressureless fluid, as explained in Sec. 12.9.1. We recall that they are not forming a real fluid, due to the absence of interactions, but since their velocity dispersion is negligible, all particles have approximately the same bulk velocity in each point (described by the divergence $\theta_{\rm cdm}$). In that case the DM component is called Cold Dark Matter (CDM).

If $p^2 f(p)$ is not very peaked near $p = 0$ or the DM mass is not very large, a significant fraction of DM particles may have a velocity p/m comparable to the speed of light, or at least to that of astrophysical objects. Typically, if the velocity dispersion is of the order of 10^{-3} (in units of c) or larger, it cannot be neglected when studying the CMB or LSS spectra. In that case, the velocity of individual particles results from two physical mechanisms: the dynamics imposed by gravitational forces, plus an intrinsic velocity distributed stochastically, according to the phase-space distribution. The notion of "velocity dispersion" refers to the second component, since the first effect tends to impose the same bulk velocity to all particles in a given point. DM with a large velocity dispersion cannot be described by fluid equations like CDM. A famous example of DM particles with a large velocity dispersion is neutrinos. The velocity dispersion is then fixed by the neutrino mass, and by their Fermi-Dirac distribution. Such DM is called Hot Dark Matter. Possible DM species with a smaller (but still significant) velocity dispersion are called Warm Dark Matter[123].

Since CDM is a very generic limit, and neutrinos are know to exist since the 1950's, the two most popular DM models until the 1990's were CDM and HDM. CDM can be motivated by many possible extensions of the standard model of particle physics (supersymmetry, Peccei-Quinn symmetry, etc.), predicting some early decoupled stable particles, with a very small velocity dispersion. HDM was motivated at that time by massive neutrinos with masses of the order of ~ 10 eV.

These two models lead to very different predictions for the matter power spectrum. We have discussed before the case of CDM. A study of perturbations in a HDM (or ΛHDM) model would show that for $k \geq k_{\rm eq}$, the power spectrum $P(k)$ would decrease with a much steeper slope. The physical explanation is that the velocity dispersion of HDM particles prevents them from clustering on small scales: hence the growth rate of $\delta_{\rm m}$ inside the Hubble radius is considerably reduced in the HDM model.

In the 90's, the first estimates of the matter power spectrum $P(k)$ (or of some related quantities) on galaxy scales and cluster scales showed clearly that the HDM scenario was ruled out. Today, using more accurate LSS data, it is firmly established that most of the DM must be cold (or eventually, warm, but with such a small velocity dispersion that WDM would be indistinguishable from CDM at least on the largest observable scales). Still, we cannot exclude

[123]The limit between HDM and WDM is usually placed in the following way: both types of particles have a sizable velocity dispersion, but most HDM particles become non-relativistic after radiation/matter equality, while most WDM particles become non-relativistic before that time.

the presence of a subdominant hot or warm component. In fact, since we know that neutrinos have a small mass (of the order of ~ 0.1 eV), we do expect that a few percents of DM is hot. Since the signature of this sub-leading component has not yet been detected, the minimal model is still ΛCDM.

In the ΛCDM case, cosmological perturbations can only measure one DM-related parameter: its relic density, parametrized by $\omega_{\rm cdm}$. This is better than nothing: the value of $\omega_{\rm cdm}$ can be used to constrain some fundamental parameters governing DM production, within the framework of a postulated model.

Measuring $\omega_{\rm cdm}$ with the CMB. Going back to Sec. 12.9.2, we see that $\omega_{\rm cdm}$ can be measured "indirectly" using CMB observations: effect **(C2)** measures $\omega_{\rm b}$, effect **(C3)** measures $\omega_{\rm m}$, and $\omega_{\rm cdm}$ is given by the difference between these two parameters.

There is no more "direct" effect of CDM that could be probed with the CMB. A priori, we could think that photons are affected by gravitational forces generated by CDM. This gravitational interaction could only be relevant before photon decoupling, and around the time of sound horizon crossing for each given wavelength (we have seen in Sec. 12.9.2 that well inside the sound horizon, photons do not feel gravitational forces anymore, because pressure forces are much larger). But around sound horizon crossing, photons feel mainly the gravitational forces generated by themselves, by neutrinos, and to a lesser extent by baryons. In fact, in this regime, CDM and other components are effectively decoupled[3].

Hence the CMB probes the CDM abundance only indirectly, as a "missing mass", exactly like in astrophysics: it measures the total matter density through the time of equality, the baryon density through the ratio between odd and even peaks, and the difference must be attributed to a non-relativistic component not interacting with the photons: Dark Matter.

In this discussion, we concentrated only on effects **(C2)** and **(C3)**. Indeed, effects **(C1)** and **(C4)** are also sensitive to $\omega_{\rm b}$ and $\omega_{\rm cdm}$, but only in combination with other parameters. So, unlike **(C2)**+**(C3)**, they do not bring a robust and nearly model-independent evidence for DM.

Measuring $\omega_{\rm cdm}$ with the matter power spectrum. The overall shape of $P(k)$ for $k \geq k_{\rm eq}$ brings very strong support for CDM (or eventually WDM with a small enough velocity dispersion). Indeed, a universe with only baryonic matter, or baryonic plus hot dark matter, would have a radically different power spectrum (with a very steep slope for $k \geq k_{\rm eq}$, and eventually huge BAOs in the purely baryonic case). More precisely, data on the matter power spectrum allows to measure the ratio $\omega_{\rm b}/\omega_{\rm cdm}$ through effect **(P2)** of Sec. 12.9.3. Hence the matter power spectrum probes the CDM abundance independently of the CMB. This probe is more direct, since $P(k)$ is really sensitive to the gravitational clustering of CDM inside the Hubble radius.

Signatures of DM annihilation or decay in the CMB. Depending on the underlying particle physics model, DM particles could either annihilate or decay. Of course, the annihilation or decay rate should be very small, since we need a stable dark matter population. Still, a small annihilation or decay rate could play a role in the recent thermal history of the universe, and in particular in the evolution of the ionized fraction of atoms, $x_e(z)$, to which the CMB is sensitive. Indeed, DM annihilation/decay products may include photons that could directly ionize hydrogen atoms, or particles that could reheat the inter-galactic medium. These mechanisms tend to increase x_e at a given time. Their efficiency depends on the mass of DM particles, on their cross-section in the case of annihilation, and on their lifetime in the case of decay.

Before recombination and photon decoupling, the ionization fraction is anyway very close to one. In the minimal cosmological model, x_e drops abruptly after recombination, freezes out when particles leave chemical equilibrium, and increases again to one at the time of reionization, due to ionizing photons emitted at star formation. DM annihilation or decay could potentially

modify this picture and play two roles:

- at high redshift (typically $600 < z < 1100$), it could slow down the drop of x_e after recombination, and increase its freeze-out value. This would appear in the CMB as a shift in the time of decoupling (impacting effect **(C1)** and **(C4)** of Sec. 12.9.2), as a broadening of the peak of the visibility function $g(\eta)$ (i.e. the last scattering "surface", which is in fact more like a shell, would be thicker, impacting **(C4)**), and as an increase of the optical depth to recombination (impacting **(C8)**, but not on the same scales as reionization).

- at small redshift (typically $z < 40$), the concentration of DM in halos could enhance its annihilation/decay rate, and annihilation/decay products could reionize atoms and increase x_e. This would be equivalent to some extra reionization: it could modify the reionization time or history, and impact **(C8)**.

The recent literature shows that for realistic models, CDM annihilation could play these two roles: hence, interesting bounds on DM annihilation can be derived from the CMB. These bounds are complementary from those inferred from cosmic ray data, or direct detection experiments.

DM decay could play the second role, but the bounds on the DM lifetime inferred in this way from the CMB are not as strong as those from cosmic ray data.

These effects had to be mentioned in this section, because they are part of the interplay between DM physics and cosmological perturbation theory. We do not discuss here bounds on DM annihilation/decay from cosmic ray data, since they have little to do with the evolution of linear cosmological perturbations.

What else could we learn on DM from CMB or LSS data? In the simplest cosmological scenario, cosmological perturbations only probe the relic abundance of DM particles. If we are lucky, they may provide some hint of DM annihilation. There are other properties of DM that can be tested with cosmological perturbations, even if at the moment, everything is consistent with the minimal ΛCDM picture:

- the small mass of neutrinos implies a small fraction of HDM, with distinct effects on the CMB and LSS power spectrum[1] (effects on $P(k)$ are larger: a small neutrino mass implies a step-like suppression of the power spectrum on small scales, by at least 4% if we assume the smallest possible neutrino masses that are consistent with neutrino oscillations experiments).

- a non-negligible velocity dispersion of the dominant DM component (like in the WDM model) or of a fraction of it (in the case of several DM species) would also induce a suppression of the small-scale matter power spectrum, at a scale and with a shape depending on the mass and momentum distribution of the warm species. However, such a feature is clearly not observed in the matter power spectrum on linear scales. Such a signature could only be found on smaller scales, affected by non-linear gravitational clustering.

- instead of being treated as collisionless, DM could experience small interactions with itself or with other species (photons, neutrinos baryons). All these possibilities require a case-by-case analysis, and usually predict observable features similar to WDM. Both the CMB and the matter power spectrum can be used to put very strong bounds on a possible DM interactions.

In conclusion of this section, it appears that CMB and LSS observations bring some crucial information concerning the abundance and the nature of Dark Matter. But as long as everything is compatible with the minimal ΛCDM scenario, CMB and LSS data are only telling us what DM cannot consist of, rather than what it is.

Effects of dark energy on CMB and matter spectrum

In this course, we assumed for simplicity that the (at least apparent) acceleration of the universe is caused by a cosmological constant. The possibility that acceleration would be caused by a dynamical component, called generically Dark Energy (DE), is one of the most investigated questions in cosmology. Of course, the acceleration could also be caused by some alternative theory of gravity, different from General Relativity on very large scales. Even in that case, everything could be reformulated in terms of a dark energy component (by moving new terms from the left to the right hand-side of the Einstein equation). Hence, measuring DE parameters is a rather generic way to look for deviations from a simple cosmological constant[124].

Possible Dark Energy parameters. At the background level, the stress-energy tensor of any species compatible with the assumption of a homogeneous and isotropic FL universe must be diagonal, involving only two function $\bar{\rho}_{\rm de}(a)$ and $\bar{p}_{\rm de}(a)$, related to each other by an equation of conservation of energy. Hence these two functions are not independent of each other, at least as long as DE is not interacting with other species. Even if in general a DE model has no reasons to obey to an equation of state, one can always define the function $w(a) \equiv \bar{p}_{\rm de}(a)/\bar{\rho}_{\rm de}(a)$.

A cosmological constant corresponds to $w(a) = -1$, i.e. to a constant density $\bar{\rho}_{\rm de} = -\bar{p}_{\rm de}$, parametrized by the value of $\Omega_\Lambda = \Omega_{\rm de} \equiv \bar{\rho}_{\rm de}/\rho_{\rm crit}$ today.

A simple extension of this model is the wCDM scenario, in which one assumes a linear and constant equation of state $\bar{p}_{\rm de} = w\bar{\rho}_{\rm de}$, leading to $\bar{\rho}_{\rm de} \propto a^{-3(1+w)}$ by virtue of energy conservation. This model has two independent DE parameters, $\Omega_{\rm de}$ and w.

However, w could in principle vary with time. In a purely phenomenological approach, after confronting the wCDM model to observations, one can test a model with one more free parameter, accounting for the variation of w today. Then, there are three DE parameters, for instance $\Omega_{\rm de}$, $w(a_0)$ and $\frac{dw}{da}(a_0)$.

At the level of perturbations, different DE models with a given background evolution could differ through their pressure perturbation, or their anisotropic stress. First, let us stress that DE cannot be a perfect fluid. If this was the case, DE would have an adiabatic sound speed obeying to

$$c_a^2 \equiv \frac{\bar{p}'}{\bar{\rho}'} = \frac{(w\bar{\rho})'}{\bar{\rho}'} = w + \frac{w'}{3\frac{a'}{a}(1+w)} \tag{273}$$

where in the last equality we have used energy conservation. Given observational constraints on w and w' today, the squared adiabatic sound speed of dark energy must be negative. If the actual squared sound speed $c_s^2 \equiv \frac{\delta p_{\rm de}}{\delta \rho_{\rm de}}$ was negative, then DE perturbations would be unstable, and would explode exponentially, in contradiction with observations. This proves that for DE, $c_s^2 \neq c_a^2$. The stability of DE perturbations imposes $c_s^2 > 0$. On the other hand, the absence of superluminal propagation in conventional physics imposes an upper bound, $c_s^2 \leq 1$.

In phenomenological studies of DE models, the default assumptions concerning DE perturbations is that *(i)* pressure perturbations are related to density perturbations through $c_s^2 = 1$, and that *(ii)* anisotropic stress is negligible. This is justified by the fact that DE could be a scalar field (often called quintessence). If this is the case, one can easily show that the sound speed is indeed equal to one[125], while anisotropic stress only appears at second order in perturbations.

[124] Note that alternatively, the observed apparent acceleration might be the consequence of a departure from the homogeneous Friemann-Lemaître model, but we will not discuss such a possibility (which is very constrained and maybe already excluded by observations) in this course.

[125] Actually, the quantity $\delta p_{\rm de}/\delta\rho_{\rm de}$ is gauge dependent. Everything becomes clearer if one decides to call "sound speed" the ratio $\delta p_{\rm de}/\delta\rho_{\rm de}$ calculated *in the frame comoving with the DE fluid*, i.e. in a gauge such that the velocity divergence $\theta_{\rm de}$ vanishes. If the sound speed of a given component derives from some fundamental principle, it should take a simple form precisely in this frame. Actually, it is only when this condition is imposed that one finds $\delta p_{\rm de} = \delta\rho_{\rm de}$ for a scalar field.

A DE model with $c_s^2 = 1$ cannot lead to sizable DE perturbations. Indeed, $c_s^2 = 1$ implies that pressure forces in the fluid resist to gravitational collapse. Astrophysicists like to formulate this problem in terms of the Jeans instability. In this language, $c_s^2 = 1$ means that the Jeans length in the fluid is as large as the Hubble radius. Then DE perturbations remain as small as in the primordial universe (of the level of 10^{-5}), and can never leave an observable signature, even during DE domination.

Only if c_s^2 is assumed to be very small, DE perturbations can grow and impact LSS observables. Indeed, if c_s^2 goes to zero, the Jeans length can be arbitrarily small, and perturbations on cosmological scales can be amplified by the Jeans instability. In that case, one needs to postulate whether DE perturbations feature anisotropic stress or not. In a phenomenological approach, one may try to parameterize the anisotropic stress using some effective parameter (like a viscosity parameter). One should stress that a very small sound speed can be assumed, but these models sound artificial: $c_s^2 \ll 1$ is not a generic prediction of "realistic" DE models (if any such model exist).

Probing DE models with CMB and LSS data. If the DE sound speed is not very small, DE models can only be probed through their impact on the background evolution, which affects indirectly the evolution of metric and matter perturbations.

Since the background expansion at late time impacts the angular diameter distance, the CMB is sensitive to DE through effects **(C1)** and **(C4)** of Sec 12.9.2, both depending on $d_{\rm a}(z_{\rm dec})$. However, **(C1)** and **(C4)** are also sensitive to many other quantities, so this probe is only useful in combination with other ones: in the best case it may remove parameter degeneracies.

The LISW effect **(C7)** probes the duration of the DE dominated era, and could bring one bit of information on DE models. In practise, it is however impossible to infer precise constraints from **(C7)** due to the large cosmic variance affecting the C_l's on the scale of the Sachs-Wolfe plateau.

We have seen in Sec 12.9.3 that LSS data can be useful especially if the matter power spectrum is measured accurately at several distinct redshifts. Then, both the evolution of the power spectrum amplitude (in other words, of the linear growth factor) and of the BAO apparent scale encode information on the expansion rate at several redshifts, and therefore on w or even $w(a)$. This is why measurements of $P(k, z)$ with future ambitious programs, like for instance the Large Synoptic Survey Telescope or the Euclid satellite, are expected to provide the most accurate tests of DE models.

In summary, the CMB is not very useful for testing DE models, but can bring a few independent pieces of information; while LSS data at different redshifts is sensitive to the evolution of the background DE density. LSS data should anyway be used in combination with other probes of homogeneous cosmology (like supernovae luminosity) in order to get a maximum amount of information on DE.

The situation that we depicted would change radically if:

- DE would account for a significant fraction of the total energy density even at early times ($z \gg 1$), like in "early dark energy" models: then, DE could affect the CMB more strongly than assumed in the previous discussion;
- large DE perturbations would develop as a consequence of a very small sound speed $c_s^2 \ll 1$: then the total power spectrum (including DE fluctuations) may depart significantly from the ΛCDM one.

Both possibilities are still open, but no such effects are indicated by current data, and there are no solid theoretical motivations for our universe to fall in one of these two cases.

Conclusions and bibliography

We hope that this short — but very dense — introduction to cosmological perturbations was clear enough, and that the reader will feel like going further.

Some readers might regret that, after exposing in details the effect of cosmological parameters on the CMB and LSS spectra (in Secs. 12.9.2, 12.9.3, 12.9.3 and 12.9.3), we did not show concrete bounds from current experiments. There are three reasons for this. First, the duration of the lectures and the length of these notes would have exploded. Second, such information would be outdated very quickly by new data releases. Third, it is actually a very good exercise (and an efficient way to digest this course) to look at published bounds on cosmological parameters (inferred from CMB data, or LSS data, or from both), and to figure out by oneself whether everything is consistent with these notes: we encourage the reader to check that the best and worst constrained parameter combinations can be related to the physical effects described here.

We promised a bibliography in the introduction. We already mentioned Ref. [1], in which one can find a similar but more detailed presentation, plus a discussion of neutrino effects. We have cited in the core of the text three seminal papers (Refs. [2, 3, 661]) providing explicit calculations on precise topics: the line-of-sight integral in Fourier space[2], the justification of the Mészáros equation and of the fact that CDM behaves effectively as a self-gravitating species during radiation domination[3], and the matching between CDM and baryon perturbations around the time of baryon drag, in view of computing analytically baryon effects in the matter power spectrum[661].

Specialists of the field refer regularly to a concise paper of Ma & Bertschinger [1218], containing a remarkably clear presentation of gauge issues and of all equations of motion in two gauges (including the Newtonian one).

Many recent textbooks cover most topics in cosmology, including linear perturbations, CMB and large scale structure. We only mention in the reference list those of S. Dodelson[7], V. Mukhanov[8] and S. Weinberg[4], all three outstanding, but there exist many excellent alternatives. For a presentation more focused on the CMB, we recommend the thesis of W. Hu[9] and the book of R. Durrer[10].

12.10 Probes of Inhomogeneities

The power spectra we have explored in the previous two chapters $-P(k)$ of the density field and C_l of the anisotropies - are the most obvious first tests of any cosmological model. The most direct way of measuring $P(k)$ is to do a redshift survey, wherein the angular positions and the redshifts (which are a measure of radial distance) of galaxies are recorded. There are, however, a number of problems with redshift surveys and their interpretation. The first is the simple fact that taking redshifts is time consuming: it is much easier to get the angular positions of galaxies than it is to also measure redshifts. In the same time that a 10,000 galaxy redshift survey, say, could be completed, a million angular positions of galaxies could be obtained. With the much greater statistics, angular surveys often compensate for the lack of radial information. Indeed, some claim that the best information we have on large-scale clustering comes from angular surveys. Clearly, then, one skill we must acquire is the ability to make predictions about the angular correlation function $w(\theta)$. In Section 9.1 we will see that the angular correlation function is an integral over the 3D power spectrum.

Redshift surveys suffer from another, more profound problem than the fact that they are time consuming. While is true that the redshift gives a reasonable estimate of radial distance (by radial distance I simply mean distance from us), it is not true that this estimate is completely accurate. Recall that a galaxy's velocity is determined solely by the Hubble expansion (and hence redshift is a perfect indicator of distance) only if the galaxy is stationary on the comoving grid. Most galaxies have nonnegligible peculiar velocities; that is, they are moving on this grid.

A galaxy's total velocity, which is measurable, is

$$\vec{v} = \vec{v}_{\rm pec} + \hat{x} v_{\rm H}$$

where the Hubble velocity $v_{\rm H} := \chi da/dt = \chi a H$. Recall that χ is the comoving distance between us and a distant galaxy, so its physical distance from us is $a\chi$. In the absence of peculiar velocities, $\hat{x} \cdot \vec{v}/H$ is a perfect distance indicator. In the real world, though, where peculiar velocities do not vanish, even an accurate measurement of a galaxy's recession velocity does not translate into an unambiguous measurement of its radial distance away from us.

The ambiguity introduced by peculiar velocities offers an opportunity. In linear theory, peculiar velocities are determined by the surrounding density field, so we can correct for the distortions induced by working with redshifts. Indeed, we can go even further and use these distortions to learn about the growth of perturbations, for the precise way in which velocities are related to the density field is determined by the rate at which perturbations grow. Since this rate is determined in cosmology by Ω_m, studying the distribution of galaxies in redshift space is one of the more promising ways to measure Ω_m.

Finally, another way of gleaning information about the underlying mass density is by studying clusters of galaxies. Although strictly speaking, this topic falls in the realm of nonlinear evolution of the density field, and therefore beyond the boundary of this book, the Press-Schechter method of approximating cluster abundances is only a very small step away from linear theory and has been shown to be quite accurate in its predictions. Further, the study of clusters is advancing at an extraordinarily rapid rate, since clusters can now be probed with many different astronomical techniques. Section 9.5 introduces the basic predictions of the Press-Schechter theory and the implications for cold dark matter models.

12.10.1 Angular Correlations

Theory

Figure 9.1 shows the angular positions of over a million galaxies from the Automated Plate Measuring (APM) Survey. In an angular survey such as this, what statistic can be computed that can be compared with theory? The simplest statistic is the two-point function: in real space it is $w(\theta)$ the angular correlation function. In Fourier space, the relevant function is the Fourier transform of w, $P_2(l)$, the two-dimensional power spectrum. In this section we compute both of these very important functions, relating them to the three-dimensional power spectrum.

First let me introduce some notation. Figure 9.2 shows the geometry: a given galaxy is at comoving distance $\chi(z)$ (Eq. (2.42)) away from us. The z-axis is typically chosen so that it points to the center of the distribution of galaxies. In the plane perpendicular to this axis, a galaxy's position is determined by the twodimensional vector $\vec{\theta} = (\theta_1, \theta_2)$. Therefore, the three-dimensional position vector $\vec{x}$ has components

$$\vec{x}(\chi(z), \vec{\theta}) = \chi(z)\,(\theta_1, \theta_2, 1)\,.$$

The assumption that all galaxies are located near the z-axis clearly breaks down if the survey measures structure on very large angular scales. As an example the APM survey in Figure 9.1 covers roughly 50 by 100 square degrees. From this data, one can measure the correlation function accurately out to about 10° or 0.17 radians, safely smaller than unity.

We measure all galaxies along the line of sight, effectively integrating over $\chi(z)$. Therefore an overdensity at angular position $\vec{\theta}$ is

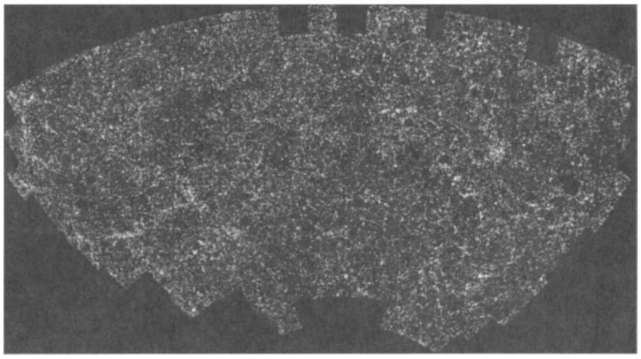


Figure 9.1. The distribution of galaxies in the APM survey. Blacked-out regions were not observed during the survey.

$$\delta_2(\vec{\theta}) = \int_0^{\chi\infty} d\chi W(\chi)\delta(\vec{x}(\chi,\vec{\theta}))$$

where the subscript $_2$ denotes the fact that δ on the left is the angular-or twodimensional - overdensity, while δ on the right is the full three-dimensional overdensity. (And I will stick with this convention: for example P_2 is the 2D power spectrum while P denotes the $3D$ spectrum.) The upper limit on the χ integral corresponds to $z \to \infty$, equal to $\chi_\infty = 2/H_0$ in a flat, matter-dominated universe. In practice, the magnitude-limited surveys that have yielded the most cosmological information to date have probed $z \lesssim 0.5$. The selection function $W(\chi)$ encodes this information: it is the probability of observing a galaxy a comoving distance χ from us. Galaxies at large distances are too faint to be included in a survey, whereas there are relatively few galaxies at very low redshift simply because the volume is small. Since it is a probability, the selection function is normalized so that $\int_0^{\chi\infty} d\chi W(\chi) = 1$

The 2D vector conjugate to $\vec{\theta}$ will be $\vec{l}$, so that the Fourier transform of $\delta_2(\vec{\theta})$ is

$$\tilde{\delta}_2(\vec{l}) = \int d^2\theta e^{-i\vec{l}\cdot\vec{\theta}}\delta_2(\vec{\theta})$$

The two-dimensional power spectrum is defined just as was the 3D:

$$< \tilde{\delta}_2(\vec{l})\tilde{\delta}_2^*(\vec{l}) >= (2\pi)^2\delta^2(\vec{l}-\vec{l})P_2(l)$$

here $\delta^2()$ is the 2D Dirac delta function, not to be confused with the overdensities $\delta_2(\vec{\theta})$. Integrating, we can therefore write the 2D power spectrum as

$$P_2(l) = \frac{1}{(2\pi)^2}\int d^2l' \left\langle \delta_2(\vec{l})\delta_2^*(\vec{l}) \right\rangle$$

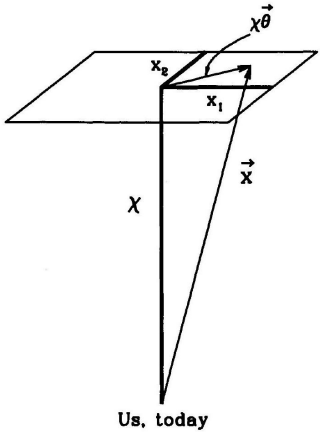


Figure 9.2. A distant galaxy is located at position $\vec{x}$ with respect to us at the origin today. This position can also be expressed in terms of $\chi(z)$, the comoving distance out to the redshift of the galaxy. The $x_1 - x_2$ plane is perpendicular to a suitably chosen x_3-axis. In this plane, a galaxy's position is given by the two-dimensional vector $\chi\vec{\theta}$, so $\vec{x} \simeq \chi(z)\,(\theta_1, \theta_2, 1)$.

$$= \frac{1}{(2\pi)^2} \int d^2l' \int d^2\theta \int d^2\theta' e^{-i\vec{l}\cdot\vec{\theta}} e^{i\vec{l}\cdot\vec{\theta}'} \times \int_0^{\chi_\infty} d\chi W(\chi) \int_0^{\chi_\infty} d\chi' W\left(\chi'\right) \left\langle \delta(\vec{x}(\chi,\vec{\theta}))\delta\left(\vec{x}'\left(\chi',\vec{\theta}'\right)\right)\right\rangle .$$

The integral over $\vec{l}'$ gives $(2\pi)^2$ times a Dirac delta function in $\vec{\theta}'$ and the brackets give the 3D correlation function,

$$\xi\left(\vec{x} - \vec{x}'\right) := \left\langle \delta(\vec{x})\delta\left(\vec{x}'\right)\right\rangle = \int \frac{d^3k}{(2\pi)^3} P(k) e^{i\vec{k}\cdot(\vec{x}-\vec{x}')}$$

The average here $\langle \ldots \rangle$ is over all realizations of the density field. At very small distance, we expect galaxies to be clustered strongly as a result of gravity, so ξ is positive. As the distance gets larger, correlations die off and ξ gets smaller and eventually goes negative. The second line follows since the correlation function is the Fourier transform of the power spectrum (Exercise 1).

Performing the integral over $\vec{\theta}'$ in Eq. (9.6) then leads to

$$P_2(l) = \int d^2\theta e^{-i\vec{l}\cdot\vec{\theta}} \int_0^{\chi_\infty} d\chi W(\chi) \int_0^{\chi_\infty} d\chi' W\left(\chi'\right) \int \frac{d^3k}{(2\pi)^3} P(k) e^{i\vec{k}\cdot\left[\vec{x}(\chi,\vec{\theta}) - \vec{x}(\chi',0)\right]}$$

The argument of the exponential at the end here is $i\left[k_1\chi\theta_1 + k_2\chi\theta_2 + k_3\left(\chi - \chi'\right)\right]$, so the integral over angles $\vec{\theta}$ gives Dirac delta functions setting $l_1 = \chi k_1$ and $l_2 = \chi k_2$. We can use these delta functions to do the k_1 and k_2 parts of the d^3k integration, remembering to divide by the derivative of the argument, in this case putting a factor of χ^2 in the denominator. We are thus left with

$$P_2(l) = \int_0^{\chi_\infty} d\chi \frac{W(\chi)}{\chi^2} \int_0^{\chi_\infty} d\chi' W\left(\chi'\right) \int_{-\infty}^{\infty} \frac{dk_3}{(2\pi)} P\left(\sqrt{k_3^2 + l^2/\chi^2}\right) e^{ik_3[\chi - \chi']}$$

Until now, we have been doing math; to complete the calculation of the power spectrum we need to introduce some physics. I claim that the only 3D Fourier modes that contribute to the integral are those with k_3 very small, much smaller than l/χ. To see why, we first need to estimate l, the variable conjugate to θ. Roughly, l^{-1} is of order the angular scales probed by the survey [1]. Since we are working in the small angle approximation, $l/\chi \sim 1/(\chi\theta) \gg 1/\chi$. Now let's consider Figure 9.3. There

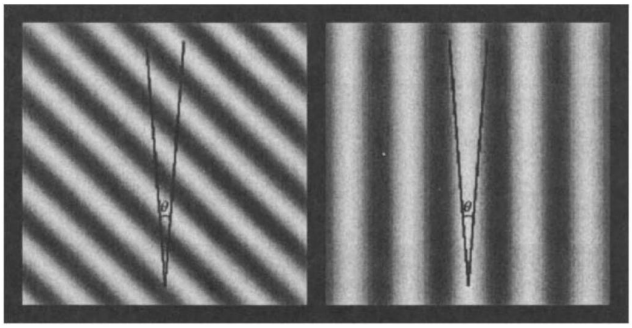


Figure 9.3. Two plane-wave perturbations and their contributions to the 2D power spectrum. Left panel shows a perturbation with longitudinal wavenumber $k_3 \gg \chi^{-1}$ (the $\hat{z}$ direction is vertical). Right panel shows a mode with $k_3 \sim \chi^{-1}$. Angular correlations due to the large k_3 mode (left panel) are negligible since there are many cancellations along the line of sight.

[1] In this sense, l is very similar to the degree of the Legendre polynomials introduced in Chapter 8 to study anisotropies. In the same sense, $P_2(l)$ is very similar to C_l; indeed in Exercise 5 you will show that they are identical on small scales. we see that modes with longitudinal wavenumber k_3 much greater than χ^{-1} do not give rise to angular correlations because of cancellations along the line of sight. Only modes with k_3 smaller than χ^{-1} lead to angular correlations. Therefore, the relevant transverse wavenumbers l/χ are much larger than the relevant longitudinal wavenumbers, and we can safely set the argument of the 3D power spectrum to l/χ (see Exercise 2 for a more systematic justification). With this approximation, the k_3 integral gives a Dirac delta function in $\chi - \chi'$ so

$$P_2(l) = \int_0^{\chi\infty} d\chi \frac{W^2(\chi)}{\chi^2} P(l/\chi)$$

This is an expression for the 2D power spectrum as an integral over the line of sight. We can change dummy variables from $\chi \to k := l/\chi$ to rewrite the integral as

$$P_2(l) = \frac{1}{l} \int_0^{\infty} dk P(k) W^2(l/k)$$

The angular correlation function is the Fourier transform of the 2D power spectrum, so

$$w(\theta) = \int \frac{d^2l}{(2\pi)^2} e^{il\cdot\vec{\theta}} P_2(l).$$

Since P_2 depends only on the magnitude of $\vec{l}$, the angular part of the integration over l is $\int_0^{2\pi} d\phi e^{il\theta\cos\phi}$, which is proportional to $J_0(l\theta)$, the Bessel function of order zero (Eq. (C.21)). Therefore,

$$\begin{aligned} w(\theta) &= \int_0^{\infty} \frac{dl}{2\pi} l P_2(l) J_0(l\theta) \\ &= \int_0^{\infty} dk k P(k) F(k,\theta), \end{aligned}$$

where the second line follows from changing the order of integration. Here the kernel for the angular correlation function is

$$F(k,\theta) := \frac{1}{k}\int_0^\infty \frac{dl}{2\pi} J_0(l\theta) W^2(l/k)$$

The kernel is plotted in Figure 9.4 for two surveys. Note that it is a function of $k\theta$ (see Exercise 3). The kernel is constant at small $k\theta$ and then begins damped oscillations. The contribution from small k though is suppressed because the integral in Eq. (9.13) is over the kernel weighted by $kP(k)$, and the latter goes to zero as $k \to 0$. Therefore, the modes that contribute most to $w(\theta)$ are typically those with wavenumbers of order the first turnover in the kernel, $k\theta \sim 0.2h\mathrm{Mpc}^{-1}$ degrees for APM and a factor of 3 smaller for the deeper Sloan Digital Sky Survey (SDSS). This means that the angular correlation function at 5° in APM is most sensitive to power at $k = 0.04h\mathrm{Mpc}^{-1}$. The wavenumbers contributing to $w(\theta)$ in a deeper

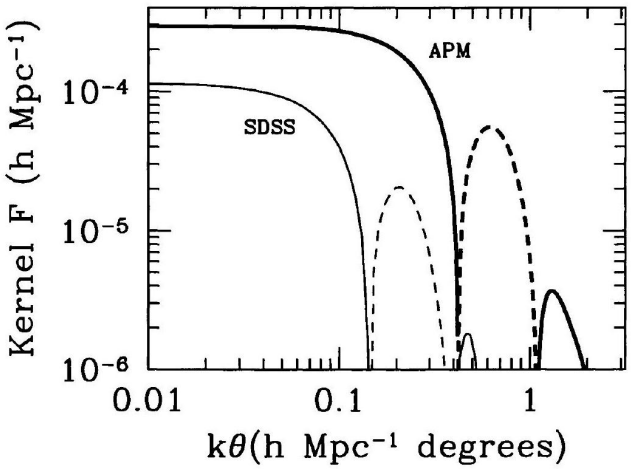


Figure 9.4. The kernel relating the angular correlation function to the 3D power spectrum in two surveys. Kernel is negative when line is dashed, positive otherwise. APM Survey probes galaxies brighter than apparent magnitude $m = 20$, while the Sloan Digital Sky Survey (SDSS) will go much deeper, potentially sensitive to galaxies brighter than $m = 23$.

angular survey are smaller. This makes sense: the same angle probes larger physical scales in a deeper survey.

Figure 9.5 shows the measurements of the angular correlation function from the APM survey. The most important conclusion from the data is that standard Cold Dark Matter - with $\Omega_m = 1$ and $h = 0.5$ - is a bad fit. To quote from the abstract of the Maddox et al. (1990) paper which measured the correlation function, "more large-scale clustering than predicted by popular versions of the Cold Dark Matter cosmogony is implied." Although sCDM has died many deaths since its inception in the early 1980s, the death from APM was perhaps its most celebrated. Despite long, hard work by many people trying to find systematic problems with this and other surveys, nothing significant has changed over the past decade to alter the conclusion that the standard Cold Dark Matter model of structure formation fails to predict accurately the observed pattern of large-scale clustering. Having said that, I want to emphasize that the situation is not quite as severe as you might imagine from a cursory examination of Figure 9.5. Consider the prediction from a Λ CDM model, also shown in Figure 9.5. Although the agreement is much better than with sCDM, there are clear discrepancies on both large and small scales. The small-scale discrepancies are completely illusory, though, because I have used the linear power spectrum to compute $w(\theta)$. Nonlinearites become important on scales of order $k \sim 0.2h\mathrm{Mpc}^{-1}$ as we saw in Chapter 7 (Exercise 10). These scales

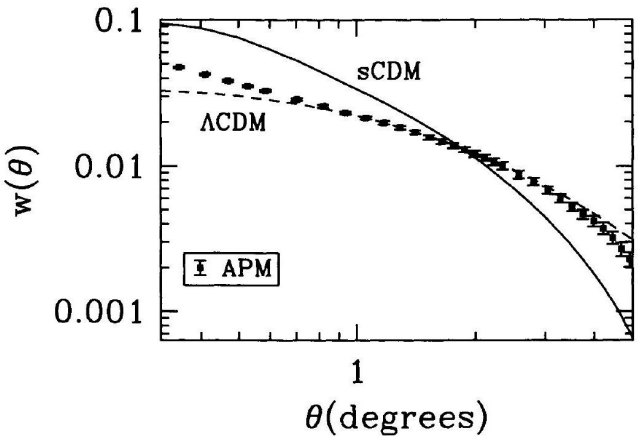


Figure 9.5. Angular correlation function in the APM survey and two theoretical models. Standard CDM (solid curve) is a bad fit to the data, while a model with a cosmological constant (here $\Omega_\Lambda = 0.7$) fits well. The apparent disagreements between the data and Λ CDM on small and large scales are illusory; see text.

contribute to $w(\theta)$ in APM when θ is of order 1°. So to compare to the data fairly, we really need to account for nonlinearities; I have not done this, so we cannot take the small-angle discrepancy seriously. On large angles, people have begun to realize that the data have been overinterpreted. The basic problem is that the points on large angles are highly correlated. So the slight discrepancy between the data and ΛCDM on large angles looks worse than it really is.

The angular correlation function can be inverted to obtain the 3D power spectrum: the results from the APM Survey are shown in the top panel of Figure 9.6. We will discuss inversion techniques in Chapter 11, but you should be aware that the points in Figure 9.6 are the result of a long process (i.e., many papers), in the midst of which error bars were often vastly underestimated. Using only the results on large scales (where nonlinear effects are irrelevant and the relation between mass overdensity and galaxy overdensity is expected to be simple), Efstathiou and Moody (2001) placed constraints on CDM models, shown in the bottom panel of Figure 9.6. We found in Chapter 7 that the CDMI transfer function depends only on $k/k_{\rm eq}$. Since $k_{\rm eq}$ scales as $\Omega_m h^2$ and since surveys measure the wavenumber k in units of hMpc^{-1}, the combination $\Omega_m h$ determines the shape of the power spectrum. For this reason, fits to large-scale structure data are often given in terms of the shape parameter $\Gamma := \Omega_m h$ (sometimes modified to account for baryons: see Exercise 8). Standard CDM with $h = 0.5$ and $\Omega_m = 1$ corresponds to $\Gamma = 0.5$; if the index of the primordial spectrum is $n = 1$, then the bottom right panel of Figure 9.6 shows

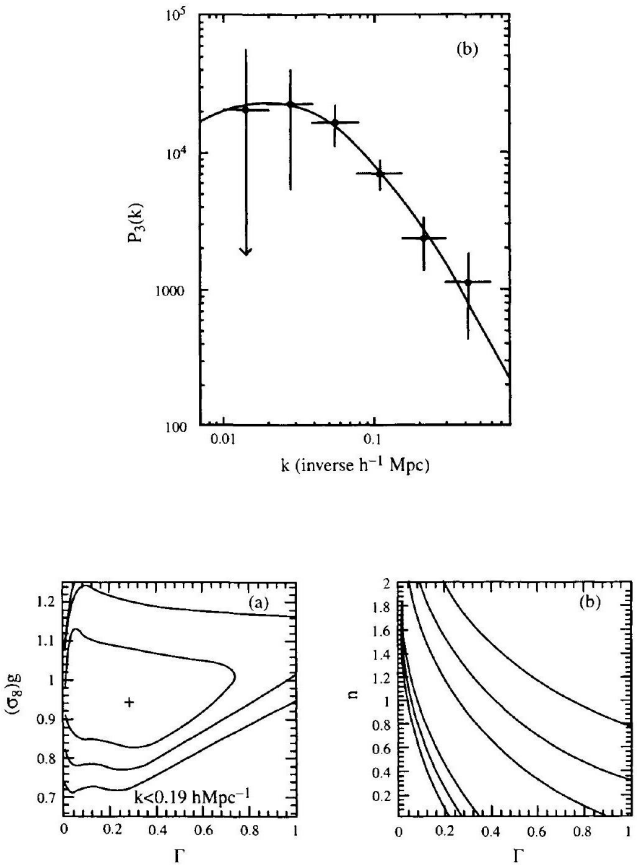


Figure 9.6. Results from the APM Survey (Efstathiou and Moody 2001). Top panel. The 3D power spectrum inferred from the angular correlation function. The curve is a CDM model with $\Gamma\,(\simeq \Omega_m h) = 0.2$, normalized to fit the data. Bottom Panel. Constraints on CDM models from the power spectrum. Three parameters - the shape parameter Γ, the amplitude σ_8, and the index of the primordial power spectrum n - were varied. Contours delineate one-, two-, and three-sigma regions. Left panel shows constraints after integrating over n; right after integrating over the amplitude. Standard CDM has $n = 1; \Gamma = 0.5$; and (COBE-normalized) $\sigma_8 = 1.15$ that sCDM is ruled out at the 2 -sigma level. ΛCDM has $\Gamma \simeq 0.2$ and is indeed a better fit to the data if n is close to 1.

12.10.2 Peculiar Velocities

Theory

In linear theory, velocities are related in a simple way to nearby overdensities. We first derive this relation in this section and then consider some of its ramifications. In particular, we will see that by measuring both the peculiar velocity field and the density field, one can infer the present value of the matter density, Ω_m.

In linear theory, we have already derived the equation which determines the velocity field. On scales well within the horizon, the continuity equation (4.103) reduces to

$$\dot{\delta} + ikv = 0.$$

At late times, though, we have solved for the evolution of δ: we know that it scales as the growth factor D_1, so

$$v(k,\eta) = \frac{i}{k}\frac{d}{d\eta}\left[\frac{\delta}{D_1}D_1\right] = \frac{i\delta(k,\eta)}{kD_1}\frac{dD_1}{d\eta}$$

A function commonly employed to relate the velocity to the density is the dimensionless linear growth rate,

$$f := \frac{a}{D_1}\frac{dD_1}{da}$$

Since $d/d\eta = a^2 H d/da$, the velocity is related to the density via

$$v(k,a) = \frac{ifaH\delta(k,a)}{k}$$

The linear growth rate can be computed from Eq. (7.77). Most probes of the velocity field to date have been limited to relatively nearby objects, $z \lesssim 0.1$, so it is a reasonable approximation to evaluate f today and neglect its evolution. Figure 9.7 shows the linear growth rate today as a function of the matter density. For small Ω_m there is less growth: mass collapsing into overdense regions has lower velocity than if there was a critical density of matter. This makes sense since an overdensity in a lower density universe has less mass and therefore exerts a weaker gravitational pull on infalling matter. Figure 9.7 shows that, for all practical purposes, the growth rate depends only on the matter density and not, say, on the cosmological constant. Also, the approximation

$$f = \Omega_m^{0.6}$$

is seen to work extremely well (see Exercise 7 for a slightly better fit to a flat universe).

There are two important points about the relation between the density and the velocity in Eq. (9.18) that should be emphasized. First, we need to remember that

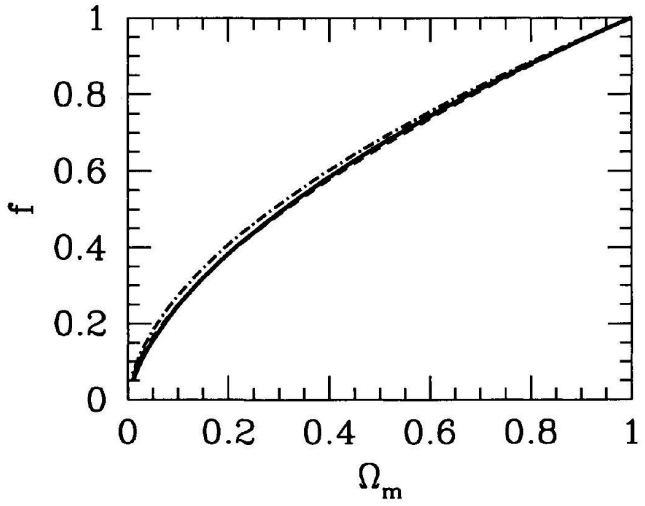


Figure 9.7. The linear growth rate as a function of matter density. There are three curves here, nearly indistinguishable. One (solid) is for an open universe with $\Omega_\Lambda = 0$; another (dot-dashed) for a flat universe with $\Omega_\Lambda = 1 - \Omega_m$; and the last (dashed) is $\Omega_m^{0.6}$.

the velocity is of course a vector and v in Eq. (9.18) is the Fourier component of the velocity parallel to $\vec{k}$. Explicitly, at low redshift we have

$$\vec{v}(\vec{k}) = ifH_0\delta(\vec{k})\frac{\vec{k}}{k^2}$$

The second point about the relation between the velocity and the density is that it holds only in linear theory. This has turned out to be a big problem for those who have tried to extract information from velocity studies. Velocities are easiest to obtain on small scales, but easiest to compare with theory on large scales.

12.10.3 Direct Measurements of Peculiar Velocities

Theory

A number of surveys have directly measured peculiar velocities. Measuring radial velocities is relatively easy: one just looks for shifted features in the spectrum of the galaxy. The hard part is breaking the radial velocity into the part due to the Hubble expansion and the remainder, the peculiar velocity. Subtracting off the Hubble expansion requires independent (i.e., other than the redshift) knowledge of the galaxy's distance from us. How accurate must a distance indicator be to be useful? Very roughly, typical peculiar velocities are of order 500 kmsec^{-1}, while the Hubble velocity is $H_0\chi = 100$ kmsec$^{-1}h(\chi/\mathrm{Mpc})$ for a galaxy a distance χ away. So a galaxy 50Mpc away has a Hubble velocity roughly 10 times as large as its peculiar velocity. To be useful, therefore, a distance indicator for such galaxies much have an accuracy of order 10%. Indeed this is roughly the best one can hope for, so 50Mpc is roughly the farthest one can hope to go in a velocity survey.

There are a number of ways of extracting cosmologically useful information from a velocity survey, but I will focus on just one of these: the two point function. With enough velocities, a survey can hope to measure the correlation function

$$\xi_v\left(\vec{x}_1, \vec{x}_2\right) := \left\langle \vec{v}\left(\vec{x}_1\right) \cdot \hat{x}_1 \vec{v}\left(\vec{x}_2\right) \cdot \hat{x}_2 \right\rangle$$

where the radial components $\vec{v} \cdot \vec{x}$ appear because these are all that can be measured using redshifts. Let us compute this correlation function using linear theory. We will see that it is an integral over the power spectrum, so - just like the angular correlation function of galaxies we considered in Section 9.1-the radial velocity correlation function is a probe of the power spectrum. The observational obstacles involved in obtaining accurate peculiar velocities are daunting. However, the promise of measuring the matter power spectrum (the velocities are due to the matter, which may or may not be aligned with the galaxies) as opposed to the galaxy power spectrum sampled by $w(\theta)$ is so alluring that it is likely that peculiar velocity surveys will continue to play an important role in cosmology.

To evaluate the velocity correlation function, we can Fourier transform the velocities appearing in Eq. (9.21) so that

$$\xi_v\left(\vec{x}_1, \vec{x}_2\right) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{x}_1} \int \frac{d^3k'}{(2\pi)^3} e^{-i\vec{k}'\cdot\vec{x}_2} \left\langle \vec{v}(\vec{k}) \cdot \hat{x}_1 \vec{v}^*\left(\vec{k}'\right) \cdot \hat{x}_2 \right\rangle$$

Using linear theory for $\vec{v}$(9.20) and the fact that $\left\langle \delta(\vec{k})\delta^*\left(\vec{k}'\right)\right\rangle = (2\pi)^3\delta^3\left(\vec{k}-\vec{k}'\right)P(k)$ leads to

$$\xi_v\left(\vec{x}_1, \vec{x}_2\right) = f^2 H_0^2 \int_0^\infty \frac{dk k^2}{(2\pi)^3} P(k) \int d\Omega_k e^{i\vec{k}\cdot\left(\vec{x}_1-\vec{x}_2\right)} \frac{\vec{k}\cdot\hat{x}_1\vec{k}\cdot\hat{x}_2}{k^4}$$

One way to do the angular integral here is to write the occurrences of $\vec{k}$ in the integrand as

$$\vec{k}e^{i\vec{k}\cdot\vec{x}} = \frac{1}{i}\frac{\partial}{\partial\vec{x}}e^{i\vec{k}\cdot\vec{x}}$$

where

$$\vec{x} := \vec{x}_1 - \vec{x}_2$$

Then taking the derivatives outside leads to an angular integral over the exponential, which has no azimuthal dependence. The integral of $e^{ikx\mu}$ over μ is is $4\pi j_0(kx)$ (Eq. (C.15)). So we have

$$\xi_v\left(\vec{x}_1,\vec{x}_2\right) = -f^2H_0^2\hat{x}_{1,2}\hat{x}_{2,j}\int_0^\infty \frac{dk}{2\pi^2k^2}P(k)\frac{\partial^2}{\partial x_i\partial x_j}j_0(kx)$$

The first partial derivative here is

$$\frac{\partial j_0(kx)}{\partial x_j} = \frac{\partial(kx)}{\partial x_j}j_0'(kx) = k\frac{x_j}{x}j_0'(kx)$$

where the prime here is derivative with respect to the argument kx. The second derivative then gives

$$\frac{\partial}{\partial x_i}\left[k\frac{x_j}{x}j_0'(kx)\right] = k^2\left[\left\{\delta_{ij}-\hat{x}_i\hat{x}_j\right\}\frac{j_0'(kx)}{kx}+\hat{x}_i\hat{x}_jj_0''(kx)\right].$$

Then, the velocity correlation function is

$$\begin{aligned}\xi_v\left(\vec{x}_1,\vec{x}_2\right) &= -f^2H_0^2\hat{x}_{1,i}\hat{x}_{2,j}\int_0^\infty\frac{dk}{2\pi^2}P(k)\left[\left\{\delta_{ij}-\hat{x}_i\hat{x}_j\right\}\frac{j_0'(kx)}{kx}+\hat{x}_i\hat{x}_jj_0''(kx)\right]\\ &:= \hat{x}_{1,i}\hat{x}_{2,j}\left\{\delta_{ij}-\hat{x}_i\hat{x}_j\right\}\xi_{v,\perp}+\hat{x}_{1,i}\hat{x}_{2,j}\hat{x}_i\hat{x}_j\xi_{v:\parallel}.\end{aligned}$$

The definitions here reflect the fact that the first term is sensitive to the component of velocity perpendicular to the line connecting two galaxies, while the second probes the velocity parallel to this line. Finally let's define the angles

$$\cos\theta_1 := \hat{x}_1\cdot\hat{x} \quad ; \quad \cos\theta_2 := \hat{x}_2\cdot\hat{x}$$

With the aid of Figure 9.8 we see that $\hat{x}_1\cdot\hat{x}_2$ is equal to $\cos\left(\theta_1-\theta_2\right)$, so performing the sums over i,j leads to

$$\xi_v\left(\vec{x}_1,\vec{x}_2\right) = \sin\theta_1\sin\theta_2\xi_{v,1}+\cos\theta_1\cos\theta_2\xi_{v,\parallel}.$$

Both components of ξ_v are integrals over the power spectrum. We can write

$$\begin{pmatrix}\xi(x)\\ \xi_{v,\perp}(x)\\ \xi_{v,\parallel}(x)\end{pmatrix} = \int_0^\infty\frac{dk}{2\pi^2k}P(k)\begin{pmatrix}k^3j_0(kx)\\ -f^2H_0^2j_0'(kx)/x\\ -f^2H_0^2kj_0''(kx)\end{pmatrix}$$

The weighted kernels are shown in Figure 9.9 and compared with the correlation function of the density. The key feature of Figure 9.9 is that, at fixed distance x, the density correlation function probes power on smaller scales than do the velocity correlation functions. Put another way, velocity surveys may be limited in how far out they go, but they pack an extra punch since they are sensitive to long wavelength modes. This is a direct result of the fact that $v\propto\delta/k$. The extra factor of $1/k$ gives additional weight to large scales.

Before going further and looking at some results from a velocity survey, I nust digress to make note of an important feature in Figure 9.9. The density correlation function, as oppossed to the velocity correlation functions, gets its contribution at a fixed distance from many Fourier modes. The figure shows the contributions to $\xi\left(50h^{-1}\mathrm{Mpc}\right)$. Naively, we would expect the main contribution to $\xi\left(50h^{-1}\mathrm{Mpc}\right)$ to come from Fourier modes with $k\sim x^{-1}=0.02h\mathrm{Mpc}^{-1}$. Since overdensities are small, i.e., still in the linear regime, on these scales, the naive expectation is

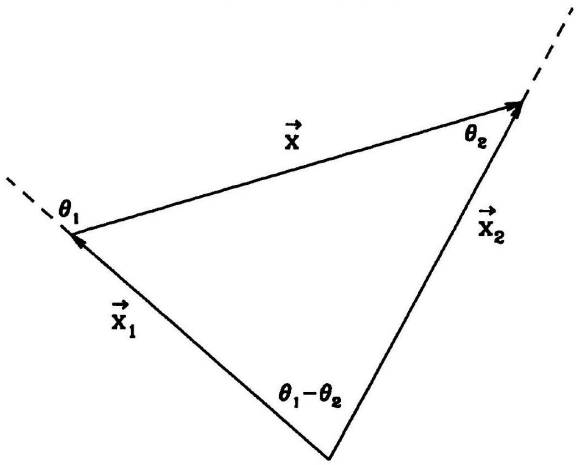


Figure 9.8. The vectors and angles of the the velocity correlation function. $\vec{x}$ is the difference vector; since the angle complementary to θ_1 is $\pi - \theta_1$, and since the three angles in the triangle must sum to π, the angle between the two galaxies is $\theta_1 - \theta_2$.

that $\xi\,(50h^{-1}\mathrm{Mpc})$ probes the linear power spectrum. Figure 9.9 shows that this expectation is incorrect. Modes with k as small as 0.02 hMpc^{-1} do contribute to the correlation function, but contributions extend out to $k \sim 0.3h\mathrm{Mpc}^{-1}$ and beyond. This means that even on scales you would think would be safely linear, the correlation function depends on the small-scale power. Ultimately we want to compare theory with observations, and we are most confident doing so for modes that are still linear. The correlation function mixes up linear and nonlinear modes, so makes it difficult to compare theory with observations. For this reason, the power spectrum has gained preeminence as the statistic of choice for large-scale structure.

Returning to the velocity correlations, let's consider Figure 9.10. It shows one attempt (Freudling et al.1999) to extract cosmological information from a velocity survey, using 1300 velocities in the all-sky SFI catalogue (Haynes et al.1999), which goes out to 70 h^{-1}Mpc. The power spectrum was parameterized by an amplitude A and the Γ parameter in the BBKS transfer function. For the analysis shown in Figure 9.10, the primordial spectral index was set to 1. The standard COBEnormalized CDM model, with critical matter density and $h = 0.5$, has an amplitude of $A = 0.29A_0$ (Eq. (7.9)) and shape parameter $\Gamma = 0.5$, so the SFI survey rules out this model at many sigma.

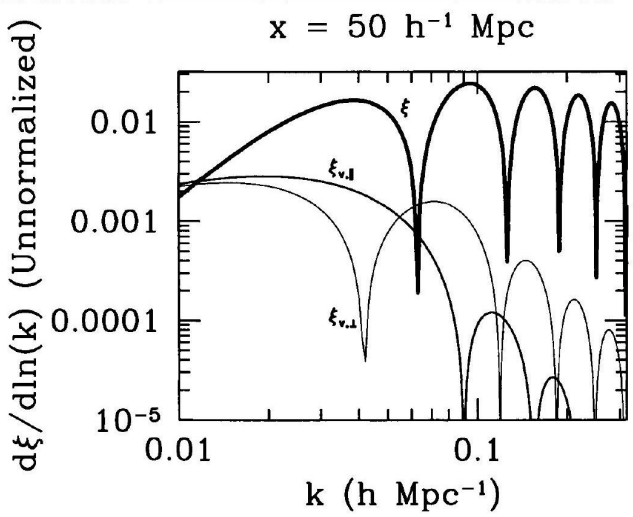


Figure 9.9. Contribution to various correlation functions from wavenumber k. Note that the velocity correlation functions get most of their contribution from smaller k than does the ordinary density correlation function.

12.10.4 Redshift Space Distortions

Theory

Redshift surveys supplement the angular information about galaxies with an estimator for the radial position, the redshift. The simplest guess about the radial position of a galaxy with redshift z is that it lies a distance

$$\chi_s(z) = \frac{z}{H_0}$$

away from us, where the subscript $_s$ denotes redshift space. Redshift space then corresponds to assigning Cartesian coordinates to a galaxy equal to

$$\vec{x}_s = \frac{z}{H_0}(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$$

This assignment neglects several unpleasant realities. First, the comoving distance out to a galaxy at redshift z is equal to z/H_0 only at relatively low redshifts. A glance back at Figure 2.3 should convince you that this approximation is off by as much as 50% at $z = 1$. Fortunately, this first problem has not yet been much of a problem since most redshift surveys to date have probed $z \lesssim 0.1$.

A second, more pernicious, problem with redshift space is that the estimate for the distance in Eq. (9.33) neglects peculiar velocities. Figure 9.11 illustrates the

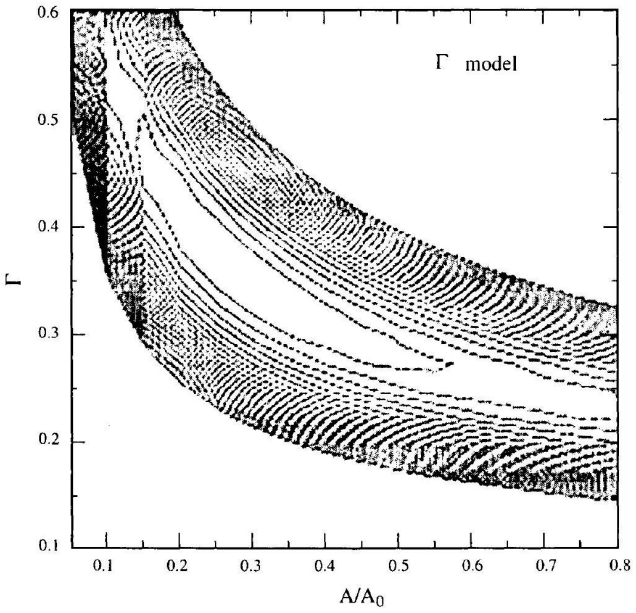


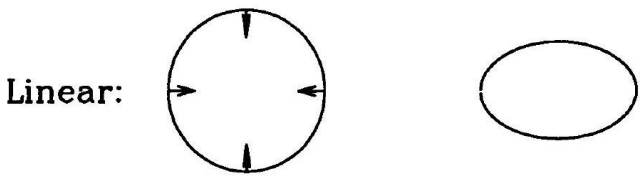
Figure 9.10. Likelihood contours on the amplitude A of the power spectrum and the shape parameter $\Gamma := \Omega_m h$ from the SFI peculiar velocity survey. The amplitude is given in units of $A_0 = 2 \times 10^6\,(h^{-1}\mathrm{Mpc})^4$, and the contours indicate shifts in the likelihood function by successive factors of $1/e$. The allowed region-delineated by banana-shaped contour in the center - is consistent with $\triangle$ CDM, but strongly disfavors standard CDM, which has $\Gamma = 0.5$ and $A = 0.29A_0$. From Freudling et al. (1999).

distortions that appear in redshift space. A slightly overdense region which is just beginning to collapse appears squashed in redshift space: the galaxies closest to us are moving toward the center of the overdense region and hence away from us, so they appear farther from us (and closer to the center of the overdense region) than they actually are. Similarly, galaxies on the "other side" of the perturbation are moving toward us, so they appear closer to us than they

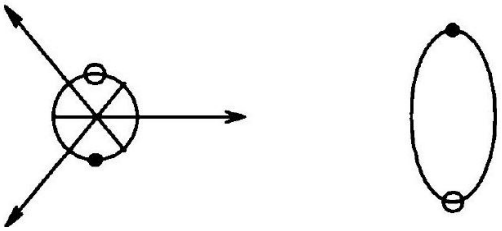
actually are. The overall effect is to induce an apparent quadrupole moment in an otherwise circular overdensity.

As a region becomes more overdense, the nature of the redshift space distortion changes. The bottom part of Figure 9.11 shows that a more collapsed object gets distorted in a much different way. It is elongated along the line of sight. More quantitatively, its quadrupole moment has the opposite sign as does a linear overdensity. It is clear then that accounting for redshift space distortions will be a tricky busi-

Real Space Redshift Space



Nonlinear Collapse:



Observer Figure 9.11. Redshift space distortions. In each case, a contour of constant density (circular in real space) is distorted in redshift space so that it looks asymmetric. Arrows denote direction and magnitude of velocity. In the case of nonlinear collapse, the velocities are so large that a point on "our side" (the bottom) of the center is mapped onto a point on the opposite side (compare the position of the solid dot on the bottom left and right).

ness, requiring careful treatment not only of linear overdensities, but also of the much more complicated effects of nonlinearities. We will content ourselves with a quantitative treatment of linear distortions, since this applies on large scales and is the starting point for all further work.

Suppose we measure the power spectrum in redshift space. How is this distorted power spectrum related to the underlying true spectrum in real space? Kaiser (1987) first solved this problem, working within the context of linear theory. The starting point is the realization that the number of galaxies in a particular region is the same, whether we use redshift-space or real-space coordinates. Therefore,

$$n_s\left(\vec{x}_s\right) d^3x_s = n(\vec{x}) d^3x$$

where n is the density of galaxies at $\vec{x}$ in real space, and n_s is the density in redshift space. The infinitesimal volume around a point in redshift space is $d^3x_s = dx_s x_s^2 \sin\theta d\theta d\phi$, while the volume around a point in real space is $d^3x = dx x^2 \sin\theta d\theta d\phi$. The angular volume elements are identical, so

$$n_s\left(\vec{x}_s\right) = n(\vec{x}) J$$

where the Jacobian J is given by

$$J := \left|\frac{d^3x}{d^3x_s}\right| = \frac{dx}{dx_s}\frac{x^2}{x_s^2}.$$

To compute the Jacobian, we use the fact that the observed redshift is the sum of two terms:

$$z = H_0 x + \vec{v} \cdot \hat{x}$$

The first term is the standard Hubble law, which says that redshift is proportional to distance; the second is the velocity along the line of sight. Recalling that redshift space corresponds to equating a galaxy's redshift with its distance from us, we see that, after dividing by H_0, this equation becomes

$$x_s = x + \frac{\vec{v} \cdot \hat{x}}{H_0}.$$

The Jacobian can be now be read off as

$$J = \left(1 + \frac{\partial}{\partial x}\left[\frac{\vec{v} \cdot \hat{x}}{H_0}\right]\right)^{-1} \left(1 + \frac{\vec{v} \cdot \hat{x}}{H_0 x}\right)^{-2}.$$

Kaiser realized that the correction term due to the derivative of the velocity is much more important than the $\vec{v} \cdot \hat{x}/H_0 x$ term. The argument goes as follows. For a plane wave perturbation, the term with the derivative of the velocity is of order $v'/H_0 \sim kv/H_0$, while the other correction is of order $v/H_0 x$. That is, the first correction term is larger than the second by a factor of order kx. Why do I say it's larger? That is, why is kx larger than unity? Kaiser's argument is that x is of order the size of the survey, while k is of order the Fourier modes we can hope to measure in the survey. Perturbations on the largest scale probed by the survey $k \sim x^{-1}$ are very poorly determined, since there are only a handful of Fourier modes with wavelength of order the survey size. Modes with smaller wavelength are much easier to measure since there are many such modes, and we effectively average over all of them to get an estimate of the power spectrum. Therefore, we are really interested only in modes with $kx \gg 1$. Expanding the remaining denominator about $v = 0$, we see that

$$J \simeq \left(1 - \frac{\partial}{\partial x}\left[\frac{\vec{v} \cdot \hat{x}}{H_0}\right]\right).$$

The number densities in real and redshift space are $n = \bar{n}(1 + \delta)$ and $n_s = \bar{n}\,(1 + \delta_s)$, respectively, with $\bar{n}$ the average number density. In light of Eq. (9.36), the overdensity in redshift space is

$$1 + \delta_s = [1 + \delta]\left(1 - \frac{\partial}{\partial x}\left[\frac{\vec{v} \cdot \hat{x}}{H_0}\right]\right).$$

Expanding to first order, we see that the overdensity in redshift space is actually a sum of the overdensity in real space and a correction due to peculiar velocity,

$$\delta_s(\vec{x}) = \delta(\vec{x}) - \frac{\partial}{\partial x}\left[\frac{\vec{v}(\vec{x}) \cdot \hat{x}}{H_0}\right]$$

Now I want to introduce the distant observer approximation. The idea is that, in most cases of interest, the direction vector $\vec{x}$ is fixed, varying little from galaxy to galaxy. To see this, go back to Figure 9.2 : $\vec{x}$ is mostly radial, with only small components in the transverse direction (proportional to the θ_1 and θ_2). As long as the galaxies are relatively close to each other in this plane, we can approximate $\hat{x} \cdot \vec{v} \to \hat{z} \cdot \vec{v}$, where $\hat{z}$ is a radial vector pointing to the center of the galaxies of interest.

In the distant observer approximation, we can compute the Fourier transform of the redshift space overdensity (here denoting the Fourier-transformed density by $\tilde{\delta}$ to avoid confusion),

$$\tilde{\delta}_s(\vec{k}) = \int d^3x e^{-i\vec{k}\cdot\vec{x}} \left[\delta(\vec{x}) - \frac{\partial}{\partial x}\left[\frac{\vec{v}(\vec{x})\cdot\hat{z}}{H_0}\right]\right]$$
$$= \tilde{\delta}(\vec{k}) - if\int d^3x e^{-i\vec{k}\cdot\vec{x}}\frac{\partial}{\partial x}\left[\int \frac{d^3k'}{(2\pi)^3}e^{i\vec{k}'\cdot\vec{x}}\tilde{\delta}\left(\vec{k}'\right)\frac{\vec{k}'}{k'^2}\cdot\hat{z}\right],$$

the first equality following from our Fourier convention and Eq. (9.43) and the second from Eq. (9.20) for linear velocities. The derivative with respect to the length x acts on the exponential, bringing down a factor of $i\vec{k}'\cdot\hat{x}$, which we again set to $i\vec{k}'\cdot\hat{z}$, so

$$\tilde{\delta}_s(\vec{k}) = \tilde{\delta}(\vec{k}) + \int \frac{d^3k'}{(2\pi)^3}\tilde{\delta}\left(\vec{k}'\right)\left[f\left(\hat{k}'\cdot\hat{z}\right)^2\right]\int d^3x e^{i\left(\vec{k}'-\vec{k}\right)\cdot\vec{x}}$$

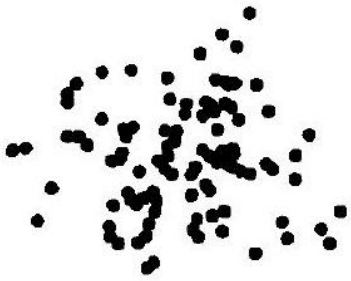
The $\vec{x}$ integral gives a Dirac delta function, equating $\vec{k}'$ with $\vec{k}$. Therefore, in the distant observer approximation,

$$\tilde{\delta}_s(\vec{k}) = \left[1 + f\mu_{\mathbf{k}}^2\right]\tilde{\delta}(\vec{k})$$

Here $\mu_{\mathbf{k}}$ is defined to be $\hat{z}\cdot\hat{k}$, the cosine of the angle between the line of sight and the wavevector $\hat{k}$. Equation (9.46) quantifies what we should have anticipated about (linear) redshift space distortions. First of all, since $f\mu_{\mathbf{k}}^2 \geq 0$, the apparent overdensity in redshift space is larger than in real space. This is clear from Figure 9.12. The central region of the galaxies is clearly more overdense in redshift space than in real space, the enhancement due to the illusion that infalling galaxies are located close to the center. The second feature of Eq. (9.46) worth noting is that the enhancement is for waves with wavevector parallel to the line of sight. A plane wave perturbation with $\vec{k}$ perpendicular to the line of sight - one in which the density along the line of sight is constant - experiences no redshift space distortion.

The power spectrum in redshift space depends not only on the magnitude of $\vec{k}$ but also on its direction, which we are parameterizing with μ. It follows immediately from Eq. (9.46) that

Real Space



Redshift Space

Figure 9.12. A hundred galaxies in real space squashed in redshift space due to linear velocities. The apparent overdensity in redshift space is much larger near the center than it is in real space. We, the observers, are sitting at the bottom of the page.

$$P_s(\vec{k}) = P(k)\left[1 + \beta\mu_{\mathbf{k}}^2\right]^2$$

Here I have introduced the parameter β, which you might think is simply equal to f, the linear growth rate. There is an additional factor in β, though, due to the fact that the mass overdensity δ is not necessarily equal to the overdensity in galaxies, δ_g. The velocity field samples the mass overdensity. So if we define the bias

$$b := \frac{\delta_g}{\delta}$$

then $\vec{v} \propto \delta \propto \delta_g/b$. Therefore, the correction due to redshift space distortions in Eq. (9.47) is proportional to

$$\beta = \frac{f}{b} \simeq \frac{\Omega_m^{0.6}}{b}$$

The redshift space distortion in the power spectrum, encoded in Eq. (9.47), is both good news and bad news. The good news is that by measuring the distortion in the redshift space power spectrum, we can hope to measure β, a combination of the density and bias. A quantitative way to do this is to measure the ratio of the quadrupole to the monopole of the power spectrum:

$$\frac{P_s^{(2)}(k)}{P_s^{(0)}(k)} := \frac{5\int_{-1}^{1} \frac{d\mu_{\mathbf{k}}}{2} \mathcal{P}_2\left(\mu_{\mathbf{k}}\right) P_s(\vec{k})}{\int_{-1}^{1} \frac{d\mu_{\mathbf{k}}}{2} \mathcal{P}_0\left(\mu_{\mathbf{k}}\right) P_s(\vec{k})}$$

Recall that $\mathcal{P}_l$ is the Legendre polynomial of order l, while P_s is the redshift-space power spectrum. Since (Exercise 11)

$$\left(1+\beta\mu_{\mathbf{k}}^2\right)^2 = \left[1+\frac{2}{3}\beta+\frac{1}{5}\beta^2\right]\mathcal{P}_0\left(\mu_{\mathbf{k}}\right) + \left[\frac{4}{3}\beta+\frac{4}{7}\beta^2\right]\mathcal{P}_2\left(\mu_{\mathbf{k}}\right) + \frac{8}{35}\beta^2\mathcal{P}_4\left(\mu_{\mathrm{k}}\right)$$

the orthogonality of the Legendre polynomials (Eq. (C.2)) implies that the ratio of the quadrupole to the monopole in linear theory is

$$\frac{P_s^{(2)}(k)}{P_s^{(0)}(k)} = \frac{\frac{4}{3}\beta+\frac{4}{7}\beta^2}{1+\frac{2}{3}\beta+\frac{1}{5}\beta^2}$$

Figure 9.13 shows Hamilton's efforts to measure the quadrupole-to-monopole ratio in two different redshift surveys. In both cases, nonlinearities are very important

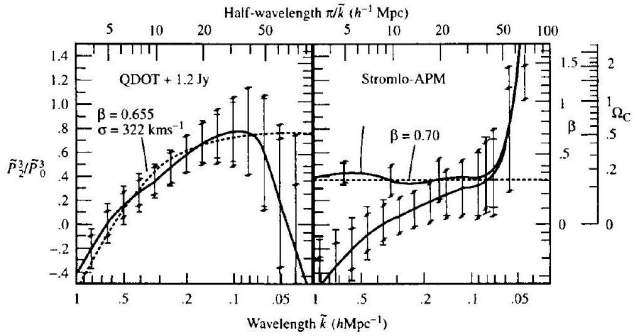


Figure 9.13. The quadrupole-to-monopole ratio for two redshift surveys (Hamilton, 1998). Left panel shows data from two redshift surveys which selected galaxies from the Infrared Astronomical Satellite (IRAS): QDOT (Lawrence et al., 1999) picked one out of six galaxies brighter than 0.6 Jansky and the 1.2Jy survey (Strauss et al., 1992) picked all galaxies above

that brightness limit. Stromlo-APM survey (Loveday et al., 1996) in right panel contains redshifts for 1 in 20 galaxies seen in APM (Figure 9.1).

and must be handled carefully. In the infrared-selected surveys depicted at left in the figure, Hamilton modeled the nonlinearities with a parameter measuring the small-scale velocity dispersion, σ. Only on scales of order $k \sim 0.1h\mathrm{Mpc}^{-1}$ does the ratio begin to asymptote to ~ 0.75, implying a value of $\beta \sim 0.7$. The right panel shows a survey of optically selected galaxies. For these, the ratio does not seem to asymptote at all (lower curve) unless nonlinear structures are removed by hand (upper curve). In that case, Hamilton finds a quadrupole-to-monopole ratio closer to 0.4, implying $\beta = 0.3$. Figure 9.14 gives a broader view of the spread in measures of β from redshift and peculiar velocity surveys.

That was the good news. The bad news is that, even if we were to give up hope of measuring β from redshift surveys, we still need to account for redshift space distortions if we want to measure the power spectrum. Blindly measuring the power spectrum by averaging over all directions $\mu_{\mathbf{k}}$ is actually a measure of

$$P_s^{(0)}(k) = \left[1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right] P(k),$$

the equality holding only in linear theory. That is, $P_s^{(0)}$ overestimates the power spectrum by as much as a factor of 2. On even moderate scales, as suggested by Figure 9.13, nonlinear effects must be taken into account. The Peacock and Dodds compilation in Figure 7.11 uses a model of the small-scale velocities to do this.

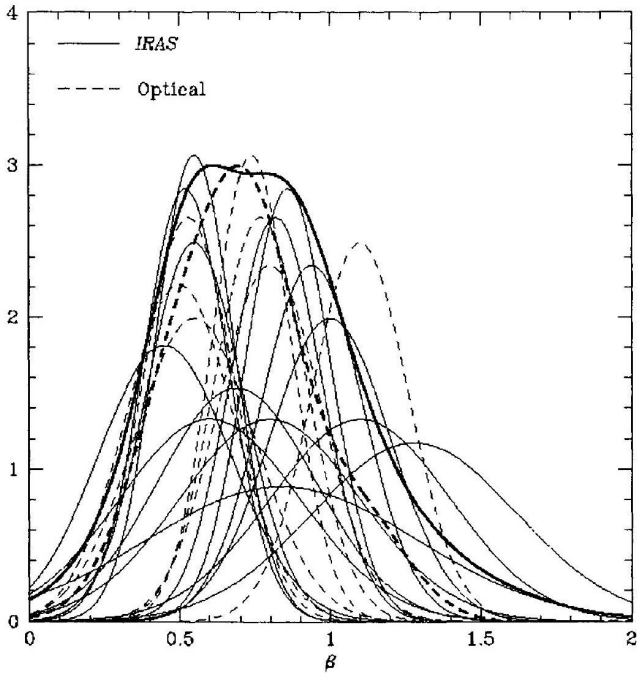


Figure 9.14. Compilation of the likelihood of β from redshift and peculiar velocity surveys (Strauss and Willick, 1995).

12.10.5 Galaxy Clusters

Theory

Until now, we have focused solely on the two-point function: the angular correlation function, the velocity covariance, and the power spectrum. You might have wondered why little has been said about one-point functions, the number density of galaxies for example. To answer this

question, first consider an extreme example. How many people are there in the universe? This clearly is an impossible question to answer with the tools we have developed. Putting aside the thorny question of the definition of a "person," we still would have to develop theories of star formation out of the gas in galaxies, then planet formation around stars, then the evolution of life via various biological processes. A prediction of the "person density" in the universe is beyond the scope of... this book, to say the least.

In a similar, but less extreme, way, a prediction of the galaxy density of the universe from what we have learned about the distribution of matter in the universe involves complicated issues we do not have the tools to address. What fraction of the matter has collapsed into nonlinear structures? How do these nonlinear structures evolve? Do galactic-size nonlinear structures merge? If so, how often? Upon collapse, how do stars form? How are stars distributed? In spiral patterns? Elliptical? One of the exciting developments of the 1990s was the evolution of a number of techniques to answer these questions. In addition to the brute-force approach of numerical simulations, several groups (e.g., Kauffman et al., 1999; Somerville and Primack, 1999; Colberg et al., 2000; Benson et al., 2001; Cooray and Sheth, 2002) developed semianalytic techniques which have been remarkably successful at predicting properties and abundances of different galaxy types. Although we will not study these techniques directly here, the one technique we will encounter - the Fress-Schechter formalism - forms the basis for much of this work.

Perhaps the fundamental difficulty encountered by one attempting to make predictions about the number density of galaxies is that the galactic scale has already gone nonlinear. Recall from Ex. 7.9 that scales smaller than $\sim 10h\mathrm{Mpc}^{-1}$ have gone non-linear. What scale in the unperturbed universe encloses the mass of a typical galaxy, $M = 10^{12}M_\odot$? The density in a spherical region of radius R is

$$\rho_{\mathrm{m}} = \frac{M}{4\pi R^3/3}$$

Since $\rho_{\mathrm{m}} = \Omega_m \rho_{\mathrm{cr}}$, we can invert this to find

$$R = 0.951 h^{-1}\mathrm{Mpc}\left(\frac{Mh}{10^{12}\Omega_m M_\odot}\right)^{1/3}$$

So a galaxy comes from matter within a radius of about 1Mpc, corresponding to fluctuations on scales of order $k \sim 1h\mathrm{Mpc}^{-1}$, well into the nonlinear regime.

This answers a question ("Why not try to predict the number density of galaxies?") but begs another: Are there objects, corresponding to scales closer to the linear regime, for which one might be able to make reliable predictions about abundances? If we invert Eq. (9.55) to get the mass enclosed within a sphere of radius R

$$M = 1.16 \times 10^{15}\Omega_m h^{-1} M_\odot \left(\frac{R}{10h^{-1}\mathrm{Mpc}}\right)^3$$

then we see that clusters of galaxies - with masses up to $10^{15}M_\odot$ - arise from perturbations on just the right scales.

How, then, to predict the abundance of galaxy clusters? The basic insight comes from a paper by Press and Schechter (1974), and the resulting framework is called Press-Schechter theory. To understand their argument, consider the onedimensional density field in Figure 9.15. The average inhomogeneity is zero, of course. There are regions with relatively large excursions in both the positive and negative direction. Underdensities cannot get smaller than -1 (when the density is zero), but there are some regions in the figure with densities more than three times the average density. It is these rare regions of large overdensity that we are interested in: they collapse, accumulate so much excess matter that local gravity becomes more important

than the Hubble flow. Particles in this region stop expanding away from each other and are trapped in the local gravitational field.

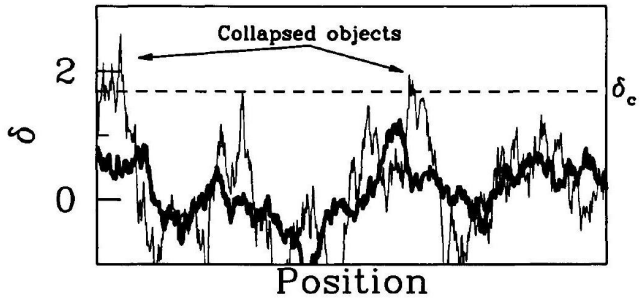


Figure 9.15. Inhomogeneities as a function of 10 position. Dark curve is the same field smoothed on larger scales. Several small-scale fluctuations have collapsed, while the large scale density field does not have any overdensities greater than δ_c, the critical value for collapse.

The Press-Schechter theory predicts the fraction of the volume that has collapsed,

$$f_{\rm coll}\,(M(R),z) = \frac{2}{\sqrt{2\pi}\sigma(R,z)}\int_{\delta_c}^{\infty} d\delta e^{-\delta^2/2\sigma^2(R,z)}.$$

Here R is the radius over which the density field has been smoothed. This radius is used to compute $\sigma(R,z)$, the rms of the smoothed density field (Ex. 7.9). As you can see from Figure 9.15, the smoothing scale matters. Typically, inhomogeneities on large scales are smaller in magnitude than those on small scales, so small scales collapse first. As time evolves, overdensities grow (proportional to the growth function), so eventually some large-scale inhomogeneities will also collapse. The right-hand side of Eq. (9.57) counts all parts of the Gaussian distribution for which the overdensity is greater than some critical density δ_c. There are several pieces of magic in this formula. First, it assumes that the distribution of inhomogeneities is Gaussian. This is impossible since δ, by definition, never gets smaller than -1. And indeed, it is possible to show that gravity skews an initially Gaussian distribution, producing more underdense regions and a nonnegligible tail of highly overdense regions. Second, the normalization is a bit of a cheat: one would naively not include the factor of 2. Finally, the rms σ in the formula is the linear rms, specifically ignoring nonlinear effects. On small scales, there is a huge difference between σ calculated with the linear power spectrum and that with the true nonlinear spectrum. Press-Schechter tells us that the collapsed fraction can be obtained using the linear σ. These peculiarities of the Press-Schechter formalism do not detract from its effectiveness. Numerical simulations (e.g., White, Efstathiou and Frenk, 1993) have shown that it works extremely well. Further, a number of groups (Peacock and Heavens, 1990; Bond et al., 1991) have justified theoretically some of the aspects of the formula that initially appeared ad-hoc.

To get the collapsed fraction into a form more comparable with observations, first differentiate $f_{\rm coll}$ with respect to M and multiply by a small interval dM. This gives the fraction of the volume collapsed into objects with mass between M and $M + dM$. Multiply this by the average number density of such objects $\rho_{\rm m}/M$ to get the the number density of collapsed objects with mass between M and $M + dM$,

$$dn(M,z) = -\frac{\rho_{\rm m}}{M}\frac{df_{\rm coll}(M(R),z)}{dM}dM.$$

The minus sign appears here since $f_{\rm coll}$ is a decreasing function of the mass M. Carry out the derivative using the fact that $dM/dR = 3M/R$. Then,

$$\frac{dn(M,z)}{dM} = \sqrt{\frac{2}{\pi}}\frac{\rho_{\rm m}\delta_c}{3M^2\sigma}e^{-\delta_c^2/2\sigma^2}\left[-\frac{R}{\sigma}\frac{d\sigma}{dR}\right].$$

The term in brackets here, the logarithmic derivative, is close to 1 for most models of interest. The dominant factor in Eq. (9.59), at least for large masses, is the exponential. If σ on a given scale is small, then the number density of collapsed objects on that scale is exponentially suppressed.

Until now, I have sidestepped the question of the numerical value of δ_c, the critical overdensity above which objects collapse. There are two approaches to obtaining δ_c; fortunately, both appear to agree. The first is to rely on a simple model of collapse, a model in which the overdensity is perfectly spherical. One can show that collapse occurs in such a model when $\delta = 1.686$ (for $\rho_{\rm m} = \rho_{\rm cr}$). The other way is simply to treat δ_c as a free parameter and calibrate it with numerical simulations (e.g., Eke, Cole and Frenk, 1996). The δ_c obtained this way is close enough to the spherical value that typically one simply adopts $\delta_c = 1.686$.

Measuring the cluster abundance, and therefore testing theories with Eq. (9.59), is a subtle business. One class of difficulties is identifying a cluster. Sophisticated algorthims have been developed to find clusters in a galaxy survey. The second set of difficulties revolves around determining the mass of the cluster. There are several methods of mass determination:

- X-Ray Temperatures Hot ionized gas in a cluster emit radiation with a cutoff frequency which is determined by the temperature of the gas. This temperature can be related to the mass of the cluster under certain assumptions.
- Sunyaev-Zeldovich Distortion Photons from the CMB pasing through clusters get scattered by the hot gas. This scattering distorts the CMB spectrum as a function of frequency, inducing a decrement at low frequency and excess at high frequency (low-energy photons gain energy from the hot electrons). The shape of the distortion is fixed; the amplitude is another measure of the temperature of the gas, which, again under certain assumptions, can be translated into a measurement of the mass. - Weak Lensing Images of background galaxies are distorted by a foreground cluster. The larger the mass of the cluster, the larger the distortions. Weak lensing is therefore becoming a fabulous tool for measuring masses of clusters without using the temperature.

At least in the first two techniques, the direct measurement is of the cluster temperature. So let's work through a relation between the mass and temperature of a cluster under a simple set of assumptions. Suppose a cluster has virialized so that its kinetic energy is equal to minus half its potential energy. Suppose also that the cluster is spherical with radius $R_{\rm vir}$ with potential (gravitational) energy is equal to $-3GM^2/5R_{\rm vir}$. Then,

$$\frac{1}{2}Mv^2 = \frac{3}{10}\frac{GM^2}{R_{\rm vir}}$$

The overdensity of the cluster $\Delta_{cl} := \rho_{cl}/\rho_{\rm m}$ then allows us to eliminate the radius, since

$$\Delta_{cl} := \frac{M}{4\pi R_{\rm vir}^3\rho_{\rm m}/3}.$$

The temperature is equally apportioned among three directions, so the average velocity squared is $v^2 = 3T/m_p$, where m_p is the proton mass. The temperature can now be expressed in terms of the total mass of the system,

$$T = \frac{m_p}{5}\left[GMH_0\sqrt{\frac{\Delta_c}{2}}\right]^{2/3}$$

when $\Omega_m = 1$. Invert to get

$$M = 8.2 \times 10^{13} h^{-1} M_\odot \left(\frac{T}{\mathrm{keV}}\right)^{3/2} \sqrt{\frac{178}{\Delta_{cl}}}.$$

Here I have normalized Δ_{cl} by its value in the spherical collapse model (with $\rho_{\rm m} = \rho_{\rm cr}$), although again simulations have verified the numerical value.

Figure 9.16 shows the cluster density as a function of temperature for the standard CDM model with $\Omega_m = 1$. One of the most important points made immediately after the COBE detections of anisotropies in 1992 was that this plain-vanilla model predicts too many clusters. Indeed, the abundance of clusters today is often used as an excellent way of normalizing a power spectrum. A typical value for σ_8 from cluster abundances (e.g., Wang and Steinhardt, 1998) is

$$\sigma_8 = 0.5 \Omega_m^{-.33-.35\Omega_m}$$

with error estimates ranging from just a few percent up to 20%, the latter probably more accurately reflecting uncertainties in the mass determinations.

Another exciting application of the Press-Schechter prediction for cluster abundances is the evolution with redshift. The basic point stems from the exponential.

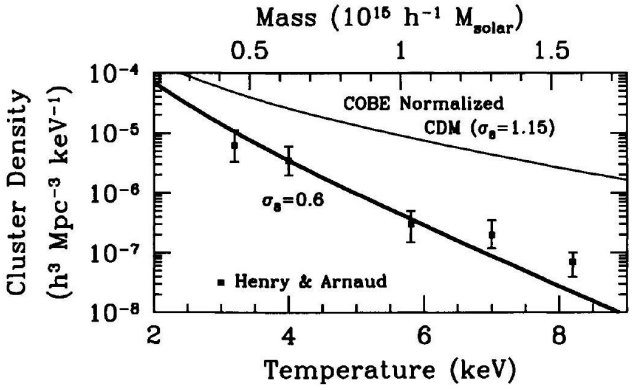


Figure 9.16. The cluster density as a function of temperature. Data from Henry and Arnaud (1991). The two theoretical curves are the Press-Schechter estimates (Eq. (9.59)) of models with $\Omega_m = 1, h = 0.5, \Omega_b = .05$. The only difference between the two is the normalization.

dependence in Eq. (9.59). At high redshift, $\sigma(R, z)$ is necessarily smaller, but how much smaller depends on the underlying cosmology. The growth is fastest in a model with $\Omega_m = 1$ (see Figure 7.12), so for a fixed abundance today, one expects many fewer clusters in such a model. Figure 9.17 illustrates this effect. An $\Omega_m = 1$ model predicts a factor of 700 fewer clusters with masses greater than $3.5 \times 10^{14} M_\odot$ at redshift 0.5 than at redshift zero. By contrast, Λ CDM predicts just a factor of 4 decline. With the observational assault on clusters just getting off the ground, we can expect strong constraints on cosmology to emerge in the coming decade.

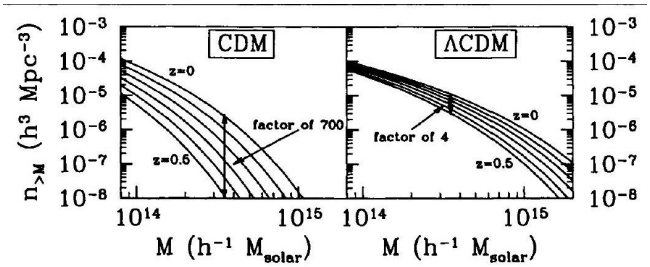


Figure 9.17. The number density of clusters with mass greater than M at different redshifts. Left panel: CDM model with $\Omega_m = 1, h = 0.5, \Omega_b = 0.05$ but normalized to give the correct abundance at $z = 0\,(\sigma_8 = 0.5)$. Curves give the abundances in redshift increments of $\delta z = 0.1$ starting from $z = 0$ going out to $z = 0.5$. Right panel: Λ CDM model which fits CMB and other data ($\Omega_m = 0.35, h = 0.7, \Omega_b h^2 = 0.02$). Note the relatively slow evolution of clusters with mass $\sim 3 \times 10^{14} M_\odot$ as compared with the critical density model.

12.10.6 Lyth bound (??)

(wiki, later I'll learn this better)

In cosmological inflation, within the slow-roll paradigm, the Lyth argument places a theoretical upper bound on the amount of gravitational waves produced during inflation, given the amount of departure from the homogeneity of the cosmic microwave background (CMB).

Summary

- During slow-roll inflation, the ratio of gravitational waves to inhomogeneities of the CMB is correlated to the inflationary potential steepness.

- Temperature inhomogeneities of the CMB were successfully and accurately measured. [1][2][3] in the CMB.

- There are current CMB polarization experiments [4][5] (see this article for instance for an overview of gravitational wave observatories) aimed at measuring the primordial gravitational wave signature in the CMB.

- However, to date, a significant signal of primordial gravitational waves was not detected. Thus the ratio cannot exceed a certain value.

- Thus the steepness of the inflationary potential is bounded.

12.11 Weak Lensing and Polarization

The traditional methods of measuring clustering - angular and redshift surveys are powerful probes of the power spectrum but share a common deficiency. They are measures of the distribution of galaxies, not the distribution of mass. Theories of the early universe can make very accurate predictions about the latter, but not about the former. A very exciting new technology which probes the mass - not the light distribution is introduced in this chapter. We will see that the inhomogeneities of the matter induce distortions in the observed shapes of distant galaxies due to gravitational lensing. Further, the statistics of these distortions are directly related to the matter power spectrum.

The anisotropies in the CMB are subject to none of the uncertainties or ambiguities which plague the density field.

- We know exactly where the CMB comes from (the surface of last scattering) so there is no analogue of peculiar velocity distortions.

- There is nothing like the mass vs light problem which afflicts the interpretation of galaxy surveys.

- In addition, the mass distribution has gone nonlinear, so a simple comparison of the linear power spectrum derived in Chapter 7 with the data is dangerous. Anisotropies in the CMB are still at the part-in-a-hundred-thousand level, so nonlinearities are for the most part irrelevant.

The C_l 's then are easy to interpret and extract information from. Nonetheless, here too we can go beyond what we have already done. Until now we have focused on anisotropies in the

temperature field. Compton scattering before decoupling also induced polarization anisotropies. Polarization opens up a new dimension in the study of the CMB. At the very least, it doubles the amount of information contained in the CMB. As we will see in this chapter, the promise of polarization goes well beyond this doubling. Gravity waves – tensor perturbations– produce a particular pattern of polarization that cannot be mimicked by scalar perturbations. Therefore, polarization offers a unique way of searching for gravity waves produced during inflation. Gravitational lensing and polarization belong in the same chapter primarily because the mathematics describing them is so similar. Both effects can be quantified with a two-by-two symmetric matrix. In lensing, this matrix is the distortion tensor encoding information about image distortion. The polarization tensor has a longer history with more famous components Q and U. It is identical mathematically, though. So, the technologies used to study both of these effects are very similar.

12.11.1 Gravitational Distortion of Images

Theory

The cosmological gravitational field distorts the paths traveled by light from distant sources to us. This fundamental fact carries with it an enormous amount of cosmological promise. Most important is the idea that light paths respond to mass. If we can measure distortions, then, we might be able to infer something about the distribution of mass in the universe. The importance of this inference cannot be overstated: most of what we think we know about this distribution comes from our observations of the galaxy distribution. We hope that, on large scales at least, the two - the mass and the galaxy distribution - are not too different. If we observe the mass distribution directly via distortion of light rays, though, then we need not rely on this hope. We can then directly compare observations with theoretical predictions. For cosmology, therefore, we expect the most important aspect of light ray distortion to be weak lensing, wherein the shapes of distant galaxies are distorted (slightly) by intervening foreground mass overdensities. We begin with an overview of image distortion along with a brief discussion of some other applications.

The idea that gravitational fields might distort distant images is as old as general relativity. Indeed, even before Einstein finalized general relativity, he understood the importance of measuring this distortion. Early notebooks of his contain calculations of the magnification of images and of the possibility of a double image of a single source (Renn, Sauer, and Stachel, 1997). And it was detection of gravitational distortion that led to the acceptance of general relativity. In 1919, Eddington led a voyage to the Southern Hemisphere to observe the deflection of starlight during a solar eclipse. The magnitude of this effect (Dyson, Eddington, and Davidson, 1920) was in good agreement with Einstein's new theory.

One of the most spectacular manifestations of gravity bending light paths is strong gravitational lensing. In 1979, Walsh, Carswell, and Weymann observed a multiply imaged quasar, thereby confirming Einstein's early speculations. Light rays leaving the quasar in different directions are focused on the same point (us) by an intervening galaxy. Since then, dozens of multiply imaged quasars have been observed, and we are on the verge of discovering many hundred more in the near future. Exactly what fraction of quasars is lensed is a question that may depend on the background cosmology. In particular, it has been argued that there should be more multiply imaged quasars in a universe with a cosmological constant than in one without (see Exercise 1 and Kochanek, 1996). There are other examples of gravitational lensing that have an impact on cosmology. Light rays that take different routes to the same endpoint typically arrive at that endpoint at different times. Therefore, two light rays emitted from the same source at the same time which we detect from different directions due to lensing typically arrive at different times. We can measure this time delay by studying sources with variable

emission. The delay turns out generally to depend on the Hubble constant, so astronomers have made very accurate measures of H_0 looking for time delays (e.g., Kundic et al., 1997). Another example is microlensing, wherein a lens moves into the line connecting a source to us. When it does, the image is magnified, so that we observe a characteristic variability in the distant source. Microlensing has been used in recent years to find massive compact halo objects (MACHOs) in our galaxy (Alcock et al., 1993). It now appears that MACHOs do not make up all, or even most, of the dark matter in our galaxy. Nonetheless, exactly what and where they are is still a mystery of cosmic significance.

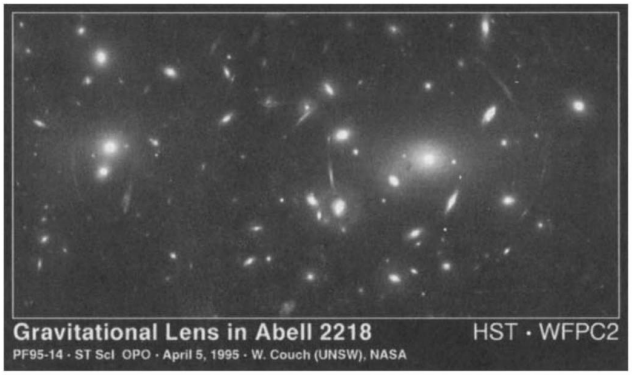
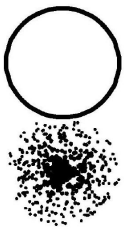
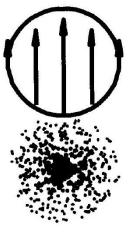


Figure 10.1. Foreground galaxies in the cluster Abell 2218 distort the images of background galaxies. Elliptical arcs surround the central region of the cluster at right.

Yet another spectacular manifestation of gravitational lensing is shown in Figure 10.1. The large cluster in the foreground, Abell 2218, distorts the shapes of the background galaxies. This leads to a distinctive pattern of elliptical arcs surrounding the central region of the cluster. Why do the background galaxies appear stretched out elliptically in Figure 10.1? Consider a circular galaxy sitting behind a large mass density with an observer out of the page as in Figure 10.2a. Since the light rays are distorted, we do not expect to see a circular image. Rather, light rays coming from the "bottom" of this source - the ones that pass closer to the central mass region - are bent more than those that do not come as close to the mass. The light rays are bent such that objects at the bottom appear to be farther away from



(a)



(b)

(c)

Figure 10.2. (a) Circular galaxy, the source, sits behind a foreground mass distribution represented by points at bottom. The observer is out of the page so that the foreground mass is between the observer and the source. (b) Light rays from source are deflected as they pass by mass distribution. Rays traveling closest to mass get deflected the most. (c) Resulting image is an arc.

the mass. (This is the only subtle part of the argument: rays are bent toward the mass distribution, so that as you extrapolate backward, the source appears farther away. See Figure 10.4.) Images will therefore be distorted as in Figure 10.2b. The net effect, therefore, is to turn a circular galaxy into the arc shown in panel c.

A very active field of research uses background galaxies to try to infer the mass distribution of clusters (e.g., Clowe et al., 1998). Most times, the images are not as dramatic as those shown in Figure 10.1. The lack of drama is offset by the huge numbers of background galaxies. By adding up many small distortions, observers have succeeded in obtaining mass estimates for a number of clusters. This idea of statistically averaging small distortions is the hailmark of weak lensing. The mass estimates are important information for cosmologists: several cosmological constraints are based on cluster masses and abundances (e.g., Section 9.5, Carlberg et al., 1997; Bahcall et al., 2000).

We will be interested in weak lensing not by a single identifiable lens such as a cluster, but rather by the generic large-scale structure in the universe. Inferring the distribution of the dark matter - i.e., pointing to a spot on the sky and identifying it as an overdense region - is not necessarily the goal. Rather, we should be satisfied if we can measure some simple statistics, for example the correlation function or its Fourier transform, the power spectrum. Indeed, these are the quantities we, as cosmologists, are most interested in anyway. We don't care where the overdense and underdense regions are; we simply want to compare theory with observations. So our main goal here is to relate the observations (which have already begun) of distortions of galaxy images to the underlying mass power spectrum. (a)

(b) (c)

Figure 10.3. Different lensing patterns. Panels (a) and (b) could be produced by a mass distribution (a) above or below the distorted images and (b) in between or on either side of the images. But the alignment in panel (c) could not be produced by lensing.

One final note concerning the correlations of distortions expected from gravitational lensing. We might expect to find the images of two (circular) galaxies to be distorted so that they look like Figure 10.3a if, for example, there is a large overdensity above or beneath this galaxy pair.

We might also expect images similar to those in Figure 10.3b if an overdensity exists between them or to either side. However, lensing cannot produce the alignment sketched in panel (c). This fact, which we will shortly prove, is often used as a check against systematic problems afflicting an observation.

12.11.2 Geodesics and Shear

Theory

We want to solve for the path of a light ray as it leaves a distant source and travels through the inhomogeneous universe. Figure 10.4 shows the geometry and notation, which will be similar to that set up in our discussion of the angular correlation function. The position of the photon at any time is given by $\vec{x}$, with the x_3 component equal to the radial distance χ and the transverse components equal to $\chi\vec{\theta}$. The intensity we observe from a source is

$$I_{\rm obs}(\vec{\theta}) = I_{\rm true}\ \left(\vec{\theta}_S\right)$$

a source whose image appears at $\vec{\theta}$ is actually at $\vec{\theta}_S$.

To solve for the path of a light ray, we need to use the machinery of general relativity. Recall that in Chapter 4, we used the time component of the geodesic equation to find dp/dt, the rate of change of the magnitude of the momentum. Here we are interested in deflections, so we will need the spatial component,

$$\frac{d^2x^i}{d\lambda^2} = -\Gamma^i_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda}$$

in particular, we will need the transverse part. Let's first consider the left side of this equation. We can express the derivatives with respect to affine parameter λ in terms of derivatives with respect to χ using the fact that

$$\frac{d\chi}{d\lambda} = \frac{d\chi}{dt}\frac{dt}{d\lambda}$$

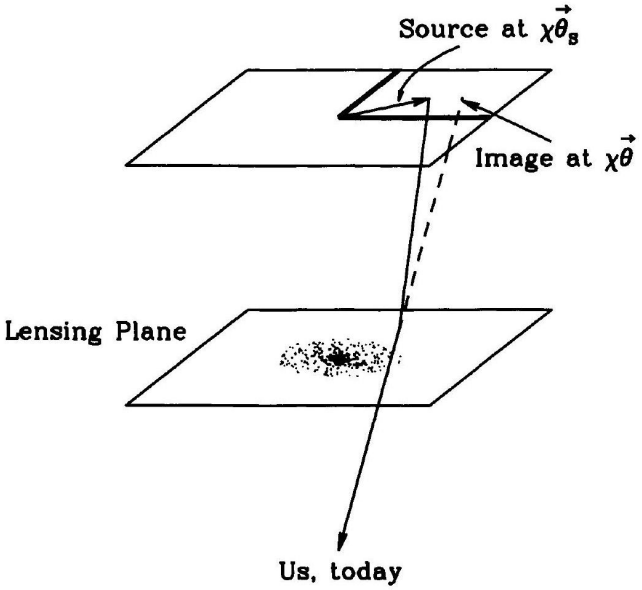


Figure 10.4. A light ray leaving a distance source is distorted as it passes by an intervening overdense region. At all times, the position of the light ray can be characterized by a $2D$ vector

specifying its angular distance from the center of the lens. The ray starts with angular vector $\overrightarrow{\theta_S}$, but appears to us to be coming from $\vec{\theta}$.

$$= \frac{-1}{a} p(1-\Psi)$$

The first part of this equality ($d\chi/dt = -1/a$) follows from Eq. (2.42), while the second part comes from Eq. (4.14). The transverse part of $\vec{x}^i$ is equal to $\chi\theta^i$, so the left-hand side of the geodesic equation is

$$\frac{d^2x^i}{d\lambda^2} = \frac{1}{a} p \frac{d}{d\chi}\left[\frac{p}{a}\frac{d}{d\chi}\left(\chi\theta^i\right)\right]$$

Here I have dropped the (small) gravitational potential because it multiplies the (small) angle θ^i. We can reduce further by remembering that (at zero order) the momentum p times a remains constant, so removing pa outside the derivative leads to

$$\frac{d^2x^i}{d\lambda^2} = p^2 \frac{d}{d\chi}\left[\frac{1}{a^2}\frac{d}{d\chi}\left(\chi\theta^i\right)\right]$$

Now let's consider the right side of the geodesic equation. Again, changing the derivatives with respect to λ to those with respect to χ leads to

$$\Gamma^i_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda} = \left(\frac{p}{a}\right)^2 (1-\Psi)^2 \Gamma^i_{\alpha\beta}\frac{dx^\alpha}{d\chi}\frac{dx^\beta}{d\chi}.$$

There are three types of terms in the sum over α and β: those with $\alpha = \beta = 0$, those with one index spatial and the other temporal, and finally those in which both α and β are spatial. We have already derived the relevant Christoffel symbols, given in Eq. (5.7). Let's work through the terms one by one:

- When $\alpha = \beta = 0$, we have

$$\Gamma^i_{00}\left(\frac{dt}{d\chi}\right)^2 = \Psi_{,i} = -\Phi_{,i}$$

where the second equality holds since in the late universe there are no anisotropic stresses so $\Phi = -\Psi$.

- When one of the indices is spatial, Γ^i_{0j} is nonzero only when $i = j$. Therefore, the spatial index j must be transverse with $x^j = \chi\theta^j$. Since θ^j is small, we can drop the potential in the Christoffel symbol leading to

$$\Gamma^i_{0j}\frac{dt}{d\chi}\frac{dx^j}{d\chi} = -aH\frac{d}{d\chi}\left[\chi\theta^i\right]$$

with of course an identical term coming from Γ^i_{j0}.

- When both indices are spatial, the Christoffel symbol is proportional to the (small) gravitational potential. When multiplied by the small transverse distance, these terms will be negligible, so we need consider only the term $\Gamma^i_{jk}\left(dx^j/d\chi\right)\left(dx^k/d\chi\right)$ in which $j = k = 3$ along the radial direction. In that case $x^3 = \chi$, the derivative is trivial, and we have

$$\Gamma^i_{jk}\frac{dx^j}{d\chi}\frac{dx^k}{d\chi} = -\Phi_{,i}$$

Collecting these terms leads to the geodesic equation for transverse motion,

$$\frac{d}{d\chi}\left[\frac{1}{a^2}\frac{d}{d\chi}\left(\chi\theta^i\right)\right] = \frac{2}{a^2}\left[\Phi_{,i} + aH\frac{d}{d\chi}\left[\chi\theta^i\right]\right]$$

The derivative on the left acting on a^{-2} exactly cancels the term proportional to aH on the right, so our final equation for the transverse displacement is

$$\frac{d^2}{d\chi^2}\left(\chi\theta^i\right) = 2\Phi_{.i}.$$

This geodesic equation tells us that in a uniform potential, the angular direction $\left(\chi\theta^i\right)'$ remains constant, whereas a changing potential perturbs it. The sign is correct: An overdensity centered at $x = y = 0$ has $\Phi > 0$ there, and therefore the derivative of Φ with respect to x ($\Phi_{,i}$ with $i = 1$) is negative for $x > 0$. As such, the force on a light ray passing the overdensity on the positive x-axis is negative, i.e., inward toward the overdensity, as we expect. Equation (10.11) can be integrated to find the image angle as a function of the source angle. Integrating once gives

$$\frac{d}{d\chi}\left(\chi\theta^i\right) = 2\int_0^\chi d\chi'\Phi_{,i}\left(\vec{x}\left(\chi'\right)\right) + \text{ constant}$$

we'll fix the constant momentarily. Integrating again leads to

$$\theta^i_S = \frac{2}{\chi}\int_0^\chi d\chi''\int_0^{\chi''} d\chi'\Phi_{,i}\left(\vec{x}\left(\chi'\right)\right) + \text{ constant}$$

since $\theta^i(\chi) := \theta^i_S$, the value of $\vec{\theta}$ at the source. We now see that the constant is equal to θ^i - the observed angle - since the angle retains its initial value if there is no perturbation. The integral in the χ', χ'' plane is restricted to the shaded region in Figure 10.5 so we can change orders of integration with the χ'' integral ranging

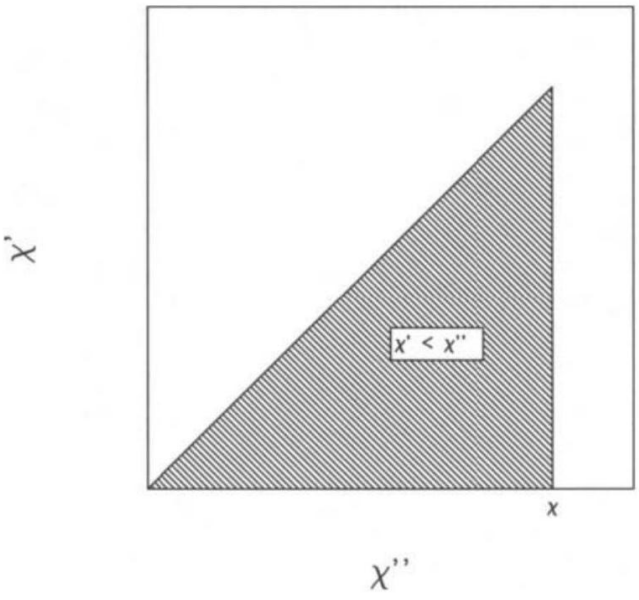


Figure 10.5. Range of integration in the double integral of Eq. (10.13). The shaded region can be expressed as $0 < \chi'' < \chi, 0 < \chi' < \chi''$ or as $\chi' < \chi'' < \chi, 0 < \chi' < \chi$. The latter is more convenient here, since the χ'' integral is then trivial.

from χ' to χ. The χ'' integral is then trivial (since $\Phi_{,i}$ depends only on χ') so

$$\theta_S^i = \theta^i + 2\int_0^\chi d\chi' \Phi_{,i}\left(\vec{x}\left(\chi'\right)\right)\left(1 - \frac{\chi'}{\chi}\right).$$

To describe the shift in the angle experienced by a light ray, it is conventional to define the 2×2 symmetric transformation matrix,

$$A_{ij} := \frac{\partial \theta_S^i}{\partial \theta^j}$$

$$:= \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}$$

The parameter κ is called the convergence; it describes how an image is magnified. Although this magnification has many important ramifications (e.g., microlensing and multiple images) it is not what is important for the distortions studied in weak lensing. Rather, these distortions are governed by the two components of the shear,

$$\gamma_1 = -\frac{A_{11} - A_{22}}{2}$$
$$\gamma_2 = -A_{12}.$$

Equation (10.15) says that the components of shear involve derivatives of Eq. (10.14) with respect to angle $\vec{\theta}$.[1] The only dependence on $\vec{\theta}$ is in the argument of the potential, where $\vec{x}\left(\chi'\right) = \chi'\vec{\theta}$ (for the transverse components). Therefore, the derivative with respect to θ^j can be written as a derivative with respect to x^j (in our notation,j) times χ'. Therefore,

$$A_{ij} - \delta_{ij} = \begin{pmatrix} -\kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & -\kappa + \gamma_1 \end{pmatrix} = 2\int_0^\chi d\chi' \Phi_{,ij}\left(\vec{x}\left(\chi'\right)\right)\chi'\left(1 - \frac{\chi'}{\chi}\right)$$

So γ_1 and γ_2 are well-defined functions of the potential. The next section shows how they influence the shapes of galaxy images.

12.11.3 Ellipticity as an Estimator of Shear

Theory

We expect lensing to turn circular images into elliptical ones. To describe this effect, then, we need to come up with quantitative measures of ellipticity, and then see how these are related to the components of shear defined above. The simplest measure of ellipticity starts with the definition of the quadrupole moments of an image. Imagine centering an image at the $\theta_x - \theta_y$ origin such that it has no dipole moment ($\langle\theta_x\rangle = \langle\theta_y\rangle = 0$ where angular brackets are averages over the intensity). Then the quadrupole moments are defined as

$$q_{ij} := \int d^2\theta I_{\rm obs}(\theta)\theta_i\theta_j$$

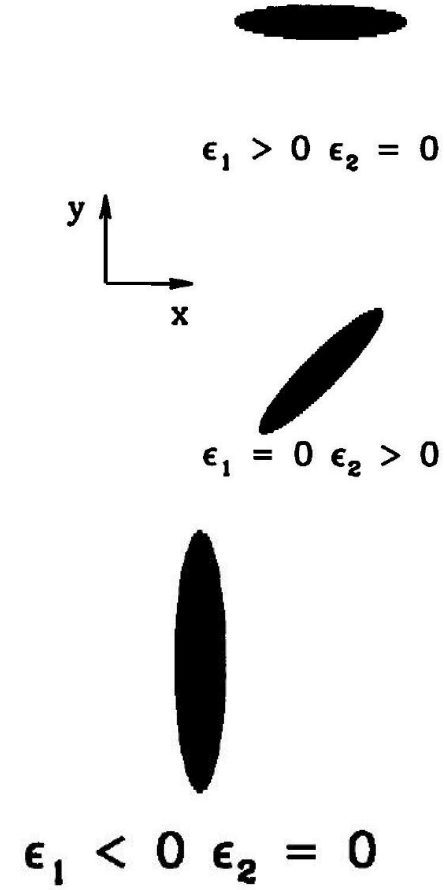
A circular image has $q_{xx} = q_{yy}$ and $q_{xy} = 0$. Therefore, two good measures of ellipticity are

$$\epsilon_1 := \frac{q_{xx} - q_{yy}}{q_{xx} + q_{yy}}$$

[1] The derivative is formally with respect to the observed angle $\vec{\theta}$, while the right-hand side of Eq. (10.14) depends on the potential at the true position of the light ray. In principle, then, the derivatives which go into the definition of A_{ij} are quite complicated. In practice, though, deflections are sufficiently small that we can ignore the distinction between the final angle $\vec{\theta}$ and the actual angle everywhere along the trajectory. Therefore, on the right-hand side of Eq. (10.14) we evaluate the potential along the undistorted path parameterized by $\vec{\theta}$.

$$\epsilon_2 := \frac{2q_{xy}}{q_{xx} + q_{yy}}$$

Figure 10.6 shows different orientations of elliptical images and the associated



$\epsilon_1 < 0 \epsilon_2 = 0$

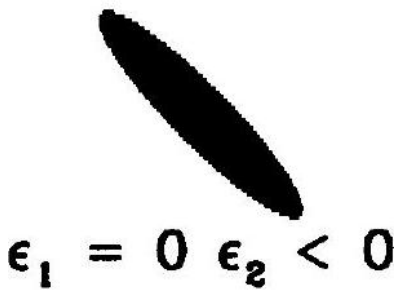


Figure 10.6. Definition of ellipticities ϵ_1 and ϵ_2. Circular images have both ellipticities equal to zero.

values of the ϵ_1 and ϵ_2. With these definitions, we can make more precise the statement at the end of Section 10.1 about correlations of ellipticities. Panel (a) in Figure 10.3 has two galaxies at $\vec{\theta}_1$ and $\vec{\theta}_2$, each with ϵ_1 positive; in panel (b) both galaxies have ϵ_1 negative. In both possible cases, then, the product $\epsilon_1\left(\vec{\theta}_1\right)\epsilon_1\left(\vec{\theta}_2\right)$ is positive if the x-axis is chosen along the direction connecting the two galaxies. The impossible case is depicted in panel (c) wherein the product is negative. Therefore, we do not expect lensing to produce $\epsilon_1\left(\vec{\theta}_1\right)\epsilon_1\left(\vec{\theta}_2\right) < 0$.

How are the ellipticities defined in Eq. (10.19) related to the shear defined in Eq. (10.16)? Let's assume that the source is spherical and compute the ellipticity of the image. Focusing on ϵ_1, we have

$$\epsilon_1 = \frac{\int d^2\theta I_{\rm true}\left(\vec{\theta}_S\right)\left[\theta_x\theta_x - \theta_y\theta_y\right]}{\int d^2\theta I_{\rm true}\left(\vec{\theta}_S\right)\left[\theta_x\theta_x + \theta_y\theta_y\right]}$$

where I have used the equality of Eq. (10.1). The integrals here are over the observed angles $\vec{\theta}$, while the integrands depend in part on the angle from which the photon started at the source, θ_S. For small angles, these are related via $\theta_i = \left(A^{-1}\right)_{ij}\theta_{Sj}$. To do the integrals, then, change dummy variables in the integral to $\vec{\theta}_S$, and write all occurences of θ_i as $\left(A^{-1}\theta_S\right)_i$. This leads to

$$\epsilon_1 = \frac{\sum\limits_{ij}\left[\left(A^{-1}\right)_{xi}\left(A^{-1}\right)_{xj} - \left(A^{-1}\right)_{yi}\left(A^{-1}\right)_{yj}\right]\int d^2\theta_S I_{\rm true}\left(\vec{\theta}_S\right)\theta_{S_i}\theta_{Sj}}{\sum\limits_{ij}\left[\left(A^{-1}\right)_{xi}\left(A^{-1}\right)_{xj} + \left(A^{-1}\right)_{yi}\left(A^{-1}\right)_{yj}\right]\int d^2\theta_S I_{\rm true}\left(\vec{\theta}_S\right)\theta_{Si}\theta_{Sj}}.$$

Now the integral over the (true) circular image vanishes unless $i = j$ and so is proportional to δ_{ij}. The proportionality constant is irrelevant since it appears in both the numerator and denominator, so

$$\begin{aligned}\epsilon_1 &= \frac{\left(A^{-1}\right)_{xi}\left(A^{-1}\right)_{xi} - \left(A^{-1}\right)_{yi}\left(A^{-1}\right)_{yi}}{\left(A^{-1}\right)_{xi}\left(A^{-1}\right)_{xi} + \left(A^{-1}\right)_{yi}\left(A^{-1}\right)_{yi}} \\ &= \frac{\left(A^{-1}_{xx}\right)^2 - \left(A^{-1}_{yy}\right)^2}{\left(A^{-1}_{xx}\right)^2 + \left(A^{-1}_{yy}\right)^2 + 2\left(A^{-1}_{xy}\right)^2}.\end{aligned}$$

It is easy to compute the inverse of the 2×2 matrix A:

$$A^{-1} = \frac{1}{(1-\kappa)^2 - \gamma_1^2 - \gamma_2^2}\begin{pmatrix} 1-\kappa+\gamma_1 & \gamma_2 \\ \gamma_2 & 1-\kappa-\gamma_1 \end{pmatrix},$$

so we see that the ellipticity ϵ_1 can be expressed in terms of the shear as

$$\begin{aligned}\epsilon_1 &= \frac{\left(1-\kappa+\gamma_1\right)^2 - \left(1-\kappa-\gamma_1\right)^2}{\left(1-\kappa+\gamma_1\right)^2 + \left(1-\kappa-\gamma_1\right)^2 + 2\gamma_2^2} \\ &= \frac{4\gamma_1(1-\kappa)}{2(1-\kappa)^2 + 2\gamma_1^2 + 2\gamma_2^2}.\end{aligned}$$

If all the distortions are small, then

$$\epsilon_1 \simeq 2\gamma_1$$

the desired result. A similar equality holds for ϵ_2. By measuring ellipticities of distant galaxies, therefore, we can get an estimate of the shear field, a field which depends manifestly on the underlying gravitational potential via Eq. (10.17).

12.11.4 Weak Lensing Power Spectrum

Theory

We can now compute the simplest statistics of the shear field, which can be estimated by measuring background galaxy ellipticities. Let's remove the identity from the transformation matrix A,

$$\psi_{ij} := A_{ij} - \delta_{ij}.$$

In the absence of inhomogeneities, the apparent angle θ is equal to the source angle θ_S, so $A = I$. Therefore, by removing the identity matrix from A, we have extracted the part describing the distortion of the light ray path due to inhomogeneities. As such, we will refer to ψ as the distortion tensor. The last term in (10.17) is ψ_{ij} for a background galaxy (or galaxies) at distance $\chi(z)$ from us. In general, a survey contains a distribution of redshifts. Let's call this distribution $W(\chi)$, just as we did when studying angular correlations in Chapter 9. Again, let's normalize W so that $\int d\chi W(\chi) = 1$. Then, the distortion tensor is

$$\psi_{ij} = 2 \int_0^{\chi_\infty} d\chi W(\chi) \int_0^{\chi} d\chi' \Phi_{,ij} \left(\vec{x}\left(\chi'\right)\right) \chi' \left(1 - \frac{\chi'}{\chi}\right).$$

We can simplify here by changing orders of integration (almost exactly as depicted in Figure 10.5). Then

$$\psi_{ij}(\vec{\theta}) = \int_0^{\chi_\infty} d\chi \Phi_{,ij}(\vec{x}(\chi)) g(\chi)$$

where I have dropped the prime and defined

$$g(\chi) := 2\chi \int_\chi^{\chi_\infty} d\chi' \left(1 - \frac{\chi}{\chi'}\right) W\left(\chi'\right)$$

On average, each of the components of the distortion tensor is zero: $\langle \psi_{ij} \rangle = 0$. To make our money, therefore, we need to do just what we did for the CMB and galaxy distributions, compute the two-point function, either the angular correlation functions of the different components of ψ_{ij} or their Fourier transforms, the power spectra. To compute these two-point functions, we will be able to essentially copy the results from Section 9.1 as long as we are careful to account for the indices on ψ_{ij}.

To compute the power spectrum of the distortion tensor, $P^{\psi}_{ijkl}(\vec{l})$, let's recall the steps we took when we analyzed the angular correlation function of galaxies (see Table 10.1).

- The distortion tensor in Eq. (10.28) is a function of the 2D vector $\vec{\theta}$ since the argument of the potential is $\vec{x} \simeq \chi\left(\theta_1, \theta_2, 1\right)$. As in the case of the galaxy density field, we can Fourier transform ψ_{ij} so that it depends on the 2D vector conjugate to $\vec{\theta}, \vec{l}$.
- In the case of the angular galaxy overdensity, we expressed the 2D overdensity as an integral over the 3D overdensity with some weighting function (Eq. (9.3)). Here the situation is identical: g in Eq. (10.29) plays the role of the selection function W there while the 3D field here is not the overdensity δ, but rather $\Phi_{,ij}$.

- Next we found that - in the small-angle limit - the 2D power spectrum is given by an integral over the 3D power spectrum, Eq. (9.10). Here, too, the 2D power spectrum of ψ_{ij} can be expressed as an integral over the 3D power spectrum of the gravitational potential Φ. The only slightly tricky part is computing the 3D power spectrum of $\Phi_{,ij}$. The Fourier transform of $\Phi_{,ij}$ is $-k_i k_j \tilde{\Phi}$ with a variance

$$k_i k_j k'_l k'_m \left\langle \tilde{\Phi}(\vec{k}) \tilde{\Phi}^* \left(\vec{k'}\right) \right\rangle = (2\pi)^3 k_i k_j k_l k_m \delta^3 \left(\vec{k} - \vec{k'}\right) P_\Phi(k).$$

So the 3D power spectrum we need - the one associated with the Fourier transform of $\Phi_{,ij}$ - is $k_i k_j k_l k_m P_\Phi(k)$. Table 10.1. Similarity between Agular Correlations of Galaxies and Weak Lensing.

	Angular galaxy distribution	Weak lensing
2D observation	$\delta_2(\vec{\theta})$	Distortion tensor $\psi_{ij}(\vec{\theta})$
Weighting function	$W(\chi)$	$g(\chi)$
3D field	δ	$\Phi_{,ij}$
3D power spectrum	$\left\langle \tilde{\delta}\tilde{\delta}^* \right\rangle \sim P(k)$	$\left\langle \tilde{\Phi}_{,ij} \dot{\Phi}^*_{,lm} \right\rangle \sim k_i k_j k_l k_m P_\Phi(k)$
2D power spectrum	$P_2(l)$	$P^\psi_{ijlm}(\vec{l})$

I will keep things in terms of P_Φ, but if you are more comfortable with the density power spectrum, you can see from Eq. (7.7) that you need only multiply P_Φ by $9\Omega_m^2 H_0^4 (1+z)^2/\left(4k^4\right)$. Applying Eq. (9.10) then leads to

$$\left\langle \tilde{\psi}_{ij}(\vec{l}) \tilde{\psi}_{lm}(\vec{l}) \right\rangle = (2\pi)^2 \delta^2(\vec{l} - \vec{l}) P^\psi_{ijlm}(\vec{l})$$

with the 2D power spectrum

$$P^\psi_{ijlm}(\vec{l}) = \int_0^{\chi\infty} d\chi \frac{g^2(\chi)}{\chi^2} \frac{l_i l_j l_l l_m}{\chi^4} P_\Phi(l/\chi)$$

Equation (10.32) is an expression for the power spectrum of the different components of the distortion tensor. We can turn these into power spectra for the convergence κ and two different components of shear by using Eq. (10.16). Let's work this out for one of the shear components explicitly; the other two are relegated to a problem. Since $\gamma_1 = \left(\psi_{22} - \psi_{11}\right)/2$, the power spectrum of γ_1 is 1/4 times $P^\psi_{2222} + P^\psi_{1111} - 2P^\psi_{2211}$. If we decompose the 2D vector $\vec{l}$ into a radial part l and an angle ϕ_l, then $l_1 = l\cos\phi_l$ and $l_2 = l\sin\phi_l$, so

$$P_{\gamma_1}(l, \phi_l) = \left(\sin^4\phi_l + \cos^4\phi_l - 2\sin^2\phi_l \cos^2\phi_l\right) \left[\frac{l^4}{4} \int_0^{\chi\infty} d\chi \frac{g^2(\chi)}{\chi^6} P_\Phi(l/\chi)\right].$$

Since $\sin^4\phi_l + \cos^4\phi_l + 2\cos^2\phi_l \sin^2\phi_l = 1$, the term in parentheses here is equal to $1 - 4\sin^2\phi_l\cos^2\phi_l = 1 - \sin^2(2\phi_l)$ or $\cos^2(2\phi_l)$. You will show in Exercise 6 that the expression in square brackets is equal to the power spectrum of the convergence,

$$P_\kappa = \frac{l^4}{4} \int_0^{\chi\infty} d\chi \frac{g^2(\chi)}{\chi^6} P_\Phi(l/\chi)$$

Therefore, the power spectrum of γ_1 is

$$P_{\gamma_1}(l, \phi_l) = \cos^2(2\phi_l)\, P_\kappa(l).$$

You will show in Exercise 6 that the power spectrum of γ_2 is also proportional to P_κ,

$$P_{\gamma_2}(l, \phi_l) = \sin^2(2\phi_l)\, P_\kappa(l).$$

Thus, the power spectra of the two components of shear depend not only on the magnitude of $\vec{l}$ but also on its direction. Figure 10.7 shows the convergence power

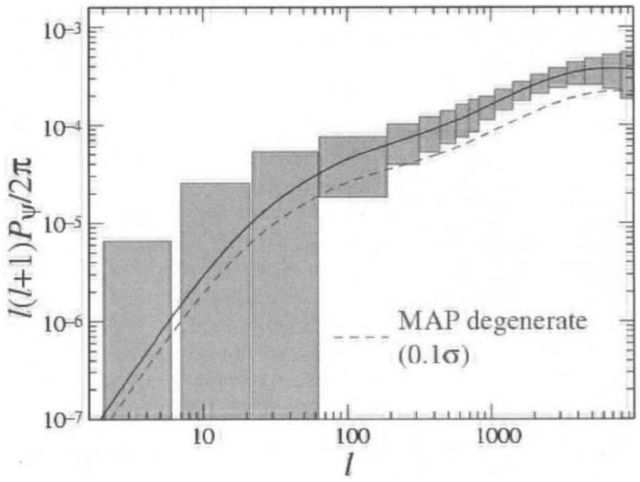


Figure 10.7. The power spectrum of the convergence for two CDM models (Hu and Tegmark, 1999). The models are indistinguishable using CMB data alone, so future weak lensing data (depicted by the error boxes) add invaluable information. The projected error boxes assume a 3° survey down to 25 th magnitude.

spectrum for two models. Note that, unlike the CMB or even the matter power spectrum, it is essentially featureless.

The power spectra of both shear components are proportional to P_κ, with a prefactor depending on the angle between $\vec{l}$ and a fixed x-axis. You might expect then that a linear combination of the two components depends only on $P_\kappa(l)$ with no dependence on the angle ϕ_l. This is correct. Even more interesting, though, is the possibility that a different linear combination would have a vanishing power spectrum. This also turns out to be correct, and extremely useful. Such a mode has no expected cosmological signal, so any nonzero value is a measure of some systematic effect. By focusing on this "zero" mode, one can identify and eliminate contaminating effects in an experiment. I want to spend some time on this decomposition into two modes - one with signal and one without - not only because of its importance in this case of weak lensing, but also because an exact analogy exists in the case of polarization of the CMB, which we will take up in the next sections. To get ahead of myself, in the case of polarization, the "zero" mode is zero only for scalar perturbations, whereas tensor perturbations do contribute to it. Therefore, we will see in Section 10.9 that this decomposition is a powerful tool with which to detect primordial gravity waves.

Consider then the following two linear combinations of the shear:

$$E(\vec{l}) := \cos(2\phi_l)\, \gamma_1(\vec{l}) + \sin(2\phi_l)\, \gamma_2(\vec{l})$$
$$B(\vec{l}) := -\sin(2\phi_l)\, \gamma_1(\vec{l}) + \cos(2\phi_l)\, \gamma_2(\vec{l}).$$

The power spectrum of each of these modes is easily obtainable in terms of the spectra of γ_1 and γ_2. First the E-mode:

$$P_E = \cos^2(2\phi_l)\, P_{\gamma_1} + \sin^2(2\phi_l)\, P_{\gamma_2} + 2\sin(2\phi_l)\cos(2\phi_l)\, P_{\gamma_1\gamma_2}.$$

This expression involves the power spectrum corresponding to $\langle\gamma_1\gamma_2\rangle$, which is equal to $\cos(2\phi_l)\sin(2\phi_l)$ times the ubiquitous convergence power spectrum. Therefore, P_E is proportional to P_κ with proportionality constant $\cos^4(2\phi_l) + \sin^4(2\phi_l) + 2\cos^2(2\phi_l)\sin^2(2\phi_l) = \left(\cos^2(2\phi_l) + \sin^2(2\phi_l)\right)^2 = 1$, or

$$P_E = P_\kappa$$

independent of the angle ϕ_l. The calculation for the B-mode is similar:

$$\begin{aligned} P_B &= \sin^2(2\phi_l)\, P_{\gamma_1} + \cos^2(2\phi_l)\, P_{\gamma_2} - 2\sin(2\phi_l)\cos(2\phi_l)\, P_{\gamma_1\gamma_2} \\ &= 0. \end{aligned}$$

You can also check that the cross power spectrum $\langle EB\rangle$ vanishes.

The field of weak lensing due to large-scale structure is its infancy. The year 2000 saw the first detections by four independent groups (Van Waerbecke et al., 2000; Bacon, Refregier, and Ellis, 2000; Wittman et al., 2000; Kaiser, Wilson, and Luppino, 2000). They presented the shear correlation function, one example of which is shown in Figure 10.10. We can easily translate the power spectra derived above into angular correlation functions that can be compared with data.

Let's focus on the angular correlation function of γ_1, the Fourier transform of P_{γ_1},

$$w_{\gamma_1}(\vec{\theta}) = \int \frac{d^2l}{(2\pi)^2} e^{i\vec{l}\cdot\vec{\theta}} \cos^2(2\phi_l) \left[\frac{l^4}{4}\int_0^{\chi_\infty} d\chi \frac{g^2(\chi)}{\chi^6} P_\Phi(l/\chi)\right]$$

The variable ϕ_l we are integrating over is the angle between the 2D vector $\vec{l}$ and an arbitrary external x-axis. If we do the angular integral over ϕ_l, then-as you can see from Figure 10.8 - the argument of the exponential is quite complicated: $il\theta\cos(\phi_l - \phi)$. Instead, let's integrate over the angle between $\vec{l}$ and $\vec{\theta}$, call it ϕ'. Then,

$$w_{\gamma_1}(\vec{\theta}) = \int_0^\infty \frac{dl}{(2\pi)^2}\frac{l^5}{4}\int_0^{\chi_\infty} d\chi \frac{g^2(\chi)}{\chi^6} P_\Phi(l/\chi) \int_0^{2\pi} d\phi' e^{il\theta\cos\phi'} \left[\cos\left(2\left(\phi' + \phi\right)\right)\right]^2.$$

The cosine squared in the integrand is equal to

$$\left[\cos(2\phi')\cos(2\phi) - \sin(2\phi')\sin(2\phi)\right]^2$$

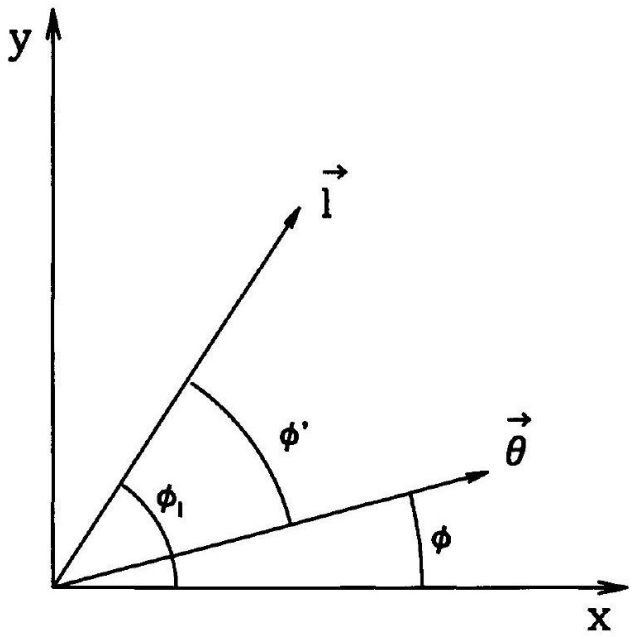


Figure 10.8. Angles made by $\vec{\theta}$ and $\vec{l}$ with respect to an external fixed $x-y$ coordinate system. The angle between $\vec{\theta}$ and $\vec{l}$ is ϕ'.

$$= \cos^2(2\phi')\cos^2(2\phi) - \sin(4\phi')\sin(4\phi)/2 + \sin^2(2\phi')\sin^2(2\phi)$$

Thus there are three terms to be integrated over. To do the integral over $\cos^2(2\phi')$, first rewrite it as $(1+\cos(4\phi'))/2$; then recall that the integral of $\cos(n\phi')\, e^{iz\cos\phi'}$ is equal to $2\pi i^n J_n(z)$ (Eq. (C.21)). Therefore, the integral of $\cos^2(2\phi')$ gives a factor of π times $J_0(l\theta) + J_4(l\theta)$. Using exactly the same arguments, you can see that the integral of $\sin^2(2\phi')$ gives π times $J_0(l\theta) - J_4(l\theta)$. Less obvious is the fact that the integral of $\sin(4\phi')$ vanishes (change integration variable to $\phi'' = \phi' - \pi$ and argue that the integrand is antisymmetric). Therefore,

$$w_{\gamma_1}(\vec{\theta}) = \frac{1}{16\pi}\int_0^\infty dl l^5 \int_0^{\chi\infty} d\chi \frac{g^2(\chi)}{\chi^6} P_\Phi(l/\chi) \times \left\{\cos^2(2\phi)\left[J_0(l\theta)+J_4(l\theta)\right] + \sin^2(2\phi)\left[J_0(l\theta)-J_4(l\theta)\right]\right\} \cdot (10$$

There are many angles floating around, so let me reiterate that $\vec{\theta} = (\theta\cos\phi, \theta\sin\phi)$; that is, ϕ is the angle that $\vec{\theta}$ makes with the x-axis. By changing the l integral into one over 3D wavenumber $k = l/\chi$, we can rewrite this angular correlation function in terms of kernels,

$$w_{\gamma_1}(\vec{\theta}) = \int_0^\infty dk k^5 P_\Phi(k)\left[F_+(k\theta)\cos^2(2\phi) + F_-(k\theta)\sin^2(2\phi)\right]$$

Note that here I have assumed that the potential remains constant with time, an assumption which breaks down at late times because of nonlinearities or non-matter domination. The kernels are integrals over radial distance χ modulated by the Bessel functions,

$$F_\pm(k\theta) := \frac{1}{16\pi}\int_0^{\chi\infty} d\chi g^2(\chi)\left[J_0(k\chi\theta) \pm J_4(k\chi\theta)\right]$$

Figure 10.9 shows these two kernels for background galaxies at redshift $z = 0.9$.

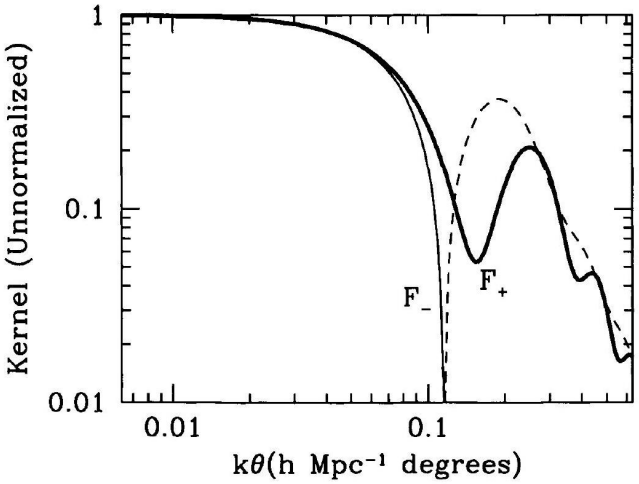


Figure 10.9. The kernels for the shear correlation function assuming all background galaxies are at $z = 0.9$. Dashed region corresponds to negative kernel. If the x-axis is chosen along the line connecting pairs of galaxies, then the + kernel is for $\langle\gamma_1\gamma_1\rangle$ and the - for $\langle\gamma_2\gamma_2\rangle$. Note that the former is always positive.

If we choose the x-axis to be along the line connecting pairs of galaxies, then we are evaluating the correlation function at $\vec{\theta} = (\theta, 0)$, that is, with $\phi = 0$. In that case, Figure 10.9 shows that w_{γ_1} is always positive, a result we anticipated pictorially in Section 10.1. The correlation function for γ_2, on the other hand, is identical to that in Eq. (10.45) except that $F_{\pm}$are interchanged. Therefore, w_{γ_2} can, and indeed does, go negative, usually on large angular scales. The final point to take away from the kernels in Figures 10.9 is the rough sense that the shear on a scale of a tenth of degree probes the power spectrum at $k \sim 1h\text{Mpc}^{-1}$ since this is where the kernel breaks.

Consider then Figure 10.10, which shows results from a survey of three "blank" (i.e., no known clusters of galaxies present) fields over a period of several years. There is a clear detection of ellipticity, presumably due to cosmic shear. The root mean square amplitude of the shear is the square root of the typical numbers on the y-axis, around 0.01. Thus, shear due to large-scale structure has been detected

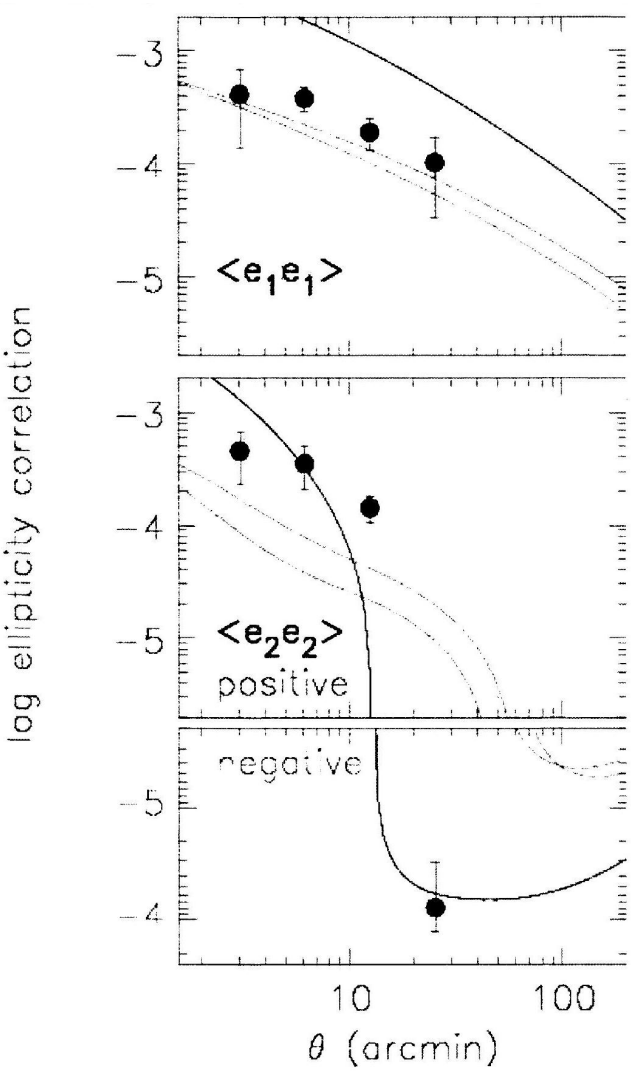


Figure 10.10. Measurement of the shear correlation functions using 145, 000 background galaxies (Wittman et al., 2000). Also shown are a variety of CDM models; topmost in top panel is standard CDM, ruled out here at many sigma. Note that $w_{\gamma_1} = \langle e_1 e_1 \rangle$ remains positive on all angular scales. See color Plate 10.10. with an amplitudeof order 10^{-2} on angular scales ranging from $1'$ out to about $1°$. Observations planned in the coming years will go far beyond these preliminary results; from these observations we will learn much about the mass distribution of the universe.

12.11.5 Polarization: The Quadrupole and the Q/U Decomposition

Theory

The radiation in the CMB is expected to be polarized because of Compton scattering at the time of decoupling. A polarization pattern shares a number of mathematical features with the shear induced by gravitational lensing that we have just studied. In addition to these mathematical similarities, they also share similar experimental histories. Whereas anisotropies in the temperature of the CMB and inhomogeneities in the density field were discovered back in the 20th century, weak lensing by large scale structure and polarization of the CMB have just been detected. They are true 21st century phenomena. Therefore, they are both fields rich for study: We are just beginning our observations of them, and they both promise to deliver much cosmological information.

Light traveling in the x-direction corresponds to electric and magnetic fields oscillating in

the y - z plane, i.e., transverse to the direction of propagation. If the intensity along the two transverse directions is equal, then the light is unpolarized. Until now, when we have considered the CMB, we have been implicitly studying this case. Now we must account for the possibility that the intensities in the two transverse directions are unequal: that the radiation is polarized.

At first glance, Compton scattering is a perfect mechanism for producing polarized radiation. It allows all transverse radiation to pass through unimpeded, while completely stopping any radiation parallel to the outgoing direction. To see this, consider Figure 10.11 which shows a ray incident from the $+\hat{x}$ direction. This (unpolarized) ray has equal intensity in the $\hat{y}$ and $\hat{z}$ directions. It scatters off an electron at the origin and gets deflected into the $+\hat{z}$ direction. [2] Since the outgoing direction is along the z-axis, none of the (incoming) intensity along the z-axis gets transmitted. By contrast, all of the intensity along the y-axis (which is perpendicular to both the incoming and outgoing directions) is transmitted. The net result is outgoing polarization in the $\hat{y}$ direction.

Obviously, we cannot content ourselves with studying one incoming ray; we must generalize to radiation incident on an electron from all directions. When we do so, we begin to realize that producing polarization will not be quite as easy as it appears from Figure 10.11. Consider first Figure 10.12, which shows a caricature of a much more relevant case: isotropic radiation incident on the electron from all directions. I say "caricature" because I have shown incoming rays from only two directions, the $+\hat{x}$ - and $+\hat{y}$-directions. The intensity of the outgoing ray along the

[2] Of course radiation gets scattered into all directions with varying probability. Here we consider just one outgoing direction for simplicity. In the next section we account for this probability.

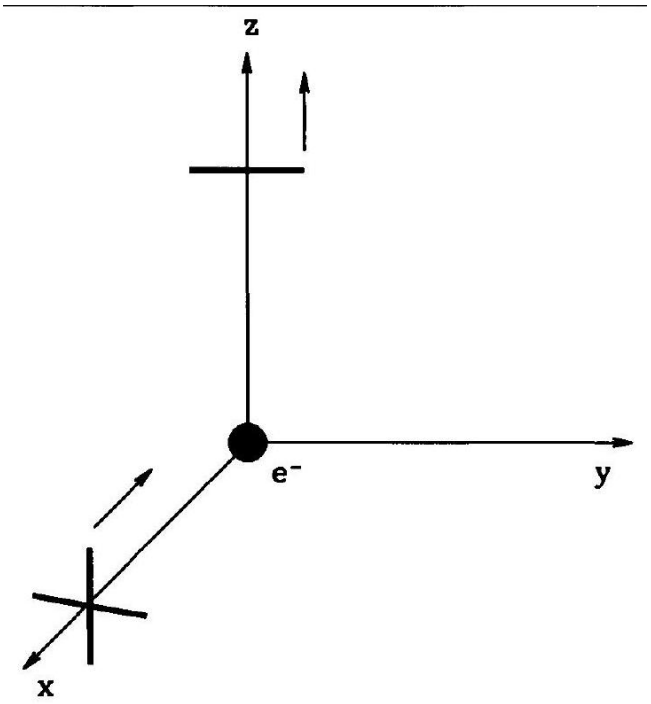


Figure 10.11. Unpolarized radiation moving toward the origin along the x-axis is scattered by an electron into the $+\hat{z}$ direction. Only the $\hat{y}$ component of the radiation remains after scattering. Since there was no incoming $\hat{x}$ polarization, the outgoing radiation is polarized in the $\hat{y}$ direction. (This and the next three figures are adapted from Hu and White, 1997b).

x-axis comes from the radiation incident from the $\hat{y}$ direction, while the outgoing y-intensity comes from the radiation incident from the $\hat{x}$-axis. Since the radiation from both directions has equal intensity (isotropic radiation), though, the outgoing wave is has equal intensity along the $\hat{x}$ - and $\hat{y}$-axes: it is unpolarized.

Can anisotropic radiation produce polarization? The simplest example of anisotropy is a dipole pattern, a caricature of which is shown in Figure 10.13. Now the outgoing intensity along the x-axis comes from the $\pm\hat{y}$-incident radiation, which has the average temperature. The outgoing intensity along the y-axis is also neither hot nor cold because it comes from a cold spot (the $-\hat{x}$-direction) and a hot spot (the $+\hat{x}$-direction). The dipole pattern leads therefore only to cancellations and unpolarized outgoing radiation.

To produce polarized radiation, the incoming radiation must have a nonzero quadrupole. Figure 10.14 illustrates the polarization produced by an incoming quadrupole. The hotter (colder) radiation incident from the $\hat{x}$ - ($\hat{y}-$) direction produces higher (lower) intensity along the y - ($x-$) axis for the outgoing wave. Therefore, the intensity of the outgoing wave is greater along the y-axis than along the x-axis: the outgoing radiation is polarized.

The fact that Compton scattering produces polarization only when the incident field has a quadrupole moment has important ramifications for cosmology. We need Compton scattering to produce the polarization, so we need to focus on the epoch

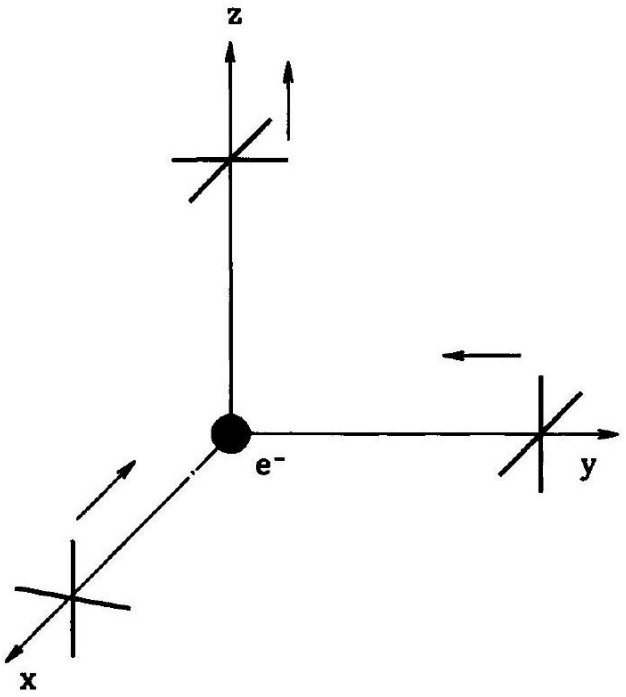


Figure 10.12. Incoming isotropic radiation produces no polarization. Here, since the incoming amplitudes from the $\hat{x}$ - and $\hat{y}$-directions are equal, the outgoing intensities along both of these directions are equal, leading to unpolarized radiation.

before electrons and photons have completely decoupled from each other. However, in this epoch electrons and photons are tightly coupled, which we have seen leads to a very small quadrupole. Therefore, we expect polarization from the standard decoupling epoch to be smaller than the anisotropies. Late reionization enhances the polarization at large scales, but does not modify the qualitative conclusion that the polarization signal is expected to be small.

Figure 10.14 depicts polarization in the $x-y$ plane, preferentially in the $\hat{y}$ direction. Alternatively, had the incoming rays been rotated by 45° in the $x-y$ plane, the outgoing polarization would have been along the axis 45° from the x and y-axes. Polarization therefore can be depicted as a headless vector, with a length corresponding to its magnitude and the orientation of the line describing the axis along which the intensity is greatest. In the 2D plane perpendicular to the direction of propagation, we therefore decompose the intensity into

$$I_{ij} = \begin{pmatrix} T+Q & U \\ U & T-Q \end{pmatrix}.$$

The diagonal elements T are the temperature we studied in Chapter 8 (with a uniform part and a perturbation Θ); the two new variables Q and U describe polarization. The pattern in Figure 10.14 has $Q < 0$ and $U = 0$. Note that these definitions of Q and U are identical to the definitions of shear and ellipticity in Section 10.2. Especially relevant is Figure 10.6 where we simply replace ϵ_1 with Q

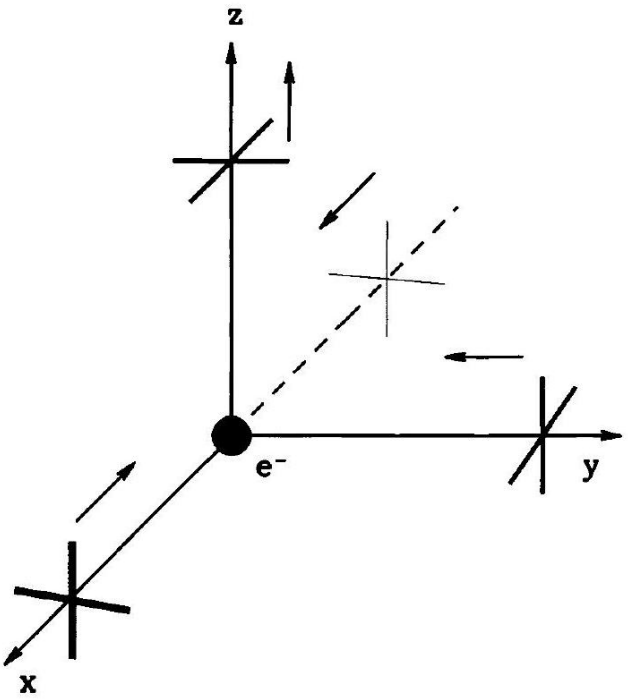


Figure 10.13. Incoming dipole radiation also produces no polarization. (See also color Plate 10.13.) Heavy (thin) lines denote hot (cold) spots. Here the incoming radiation is hotter than average (average is medium thickness) from the $+\hat{x}$-direction, and colder than average from the $-\hat{x}$-direction. The two rays from the $\pm\hat{x}$-directions therefore produce the average intensity for the outgoing ray along the $\hat{y}$-direction. The outgoing intensity along the $\hat{x}$-direction is produced by the ray incident from the $\pm\hat{y}$-directions. Since these have the average intensity, the outgoing intensity is also the average along the $\hat{x}$-direction. The net result is outgoing unpolarized light.

and ϵ_2 with U. A final note: students of electricity and magnetism will no doubt recognize T, Q, and U as three of the four Stokes parameters used to describe polarization. The fourth, V, is nonzero only if polarization is circularly polarized, a phenomenon we do not expect in the early universe, so I have implicitly set $V = 0$ here.

12.11.6 Polarization from a Single Plane Wave

Theory

The pictures of the previous subsection are important to gain a qualitative understanding of how Compton scattering produces polarization, but they are inefficient tools with which to study the phenomenon quantitatively. The proper tool is the Boltzmann equation. We could proceed now by simply writing down the Boltzmann equation for the Q and U polarization states. In doing so, however, we would lose some of the intuition just gained, so I will take an intermediate tack. We will generalize the discussion of the previous section by summing o: er all incident rays, not just a handful. This will enable us to make the connection with the distribution Θ

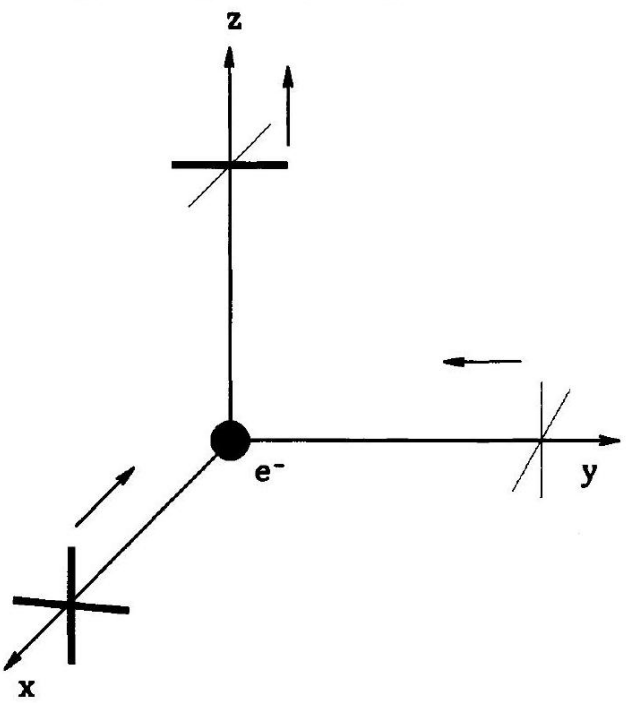


Figure 10.14. Incoming quadrupole radiation produces outgoing polarized light. See also color Plate 10.14.) The outgoing radiation has greater intensity along the y-axis than in the $\hat{x}$-direction. This is a direct result of the hotter radiation incident from the $\hat{x}$-direction.

we have used until now to characterize the photons.

We first need to define the polarization axes in the most general case when the incoming photon arrives from direction $\hat{n}'$. When that direction was $\hat{x}$, as in the previous section, it was clear that polarization was defined as the difference in the intensity along the two perpendicular directions, $\hat{y}$ and $\hat{z}$. In the general case, depicted in Figure 10.15, the direction of the incoming photon is depicted by $\hat{n}'$, and we must integrate over all incoming directions. The two axes perpendicular to this direction are most conveniently taken to be $\hat{\theta}'$ and $\hat{\phi}'$, the standard unit vectors perpendicular to the position vector. These are called $\hat{\epsilon}'_1$ and $\hat{\epsilon}'_2$. We still are interested in the polarization of outgoing photons in the $\hat{z}$-direction, so we can choose the two outgoing polarization axes as $\hat{\epsilon}_1 = \hat{x}$ and $\hat{\epsilon}_2 = \hat{y}$. In short, the incoming polarization vectors are $\hat{\epsilon}'_i$, the outgoing are $\hat{\epsilon}_i$.

The idea that Compton scattering allows the fields transverse to the outgoing direction to pass through unimpeded, while stopping those parallel to the outgoing direction, can be encapsulated by saying that the cross-section for outgoing photons polarized in the $\hat{\epsilon}_i$ direction is proportional to

$$\sum_{j=1}^{2} \left| \hat{\epsilon}_i(\hat{n}) \cdot \hat{\epsilon}'_j\left(\hat{n}'\right) \right|^2 .$$

The Q polarization is the difference between this cross-section for $i = 1$ and $i = 2$, i.e., the difference between the field strength in $\hat{x}$ - and $\hat{y}$-directions:

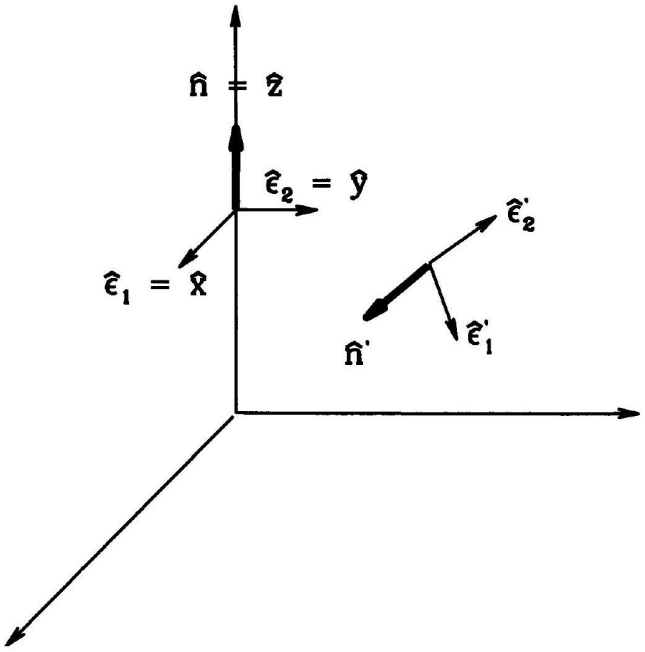


Figure 10.15. Incoming photon from direction $\hat{n}'$ Compton scatters off an electron at the origin producing outgoing photon in direction $\hat{n} = \hat{z}$. The plane perpendicular to the incoming direction is spanned by the two polarization vectors, $\hat{\epsilon}'_1 = \hat{\theta}'$ and $\hat{\epsilon}'_2 = \hat{\phi}'$. The outgoing photon is in the $\hat{z}$ direction, so the polarization vectors are $\hat{\epsilon}_1 = \hat{x}$ and $\hat{\epsilon}_2 = \hat{y}$.

$$\sum_{j=1}^{2} \left|\hat{\epsilon}_1(\hat{n}) \cdot \hat{\epsilon}'_j\left(\hat{n}'\right)\right|^2 - \sum_{j=1}^{2} \left|\hat{\epsilon}_2(\hat{n}) \cdot \hat{\epsilon}'_j\left(\hat{n}'\right)\right|^2 = \sum_{j=1}^{2} \left(\left|\hat{x} \cdot \hat{\epsilon}'_j\left(\hat{n}'\right)\right|^2 - \left|\hat{y} \cdot \hat{\epsilon}'_j\left(\hat{n}'\right)\right|^2\right)$$

Integrating over all incoming $\hat{n}'$ directions leads to

$$Q(\hat{z}) = A \int d\Omega_{n'} f\left(\hat{n}'\right) \sum_{j=1}^{2} \left(\left|\hat{x} \cdot \hat{\epsilon}'_j\left(\hat{n}'\right)\right|^2 - \left|\hat{y} \cdot \hat{\epsilon}'_j\left(\hat{n}'\right)\right|^2\right)$$

Here A is a normalization constant which will not concern us for now, and $f\left(\hat{n}'\right)$ is the intensity of the radiation incoming from the $\hat{n}'$-direction, and we integrate over all such directions. Note that f depends only on $\hat{n}'$, but not on ϵ'_j: this corresponds to the assumption that the incident radiation is unpolarized.

To take the dot products in Eq. (10.50), we will find it useful to express $\hat{\epsilon}'_1$ and $\hat{\epsilon}'_2$ in terms of their Cartesian coordinates. Since they are equal to $\hat{\theta}'$ and $\hat{\phi}'$, respectively, we have

$$\begin{aligned}
\hat{\epsilon}'_1\left(\theta', \phi'\right) &= \left(\cos\theta' \cos\phi', \cos\theta' \sin\phi', -\sin\theta'\right) \\
\hat{\epsilon}'_2\left(\theta', \phi'\right) &= \left(-\sin\phi', \cos\phi', 0\right)
\end{aligned}$$

Now, the dot products become trivial, and we find

$$\begin{aligned}
Q(\hat{z}) &= A \int d\Omega_{n'} f\left(\hat{n}'\right) \left[\cos^2\theta' \cos^2\phi' + \sin^2\phi' - \cos^2\theta' \sin^2\phi' - \cos^2\phi'\right] \\
&= -A \int d\Omega_{n'} f\left(\hat{n}'\right) \sin^2\theta' \cos 2\phi'.
\end{aligned}$$

You might recognize the combination of angles here as being propotional to the sum of the spherical harmonics $Y_{2,2} + Y_{2,-2}$ (Eq. (C.10)). Since the spherical harmonics are orthogonal, the integral will pick out the $l = 2, m = \pm 2$ components of the distribution f. That is, nonzero Q will be produced only if the incident radiation has a quadrupole moment. This verifies

the argument-by-pictures given in the previous subsection. It is straightforward to derive the corresponding expression for the U component of polarization (Exercise 10),

$$U(\hat{z}) = -A \int d\Omega_{n'} f(\hat{n}') \sin^2\theta' \sin(2\phi')$$

The combination of sines here is proportional to $Y_{2,2} - Y_{2,-2}$. Again, only an incident quadrupole produces U polarization.

We can now relate the outgoing Q and U to the moments of the incident unpolarized distribution. We'll do this in four steps, in increasing generality.

- First, we'll consider the polariztion induced by a wavevector $\vec{k}$ in the $\hat{x}$-direction. - Next, we allow $\vec{k}$ to lie anywhere in the $\hat{x} - \hat{z}$ plane.
- Then, we consider the most general possible wavevector.
- The first three steps will give us Q and U of the outgoing radiation along the z-axis. We need to generalize this to arbitrary outgoing directions.

The reason that we need to move so slowly is that the photon distribution, $f(\hat{n}')$, takes its cue from the direction of the wavevector. Recall that, in Chapter 4, we wrote the photon distribution as the sum of a zero-order piece - the uniform Planck distribution - and a perturbation, characterized by $\Theta(k,\mu)$ (e.g., Eq. (4.35)). There μ was the dot product of the wave vector $\hat{k}$ and the direction of propagation. Here we have labeled the direction of propagation of the incident photon as $\hat{n}'$, so $\mu = \hat{k}\cdot\hat{n}'$. Thus, $f(\hat{n}')$ in Eq. (10.52) will be an expansion in Legendre polynomials with argument $\hat{k}\cdot\hat{n}'$. This argument is not equal to the cosine of θ', since θ' is the angle between the external z-axis and $\hat{n}'$. Relating μ to θ' and ϕ' therefore is not trivial, and we will proceed slowly. Let's first consider the wave vector $\vec{k}$ to lie in the $\hat{x}$-direction. Then,

$$\begin{aligned}\mu := \hat{k}\cdot\hat{n}' &= \left(\hat{n}'\right)_x \\ &= \sin\theta'\cos\phi'.\end{aligned}$$

Recall that we decomposed the perturbation Θ into a sum over Legendre polynomials, so

$$\begin{aligned}\Theta\left(k, \hat{k}\cdot\hat{n}'\right) &= \sum_l (-i)^l (2l+1)\Theta_l(k)\mathcal{P}_l\left(\hat{k}\cdot\hat{n}'\right) \\ &\rightarrow -5\Theta_2(k)\mathcal{P}_2\left(\sin\theta'\cos\phi'\right),\end{aligned}$$

where the last line follows by substituting our expression for μ (Eq. (10.54)) and considering only the relevant quadrupole part of the sum.

A plane wave with wavevector $\vec{k}$ pointing in the $\hat{x}$-direction therefore has

$$Q(\hat{z}, \vec{k}\|\hat{x}) = 5A\Theta_2(k)\int_0^\pi d\theta' \sin\theta' \int_0^{2\pi} d\phi' \mathcal{P}_2\left(\sin\theta'\cos\phi'\right)\sin^2\theta'\cos 2\phi'$$

Recall that $\mathcal{P}_2(\mu) = (3\mu^2 - 1)/2$. The $-1/2$ part of this gives no contribution to the integral since the ϕ' integral over $\cos(2\phi')$ vanishes. Therefore, we are left with

$$Q(\hat{z}, \vec{k}\|\hat{x}) = \frac{15A\Theta_2(k)}{2}\int_0^\pi d\theta' \sin^5\theta' \int_0^{2\pi} d\phi' \cos^2\phi' \cos 2\phi'$$

The ϕ' integral is $\pi/2$, while the θ' integral — easily done by defining $\mu' = \cos\theta'$??? is 16/15. So

$$Q(\hat{z}, \vec{k} \| \hat{x}) = 4\pi A \Theta_2(k)$$

We've now made part of the connection between polarization - represented by Q here - and the formalism of anisotropies - described by Θ in general and Θ_2 specifically for the quadrupole. This expression though applies only in the very simple case that the wavevector points along the x-axis, perpendicular to the line of sight.

Let's generalize this expression to wavevectors pointing in an arbitrary direction in the $\hat{x}$ - $\hat{z}$ plane: $\hat{k} = (\sin\theta_k, 0, \cos\theta_k)$. In this case, the factor of $\left(\hat{k} \cdot \hat{n}'\right)^2$ coming from $\mathcal{P}_2$ is $\sin^2\theta_k \sin^2\theta' \cos^2\phi' + \cos^2\theta_k \cos^2\theta'$. The first term is identical to the $\hat{k} \| \hat{x}$ case just derived, multiplied by $\sin^2\theta_k$. The second term introduces no new ϕ' dependence; since the integral over $\cos(2\phi')$ vanishes, it does not contribute. Therefore

$$Q(\hat{z}, \vec{k} \perp \hat{y}) = 4\pi A \sin^2\theta_k \Theta_2(k)$$

In Exercise 10 you will show that there is no U-polarization from this type ($\vec{k}$ in the $\hat{x} - \hat{z}$ plane): the polarization is all Q.

For any single plane wave, we can always rotate our coordinate system around the z-axis to ensure that the plane wave lies in the $\hat{x}$ - $\bar{z}$ plane, so that Eq. (10.59) applies. When we come to consider the real universe, however, with its superposition of many plane-wave perturbations, we won't have this luxury. Instead, we will need to account for the most general wavevector with orientation $\hat{k} = (\sin\theta_k \cos\phi_k, \sin\theta_k \sin\phi_k, \cos\theta_k)$. For this more general perturbation, you can go through (Exercise 9) exactly the same types of calculations as went into Eq. (10.59) to show that

$$Q(\vec{z}, \vec{k}) = 4\pi A \sin^2\theta_k \cos(2\phi_k) \Theta_2(k)$$
$$U(\vec{z}, \vec{k}) = 4\pi A \sin^2\theta_k \sin(2\phi_k) \Theta_2(k).$$

We must make one final generalization. Until now, we have focused solely on the outgoing radiation along the z-axis. Of course, not all outgoing rays will be along the z-axis. (This is what we are looking for: difference in polarization as a function of angle.) To account for arbitrary directions, we need to allow the polar angle θ_k in Eq. (10.60) to be the angle between the observation direction $\hat{n}$ and $\hat{k}$. So $\cos\theta_k \to \hat{n} \cdot \hat{k}$, and of course $\sin^2\theta_k$ becomes $1 - (\hat{n} \cdot \hat{k})^2$. Therefore, for observations near the z-axis, the outgoing polarization induced by incoming unpolarized incident radiation is

$$Q(\hat{n}, \vec{k}) = 4\pi A \left[1 - (\hat{n} \cdot \hat{k})^2\right] \cos(2\phi_k) \Theta_2(k)$$
$$U(\hat{n}, \vec{k}) = 4\pi A \left[1 - (\hat{n} \cdot \hat{k})^2\right] \sin(2\phi_k) \Theta_2(k).$$

These expressions are valid only for directions $\hat{n}$ near the z-axis. This restriction is due to the dependence on the azimuthal angle, ϕ_k. Far from the z-axis, $\cos(2\phi_k)$ and $\sin(2\phi_k)$ give way to much more complicated expressions depending on both $\hat{n}$ and $\hat{k}$. Near the z-axis, though, the relatively simple sine and cosine describe the dependence on azimuthal angle. Thus, we will work in the small angle limit, where all observation directions are close to one another, clustered around the z-axis.

Equations (10.61) allow us to draw polarization patterns around the z-axis for arbitrary $\hat{k}$ modes. Consider the four patterns in Figure 10.16. In each case, the z-axis is out of the page in the center of the panel. For $\vec{k}$ in the $\hat{x} - \hat{y}$ plane ($\theta_k = 90°$), Eq. (10.61) says that the strength of the polarization as a function of $\hat{n}_x$ and $\hat{n}_y$ varies as

$$1 - (\hat{n} \cdot \hat{k})^2 = 1 - \left(\hat{n}_x \hat{k}_x + \hat{n}_y \hat{k}_y\right)^2$$

That is, deviations from the maximum at $\hat{n}_x = \hat{n}_y = 0$ are small, quadratic in $\hat{n}_x, \hat{n}_y$. The orientation of the polarization in these cases can be either Q (top left panel, $\phi_k = 0$) or U (bottom left panel, $\phi_k = 45°$). For $\vec{k}$ out of the $\hat{x} - \hat{y}$ plane, we begin to observe changes in the polarization strength. The two right panels in Figure 10.16 illustrate these changes. Again there can be either Q or U polarization. The most important feature of these patterns is that the polarization strength is always changing in the direction parallel or perpendicular to the sense of the

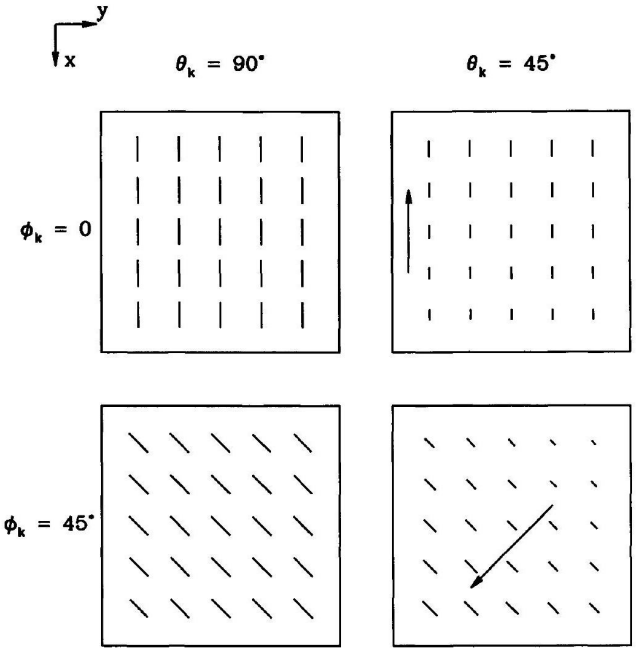


Figure 10.16. Polarization patterns near the z-axis arising from four plane wave perturbations with different $\vec{k}$. E.g., upper left arises from $\vec{k}$ along the x-axis ($\theta_k = 90°, \phi_k = 0$). For $\hat{k}$ in the $\hat{x} - \hat{y}$ plane (two left panels with $\theta_k = 90°$), polarization is at a maximum at $\hat{z}$, so little variation is seen. Arrows in right panel show direction in which polarization strength increases. This direction (or the direction perpendicular to it along which polarization remains constant) is aligned with the polarization pattern. This alignment is the hallmark of an E-mode.

polarization. In the top right panel, polarization is aligned with the x-axis, and this is the direction in which the polarization strength is changing. In the bottom right panel, polarization is aligned along $\hat{x} + \hat{y}$, and the change is along the perpendicular direction $\hat{x} - \hat{y}$. We will soon decompose polarization into E and B modes, just as we did the shear pattern in weak lensing. The patterns observed here are all pure E. Indeed, scalar perturbations generate only E modes.

We can also begin to understand the E/B decomposition. The polarization generated by scalar perturbations, the E mode, varies in strength in the same direction as its orientation. This conjures images of an electric field. An electric field from a point source, $\vec{E} = q\hat{r}/r^2$, varies in strength as one moves away from the point source. The electric field is pointed in the same direction: radially away from the source. As one moves in the direction of the field, the strength of the field decreases. In Section 10.9, we will encounter the B mode, and see that - just like a magnetic field - the B mode of polarization varies in strength in a different direction from that in which it is pointing.

12.11.7 Boltzmann Solution

Theory

To make quantitative predictions for the polarization expected in the CMB, we must go beyond the treatment of the previous section. There, we sat a single electron at the origin and

considered the polarization emerging from incoming radiation with a given distribution. The real problem has lots of electrons coupled to an evolving photon distribution. For this, we need the Boltzmann equation. We wrote down the relevant equations in Chapter 4, although it will take a little bit of work to relate the variable we used there, Θ_P, to Q and U introduced above. The relevant equations ((4.101) and (4.102)) from Chapter 4 are

$$\dot{\Theta}_P + ik\mu\Theta_P = -\dot{\tau}\left[-\Theta_P + \frac{1}{2}\left(1 - \mathcal{P}_2(\mu)\right)\Pi\right]$$
$$\Pi = \Theta_2 + \Theta_{P2} + \Theta_{P0}$$

where $\mu := \hat{k}\cdot\hat{n}$, and Θ_{P0} and Θ_{P2} are the monopole and quadrupole, respectively, of the polarization field.

We are left with the question of the relationship between Θ_P and $Q, U.\Theta_P$ is the strength of the polarization, while Q and U together describe both the strength and the orientation. In Chapter 4, we implicitly chose $\hat{k}$ to to lie in the $\hat{x}$ - $\hat{z}$ plane in Chapter 4. In that case, we have just seen that the polarization is all Q, so

$$\begin{aligned} Q(\hat{z},\vec{k}) &= \Theta_P(\hat{z},\vec{k}) \\ U(\hat{z},\vec{k}) &= 0 \qquad\qquad \hat{k}\perp\hat{y}. \end{aligned}$$

More generally, for arbitrary $\vec{k}$, at least for directions $\hat{n}$ close to $\hat{z}$, we have

$$\begin{aligned} Q(\vec{k},\hat{n}) &= \Theta_P(\hat{k}\cdot\hat{n})\cos\left(2\phi_k\right) \\ U(\vec{k},\hat{n}) &= \Theta_P(\hat{k}\cdot\hat{n})\sin\left(2\phi_k\right). \end{aligned}$$

Equation (10.66) is a crucial connection between the polarization pattern $Q(\hat{n}), U(\hat{n})$ we are interested in and $\Theta_P(\mu := \hat{k}\cdot\hat{n})$ for which we have Boltzmann equations. Now all we need to do is solve the Boltzmann equations for Θ_P, and then use Eq. (10.66) to construct power spectra for Q and U. We attack the first task in this section and the second in the next.

First, though, to solidify this connection between Θ_P and Q, U, it is instructive to rederive the result of the previous section for an incoming unpolarized wave using the Boltzmann equation. We found there (Eq. (10.59)) that the outgoing polarization (for $\hat{k}$ in the $\hat{x}-\hat{z}$ plane) was proportional to $(1-\mu^2)\,\Theta_2$ where μ is the cosine of the angle between $\hat{k}$ and $\hat{z}$. Can we get this from the Boltzmann equation? In the absence of any prior polarization, Eq. (10.63) reduces to

$$\begin{aligned} \dot{\Theta}_P &= -\dot{\tau}\frac{1-\mathcal{P}_2(\mu)}{2}\Pi \\ &= \frac{-3\dot{\tau}}{2}\left(1-\mu^2\right)\Theta_2. \end{aligned}$$

Integrating over η, we find that the polarization induced by Compton scattering from incident unpolarized radiation is

$$\Theta_P = \frac{3\tau}{2}\left(1-\mu^2\right)\Theta_2$$

i.e. the optical depth times the quadrupole modulated by the geometric factor $1-\mu^2$, in agreement with the less formal derivation advanced above. We also see that the strength of the polarization generated is proportional to the optical depth, τ, the integral along the line of sight of the free electron density times the Thomson cross-section.

Now let's solve the Boltzmann equation for the polarization. In analogy to Eq. (8.46), the formal solution to Eq. (10.63) for Θ_P is

$$\Theta_P(\hat{n},\vec{k}) = \int_0^{\eta_0} d\eta e^{i\vec{k}\cdot\hat{n}(\eta-\eta_0)-\tau(\eta)} S_P(k,\mu,\eta),$$

where the source term is

$$S_P(k,\mu,\eta) = -\frac{3}{4}\dot{\tau}\left(1-\mu^2\right)\Pi.$$

Remember that the visibility function is defined as $-\dot{\tau}e^{-\tau}$, so

$$\Theta_P(\hat{n},\vec{k}) = \frac{3}{4}\left(1-\mu^2\right)\int_0^{\eta_0} d\eta e^{i\vec{k}\cdot\hat{n}(\eta-\eta_0)} g(\eta)\Pi(k,\eta).$$

A reasonable approximation is to assume that we can evaluate the integrandexcept for the rapidly changing visibility function - at the time of decoupling (for standard recombination). Then, since the visibility function integrates to unity,

$$\Theta_P(\hat{n},\vec{k}) \simeq \frac{3\Pi\left(k,\eta_*\right)}{4}\left(1-\mu^2\right)e^{i\vec{k}\cdot\hat{n}(\eta_*-\eta_0)}.$$

Neglecting η_* compared with η_0 and rewriting the factors of μ as derivatives leads to

$$\Theta_P(k,\mu) \simeq \frac{3\Pi\left(k,\eta_*\right)}{4}\left(1+\frac{\partial^2}{\partial\left(k\eta_0\right)^2}\right)e^{-ik\eta_0\mu}.$$

To get the moments Θ_{Pl}, we must multiply Eq. (10.73) by $\mathcal{P}_l(\mu)$ and integrate over all μ as in Eq. (4.99). This gives (Eq. (C.15))

$$\Theta_{Pl}(k) \simeq \frac{3\Pi\left(k,\eta_*\right)}{4}\left(1+\frac{\partial^2}{\partial\left(k\eta_0\right)^2}\right)j_l\left(k\eta_0\right).$$

The sum of the spherical Bessel function and its second derivative can be rewritten using the spherical Bessel equation (C.13) as

$$j_l + j_l'' = -\frac{2j_{l-1}}{k\eta_0} + \frac{2(l+1)}{\left(k\eta_0\right)^2}j_l + \frac{l(l+1)}{\left(k\eta_0\right)^2}j_l.$$

Of the three terms on the right, the last one dominates on small scales. To see this, remember that the spherical Bessel function peaks roughly at $k\eta_0 \sim l$. Physically, this means that anisotropy on an angular scale l is determined by perturbations with wavelength $k^{-1} \sim \eta_0/l$. For our order-of-magnitude estimate, this means that we can take $k\eta_0$ to be of order l in the three terms on the right-hand side. The first is then of order l^{-1}, the second of order l^{-1}, and the last of order $l^2/\left(k\eta_0\right)^2 \sim 1$: the last term dominates. Therefore,

$$\Theta_{Pl}(k) \simeq \frac{3\Pi\left(k,\eta_*\right)}{4}\frac{l^2}{\left(k\eta_0\right)^2}j_l\left(k\eta_0\right).$$

In the tight coupling limit, we can express Π in terms of the quadrupole, which in turn is related to the dipole. As you can show in Exercise 12, $\Pi = 5\Theta_2/2$. Therefore, the polarization moments today are

$$\Theta_{Pl}(k) \simeq \frac{15\Theta_2\left(k,\eta_*\right)}{8}\frac{l^2}{\left(k\eta_0\right)^2}j_l\left(k\eta_0\right).$$

We can go one step further by noting that - in the tightly coupled limit - the quadrupole is proportional to the dipole (Eq. (8.34)). Therefore,

$$\Theta_{Pl}(k) \simeq -\frac{5k\Theta_1(k,\eta_*)}{6\dot{\tau}(\eta_*)}\frac{l^2}{(k\eta_0)^2}j_l(k\eta_0).$$

Equation (10.78) is a final expression for the polarization moments today assuming the tightly coupled limit. Three features are worthy of note. First, and most important, the polarization spectrum is seen to be smaller than the anisotropy spectrum by a factor of order k/τ at the time of decoupling. We will quantify this in the next section, but we now understand that it is a direct result of the twin facts that polarization is generated by a quadrupole moment and the quadrupole is suppressed in the early universe due to Compton scattering. Second, we expect there to be oscillations in the polarization power spectrum because $\Theta_{Pl} \propto \Theta_1$, which undergoes acoustic oscillations. More quantitatively, we expect the polarization oscillations, just like the dipole, to be out of phase with the monopole. The peaks and troughs in the temperature anisotropy spectrum, arising primarily from oscillations in the monopole, should then be out of phase with the peaks and troughs in the polarization power spectrum. Finally, there is no analogue here to the integrated Sachs-Wolfe effect which impacts the temperature anisotropy spectrum. Polarization cannot be induced by photons moving through changing gravitational potentials. Therefore, the polarization spectrum today is in some senses a more pristine view of the early universe, uncontaminated by later developments.

12.11.8 Polarization Power Spectra

Theory

Equation (10.78) is an expression for the polarization moments from a single plane wave. In the real universe, we have not just one plane wave, but a superposition of many waves, all with differing amplitudes $\Theta_P(\vec{k},\hat{n})$. The angular power spectrum from a superposition of plane waves follows from the identical calculation on the temperature anisotropies (Eq. (8.68)):

$$C_{P,l} = \frac{2}{\pi}\int_0^\infty dk k^2 |\Theta_{Pl}(k)|^2$$

For quite a while, cosmologists computed this power spectrum without reference to Q or U. In 1997, a flurry of papers appeared which derived the power spectra for Q and U. These exploited Eq. (10.66) or large-angle generalizations of it.

Based on our solution for the power spectra of the different components of shear in Section 10.4, we have a sense of what to expect for the power spectra of Q and U. Consider first Figure 10.17. In the small angle limit, Q for example is a function of the 2D vector $\vec{\theta}$, the projection of $\hat{n}$ onto the plane perpendicular to the $\hat{z}$ axis. Thus, we can Fourier transform Q just as we Fourier transformed the shear fields above; its transform will depend on $\vec{l}$, the vector conjugate to $\vec{\theta}$. Based on our experience with weak lensing, we expect the power spectrum of $\tilde{Q}$ to depend not only on the magnitude of $\vec{l}$ but also on its orientation. Looking back at Eq. (10.66), we will not be surprised to find that this power spectrum, $C_{QQ}(\vec{l})$ is proportional to $C_{P,l}$. The proportionality constant is $\cos^2(2\phi_l)$, where ϕ_l is the angle $\vec{l}$ makes with the x-axis. Thus, the factor of $\cos(2\phi_k)$ in Eq. (10.66) becomes $\cos(2\phi_l)$ when we sum over all $\vec{k}$. Similarly, the power spectrum of $\tilde{U}$ is $\sin^2(2\phi_l)C_{P,l}$.

Let's derive this connection between the power spectra of Q and U and that of Θ_P explicitly. We can write the Q polarization as a sum over all plane waves:

$$Q(\vec{\theta}) = \int \frac{d^3k}{(2\pi)^3}\Theta_P(\vec{\theta},\vec{k})\cos(2\phi_k)$$

The modulating factor $e^{i\vec{x}\cdot\vec{k}}$ is set to 1 here, since we observe from only one position, and we are calling that position $\vec{x} = 0$. To deal with the $\cos(2\phi_k)$ factor, first note that it is equal to $\cos^2\phi_k - \sin^2\phi_k$, or in terms of the Cartesian components of $\vec{k}$:

$$\cos(2\phi_k) = \frac{k_x^2 - k_y^2}{k_x^2 + k_y^2}.$$

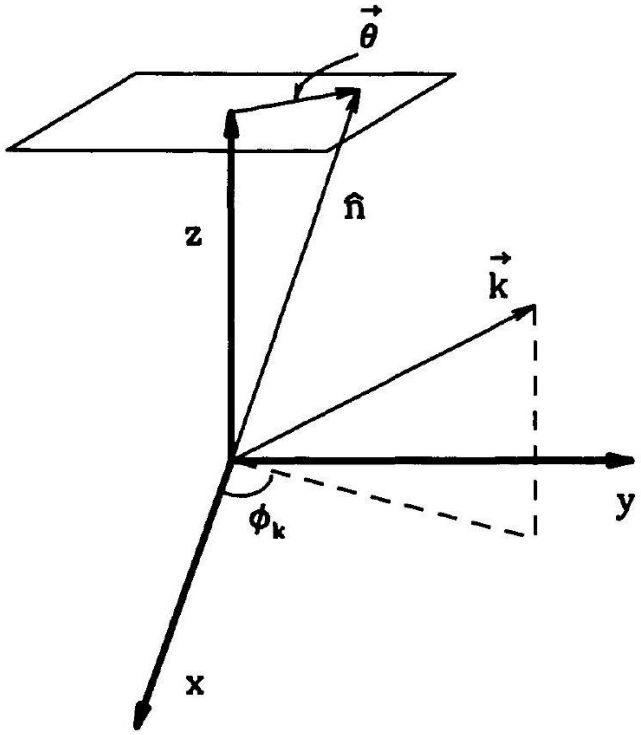


Figure 10.17. Different vectors in polarization. We observe radiation with incoming direction $\hat{n}$, also parameterized by 2D angle θ. Wavevector $\vec{k}$ has an azimuthal angle ϕ_k.

Since Θ_P has the exponential factor $e^{-i\vec{k}\cdot\hat{n}\eta_0}$ (e.g. Eq. (10.69)), and since $\hat{n}_x = \theta_x$ and $\hat{n}_y = \theta_y$, we can rewrite these Cartesian coordinates as derivatives with respect to the $\vec{\theta}$. For example, $k_x \to [-i\eta_0]^{-1}\,\partial/\partial\theta_x$. The full $\cos(2\phi_k)$ factor therefore can be written solely as derivatives with respect to $\vec{\theta}$:

$$\cos(2\phi_k) = D_Q(\vec{\theta}) := \left[\frac{\partial^2}{\partial\theta_x^2} + \frac{\partial^2}{\partial\theta_y^2}\right]^{-1}\left(\frac{\partial^2}{\partial\theta_x^2} - \frac{\partial^2}{\partial\theta_y^2}\right)$$

This expression looks formidable, but it is extremely useful for summing up many different k-modes. We can replace Eq. (10.80) with

$$\begin{aligned} Q(\vec{\theta}) &= D_Q(\vec{\theta})\int\frac{d^3k}{(2\pi)^3}\Theta_P(\vec{\theta},\vec{k}) \\ &= D_Q(\vec{\theta})\Theta_P(\vec{\theta}) \end{aligned}$$

Both $Q(\vec{\theta})$ and $\Theta_P(\vec{\theta})$ can be written in terms of their Fourier transforms, so that Eq. (10.83) becomes

$$\int\frac{d^2l}{(2\pi)^2}e^{i\vec{l}\cdot\vec{\theta}}Q(\vec{l}) = D_Q(\vec{\theta})\int\frac{d^2l}{(2\pi)^2}e^{i\vec{l}\cdot\vec{\theta}}\Theta_{Pl}$$

Now D_Q, which is so very complicated in θ-space, becomes very simple, for we know exactly what it looks like when it acts on the exponential $e^{i\vec{l}\cdot\vec{\theta}}$. In that case, it simply becomes $\cos(2\phi_l)$,

where ϕ_l is the angle that the 2D vector $\vec{l}$ makes with the x-axis. Therefore, the Fourier transform of $Q(\vec{\theta})$ is

$$\tilde{Q}(\vec{l}) = \tilde{\Theta}_{Pl} \cos(2\phi_l)$$

An identical argument says that $U(\vec{l}) = \Theta_{Pl} \sin(2\phi_l)$. Therefore, the power spectra of $\tilde{Q}$ and $\tilde{U}$ are

$$C_{QQ}(\vec{l}) = C_{P,l} \cos^2(2\phi_l)$$
$$C_{UU}(\vec{l}) = C_{P,l} \sin^2(2\phi_l) .$$

Recall that in the case of weak lensing, we noticed that one could take linear combinations of γ_1 and γ_2 such that the power spectrum of one of the linear combinations vanishes (Eq. (10.40)), while the other is equal to the convergence power spectrum (Eq. (10.39)). Here we can do exactly the same thing. If we define

$$E(\vec{l}) := \tilde{Q}(\vec{l}) \cos(2\phi_l) + \tilde{U}(\vec{l}) \sin(2\phi_l)$$
$$B(\vec{l}) := -\tilde{Q}(\vec{l}) \sin(2\phi_l) + \tilde{U}(\vec{l}) \cos(2\phi_l)$$

then

$$C_{BB}(\vec{l}) = 0$$

In the small-scale limit, the power in the E-mode is precisely equal to C_P:

$$\lim_{l \gg 1} C_{EE}(\vec{l}) = C_{P,l}$$

Figure 10.18 shows the resultant power spectrum, both the exact numerical result and the approximation of Eq. (10.78) integated over all modes as dictated by Eq. (10.79). Also shown is the spectrum of temperature anisotropies from Chapter 8. As expected it is higher in amplitude, since polarization is suppressed in the tightly coupled limit. Also as anticipated, the oscillations in the polarization spectrum are out of phase with those in the temperature spectrum. In 2002, the DASI experiment announced the first detection of polarization, a detection shown in Figure 10.18. This is akin to the first detection of shear by large scale structure, the beginning of our journey down promising new paths in cosmology.

The spectrum in Figure 10.18 is shown only on small scales; on larger scales, the treatment of this section needs to be modified (Kamionkowski, Kosowsky, and Stebbins, 1997a; Seljak and Zaldarriaga, 1997). The resultant spectrum has no surprises: it falls off very rapidly on large scales. Since the polarization is proportional to the dipole, which vanishes for large-scale modes, we could have anticipated this result as well. Although I won't go into the technical details of this large-angle

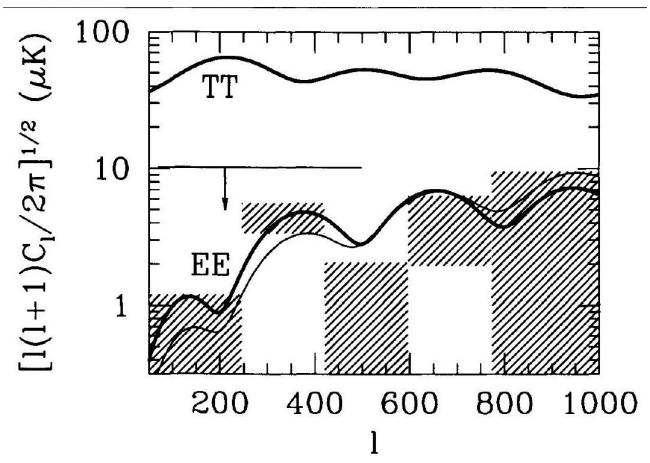


Figure 10.18. Power spectra of temperature and E-mode polarization for the standard CDM model. Thick curves show exact results; thin curve is the tight coupling approximation of Eq. (10.78). Only scalar perturbations have been assumed, so there is no power in the B-mode. Straight line at 10μK is an upper limit from Hedman et al., 2002, while the hatched boxes are the first detection by the DASI experiment (Kovac et al., 2002).

result, the basic idea is that instead of expanding polarization in terms of Legendre polynomials, or ordinary spherical harmonics, one must use tensor spherical harmonics.

One final comment: we have been implicitly assuming until now that the perturbations of interest are scalar. We inserted this assumption early on by writing the plane-wave perturbation as Eq. (10.55). If the perturbations were tensor, the decomposition would have included an azimuthal dependence; recall Eq. (4.115).

12.11.9 Detecting Gravity Waves

Theory

There is a fundamental difference between the scalar perturbations we have considered in the previous sections and tensor perturbations. A scalar plane-wave perturbation has one direction associated with it: the direction of the wavevector $\vec{k}$. Once this direction is specified, all photon moments depend only on the angle between the incoming photon and the wavevector. Once this angle is specified, there is an azimuthal symmetry about the $\hat{k}$ direction. This rotational symmetry is the reason that only the E mode is produced by scalar perturbations. There are two directions in a polarization field: (i) the direction in which the polarization strength is changing and (ii) the orientation of the polarization. For scalar perturbations, we saw in Figure 10.16 that these directions must be aligned (or perpendicular to each other). Intuitively, each direction looks to the only vector it knows $-\vec{k}$ - for guidance, and they each arrive at the same end. This alignment is the salient characteristic of the E mode.

The photon distribution from tensor perturbations is not rotationally symmetric about the $\vec{k}$-direction. Gravity waves are pulsations in the metric; these induce an azimuthal dependence to the photon distribution. Recall from Eq. (4.115) that the resultant distribution varies as $\sin(2\phi)$ or $\cos(2\phi)$, where ϕ is the azimuthal angle about the $\vec{k}$-axis. This dependence on ϕ means that there is an additional direction to choose from when the polarization field gets induced. We might expect then that the orientation of the polarization will not necessarily be aligned with the direction of changing polarization strength. That is, we might expect that gravity waves will produce B-mode polarization. This is exactly what we will show in this section.

Before working through the algebra, we need to pause to understand the importance of the B-mode generated by tensor perturbations. Let's start with the difficulty of detecting tensors through the E-mode. Both scalars and tensors contribute to the E-mode, so the only way to disentangle them is to take advantage of differences in their spectra as a function of l. We saw in the case of temperature anisotropies that this is a tricky game, though, for other parameters can change spectra in ways similar to tensors. So even if we had perfect knowledge of the $C_{E,l}$ spectrum (no noise), we would still not necessarily know whether tensors were present. The B-mode is different. There is no contamination from scalar perturbations, so if we observe a B-mode in polarization, we know that it comes from gravity waves. In principle, this realization has unlimited power: no matter how small the tensor signal from inflation (no matter how small $H/m_{\rm Pl}$), we can ultimately detect this signal by searching for a B-mode. In practice, there are contaminants due to nonlinearities, but these are quite small. Estimates (Knox and Song, 2002; Kesden, Cooray, and Kamionkowski, 2002) suggest that the lowest obtainable limit on r, the tensor-to-scalar ratio, is of order 10^{-4}.

Let's compute the polarization pattern from a single plane wave generated by tensor perturbations. This problem is identical to that treated in Section 10.6. To find the outgoing polarization near the z-axis, we need to integrate over the incoming photon distribution. As in Eqs. (10.52) and (10.53), we want

$$\begin{aligned}\begin{pmatrix} Q \\ U \end{pmatrix} &\propto -\int d\Omega' \Theta^T(\Omega') \sin^2\theta' \begin{pmatrix} \cos(2\phi') \\ \sin(2\phi') \end{pmatrix} \\ &\propto -\int d\Omega' \Theta^T(\Omega') \begin{pmatrix} Y_{2,2}(\Omega') + Y_{2,-2}(\Omega') \\ \frac{1}{i}\left[Y_{2,2}(\Omega') - Y_{2,-2}(\Omega')\right] \end{pmatrix},\end{aligned}$$

where I have inserted the photon distribution due to tensor perturbations, Θ^T; recognized the combination of $\sin^2\theta'$ and the azimuthal dependence as $Y_{2,2} \pm Y_{2,-2}$; and neglected the absolute normalization of the polarization.

To complete the calculation, we need to find the angular dependence of Θ^T. This is a bit more difficult than one might expect. Although we know that this angular dependence is $\sin^2\theta' \cos(2\phi')$ (for h_+) or $\sin^2\theta' \sin(2\phi')$ ($h_\times$)for $\vec{k}$ lying along the z-axis, we need the dependence for a general wavevector $\vec{k}$. One way of finding this dependence is to rotate the coordinate system so that a unit vector pointing in the $\hat{z}$-direction gets rotated so that it points in the $\hat{k}$-direction. The relevant rotation matrix is

$$R = \begin{pmatrix} \cos\theta_k \cos\phi_k & -\sin\phi_k & \sin\theta_k \cos\phi_k \\ \cos\theta_k \sin\phi_k & \cos\phi_k & \sin\theta_k \sin\phi_k \\ -\sin\theta_k & 0 & \cos\theta_k \end{pmatrix}.$$

You should verify that R really does take $\hat{z} \to \hat{k}$ and work through a simple derivation of R (Exercise 13). We want to know what R does to Θ^T. To be concrete, let's focus on h_χ, so that $\Theta^T \propto \sin^2\theta' \sin(2\phi')$. First, we can reexpress this angular dependence in terms of the unit vector $\hat{n}'$ describing the direction of the incident photon:

$$\begin{aligned}\sin^2\theta' \sin(2\phi') &= 2\sin^2\theta' \sin\phi' \cos\phi' \\ &= 2\hat{n}'_x \hat{n}'_y.\end{aligned}$$

Now let's rotate the coordinate system so that the z-axis points in the direction of $\vec{k}$. The anisotropies due to the h_x-mode used to be proportional to $\hat{n}'_x \hat{n}'_y$. In the new coordinate system, they become

$$\Theta^T \propto \left(R^t \hat{n}'\right)_x \left(R^t \hat{n}'\right)_y$$

where ^{t} denotes transpose.

Now we work through the matrix multiplication and find

$$\begin{aligned}&\Theta^T \propto (\cos\theta_k \cos\phi_k \sin\theta' \cos\phi' + \cos\theta_k \sin\phi_k \sin\theta' \sin\phi' - \sin\theta_k \cos\theta') \\ &\times (-\sin\theta' \sin\phi_k \cos\phi' + \cos\phi_k \sin\theta' \sin\phi') \\ &= \sin\theta' \sin(\phi' - \phi_k)(\cos\theta_k \sin\theta' \cos(\phi' - \phi_k) - \sin\theta_k \cos\theta') \\ &= \frac{1}{2}\cos\theta_k \sin^2\theta' \sin[2(\phi' - \phi_k)] - \sin\theta' \cos\theta' \sin\theta_k \sin(\phi' - \phi_k).\end{aligned}$$

This last combination can be reexpressed in terms of spherical harmonics: it is a linear combination of $Y_{2,\pm2}, Y_{2,\pm1}$, and $Y_{2,0}$. That is, the anisotropy pattern about the wavevector $\vec{k}$ due to gravity waves (the $h_\times$mode) has a $Y_{2,2} - Y_{2,-2}$ dependence when $\hat{k}$ is along the z-axis. When $\hat{k}$ is general, this dependence gets mixed up among all the $Y_{2,m}$'s. We are interested in the polarization pattern generated by Θ^T; from Eq. (10.90) and the orthogonality property of

the spherical harmonics, this means we are interested only in the $Y_{2,\pm 2}$ components of Θ^T. We can now extract these from Eq. (10.94). The last term on the right has a factor of $\sin(\phi' - \phi_k)$, so it is proportional to $Y_{2,\pm 1}$, and we can neglect it. The first term is

$$\frac{1}{2}\cos\theta_k \sin^2\theta' \sin\left[2\left(\phi' - \phi_k\right)\right] = -\sqrt{\frac{32\pi}{15}}\frac{\cos\theta_k}{4i}\left[e^{-2i\phi_k}Y^*_{2,-2}\left(\Omega'\right) - e^{2i\phi_k}Y^*_{2,2}\left(\Omega'\right)\right]$$

To find Q and U, we dot this into $Y_{2,2} \pm Y_{2,-2}$, so that the integral in Eq. (10.90) leads to

$$\begin{pmatrix} Q(\hat{z}) \\ U(\hat{z}) \end{pmatrix} \propto \Theta_0^T \cos\theta_k \begin{pmatrix} -\sin\left(2\phi_k\right) \\ \cos\left(2\phi_k\right) \end{pmatrix}$$

If we move small angles away from the $\hat{z}$-direction, then the azimuthal dependence does not change, and $\cos\theta_k \to \hat{n}\cdot\hat{k}$, so that

$$\begin{pmatrix} Q(\hat{n}) \\ U(\hat{n}) \end{pmatrix} \propto \Theta_0^T \hat{n}\cdot\hat{k} \begin{pmatrix} -\sin\left(2\phi_k\right) \\ \cos\left(2\phi_k\right) \end{pmatrix}$$

The polarization pattern described by Eq. (10.97) has a nonzero B-mode. To see this, first consider the definition of B in Eq. (10.87). This definition is in Fourier space, but we remember that using the operator $D_{Q,U}$, we can replace $\vec{l}$ with $\vec{k}$, the wavevector. Therefore, for $\vec{k}$ in the $\hat{x}$ - $\hat{z}$ plane ($\phi_k = 0$), the B-mode corresponds to only U polarization. Indeed this is precisely what Eq. (10.97) says is produced by the $h_\times$mode of gravity waves. So the anisotropies due to gravity waves do produce the B-mode of polarization.

Figure 10.19 shows the polarization patterns due to a single plane wave $h_\times \propto e^{i\vec{k}\cdot\vec{x}}$ for four different wavevectors $\vec{k}$. For example, the top left panel considers $\vec{k}$ lying along the x-axis. In that case, since $\phi_k = 0$, Eq. (10.97) says that the polarization is all U and that the strength scales as

$$\hat{n}\cdot\hat{k} = \hat{n}_x$$

The strength of the polarization therefore increases as one moves away from the y-axis. The important feature of this pattern, which characterizes the B-mode, is that the strength of the polarization varies in the $\hat{x}$-direction, while the orientation is in the $\hat{x} \pm \hat{y}$ direction. These two directions (varying strength and polarization orientation) are not aligned or perpendicular to each other. The other panels show the same feature.

Figure 10.20 shows the anisotropy spectrum in a standard CDM model with an equal amount of tensors and scalars. The T and E spectra are similar to the tensorless case depicted in Figure 10.18. With tensors, the B spectrum is now nonzero, albeit small. Studies suggest that polarization searches will help significantly in the quest to detect small r.

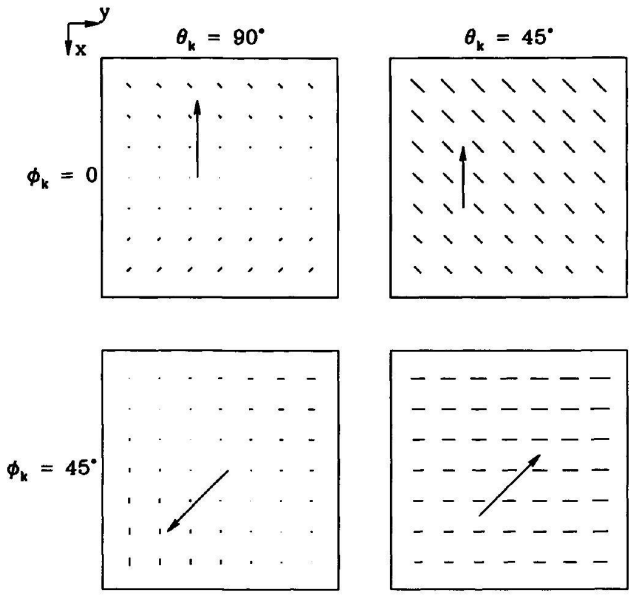


Figure 10.19. Polarization patterns from a single plane wave $h_\times(\vec{k})$ in a plane perpendicular to the z-axis ($\hat{z}$ out of the paper). Patterns from four different $\vec{k}$ are shown. Arrows depict direction of increasing polarization strength. This direction is not aligned with the orientation of the polarization.

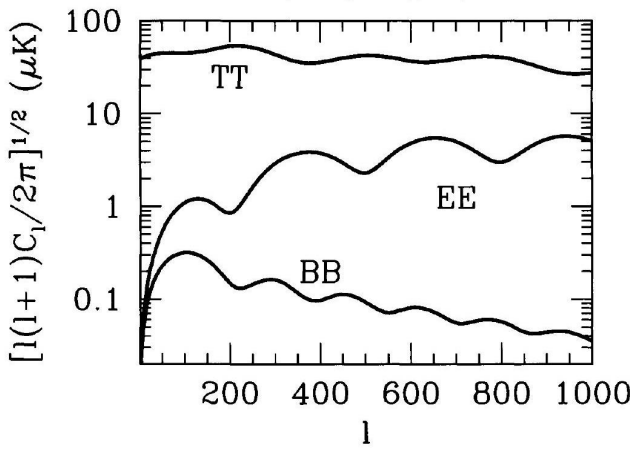


Figure 10.20. Anisotropy spectrum from a standard CDM model with equal amounts of scalar and tensor perturbations ($r = 1$). The T and E spectra come from both scalars and tensors, whereas the B-mode is due solely to tensors.

Part V
Problems

13 Typical problems of modern cosmology

13.1 Problems about Homogeneous Dynamics and Thermal History

13.1.1 Mathematical Problems for Basic Preparation to Cosmology

(я не уверен, что тут, но мб там диффгем или кинетика какая-то нужна будет)

13.1.2 Typical Problems about Friedman Universe (?!??!)

Leuven-2023hw1 **1.2. Universe with matter and a cosmological constant**

Consider a spatially flat Friedmann-Lemaitre-Robertson-Walker (FLRW) universe filled with only matter and a cosmological constant (i. e. no radiation).

a) Show that solving the Friedmann equation (normalized to $a(t_0) = 1$ at present time t_0) leads to the following solution for the scale factor:

$$a(t) = \left[\frac{\Omega_{\mathrm{m},0}}{1-\Omega_{\mathrm{m},0}} \sinh^2\left(\frac{3}{2}\left(1-\Omega_{\mathrm{m},0}\right)^{1/2} H_0 t\right)\right]^{1/3},$$

where $\Omega_{\mathrm{m},0}$ is the current density parameter for matter and H_0 is today's Hubble parameter. Discuss the evolution of the scale factor for $t \to \infty$ and $t \to 0$.

b) Derive an expression for the age of the universe t_0 in terms of H_0 and $\Omega_{\mathrm{m},0}$. Sketch the dependence $t_0(\Omega_{\mathrm{m},0})$ and evaluate it explicitly for the case $\Omega_{\mathrm{m},0} = 0.3$ and for an Einstein-de Sitter universe ($\Omega_{\mathrm{m},0} = 1$). Which universe is older (for fixed H_0)?

c) Show that the energy density of the universe becomes dominated by the cosmological constant at the following redshift: $z_\Lambda = \left(\Omega_{\mathrm{m},0}^{-1} - 1\right)^{1/3} - 1$.

a) Solving Friedman equations

Basically the dependence is $a(t) \sim \mathrm{sh}^{2/3}(\mathrm{const} \cdot t)$. The Friedmann-Lemaitre equation will give us:

$$\frac{\dot{a}^2}{a} = H_0^2\left(\frac{\Omega_{m,0}}{a^3} + 1 - \Omega_{m,0}\right) \qquad \text{because } \Omega_\lambda + \Omega_{m,0} = 1.$$

$$\frac{da}{dt} = H_0 a\left(\frac{\Omega_{m,0}}{a^3} + 1 - \Omega_{m,0}\right)^{1/2},$$

$$\int \frac{\sqrt{a}}{\sqrt{\Omega_{m,0} + \left(1-\Omega_{m,0}a^3\right)}} da = H_0 t.$$

Defining x such that $x^2 = \frac{1-\Omega_{m,0}}{\Omega_{m,0}} a^3 \equiv ca^3$, we have $da = 2xdx/(3ca^2)$, so $\sqrt{a}da = \sqrt{a}2\sqrt{c}a^{3/2}dx/(3ca^2) = 2dx/(3\sqrt{c})$ and

$$\frac{2}{3\sqrt{1-\Omega_{m,0}}} \int \frac{dx}{\sqrt{1+x^2}} = H_0 t.$$

Since $\int \frac{dx}{\sqrt{1+x^2}} = \mathrm{sh}\, x + C$ and $a(0) = 0$, we have

$$a(t) = \left[\frac{\Omega_{m,0}}{1-\Omega_{m,0}} \mathrm{sh}^2\left(\frac{3}{2}\left(1-\Omega_{m,0}\right)^{1/2} H_0 t\right)\right]^{1/3}.$$

For $t \to 0$ curvature $a \to 0$, so $\Omega_m \to 1$, so we would expect the universe to be matter dominated and hence from $H^2(a) \sim 1/a^3$ we expect $a \sim t^{2/3}$. From our solution we have the same

$$a \sim \left(\frac{(1-\Omega_{m,0})t^2}{1-\Omega_{m,0}} \right)^{1/3} \sim t^{2/3}.$$

If time gets large, the cosmological constant starts to dominate, $\Omega_m = 0$ and from $H^2(a) \sim$const we expect an exponential dependence. From our formula since for t large $\operatorname{sh} t \sim e^t$,

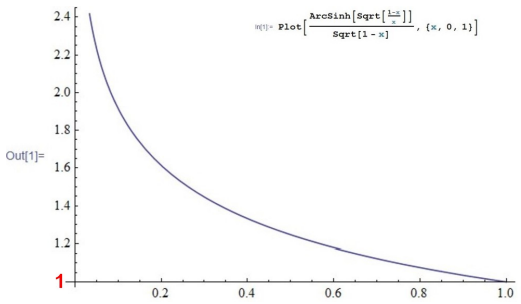
$$a \sim e^{(1-\Omega_{m,0})^{1/2}t}.$$

b) Finding age of the Universe

Now $a(t_0) = 1$, so after inverting $a(t)$ we get

$$t_0 = \frac{2}{3\left(1-\Omega_{m,0}\right)^{1/2} H_0} \operatorname{ArcSh}\left(\sqrt{\frac{1-\Omega_{m,0}}{\Omega_{m,0}}} \right).$$

The sketch:



For the case $\Omega_{\mathrm{m},0} = 1$ we should or solve Friedmann equations again or take the limit $\Omega_{m,0} \to 1$:

$$t_0 = \frac{0.66}{H_0} \approx 0.66 \cdot 14.4\text{bl years} \approx 9.5 \text{ bl years}.$$

For $\Omega_{m,0} = 0,3$

$$t_0 = \frac{0.66 \cdot 1.44}{H_0} \approx 0.95 \cdot 14.4\text{bl years} \approx 13.7 \text{ bl years}.$$

We see that if $\Omega_{\mathrm{m},0}$ greater, than its lifetime is smaller.

c) Critical energy density

Critical point we have $\Omega_{m,0}/a^3 = \Omega_\Lambda$. The redshift parameter z is such that $1+z = 1/a$, so from the main

$$\Omega_m(1+z)^3 = 1-\Omega_m$$

we get

$$z = \sqrt[3]{\frac{1}{\Omega_m} - 1} + 1.$$

Leuven-2023hw1.3. Horizon scale at matter-radiation equality

Solve the Friedmann-Lemaitre equation for a flat universe containing only matter and radiation in terms of the conformal time η. Then show that the conformal time at matter-radiation equality a_{eq} is given by:

$$\eta_{\mathrm{eq}} = 2(\sqrt{2}-1) H_0^{-1} \Omega_{\mathrm{m},0}^{-1/2} a_{\mathrm{eq}}^{1/2}$$

Solution

Friedman'e equations:

$$\frac{H^2}{H_0^2} = \frac{\Omega_{r,0}}{a^4} + \frac{\Omega_{m,0}}{a^3}.$$

at equality $\frac{\Omega_{r,0}}{a_{eq}^4} = \frac{\Omega_{m,0}}{a_{eq}^3}$, so $a_{eq} = \frac{\Omega_r}{\Omega_m}$ and from $d\eta := \frac{dt}{a}$ we have $\frac{\dot{a}}{a} = \frac{da}{a^2 d\eta}$ and

$$\frac{1}{H_0^2}\left(\frac{da}{a^2 d\eta}\right)^2 = \Omega_m\left(\frac{a_{eq}}{a^4} + \frac{1}{a}\right)$$

$$d\eta = \frac{da}{\Omega_m^{1/2} H_0 (a_{eq} + a)^{1/2}}$$

$$\eta_{eq} - 0 = \frac{a_{eq}^{1/2}}{\Omega_m^{1/2} H_0} \int_1^2 \frac{d(1 + a/a_{eq})}{1 + a/a_{eq}}$$

$$\eta_{eq} = \frac{2(\sqrt{2} - 1)\sqrt{a_{eq}}}{\Omega_m^{1/2} H_0 a_{eq}}$$

(???? дальше нужно додумывать!!!)
I didn't understand the second question, but we can obtain a particle horizon of about

$$H_p\left(t_{\rm CMB}\right) = c\eta_{\rm CMB} = 284{\rm Mpc} = 8.9 \times 10^{-3} H_p\left(t_0\right)$$

which corresponds to a proper size at that time of:

$$a_{\rm CMB} H_p\left(t_{\rm CMB}\right) = 261{\rm kpc}$$

Since we observe the CMB to be emitted essentially from our particle horizon (284Mpc $\ll$ 14.4Gpc), our expectation is that parts of the cosmic microwave background (CMB) that are separated by about a fraction of a great circle across the sky of

$$f = \frac{H_p\left(t_{\rm CMB}\right)}{H_p\left(t_0\right)}$$

(an angular size of $\theta \sim 1.7^\circ$) should be out of causal contact with each other.

Leuven-2023hw1.4. Unstability of matter and Λ

Identify what conditions have to be fulfilled such that a Universe with matter and a cosmological constant is static.
a) Show that this universe is unstable.
b) Find the radius of the Universe.

Solution

The acceleration equation

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3c^2}(\rho + 3P)$$

tells us that a static universe is only possible if $\rho = -3P$. Obviously this is not possible if we have only matter ρ_m with $P_m = 0$ or only a cosmological constant $\rho_\Lambda = -P_\Lambda$. But in a universe with both, we can have

$$\rho = \rho_m + \rho_\Lambda = -3P = 3\rho_\Lambda \quad \Rightarrow \quad \rho_m = 2\rho_\Lambda$$

The Friedmann equation $H^2 = \frac{8\pi G}{3c^2}\rho + \frac{\Lambda}{3} - \frac{kc^2}{R^2 a^2}$ is then

$$H^2 = \frac{8\pi G}{3c^2}\left(\rho_m + \rho_\Lambda\right) - \frac{kc^2}{R^2 a^2}$$

and the right-hand side vanishes if we take a positively curved universe, $k = +1$, with radius

$$Ra = \frac{c^4}{8\pi G \rho_\Lambda} = \frac{c^2}{\Lambda}$$

This is the Einstein static universe. It is unstable. If a is a little smaller than the critical value Ra then $\rho_m \sim a^{-3}$ is a little larger and the acceleration equation $\frac{\ddot{a}}{a} = -\frac{4\pi G}{3c^2}(\rho + 3P)$ says that a will decrease further. Similarly, if a is larger than the critical value it will increase further.

(??? про радиус так и не понял, откуда он в этом уравнении??? не верное уравнение - удалю этот вопрос!!!)

Leuven-2023hw1.5. Singularity theorem for FLRW universes (????)

Ordinary matter with energy density ρ and pressure p satisfies the strong energy condition (SEC), which reads:

$$\rho + 3p \geq 0$$

Prove that in a FLRW universe, if the matter content always satisfies the SEC, then there must have been a Big Bang singularity $a(t) = 0$ at a time t such that $t_0 - t \leq H_0^{-1}$, where $H_0 > 0$ is the present Hubble parameter.

This shows that General Relativity does not allow for a universe that transitions from a contracting to an expanding phase (assuming SEC).

Solution

(я по итогу забил на эту задачу вообще, так и не разобрался, потом подумаю мб)

Solution We need to prove that in $t > 0$ there is another big bang in this model. This is obvious, because from

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3c^2}(\rho + 3P)$$

the second derivative will be bigger and bigger in it absolute value and negative, so our convex expansion will be never to infinity, but definitely will come back to 0.

(???? дальше какой-то текст, я забыл)

We have the equation

$$\ddot{a} = -\frac{4\pi G}{3}(\rho + 3p)a.$$

If the strong energy condition is satisfied, then this informs us that the second derivative of the scale factor is always negative. We now show that this leads to a singularity. Because the universe is expanding, we know that if we go back in time, the scale factor decreases. To avoid reaching zero in finite time, we have two possibilities:

1. The scale factor has a local minimum at some point. However, this is not possible because at this point the second derivative will be positive.

2. The scale factor asumptotes to zero at minus infinity. Thus, the first derivative also approaches zero. Because we know this is positive for all times up to present day (since there are no local minima), this implies that the second derivative should be positive for some early times, which is also not possible.

Hence, this proves that we reach zero in a finite amount of time. To estimate this time, we use Taylor's theorem. Denote by t_0 the present time. Then for all times $t < t_0$ we have by the theorem some $t < t_1 < t_0$ which satisfies

$$a(t) = 1 + H_0\,(t - t_0) + \frac{\ddot{a}\,(t_1}{2}\,(t - t_0)^2\,.$$

Since the second derivative is negative, we thus have that for all times $t < t_0$

$$a(t) < 1 + H_0\,(t - t_0)\,.$$

The right hand side becomes 0 at the time t with $t_0 - t = H_0^{-1}$. Hence, it follows that a should become zero at some earlier time. (Note that this also gives a more concrete less descriptive proof of the assertion that we reach zero at some finite time.)

(ниже что-то не про задачу написано.)

Eq. $\chi_{\rm ph}(\tau) = \tau - \tau_i = \int\limits_{t_i}^{t} \frac{\mathrm{d}t}{a(t)}$ can be written as

$$\chi_{\rm ph}(\tau) = \int\limits_{t_i}^{t} \frac{\mathrm{d}t}{a} = \int\limits_{a_i}^{a} \frac{\mathrm{d}a}{a\dot{a}} = \int\limits_{\ln a_i}^{\ln a} (aH)^{-1}\ \mathrm{d}\ln a,$$

where $a_i \equiv 0$ corresponds to the Big Bang singularity. The causal structure of the spacetime can hence be related to the evolution of the comoving Hubble radius $(aH)^{-1}$. For a universe dominated by a fluid with constant equation of state $w \equiv P/\rho$, we get

$$(aH)^{-1} = H_0^{-1} a^{\frac{1}{2}(1+3w)}.$$

With SEC $1+3w > 0$ the integral above is dominated by the upper limit and receives vanishing contributions from early times. We see this explicitly in the example of a perfect fluid. We find

$$\chi_{\rm ph}(a) = \frac{2H_0^{-1}}{(1+3w)}\left[a^{\frac{1}{2}(1+3w)} - a_i^{\frac{1}{2}(1+3w)}\right] \equiv \tau - \tau_i.$$

The fact that the comoving horizon receives its largest contribution from late times can be made manifest by defining

$$\tau_i \equiv \frac{2H_0^{-1}}{(1+3w)} a_i^{\frac{1}{2}(1+3w)} \overset{a_i \to 0, w > -\frac{1}{3}}{\longrightarrow} 0.$$

Leuven.exam.1.1 Recombination (8 p)

Around 380,000 years after the Big Bang, protons and electron combined to form hydrogen atoms, an epoch called recombination.

(i) Given the interaction rate Γ for the process $e^- + p^+ \leftrightarrow H + \gamma$, describe how we can determine when the reaction is in thermal equilibrium in the context of an expanding universe.

(ii) Show that, using the Boltzmann equation in equilibrium (the Saha equation), the electron fraction X_e prior to recombination is given by

$$\frac{X_e^2}{1-X_e} = \frac{\pi^2}{2\zeta(3)}\eta\left(\frac{m_e}{2\pi T}\right)^{3/2} e^{-\varepsilon_0/T}$$

where $\varepsilon_0 = m_e + m_p - m_H$ is the binding energy of H. You can assume that the universe is neutral and that $n_b = n_p + n_H$. (Masses and the value of ϵ_0 were given).

(iii) Define the temperature of recombination $T_{\rm rec}$ such that $X_e\left(T_{rec}\right) = 10^{-1}$. Show that $T_{\rm rec} \sim 0.3$eV. You can use that $\ln\left(10^{17}\right) \sim 40$. Why is this much lower than the value of ϵ_0?

(iv) Discuss qualitatively electron freeze-out and photon decoupling, two important consequences of recombination. Make a sketch of $X_e(T)$ (from high to low temperature) and clearly denote $T_{\rm rec}$ and $T_{\rm dec}$ (temperature of photon decoupling) in your plot. Also show the prediction by the Saha equation.

Solution

13.1.3 Other Problems about Friedman Universe (?!??!)

Weak field limit

Show that in the weak field limit of General Relativity the source of gravity is the quantity $(\rho + 3p)$.

Solution

Then we linearize the Einstein equations

$$R_{\mu\nu} = 8\pi G\left(T_{\mu\nu} - \frac{1}{2}T g_{\mu\nu}\right)$$

with respect to $h_{\mu\nu}$. In the first order of approximation we can neglect terms quadratic on Γ in the Ricci tensor

$$R_{\nu\sigma} = \Gamma^{\mu}_{\nu\sigma,\mu} - \Gamma^{\mu}_{\nu\mu,\sigma} + \Gamma^{\alpha}_{\nu\sigma}\Gamma^{\mu}_{\alpha\mu} - \Gamma^{\alpha}_{\nu\mu}\Gamma^{\mu}_{\alpha\sigma}$$

Also, for stationary metrics the second term in the 00 component equals to zero, so

$$R_{00} = \partial_\alpha\Gamma^{\alpha}_{00} = \frac{1}{2}\Delta g_{00} = \Delta\Phi.$$

It is easy to see that $T_{\mu\nu}$ is of the first order of smallness with respect to Φ, so from the 00 component of the Einstein equation we obtain

$$\Delta\Phi = 8\pi G\left(T_{00} - \frac{1}{2}T\right) = 8\pi G\left(\rho - \frac{1}{2}(\rho - 3p)\right) = 4\pi G(\rho + 3p).$$

Therefore the source of gravitational field is not ρ, but the combination $\rho + 3p$. In particular, an object satisfying the condition $\rho + 3p < 0$, will repel, rather than attract, non-relativistic particles.

Leuven-exam.1.2 Big Bang and Quantum Cosmology (4 p)

Given that the matter content of the universe always satisfied the strong energy condition $\rho + 3P \geq 0$, prove that the universe always had a Big Bang in the past at a time $t(a(t) = 0)$ such that $t_0 - t < 1/H_0$. How does quantum cosmology solve this problem? Describe what form the solution would take, and what we might learn from this (max 1/2 page).

Solution

13.1.4 Solutions by Dodelson to Chapter 1

Solution to Dodelson-1.1

The ratio

$$\frac{\rho_\Lambda}{3H^2/(8\pi G)} = (\rho_\Lambda/\rho_c)_0 \left(\frac{H_0}{H}\right)^2$$

where subscript 0 means evaluate today, where it is assumed to be 0.7. Again, by assumption, the universe is forever radiation dominated (clearly not true today, but a good approximation early on), so $H/H_0 = a^{-2}$. The temperature also scales as a^{-1}, so $H/H_0 = (T/T_0)^2$ with $T_0 = 2.7$ K $= 2.3 \times 10^{-4}$eV. So,

$$\frac{\rho_\Lambda}{3H^2/(8\pi G)} = 0.7 \left(\frac{T_0}{T}\right)^4$$

At the Planck scale, $T_0/T = 2.3 \times 10^{-4}/1.22 \times 10^{28}$, so

$$\frac{\rho_\Lambda}{3H^2/(8\pi G)} = 9 \times 10^{-128}$$

This is the so-called fine-tuning problem: for the cosmological constant to be important today, it had to have been fine-tuned to an absurdly small value at early times. It's a deep problem.

Solution to Dodelson-1.2

We need to do the integral

$$t_0 = \frac{1}{H_0} \int_0^1 \frac{da}{a} \left[\Omega_\Lambda + \frac{1-\Omega_\Lambda}{a^3}\right]^{-1/2}$$

for $\Omega_\Lambda = 0.7$ and 0. The latter case can be done analytically:

$$\int_0^1 \frac{da}{a} a^{3/2} = \frac{2}{3}$$

So $t_0 = 2/3H_0 = 0.67 \times 10^{10} h^{-1}$ yrs. When Ω_Λ is not zero, the integral needs to be done numerically. I find

$$\int_0^1 \frac{da}{a} \left[0.7 + \frac{0.3}{a^3}\right]^{-1/2} = 0.96$$

So for fixed Hubble constant, a cosmological constant universe is older than a matter-dominated one, older by a factor of $0.96/0.67 = 1.43$. For $h = 0.7$, a cosmological constant universe has an age of 14 billion years, in accord with other observations of the age of the universe.

Solution to Dodelson-1.4

An inverse wavelength is ν/c, so replacing ν everywhere in Eq. (1.8) by c/λ leads to

$$I_\nu = \frac{4\pi\hbar c}{\lambda^3} \frac{1}{\exp\{2\pi\hbar c/\lambda k_B T\} - 1}.$$

This is energy per Hz; we want energy per cm^{-1}, so we need to multiply by c, leaving

$$I_{1/\lambda} = \frac{4\pi\hbar c^2}{\lambda^3} \frac{1}{\exp\{2\pi\hbar c/\lambda k_B T\} - 1}.$$

Plugging in numbers leads to

$$I_{1/\lambda} = 1.2 \times 10^{-5} \text{ergsec}^{-1}\ \text{cm}^{-1}\text{sr}^{-1} \left(\frac{\text{cm}}{\lambda}\right)^3 \frac{1}{\exp\{0.53\ \text{cm}/\lambda\} - 1}.$$

A quick check verifies that this agrees with Figure 1.10.
To find the peak, differentiate I with respect to $1/\lambda$ and set equal to zero. This leaves

$$\lambda = \frac{1}{3}\frac{(2\pi\hbar c/k_B T)}{1-\exp\{-2\pi\hbar c/\lambda k_B T\}}.$$

So $1/\lambda_{\text{peak}}$ is $3/.53\ \text{cm}^{-1}$. The exact coefficient, accounting for the exponential is 2.82, so $1/\lambda_{\text{peak}} = 5.3\ \text{cm}^{-1}$, exactly where it occurs in Figure 1.10.

13.1.5 Problems on thermal history (!!?)

(добавлю в условия теор указания также потом про то, что тут происходит)

Leuven-2023hw2.1. Graviton decoupling

1) Estimate the temperature today of primordial gravitons which decoupled at $T \sim m_{\text{pl}}$ in
(i) the Standard Model with $g_{*S} = 106.75$ for all $T > 1\text{TeV}$
(ii) a large GUT model which has $g_{*S} \approx 10^3$ for $T > 10^{16}\text{GeV}$.
2) What was the graviton temperature at $T = 1\text{MeV}$?
Hint: (?? here is a little theory about what is going on????)

Solution

(???? пока не понял, что нужно понимать, чтобы это было очевидно???)
If the thermal degrees of freedom before decoupling are $g^{th}_{*S,b}$ and those after are $g^{th}_{*S,a}$, then $T_g = T_{\gamma,b}$. Indices "a" and "b" mean "after" and "before".
Because $g^{\text{th}}_{*S} \cdot (aT_\gamma)^3$ is constant, we find that

$$T_g = \left(\frac{g^{th}_{*S,a}}{g^{th}_{*S,b}}\right)^{1/3} T_{\gamma,a}.$$

But what if not all the relativistic degrees of freedom decouple at the same time? Then suppose we have a series of two decouplings. Denoting things at these two stages by the subscript a_i for $i = 1, 2$ and doing the calculation above, we find

$$T_g = \left(\frac{g^{th}_{*S,a_2}}{g^{th}_{*S,a_1}}\right)^{1/3}\left(\frac{g^{th}_{*S,a_1}}{g^{th}_{*S,b}}\right)^{1/3} T_{\gamma,a},$$

which yields the same result. Hence, we can just calculate as if all the degrees of freedom are decoupling at the same stage.
1. In the case of the standard model, we have $2 + 106,75$ degrees of freedom before decoupling. (The extra 2 are those of the graviton) After decoupling, the only degrees of freedom left are the 2 of the photon. Using $T_\gamma = 2,725K$ for the cosmic microwave background, we get

$$T_g = 0,719K$$

2. For the large GUT model, we have $2 + 10^3 \sim 10^3$ degrees of freedom before decoupling, and still only 2 after. This gives

$$T_g = 0,343K$$

We now repeat the calculation for $T = 1\text{MeV}$. This corresponds to $T = 1,16 \cdot 10^{10}$ K. Another change is that the electron and positron have not yet decoupled at this temperature. Thus the degrees of freedom after is now 5,5. Repeating the calculation then gives

$$T_g = 4,29 \cdot 10^9\ \text{K} \qquad \text{(Standard Model)}$$

$$T_g = 2,05 \cdot 10^9\ \text{K} \qquad \text{(Large GUT theories)}$$

Leuven-2023hw2.2. Relic baryon density

Consider massive particles b and antiparticles $\bar{b}$ with mass m and number densities $n(t)$ and $\bar{n}(t)$ respectively. Consider an interaction of the form

$$b + \bar{b} \leftrightarrow \ell + \bar{\ell}$$

where the particle and antiparticle are annihilated creating a lepton ℓ and an antilepton ℓ with a thermally averaged cross section $\langle \sigma v \rangle$, while the opposite reaction is also possible. Consider also that the leptons are at equilibrium, i.e., their number density is at the chemical equilibrium value $n_\ell = n_\ell^{\rm eq}$.

a) Express the Boltzmann equation for the particle density n as follows

$$\frac{\mathrm{d}n}{\mathrm{d}t} = -\frac{3\dot{a}}{a} n - \langle \sigma v \rangle \left(n\bar{n} - n^{eq}\bar{n}^{eq} \right),$$

where a is the scale factor of the FLRW metric, and $n^{\rm eq}$ and $n^?$ are the equilibrium particle (b) and antiparticle ($\bar{b}$) densities, respectively. Identify the significance of each term appearing in this equation. By considering also the evolution of the antiparticles, show that

$$(n - \bar{n})a^3 = \text{ const.}$$

b) Assuming particle-antiparticle symmetry, show that

$$\frac{1}{a^3}\frac{\mathrm{d}\left(na^3\right)}{\mathrm{d}t} = -\langle \sigma v \rangle \left[n^2 - \left(n^{(a)} \right)^2 \right]$$

Define $Y := n/T^3$ and $x := m/T$, where T is the temperature, and show that

$$\frac{\mathrm{d}Y}{\mathrm{d}x} = -\frac{\lambda}{x^2}\left[Y^2 - (Y^{\rm eq})^2 \right], \qquad \text{with} \qquad \lambda := \frac{m^3 \langle \sigma v \rangle}{H_{(T=m)}},$$

in a radiation dominated cosmology.

c) If λ is constant, show that at late times Y approaches a constant value given by

$$Y_\infty = \frac{x_f}{\lambda}$$

where x_f corresponds to the freeze-out temperature. Explain the dependence of Y_∞ on $\langle \sigma v \rangle$ and sketch the schematic evolution of Y versus x for both a strongly and a weakly interacting population of annihilating particles and antiparticles. If there was a speed-up in the expansion rate of the universe caused by the addition of extra low-mass neutrino species what would happen to the abundance of the surviving massive particles b and why?

d) Now consider that the particles in question are protons and antiprotons. You may use that $\langle \sigma v \rangle \sim 100\mathrm{GeV}^{-2}$. Show that the freeze-out temperature is

$$T_f \sim 20\mathrm{MeV}$$

e) Also compute

$$\frac{n}{n_\gamma} = \frac{\bar{n}}{n_\gamma} \sim 10^{-19}$$

where n_γ is the density of photons. How does this compare to observational data? What do you conclude about the abundances of protons and antiprotons in the early universe?

Solution a)

The Boltzmann equation for the baryons is given by

$$\frac{1}{a^3}\frac{d\left(na^3\right)}{dt} = -\langle \sigma v \rangle n\bar{n} + \langle \sigma v \rangle \cdot \left(\frac{n\bar{n}}{n_l \bar{n}_l} \right)_{eq} n_l \bar{n}_l,$$

Because the leptons are in equilibrium $n_l\bar{n}_l, = (n_l\bar{n}_l)_{eq}$,

$$\frac{dn}{dt} = -3n\frac{\dot{a}}{a} - \langle \sigma v \rangle \left(n\bar{n} - n_{eq}\bar{n}_{eq} \right).$$

The density of antiparticles $\bar{n}$ obeys the similar equation, so after subtraction

$$\frac{dn}{dt} - \frac{d\bar{n}}{dt} = -3\frac{\dot{a}}{a}(n - \bar{n}).$$

$$\frac{d}{dt}\ln(n - \bar{n}) + 3\frac{d}{dt}\ln a = 0.$$

$$\frac{d}{dt}\left(a^3(n - \bar{n}) \right) = 0.$$

Solution b)

With particle antiparticle symmetry $n = \bar{n}$ the Boltzmann equation is

$$\frac{1}{a^3}\frac{d\left(na^3\right)}{dt} = -\langle\sigma v\rangle\Big(n^2 - \left(n_{eq}\right)^2\Big).$$

With $Y := \frac{n}{T^3}$ and $x := \frac{m}{T}$ and knowing that $T(a) = T_0/a$, we then find

$$\begin{aligned}\frac{dY}{dx} &= \frac{dY}{dT}\left(\frac{dx}{dT}\right)^{-1} = \left(\frac{dn/dT}{T^3} - \frac{3n}{T^4}\right)\left(-\frac{m}{T^2}\right)^{-1} = \frac{1}{m}\left(-\frac{dn/dT}{T} + \frac{3n}{T^2}\right) = \\ &= \frac{1}{m}\left(-\frac{dn}{da}\frac{da}{T} + \frac{3na^2}{T_0^2}\right) = \frac{1}{m}\left(\frac{dn}{da}\frac{a^2}{T_0}\frac{a}{T_0} + \frac{3na^2}{T_0^2}\right) = \frac{1}{mT_0^2}\left(\frac{dn}{da}a^3 + 3na^2\right) = \\ &= \frac{d(a^3n)}{mT_0 da \cdot \frac{dt}{dt}\frac{a}{a}} = \frac{d(a^3n)}{dt}\frac{1}{mT_0 Ha}\end{aligned}$$

Then the Boltzmann equation gives

$$\frac{dY}{dx} = -\frac{T^4}{mH}\langle\sigma v\rangle\left(Y^2 - \left(Y_{eq}\right)^2\right).$$

Because in the radiation era $a \sim t^{1/2}$ and always $T \sim 1/a$, we should have $H \sim T^2$ or $H = H(m)/x^2$, because $T \equiv m/x$. We get

$$\frac{dY}{dx} = -\frac{\lambda}{x^2}\left(Y^2 - \left(Y_{eq}\right)^2\right), \qquad \lambda := \frac{m^3\langle\sigma v\rangle}{H_{(T=m)}}.$$

Solution c)

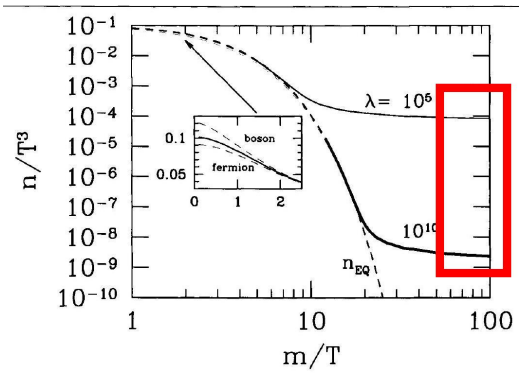
The solution is:

$$\frac{1}{Y_\infty} - \frac{1}{Y_f} = -\lambda\left(\frac{1}{Y_\infty} - \frac{1}{x_f}\right).$$

For late times $T \to 0$, so $1/Y_f \to 0$,. Integrating the resulting equation gives

$$Y_\infty = \frac{x_f}{\lambda}.$$

Because $\lambda \sim \langle\sigma v\rangle$, it is clear that Y_∞ will be inversely proportional to the interaction cross section. This is also logical since an efficient interaction means that more particles have annihilated. The dependence is sketched here:



If the expansion sped up, then the freezeout temperature would be reached earlier, which gives a larger abundance.

Solution d)

$$\Gamma \sim H$$

$$\Gamma = n\langle\sigma v\rangle, \qquad H \sim T^2/M_{pl}, \qquad n_p^{\rm eq} = g_p \left(\frac{m_p T}{2\pi}\right)^{3/2} \exp\left(\frac{\mu_i - m_p}{T}\right)$$

we look for a temperature T_f such that

$$g_p \left(\frac{mT_f}{2\pi}\right)^{3/2} e^{-m/T_f} \frac{M_{pl}}{T_f^2} \sim 1.$$

This gives

$$T_f^{-1/2} e^{-m/T_f} \sim 10^{-19}$$

which is solved by $T_f \sim 20$MeV.

Solution e)

Using the formulas we have derived, we find

$$\frac{n}{n_\gamma} = \frac{\pi^2}{2\zeta(3)} \frac{x_f}{\lambda}$$

Using $T_f = 20$MeV and filling in the numbers, we get

$$\frac{n}{n_\gamma} \sim 10^{-19}.$$

This is nowhere near the observed 10^{-9}. Hence, our assumption of baryon symmetry in the early universe is wrong.

Вопросы на додумать

(!!!??? там какая-то идея есть за этими преобразованиями, нужно ее продумать еще раз!!! пока слабо понимаю, почему именно в таких переменных решаем?)

Leuven-2023hw2.3a. Primordial Nucleosynthesis

Compute the ratio of the ^{3}He and proton number densities at equilibrium and discuss its dependence on the baryon-to-photon ratio η.
Hint: ^{3}He is produced in the reaction $D + p^+ \leftrightarrow He^3 + \gamma$.

Solution

The Saha equilibrium formula for the reaction $D + p^+ \leftrightarrow {}^3\mathrm{He} + \gamma$ gives us

$$\left(\frac{n_{He}}{n_D n_p}\right)_{eq} = \frac{g_{He}}{g_D g_p} \left(\frac{m_{He}}{m_D m_p} \frac{2\pi}{T}\right)^{3/2} e^{-(m_{He} - m_D - m_p)/T}.$$

$$(n_{He})_{eq} = (n_D n_p)_{eq} \frac{4}{3 \cdot 2} \left(\frac{3m_p}{2m_p \cdot m_p} \frac{2\pi}{T}\right)^{3/2} e^{-(m_{He} - m_D - m_p)/T}.$$

And from the reaction $n + p^+ \leftrightarrow \mathrm{D} + \gamma$ we know that

$$\left(\frac{n_{\rm D}}{n_p}\right)_{\rm eq} \approx \eta \left(\frac{T}{m_p}\right)^{3/2} e^{B_{\rm D}/T}; \qquad n_b = \eta n_\gamma = \eta \cdot \frac{2\zeta(3)}{\pi^2} T^3,$$

where $\eta := 5.5 \times 10^{-10} \left(\Omega_b h^2 / 0.020\right)$ is the baryon-to-photon ratio.

We get

$$(n_{He})_{eq} = \eta \left(\frac{T}{m_p}\right)^{3/2} \left(n_p^2\right)_{eq} \frac{2}{3} \left(\frac{3\pi}{m_p T}\right)^{3/2} e^{B_{He}/T} =$$
$$= \frac{2}{3}\eta \left(\frac{3\pi}{m_p^2}\right)^{3/2} \left(n_p^2\right)_{eq} e^{B_{He}/T}$$

Because η is very small, we need to drop to temperatures below B_{He} to get a significant contribution of the exponential. Hence, if η gets smaller, the freezeout temperature also gets lower.

Leuven-2023hw2.3b. Factors of ^{4}He and deuterium production

Discuss briefly the effect of the following on the production of ^{4}He and deuterium during primordial nucleosynthesis:
(i) The baryon density today (and thus the parameter η) is increased.
(ii) The neutron decay time is longer than supposed.
(iii) There are many more neutrinos than antineutrinos or photons in the universe today or, conversely, there are many more antineutrinos than neutrinos or photons.
(iv) The neutron mass is increased (but the binding energies of ^{3}He, D and ^{4}He are the same).
(v) The gravitational constant G varies with time and was larger in the past.

Solution

Our basic reactions and main equations for them are:

$$\mathrm{D} + p^+ \leftrightarrow {}^3\mathrm{He} + \gamma,$$
$$\mathrm{D} + {}^3\mathrm{He} \leftrightarrow {}^4\mathrm{He} + p^+.$$

$$\left(\frac{n_\mathrm{D}}{n_p}\right)_\mathrm{eq} \approx \eta \left(\frac{T}{m_p}\right)^{3/2} e^{B_\mathrm{D}/T}$$

$$(n_{He})_{eq} = \frac{2}{3}\eta \left(\frac{3\pi}{m_p^2}\right)^{3/2} \left(n_p^2\right)_{eq} e^{B_{He}/T}$$

So we just look at them and get all the answers. Ans also from some common sense.
1. If η increases, then because the freezeout temperature becomes earlier, we have more He^3 left. Hence, the quantities of He^4 produced will also get larger.
2. If the neutron lifetime is longer, more neutrons get to participate in the reactions, and hence the produced He^4 will increase.
3. The (anti)neutrinos do not participate in the required interactions and hence will not influence the produced quantities.
4. Similar to the neutrinos, the gravitons do not interact (at the relevant temperatures, the gravitational interaction becomes too weak), so they do not influence it.
5. Baumann shows that $G_F^2 T_f^5 \sim \sqrt{G_N g_\star} T_f^2$, so the freezout temperature would be reached earlier resulting in more production of He^4.

Leuven-2021hw-2.2. Chemical potential for electrons

a) If the chemical potential is vanishing $\mu \sim 0$, there is an equal number of particles and antiparticles. If $\mu \neq 0$, show that the net particle number $n - \bar{n}$ of a given particle species in the relativistic limit ($T \gg m$) is given by

$$n - \bar{n} = \frac{gT^3}{6\pi^2}\left[\pi^2 \left(\frac{\mu}{T}\right) + \left(\frac{\mu}{T}\right)^3\right].$$

You may use the integral $\int_0^\infty \mathrm{d}y \frac{y}{\mathrm{e}^y+1} = \frac{\pi^2}{12}$.

b) Use this result to evaluate the chemical potential of electrons in the relativistic limit. You may use the fact that the universe is neutral.

Solution

(типичная задача, допишу скоро, у Баумана тоже решена)

13.1.6 Problems on homogeneous inflation 1

(todo: solve them again!!!)

Leuven-2023hw2.4. The flatness problem

Assume that the universe is dominated by a perfect fluid with energy density $\rho(a)$, pressure $P(a)$ and equation of state $P = w\rho$, where $w = \text{const}$. Show that the density parameter $\Omega(a)$ defined with respect to the critical density $\rho_{\text{crit}}\,(a) = 3H^2(a)/(8\pi G)$ as $\Omega := \rho/\rho_{\text{crit}}$ fulfills the following differential equation:

$$\frac{\mathrm{d}\Omega}{\mathrm{d}(\ln a)} = (1+3w)\Omega(\Omega-1).$$

Show that the value $\Omega = 1$ is an unstable fixed point for $w > -1/3$, but corresponds to an attractor for $w < -1/3$. Describe briefly how this relates to the flatness problem and its resolution by an inflationary phase.

Solution

To derive the equation, we use the FL equations

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p),$$
$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \qquad \Leftrightarrow \qquad \dot{a}^2(\Omega-1) = k. \qquad \left(\Omega := \frac{8\pi G}{3H^2}\rho\right)$$

Taking a time derivative of this equation and using $\frac{d\Omega}{dt} = \dot{a}\frac{d\Omega}{da}$, we find

$$2\ddot{a}(\Omega-1) + \dot{a}^2\frac{d\Omega}{da} = 0$$

Inserting the second Friedmann equation and using our equation of state, we arrive at

$$\frac{8\pi G}{3}\rho a(1+3w)(\Omega-1) = \frac{d\Omega}{da}\dot{a}^2.$$

Eliminating ρ in favour of Ω yields

$$\frac{d\Omega}{d\ln a} = a\frac{d\Omega}{da} = (1+3w)\Omega(\Omega-1).$$

We can easily solve this differential equation by integrating

$$\int \frac{1}{\Omega(\Omega-1)}d\Omega = \int (1+3w)d(\ln a).$$

This equation tells us that when $\Omega \neq 0, 1$ (in these cases, the solution is that Ω remains constant of course)

$$\left|\frac{\Omega-1}{\Omega}\right| = e^{(1+3w)\ln a}.$$

As we go forward in time, $\ln a$ increases. Hence, we see that if $w < -\frac{1}{3}$ the value of Ω will be driven towards 1, whereas in the other case, it will move away from 1. Thus, in the first case we have an attractor, and in the other case an unstable fixed point. Thus, if we had an inflationary phase with some strange matter satisfying $w < -\frac{1}{3}$, this phase would have driven Ω close to 1, which would explain why it is still so close to 1 today.

Leuven-2023hw2.5. The horizon problem

Show that if you assume that the universe was always dominated by ordinary matter or radiation early on, the comoving horizon at the (very small) scale factor a_{E} can be expressed in terms of the comoving horizon today, a_0 and the Hubble parameters at a_{E} and a_0 as follows:

$$\chi_{\text{ph}}\,(a_{\text{E}}) = \frac{a_0 H_0}{a_{\text{E}} H_{\text{E}}}\chi_{\text{ph}}\,(a_0)\,.$$

Compute the ratio $a_0 H_0/\,(a_{\text{E}} H_{\text{E}})$ assuming that the universe has been dominated by radiation since a_{E} and that the temperature at a_{E} was $T_{\text{E}} \sim 10^{15}\text{GeV}$. For today's temperature you can use the temperature of the CMB. The horizon problem is solved if there was an inflationary phase before the radiationdominated phase, such that the comoving Hubble radius at the beginning of inflation a_{I} was larger than the comoving Hubble radius today

$$(a_{\text{I}} H_{\text{I}})^{-1} > (a_0 H_0)^{-1}\,.$$

Use this criterion to give an estimate for the number of e-folds the universe has to expand during inflation to solve the horizon problem. To do so, assume that a_{E} with $T_{\text{E}} \sim 10^{15}\text{GeV}$ corresponds to the end of inflation and that the Hubble parameter during inflation was almost constant.

Solution

Since we assume that the universe is dominated by ordinary matter, we know that the strong energy condition holds $1+3w>0$, and

$$(aH)^{-1}=H_0^{-1}a^{\frac{1}{2}(1+3w)}.$$

The particle horizon $\chi_{ph}(a):=\int\limits_{\ln a_i}^{\ln a}(aH)^{-1}d\ln a$ is:

$$\chi_{ph}(a)=\frac{2}{1+3w}(aH)^{-1}.$$

If a_E corresponds to the end of inflation and a_0 to our present day, then we thus have

$$\chi_{ph}\left(a_E\right)=\chi_{ph}\left(a_0\right)\frac{a_0H_0}{a_EH_E}.$$

Next, we suppose that the universe is radiation dominated, and in this case, we know that $H\sim a^{-2}$. Thus, we find that

$$\frac{a_0H_0}{a_EH_E}=\frac{a_E}{a_0}.$$

Using that $T\sim a^{-1}$ we have

$$\frac{a_0H_0}{a_EH_E}=\frac{T_0}{T_E}\sim 10^{-28}.$$

Now, inflation will solve the horizon problem if the observable universe fits inside the Hubble radius before inflation. Denoting by I times before inflation, we thus need

$$\left(a_IH_I\right)^{-1}>\left(a_0H_0\right)^{-1}\sim\left(a_EH_E\right)^{-1}10^{28}.$$

Since we assume the Hubble rate to be constant during inflation, we thus have

$$\frac{a_E}{a_I}\sim 10^{28}.$$

The universe should have expanded at least $\ln 10^{28}\approx 64$ many e-folds.

Leuven-2023hw4.1 Derive $T_{\mu\nu}=\partial_\mu\phi\partial_\nu\phi-\frac{1}{2}g_{\mu\nu}g^{\rho\sigma}\partial_\rho\phi\partial_\sigma\phi-g_{\mu\nu}V(\phi)$ and $P=\frac{1}{2}\left(\frac{d\phi}{dt}\right)^2-V(\phi)$ for a typical model of a scalar field inflation

Solution

The energy-momentum tensor is obtained by variation by metric of the action of the scalar field in a general curved space-time, then pressure - is just its component.

$$S=\int d^4x\sqrt{-g}\left(-\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi-V(\phi)\right)$$

$$T_{\mu\nu}\equiv-\frac{2}{\sqrt{-g}}\frac{\delta S}{\delta g^{\mu\nu}}=-\frac{2}{\sqrt{-g}}\left(\frac{\delta\sqrt{-g}}{\delta g^{\mu\nu}}\left(-\frac{1}{2}g^{\rho\sigma}\partial_\rho\phi\partial_\sigma\phi-V(\phi)\right)-\sqrt{-g}\frac{\delta\left(\frac{1}{2}g^{\rho\sigma}\partial_\rho\phi\partial_\sigma\phi\right)}{\delta g^{\mu\nu}}\right)$$

Let's use $\delta g/\delta g^{\mu\nu}=-gg_{\mu\nu}$, so $\frac{\delta\sqrt{-g}}{\delta g^{\mu\nu}}=-\frac{\delta g/\delta g^{\mu\nu}}{2\sqrt{-g}}=-\sqrt{-g}g_{\mu\nu}/2$ and

$$T_{\mu\nu}=-\frac{2}{\sqrt{-g}}\left(g_{\mu\nu}\frac{\sqrt{-g}}{2}\left(\frac{1}{2}g^{\rho\sigma}\partial_\rho\phi\partial_\sigma\phi+V(\phi)\right)-\frac{\sqrt{-g}}{2}\delta^\rho_\mu\delta^\sigma_\nu\partial_\rho\phi\partial_\sigma\phi\right)$$

So

$$T_{\mu\nu}=\partial_\mu\phi\partial_\nu\phi-\frac{1}{2}g_{\mu\nu}g^{\rho\sigma}\partial_\rho\phi\partial_\sigma\phi-g_{\mu\nu}V(\phi).$$

$$T^\nu{}_\mu=g^{\nu\lambda}\partial_\lambda\phi\partial_\mu\phi-\frac{1}{2}g^\nu{}_\mu g^{\rho\sigma}\partial_\rho\phi\partial_\sigma\phi-g^\nu{}_\mu V(\phi).$$

Then let's substitute $\partial_i\phi=0,(i=1,2,3)$ and $ds^2=-dt^2+a^2(t)d\mathbf{x}^2$

$$\rho\equiv-T^0{}_0=-\left((-1)\partial_0\phi\partial_0\phi-\frac{1}{2}\cdot1\cdot(-1)\cdot\partial_0\phi\partial_0\phi-(+1)V(\phi)\right)=\frac{1}{2}\left(\frac{d\phi}{dt}\right)^2+V(\phi)$$

$$P\equiv\delta^i_jT^j{}_i=0-\frac{1}{2}\cdot1\cdot(-1)\cdot\partial_0\phi\partial_0\phi-(+1)V(\phi)=\frac{1}{2}\left(\frac{d\phi}{dt}\right)^2-V(\phi)$$

Leuven-2023hw4 **4.2 Derive** $\frac{d^2\phi}{dt^2} + 3H\frac{d\phi}{dt} = -V'(\phi)$

Solution

From $\mathcal{L} = \left(-\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right)\sqrt{-g}$ the Euler-Lagrange equations are:

$$-\partial_\mu\left(\sqrt{-g}g^{\mu\nu}\partial_\nu\phi\right) + \sqrt{-g}V'(\phi) = 0$$

which you usually rearrange as:

$$-\frac{1}{\sqrt{-g}}\partial_\mu\left(\sqrt{-g}g^{\mu\nu}\partial_\nu\phi\right) + V'(\phi) = 0$$

If $\phi(t)$, and considering $g \propto -a^6$, we have

$$-\frac{1}{\sqrt{-g}}\frac{d}{dt}(-\sqrt{-g}\dot{\phi}) + V'(\phi) = 0$$
$$\ddot{\phi} + 3H\phi + V'(\phi) = 0$$

where $H = \frac{\dot{a}}{a}$ and dot is derivative wrt ordinary time t.

Leuven-2023hw4 **4.3. Show that** $\epsilon = \frac{1}{16\pi G}\left(\frac{V'}{V}\right)^2, \quad \delta = \epsilon - \frac{1}{8\pi G}\frac{V''}{V}$

Solution

a)

$\epsilon = -\frac{H'}{H^2}$, so we need to find H' and H.

If we substitute $\rho_\phi = \frac{1}{2}\left(\frac{d\phi}{dt}\right)^2 + V(\phi)$ and $P_\phi = \frac{1}{2}\left(\frac{d\phi}{dt}\right)^2 - V(\phi)$ into the second Friedmann equation $H' = -\left(\rho_\phi + P_\phi\right)/\left(2M_{\rm pl}^2\right)$, we get

$$H' = -\frac{1}{2}\frac{\phi'^2}{M_{\rm pl}^2}.$$

Equation of motion is approximately $3H\phi' \approx -V'$, so $\phi' = -V'/(3H)$ and

$$H' = -\frac{1}{18}\frac{V'^2}{M_{\rm pl}^2 H^2}.$$

From Friedmann equation $H^2 = \rho_\phi/\left(3M_{\rm pl}^2\right)$ with $\rho_\phi = \frac{1}{2}\left(\frac{d\phi}{dt}\right)^2 + V(\phi)$ and $\frac{\partial\phi}{\partial t} \ll V$ we get

$$H^2 = \frac{1}{3M_{\rm pl}^2}\left[\frac{1}{2}\phi'^2 + V\right] \approx \frac{V}{3M_{\rm pl}^2}.$$

Finally

$$\epsilon = -\frac{H'}{H^2} = \frac{1}{18}\frac{V'^2}{M_{\rm pl}^2 H^4} = \frac{1}{18}\frac{V'^2}{M_{\rm pl}^2}\frac{9M_{\rm pl}^4}{V^2} = \frac{M_{\rm pl}^2}{2}\frac{V'^2}{V^2} = \frac{1}{16\pi G}\left(\frac{V'}{V}\right)^2.$$

b)

Taking the time-derivative of Klein Gordon equation $3H\dot{\phi} \approx -V'$, we have

$$3\dot{H}\dot{\phi} + 3H\ddot{\phi} = -V''\dot{\phi},$$

$$\ddot{\phi} = \frac{-V''\dot{\phi}}{3H} - \frac{\dot{H}\dot{\phi}}{H}$$

so

$$\delta \equiv -\frac{\ddot{\phi}}{H\dot{\phi}} = -\frac{V''}{3H^2} - \frac{\dot{H}}{H^2} = -M_{\rm pl}^2\frac{V''}{V} + \epsilon = \epsilon - \frac{1}{8\pi G}\frac{V''}{V}.$$

13.1.7 Problems on homogeneous inflation 2

Leuven-2023hw4.4. Derive $S = S\left[\phi^{(0)}\right] + \frac{1}{2}\int d^4x \left(\dot{\chi}^2 - (\nabla\chi)^2 + \frac{\ddot{a}}{a}\chi^2\right)$

Solution

We know that we have equations

$$\ddot{\phi}^{(0)} + 2\frac{\dot{a}}{a}\dot{\phi}^{(0)} = -a^2 V'\left(\phi^{(0)}\right)$$

and

$$\delta\ddot{\phi} + 2\frac{\dot{a}}{a}\delta\dot{\phi} - \nabla^2\delta\phi = -a^2 V''\left(\phi^{(0)}\right)\delta\phi \approx 0$$

So to prove that the action $S = S\left[\phi^{(0)}\right] + \frac{1}{2}\int d^4x \left(\dot{\chi}^2 - (\nabla\chi)^2 + \frac{\ddot{a}}{a}\chi^2\right)$ can give them, we need 1) prove that from it comes the equation of motion $\ddot{\chi} = \nabla^2\chi + \frac{\ddot{a}}{a}\chi$, and then 2) prove that it is the same equation as one for $\delta\phi$.

Preparation:

$$\frac{\partial L}{\partial\chi} = \frac{2}{2}\frac{\ddot{a}}{a}\chi, \qquad \frac{\partial L}{\partial\partial_t\chi} = \frac{2}{2}\dot{\chi} \qquad \frac{\partial L}{\partial\partial_x\chi} = -\frac{2}{2}\partial_x\chi$$

Euler-Lagrangian equations $-\frac{\partial L}{\partial\chi} + \partial_t\frac{\partial L}{\partial\partial_t\chi} + \nabla\frac{\partial L}{\partial\nabla\chi} = 0$ are

$$-\frac{\ddot{a}}{a}\chi + \ddot{\chi} - \nabla^2\chi = 0$$

Then let's imply $\chi = a\delta\phi$. Preparation:

$$\dot{\chi} = \dot{a}\delta\phi + a\delta\dot{\phi}, \qquad \ddot{\chi} = \ddot{a}\delta\phi + 2\dot{a}\delta\dot{\phi} + a\delta\ddot{\phi}, \qquad \nabla^2\chi = a\nabla^2\delta\phi$$

and we get

$$-\frac{\ddot{a}}{a}a\delta\phi + \ddot{a}\delta\phi + 2\dot{a}\delta\dot{\phi} + a\delta\ddot{\phi} - a\nabla^2\delta\phi = 0$$

$$\delta\ddot{\phi} + 2\frac{\dot{a}}{a}\delta\dot{\phi} - \nabla^2\delta\phi = 0$$

So indeed, the action should be as we stated.

Leuven-2023hw4.5. Equivalency of commutators

a)

Let's show that $[\chi(\eta,\mathbf{x}),\chi(\eta,\mathbf{y})] = [\pi(\eta,\mathbf{x}),\pi(\eta,\mathbf{y})] = 0, [\chi(\eta,\mathbf{x}),\pi(\eta,\mathbf{y})] = i\delta(\mathbf{x}-\mathbf{y}) \Longrightarrow [a_\mathbf{k}, a_\mathbf{l}] = 0, \left[a_\mathbf{k}, a_\mathbf{l}^\dagger\right] = \delta(\mathbf{k}-\mathbf{l})$

b)

Let's show that $[a_\mathbf{k}, a_\mathbf{l}] = 0, \left[a_\mathbf{k}, a_\mathbf{l}^\dagger\right] = \delta(\mathbf{k}-\mathbf{l}) \Longrightarrow [\chi(\eta,\mathbf{x}),\chi(\eta,\mathbf{y})] = [\pi(\eta,\mathbf{x}),\pi(\eta,\mathbf{y})] = 0, [\chi(\eta,\mathbf{x}),\pi(\eta,\mathbf{y})] = i\delta(\mathbf{x}-\mathbf{y})$

Leuven-2023hw4.6. Verify that comoving curvature perturbation ζ equals a) during inflation $\zeta = -aH\frac{\delta\phi}{\dot{\phi}^{(0)}}$ and b) after inflation $\zeta = -\frac{3}{2}\Psi$

(here dot is a derivative wrt. conformal time)

Solution to a)

By definition $\zeta \equiv -\Psi + \frac{1}{3}\frac{\delta\rho}{\rho+P}$.
During slow-roll inflation, for modes inside the horizon and those which have just left the horizon, $\Psi \approx 0$.
From $P = \frac{1}{2}\left(\frac{d\phi}{dt}\right)^2 - V(\phi), \rho = \frac{1}{2}\left(\frac{d\phi}{dt}\right)^2 + V(\phi)$ and $\frac{d}{dt} = \frac{1}{a}\frac{d}{d\eta}$, we have $\frac{1}{\rho+P} = \frac{a^2}{\dot{\phi}^{(0)2}}$.
Using $\delta\rho = V'\left(\phi^{(0)}\right)\delta\phi$ and from KG equation $V' \approx 3H\frac{d\phi^{(0)}}{dt}$, where $\frac{d\phi}{dt} = \frac{1}{a}\frac{d\phi}{d\eta} \equiv \frac{\dot{\phi}}{a}$.
Totally:

$$\zeta = 0 + \frac{1}{3}\frac{a^2}{\dot{\phi}^{(0)2}}V'\left(\phi^{(0)}\right)\delta\phi = -\frac{1}{3}\frac{a^2}{\dot{\phi}^{(0)2}}\cdot 3H\frac{\dot{\phi}^{(0)}}{a}\delta\phi = -aH\frac{\delta\phi}{\dot{\phi}^{(0)}}$$

Solution to b)

By definition $\zeta \equiv -\Psi + \frac{1}{3}\frac{\delta\rho}{\rho+P}$. After inflation $P = \rho/3, \delta_\gamma = \frac{\delta\rho}{\rho}$ and $\Psi = -\frac{1}{2}\delta_\gamma$, so

$$\zeta = -\Psi + \frac{1}{3}\frac{\delta\rho}{\rho+P} = -\Psi + \frac{1}{4}\frac{\delta\rho}{\rho} = -\Psi + \frac{1}{4}\delta_\gamma = -\Psi - \frac{\Psi}{2} = -\frac{3}{2}\Psi.$$

(??? need to add a comment, why such relations between deltas and Ψ)

Long solution to b) by Dodelson (???)

Let's follow Dodelson's notation. We know that we can also defint ζ as

$$\zeta \equiv -\frac{ik_i\delta T_i^0 H}{k^2(\rho+\mathcal{P})} - \Psi$$

After inflation ends, $ik_i\delta T_i^0 = 4ak\rho_r\Theta_1$, proportional to the dipole of the radiation. Since $\mathcal{P} = \rho/3$,

$$\zeta = -\frac{3aH\Theta_1}{k} - \Psi$$

It is well known that $\Theta_1 = \mathcal{N}_1 = \frac{iv_b}{3} = \frac{iv}{3} = -\frac{k\Phi}{6aH}$.
The Eq. (5.33) of Dodelson describes how the higher moments of the photon and neutrino distributions cause $\Psi + \Phi$ to be nonzero. Let us here neglect these higher order moments, which cause the sum of the gravitational potentials to be slightly nonzero. Under this approximation, we can eliminate Ψ everywhere by simply setting it to $-\Phi$. Then,

$$\zeta = \frac{3aH}{k}\frac{k\Phi}{6aH} - \Psi = \frac{\Phi}{2} - \Psi = -\frac{3}{2}\Psi$$

Leuven-2023hw4 **4.7. Verify** $4\pi G\left(\dot{\phi}^{(0)}\right)^2 = \epsilon a^2 H^2$

(Here "(0)" means 0-th order, dot means derivative wrt. conformal time)

Solution "in two lines"

Second Friedmann equation $\frac{dH}{dt} = -\frac{(\rho_\phi + P_\phi)}{\left(2M_{\rm pl}^2\right)}$ with $\rho_\phi = \frac{1}{2}\left(\frac{d\phi}{dt}\right)^2 + V(\phi)$ and $P_\phi = \frac{1}{2}\left(\frac{d\phi}{dt}\right)^2 - V(\phi)$ is

$$H_t' = -4\pi G(\phi_t')^2.$$

We just put here $\epsilon := \frac{d}{dt}\left(\frac{1}{H}\right) = -\frac{H_t'}{H^2}$ and $\phi_t' = \frac{1}{a}\frac{d\phi}{d\eta} \equiv \frac{1}{a}\dot{\phi}$, so

$$4\pi G\dot{\phi}^2 = \epsilon a^2 H^2.$$

Solution by rewriting left hand side

$$\epsilon a^2 H^2 = -a^2 H^2 \left(\frac{1}{H^2}\frac{dH}{dt}\right) = -a^2 \frac{d}{dt}\left(\frac{da}{dt}\frac{1}{a}\right) = -a^2 \left(\left(\frac{1}{a}\frac{d^2a}{dt^2} - \frac{da}{dt}\right)^2 \frac{1}{a^2}\right).$$

Insert the Friedmann equations $\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho$; $\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P)$ and get

$$\epsilon a^2 H^2 = 4\pi G a^2 (\rho + P).$$

Scalar field has the property $\rho + P = \left(\frac{d\phi}{dt}\right)^2 = \frac{\dot{\phi}^2}{a^2}$, thus finally

$$\epsilon a^2 H^2 = 4\pi G \dot{\phi_0}^2$$

Leuven-2023hw4 **4.8. Verify $\frac{d\epsilon}{d\eta} = 2aH\epsilon(\epsilon + \delta)$, where $\epsilon \equiv \frac{d}{dt}\left(\frac{1}{H}\right)$, $\delta \equiv \frac{1}{H}\frac{d^2\phi/dt^2}{d\phi/dt}$**

Solution

Second Friedmann equation $\frac{dH}{dt} = -\frac{(\rho_\phi + P_\phi)}{(2M_{\rm pl}^2)}$ with $\rho_\phi = \frac{1}{2}\left(\frac{d\phi}{dt}\right)^2 + V(\phi)$ and $P_\phi = \frac{1}{2}\left(\frac{d\phi}{dt}\right)^2 - V(\phi)$ is

$$H_t' = -4\pi G(\phi_t')^2.$$

We differentiate and use $H_t' \equiv -\varepsilon H^2$:

$$4\pi G \cdot 2\phi_t'\phi_t'' = \frac{d\varepsilon}{a d\eta}H^2 + \varepsilon \cdot 2H\frac{dH}{dt}$$

$$4\pi G(\phi_t')^2 \cdot 2H\frac{\phi_t''}{H\phi_t'} = \frac{d\varepsilon}{a d\eta}H^2 - \varepsilon H^2 \cdot 2\varepsilon H$$

$$\frac{d\varepsilon}{d\eta}\frac{H^2}{a} = H^2\varepsilon \cdot 2H\delta + 2\varepsilon H^2 \cdot \varepsilon H$$

$$\frac{d\varepsilon}{d\eta} = 2\varepsilon a H(\delta + \varepsilon).$$

13.1.8 Problems on Einstein equations

D-1.

Derive the Christoffel symbols, $\Gamma^i{}_{\mu\nu}$, given in Eq. (5.7).
(?? впишу)
When doing this, you will need g^{ij}. Show that it is equal to $\delta_{ij}(1 - 2\Phi)/a^2$.

D-2.

Show that R_{ij} is given by Eq. (5.14).

D-3.

Use the full general relativistic expression for the energy momentum tensor given in Eq. (2.101), which holds even in the presence of metric perturbations. Show that, with scalar perturbations to the metric, the phase space integral for the time-time component reduces to that in Eq. (5.22). Show that the contribution from species α to T_i^0 is

$$T_i^0 = g_\alpha a \int \frac{d^3p}{(2\pi)^3} p_i f_\alpha(\vec{p}, \vec{x}, t)$$

Note the extra factor of a.

D-4.

Consider a 3D matrix with components $G_{ij} = \left(\hat{k}_i\hat{k}_j - \delta_{ij}/3\right) G^L$. Show that this form is traceless and satisfies $\epsilon_{ijk}G_{kl,jl} = 0$ so it is the proper form for the longitudinal component.

D-5.

Compute the time-space component of the Einstein tensor. Show that, in Fourier space,

$$G^0_i = 2ik_i \left(\frac{\dot{\Phi}}{a} - H\Psi\right)$$

Combine with the energy-momentum tensor derived in Exercise 3 to show that

$$\dot{\Phi} - aH\Psi = \frac{4\pi G a^2}{ik} \left[\rho_{\rm dm} v + \rho_b v_{\rm b} - 4i\rho_\gamma \Theta_1 - 4i\rho_\nu \mathcal{N}_1\right]$$

The time-space component of Einstein's equations adds no new information once we already have the two equations derived in the text. Deciding which two to use is a matter of convenience.

D-6.

Take the Newtonian limit of Einstein's equations. Combine the timetime equation (5.27) with the time-space equation of Exercise 5 to obtain the algebraic (i.e., no time derivatives) equation for the potential given in Eq. (5.81). Show that this reduces to Poisson's equation (with the appropriate factors of a) when the wavelength is much smaller than the horizon ($k\eta \gg 1$).

D-7.

Fill in the blanks in the derivation of the tensor equation.
(a) Show that Γ^i_{jk} is given by Eq. (5.43) in the presence of tensor perturbations.
(b) Show that the last term in Eq. (5.48) is given by Eq. (5.54).

D-8.

We defined the perturbations to the photon distribution function via Eq. (4.34). Show that, if Θ depends only on μ, the cosine of the angle between $\hat{k}$ ($:= \hat{z}$ here) and $\hat{p}$, then $T^1_1 - T^2_2$ vanishes. This is indeed the dependence we have been dealing with so far. This is yet another aspect of the decomposition theorem: the terms Θ that source the scalar perturbations (and are sourced by them) do not affect tensor perturbations. Anisotropies induced by tensor perturbations will have Θ of the form

$$\Theta(\mu, \phi) = \left(1 - \mu^2\right) \cos(2\phi)\Theta_+(\mu)$$

for those perturbations generated by $h_+$and a similar expression for $h_\times$with the cos replaced by a sin. These, however, have a negligible impact on the evolution of the gravity waves, so we are justified in setting the right-hand side of Eq. (5.63) to zero.

D-9.

Use the $\frac{1}{2}$ component of the Einstein equations to show that $h_\times$obeys the same equation as does h_+.

D-10.

Show that scalar perturbations (Φ and Ψ) do not contribute to either $G^1{}_1 - G^2_2$ or to G^1_2. This completes the demonstration of the decomposition theorem for scalars and tensors.

D-11.

Consider vector perturbations to the metric. These can be described by two function h_{xz} and h_{yz} where again only the spatial part of the metric is perturbed. The perturbative part of g_{ij} is

$$h^V_{ij} = \begin{pmatrix} 0 & 0 & h_{xz} \\ 0 & 0 & h_{yz} \\ h_{xz} & h_{yz} & 0 \end{pmatrix}$$

Show that h_{xz} and h_{yz} do not affect any of the equations we have derived so far for scalar or tensor evolution - yet another aspect of the decomposition theorem.

D-12.

Solve the wave equation (5.63) if the universe is purely matter dominated. Do the same for the radiation-dominated case.

D-13.

Define the transfer function for gravity wave evolution as

$$T(k) := \frac{h_\alpha(k,\eta)}{h_\alpha(k,\eta=0)} \left(\frac{k\eta}{3j_1(k\eta)} \right).$$

You should recognize the term in parentheses as the inverse of the matter-dominated solution you derived in

D-12.

Solve Eq. (5.63) numerically and compute the transfer function. Compare your solution with the fit of Turner, White, and Lidsey (1993),

$$T(y) = \left[1 + 1.34y + 2.5y^2\right]^{1/2}$$

where $y := (k\eta_0/370h)$ (with h parametrizing the Hubble constant). Assume the universe today is flat and matter dominated, but account for transition from matter to radiation.

D-14.

Derive the transformations in the metric components given by Eq. (5.75). Show that Φ_A and Φ_H do not change under a general coordinate transformation.

13.2 Problems about Inhomogeneities and Anisotropies

13.2.1 Problems about initial conditions

D-1.

Find the ratio of neutrino to radiation energy density, f_ν. Assume that there are three species of massless neutrinos.

D-2.

Account for the neutrino quadrupole moment when setting up initial conditions.
(a) Start with Eq. (4.107). This is an equation for $\mathcal{N}(\mu)$. Turn this into a hierarchy of equations for the neutrino moments:

$$\begin{aligned} \dot{\mathcal{N}}_0 + k\mathcal{N}_1 &= -\dot{\Phi} \\ \dot{\mathcal{N}}_1 - \frac{k}{3}(\mathcal{N}_0 - 2\mathcal{N}_2) &= \frac{k}{3}\Psi \\ \dot{\mathcal{N}}_2 - \frac{2}{5}k\mathcal{N}_1 &= 0. \end{aligned}$$

To do this, you need to recall the definition of these moments, which is equivalent to that for photons, Eq. (4.99). A good way to reduce Eq. (4.107) into this hierarchy is to multiply it first by $\mathcal{P}_0$ and then integrate over $\int_{-1}^{1} d\mu$. This leads to the first equation above. Then multiply Eq. (4.107) by $\mathcal{P}_1$ to get the second and $\mathcal{P}_2$ to get the third. More details are given in Section 8.3, where we go through the same exercise for the photon moments. In the third equation you may neglect $\mathcal{N}_3$ because it is smaller than $\mathcal{N}_2$ by a factor of order $k\eta$ (prove this!).
(b) Eliminate $\mathcal{N}_1$ from these equations and show that

$$\ddot{\mathcal{N}}_2 = \frac{2k^2}{15}(\Psi + \mathcal{N}_0 - 2\mathcal{N}_2)$$

Drop $\mathcal{N}_2$ on the right-hand side because it is much smaller than $\Psi + \mathcal{N}_0$.

(c) Rewrite Einstein's equation (5.33) as

$$\mathcal{N}_2 = -(k\eta)^2 \frac{\Phi + \Psi}{12 f_\nu}$$

This neglects the photon quadrupole. Argue that Compton scattering sets $\Theta_2 \ll \mathcal{N}_2$ so this is a reasonable assumption.

(d) Now differentiate this form of Einstein's equation twice to get an expression for $\ddot{\mathcal{N}}_2$. Equate this to the expression for $\ddot{\mathcal{N}}_2$ derived in part (b). (You may drop all derivatives of Φ and Ψ when doing this since the mode of interest is the $p = 0$ constant mode.) Use this equation to express $\mathcal{N}_0$ in terms of Φ and Ψ.

(e) Finally assume that $\Theta_0 = \mathcal{N}_0$ and use your expression for $\mathcal{N}_0$ to rewrite Eq. (6.12) as a relation between the two gravitational potentials. Show that this relation is

$$\Phi = -\Psi \left(1 + \frac{2 f_\nu}{5}\right)$$

D-3.

Show that the initial conditions for the velocities and dipoles of matter and radiation are as given in Eq. (6.16).

D-4.

Inflation also solves the flatness problem. This is the question of why the energy density today is so close to critical.

(a) Suppose that

$$\Omega(t) := \frac{8\pi G \rho(t)}{3H^2(t)}$$

is equal to 0.3 today, where ρ counts the energy density in matter and radiation (assume zero cosmological constant). From Eq. (1.2), plot $\Omega(t) - 1$ as a function of the scale factor. How close to one would $\Omega(t)$ have been back at the Planck epoch (assuming no inflation took place so that the scale factor at the Planck epoch was of order 10^{-32})? This fine-tuning of the initial conditions is the flatness problem. If not for the fine tuning, an open universe would be obviously open (i.e., Ω would be almost exactly zero) today.

(b) Now show that inflation solve the flatness problem. Extrapolate $\Omega(t) - 1$ back to the end of inflation, and then through 60 e-folds of inflation. What is $\Omega(t) - 1$ right before these 60 e-folds of inflation?

D-5.

Another way of looking at the problems that inflation solves is to consider the entropy within our Hubble volume. This is proportional to the total number of particles in the volume, with a proportionality constant of order unity. How many photons are there within our Hubble volume today? Explain how inflation produces entropy this large.

D-6.

We showed that, if the universe was always dominated by ordinary matter or radiation early on, then the comoving horizon when the scale factor was a_e (very small) was $a_0 H_0 / a_e H_e$ times the comoving Hubble radius today. Compute this ratio assuming that the temperature was equal to 10^{15}GeV at a_e. Account for the radiation-to-matter transition at $a \sim 10^{-4}$.

D-7.

Consider a free, homogeneous scalar field with mass m. The potential for this field is $V = m^2\phi^2/2$. Show that, if $m \gg H$, the scalar field oscillates with frequency equal to its mass. Also show that its energy density falls off as a^{-3}, so it behaves exactly like ordinary nonrelativistic matter.

D-8.

Show that Eq. (6.33) follows from Eq. (6.32) by changing variables from t to η.

D-9. Well-known properties of the quantized harmonic oscillator.

(a) The momentum of the harmonic oscillator with unit mass is $p = dx/dt$. Compute

$$[\hat{x}, \hat{p}]$$

and show that it is equal to i. You can obtain the operator $\hat{p}$ by differentiating $\hat{x}$ (Eq. (6.38)) with respect to time.

(b) Compute the zero-point energy of the harmonic oscillator with unit mass. Do this by quantizing the energy

$$E = \frac{p^2}{2} + \frac{\omega^2 x^2}{2}$$

and then computing its expectation value in the ground state: $\langle 0|\hat{E}|0\rangle$.

D-10.

Show that gravity waves are not sourced by the scalar field during inflation. To do this, recall that the right-hand side of Eq. (6.45) is

$$\delta T_1^1 - \delta T_2^2$$

where δT is the perturbation to the energy-momentum tensor (assumed to be dominated by ϕ) and, as in the derivation of Eq. (5.63), I have chosen $\vec{k}$ to be in the $\hat{z}$ direction. Show that this right-hand side is indeed zero for the scalar field.

D-11.

Show that Eq. (6.57) is the appropriate solution to Eq. (6.56).

(a) Define $\tilde{v} = v/\eta$ and rewrite Eq. (6.56) in terms of $\tilde{v}$.

(b) The resulting equation is the spherical Bessel equation. Write down the general solution to this as a linear combination of two functions of $k\eta$.

(c) Use the boundary conditions of Eq. (6.58) to determine the coefficients of part (b). Show that Eq. (6.57) is the correct solution for these boundary conditions.

D-12.

Derive some useful identities involving the slow-roll parameters during inflation.

(a) Show that

$$\frac{d}{d\eta}\left(\frac{1}{aH}\right) = \epsilon - 1$$

(b) Show that

$$4\pi G\left(\dot{\phi}^{(0)}\right)^2 = \epsilon a^2 H^2$$

(c) Using the definitions of ϵ and δ, show that

$$\frac{d\epsilon}{d\eta} = -2aH\epsilon(\epsilon + \delta)$$

Use this to show that $d\ln\epsilon|_{aH=k}/d\ln(k) = -2(\epsilon + \delta)$.

D-13.

Show that on large scales Eq. (6.85) holds. One way to do this is to combine Einstein's equations, the time-time (5.27) and time-space (Exercise 5 of Chapter 5) components, and take the large-scale limit.

D-14.

Express the slow-roll parameters ϵ and η in terms of the potential V and its derivatives with respect to ϕ. Show that, to lowest order,

$$\epsilon = \frac{1}{16\pi G}\left(\frac{V'}{V}\right)^2$$

and

$$\delta = \epsilon - \frac{1}{8\pi G}\frac{V''}{V}$$

where primes denote derivatives with respect to $\phi^{(0)}$.

D-15.

There are a number of ways of describing pressure in the universe and of relating the pressure to the energy density. One was introduced back in Chapter 2, the equation of state,

$$w := \frac{\mathcal{P}}{\rho}$$

The second is the sound speed,

$$c_s^2 := \frac{d\mathcal{P}}{d\rho}$$

The way to compute c_s^2 is to differentiate both $\mathcal{P}$ and ρ with respect to time and take the ratio. Finally, there is the ratio of perturbations in the energy density to those in the pressure,

$$\frac{\delta\mathcal{P}}{\delta\rho} = \frac{-3\delta T_0^0}{\delta T_i^i}$$

where the minus sign accounts for the fact the the time-time component of the energy-momentum tensor is minus the energy density with our convention, and the factor of 3 negates the sum over the three spatial indices. For adiabatic perturbations, $\delta\mathcal{P}/\delta\rho = c_s^2$. Show that this holds for three separate cases: matter, radiation, and a single scalar field during inflation at the time of horizon crossing. For the last case, it is enough to show that the difference $\delta\mathcal{P}/\delta\rho - c_s^2$ is of order the slow-roll parameters ϵ and δ.

Solution

D-16.

Show that in a gauge given by Eq. (6.91), the equation governing the perturbations to a scalar field $\delta\phi$ is Eq. (6.71).
(a) Bardeen's equation for the gauge-invariant density in the absence of anisotropic stress is

$$\frac{d}{d\eta}\left(a^3\rho\epsilon_m\right) = -(\rho + \mathcal{P})a^3 kv$$

with gauge-invariant density defined as

$$\rho\epsilon_m = -\rho - \delta T_0^0 + \frac{3iH}{k^2}k_i\delta T_i^0$$

and velocity v via Eq. (6.92). Compute $\rho\epsilon_m$ for a scalar field in a gauge with spatially flat slicing. Show that, to lowest order in slow-roll parameters ϵ (not ϵ_m) and δ, the equation reduces to

$$\frac{d}{d\eta}\left(a^3\dot{\phi}^{(0)}\dot{\delta\phi}\right) = -k^2 a\dot{\phi}^{(0)}\delta\phi$$

(b) Again using the slow-roll approximation reduce Eq. (6.119) to the form of Eq. (6.45).

Solution

D-17.

Show that the curvature in conformal Newtonian gauge is equal to $4k^2\Phi/a^2$. To do this, compute the three-dimensional Ricci scalar arising from the spatial part of the metric $g_{ij} = \delta_{ij}a^2(1+2\Phi)$.

Solution

D-18.

Determine the predictions of an inflationary model with a quartic potential,

$$V(\phi) = \lambda\phi^4$$

(a) Compute the slow roll parameters ϵ and δ in terms of ϕ.
(b) Determine ϕ_e, the value of the field at which inflation ends, by setting $\epsilon = 1$ at the end of inflation.
(c) To determine the spectrum, you will need to evaluate ϵ and δ at $-k\eta = 1$. Choose the wavenumber k to be equal to $a_0 H_0$, roughly the horizon today. Show that the requirement $-k\eta = 1$ then corresponds to

$$e^{60} = \int_0^N dN' \frac{e^{N'}}{\left(H\left(N'\right)/H_e\right)}$$

where H_e is the Hubble rate at the end of inflation, and N is defined to be the number of e-folds before the end of inflation:

$$N := \ln\left(\frac{a_e}{a}\right).$$

(d) Take the Hubble rate to be a constant in the above with H/H_e equal to 1. This implies that $N \simeq 60$. Turn this into an expression for ϕ. The simplest way to do this is to note that $N = \int_t^{T_e} dt' H\left(t'\right)$ and assume that H is dominated by potential energy. Show that this mode leaves the horizon when $\phi^2 = 60 m_{\rm Pl}^2/\pi$.
(e) Determine the predicted values of n and n_T.
(f) Estimate the scalar amplitude in terms of λ. As a rough estimate, assume that $k^3 P_\Phi(k)$ for this mode is equal to 10^{-8} (we will find a more precise value when we normalize to large-angle anisotropies in Chapter 8). What value does this imply for λ?
This model illustrates many of the features of contemporary models. In it, (i) the field is of order - even greater than – the Planck scale, but (ii) the energy scale V is much smaller because of (iii) the very small coupling constant.

Solution

Baumann-4.3. Tensors and the Lyth Bound

(a) Show that the tensor-to-scalar ratio predicted by slow-roll inflation is

$$r \equiv \frac{A_t}{A_s} = \frac{8\dot{\phi}^2}{M_{\rm pl}^2 H^2}.$$

(b) Show that the inflaton field travels a "distance" $\Delta\phi \equiv |\phi_E - \phi_\star|$ during (observable) inflation

$$\frac{\Delta\phi}{M_{\rm pl}} = \frac{N_\star}{60}\sqrt{\frac{r}{0.002}},$$

where $N_\star$ is the total number of e-folds between the time when the CMB scales exited the horizon and the end of inflation. [You may assume that $\varepsilon \approx$ const. during inflation] Comment on the implication of this result for observable gravitational waves. [Realistically, we require $r > 0.001$ to have a fighting chance of detecting gravitational waves via CMB polarisation.]
(c) Derive the following relationship between the energy scale of inflation, $V^{1/4}$, and the tensor-to-scalar ratio,

$$V^{1/4} = \left(\frac{3\pi^2}{2} r A_s\right)^{1/4} M_{\rm pl}.$$

Use $A_s = 2.5 \times 10^{-9}$ to determine $V^{1/4}$ for $r = 0.01$. How does that compare to the energy scales probed by the LHC?

Solution

Solution some problems by Dodelson chapter 6

Solution to

5 There are 411 photons per cm^{-3} today; the Hubble volume is $(4\pi/3)\left[3000h^{-1}\mathrm{Mpc}\right]^3 = 3.3\times10^{84}h^{-3}$ cm^3. So the total number of photons is $1.4\times10^{87}h^{-3}$. This number remains roughly constant throughout the matter and radiation eras since the number density scales as T^3, the physical volume as a^3, and the temperature as a^{-1}. So another problem of the classical cosmology is: Why is the entropy of the universe so large?

Inflation solves this problem. At first the solution seems obvious: inflation makes the scale factor grow exponentially fast, thereby increasing the product aT and hence the entropy. In fact, the solution is not quite that simple because during inflation, the exponential expansion is adiabatic: the temperature still falls as a^{-1}. So near the end of inflation the temperature has dropped rapidly enough so that if the entropy was initially of order unity, it remained of order unity.

The production of entropy actually takes place at the end of inflation during the reheating process. Even though the temperature at the end of inflation is extremely small, the energy density (which is almost completely in the scalar field) is not. When the energy in the scalar field transforms into radiation, the temperature of the radiation shoots up from its very low value of T to $\rho^{1/4} \gg T$. Thus, the reheating process is responsible for the large entropy we see today. Another way to say this is to point out that inflation is a very ordered state: the universe supercools while the field is trapped in a false vacuum. The transition to the true vacuum is a transition to the very disordered state of equilibrium.

Solution to

11 (a) With this substitution, the equation becomes

$$\frac{d^2\tilde{v}}{d\eta^2} + \frac{2}{\eta}\frac{d\tilde{v}}{d\eta} + \left(k^2 - \frac{2}{\eta^2}\right)\tilde{v}$$

Defining $x \equiv k\eta$, we see that $\tilde{v}$ satisfies the spherical Bessel equation of order 1 (Eq. (C.13)). (b) The two general solutions are $j_1(x)$ and $y_1(x)$. The general solution is therefore $Aj_1 + By_1$. Writing these out explicitly leads to

$$\begin{aligned} v = \eta\tilde{v} &= \eta\left(A\frac{\sin x - x\cos x}{x^2} - B\frac{\cos x + x\sin x}{x^2}\right) \\ &= \frac{1}{2k^2\eta}\left(e^{ik\eta}[-iA - Ak\eta - B + iBk\eta]\right. \\ &\quad \left. + e^{-ik\eta}[iA - Ak\eta - B - iBk\eta]\right). \end{aligned}$$

When $k\eta$ is very large and negative, we want $v \to e^{-ik\eta}/\sqrt{2k}$, so the coefficient of $e^{+ik\eta}$ in this limit, proportional to $-A + iB$, must vanish. Thus, $A = iB$. The coefficient of $e^{-ik\eta}$ is

$$\frac{1}{2k^2\eta}[-2Ak\eta] = \frac{-A}{k}$$

This must equal $(2k)^{-1/2}$, so $A = -(k/2)^{1/2}$. Therefore the correct solution is

$$v = \frac{-1}{\sqrt{2k}k\eta}\left(e^{-ik\eta}[i - k\eta]\right)$$

in agreement with Eq. (6.57).

Solution to

13 The two components of Einstein's equations are

$$\begin{aligned} k^2\Psi + 3aH(\dot{\Psi} + aH\Psi) &= 4\pi Ga^2\delta T_0^0 \\ ik_i(\dot{\Psi} + aH\Psi) &= -4\pi Ga\delta T_i^0 \end{aligned}$$

Here I have simply copied the results from Chapter 5, replacing Φ with $-\Psi$. Multiply the second of these by $3iaHk_i/k^2$, and then add the two equations to get

$$k^2\Psi = 4\pi Ga^2\left[\delta T_0^0 - \frac{3Hk_i\delta T_i^0}{k^2}\right]$$

On large scales, the left-hand side is negligible, so the terms in brackets on the right must sum to zero, giving Eq. (6.85).

13.2.2 Problems on inhomogeneities

D-1.

Derive Eqs. (7.11) and (7.12).
(a) First neglect the scattering term in Eq. (4.100), the one proportional to $\dot{\tau}$. Then the photon evolution equation is identical to the neutrino evolution equation (4.107). Show that this collisionless equation reduces to the two equations for the monopole and dipole. To get the monopole equation, multiply Eq. (4.107) by $(d\mu/2)\mathcal{P}_0(\mu) = d\mu/2$ and integrate from $\mu = -1$ to 1. To get the dipole, multiply by $(d\mu/2)\mathcal{P}_1(\mu)$ and integrate.
(b) Show that, in the limit of small baryon density, the scattering term in Eq. (4.100) can indeed be neglected. Neglect Π, since the quadrupole and polarization are very small. Then show that the scattering term is proportional to $R, 3/4$ times the baryon-to-photon ratio. You will want to use Eq. (4.106). It cannot be emphasized enough that this series of approximations is valid only for the purposes of this chapter, wherein we are interested in the matter distribution.

D-3.

The four subsections in Sections 7.2 and 7.3 correspond to four different approximations to the full set of Einstein-Boltzmann equations. In the following table, fill in the regime of validity for each approximation:

	$a \ll a_{\rm eq}$	$a \sim a_{\rm eq}$	$a \gg a_{\rm eq}$
$k\eta \ll 1$			
$k\eta \sim 1$			
$k\eta \gg 1$			

For example, the super-horizon solution of Section 7.2.1 is valid along the whole top row, since it sets $k\eta \to 0$. Note that time evolves from upper left to bottom right, so the fact that none of the approximations work in the center square means that only those scales that enter the horizon well before or well after equality will be subject to analytic techniques.

D-4.

Fill in some of the algebraic detail left out of Section 7.2.1.
(a) Show that Eq. (7.24) leads to Eq. (7.25) by carrying out the differentiation.
(b) Show that Eq. (7.25) is equivalent to Eq. (7.27) when the definition of u from Eq. (7.26) is used.
(c) Show that the integral in Eq. (7.31) can be done analytically with the result given in Eq. (7.32). One way to do the integral is to define a dummy variable $x := \sqrt{1+y}$

D-5.

Find the wavenumber of the mode which equals the inverse comoving Hubble radius at equality. That is, define $k_{\rm eq}$ to be equal to $a_{\rm eq}H\,(a_{\rm eq})$. Show that this definition implies

$$k_{\rm eq} = \sqrt{\frac{2\Omega_m H_0^2}{a_{\rm eq}}}$$

Then use Eq. (2.87) to show that $k_{\rm eq}$ is given by Eq. (7.39). Show that if you define $k_{\rm eq}$ by setting it to $1/\eta_{\rm eq}$, you get a number 17% lower.

D-7.

Show that $\delta \propto H$ is a solution to the evolution equation (7.73) if the universe is flat with a cosmological constant. You will need to use Eq. (1.2). Show also that the solution is valid if the universe has zero cosmological constant, but is open with $\Omega_m < 1$.

D-8.

Derive the growth factor for an open universe with $\Omega_m < 1$:

$$D_1\,(a, \Omega_m) = \frac{5\Omega_m}{2\,(1-\Omega_m)}\left[3\frac{\sqrt{1+x}}{x^{3/2}}\ln(\sqrt{1+x}-\sqrt{x}) + 1 + \frac{3}{x}\right]$$

where $x := (1-\Omega_m)\,a/\Omega_m$. There may be easier ways to do this (e.g., you might want to check The Large Scale Structure of the Universe, Section 11), but I found it easiest to define a dummy variable $y := \Omega_m/a$; write the integral of Eq. (7.77) as

$$\int_{\Omega_m/a}^{\infty} \frac{dy}{y^2\,(y+1-\Omega_m)^{3/2}} = 2\left[\frac{d}{d\epsilon}\frac{d}{d\lambda}\int_{\Omega_m/a}^{\infty}\frac{dy}{(y+\epsilon)\sqrt{y+\lambda}}\right]_{\epsilon=0,\lambda=1-\Omega_m};$$

and then use 2.246 from Gradshteyn and Ryzhik.

D-9.

One popular way to characterize power on a particular scale is to compute the expected RMS overdensity in a sphere of radius R,

$$\sigma_R^2 := \left\langle \delta_R^2(x)\right\rangle$$

Here

$$\delta_R(\vec{x}) := \int d^3x'\delta\left(\vec{x}'\right) W_R\left(\vec{x}-\vec{x}'\right)$$

where $W_R(x)$ is the tophat window function, equal to 1 for $x < R$ and 0 otherwise; the angular brackets denote the average over all space.

(a) By Fourier transforming, express σ_R in terms of an integral over the power spectrum.

(b) Use the BBKS transfer function to compute $\sigma_8\left(R = 8h^{-1}\mathrm{Mpc}\right)$ for a standard CDM model ($h = 0.5, n = 1, \Omega_m = 1$). We will see in Chapter 8 that COBE normalization for this model is

$$\delta_H = 1.9\times 10^{-5}.$$

The value of σ_8 you find is yet another sign of the sickness of the model. For galaxies, σ_8 is known to be unity (or less, depending on galaxy type). A model with $\sigma_8 > 1$ then requires galaxies to be less clustered than the dark matter. Present models of galaxy formation suggest that this is unlikely. There are even direct measures of σ_8 of the mass (e.g., Section 9.5); these too constrain σ_8 to be less than one. (c) In the same model, plot σ_R as a function of R. Since σ_R monotically increases, small scales tend to go nonlinear before large scales, the signature of a hierarchical model.

D-10.

Rewrite σ from Exercise 9 as

$$\sigma_R^2 = \int_0^{\infty}\frac{dk}{k}\Delta^2(k)\tilde{W}_R^2(k),$$

where $\tilde{W}_R$ is the Fourier tranform of the tophat window function and $\Delta^2 = d\sigma^2/d\ln(k)$ is the contribution to the variance per $\ln(k)$. A useful transition point is the value of k at which Δ exceeds 1. Scales larger than this are linear, while smaller scales have gone nonlinear. Find $k_{\rm nl}$ defined in this way for the sCDM model described in

D-9.

D-11.

Compute the growth factors in a universe with $\Omega_{\rm de} = 0.7, \Omega_m = 0.3$, and $w = -0.5$. Plot as a function of a. Compare with the cosmological constant model ($w = -1$) with the same $\Omega_{\rm de}, \Omega_m$.

Leuven-2023hw3.1. Gauge-invariant variables $(4p)$

Show that the Bardeen potentials Φ_A and Φ_H defined as

$$\begin{cases}\Phi_A = A + \frac{1}{a}\frac{\partial}{\partial\eta}(a(\dot{E}-B))\\ \Phi_H = -\psi + aH(B-\dot{E})\end{cases}$$

(dot is $\frac{d}{d\eta}$) in a notation $g_{00} \equiv -(1+2A), g_{0i} \equiv -aB_{.i}, g_{ij} \equiv a^2\left(\delta_{ij}[1+2\psi] - 2E_{.ij}\right)$ are invariant under infinitesimal coordinate transformations of the form

$$t \to t + \zeta(t,\vec{x}), \quad x^i \to x^i + \partial^i\xi(t,\vec{x}),$$

Solution

We consider a coordinate transformation given by

$$\begin{cases} \tilde{t} = t + \xi^0(t, \vec{x}) \\ \tilde{x^i} = x^i + \delta^{ij}\partial_j \xi(t, \vec{x}) \end{cases}$$

where ξ and ξ^0 are small perturbations. Under such a transformation, the scalars transform as

$$\begin{cases} \tilde{A} = A - \frac{1}{a}\dot{\xi}, \\ \tilde{\psi} = \psi - H\xi^0, \\ \tilde{B} = B - \frac{\xi^0}{a} + \dot{\xi}, \\ \tilde{E} = E + \xi, \end{cases}$$

as indicated on p. 133-134 of Dodelson. Filling these expressions into the expressions for the Bardeen potentials, we easily calculate

$$\begin{aligned} \tilde{\Phi}_A &= A - \frac{1}{a}\dot{\xi}^0 + \frac{1}{a}\frac{\partial}{\partial\eta}\left[a\left(\dot{E} + \dot{\xi} - B + \frac{\xi^0}{a} - \dot{\xi}\right)\right] \\ &= A - \frac{1}{a}\dot{\xi}^0 + \frac{1}{a}\frac{\partial}{\partial\eta}\left[a(\dot{E} - B) + \xi^0\right] \\ &= A + \frac{1}{a}\frac{\partial}{\partial\eta}(a(\dot{E} - B)) \\ &= \Phi_A \end{aligned}$$

and

$$\begin{aligned} \tilde{\Phi}_H &= -\psi + H\xi^0 + aH\left(B - \frac{\xi^0}{a} + \dot{\xi} - \dot{E} - \dot{\xi}\right) \\ &= -\psi + aH(B - \dot{E}) \\ &= \Phi_H. \end{aligned}$$

This proves the invariance of the Bardeen potentials.

Leuven-2023hw3 3.2. Geodesic equation for photons (4p)

Consider the perturbed FLRW metric in the Newtonian gauge

$$\mathrm{d}s^2 = -(1 + 2\Psi)\mathrm{d}t^2 + a^2(t)(1 + 2\Phi)\delta_{ij}\ \mathrm{d}x^i\ \mathrm{d}x^j.$$

Use the zeroth component of the geodesic equation for massless particles to derive the equation for the time variation of the magnitude of the physical momentum p as a function of the perturbations in the metric

$$\frac{\mathrm{d}}{\mathrm{d}t}\ln(pa) = -\partial_t\Phi - \frac{\hat{p}^i}{a}\partial_i\Psi,$$

where $p = \sqrt{g_{ij}P^iP^j}$, P^i is the space component of the four-momentum $P^\mu = \mathrm{d}x^\mu/\mathrm{d}\lambda$, λ is the affine parameter of the geodesic, and $\hat{p}^i$ is the unit vector (i.e. $\delta_{ij}\hat{p}^i\hat{p}^j = 1$) determining the direction of the momentum.

Solution

We start from the line element

$$ds^2 = -(1 + 2\Psi)dt^2 + a^2(1 + 2\Phi)\delta_{ij}dx^idx^j.$$

Because the photons are constrained by $P^2 = 0$, we find that

$$P^0 = \frac{p}{\sqrt{1 + 2\Psi}} \sim p(1 - \Psi),$$

to first order. The first component of the geodesic equation is

$$\frac{dP^0}{d\lambda} = -\Gamma^0_{\alpha\beta}P^\alpha P^\beta.$$

Since $\frac{dt}{d\lambda} = P^0$, we thus have

$$p(1-\Psi)\frac{d}{dt}[p(1-\Psi)] = -\Gamma^0_{\alpha\beta}P^\alpha P^\beta,$$

or to first order

$$\frac{d}{dt}[p(1-\Psi)] = -\Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p}(1+\Psi).$$

Expanding the derivative on the left hand side and multiplying by $(1+\Psi)$, we find

$$\frac{dp}{dt} = p\left[\frac{\partial\psi}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}\right] - \Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p}(1+2\Psi).$$

Dodelson calculates on p. 91-92 that

$$\Gamma^0_{\alpha\beta}\frac{P^\alpha P^\beta}{p} = [-1+2\Psi]\left[-\frac{\partial\Psi}{\partial t} - 2\frac{\partial\Psi}{\partial x^i}\frac{\hat{p}^i}{a} - p\left(\frac{\partial\Phi}{\partial t} + H\right)\right].$$

Inserting this expression and dividing by p, we arrive at

$$\frac{1}{p}\frac{dp}{dt} = -H - \frac{\partial\Phi}{\partial t} - \frac{\hat{p}^i}{a}\frac{\partial\Psi}{\partial x^i}.$$

The left hand side is just the derivative of $\ln p$. Furthermore from the definition of H, it is also clear that this is the derivative of $\ln a$ with respect to time. Hence, we indeed arrive at

$$\frac{d\ln(pa)}{dt} = -\partial_t\Phi - \frac{\hat{p}^i}{a}\partial_i\Psi$$

Leuven-2023hw3 3.3. Large-scale super-horizon potential modes

In this exercise we consider the evolution of the gravitational potential Φ for largescale super-horizon modes.
a) Show that the set of evolution equations for large-scale perturbations ($k\eta \ll 1$)

$$\dot{\Theta}_0 = -\dot{\Phi}, \quad \dot{\delta} = -3\dot{\Phi}, \quad \frac{3\dot{a}}{a}\left(\dot{\Phi} + \Phi\frac{\dot{a}}{a}\right) = 4\pi G a^2\left(\rho_{\rm m}\delta + 4\rho_{\rm r}\Theta_0\right)$$

can be combined into a single second-order equation of the following form

$$\Phi'' + \frac{21y^2+54y+32}{2y(y+1)(3y+4)}\Phi' + \frac{\Phi}{y(y+1)(3y+4)} = 0$$

by using the variable

$$y := \frac{a}{a_{\rm eq}} = \frac{\rho_{\rm m}}{\rho_{\rm r}},$$

where a dot indicates a derivative with respect to η and a prime indicates a derivative with respect to y.
b) Introduce the variable

$$u := \frac{y^3}{\sqrt{1+y}}\Phi$$

and show that (1) can be written as

$$u'' + u'\left[-\frac{2}{y} + \frac{3}{2(1+y)} - \frac{3}{3y+4}\right] = 0.$$

c) Show by means of (3) that a solution to (1) can be expressed as

$$\frac{y^3}{\sqrt{1+y}}\Phi = A\int_0^y \mathrm{d}\tilde{y}\frac{\tilde{y}^2(3\tilde{y}+4)}{(1+\tilde{y})^{3/2}}.$$

Show that this integral can be computed analytically and gives

$$\Phi = \frac{2A}{15}\frac{1}{y^3}\left(16\sqrt{1+y} + 9y^3 + 2y^2 - 8y - 16\right).$$

Hint: Use the dummy variable $x = \sqrt{1+y}$.
d) Show that (4) asymptotes to a constant for $y \to 0$ and $y \to \infty$. What is the ratio of Φ at late times to Φ at early times?

Solution

In this exercise we consider the evolution of the gravitational potentials Φ for large-scale superhorizon modes and fill some gaps that were left in the lecture a) Consider the set of evolution equations for large-scale perturbations

$$\begin{aligned}\dot{\Theta}_0 &= -\dot{\Phi},\\ \dot{\delta} &= -3\dot{\Phi},\\ \frac{3\dot{a}}{a}\left(\dot{\Phi}-\Phi\frac{\dot{a}}{a}\right) &= 4\pi G a^2\left(\rho_{\rm m}\delta+4\rho_{\rm r}\Theta_0\right),\end{aligned}$$

where a dot indicates a derivative with respect to η. First of all, consider the variable $y\equiv\frac{a}{a_{\rm eq}}=\frac{\rho_{\rm m}}{\rho_{\rm x}}$ as defined in the assignment. A prime will indicate a derivative with respect to y. Let us first rewrite the equations in (1-3) in the variable y. For this purpose, note that by chain rule we have

$$\frac{\mathrm{d}}{\mathrm{d}\eta}=\frac{\mathrm{d}y}{\mathrm{d}\eta}\frac{\mathrm{d}}{\mathrm{d}y}=\frac{\dot{a}}{a_{\rm eq}}\frac{\mathrm{d}}{\mathrm{d}y}=aHy\frac{\mathrm{d}}{\mathrm{d}y}$$

where we used that $H=\frac{da/dt}{a}=\frac{\dot{a}}{a^2}$. Now using this the left-hand side of the third evolution equation (3) gives

$$\frac{3\dot{a}}{a}\left(\dot{\Phi}+\Phi\frac{\dot{a}}{a}\right)=3Hya'\left(aHy\Phi'+\Phi Hya'\right)=3H^2y^2\left(a'\Phi'+\left(a'\right)^2\Phi\right)$$

Next combining (1) and (2) yields

$$\dot{\delta}-3\dot{\Theta}_0=0\Rightarrow\delta-3\Theta_0=c$$

where c is an constant. However, using the adiabatic initial conditions, we have that $c=0$. If we now use (6) in the third evolution equation (3) we have

$$3H^2y^2\left(a'\Phi'+\left(a'\right)^2\Phi\right)=4\pi Ga^2\left(\rho_{\rm m}\delta+4\rho_{\rm r}\Theta_0\right)=4\pi Ga^2\rho_m\delta\left(1+\frac{4}{3y}\right).$$

Next, rewrite the first Friedmann equation as

$$\frac{8\pi G\rho_m}{3}=\frac{8\pi Gp}{3}\frac{y}{y+1}=H^2\frac{y}{y+1}$$

where we used that $\rho=\rho_m+\rho_r$ and the definition for the variable y. Substituting this in the right-hand side expression from (7) yields

$$3H^2y^2\left(a'\Phi'+\left(a'\right)^2\Phi\right)=\frac{3H^2}{2}\frac{y}{y+1}a^2\delta\left(1+\frac{4}{3y}\right)$$

Rewriting this gives

$$\frac{a'}{a}y\Phi'+\left(\frac{a'}{a}\right)^2y\Phi=\frac{\delta}{2(y+1)}\left(1+\frac{4}{3y}\right)$$

Now by using the definition of the variable y and setting δ to one side yields

$$\delta=\frac{6(y+1)}{3y+4}\left(y\Phi'+\Phi\right)$$

In the second evolution equation (2) we can interchange the η-derivative with y-derivative so that the above intermediate expression (11) becomes

$$-3\Phi'-\frac{\mathrm{d}}{\mathrm{d}y}\left[\frac{6(y+1)}{3y+4}\left(y\Phi'+\Phi\right)\right]=\frac{6(3y+4)-18(y+1)}{(3y+4)^2}\left(y\Phi'+\Phi\right)+\frac{6(y+1)}{3y+4}\left(2\Phi'+y\Phi''\right)$$

which by some simple algebraic manipulation reduces to the requested equation

$$\Phi''+\frac{21y^2+54y+32}{2y(3y+4)(y+1)}\Phi'+\frac{1}{y(3y+4)(y+1)}\Phi=0$$

Next, introduce the new variable

$$u \equiv \frac{y^3}{\sqrt{1+y}}\Phi$$

The derivatives with respect to the variable y of above expression yield

$$\Phi' = -\frac{5y+6}{2\sqrt{1+y}y^4}u + \frac{\sqrt{1+y}}{y^3}u'$$
$$\Phi'' = \frac{-10(1+y)y+(5y+6)(5y+8)}{4(1+y)y^5\sqrt{1+y}}u - \frac{5y+6}{y^4\sqrt{1+y}}u' + \frac{\sqrt{1+y}}{y^3}u''$$

Now, if we substitute these derivatives into (13) and collect all terms proportional to u one finds

$$\frac{-10(1+y)y+(5y+6)(5y+8)}{4(1+y)y^5\sqrt{1+y}} - \frac{5y+6}{2\sqrt{1+y}y^4}\frac{21y^2+54y+32}{2y(y+1)(3y+4)} + \frac{1}{y(3y+4)(y+1)}\frac{\sqrt{1+y}}{y^3}$$

Letting Mathematica come to the rescue yields that these vanishes. The remaining expression to reduce is

$$\frac{\sqrt{1+y}}{y^3}u'' + \left[\frac{\left(21y^2+54y+32\right)}{2y^4(y+1)(3y+4)} - \frac{2(5y+6)(3y+4)}{2y^4(y+2)(3y+4)}\right]\sqrt{1+y}u' = 0$$

Further reducing the denominator yields

$$u'' - \frac{\left(9y^2+22y+16\right)}{2y(y+1)(3y+4)}u' = 0$$

which again by invoking Mathematica yields

$$u'' + u'\left[-\frac{2}{y} + \frac{3}{2(1+y)} - \frac{3}{3y+4}\right]$$

c) Note that equation (19) is an first order differential equation in u' which can be expressed as

$$\frac{du'}{u'} = dy\left[\frac{2}{y} - \frac{3/2}{1+y} + \frac{3}{3y+4}\right]$$

Integrating both sides yields

$$\ln(u') = C + 2\ln(y) - (3/2)\ln(1+y) + \ln(3y+4),$$

where C is an constant. If we now exponentiate both side we have

$$u' = A\frac{y^2(3y+4)}{(1+y)^{3/2}}$$

where $A = \exp(C)$. Again integrating both sides and using the definition of u (14)

$$\frac{y^3}{\sqrt{1+y}}\Phi = A\int_0^y dy'\frac{y'^2(3y'+4)}{(1+y')^{3/2}}$$

As explained in Dodelson there should be another constant here, arising from the integration. However, because $y^3\Phi \to 0$ early on, one can take it to be zero. Let us now perform the integral. Given as hint in the assignment introduce the variable $x = \sqrt{1+y}$ which gives

$$\int_0^y dy'\frac{y'^2(3y'+4)}{(1+y')^{3/2}} = 2\int_1^{\sqrt{1+y}}\frac{\left(x^2-1\right)^2\left(3x^2+1\right)}{x^2}\,\mathrm{d}x$$
$$= \int_1^{\sqrt{1+y}}\left(6x^4-10x^2+2+\frac{2}{x^2}\right)\mathrm{d}x$$
$$= \frac{6}{5}(1+y)^{5/2} - \frac{10}{3}(1+y)^{3/2} + 2(1+y)^{1/2} - \frac{2}{(1+y)^{1/2}} - \frac{32}{15}.$$

Substituting this back into (23) yiels

$$\begin{aligned}\Phi &= A\frac{\sqrt{1+y}}{y^3}\left[\frac{6}{5}(1+y)^{5/2}-\frac{10}{3}(1+y)^{3/2}+2(1+y)^{1/2}-\frac{2}{(1+y)^{1/2}}-\frac{32}{15}\right]\\ &= \frac{2A}{15}\frac{1}{y^3}\left[9\left(y^3+3y^2+3y+1\right)-25\left(1+2y+y^2\right)+15(1+y)-16\sqrt{1+y}\right]\\ &= \frac{2A}{15}\frac{1}{y^3}\left[16\sqrt{1+y}+9y^3+2y^2-8y-16\right].\end{aligned}$$

d) We must calculate the limit

$$\lim_{y\to 0}\Phi(y)=\lim_{y\to 0}\frac{2A}{15}\left(\frac{16\sqrt{1+y}+9y^3+2y^2-8y-16}{y^3}\right)$$

Looking at the dominant behavior for small y, we notice that the limit is of the form $\frac{0}{0}$. Luckily, we can use l'Hopital rule [1], as both nominator and denominator are differentiable. This gives the limit

$$\lim_{y\to 0}\frac{2A}{15}\left(\frac{\frac{8}{\sqrt{1+y}}+27y^2+4y-8}{3y^2}\right)$$

Unfortunately, the limit again yields $\frac{0}{0}$. Nevertheless, using l'Hopital rule two more times yields

$$\lim_{y\to 0}\Phi(y)=\frac{2A}{15}\lim_{y\to 0}\frac{6(1+y)^{-5/2}+54}{6}=\frac{4A}{3}$$

the same result that can be found in Dodelson.

Leuven-2023hw3.4. Hubble radius and equality scale

(a) Find the wavenumber of the mode that equals the inverse comoving Hubble radius at matter-radiation equality, i.e. define $k_{\rm eq}:=a_{\rm eq}H\left(a_{\rm eq}\right)$. Show that this definition implies

$$k_{\rm eq}=\sqrt{\frac{2\Omega_m H_0^2}{a_{\rm eq}}}.$$

(b) Now define a_H, the scale factor at which wavelength k equals the comoving Hubble radius, via $a_H H\left(a_H\right)=k$. Express a_H/a_{eq} only in terms of k and k_{eq}. Show that in the limit $k\gg k_{\rm eq}$, this expression reduces to

$$\frac{a_H}{a_{eq}}\overset{k\gg k_{eq}}{\sim}\frac{k_{eq}}{\sqrt{2}k}.$$

Solution to (a)

Let us try to find the wavenumber of the mode that equals the inverse comoving Hubble radius at matter-radiation equality, i.e. $k_{\rm eq}\equiv a_{\rm eq}H\left(a_{\rm eq}\right)$. At matter-radiation equality we have $\rho_m=\rho_r$ or equivalently $\frac{\Omega_m}{a_{eq}}=\frac{\Omega_\gamma}{a_{cq}}$. Hence, by the first Friedmann equation

$$H^2\left(a_{\rm eq}\right)=H_0^2\left(\frac{\Omega_r}{a_{\rm eq}^4}+\frac{\Omega_m}{a_{\rm eq}^3}\right)=\frac{2\Omega_m H_0^2}{a_{\rm eq}^3}$$

Therefore, we have

$$k_{eq}=\sqrt{\frac{2\Omega_m\left(\frac{H_o}{c}\right)^2}{a_{eq}}}.$$

where we have reinstated the factor for the speed of light c to have the right units. Remember that $a_{eq}=4.15\times10^{-5}\left(\Omega_m h^2\right)^{-1}$ and $H_0=100$ km skpc^{-1} so that the wavenumber becomes

$$k_{\rm eq}=\sqrt{\frac{2(100/299792.458)^2}{4.15\times10^{-5}}}{\rm Mpc}^{-1}\Omega_m h^2=0.073{\rm Mpc}^{-1}\Omega_m h^2.$$

Suppose that we had defined $k_{eq}=\eta_{eq}^{-1}$. Remember that in first homework we found

$$\eta_{\rm eq}=2(\sqrt{2}-1)H_0^{-1}\Omega_m^{-1/2}a_{\rm eq}^{1/2}$$

and indeed when reinstating the speed of light c we find

$$k_{\rm eq}=\frac{H_0\Omega_m^{1/2}}{2(\sqrt{2}-1)a_{\rm eq}^{1/2}}=\frac{100/299792.458}{2(\sqrt{2}-1)\left(4.15\times10^{-5}\right)^{1/2}}{\rm Mpc}^{-1}\Omega_m h^2=0.0625{\rm Mpc}^{-1}\Omega_m h^2$$

Solution to (b)

Let us define a_H the scale factor at which wavelength k equals the comoving Hubble radius, via $k = a_H H(a_H)$. Furthermore, the first Friedmann equation gives

$$H^2(a_H) = H_0^2\left(\frac{\Omega_m}{a_H^3} + \frac{\Omega_r}{a_H^4}\right).$$

and therefore

$$k = H_0\sqrt{\frac{\Omega_m}{a_H} + \frac{\Omega_r}{a_H^2}}$$

Using (30) we can write

$$\frac{k^2}{k_{eq}^2} = \frac{\frac{\Omega_m}{a_H} + \frac{\Omega_r}{a_H^2}}{\frac{2\Omega_m H_0^2}{a_{eq}}} = \frac{a_{eq}}{2}\frac{(a_H\Omega_m + \Omega_r)}{a_H^2 H_0^2\Omega_m} = \frac{a_{eq}a_H + a_{eq}^2}{a_H^2} = \frac{1}{2}\left(\frac{a_{eq}}{a_H}\right)^2 + \frac{1}{2}\frac{a_{eq}}{a_H}$$

where we used that $\Omega_r = \Omega_m a_{eq}$. The expression above is quadratic with a positive solution given by

$$\frac{a_{eq}}{a_H} = \frac{1}{2}\sqrt{1 + 8\frac{k^2}{k_{eq}^2}} - 1$$

In the limit

$$\lim_{k\gg k_{eq}}\frac{a_H}{a_{eq}} = \lim_{k\gg k_{eq}}\frac{1}{2}\sqrt{1 + 8\frac{k^2}{k_{eq}^2}} - 1 = \frac{1}{2}\sqrt{8\frac{k^2}{k_{eq}^2}} = \frac{k_{eq}}{\sqrt{2}k}$$

D-6.

Define a_H, the scale factor at which wavelength k equals the comoving Hubble radius, via $a_H H(a_H) := k$. Express $a_H/a_{\rm eq}$ in terms of k and $k_{\rm eq}$. Show that in the limit $k \gg k_{\rm eq}$, this expression reduces to

$$\lim_{k\gg k_{\rm eq}}\frac{a_H}{a_{\rm eq}} = \frac{k_{\rm eq}}{\sqrt{2}k}$$

Baumann 3.4. Growth of Matter Perturbations I: Early Times and Meszarous equation

At early times, the universe was dominated by radiation (r) and pressureless matter (m). You may ignore baryons.

(a) Show that the conformal Hubble parameter satisfies

$$\mathcal{H}^2 = \frac{\mathcal{H}_0^2\Omega_m^2}{\Omega_r}\left(\frac{1}{y} + \frac{1}{y^2}\right),$$

where $y \equiv a/a_{\rm eq}$ is the ratio of the scale factor to its value when the energy density of the matter and radiation are equal.

(b) Describe qualitatively the behaviour of the Newtonian-gauge fractional density perturbations δ_r and δ_m for a scalar perturbation at scale k, with adiabatic initial conditions, that re-enters the Hubble radius well before $a_{\rm eq}$.

(c) For perturbations on scales much smaller than the Hubble radius, the fluctuations in the radiation can be neglected. Assuming that Φ evolves on a Hubble timescale, show that

$$\delta_m'' + \mathcal{H}\delta_m' - 4\pi G a^2\bar{\rho}_m\delta_m \approx 0.$$

(d) Show that, in terms of the variable y, eq. ($\star$) becomes

$$\frac{d^2\delta_m}{dy^2} + \frac{2+3y}{2y(1+y)}\frac{d\delta_m}{dy} - \frac{3}{2y(1+y)}\delta_m = 0.$$

Hence verify that the solutions are

$$\delta_m \propto \begin{cases} 2+3y \\ (2+3y)\ln\left(\frac{\sqrt{1+y}+1}{\sqrt{1+y}-1}\right) - 6\sqrt{1+y} \end{cases}$$

Determine how δ_m grows with y for $y \ll 1$ (RD) and $y \gg 1$ (MD).

(a) Friedmann Equation

The conformal-time Friedmann equation is

$$\mathcal{H}^2 = \frac{8\pi G}{3} a^2 \left(\rho_m + \rho_r\right)$$

Defining $y = a/a_{\rm eq}$, we have

$$\rho_m = \rho_{\rm eq} y^{-3}$$
$$\rho_r = \rho_{\rm eq} y^{-4}$$

so that the Friedmann equation is

$$\begin{aligned}\mathcal{H}^2 &= \frac{8\pi G}{3} a^2 \rho_{\rm eq} \left(y^{-3} + y^{-4}\right) \\ &= \frac{8\pi G}{3} a_{\rm eq}^2 \rho_{\rm eq} \left(y^{-1} + y^{-2}\right).\end{aligned}$$

As usual, we can define

$$\Omega_m = \frac{8\pi G}{3\mathcal{H}_0^2}\rho_{m,0},$$
$$\Omega_r = \frac{8\pi G}{3\mathcal{H}_0^2}\rho_{r,0},$$

where $a = 1$ today. We can write these as

$$\Omega_m = \frac{8\pi G}{3\mathcal{H}_0^2}\rho_{\rm eq} a_{\rm eq}^3,$$
$$\Omega_r = \frac{8\pi G}{3\mathcal{H}_0^2}\rho_{\rm eq} a_{\rm eq}^4.$$

Their ratio is

$$\frac{\Omega_m}{\Omega_r} = \frac{1}{a_{\rm eq}}.$$

Putting these together, we obtain

$$\mathcal{H}^2 = \mathcal{H}_0^2 \frac{\Omega_m^2}{\Omega_r}\left(y^{-1} + y^{-2}\right).$$

(b) Qualitative Behaviours

Radiation perturbations oscillate with constant amplitude around $\delta_r = 0$. Dark matter perturbations in the radiation era on subhorizon scales grow logarithmically, as the universe expands too quickly for matter to cluster much. In particular, perturbations should be frozen, but because they enter the horizon with non-negligible $\dot{\delta}, \delta$ continues to grow a bit after horizon entry until its initial velocity is redshifted away, hence the log growth. (c) Matter Perturbations Combining the continuity and Euler equations for pressureless perturbations,

$$\delta_m' = -\nabla \cdot \boldsymbol{v}_m$$
$$\boldsymbol{v}_m' = -\mathcal{H}\boldsymbol{v}_m - \nabla\Phi$$

we find

$$\delta_m'' + \mathcal{H}\delta_m' = \nabla^2\Phi.$$

Because radiation oscillates quickly compared to the Hubble time on small scales, averaging over time the gravitational potential is only sourced by matter fluctuations, so the Poisson equation is only sensitive to matter,

$$\nabla^2\Phi = 4\pi G a^2 \bar{\rho}_m \delta_m.$$

Therefore, we find

$$\delta_m'' + \mathcal{H}\delta_m' - 4\pi G a^2 \bar{\rho}_m \delta_m = 0.$$

This next step involves a bit of algebra, but it's nothing outside your powers. Starting from (8.3.239), there are two things we want to do before anything else. One is that, of course, we want to switch from τ to y as the time coordinate. The other is that we want to replace the $4\pi G a^2 \bar{\rho}$ term with something in the Ω and y

language that we've used elsewhere in this problem. Let's do that first: using (8.3.229) and (8.3.233), we can write

$$\begin{aligned} 4\pi G a^2 \rho_m &= \frac{3}{2}\mathcal{H}_0^2 a^2 \frac{\rho_m}{\rho_{m,0}}\Omega_m \\ &= \frac{3}{2}\mathcal{H}_0^2 \Omega_m a^{-1} \\ &= \frac{3}{2}\mathcal{H}_0^2 \Omega_m a_{\rm eq}^{-1} y^{-1} \\ &= \frac{3}{2}\mathcal{H}\mathcal{H}_0^2 \frac{\Omega_m^2}{\Omega_r} y^{-1} \end{aligned}$$

Let's keep this in our back pockets. Next, we'll do the variable change, noting first that, differentiating (8.3.234) with respect to y, we have

$$\frac{d\mathcal{H}}{dy} = -\frac{\mathcal{H}_0^2}{2\mathcal{H}}\frac{\Omega_m^2}{\Omega_r}\left(y^{-2}+2y^{-3}\right).$$

With this in mind, we can easily take derivatives to find

$$\begin{aligned} \delta_m' &= \mathcal{H} y \frac{d\delta_m}{dy} \\ \delta_m'' &= \mathcal{H}^2 y^2 \left[\frac{d^2\delta_m}{dy^2} + \frac{1}{y}\frac{d\delta_m}{dy} - \frac{\mathcal{H}_0^2}{2\mathcal{H}^2}\frac{\Omega_m^2}{\Omega_r}\left(y^{-2}+2y^{-3}\right)\frac{d\delta_m}{dy}\right] \end{aligned}$$

Plugging all this into (8.3.239) and dividing out $\mathcal{H}^2 y^2$, we obtain

$$\frac{d^2\delta_m}{dy^2} + \frac{2}{y}\frac{d\delta_m}{dy} - \frac{\mathcal{H}_0^2}{2\mathcal{H}^2}\frac{\Omega_m^2}{\Omega_r}\left(y^{-2}+2y^{-3}\right)\frac{d\delta_m}{dy} - \frac{3}{2y^3}\frac{\mathcal{H}_0^2}{\mathcal{H}^2}\frac{\Omega_m^2}{\Omega_r}\delta_m = 0.$$

We can simplify two of these terms with the help of (8.3.234), finding

$$\frac{\mathcal{H}_0^2}{\mathcal{H}^2}\frac{\Omega_m^2}{\Omega_r} = \frac{y^2}{1+y}$$

Plugging this into our equation and combining terms, we obtain the Meszaros equation,

$$\frac{d^2\delta_m}{dy^2} + \frac{2+3y}{2y(1+y)}\frac{d\delta_m}{dy} - \frac{3}{2y(1+y)}\delta_m = 0.$$

By plugging and checking, we can confirm that the two solutions are

$$\begin{aligned} \delta_m &\propto 2+3y \\ \delta_m &\propto (2+3y)\ln\left(\frac{\sqrt{1+y}+1}{\sqrt{1+y}-1}\right) - 6\sqrt{1+y}. \end{aligned}$$

During radiation domination ($y \ll 1$), the first mode is clearly constant while the second, Taylor expanding, contains a constant term and a piece evolving as $\ln y \sim \ln a$. During matter domination ($y \gg 1$), the first solution is $\delta \sim y \sim a$, while the second decays as $y^{-3/2}$. These confirm the results from the lectures: matter clusters very slowly, growing logarithmically, during radiation domination, and then grows, $\delta \sim a$, during matter domination.

Baumann 3.5. Growth of Matter Perturbations II: Late Times

At late times, the universe is dominated by pressureless matter (m) and dark energy (Λ). Assuming that the dark energy doesn't cluster, the gravitational potential is only sourced by the matter.

(a) Use the Einstein equations to show that the comoving-gauge matter density contrast, $\Delta_m \equiv \delta_m - 3\mathcal{H}v_m$, evolves as

$$\Delta_m'' + \mathcal{H}\Delta_m' - 4\pi G a^2 \bar{\rho}_m \Delta_m = 0.$$

(b) Show that $u \equiv \Delta_m/H$ satisfies

$$\frac{d^2u}{da^2} + 3\frac{d\ln(aH)}{da}\frac{du}{da} = 0.$$

Confirm that the decaying mode is $\Delta_m \propto H$, while the growing mode can be written as

$$\Delta_m \propto H\int_a^{a_i}\frac{{\rm d}\tilde{a}}{(\tilde{a}\tilde{H})^3}.$$

(c) What are the growing and decaying modes of Δ_m in the matter-dominated era? What is the asymptotic limit ($a \to +\infty$) of the growing mode solution in the dark energy-dominated era?

Solution for (a) Evolution Equation

Recall that the Poisson equation can be conveniently written in terms of Δ_m as

$$\nabla^2\Phi = 4\pi G a^2 \bar{\rho}_m \Delta_m,$$

where we have defined $\Delta_m \equiv \delta_m - 3\mathcal{H}v_m$. Because $\bar{\rho}_m \sim a^{-3}$, this implies (for a given k mode)

$$\Phi \sim \frac{\Delta_m}{a}.$$

Plugging this into (E3),

$$\Phi'' + 3\mathcal{H}\Phi' + \left(2\mathcal{H}' + \mathcal{H}^2\right)\Phi = 0,$$

we find that the matter density perturbations in comoving gauge obey the equation

$$\Delta_m'' + \mathcal{H}\Delta_m + \left(\mathcal{H}' - \mathcal{H}^2\right)\Delta_m = 0.$$

Let's simplify this using the Friedmann equation,

$$\mathcal{H}^2 = \frac{8\pi G}{3} a^2 \left(\bar{\rho}_m + \bar{\rho}_\Lambda\right).$$

Taking a derivative of this, and using the fact that $\bar{\rho}_m' + 3\mathcal{H}\bar{\rho}_m = 0$ while $\bar{\rho}_\Lambda' = 0$, we find

$$\mathcal{H}' = \mathcal{H}^2 - 4\pi G a^2 \bar{\rho}_m.$$

Therefore, we see that Δ_m obeys the equation

$$\Delta_m'' + \mathcal{H}\Delta_m - 4\pi G a^2 \bar{\rho}_m \Delta_m = 0.$$

The second term is the usual friction due to the expansion of the universe, while the last term describes gravitational instability.

Solution for (b) Matter and Dark Energy

Let's define $u \equiv \Delta/H$, where the cosmic-time Hubble parameter H is related to the conformal-time version by $\mathcal{H} = aH$. We'll use overdots to denote derivatives with respect to a, so we have

$$\begin{aligned}
\Delta &= \frac{\mathcal{H}}{a} u, \\
\Delta' &= \mathcal{H}^2 \dot{u} - \frac{\mathcal{H}^2}{a} u + \mathcal{H}\dot{\mathcal{H}} u \\
\Delta'' &= a\mathcal{H}\left(\mathcal{H}^2 \ddot{u} + 3\mathcal{H}\dot{\mathcal{H}}\dot{u} - \frac{\mathcal{H}^2}{a}\dot{u} + \frac{\mathcal{H}^2}{a^2} u - \frac{2\mathcal{H}\dot{\mathcal{H}}}{a} u + \dot{\mathcal{H}}^2 u + \mathcal{H}\ddot{\mathcal{H}} u\right)
\end{aligned}$$

Plugging these into our master equation, many terms cancel and we find

$$\mathcal{H}^2 \ddot{u} + 3\mathcal{H}\dot{\mathcal{H}}\dot{u} + \left(\dot{\mathcal{H}}^2 - \frac{\mathcal{H}\dot{\mathcal{H}}}{a} + \mathcal{H}\ddot{\mathcal{H}} - 4\pi G\bar{\rho}_m\right) u = 0$$

Taking a derivatives of the Friedmann equation, $3\mathcal{H}^2 = 8\pi G\left(\rho_m + \rho_\Lambda\right)$, with $\rho_m = \rho_{m,0} a^{-3}$ and ρ_Λ constant, we find

$$\begin{aligned}
\mathcal{H}\dot{\mathcal{H}} &= \frac{\mathcal{H}^2}{a} - 4\pi G a \bar{\rho}_m, \\
\dot{\mathcal{H}}^2 + \mathcal{H}\ddot{\mathcal{H}} &= \frac{\mathcal{H}^2}{a^2}.
\end{aligned}$$

Plugging these back into our fluctuation equation, we find that the terms in square brackets all cancel, leaving us with

$$\mathcal{H}^2 \ddot{u} + 3\mathcal{H}\dot{\mathcal{H}}\dot{u} = 0.$$

Dividing by $\mathcal{H}^2$ and using $\mathcal{H} = aH$, we have this in the desired form,

$$\frac{d^2u}{da^2} + 3\frac{d\ln(aH)}{da}\frac{du}{da} = 0.$$

There are two solutions, in general, to (8.3.263). One is obvious: $u = $ const. is a solution, corresponding to $\Delta = uH \sim H$. This is generally a decaying mode, as H decreases over time in all reasonable (i.e., $w > -1$) cosmologies. The second solution can be obtained by straighforwardly integrating (8.3.263). Solving for $\dot{u}$, we find

$$d\ln\dot{u} = d\ln(aH)^{-3},$$

which has the solution

$$\dot{u} \sim (aH)^{-3}.$$

Integrating this we obtain the growing-mode solution,

$$\Delta \sim H\int_a^{a_i}\frac{d\tilde{a}}{(\tilde{a}H)^3}.$$

Solution for (c) Solutions and Behaviours

In the matter-dominated era, with $\rho_\Lambda \approx 0$, we have $H^2 \propto \rho_m \propto a^{-3} \propto (1+z)^3$, so the decaying mode behaves as

$$D_-(z) \propto H(z) \propto a^{-3/2}.$$

The growing mode instead goes as

$$D_+(z) \propto (1+z)^{3/2}\int_z^\infty \frac{1+z'}{(1+z')^{9/2}}\,\mathrm{d}z' \propto \frac{1}{1+z} = a.$$

This results confirm the expected behaviours for a matter-dominated universe. Now let us consider a universe with $\rho_\Lambda \neq 0$ so that eventually the universe tends to de Sitter with $H(z) \to$ const in the asymptotic future. The "decaying mode" then trivially goes to a constant at late times. In fact, the "growing mode" has the same behaviour. From eq. (8.3.266), we see that the integral over $\tilde{z}$ gives a finite answer as $z \to -1$ (the asymptotic future) because the integrand goes to zero near the lower limit of integration. (The asymptotic growth in the matter-dominated case happens because the integrand diverges near the lower limit.) Thus we see that the growth of structure eventually halts as dark energy comes to dominate.

Leuven-2023hw6.5. The evolution of small-scale sub-horizon modes and Meszarous equations

a) Rewrite the set of evolution equations for dark matter overdensities δ on small sub-horizon scales (Θ_0, Θ_1 neglected; $aH/k \ll 1$),

$$\dot{\delta} + \mathrm{i}kv = -3\dot{\Phi}, \quad \dot{v} + \frac{\dot{a}}{a}v = \mathrm{i}k\Phi, \quad k^2\Phi = 4\pi G a^2\rho_\mathrm{m}\delta,$$

in terms of the variable y defined above in (10) and use them to derive the Meszaros equation:

$$\delta'' + \frac{2+3y}{2y(1+y)}\delta' - \frac{3}{2y(y+1)}\delta = 0.$$

b) Verify that the solutions to the Meszaros equation are

$$\delta \propto \begin{cases} y + \frac{2}{3} \\ \left(y+\frac{2}{3}\right)\ln\left(\frac{\sqrt{1+y}+1}{\sqrt{1+y}-1}\right) - 2\sqrt{1+y} \end{cases}$$

and discuss their behavior for late times ($y \gg 1$).

Solution

Leuven-2023hw3.5. Transfer function of dark matter

Solve the set of five equations

$$\begin{aligned}\dot{\Theta}_{r,0} + k\Theta_{r,1} &= -\dot{\Phi},\\ \dot{\Theta}_{r,1} - \frac{k}{3}\Theta_{r,0} &= -\frac{k}{3}\Phi,\\ \dot{\delta} + ikv &= -3\dot{\Phi},\\ \dot{v} + \frac{\dot{a}}{a}v &= ik\Phi\end{aligned}$$

and

$$k^2\Phi + 3\frac{\dot{a}}{a}\left(\dot{\Phi} + \frac{\dot{a}}{a}\Phi\right) = 4\pi G a^2\left(\rho_{dm}\delta + 4\rho\Theta_{r,0}\right)$$

numerically to obtain the transfer function for dark matter. Use the initial conditions derived in Chapter 6 of Dodelson's book. The one numerical problem you may encounter using (5) occurs on small scales when you try to evolve all the way to the present. The photon moments then become difficult to track, and even a good differential equation solver will balk at late times. However, there are several simple solutions to this: (i) by the late times in question, the potential is constant so there is no need to evolve all the way to the present or (ii) stop following the photon moments after a certain time; they don't have any effect on the matter distribution at late times anyway. Plot the evolution of the gravitational potential Φ for a few different values of k for sCDM (with Hubble constant $h = 0.5$) and ΛCDM (with $\Omega_\Lambda = 0.7$ and $h = 0.7$). Normalize Φ to 1 at early times in this plot. Also plot the transfer function for sCDM and ΛCDM and compare with the BBKS transfer function

$$T\left(x \equiv k/k_{eq}\right) = \frac{\ln[1+0.171x]}{0.171x}\left[1 + 0.284x + (1.18x)^2 + (0.399x)^3 + (0.490x)^4\right]^{-0.25}.$$

Solution

(no one has time for it)

D-2.

Solve the set of five equations ((7.11)-(7.14) and (7.15)) numerically to obtain the transfer function for dark matter. Use the initial conditions derived in Chapter 6. The one numerical problem you may encounter using Eq. (7.15) occurs on small scales when you try to evolve all the way to the present. The photon moments then become difficult to track, and even a good differential equation solver will balk at late times. However, there are several simple solutions to this: (i) by the late times in question, the potential is constant so there is no need to evolve all the way to the present or (ii) stop following the photon moments after a certain time; they don't have any effect on the matter distribution at late times anyway. Plot the transfer function for sCDM (with Hubble constant $h = 0.5$) and $\triangle CDM$ (with $\Omega_\Lambda = 0.7$ and $h = 0.7$). Compare with the BBKS transfer function of Eq. (7.70).

Solution

Solution some problems by Dodelson chapter 7

Solution to

4 (c) To do the integral, introduce a new dummy variable $x \equiv \sqrt{1+y}$. Then Eq. (7.31) becomes

$$\Phi = \frac{3\Phi(0)}{2}\frac{\sqrt{1+y}}{y^3}\int_1^{\sqrt{1+y}} dx \frac{\left(x^2-1\right)^2\left(3x^2+1\right)}{x^2}.$$

Now integrate by parts using the fact that the integral of $1/x^2$ is equal to $-1/x$. The surface term is proportional to the numerator and so vanishes at the lower limit, when $x = 1$. Therefore,

$$\begin{aligned}\Phi &= \frac{3\Phi(0)}{2}\frac{\sqrt{1+y}}{y^3}\left[-\frac{y^2(4+3y)}{\sqrt{1+y}} + \int_1^{\sqrt{1+y}} dx\left(18x^4 - 20x^2 + 2\right)\right]\\ &= \frac{3\Phi(0)}{2}\frac{\sqrt{1+y}}{y^3}\left[-\frac{y^2(4+3y)}{\sqrt{1+y}} + \left(\frac{18}{5}x^5 - \frac{20}{3}x^3 + 2x\right)\Big|_1^{\sqrt{1+y}}\right].\end{aligned}$$

Evaluating the terms in parentheses at the upper and lower limits leads to Eq. (7.32).

Solution to

9

$$\sigma_R^2 = \left\langle \left[\int d^3x \delta(x) W_R(x) \right]^2 \right\rangle$$
$$= \left\langle \left[\frac{d^3k}{(2\pi)^3} \tilde{\delta}(\vec{k}) \tilde{W}_R^*(\vec{k}) \right]^2 \right\rangle$$

where denotes Fourier transform, and I have used the fact that since $W_R(x)$ is real, $\tilde{W}_R(\vec{k}) = \tilde{W}_R^*(-\vec{k})$. Also I have evaluated δ_R at the origin; and the angular brackets denote the average, now over all realizations of $\tilde{\delta}(\vec{k})$. Squaring and using the fact that

$$\left\langle \delta(\vec{k}) \delta\left(\vec{k}'\right) \right\rangle = (2\pi)^3 \delta^3\left(\vec{k} + \vec{k}'\right) P(k)$$

leads to

$$\sigma_R^2 = \int \frac{d^3k}{(2\pi)^3} P(k) \left| \tilde{W}_R(\vec{k}) \right|^2$$

It remains only to compute the Fourier transform of the top-hat window function,

$$\tilde{W}_R(\vec{k}) = \int d^3x W_R(\vec{x}) e^{-i\vec{k}\cdot\vec{x}}$$
$$= \frac{2\pi}{V_R} \int_0^R dx x^2 \int_{-1}^1 d\mu e^{ikx\mu}.$$

Note that I have normalized the window function so that the integral over it is unity; hence the factor of $V_R = 4\pi R^3/3$. Carrying out the remaining angular and radial integrals leads to

$$\tilde{W}_R(k) = \frac{3}{kR^3} \int_0^R dx x \sin(kx)$$

$$= \frac{3}{k^3 R^3} [-kR\cos(kR) + \sin(kR)]$$

By way of solving

Solution to

10, note that

$$\Delta^2(k) = \frac{4\pi}{(2\pi)^3} k^3 P(k)$$

13.2.3 Problems on Anisotropies

D-1.

Most of this book is devoted to understanding adiabatic perturbations with the initial conditions derived in Chapter 6. Another class of perturbations are isocurvature perturbations with initial conditions $\Theta_0 = \Psi = \Phi = 0$. Show that these initial conditions imply that

$$\Theta_0(\eta_*) + \Psi(\eta_*) = 2\Psi(\eta_*).$$

D-2.

The equation for a damped harmonic oscillator is

$$m\ddot{x} + b\dot{x} + kx = 0.$$

Find the solutions to this equation if $k/m > (b/2m)^2$. What is the frequency of oscillations? How does this differ from the undamped ($b = 0$) solution? What is the other effect of nonzero b besides the change in frequency?

D-3.

Determine $R(\eta_*)$ when $\Omega_b h^2 = 0.01, 0.02$. Plot the sound speed as a function of the scale factor for these two values of $\Omega_b h^2$.

D-4.

Show that the sound horizon can be expressed in terms of the conformal time as

$$r_s(\eta) = \frac{2}{3k_{\rm eq}}\sqrt{\frac{6}{R(\eta_{\rm eq})}}\ln\left\{\frac{\sqrt{1+\bar{R}}+\sqrt{R+R(\eta_{\rm eq})}}{1+\sqrt{R(\eta_{\rm eq})}}\right\},$$

where $k_{\rm eq}$ is given in Eq. (7.39).

D-5.

Obtain the WKB solution to Eq. (8.18). Write

$$\Theta_0 = Ae^{iB}$$

with A and B real. Show that the homogeneous part of Eq. (8.18) breaks up into two equations, coming from the real and imaginary part:

$$\text{Real: } -(\dot{B})^2 + \frac{\ddot{A}}{A} + \frac{\dot{R}}{1+R}\frac{\dot{A}}{A} + k^2c_s^2 = 0$$
$$\text{Imaginary } : 2\dot{B}\frac{\dot{A}}{A} + \ddot{B} + \frac{\dot{R}}{1+R}\dot{B} = 0.$$

Find B using the real part and the fact that B changes much more rapidly than A. Then, use the imaginary equation to determine A. Show that the homogeneous solutions obtained in this way differ from the simple oscillatory solutions of Eq. (8.21) by a factor of $(1+R)^{1/4}$.

D-6.

Obtain a semianalytic solution for $\Theta_0 + \Psi$ and Θ_1 at recombination by carrying out the integrals in Eqs. (8.24) and (8.26). To do this you will need expressions for the gravitational potentials. Hu and Sugiyama (1995) provided the following convenient fits:

$$\Phi(k,y) = \bar{\Phi}(k,y)\left\{[1-T(k)]\exp\left[-0.11\,(ky/k_{\rm eq})^{1.6}\right] + T(k)\right\}$$
$$\Psi(k,y) = \bar{\Psi}(k,y)\left\{[1-T(k)]\exp\left[-0.097\,(ky/k_{\rm eq})^{1.6}\right] + T(k)\right\}$$

where $y := a/a_{\rm eq}, T(k)$ is the BBKS transfer function and the large-scale potentials are

$$\bar{\Phi}(k,y) = \frac{3}{4}\left(\frac{k_{\rm eq}}{k}\right)^2\frac{y+1}{y^2}\bar{\Delta}_T(y)$$
$$\bar{\Psi}(k,y) = -\frac{3}{4}\left(\frac{k_{\rm eq}}{k}\right)^2\frac{y+1}{y^2}\left(\Delta_T(y) + 0.65N_2/(1+y)\right).$$

Finally the two functions N_2 and Δ_T are

$$N_2(y) = -0.1\frac{20y+19}{3y+4}\frac{y^2}{y+1}\Phi_{ls} - \frac{8}{3}\frac{y}{3y+4} + \frac{8}{9}\ln[3y/4+1]$$
$$\Delta_T = \left[1.16 - \frac{0.48y}{y+1}\right]\Phi_{ls}\frac{y^2}{y+1}.$$

Here Φ_{ls} is the large-scale solution of Eq. (7.32).

D-7.

Our treatment of diffusion damping neglected the effect of polarization. Go through the same expansion in $\dot{\tau}^{-1}$ that we carried out in Section 8.4 this time accounting for polarization. Show that this changes the factor of 8/9 in Eq. (8.40) to 16/15. This beautiful result was obtained by Zaldarriaga and Harari (1996) when the first author was an undergraduate!

D-8.

Assume that all electrons associated with hydrogen stay ionized and set $R = 0$. Evaluate the damping scale, k_D, defined in Eq. (8.40). Show that in this limit, the damping scale is given by Eq. (8.42), where

$$f_D(y) = 5\sqrt{1+1/y} - \frac{20}{3}(1+1/y)^{3/2} + \frac{8}{3}\left[(1+1/y)^{5/2} - 1/y^{5/2}\right].$$

D-9.

Show that

$$\int d\Omega Y_{lm}(\hat{p})\mathcal{P}_l(\hat{p}\cdot\hat{k}) = \frac{4\pi}{2l+1}$$

D-10.

There is a different way to go from the inhomogeneous temperature field at recombination, $\Theta_0\left(\vec{x},\eta_*\right)$ or $\Theta_0\left(\vec{k},\eta_*\right)$, to the anisotropy pattern today, a_{lm}, than that given in the text.

(a) Assume that the photons we see today from direction $\hat{p}$ come from the surface of last scattering: $\Theta\left(\vec{x}_0,\hat{p},\eta_0\right) = \left(\Theta_0+\Psi\right)\left(\vec{x}=\chi_*\hat{p},\eta_*\right)$ where x_0 is our position. Fourier transform the right-hand side and expand the left in terms of spherical harmonics to get

$$\sum_{lm} a_{lm}Y_{lm}(\hat{p}) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\hat{p}\chi_*}\left(\tilde{\Theta}+\tilde{\Psi}\right)\left(\vec{k},\eta_*\right).$$

Now expand the exponential using Eq. (C.16) and then expand the resulting Legendre polynomial using Eq. (C.12). Equate the coefficients of $Y_{lm}(\hat{p})$ to get an expression for a_{lm}.

(b) Square the a_{lm} you got in (a) and take the expectation value to get an expression for C_l. You should recapture the expression in Eq. (8.68) with Θ_l given by the first term in Eq. (8.56).

D-11.

A simple way to estimate the COBE normalization of δ_H is to fix C_{10}. From Figure 8.12, estimate C_{10}. Use this and the Sachs-Wolfe formula, Eq. (8.75), to estimate δ_H for a flat, matter-dominated universe. Compare with the number given in Eq. (8.76)

D-12.

Show that the cross-terms from the monopole and dipole vanish when summing over all modes. The monpole is proportional to $j_l\left(k\eta_0\right)$ while the dipole is proportional to $j_l'\left(k\eta_0\right)$. Compute the three possible integrals

$$\int_0^\infty dx j_l j_l \quad ; \quad \int_0^\infty dx j_l j_l' \quad ; \quad \int_0^\infty dx j_l' j_l'.$$

Show that the integrals of the squares $\left(j_l^2 \text{ and } \left(j_l'\right)^2\right)$ are much larger than the integral of the cross-term $j_l j_l'$. Do the integrals for $l = 10$ up to $l = 50$.

D-13.

Determine the shift in the locations of the peaks and troughs in the CMB anisotropy spectrum if the universe is flat with a cosmological constant as opposed to flat and matter dominated. Keep the sound horizon fixed in this calculation by fixing $\Omega_m h^2 = 0.15$. The peak positions then depend only on the distance to the last scattering surface, $\eta_0 - \eta_*$. Consider two flat models: (i) $\Omega_\Lambda = 0$ (so that $\Omega_m = 1$) and (ii) $\Omega_\Lambda = 0.7$ (so that $\Omega_m = 0.3$). What value of h is needed in the two cases to keep $\Omega_m h^2$ fixed? Determine $\eta_0 - \eta_*$ in each case (in the the cosmological constant case, you will have to do the integral numerically). Compare your result with the fitting formula: $\eta_0 \propto 1 + \ln\left(\Omega_m^{0.085}\right)$ and with the shift in Figure 8.19.

D-14.

Compute the conformal time today in a flat model with dark energy $\Omega_{\rm de} = 0.7$ today with $w = -0.5$. Compare the expected shift in the anisotropy spectrum with the cosmological constant model of the previous problem.

D-15.

Determine the effects of reionization using the Boltzmann equation. Neglect the gravitational potentials, the velocity, and Θ_0 in the Boltzmann equation for photons. Start with a spectrum $\Theta_l(\eta)$ and evolve till today. Show that the moments are indeed suppressed by $e^{-\tau}$.

D-16.

Assume that recombination took place instantaneously. Show that the solution for the lth moment due to tensor perturbations (Eq. (4.116)) is

$$\Theta_l^T = -\frac{1}{2}\int_{\eta_*}^{\eta_0} d\eta \dot{h} j_l \left[k\left(\eta_0 - \eta\right)\right]$$

D-17.

Using the decomposition for tensor modes given in Eq. (4.115), find the contribution to the C_l 's from $\Theta_l^T(k)$. That is, show that the analogue of Eq. (8.68) for tensors is

$$C_{l,i}^T = \frac{(l-1)l(l+1)(l+2)}{\pi}\int_0^\infty dk k^2 \left|\frac{\Theta_{l-2,i}^T}{(2l-1)(2l+1)} + 2\frac{\Theta_{l,i}^T}{(2l-1)(2l+3)} + \frac{\Theta_{l+2,i}^T}{(2l+1)(2l+3)}\right|^2,$$

where i denotes the two different components $+$ and x.

D-18.

Determine the spectrum of anisotropies due to gravity waves produced during inflation.
(a) Combine the results of the previous two problems, your solution to

D-5.

12, and the primordial amplitude of gravity waves in Eq. (6.100) to find the large-angle C_l 's due to inflation-produced gravity waves.
(b) Tensor anisotropies are often parametrized by

$$r := \frac{C_2^T}{C_2^S}$$

where C_2^T is the variance of the quadrupole due to tensors and C^S is the same due to scalars. [8] We already derived an expression for the scalar C_2 in Eq. (8.75). Find C_2^T and compute r to first order in the slow roll parameter ϵ.
(c) The results of part (b) and Eq. (6.104) imply a consistency relation - a robust prediction of inflation - between the two observables n_T and r. What is the consistency relation?
[8] Note that this convention is not universal; r is sometimes defined to the (more precisely constrained) ratio at $l \sim 10$.

Solution some problems by Dodelson chapter 8

Solution to

2 Assume a solution of the form $x = e^{i\omega t}$. The damping equation then becomes a quadratic equation for ω:

$$\omega^2 - \frac{ib}{m}\omega - \frac{k}{m} = 0$$

Solving with $k/m > \gamma^2 \equiv (b/2m)^2$ leads to

$$\omega = i\gamma \pm \omega_1$$

The frequency is now $\omega_1 \equiv \left[k/m - \gamma^2\right]^{1/2}$, smaller than in the undamped case. The amplitude is also damped by $e^{-\gamma t}$.

Solution to

9 Use the addition theorem of spherical harmonics (C.12) to write

$$\mathcal{P}_{l'}(\hat{\gamma}\cdot\hat{k}) = \frac{4\pi}{2l+1}\sum_{m'} Y^*_{l'm'}(\hat{\gamma})Y_{l'm'}(\hat{k})$$

Then the angular integral becomes an integral over the product of two spherical harmonics, which - because of orthogonality - is equal to 1 if $l' = l$ and $m' = m$ and zero otherwise. This leads directly to the desired reult.

Solution to

12 I get the result show in Figure A.1. The integral of the cross-term is significantly smaller than that of either of the squares, so there is no interefence between the monopole and dipole.

Solution to

17 The generalization of Eq. (8.67) to tensors gives

$$C_l^T = \sum_{l'l''}(-i)^{l'+l''}\,(2l'+1)\,(2l''+1)\int\frac{d^3k}{(2\pi)^3}\Theta_{l'}^T(k)\Theta_{l''}^{T,*}(k)I_{lml'}(k)I^*_{lml''}(k)$$

where I have defined

$$I_{lml'}(k) \equiv \sqrt{\frac{8\pi}{15}}\int d\Omega\mathcal{P}_{l'}(\hat{k}\cdot\hat{\gamma})Y_{lm}(\Omega)\left[Y_{22}(\Omega)+Y_{2-2}(\Omega)\right]$$

The factor of $[8\pi/15]^{1/2}\left[Y_{22}+Y_{2-2}\right]$ is the combination $\sin^2\theta\cos(2\phi)$ which appears in Eq. (4.115), so this expression is valid only for the + mode. However, the × mode gives exactly the same result.

The integral $I_{lml'}$ is not trivial. By rewriting the Legendre polynomial as $\left[4\pi/\left(2l'+1\right)\right]^{1/2}Y_{l'0}/i^{l'}$, we can turn $I_{ll'}$ into an integral over the product of three

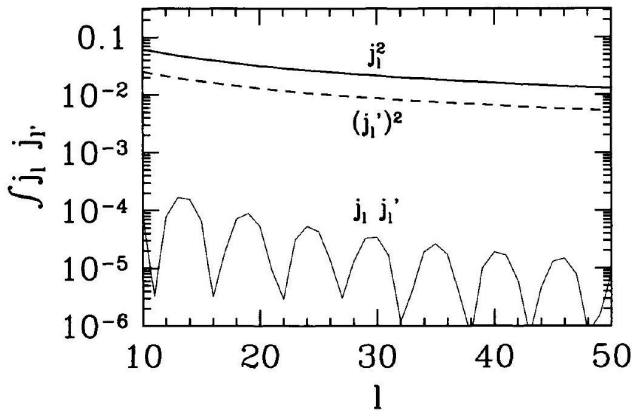


Figure A.1. The integrals of products of spherical Bessel functions.

spherical harmonics. Such integrals are intensively studied in quantum mechanics and can be expressed in terms of the Wigner 3-j symbols. By the way, my favorite reference for these things - especially useful for this integral-is Quantum Mechanics (Landau and Lifshitz), like all the other texts in their Course of Theoretical Physics a wonderful investment. The integral is then

$$I_{lml'} = \sqrt{\frac{32\pi^2}{15\left(2l'+1\right)}}\frac{1}{i^{l'}}\left\langle lm\left|Y_{22}+Y_{2-2}\right|l'0\right\rangle$$

which vanishes unless $m = 2$ or $m = -2$. When m takes on one of these two values, the matrix element is

$$\left\langle l2\left|Y_{22}+Y_{2-2}\right|l'0\right\rangle = i^{l'-l}\begin{pmatrix} l & 2 & l' \\ 0 & 0 & 0\end{pmatrix}\left[\frac{5\left(2l'+1\right)\left(2l+1\right)}{4\pi}\right]^{1/2}\begin{pmatrix} l & 2 & l' \\ -2 & 2 & 0\end{pmatrix}.$$

The first 3 -j symbol here, the one with the bottom row all zero, vanishes unless the sum of the elements in the top row $l + l' + 2$ is even. And of course l' cannot differ from l by more than 2 since the combination of $Y_{22}Y_{l'0}$ leads to angular momenta ranging from $l' - 2$ to $l' + 2$. So the only time the matrix element is nonzero is when $l' = l - 2, l, l + 2$. Using Table 9 in Section 106 of Quantum Mechanics leads to the final result:

$$I_{lml'} = \sqrt{\frac{8\pi}{3}}\sqrt{2l+1}i^{-l}\left(\delta_{m,2} + \delta_{m,-2}\right)\left[c_{-2}\delta_{l',l-2} + c_0\delta_{l',l} + c_2\delta_{l',l+2}\right]$$

where here $\delta_{m,2}$ (and all other δ 's) is the Kronecker delta, equal to 1 if $m = 2$ and zero otherwise. The coefficients are

$$\begin{aligned} c_{-2} &= \frac{\sqrt{6}}{4}\frac{\left[(l-1)l(l+1)(l+2)\right]^{1/2}}{(2l-3)(2l-1)(2l+1)} \\ c_0 &= \frac{-2\sqrt{6}}{4}\frac{\left[(l-1)l(l+1)(l+2)\right]^{1/2}}{(2l-1)(2l+1)(2l+3)} \\ c_2 &= \frac{\sqrt{6}}{4}\frac{\left[(l-1)l(l+1)(l+2)\right]^{1/2}}{(2l+1)(2l+3)(2l+5)}. \end{aligned}$$

The result in Eq. (8.93) then follows.

Solution to

18
(a) On large scales, we can take the matter-dominated solution for h, so

$$\Theta_{l,i} = \frac{-1}{2}\int_{\eta_*}^{\eta_0} d\eta j_l\left[k\left(\eta_0 - \eta\right)\right]\frac{d}{d\eta}\left[\frac{3j_1(k\eta)}{k\eta}\right]P_h^{1/2}$$

Here I have used the fact that the initial amplitude of the gravity waves is $P_h^{1/2}$ with the time dependence given in the square brackets. Plug this into Eq. (8.93) to get

$$\begin{aligned} C_l^T =& 2\frac{9(l-1)l(l+1)(l+2)}{4\pi}\int_0^\infty dk k^2 P_h(k)\Big|\int_0^{\eta_0} d(k\eta)\frac{j_2(k\eta)}{k\eta} \\ &\times\left[\frac{j_{l-2}\left(k\left[\eta_0-\eta\right]\right)}{(2l-1)(2l+1)} + 2\frac{j_l\left(k\left[\eta_0-\eta\right]\right)}{(2l-1)(2l+3)} + \frac{j_{l+2}\left(k\left[\eta_0-\eta\right]\right)}{(2l+1)(2l+3)}\right]\Big|^2, (\end{aligned}$$

where I have set the lower limit on the time integral to zero since $\eta_* \ll \eta_0$. Also, I have used the identity $(j_1/x)' = -j_2/x$. The factor of 2 out in front comes from the sum over the + and × components. Using Eq. (6.100) for P_h (in the slow-roll approximation $\epsilon = 0$ and $\nu = 3/2$) and defining new integration variables $y \equiv k\eta_0$ and $x \equiv k\eta$ leads to

$$\begin{aligned} C_l^T =& 36\left(\frac{H_{\rm inf}}{m_{\rm Pl}}\right)^2(l-1)l(l+1)(l+2)\int_0^\infty\frac{dy}{y}\Big|\int_0^y dx\frac{j_2(x)}{x} \\ &\times\left[\frac{j_{l-2}(y-x)}{(2l-1)(2l+1)} + 2\frac{j_l(y-x)}{(2l-1)(2l+3)} + \frac{j_{l+2}(y-x)}{(2l+1)(2l+3)}\right]\Big|^2. \end{aligned}$$

Here $H_{\rm inf}$ denotes the Hubble rate during inflation, or more precisely the Hubble rate when the modes in question crossed the horizon (when $k\eta = -1$ early on). This expression does well on the low multipoles. To get even better results stick in the transfer function of Eq. (5.88).

(b) For the $l = 2$ mode, the double integral in Eq. (A.107) is equal to 2.139×10^{-4}, so $C_2^T = 0.185\left(H/m_{\rm PI}\right)^2$. The scalar C_2 is equal to $\pi\delta_H^2/12$. Using Eq. (6.100) for δ_H leads to

$$r = 13.86\epsilon$$

(c) Combining with Eq. (6.104), we expect

$$r = -6.93 n_T.$$

For many models, the inflationary parameter $\delta = -\epsilon$, so

$$n - 1 = -n_T$$

13.2.4 Problems from probes of inhomogeneities by Dodelson (??????)

(вообще таким не занимался, ничего про это пока не знаю.)

D-1.

Suppose the correlation function is defined as

$$\xi(\vec{r}) := \langle \delta(\vec{x})\delta(\vec{x}+\vec{r})\rangle$$

By Fourier expanding each of the δ 's and using Eq. (C.20), show that this definition implies that the correlation function is the Fourier transform of the power spectrum.

D-2.

Expand the 3D power spectrum in the integral of Eq. (9.9) about $k_3 = 0$. The leading term is the one we considered. Show that the next term is of order $(1/l)^2$, compared with the leading term.

D-3.

Rewrite the kernel in Eq. (9.14) as an integral over χ. Show that F is a function of $k\theta$ only.

D-4.

Give an order-of-magnitude estimate for the kernel of the angular correlation function.
(a) Consider a shell in Fourier space with radius k and width dk. What fraction of the volume of this shell has $|k_3| < \chi^{-1}$?
(b) Argue that only Fourier modes with $|k_3| < \chi^{-1}$ contribute to the angular correlation function with a weight $\Delta^2(k) = k^3P(k)/2\pi^2$. Combine this argument with the fraction computed in part (a) to estimate the kernel relating $w(\theta)$ to $P(k)$. Compare this estimate with Eq. (9.13).

D-5.

Decompose the angular correlation function into a sum over spherical harmonics,

$$w(\theta) = \sum_l \frac{2l+1}{4\pi} C_l^{\rm matter} \mathcal{P}_l(\cos\theta)$$

where the superscript matter distinguishes these C_l 's from the ones characterizing anisotropies in the CMB, and $\mathcal{P}_l$ here are the Legendre polynomials. Express $C_l^{\rm matter}$ as an integral over the 3D power spectrum. Show that on small scales $C_l^{\rm matter} = P_2(l)$, where P_2 is the 2D power spectrum introduced in Section 9.1.

D-6.

In Section 9.1 we implicitly neglected the evolution of the power spectrum. That is, we assumed that $P(k)$ remains constant with time. Allow $P(k)$ to scale as $(1+z)^\beta$. What is β for linear modes in a flat, matter-dominated universe? Rewrite the kernel in terms of an integral over z, accounting for this evolution.

D-7.

Compute (numerically) the linear growth rate f today in an open universe and compare with the approximation $\Omega_m^{0.6}$. What is the fractional error between the approximation and the exact result? Now assume that the universe is flat, with $\Omega_m + \Omega_\Lambda = 1$. Again compare the exact linear growth rate with $\Omega_m^{0.6}$. Show that

$$f = \Omega_m^{0.6} + \frac{\Omega_\Lambda}{70}\left(1 + \frac{\Omega_m}{2}\right)$$

is a better approximation, with no worse than 4% accuracy for $\Omega_m > 0.025$.

D-8.

Using CMBFAST, compute the transfer function for standard CDM ($\Omega_m = 1; h = 0.5$) with $\Omega_b = 0.01, 0.05$, and 0.1. Show that the BBKS transfer function is still a reasonable fit as long as

$$\Gamma = \Omega_m h \to \Omega_m h e^{-2\Omega_b}.$$

D-9.

Using CMBFAST or the BBKS transfer function, compute COBEnormalized σ_8 for ΛCDM with $h = 0.7, \Omega_\Lambda = 0.7, \Omega_m = 0.26$, and $\Omega_b = 0.04$. Locate the model on the bottom left panel of Figure 9.6. What does this imply about the relation between σ_8 (of the mass, which you have just computed) and $(\sigma_8)_g$ (of the galaxies, which APM is sensitive to)?

D-10.

Assume that the universe is flat with matter and a cosmological constant. Expand the comoving distance out to a galaxy at redshift z (neglecting peculiar velocities) about $z = 0$. The first-order term in the expansion should give back the redshift space answer. What is the second-order term, the leading correction to redshift space? Express your answer in terms of Ω_m.

D-11.

Derive (9.51), using the fact that $\mathcal{P}_4(x) = 35x^4/8 - 15x^2/4 + 3/8$. Show that the definition of the moments in Eq. (9.50) $- P_s^{(l)}(k) = (2l+1) \int_{-1}^{1} (d\mu_{\mathbf{k}}/2)\, \mathcal{P}_l(\mu_{\mathbf{k}})\, P_s(k, \mu_{\mathbf{k}})$ - means that

$$P_s(k, \mu_{\mathbf{k}}) = \sum_l \mathcal{P}_l(\mu_{\mathbf{k}})\, P_s^{(l)}(k)$$

D-12.

In the text we showed how redshift space distortions affect the power spectrum. Show how the redshift space distortions affect the correlation function. Assume linear theory. You will probably need to consult Hamilton (1992), which transforms Kaiser's result to the correlation function in a single (dense) paragraph.

Solution some problems by Dodelson chapter 9

Solution to

2 Expand the power spectrum about $k_3 = 0$:

$$P\left(\sqrt{k_3^2 + (H_0\kappa/\chi)^2}\right) = P(H_0\kappa/\chi) + \frac{1}{2k}\frac{dP}{dk}\bigg|_{k_3=0} k_3^2 + \ldots$$

For a smooth power spectrum dP/dk is of order P/k, so the coefficient of k_3^2 is of order P/k^2. For us, $k = H_0\kappa/\chi$, so this coefficient is of order $P\chi^2/(H_0\kappa)^2$. We can write k_3^2 as $-H_0^2\partial^2/\partial\chi^2$ acting on the exponential of Eq. (9.9). Assuming the selection function is relatively smooth, this is of order H_0^2/χ^2. So the first correction to the leading term is of order $1/\kappa^2$, which is small as long as the angular scales probed are not too large.

Solution to

3 Define the dummy variable $\chi \equiv H_0\kappa/k$. Then

$$F = H_0 \int \frac{d\chi}{2\pi\chi} J_0(k\theta\chi/H_0)\, W^2(\chi)$$

an expression which clearly depends only on the combination $k\theta$.

Solution to

5 To express $C_l^{\rm matter}$ in terms of w, multiply both sides of Eq. (9.66) by $\mathcal{P}_{l'}(\cos\theta)$ and integrate over $\mu \equiv \cos\theta$. This gives

$$C_l^{\rm matter} = 2\pi \int_{-1}^{1} d\cos\theta \mathcal{P}_l(\cos\theta) w(\theta)$$

Express w as an integral over the 2D power spectrum as in the first line of Eq. (9.13). Then,

$$C_l^{\rm matter} = \int_0^\infty dl' l' P_2(l') \int_{-1}^{1} d\cos\theta \mathcal{P}_l(\cos\theta) J_0(l'\theta)$$

Note the difference in P 's: the first P_2 here is the 2D power spectrum, the second $\mathcal{P}_l$ is the Legendre polynomial. In the limit that l' is large, the Bessel function becomes

$$J_0(l'\theta) \to \mathcal{P}_{l'}(\cos\theta)$$

Therefore, the integral over θ vanishes unless $l = l'$, in which case it is equal to $2/(2l+1)$. The integral over l' is identical to a sum over l' at large l' since $dl' \to 1$. The factor of $2/(2l+1)$ in the denominator cancels the factor of l' in the numerator, leaving the desired equality between the 2D power spectrum and $C_l^{\rm matter}$.

13.2.5 *Problem Sets* by Baumann's TASI lectures

Problem 1 (Homogeneous and Isotropic Spaces)**.** *Homogeneous and isotropic spaces are characterized by translational and rotational invariance. Convince yourself that in three dimensions there exist only three types of homogeneous and isotropic spaces with simple topology:*

i) flat space

ii) a three-dimensional sphere with constant positive curvature

iii) a three-dimensional hyperbolic space with constant negative curvature.

It is easier to visualize the two-dimensional analogues. Consider the embedding of a two-dimensional sphere in a three-dimensional Euclidean space

$$x^2 + y^2 + z^2 = a^2\,, \tag{274}$$

where a is the radius of the sphere. Show that the induced metric on the surface of the sphere is

$$\mathrm{d}\ell_2^2 = \frac{\mathrm{d}r'^2}{1-(r'^2/a^2)} + r'^2\mathrm{d}\phi^2\,, \tag{275}$$

where $x = r'\cos\phi$ and $y = r'\sin\phi$. The limit $a^2 \to \infty$ corresponds to flat space (a plane). Negative a^2 corresponds to a space with constant negative curvature. It cannot be embedded in three-dimensional Euclidean space. (Consider the embedding of $x^2+y^2-z^2 = -a^2$ in a space with metric $\mathrm{d}\ell_2^2 = \mathrm{d}x^2 + \mathrm{d}y^2 - \mathrm{d}z^2$ instead.)

By rescaling the radial coordinate the metric can be brought into the form

$$\mathrm{d}\ell_2^2 = |a^2|\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\phi^2\right)\,, \tag{276}$$

where $k = +1$ for the sphere ($a^2 > 0$), $k = -1$ for the hyperbolic space ($a^2 < 0$) and $k = 0$ for the plane ($a^2 = 0$).

Generalize the above argument to the embedding of homogeneous and isotropic three-dimensional spaces in four-dimensional Euclidean space. Show that their metric is

$$\mathrm{d}\ell_3^2 = a^2\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega^2\right)\,, \qquad \mathrm{d}\Omega^2 \equiv \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\,, \tag{277}$$

or

$$\mathrm{d}\ell_3^2 = a^2\left[\mathrm{d}\chi^2 + \begin{pmatrix}\sinh^2\chi\\ \chi^2\\ \sin^2\chi\end{pmatrix}\mathrm{d}\Omega^2\right] \quad \begin{matrix}k=-1\\k=0\\k=+1\end{matrix}\,. \tag{278}$$

Convince yourself that the only time-dependent four-dimensional spacetime that preserves homogeneity and isotropy of space is the FRW metric

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega^2\right)\,. \tag{279}$$

Problem 2 (Conformal Time)**.** *Derive some simple expressions for the conformal time τ as a function of a.*

1. *Show that $\tau \propto a^{1/2}$ in a matter-dominated universe and $\tau \propto a$ in one dominated by radiation.*

2. *Consider a universe with only matter and radiation, with equality at* $a_{\rm eq}$. *Show that*

$$\tau = \frac{2}{\sqrt{\Omega_m H_0^2}}\left[\sqrt{a+a_{\rm eq}} - \sqrt{a_{\rm eq}}\right] . \tag{280}$$

What is the conformal time today? At decoupling?

3. *Use your favorite software (say Mathematica or Maple) to compute the conformal time numerically for our universe (filled with dark energy, matter and radiation). Compute the conformal time today and at decoupling. What is the percentage error between this result and the analytical result for a matter/radiation only universe, Eqn. (280)?*

Problem 3 (Friedmann Equations)**.** *Derive the Ricci tensor and the Ricci scalar for the FRW spacetime (328)*

$$R_{00} = -3\frac{\ddot{a}}{a}\,, \qquad R_{ij} = \delta_{ij}\left[2\dot{a}^2 + a\ddot{a} + 2\frac{k}{a^2}\right], \qquad R = g^{\mu\nu}R_{\mu\nu} = 6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2}\right] . \tag{281}$$

Confirm that the 00-component of the Einstein Equation (340) gives the Friedmann Equation

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}\,. \tag{282}$$

Confirm that the trace of the Einstein Equation (340) gives the acceleration equation

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p)\,. \tag{283}$$

Show that the two Friedmann Equations imply the continuity equation

$$\dot{\rho} = -3H(\rho + p)\,. \tag{284}$$

Derive the continuity equation from

$$\nabla_\mu T^\mu_{\ \nu} = 0\,. \tag{285}$$

(Hint: Contract Eqn. (285) with U^μ*, use the energy-momentum tensor for a perfect fluid and the properties of the 4-velocity. No need for Christoffel symbols!)*

Problem 4 ($\lambda\phi^4$ Inflation)**.** *Derive the slow-roll dynamics for* $\lambda\phi^4$ *inflation.*

Problem 5 (The Phase Space of $m\phi^2$ Inflation)**.** *Read about the attractor behavior of* $m^2\phi^2$ *inflation in Mukhanov's book [9].*

Problem 6 (Vacuum Selection)**.** *Read about the* **Unruh effect** *in your favorite resource for QFT in curved spacetime.*

Problem 7 (Slow-Roll Mode Functions)**.** *In this problem we compute the mode functions and the power spectrum of curvature perturbations to first order in the slow-roll approximation.*

Recall the mode equation

$$v_k'' + \left(k^2 - \frac{z''}{z}\right)v_k = 0\,, \qquad z^2 = 2a^2\varepsilon\,. \tag{286}$$

1. *Show that*

$$\frac{z''}{z} = \frac{\nu^2 - 1/4}{\tau^2}\,, \qquad \nu \approx \frac{3}{2} + 3\varepsilon - \eta\,, \tag{287}$$

at first order in the slow-roll parameters

$$\varepsilon \equiv -\frac{\dot{H}}{H^2}\,, \qquad \eta \equiv 2\varepsilon - \frac{\dot{\varepsilon}}{2H\varepsilon}\,. \tag{288}$$

The solution can then be expressed as a linear combination of Hankel functions

$$v_k(\tau) = x^{1/2}\left[c_1 H_\nu^{(1)}(x) + c_2 H_\nu^{(2)}(x)\right], \qquad x \equiv k|\tau|\,. \tag{289}$$

In the far past, $x = k|\tau| \to \infty$*, the Hankel functions have the asymptotic limit*

$$H_\nu^{(1,2)}(x) \to \sqrt{\frac{2}{\pi x}}\exp\left[\pm i\left(x - \frac{\nu\pi}{2} - \frac{\pi}{4}\right)\right] \tag{290}$$

2. *Show that the boundary condition (506) implies*

$$v_k = a_1(\pi x/4k)^{1/2} H_\nu^{(1)}(x)\,, \tag{291}$$

where

$$a_1 = \exp[i(2\nu+1)\pi/4] \tag{292}$$

is a k-independent complex phase factor.

3. *Compute the power spectrum of $\mathcal{R} = v/z$ at large scales, $k \ll aH$.*
Hint: Use the identity

$$H_\nu^{(1)}(k|\tau|) \;\to\; \frac{i}{\pi}\Gamma(\nu)\left(\frac{k|\tau|}{2}\right)^{-\nu}, \quad \text{for} \quad k\tau \to 0\,, \tag{293}$$

and $\Gamma(3/2) = \sqrt{\pi}/2$.
Show that this reproduces the result of perfect de Sitter in the limit $\varepsilon = \eta = 0$.

4. *Read off the scale-dependence of the spectrum.*

Problem 8 (Predictions of $\lambda\phi^4$ Inflation)**.** *Determine the predictions of an inflationary model with a quartic potential*

$$V(\phi) = \lambda\phi^4\,. \tag{294}$$

1. *Compute the slow-roll parameters ϵ and η in terms of ϕ.*

2. *Determine $\phi_{\rm end}$, the value of the field at which inflation ends.*

3. *To determine the spectrum, you will need to evaluate ϵ and η at horizon crossing, $k = aH$ (or $-k\tau = 1$). Choose the wavenumber k to be equal to a_0H_0, roughly the horizon today. Show that the requirement $-k\tau = 1$ then corresponds to*

$$e^{60} = \int_0^N {\rm d}N' \frac{e^{N'}}{H(N')/H_{\rm end}}\,, \tag{295}$$

where $H_{\rm end}$ is the Hubble rate at the end of inflation, and N is defined to be the number of e-folds before the end of inflation

$$N \equiv \ln\left(\frac{a_{\rm end}}{a}\right)\,. \tag{296}$$

4. *Take this Hubble rate to be a constant in the above with $H/H_{\rm end} = 1$. This implies that $N \approx 60$. Turn this into an expression for ϕ. This simplest way to do this is to note that $N = \int_t^{t_{\rm end}} {\rm d}t' H(t')$ and assume that H is dominated by potential energy. Show that this mode leaves the horizon when $\phi = 22M_{\rm pl}$.*

5. *Determine the predicted values of $n_{\rm s}$, r and $n_{\rm t}$. Compare these predictions to the latest WMAP5 data (see* **Lecture 3***).*

6. *Estimate the scalar amplitude in terms of λ. Set $\Delta_{\rm s}^2 \approx 10^{-9}$. What value does this imply for λ?*

This model illustrates many of the features of generic inflationary models: (i) the field is of order – even greater than – the Planck scale, but (ii) the energy scale V is much smaller than Planckian because of (iii) the very small coupling constant.

Problem 9 (Plots of Bispectra)**.** *Reproduce the plots of the bispectra for the local and equilateral shapes (Figs. 83 and 84, respectively). Then plot the bispectra for slow-roll inflation, the DBI model and for models with excited initial states, i.e. $\mathcal{S}^{\rm SR}(1,x_2,x_3)$ (Eqn. (627)), $\mathcal{S}^{\rm DBI}(1,x_2,x_3)$ (Eqn. (642)) and $\mathcal{S}^{\rm folded}(1,x_2,x_3)$ (Eqn. (643)).*

13.2.6 Solution some problems by Dodelson chapter 2

Solution to

1 (a) To get from kelvin to eV, use $k_B = \text{eV}/(11605\text{ K})$. So $2.725\text{ K} \to k_B 2.725\text{ K} = (2.725/11605)\text{eV}$. Or $2.348 \times 10^{-4}\text{eV}$. (b) Since $T_0 = 2.348 \times 10^{-4}\text{eV}$,

$$\rho_\gamma = \frac{\pi^2 T_0^4}{15} = 2.000 \times 10^{-15}\text{eV}^4$$

To get this in gcm^{-3}, first divide by $(\hbar c)^3 = \left(1.973 \times 10^{-5}\text{eVcm}\right)^3$ to get 0.2604eVcm^{-3}. Then to change from eV to grams, remember that the mass of the proton is either 1.673×10^{-24} g or 0.9383×10^9eV, so $1\text{eV} = 1.783 \times 10^{-33}$ g. Therefore, $\rho_\gamma = 4.643 \times 10^{-34}$ g cm^{-3}.

(c) We have parametrized $H_0 = 100h$ kmsec^{-1}Mpc^{-1}, or using the fact that one Mpc is equal to 3.1×10^{19} km, $H_0 = 3.23h \times 10^{-18}\text{sec}^{-1}$. To get this into inverse cm, divide by the speed of light, $c = 3 \times 10^{10}$ cmsec^{-1}; then $H_0 = 1.1\text{ h} \times 10^{-28}$ cm. Or $H_0^{-1} = 9.3h^{-1} \times 10^{27}$ cm.

(d) To get the Planck mass $\left(1.2 \times 10^{28}\text{eV}\right)$ into kelvins, multiply by $k_B^{-1} = 11605\text{ K/eV}$; then $m_{\rm Pl} = 1.4 \times 10^{32}$ K. To get it into inverse cm, divide by $\hbar c = 1.97 \times 10^{-5}$eVcm to get $m_{\rm Pl} = 6.1 \times 10^{32}$ cm^{-1}. To get this is units of time, multiply by the speed of light to get $m_{\rm Pl} = 6.1 \times 10^{32} \times 3 \times 10^{10}$ cmsec^{-1}, or $m_{\rm Pl} = 1.8 \times 10^{43}\text{sec}^{-1}$.

Solution to

7
Start with

$$\Gamma^0_{\mu\nu} = \frac{g^{0\alpha}}{2}\left[\frac{\partial g_{\alpha\mu}}{\partial x^\nu} + \frac{\partial g_{\alpha\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial x^\alpha}\right]$$

where μ, ν range from 0 t 02,0 being the time index, 1 corresponding to θ, and 2 to ϕ. Since the metric is diagonal, $g^{0\alpha}$ is nonzero only when $\alpha = 0$ in which case it is -1. So

$$\Gamma^0_{\mu\nu} = \frac{-1}{2}\left[\frac{\partial g_{0\mu}}{\partial x^\nu} + \frac{\partial g_{0\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial t}\right]$$

All of these terms vanish: the first two since g_{00} is a constant, and the last because none of the metric elements depend on $x^0 = t$. So $\Gamma^0_{\mu\nu} = 0$ for all μ, ν.

Next consider

$$\Gamma^\theta_{\mu\nu} = \frac{g^{\theta\alpha}}{2}\left[\frac{\partial g_{\alpha\mu}}{\partial x^\nu} + \frac{\partial g_{\alpha\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial x^\alpha}\right]$$

Again since the metric is diagonal, and $g^{\theta\theta} = 1/r^2$, this reduces to

$$\Gamma^\theta_{\mu\nu} = \frac{1}{2r^2}\left[\frac{\partial g_{\theta\mu}}{\partial x^\nu} + \frac{\partial g_{\theta\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial \theta}\right].$$

Only the $g_{\phi\phi}$ component depends on one of our variables, so only it is nonzero when differentiated. Therefore, the first two terms vanish and the last is nonzero only when $\mu = \nu = \phi$, in which case it is

$$\Gamma^\theta_{\phi\phi} = \frac{1}{2r^2}\left[-r^2\frac{\partial \sin^2\theta}{\partial\theta}\right] = -\sin\theta\cos\theta.$$

Finally, when the upper index is ϕ, we have

$$\Gamma^\phi_{\mu\nu} = \frac{1}{2r^2\sin^\theta}\left[\frac{\partial g_{\phi\mu}}{\partial x^\nu} + \frac{\partial g_{\phi\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial \phi}\right]$$

The last term vanishes since none of the metric elements depend on ϕ; the first two are nonzero only if one of the indices μ, ν is equal to ϕ and the other is θ, so

$$\Gamma^\phi_{\phi\theta} = \Gamma^\phi_{\theta\phi} = \frac{\cos\theta}{\sin\theta}$$

The geodesic equation is

$$\frac{d^2x^\mu}{d\lambda^2} = -\Gamma^\mu_{\alpha\beta}P^\alpha P^\beta$$

with

$$P^\mu \equiv \frac{dx^\mu}{d\lambda}$$

Let's apply this to the $\mu = \theta$ component. The left-hand side is

$$\frac{d^2\theta}{d\lambda^2} = \frac{d}{\lambda}\frac{dt}{d\lambda}\dot{\theta} = E^2\ddot{\theta}$$

since $E = dt/d\lambda$ is constant. The Christoffel symbol on the right-hand side $\Gamma^\theta_{\alpha\beta}$ is nonzero only when $\alpha = \beta = \phi$ in which case it is $-\sin\theta\cos\theta$. So,

$$\ddot{\theta} - \sin\theta\cos\theta(\dot{\phi})^2 = 0$$

For the second equation, consider the ϕ component of the geodesic equation,

$$\frac{d^2\phi}{d\lambda^2} = -\Gamma^\phi_{\alpha\beta}P^\alpha P^\beta$$

Again the left-hand side is simply $E^2\ddot{\phi}$. The right-hand side gets nonzero contributions when $\alpha = \theta, \beta = \phi$ or an identical term when $\alpha = \phi, \beta = \theta$. Therefore,

$$\ddot{\phi} + 2\frac{\cos\theta}{\sin\theta}\dot{\theta}\dot{\phi} = 0$$

Incidentally this is equivalent to

$$\frac{d}{dt}\left(\dot{\phi}\sin^2\theta\right) = 0$$

and the conserved quantity in parentheses is the angular momentum.
The Ricci scalar is

$$\mathcal{R} = g^{\mu\nu}R_{\mu\nu} = -R_{00} + \frac{1}{r^2}R_{\theta\theta} + \frac{1}{r^2\sin^2\theta}R_{\phi\phi}$$

The time-time component vanishes since all Γ 's with time components are zero. We need to compute the two spatial components. First, consider

$$R_{\theta\theta} = \frac{\partial\Gamma^\alpha{}_{\theta\theta}}{\partial x^\alpha} - \frac{\partial\Gamma^\alpha{}_{\theta\alpha}}{\partial\theta} + \Gamma^\alpha{}_{\beta\alpha}\Gamma^\beta{}_{\theta\theta} - \Gamma^\alpha{}_{\beta\theta}\Gamma^\beta{}_{\theta\alpha}.$$

The first and third terms vanish since the Christoffel symbol with two lower θ 's vanishes. For the same reason, the index α in the second term must be equal to ϕ, and both β and α in the last term must equal ϕ:

$$R_{\theta\theta} = -\frac{\partial(\cos\theta/\sin\theta)}{\partial\theta} - \left(\frac{\cos\theta}{\sin\theta}\right)^2.$$

Carrying out the derivative then gives

$$R_{\theta\theta} = \left[1 + \frac{\cos^2\theta}{\sin^2\theta}\right] - \left(\frac{\cos\theta}{\sin\theta}\right)^2 = 1.$$

The other spatial component is

$$R_{\phi\phi} = \frac{\partial\Gamma^\alpha_{\phi\phi}}{\partial x^\alpha} - \frac{\partial\Gamma^\alpha_{\phi\alpha}}{\partial\phi} + \Gamma^\alpha_{\beta\alpha}\Gamma^\beta_{\phi\phi} - \Gamma^\alpha_{\beta\phi}\Gamma^\beta_{\phi\alpha}.$$

The Christoffel symbol in the first term is nonzero only if $\alpha = \theta$, while the one in the second term is always zero. In the third term β must be equal to θ to make the second Christoffel symbol be nonzero, and then $\alpha = \phi$. In the last term β can be θ and $\alpha = \phi$ or vice versa, so

$$R_{\phi\phi} = \frac{\partial\Gamma^\theta_{\phi\phi}}{\partial\theta} + \Gamma^\phi{}_{\theta\phi}\Gamma^\theta_{\phi\phi} - \Gamma^\phi{}_{\theta\phi}\Gamma^\theta_{\phi\phi} - \Gamma^\theta_{\phi\phi}\Gamma^\beta_{\phi\theta}.$$

The middle two terms cancel leaving

$$R_{\phi\phi} = -\frac{\partial(\sin\theta\cos\theta)}{\partial\theta} + \sin\theta\cos\theta\frac{\cos\theta}{\sin\theta}.$$

Carrying out the derivative gives

$$R_{\phi\phi} = -\cos^2\theta + \sin^2\theta + \cos^2\theta = \sin^2\theta.$$

Summing up, we get

$$\mathcal{R} = \frac{1}{2r^2}$$

The Ricci scalar is therefore a measure of the curvature of the space.

Solution to

9 Accumulating the various Γ 's leads to

$$\frac{d^2x^i}{d\lambda^2} = -2\frac{\dot{a}}{a}\frac{dt}{d\lambda}\frac{dx^i}{d\lambda}$$

Change to differentiation with respect to η using the facts that $dt/d\lambda = E$ and $d\eta/d\lambda = E/a$. Then the geodesic equation becomes

$$\frac{E}{a}\frac{d}{d\eta}\left(\frac{E}{a}\frac{dx^i}{d\eta}\right) = -2\frac{\dot{a}}{a}\frac{E^2}{a}\frac{dx^i}{d\eta}$$

Since $E/a \propto a^{-2}$, when the derivative on the left acts on E/a, the resulting term (proportional to $dx^i/d\eta$) exactly cancels the term on the right, leaving the result of Eq. (2.99).

Solution to

10 The age integral is

$$t(a) = \int_0^a \frac{da'}{a'H\left(a'\right)}.$$

Since we are assuming only matter and radiation, we can take

$$H(a) = H_0\sqrt{\rho/\rho_{\rm cr}} = \sqrt{\frac{1}{a^3} + \frac{\Omega_r}{a^4}}$$

where the $1/a^3$ term is from matter with density equal to the critical density. When the density in matter is equal to critical, $\Omega_r = a_{\rm eq} = 4.15 \times 10^{-5}h^{-2}$. Therefore, the age integral is

$$t = \frac{1}{H_0}\int_0^a \frac{da'a'}{\sqrt{a' + a_{\rm eq}}}$$

Integrate by parts to get

$$H_0 t = 2a\sqrt{a + a_{\rm eq}} - 2\int_0^a da'\sqrt{a' + a_{\rm eq}}.$$

Carrying out the last integral leads to

$$H_0 t = 2a\sqrt{a + a_{\rm eq}} - \frac{4}{3}\left\{\left[a + a_{\rm eq}\right]^{3/2} - a_{\rm eq}^{3/2}\right\}.$$

At very early times, such as when the temperature was $0.1{\rm MeV}$, a is much smaller than $a_{\rm eq}$, so

$$t \to \frac{a^2}{2H_0\sqrt{a_{\rm eq}}} \quad ; \quad a \ll a_{\rm eq}.$$

This limit is easiest to see directly in the integral of Eq. (A.39), but you can also get it by Taylor expanding Eq. (A.41). When the temperature is 0.1MeV, the scale factor is $2.35 \times 10^{-4}{\rm eV}/0.1{\rm MeV} = 2.35 \times 10^{-9}$, the temperature today divided by 0.1MeV. Plugging in numbers leads to

$$t(0.1{\rm MeV}) = 4.28 \times 10^{-16} \times 9.78 \times 10^9{\rm yr} = 130{\rm sec}.$$

At $T = 1/4{\rm eV}, a = 9.4 \times 10^{-4}$, significantly larger than $a_{\rm eq} = 8.5 \times 10^{-5}$ with $h = 0.7$, so $H_0 t \to (2/3)a^{3/2}$. So,

$$t(1/4{\rm eV}) = 270,000{\rm yr}$$

Solution to

12 The angle subtended is the physical distance divided by the angular diameter distance

$$\theta(z) = \frac{5\text{kpc}(1+z)}{\chi(z)}$$

In a flat, matter-dominated universe, χ is given by Eq. (2.43). When $z = 0.1$ (1), the term in brackets in Eq. (2.43) is equal to 0.0465(0.293). The comoving distance out to z is, therefore,

$$\chi = \begin{cases} 280h^{-1}\text{Mpc} & z = 0.1 \\ 1760h^{-1}\text{Mpc} & z = 1 \end{cases} .$$

Carrying out the division and converting radians to arcsec)1 radian equals 2.06×10^5 arcsec) leads to

$$\theta = \begin{cases} 4.0h'' & z = 0.1 \\ 1.2h'' & z = 1 \end{cases} .$$

In a universe with $\Omega_\Lambda = 0.7, \Omega_m = 0.3, \chi$ must be computed numerically. At $z = 1$, I find χ to be larger than in the flat, matter-dominated case by a factor of 1.3, so the angular size will be smaller by this factor, down to $0.9h''$. At $z = 0.1$ the difference in comoving distances is only 5%, so the angular size goes down to $3.8h''$ in the cosmological constant case.

Solution to

13 Rewriting Eq. (1.8) in terms of momentum $p = h\nu/c = 2\pi\hbar\nu/c$ and recognizing the denominator there as $1/f$ leads to

$$I_\nu = f\frac{4\pi p^3}{(2\pi)^3}$$

with $\hbar = c = 1$. So the energy density is the integral of this over all frequencies, with a factor of 4π to count photons from all directions (i.e., I_ν is per steradian):

$$\rho_\gamma = 4\pi \int_0^\infty d\nu I_\nu$$

This can be converted into an integral over momentum, with $d\nu = dp/(2\pi)$:

$$\rho_\gamma = 2 \int_0^\infty dp I_\nu.$$

Solution to

15 We want to compute $\rho = -T^0_0$. Setting $\mu = \nu = 0$ leads to

$$T^0_0 = -g_i \int \frac{dP_1 dP_2 dP_3}{(2\pi)^3} \left(-\det\left[g_{\mu\nu}\right]\right)^{-1/2} P^0 f_i$$

The matrix $g_{\mu\nu}$ is diagonal, so the determinant is simply the product of the diagonal elements, $-a^6$. By definition, $p^2 = g^{ij}P_iP_j = a^{-2}\delta_{ij}P_iP_j$. So $p_i \equiv \hat{p}_i p = P_i/a$ with $\hat{p}_i$ a unit vector pointing in the direction of the momentum. Therefore, $d^3P = a^3 d^3p$ and the factors of a precisely cancel those coming from the determinant. We're left with

$$T^0_0 = -g_i \int \frac{d^3p}{(2\pi)^3} P^0 f_i$$

The four vector P_μ squared is equal to $-m^2$, the mass of the particle, so $g_{00}\left(P^0\right)^2 = -m^2 - g_{ij}P^iP^j = -m^2 - p^2$. Since $g_{00} = -1, P^0 = \sqrt{p^2 + m^2}$, in accord with Eq. (2.59).

Solution to

17 The energy density of a massless boson is $g\pi^2T^4/30$, while that of a fermion is 7/8 times this. So,

$$s = \frac{2\pi^2}{45}\left[\sum_{i=\text{ bosons}} g_i T_i^3 + \frac{7}{8}\sum_{i=\text{ fermions}} g_i T_i^3\right]$$

accounting for the possibility that different species have different temperatures.

13.2.7 Solution some problems by Dodelson chapter 3

Solution to

1 The number density of a species with degeneracy $g = 2$ is

$$n = 2\int \frac{d^3p}{(2\pi)^3} f(p)$$

For the distributions we will consider, the phase space density f depends only on the magnitude of the momentum, so the angular part of the integral can be performed leading to the a factor of 4π; therefore,

$$n = \frac{1}{\pi^2}\int_0^\infty dp p^2 f(p)$$

First let's consider the high m/T limit. In this case, the limit of the Boltzmann distribution is $\exp\left[-\left(m + p^2/2m\right)/T\right]$. I claim, though, that this is precisely the limit of both the Fermi-Dirac and Bose-Einstein distributions:

$$\frac{1}{e^{E/T} \pm 1} \to e^{-E/T}$$

since $E \simeq m \gg T$ so that the exponential in the denominator dwarfs the 1. Therefore the low-temperature limit of all three distributions is

$$n^{\text{low T}} = \frac{e^{-m/T}}{\pi^2}\int_0^\infty dp p^2 e^{-p^2/2mT}$$

To do the integral, define a dimensionless parameter $x \equiv p/\sqrt{2mT}$. In terms of the variable $dpp^2 = [2mT]^{3/2} dx x^2$, so

$$n^{\text{low T}} = \frac{e^{-m/T}}{\pi^2}[2mT]^{3/2}\int_0^\infty dx x^2 e^{-x^2}$$

But the integral is equal to $\sqrt{\pi}/2$, so we have

$$n^{\text{lowT}} = 2e^{-m/T}\left(\frac{mT}{2\pi}\right)^{3/2}$$

The high-temperature Boltzmann limit is

$$n^{\text{HiT, Boltz}} = \frac{1}{\pi^2}\int_0^\infty dp p^2 e^{-p/T}.$$

Defining the dummy variable $x \equiv p/T$ leads to

$$n^{\text{HiT, Boltz}} = \frac{1}{\pi^2}T^3\int_0^\infty dx x^2 e^{-x}.$$

The x integral is equal to 2. So,

$$n^{\text{Hi T. Boltz}} = \frac{2T^3}{\pi^2}$$

The Bose-Einstein and Fermi-Dirac integrals similarly are

$$n^{\rm Hi}{\rm T,BE/FD} = \frac{T^3}{\pi^2}\int_0^\infty \frac{dx x^2}{e^x \mp 1}$$

The integrals can be written in terms of the Riemann zeta function, via Eq. (C.27). So the integral in Eq. (A.63) with the minus sign - the Bose-Einstein distribution-is $\zeta(3)\Gamma(3) = 2\zeta(3)$. The integral with the plus sign - the Fermi-Dirac distribution - is $3\zeta(3)\Gamma(3)/4 = 3\zeta(3)/2$, so

$$n^{\rm HiT} = \frac{\zeta(3)T^3}{\pi^2}\begin{cases} 2 & \text{Bose-Einstein} \\ 3/2 & \text{Fermi-Dirac} \end{cases}.$$

By the way, $\zeta(3) \simeq 1.202$, so there are more bosons than fermions for the same temperature, and these bracket the Boltzman amount. All of course are proportional to T^3.

Solution to

6 The photon number density is 411 cm^{-3}, while the baryon number density is $n_b = \rho_b/m_p = \rho_{\rm cr}\Omega_b/m_p$. Plugging in numbers gives

$$n_b = \Omega_b \frac{1.879h^2 \times 10^{-29}\ {\rm g\ cm^{-3}}}{1.673 \times 10^{-24}\ {\rm g}} = 1.12 \times 10^{-5}\Omega_b h^2\ {\rm cm^{-3}}.$$

So η_b, the ratio of the baryon to the photon number density, is indeed given by Eq. (3.11).

Solution to

11 To find this ratio, we compute the entropy density $(\mathcal{P}+\rho)/T$ at the two times. In both cases, only relativistic particles contribute to the entropy density significantly so that Eq. (A.53) holds. At high temperatures, the following particles contribute to the energy density: quarks ($g_* = 5 \times 3 \times 2$ for the five least massive types - up, down, strange, charm, bottom - each with three colors and two spin states); anti-quarks ($g_* = 30$ again); leptons ($g_* = 6 \times 2$ for the six types - $e, \nu_e, \mu, \nu_\mu, \tau, \nu_\tau$ - each with two spin states); anti-leptons ($g_* = 12$ again); photons (2); and gluons ($g_* = 8 \times 2$ for eight possible colors each with two spin states). This totals up to

$$g_* = 2 + 16 + \frac{7}{8}(30 + 30 + 12 + 12) = 91.5$$

The sixth quark, the top quark, does not contribute because it is too heavy to be around at these temperatures $m_t \simeq 175$GeV. Today entropy comes only from photons and neutrinos. The former contribute 2 to g_*; the latter contribute $(7/8)\times\ 3 \times 2 \times (4/11)^{4/3} = 1.36$, so today $g_* = 3.36$. Since the product sa^3 remains constant, we have
Therefore,

$$\left[g_*(aT)^3\right]\big|_{T=10{\rm Gev}} = \left[g_*(aT)^3\right]\big|_{T_0}$$

$$\frac{(aT)^3\big|_{T=10{\rm Gev}}}{(a_0T_0)^3} = \frac{3.36}{91.5} = \frac{1}{27}$$

13.2.8 Solution some problems by Dodelson chapter 4

Solution to

1 First integrate Eq. (4.6) over all momentum. This gives

$$\frac{\partial n}{\partial t} + \frac{\partial (nv)}{\partial x} = 0$$

the $\partial f/\partial p$ term vanishing after integrating by parts and noticing that $f = 0$ at $p = \pm\infty$ (there are no particles with infinite momentum). This is the continuity equation. To get the Euler equation, first multiply by p/m and then integrate over all momentum. This gives

$$\frac{\partial (nv)}{\partial t} + \frac{\partial}{\partial x}\int_{-\infty}^{\infty}\frac{dp}{2\pi}\frac{p^2}{m^2} + \frac{kx}{m}n = 0$$

where the last term follows from an integration by parts. The integral over p^2 yields two terms, one a bulk velocity term, v^2, and the second a pressure term, P. Using the continuity equation reduces this to

$$\dot{v} + v\frac{\partial v}{\partial x} + \frac{1}{n}\frac{\partial P}{\partial x} + \frac{kx}{m} = 0.$$

Solution to

4 From Eq. (3.3), the electron distribution function peaks at zero momentum, with a maximum value of $e^{(\mu - m_e)/T}$. To relate the chemical potential to the density, recall that $n = e^{\mu/T} n^{(0)}$, so in the low-temperature limit (Eq. (3.6)):

$$e^{\mu/T} = \frac{n_e}{2}\left(\frac{2\pi}{m_e T}\right)^{3/2} e^{m_e/T}.$$

So the maximum value of f_e is $(n_e/2)\,(2\pi/m_e T)^{3/2}$. Divide Eq. (3.44) by the Thomson cross-section to get $n_e = 1.12 \times 10^{-5}\Omega_B h^2$ cm^{-3} today including both ionized and captured electrons. Taking the electron temperature to be equal to the photon temperature today gives $2\pi/m_e T = 2.04 \times 10^{-11}$ cm^2. Putting back in the factors of a leads to

$$f_e^{\rm MAX} = 10^{-21}\Omega_B h^2 a^{-3/2}.$$

This expression holds only up to $T \leq m_e$, corresponding to $a \simeq 4.6 \times 10^{-10}$. So, as long as the temperature is well below the electron mass, f_e is very small.

Solution to

7 The difference between the amplitude we used in the derivation in Section 4.3 and the more accurate one given in the problem is $2\pi\sigma_T m_e^2[3\cos(\hat{p}.\ \hat{p}') - 1]$. The combination is square brackets is twice the second Legendre polynomial. Rewrite using the addition formula of spherical harmonics; then the difference becomes

$$2\pi\sigma_T m_e^2 \frac{8\pi}{5}\sum_{m=-2}^{2} Y_{2m}(\hat{p}) Y_{2m}^*(\hat{p}')\,.$$

This is the quantity we need to insert into the multiple integral in Eq. (4.49) in place of $\mathcal{M}^2$. When we do this, only the $m = 0$ term will contribute since all other $Y_{2m}(\hat{p}')$ have an azimuthal dependence which integrates to zero. Therefore, the new collision term due to anisotropic Compton scattering is

$$\delta C[f(\vec{p})] = \frac{\pi^2 n_e \sigma_T}{p}\mathcal{P}_2(\mu)\int \frac{d^3p'}{(2\pi)^3 p'}\mathcal{P}_2\left(\hat{\gamma}'\cdot\hat{k}\right) \times\left\{\delta\left(p-p'\right) + \left(\vec{p}-\vec{p}'\right)\cdot\vec{v}\frac{\partial\delta\left(p-p'\right)}{\partial p'}\right\}\left\{f\left(\vec{p}'\right) - f(\vec{p})\right\},$$

where I have used the fact that $Y_{20} = -\sqrt{5}\mathcal{P}_2/\sqrt{4\pi}$. The only term which survives the angular integral is the one proportional to $\delta\left(p-p'\right) f\left(\vec{p}'\right)$, leaving

$$\delta C[f(\vec{p})] = -\frac{n_e\sigma_T}{2p}\mathcal{P}_2(\mu)\int_0^\infty dp' p'\delta\left(p-p'\right)p'\frac{\partial f^{(0)}}{\partial p'} \times\int_{-1}^{1}\frac{d\mu}{2}\mathcal{P}_2(\mu)\Theta(\mu).$$

The angular integral gives $-\Theta_2$. Then integrating over the Dirac δ-function yields

$$\delta C[f(\vec{p})] = +p\frac{\partial f^{(0)}}{\partial p}\frac{n_e\sigma_T}{2p}\mathcal{P}_2(\mu)\Theta_2$$

This adds a factor of $-\mathcal{P}_2\Theta_2/2$ inside the square brackets of Eq. (4.54) and explains the corresponding factor in Eq. (4.100).

13.2.9 Solution some problems by Dodelson chapter 5

Solution to

4 In Fourier space,

$$\begin{aligned}\epsilon_{ijk}\left(\hat{k}_k\hat{k}_l-\delta_{kl}/3\right)G^L_{.jl} &= -k^2\epsilon_{ijk}\left(\hat{k}_k\hat{k}_j-\hat{k}_j\hat{k}_k/3\right)G^L\\ &= -2k^2/3\epsilon_{ijk}\hat{k}_j\hat{k}_kG^L=0\end{aligned}$$

since ϵ_{ijk} is antisymmetric under interchange of j and k while $\hat{k}_j\hat{k}_k$ is symmetric. The combination is also traceless since $\delta_{ij}\left(\hat{k}_i\hat{k}_j-\delta_{ij}/3\right)=0$.

Solution to

7 (a) By definition,

$$\Gamma^i_{jk}=\frac{g^{ii'}}{2}\left[g_{i'j,k}+g_{i'k,j}-g_{jk,i'}\right].$$

All derivatives here are spatial, and the only spatially varying part of the metric is the first-order piece $\mathcal{H}$. Therefore, we can again use the zero-order $g^{ii'}=\delta_{ii'}/a^2$, leaving Eq. (5.43).

(b) The product $\Gamma^\alpha_{\beta j}\Gamma^\beta_{i\alpha}$ vanishes when both indices α and β are zero (because $\Gamma^0_{0i}=0$) and when both indices are spatial (because then each Christoffel symbol is first order). Therefore, this product is

$$\begin{aligned}\Gamma^\alpha_{ij}\Gamma^\beta_{i\alpha} &= \Gamma^0_{hj}\Gamma^k_{i0}+\Gamma^k_{0j}\Gamma^0_{ik}\\ &= \Gamma^0_{kj}\Gamma^k_{i0}+(i\leftrightarrow j).\end{aligned}$$

But

$$\begin{aligned}\Gamma^0_{kj}\Gamma^k_{i0} &= \frac{1}{2}\left(2Hg_{jk}+a^2\mathcal{H}_{jk,0}\right)\left(H\delta_{ik}+\frac{1}{2}\mathcal{H}_{ik,0}\right)\\ &= (H)^2g_{ij}+a\frac{da}{dt}\mathcal{H}_{ij,0}.\end{aligned}$$

We must remember to add back in the same set of terms with i and j interchanged. This just introduces a factor of 2, so

$$\Gamma^\alpha_{\beta j}\Gamma^\beta_{i\alpha}=2(H)^2g_{ij}+2a\frac{da}{dt}\mathcal{H}_{ij,0}$$

13.2.10 Solution some problems by Dodelson chapter 11

Solution to 4

The noise matrix is

$$(C_N)_{ij}=\frac{1}{\bar{n}}\int d^3x\psi_i(\vec{x})\psi_j(\vec{x})$$

in this case of constant $\bar{n}$. Let's consider first the diagonal elements of the matrix. For these, both ψ_i and ψ_j require x to be within a radius R of the center of the i th cell, so

$$(C_N)_{ii}=\bar{n}\int\limits_{x<R}d^3x=\frac{4\pi\bar{n}R^3}{3}$$

(no sum over i intended). For cells separated by more than $2R$, the integral vanishes since $\vec{x}$ cannot be within a distance R of both cell centers. For distances less than $2R$ there is some overlap and the integral becomes

$$\bar{n}\int\limits_{x<R}d^3x\Theta(R-|\vec{x}-\vec{r}|)=2\pi\bar{n}\int\limits_0^R dxx^2\int\limits_{-1}^1 d\mu\Theta\left(R-\sqrt{x^2+r^2-2xr\mu}\right)$$

where $\vec{r}$ is the difference between the positions of the two cell centers and Θ is the step function equal to 1 if its argument is positive and zero otherwise. The μ integral therefore goes runs from $\left(x^2+r^2-R^2\right)/2xr$ up to 1. If this lower limit is greater than 1, then the μ integral vanishes; otherwise it is unity. The only contribution then comes when the lower limit is less than 1, which happens when x lies between $r \pm R$. The integral is therefore

$$\int_{r-R}^{R} dx x^2 \left[1 - \frac{x^2+r^2-R^2}{2xr}\right] = \frac{1}{2r}\int_{r-R}^{R} dx x \left[2xr - \left(x^2+r^2-R^2\right)\right].$$

The x integral here is then tedious but completely straightforward since the integrand is simply powers of x. I find that

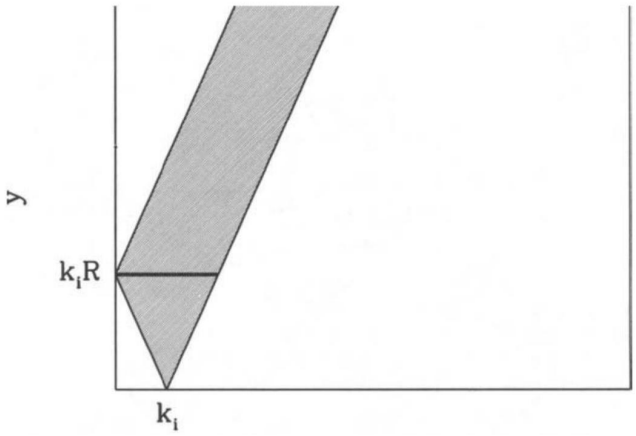
$$(C_N)_{ij} = \frac{\pi \bar{n} R^3}{12}\left(2 - \frac{r_{ij}}{R}\right)^2\left(4 + \frac{r_{ij}}{R}\right).$$

Solution to 7

We need to compute the integral of Eq. (11.55). Since the window function is sharply peaked for small scale modes, we can set k everywhere to k_i. Then inserting our explicit expression for the window function in a volume-limited survey (Eq. (11.59)), we are left with

$$(C_S)_{ii} \simeq \frac{9P(k_i)}{4\pi^2 R^2}\int_0^\infty dk \int_{|k-k_i|R}^{(k+k_i)R} \frac{dy}{y} j_1^2(y)$$

The best way to do the integrals here is to switch orders of integration. Consider the Figure A.2, which shows the region of integration. The region below the horizontal



k

Figure A.2. Region of integration for the double integral in Eq. (A.128). The region below the horizontal line constitutes the first term, above the second term.

line corresponds to $y < k_i R$ and $k_i - y/R < k < k_i + y/R$. In the region above k is bounded by $y/R \pm k_i$. Therefore,

$$\int_0^\infty dk \int_{|k-k_i|R}^{(k+k_i)R} \frac{dy}{y} j_1^2(y) = \int_0^{k_i R} \frac{dy}{y} j_1^2(y) \int_{k_i - y/R}^{k_i + y/R} dk + \int_{k_i R}^\infty \frac{dy}{y} j_1^2(y) \int_{y/R - k_i}^{k_i + y/R} dk$$
$$= \frac{2}{R}\int_0^{k_i R} dy j_1^2(y) + 2k_i \int_{k_i R}^\infty \frac{dy}{y} j_1^2(y).$$

In the limit $k_i R \gg 1$, the first term here is much larger than the second (since $j_1^2(y)$ goes as $1/y^2$ for large y). In the first integral, we may replace the upper limit by infinity, again since $k_i R$ is large. The resulting integral is (Eq. (C.17)) $\pi\Gamma(3/2)/4\Gamma(5/2) = \pi/6$. Multiplying this by $2/R$ and then by $9/4\pi^2 R^2$ leads to a factor of $1/V$.

Solution to 9

We want to compute the variance

$$\left\langle (\hat{\lambda} - \bar{\lambda})^2 \right\rangle = \left\langle \left(F^{-1}_{\lambda\lambda} \frac{\Delta C^{-1} C_{,\lambda} C^{-1} \Delta - \mathrm{Tr}\left[C^{-1}\mathrm{C}_{,\lambda}\right]}{2} \right)^2 \right\rangle$$

where the estimator $\hat{\lambda}$ is given by Eq. (11.92), and I have assumed that $\lambda^{(0)} = \bar{\lambda}$, i.e., we are at the true maximum of the likelihood function. Upon squaring there are terms with no Δ; those with two Δ 's; and those with four. The ones with two Δ 's can be simply evaluated by using

$$\langle \Delta_i \Delta_j \rangle = C_{ij}$$

where the indices label the pixels. Since the distribution is assumed Gaussian, the expectation value of four Δ 's is

$$\langle \Delta_i \Delta_j \Delta_k \Delta_l \rangle = C_{ij}C_{kl} + C_{ik}C_{jl} + C_{il}C_{jk}$$

Putting in these expectation values leads to

$$\left\langle (\hat{\lambda} - \bar{\lambda})^2 \right\rangle = \frac{F^{-2}_{\lambda\lambda}}{4} \left\{ \left[C^{-1}C_{,\lambda}C^{-1}\right]_{ij} \left[C^{-1}C_{,\lambda}C^{-1}\right]_{kl} (C_{ij}C_{kl} + C_{ik}C_{jl} + C_{il}C_{jk}) - \left(\mathrm{Tr}\left[C^{-1}C_{,\lambda}\right]\right)^2 \right\}.$$

The $C_{ij}C_{kl}$ terms lead to $\left(\mathrm{Tr}\left[\mathrm{C}^{-1}\mathrm{C}_{,\lambda}\right]\right)^2$, cancelling the similar term on the last line, so

$$\left\langle (\hat{\lambda} - \bar{\lambda})^2 \right\rangle = \frac{F^{-2}_{\lambda\lambda}}{4} \left[C^{-1}C_{,\lambda}C^{-1}\right]_{ij} \left[C^{-1}C_{,\lambda}C^{-1}\right]_{kl} (C_{ik}C_{jl} + C_{il}C_{jk})$$

Both terms here contribute identically (giving one factor of 2). The matrix multiplication simplifies since all matrices are symmetric. For example,

$$\left[C^{-1}C_{,\lambda}C^{-1}\right]_{ij} \left[C^{-1}C_{,\lambda}C^{-1}\right]_{kl} C_{ik}C_{jl} = \mathrm{Tr}\left[\mathrm{C}^{-1}\mathrm{C}_{,\lambda}\mathrm{C}^{-1}\mathrm{C}_{,\lambda}\right]$$

and we recognize the right-hand side as $2F_{\lambda\lambda}$ (another factor of 2). Therefore,

$$\left\langle (\hat{\lambda} - \bar{\lambda})^2 \right\rangle = F^{-1}_{\lambda\lambda}$$

It is important to keep in mind that this equality holds only if the overdensities are distributed as Gaussians and if we truly have reached the point in parameter space which is the true maximum.

Solution to

15 (a) Use a likelihood approach. The likelihood function for the parameters, here the amplitudes of the different components Θ^α, is proportional to $e^{-\chi^2/2}$ with

$$\chi^2 = (d - W\Theta)N^{-1}(d - W\Theta) + \sum_{\alpha=1}^{N_{\mathrm{foregrounds}}} (\Theta^\alpha)^2 / C^\alpha$$

The first term is identical to that generated by Eq. (11.144). The second accounts for the prior, that Θ^α has mean zero and variance C^α. Maximizing the likelihood corresponds to minimizing the χ^2, with respect to the parameters Θ. Since the χ^2 is quadratic, the minimization leads to a linear equation for Δ^α, the minimum variance estimator of Θ^α:

$$\Delta = \left(WN^{-1}W + C^{-1}\right)^{-1} WN^{-1}d$$

The new covariance matrix is the first term on the right,

$$C_N = \left(WN^{-1}W + C^{-1}\right)^{-1}$$

Here the matrix C is diagonal with $_{00}$ element equal to zero, and the other diagonal elements equal to the assumed power spectra of the foregrounds.

(b) If there is only one foreground with shape vector $W^1 = (1, 1/2)$, and if this foreground has assumed power equal to the noise, then the new inverse covariance matrix goes from that in Eq. (11.151) to the same matrix with $1/\sigma_n^2$ added to the 11 component. Thus,

$$C_N^{-1} = \frac{1}{\sigma_n^2} \begin{pmatrix} 2 & 3/2 \\ 3/2 & 9/4 \end{pmatrix}$$

Note that the 0_0 component of this is unchanged, as it must be since it is the inverse covariance if all foregrounds are known. The inverse of this gives the new covariance matrix,

$$C_N = \frac{4\sigma_n^2}{9}\begin{pmatrix} 9/4 & -3/2 \\ -3/2 & 2 \end{pmatrix}$$

Immediately, we see that the noise in the presence of foregrounds is σ_n. This is a factor of $\sqrt{5}$ smaller than if we had no prior knowledge of the foreground amplitude. It is only a factor of $\sqrt{2}$ larger than the case without foregrounds; thus the new FDF in this case is $\sqrt{2}$. The minimum variance estimator is

$$\begin{aligned}\Delta^0 &= \frac{4\sigma_n^2}{9}(9/4, -3/2)\begin{pmatrix} 1 & 1 \\ 1 & 1/2 \end{pmatrix}\begin{pmatrix} 1/\sigma_n^2 & 0 \\ 0 & 1/\sigma_n^2 \end{pmatrix}\begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \\ &= \frac{d_1 + 2d_2}{3}.\end{aligned}$$

14 Extra Problems about Basic Models

(some not very important problems are here)

14.1 General questions

14.1.1 Short problems for checking basics of modern cosmology

(многое можно спросить, вот и напишу потом)

Typical questions about geometry and dynamics

(a) What is conformal time? Why is it useful?

(b) How do the energy densities in radiation (ρ_r), matter (ρ_m) and a cosmological constant (ρ_Λ) evolve with the scale factor $a(t)$?

(c) What is $a(t)$ for a flat universe dominated by radiation, matter or a cosmological constant?

(d) What is the redshift of matter-radiation equality if $\Omega_r = 9.4 \times 10^{-5}$ and $\Omega_m = 0.32$?

(e) What is H_0^{-1} in sec and in cm?

Typical questions about thermal history

(c) Describe $g_\star(T)$ given the particle content of the Standard Model.

(d) Why does the neutrino temperature scale as $T_\nu \propto a^{-1}$ after decoupling?

(e) Why is today's photon temperature larger than that of neutrinos?

(f) Why is the recombination temperature much lower than the ionization energy of hydrogen?

Typical questions about structure formation

(a) How do density perturbations in a pressureless fluid grow (i) in a static space and (ii) in an expanding space?

(b) What is the gauge problem?

(c) What are adiabatic fluctuations?

(d) Explain the Meszaros effect.

(e) Explain the relevance of the conservation of the comoving curvature perturbation $\mathcal{R}$ on superhorizon scales.

14.1.2 Short questions and problems for knowledge checking

(!!! later I'll make sure that all these questions are answered in 1st part or in theory)

Discuss the methods to achieve an accurate determination of the Hubble parameter

Discuss the results of the WMAP satellite. What else can PLANCK add to this?

How did it come to be that a significant mass of the universe is made up of dark matter?

Use the action to derive the equation of motion of the scalar field $phi(t, x^{dot})$ in a flat LFRW universe.What does the equation of motion reduce to if V(phi) is homogenous.

Inflation predicts a nearly scalar invarious spectrum of density pertubations. Give the 2 main reasons why this is a very important result.

Discuss the different evidences for dark matter

In which way can one derive different cosmological parameters from the frequency and intensity of the CMB?

Speeds (???)

The max velocity of matter is c, and 13.7 Gyr ago everything was together in the big bang. The radiation of the CMB travels with the velocity of light. Why is it possible that we can see it, and that it did not already pass us? You may chose in which way you want to explain this to or a 10 year old kid, a student of a different master or a 75 old person.

Use the formula for f(x), the Gaussian expectation value of 2 fourier coefficients to calculate the 2 point correlation function. (Both equations were given)

Derive the formula for the redshift

Why is the observation of the cosmological constant a relief for astrophysicists and a problem for theoretical physicists?

(??? do we observe it???)

How can the He abundance tell us about the conditions of the early Universe?

Short questions about homogeneous inflation

What is the horizon problem? How does inflation solve it?

How can we prove that indeed there was an inflation?

What are the conditions for successful slow-roll inflation?

Why do we need slow-rolling of inflation?

Slow roll inflation is quantified in terms of a sequence of parameters which measure the local smoothness of the inflaton potential. The lowest-order parameter is $\epsilon \propto (V'/V)^2$, the next-lowest is $\eta \propto V''/V$, and so on. Higher-order parameters measure ever more localized kinks in the potential.

The values of ϵ, η, etc are constrained by a few different things. For one, inflation needs to last long enough to solve the horizon and flatness problems. Inflation occurs as long as $\epsilon < 1$, and its rate of change is

$$\frac{\mathrm{d}\epsilon}{\mathrm{d}N} = 2\epsilon(\eta - \epsilon)$$

where N is a time variable denoting the number of e-folds before the end of inflation: current estimates require that ΔN be at least around 60. So if ϵ is too large (the potential has too steep a grade), then the field rolls down too quickly to drive sufficient inflation. Also, if η is too large (the potential has too much curvature) then a similar fate awaits. So the first reason we want inflation to be "slow roll" is that we want to get enough inflation, and we need a reasonably flat potential to achieve that.

The other big reason that slow roll inflation is desired is due to the influence that inflationary dynamics have on the shape of the density perturbation spectrum, $P(k)$. Measurements of the cosmic microwave background and large scale structure surveys have revealed that $P(k)$ is very nearly scale invariant, $P(k) \sim k^{n-1}$ with $n \approx 1$. A lengthy and involved (though fun and rewarding) calculation reveals that we can connect the spectral parameters directly to the inflationary dynamics in terms of the slow roll parameters. In particular,

$$n - 1 = 4\epsilon - 2\eta,$$

revealing that slow roll inflation leads to a nearly scale invariant power spectrum.

None of this is to say that there cannot be non-slow roll behavior during inflation, just that it needs to be the exception rather than the rule. There are many models involving localized, transient kinks in the potential that create interesting features in the spectrum on specific scales. People have also looked at "fast roll" inflation in which, though the potential may be flat, the field velocity is high compared to the expansion rate. This is also non-slow roll behavior and one finds that the resulting power spectrum deviates from a power law on the affected scales (fast rolling fields tend to slow down even as they roll down hill during inflation, due to Hubble drag, and so slow roll inflation can be achieved even with initially fast-rolling fields).

Lastly, we only have observations across a limited range of scales (maybe the first dozen or so e-folds of inflation), and so the potential could do all sorts of non-slow roll tricks on these scales (both larger and smaller) as long as sufficient inflation is achieved and there aren't dramatic spikes in the power spectrum (that might, for example, generate unacceptable amounts of primordial black holes.)

Can we produce the CMB power spectrum from a given model of inflation, theoretically? If yes, how?

Absolutely! This is how we learn about the physics of inflation. Given an inflaton potential, we can work out the spectrum of density perturbations it generates. With this power spectrum (plus several so-called "late-time" parameters), we can compute the expected CMB temperature and polarization anisotropies. This is a very involved, multi-step analysis so I will only highlight the main parts here.

Given the scalar inflaton field with potential $V(\phi)$, we write the field as a perturbation about a homogeneous value, $\phi(\mathbf{x}, t) = \phi_0(t) + \delta\phi(\mathbf{x}, t)$, drop it into the Klein Gordon Equation, and obtain,

$$\ddot{\delta\phi} + 3H\dot{\delta\phi} - \left(\frac{\nabla^2}{a^2} - \left.\frac{\mathrm{d}^2 V(\phi)}{\mathrm{d}\phi^2}\right|_{\phi=\phi_0}\right)\delta\phi = 0.$$

Moving to Fourier space gives

$$\ddot{\delta\phi_k} + 3H\dot{\delta\phi_k} + \left(\frac{k}{a}\right)^2 \delta\phi_k = 0.$$

This equation has no general analytic solution, but we can solve it exactly for de Sitter expansion ($H = const$) and then add perturbations to the solution to study the quasi-de Sitter expansion observed during real inflation (so-called *slow roll* inflation, in which time derivatives of H are increasingly suppressed at higher-order). The quantity of interest is the variance (or mean squared amplitude) of the fluctuation,

$$\langle|\delta\phi_k|^2\rangle = \frac{H^2}{2k^3}.$$

This is the variance of the field on horizon scales, where the quantum fluctuation is becoming classical. The variance of the full fluctuation also includes a piece from the *classical* evolution of the field, which is $H^2/\dot{\phi}^2$, giving

$$\langle|\delta_k|^2\rangle = \frac{H^2}{\dot{\phi}^2}\langle|\delta\phi_k|^2\rangle = \frac{H^4}{2\dot{\phi}^2 k^3}.$$

From the total variance, we can form the *power spectrum*, defined

$$P(k) = \frac{k^3}{2\pi^2}\langle|\delta_k|^2\rangle = \frac{1}{4\pi}\frac{H^4}{\dot{\phi}^2}.$$

Notice that the k-dependence is not explicit in this expression: it actually arises from the time-dependence of the cosmological quantities H and $\dot{\phi}$. As modes evolve in time, their amplitudes freeze once they cross the horizon, $k = aH$, and so this expression is to be evaluated at that time. Once we have the power spectrum across the scales of interest, we use it to "seed" the density perturbations in the baryon-photon plasma. These perturbations are what are observed as anisotropies in the temperature and polarization of the CMB.

Since the observed CMB is influenced by more than simply the initial spectrum (the growth of the perturbations from their birth up until last scattering depends on things like the baryon and CDM densities, for example), we must specify several more quantities to obtain the final CMB spectrum. Thankfully, there is publicly-available software that does this (e.g. CAMB). So, for a given initial spectrum $P(k)$ and values for these other quantities, we obtain a set of CMB spectra we can use to compare our model with observations.

Slow roll inflation gives rise to a *power law* spectrum,

$$P(k) = P(k_0)\left(\frac{k}{k_0}\right)^{n-1},$$

where the various parameters can be written in terms of the inflaton potential and field. For example, the *spectral index*, which governs the *tilt* of the spectrum, is determined by the first and second derivatives of the potential,

$$n - 1 = \frac{m_{\rm Pl}^2}{4\pi}\left[\frac{3}{2}\left(\frac{V'}{V}\right)^2 - \frac{V''}{V}\right].$$

This is an example of how a determination of the spectral index (inferred through the CMB anisotropies) can be used to learn about the shape of the inflaton potential. Other spectral parameters, like the overall amplitude and higher-order k-dependence of the spectrum, tell us about the height of the potential and the higher-order ϕ-terms in its Taylor expansion, respectively.

This has been a very high-level and terse answer, since easily an entire textbook could be written that addresses just this topic.

Can there be a non-slow rolling inflation?

What is the value of the pressure in the slow-roll model? Compare to the value in the regular model.

(?? we need negative pressure (? why?) but it is not only in SR, right???)

Short questions about inhomogeneous inflation and initial conditions

Why is the power spectrum of the perturbed scalar field in slow roll inflation related to the power spectrum of density perturbations? (Use equations to proof this)

Short questions about thermal history

Short questions about structure formation

Short questions about anisotropies and CMB

(i) Why do the simplest inflation models predict nearly Gaussian-like perturbations?

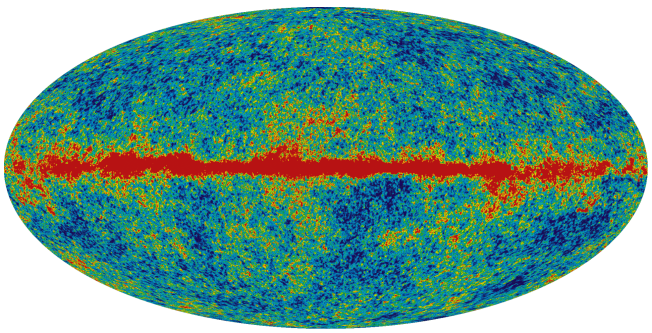
(ii) Given a field $f(\mathbf{x})$ and its Fourier expansion with modes $f_{\mathbf{k}}$, with $f_{\mathbf{x}} = a_{\mathbf{k}} + ib_{\mathbf{k}}$, and given the distribution $p\,(a_{\mathbf{k}}, b_{\mathbf{k}})$, argue why p is homogeneous and isotropic.

(iii) For the temperature anisotropy in a given direction, the integrated Sachs-Wolfe term is given by the integral $\frac{\delta T}{T}(\hat{p}) = \int_{\eta_{\mathbf{dec}}}^{\eta_0} (\dot{\psi} - \dot{\phi}) d\eta$. Explain all the symbols in the right hand side. Naively, one might think that the integral over a derivative reduces to the contribution from the two boundaries. Discuss why this is not the case.

(iv) Discuss the significance of the polarization spectra to determine the optical depth to recombination.

Why is statistics necessary to compare the CMB to theoretical predictions?

Below you see a plot of CMB. What should we do to get some information about the Universe?



14.1.3 Problems to Check Deep Understanding of Cosmology

(пока не до этого, в теории еще писать про это много когда-то буду)

14.1.4 Typical general questions from curious people

(все это написано в теории, тут просто предложение попрактиковаться и написать качественные, интересные и короткие ответы на эти вопросы.)
https://www.astro.ucla.edu/~wright/cosmology_faq.html#RB

What is the currently most accepted model for the Universe?

What is the evidence for the Big Bang?

What happened during the Big Bang?
What is this "anti-gravity"? [The cosmological constant]
Why do we think that the expansion of the Universe is accelerating?
What is quintessence?
How old is the Universe?
If the Universe is only 14 billion years old, why isn't the most distant object we can see 7 billion light years away?
If the Universe is only 14 billion years old, how can we see objects that are now 47 billion light years away?
Is the Universe really infinite or just really big?
How can the Universe be infinite if it was all concentrated into a point at the Big Bang?
How can the oldest stars in the Universe be older than the Universe?
Can objects move away from us faster than the speed of light?
What is the redshift?
Are quasars really at the large distances indicated by their redshifts?
What about objects with discordant redshifts, like Stephan's Quintet?
Has the time dilation of distant source light curves predicted by the Big Bang been observed?
Are galaxies really moving away from us or is space just expanding?
Why doesn't the Solar System expand if the whole Universe is expanding?
Is the Universe expanding or is it just that our definitions of length and time are changing?
Why haven't the CMBR photons outrun the galaxies in the Big Bang?
Where was the center of the Big Bang?
What is meant by a flat Universe?
Is the Big Bang a Black Hole?
What is the Universe expanding into?
What came before the Big Bang?
Doug Scott's Cosmic Microwave Background Radiation (CMBR) FAQ
Can the CMBR be redshifted starlight?
Why is the sky dark at night?
Will the Universe expand forever or recollapse?
Does entropy prevent a Big Crunch?
What about the oscillating Universe?
What is the dark matter?

What about MOND?

What is the value of the Hubble constant?

What can a layperson do in cosmology?

15 Problems about Special Modesl

15.0.1 Problems about Weak Lensing and Polarization 1

D-1.

The probability that there will be a galaxy massive enough to act as a lens between a quasar at redshift z and us is roughly proportional to the volume between us. Compute

$$V(z) := \int_{x<\chi(z)} d^3x$$

The integral is trivial, but the dependence on z is not. Numerically compute $V(z)$ in a flat universe with cosmological constant. Plot $V(z)$ vs Ω_Λ for $z = 2, 3, 4$. If the galaxy density does not depend on cosmology, then the expected number of lenses scales simply as this volume. For $z = 3$, what is the ratio of lenses expected in a flat universe with $\Omega_\Lambda = 0.7$ as compared with a flat, matter-dominated universe? Be warned that Keeton (2002), among others, has made a strong case that the expected lensing frequency does not vary this dramatically with cosmology because of differences in the galaxy densities in the different models.

D-2.

Compute the magnification of an image in terms of the convergence κ and shear γ_1 and γ_2. Show that, in the limit of weak fields, the magnification μ is related to the convergence via

$$\mu \simeq 1 + 2\kappa$$

D-3.

It is often useful to write observable properties of lenses - such as deflection angles and shear - in terms of a projected potential ϕ.

(a) Using Eq. (10.14), determine the ϕ such that $\vec{\theta}_S = \vec{\theta} + \nabla\phi$ where ∇ is the gradient with respect to the 2D angular variable $\vec{\theta}$.

(b) Express the transformation matrix defined in Eq. (10.15) in terms of the projected potential.

D-4.

When the lens is at a fixed redshift (e.g., a single galaxy or a cluster as opposed to large-scale structure in general) z_L corresponding to comoving distance χ_L from us, show that the projected potential ϕ from the previous problem reduces to

$$\phi\left(\vec{\theta}; z_L\right) = 4G\frac{\chi_S - \chi_L}{\chi_S\chi_L}\int d^2R\Sigma(\vec{R})\ln\left|\vec{R} - \chi_L\vec{\theta}\right|$$

Here χ_S is the comoving distance out to the source; $\vec{R}$ is the radius in the plane perpendicular to the line of sight; and $\Sigma(\vec{R})$ is the projected surface density in this plane.

D-5.

Compute the observed component of ellipticity ϵ_2 from an intrinsically circular source; express it in terms of the components of the transformation matrix, κ, γ_1, and, most importantly, γ_2.

D-6.

(a) Show that the power spectrum of the convergence is given by Eq. (10.35). Show that the power spectrum of γ_2 is given by Eq. (10.36).

(b) Using CMBFAST or the BBKS transfer function, compute numerically P_κ for standard CDM with $\Omega_m = 1, h = 0.5, n = 1$. Assume all background galaxies are at redshift $z = 1$. At what l do you expect your result - based on the linear power spectrum - to lose validity due to nonlinear effects?

D-7.

Equation (10.34) gives the power spectrum of the convergence in the small angle limit ($l \gg 1$). The more general expression is (Stebbins, 1996, the extra factor of $(2\pi)^3$ here due to differing power spectrum conventions)

$$C_{\kappa,l} = 4\pi l^2(l+1)^2\int\frac{d^3k}{(2\pi)^3}k^2I_l^2(k\chi)P_\Phi(k)$$

when all background galaxies are at comoving distance χ and

$$I_l(x) := \int_0^1\frac{dy}{y}(1-y)j_l(xy)$$

(a) Verify, either analytically or numerically, that in the small-angle limit Eq. (10.102) reduces to the expression for $P_\kappa(l)$ in Eq. (10.34).

(b) Redo the calculation of the convergence power spectrum for the sCDM model of Exercise 6, this time using the general expression in Eq. (10.102).

D-8.

In the text, we computed the angular correlations of galaxies (Chapter 9) and the weak lensing correlation function (this chapter). One can also compute the cross-correlation, which measures how correlated the galaxies are with the mass. One way to measure this is to separate a galaxy sample into foreground and background galaxies and measure the cross-correlation between the two samples. Since they are separated by such large distances, the only possible correlation arises because the background galaxies have been magnified by the foreground mass. This problem allows you to work out the background/foreground correlation function (e.g., Moessner and Jain, 1998). Incidentally, this cross-correlation function can also be measured by the QSO/galaxy correlation function. Suppose the 2D overdensity of foreground galaxies is due solely to intrinsic inhomogenities, so that it is given by Eq. (9.3). Assume that the 2D overdensity of background galaxies arises only from magnification. That is, galaxies that should not be included in the survey because they are intrinsically fainter than the magnitude limit are magnified and so appear brighter, thereby making the cut. If the magnification is μ, then the number of background galaxies in an angular patch is

$$n_b = \bar{n}_b\mu^{2.5s-1}$$

Here $\bar{n}_b$ is the average number of background galaxies, and s is defined as $d\log N(m)/dm$ where $N(m)$ is the number of galaxies at the magnitude limit m. For the present problem, don't worry about where this relation comes from (see Broadhurst, Taylor, and Peacock, 1995, for an explanation).

(a) Express δ_b, the background overdensity, in terms of κ and s using Eq. (10.100). (b) Find an expression for the convergence $\kappa(\vec{\theta})$ in terms of the mass overdensity. First express it in terms of the relevant components of transformation matrix A of Eq. (10.17), but then eliminate the potential there in favor of the density field δ.

(c) Using these two expressions - Eq. (9.3) and your answer in (b) - for the foreground and background overdensities, compute the angular cross-correlation function $w_{bg}(\theta) := \langle\delta_b(\theta)\delta_g(0)\rangle$

D-9.

As the wavevector $\vec{k}$ moves out of the $\hat{x}-\hat{z}$ plane, show that the Q polarization (for outgoing radiation in the $\hat{z}$-direction) changes as $\cos(2\phi_k)$. To do this, first compute $\hat{k}\cdot\hat{n}'$, and then integrate $\mathcal{P}_2\left(\hat{k}\cdot\hat{n}'\right)$ over solid angle, with the weighting factor $\sin^2\theta'\cos(2\phi')$ derived in Eq. (10.52).

15.0.2 Problems about Weak Lensing and Polarization 2

D-10.

This problem focuses on the U-component of polarization.

(a) We showed that the Q-component of polarization from unpolarized incident radiation is given by Eq. (10.52), which stems from Eq. (10.50). The Q-component thus depends on the difference between $|\hat{\epsilon}_i\cdot\hat{x}|^2$ and $|\hat{\epsilon}_i\cdot\hat{y}|^2$. For the U-component, $\hat{x}$ and $\hat{y}$ must be replaced by unit vectors rotated 45°, i.e., $(\hat{x}+\hat{y})/\sqrt{2}$ and $(\hat{x}-\hat{y})/\sqrt{2}$. With this replacement, derive Eq. (10.53).

(b) Show that a plane-wave perturbation with wavevector $\vec{k}$ lying in the $\hat{x}-\hat{z}$ plane does not produce any U-polarization in the outgoing $\hat{z}$-direction.

(c) For the most general orientation of the wavevector, $\hat{k} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$, show that U-polarization is given by Eq. (10.60).

D-11.

Draw the polarization patterns near the z-axis arising from a planewave scalar perturbation with (a) $\theta_k = \pi/8, \phi_k = \pi/8$; (b) $\theta_k = 3\pi/4, \phi_k = \pi/4$; (c) $\theta_k = 3\pi/4, \phi_k = 0$; and (d) $\theta_k = 3\pi/2, \phi_k = 0$. In each case, show that the sense of polarization is aligned with (or perpendicular to) the direction in which the polarization strength is changing.

D-12.

In the tight coupling limit, find an expression for $\Pi := \Theta_2 + \Theta_{P2} + \Theta_{P0}$.

(a) When $\dot{\tau}$ is very large, the terms multiplying it on the right hand side of Eq. (10.63) must cancel. Write down this equality for $\Theta_P(\mu)$ in terms of the moments, Θ_2, Θ_{P2}, and Θ_{P0}.

(b) Expand $\Theta_P(\mu)$ in terms of Legendre polynomials, keeping only the monopole and the quadupole. Then equate the coefficients of $\mathcal{P}_0$ and $\mathcal{P}_2$.

(c) This leads to two equations for three unknowns. Show that solving for the two polarization moments in terms of the temperature quadrupole gives $\Theta_{P0} = 5\Theta_2/4$ and $\Theta_{P2} = \Theta_2/4$

(d) Use the results of (c) to determine Π in terms of Θ_2.

D-13.

This problem concerns the rotation matrix R given in Eq. (10.91). (a) Act with R on the unit vector $(0,0,1)$ and show that it gets transformed into $\hat{k}$.

(b) Derive R. One way to do this is to first rotate the $x-y$ - z frame about the z-axis by an angle ϕ'. Then, rotate about the y-axis by an angle $-\theta'$. The product of these two rotations is R.

D-14.

In the text we considered polarization patterns from a single planewave perturbation due to gravity waves. There are actually two such orientations. We considered only $h_\times$. In this problem, consider polarization from h_+. In a frame in which $\hat{k}$ is along the z-axis the anisotropies have a $\sin^2\theta\cos(2\phi)$ dependence. (a) Find the dependence of Θ^T on angle in the more general frame in which $\hat{k}$ does not lie along the z-axis.

(b) Determine Q and U from this incoming distribution. (c) Plot the anisotropy pattern near the outgoing $\hat{z}$-direction for the four sets of θ_k, ϕ_k shown in Figure 10.19.

D-15.

Find expressions for the cross-correlation spectra between the temperature anisotropy and polarization anisotropy $C_{TQ}(\vec{l})$ and $C_{TU}(\vec{l})$ in terms of Θ_l and Θ_{Pl}. Assume scalar perturbations only. Express $C_{TE,l}$ in terms of these.

Solution some problems by Dodelson chapter 10

Solution to

2 The total intensity received at the detector is the angular integral of $I_{\rm obs}\,(\theta)$ over θ. The total intensity emitted is the angular integral of $I_{\rm true}\,(\theta_S)$ over θ_S. The magnification μ is the ratio of the two:

$$\mu \equiv \frac{\int d^2\theta I_{\rm obs}(\theta)}{\int d^2\theta_S I_{\rm true}\,(\theta_S)}$$

Change variables in the denominator to θ, leading to a factor of $\det(A)$ where A is defined in Eq. (10.15). Recall now that $I_{\rm true}\,(\theta_S) = I_{\rm obs}\,(\theta)$, so except for the determinant, the numerator and denominator cancel. This leaves

$$\mu = \frac{1}{\det(A)} = \frac{1}{(1-\kappa)^2 - (\gamma_1^2+\gamma_2^2)}.$$

If all the perturbations are small, then the magnification depends only on κ:

$$\mu \simeq 1 + 2\kappa.$$

Solution to

3 (a) Reading off from Eq. (10.14), we see immediately that

$$\phi = 2\int_0^{\chi S} d\chi \frac{\chi_S - \chi}{\chi_S \chi}\Phi(\chi\vec{\theta},\chi)$$

where I have let $\chi \to \chi_S$ in Eq. (10.14) and replaced the dummy variable χ' there with χ. The only subtlety here is the extra factor of χ in the denominator. This comes from changing the derivative with respect to position (the comma in Eq. (10.14)) to an angular derivative.

Solution to

4 Recall that, in the Newtonian limit, the gravitational potential can be written in terms of the mass density:

$$\Phi(\vec{x}) = -G\int \frac{d^3x'}{|\vec{x}-\vec{x}'|}\rho\,(\vec{x}')$$

We will do this integral in cylindrical coordinates, so that $\vec{x}' = \left(\vec{R},\chi'\right)$. Thus,

$$\phi \simeq -2G\frac{\chi_S - \chi_L}{\chi_S\chi_L}\int d^2R\int d\chi'\rho\left(\vec{R},\chi'\right)\int_0^{\chi s}\frac{d\chi}{\sqrt{\left(\vec{R}-\chi_L\vec{\theta}\right)^2+(\chi-\chi')^2}}$$

where I have set $\chi = \chi_L$ in the slowly varying factors out front. The innermost integral can be done analytically: it is equal to

$$2\ln\left|x+\sqrt{\left(\vec{R}-\chi_L\vec{\theta}\right)^2+x^2}\right)\right|_0^\infty$$

where I have set the upper limit to infinity because there is no contribution to the relevant part of the projected potential from large x. In fact, the only part which depends on $\vec{\theta}$ (and hence is relevant when derivatives are taken) comes from the lower limit: $-2\ln\left|\vec{R}-\chi_L\vec{\theta}\right|$. The integral over χ' then becomes the surface density leaving the desired result.

15.0.3 Problems about Quantum Cosmology (???)

Leuven-2023hw7. Inflation in quantum cosmology (????????)

(?????? тут у меня идейного характера проблемы, так что отдельно нужно засесть в эту задачу!!!)

In this long exercise, we consider the semiclassical approximation of the no-boundary wave function (NBWF). We restrict our attention to FLRW metrics with spherical spatial slices and consider a homogeneous scalar field. Thus, the NBWF is a function of the (real) scale factor b and scalar field value χ. In the semiclassical approxmation, it is given by

$$\Psi(b,\chi) \approx \mathrm{e}^{-\bar{I}(b,\chi)}$$

where $I(b,\chi)$ is the on-shell action of a saddle point. (There could potentially be multiple terms here, but we will only consider one.) The no-boundary proposal is based on a Euclidean path-integral but, as we will see, the saddle points will generically be complex. Hence, they are described by two complex functions $a(\tau)$ and $\phi(\tau)$ taking on the real values b and ϕ, respectively, at some point in the complex τ-plane. The "no-boundary condition" that the saddle should describe a compact, regular geometry reduces to the conditions

$$a(\tau) = \tau + \mathcal{O}\left(\tau^2\right), \quad \phi(\tau) = \phi_0 + \mathcal{O}\left(\tau^2\right)$$

around $\tau = 0$. In the exercise, you will derive some consequences and illustrate some properties of the no-boundary proposal that are valid quite generally in slow-roll inflation. For concreteness, we will consider a quadratic potential for the inflation. The (Euclidean) action of this model is

$$I = \frac{3\pi}{4H^2}\int \mathrm{d}\tau \left[-a\dot{a}^2 - a + a^3 + a^3\left(\dot{\phi}^2 + \mu^2\phi^2\right)\right],$$

which is obtained from

$$I = \int\limits_M \mathrm{d}^4x\sqrt{g}\left[-\frac{1}{16\pi G_N}(R - 2\Lambda) + \frac{1}{2}(\partial\varphi)^2 + \frac{1}{2}m^2\varphi^2\right],$$

by considering Euclidean FLRW metrics $\mathrm{d}s^2 = \mathrm{d}\tilde{\tau}^2 + \tilde{a}^2(\tilde{\tau})\mathrm{d}\Omega_3^2$, where $\mathrm{d}\Omega_3^2$ is the usual metric on the unit three-sphere, and homogeneous scalar fields $\varphi = \varphi(\tilde{\tau})$. To arrive at (3) we have also rescaled the quantities such that everything in (3) is dimensionless, for instance $H^2 = \Lambda/\left(3G_N\right)$.

a) Derive the scalar equation of motion

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} - \mu^2\phi = 0$$

from (3). Here and below, a dot represents a τ-derivative. Instead of the equation of motion for a, we use the Hamiltonian constraint, which together with the scalar field equation implies the second-order equation for a. To derive the Hamiltonian constraint, change variables from τ to λ by $\mathrm{d}\tau = N\,\mathrm{d}\lambda$ where N is such that the range of λ is $\lambda \in [0,1]$. Since the range of τ was not fixed to begin with, N should be treated as dynamical (even though it does not depend on λ). Show that the Hamiltonian constraint can be written as

$$\left(\frac{\dot{a}}{a}\right)^2 - \frac{1}{a^2} + 1 - \dot{\phi}^2 + \mu^2\phi^2 = 0$$

by demanding that the action is stationary with respect to a variation in N. Interpret the terms in this equation by comparing it to the first Friedmann-Lemaitre equation.

b) The simplest solution to these equations is obtained by setting $\phi = 0$, which reduces the problem to the case with only a cosmological constant and no scalar field. Solve the equation for a in this case. Find the points in the complex τ-plane where $a(\tau) = b \in \mathbb{R}_+$.

Consider now the slow-roll scenario. Recall from the lecture that slow-roll is achieved in a quadratic potential when ϕ is sufficiently large. We will find that many of the properties of the $\phi = 0$ solution generalise to this case. Specifically, the solution will be approximately Euclidean along a horizontal segment starting at $\tau = 0$ and approximately Lorentzian along a vertical line where the scalar field is approximately real as well.

c) Solve (6) close to $\tau = 0$ by using the "no-roll" approximation $\phi \approx \phi_0$. Use (5) to derive the second term in the expansion $\phi = \phi_0 + \mathcal{O}\left(\tau^2\right)$.

d) Show that the slow-roll solution of (5) and (6) is

$$a = A\exp\left[-i\mu\phi_0\tau + \frac{\mu^2}{6}\tau^2\right], \quad \phi = \phi_0 + i\frac{\mu}{3}\tau$$

by assuming that $\ddot{\phi}$ is negligible compared to the other terms in (5) and that the scalar potential term $\mu^2\phi^2$ gives the dominant contribution to $\dot{a}/a$. Comment on the behaviour of a around $\tau = 0$.

g) Show that the classicality condition

$$\left(\nabla \bar{I}^{\mathrm{Re}}\right)^2 \ll \left(\nabla \bar{I}^{\mathrm{Im}}\right)^2$$

is satisfied in the large volume limit. Here,

$$(\nabla \bar{I})^2 = f^{\alpha\beta}\partial_\alpha \bar{I}\partial_\beta \bar{I} = -\frac{2H^2}{3\pi b}\left(\frac{\partial \bar{I}}{\partial b}\right)^2 + \frac{2H^2}{3\pi b^3}\left(\frac{\partial \bar{I}}{\partial \chi}\right)^2$$

where $f_{\alpha\beta}$ is the metric on "minisuperspace" [4] deduced from the kinetic term $\frac{1}{2}f_{\alpha\beta}\dot{q}^\alpha\dot{q}^\beta$ in the action and $f^{\alpha\beta}$ is the inverse of $f_{\alpha\beta}$

$$f_{\alpha\beta} = \frac{3\pi}{2H^2}\begin{pmatrix} -a & 0 \\ 0 & a^3 \end{pmatrix}, \quad f_{\alpha\beta} = \frac{2H^2}{3\pi}\begin{pmatrix} -a^{-1} & 0 \\ 0 & a^{-3} \end{pmatrix}.$$

h) The classicality condition implies that the wave function predicts (approximately) classical histories. These are the integral curves of the $\bar{S} = -\bar{I}^{\mathrm{mm}}$, that is, the solutions to

$$p_\alpha^{\mathrm{cl}} \equiv f_{\alpha\beta}\frac{\mathrm{d}q_{\mathrm{cl}}^\beta}{\mathrm{d}t} = \partial_\alpha \bar{S}\left(q_{\mathrm{cl}}\right)$$

Show, using (12), that these reduce to the classical (Lorentzian) slow-roll equations

$$\frac{\dot{a}_{\mathrm{cl}}}{a_{\mathrm{cl}}} = \mu\phi_{\mathrm{cl}}, \quad \dot{\phi}_{\mathrm{cl}} = -\frac{1}{3}\mu.$$

What does this say about the assumption $\chi \sim \phi_0^{\mathrm{Re}}$ in the calculation of $\bar{I}^{\mathrm{Im}}$?

j) In the approximation we are working

$$\Psi \sim \exp\left[-i\frac{\pi}{2}\frac{\mu\chi b^3}{H^2} + \frac{\pi}{2}\frac{1}{(H\mu\chi)^2}\right].$$

Interpret the weighting $|\Psi|^2$ assigned to the different classical histories. Does this wave function seem to predict many or few efolds of inflation?

e) Next, we want to find a point in the complex τ-plane where $(a, \phi) = (b, \chi)$

i) Assume that ϕ_0 is approximately real and show that the solution valid close to the origin is approximately real along the vertical

$$\tau = \frac{\pi}{2\mu\phi_0^{\mathrm{Re}}} + iy, \quad \text{where} \quad \phi_0^{\mathrm{Re}} = \mathrm{Re}\left(\phi_0\right), \quad y \in \mathbb{R}.$$

Verify that the approximate solution is valid at $y = 0$ by checking that the $\mathcal{O}\left(\tau^2\right)$ term in ϕ is small. ii) Now use the slow-roll solution and show that requiring that ϕ is approximately real along the same vertical implies that

$$\phi_0^{\mathrm{Im}} = -\frac{\pi}{6\phi_0^{\mathrm{Rm}}}, \quad \text{where } \phi_0^{\mathrm{Im}} = \mathrm{Im}\,\phi_0,$$

consistent with the assumption that the imaginary part of ϕ_0 is small since we are assuming that ϕ_0^{Re} is large to be in the slow-roll regime. iii) Show that A in the slow-roll solution is given by

$$A \approx \frac{i}{\mu\phi_0^{\mathrm{Re}}}$$

by matching the two solutions at $y = 0$. Then verify that the slow-roll solution is approximately real along the same vertical. [3]

We have now found approximate saddles describing complex slow-roll saddles. The two parameters that are left unfixed, ϕ_0^{Re} and y, determine the arguments of the wave function, χ and b.

f) At last, we compute the on-shell action (and hence Ψ in the semiclassical approximation) using the approximate solutions.

i) First simplify the on-shell action by using (6) to eliminate $\dot{a}$ in (3).

ii) The dominant contribution to the real part comes from the approximately Euclidean region. Use the approximate solution valid there to show that

$$\bar{I}^{\mathrm{Re}} \approx -\frac{\pi}{2\left(H\mu\phi_0^{\mathrm{Re}}\right)^2}$$

iii) The main contribution to the imaginary piece of the action evaluated on the saddle, on the other hand, comes from the approximately Lorentzian region. To simplify the calculation, assume that $\chi \approx \phi_0^{\mathrm{Re}}$. Show that

$$\bar{I}^{\mathrm{Im}} \sim \frac{\pi\mu\chi b^3}{2H^3}$$

Solution to a)

Solution to b)

Solution to c)

Solution to d)

Solution to e)

Solution to f)

Solution to g)

Solution to h)

Solution to j)

Comments about this problem

One might think that initially $\psi_0 = 0$ because there diff. eq. tells us this. (??? I'll add explanation of this, didn't finish thinking about it)

Part VI
Other Topics in Cosmology

(мб потом отдельные части про них сделаю, пока всё остальное тут.)

16 Stochastic backgrounds of cosmological origin

22.1 Characteristic frequency of relic GWs
22.2 GW production by classical fields
22.3 GWs from preheating after inflation
22.4 GWs from first-order phase transitions
22.5 Cosmic strings
22.6 Alternatives to inflation
22.7 Bounds on primordial GW backgrounds

In Chapter 21 we have studied in detail the production of a stochastic GW background due to the amplification of vacuum fluctuations during inflation. Because of its potential importance, and of the many technical steps needed in the relevant computations, we devoted a full chapter to this topic. In this chapter we will examine several other cosmological production mechanisms that have been proposed in the literature, and we will then summarize the present knowledge on stochastic GW backgrounds.

16.1 22.1 Characteristic frequency of relic GWs

We first discuss what is the typical frequency today of a GW signal of cosmological origin. Part of the answer depends on the dynamics of the production mechanism, and therefore is model-dependent, and part is kinematic, depending on the redshift from the production era to the present time. First of all, it is useful to separate the kinematics from the dynamics.

We consider a graviton produced with a frequency f_*, at a production time $t = t_*$, when the temperature had a value T_*, within the RD or MD phases. Its frequency today is redshifted to a value $f_0 = f_* a(t_*)/a(t_0)$. To compute the ratio $a(t_*)/a(t_0)$ we use eq. (17.113), setting $a_r = a_0 = 1$ and $T_r = T_0 \simeq 2.7255(6)$K. This gives

$$f_0 = f_* \frac{a(t_*)}{a(t_0)} \simeq 7.80 \times 10^{-14} f_* \left(\frac{106.75}{g_*^S(T_*)}\right)^{1/3} \left(\frac{1\text{GeV}}{T_*}\right) \tag{22.1}$$

where, as usual, we have normalized the effective number of relativistic species to its high-energy value in the Standard Model; see Section 17.5. We now want to estimate the characteristic value of the frequency f_* of a graviton produced in the early Universe at time t_*, when the temperature was T_*. Here of course the dynamics of the specific production mechanism enters, but still some general considerations are possible. One of the relevant parameters in this estimate is certainly the Hubble parameter at the time of production, $H(t_*) \equiv H_*$. It can then be useful to write the reduced wavelength λ_* of a graviton at the time of production as

$$\lambda_* = \epsilon_* H_*^{-1} \tag{22.2}$$

Since H_*^{-1} is the size of the horizon at time t_*, $\epsilon_* > 1$ corresponds to super-horizon modes at the time of production. Because of causality, the generation of super-horizon modes will be suppressed. This does not mean that the production of modes with $\epsilon_* > 1$ is strictly

forbidden by causality. The spectrum of modes produced by causal mechanisms is continuous with frequency, and does not drop suddenly to zero below a critical frequency. Thus, to some extent, even super-horizon modes are produced. However, at sufficiently low frequencies, such that the corresponding wavelengths are well outside the horizon, causality requires that their spectrum goes to zero, as $f \to 0$, as

$$h_0^2 \Omega_{\rm gw}(f) \propto f^3 \tag{22.3}$$

This follows from the fact, for super-horizon modes, the source term $\tilde{\sigma}_A^{\rm TT}(\mathbf{k})$ in eq. (19.220) should have a flat power spectrum, i.e. $\left\langle \left|\tilde{\sigma}_A^{\rm TT}(\mathbf{k})\right|^2 \right\rangle$ will be independent of $\mathbf{k}$. This corresponds to white noise and reflects the absence of causal correlations. The same will then be true for $\tilde{h}_A(\mathbf{k})$. This means that, for super-horizon modes, the power spectrum $\mathcal{P}_T(k)$ at the time of production satisfies

$$\mathcal{P}_T(k) \propto k \tag{22.4}$$

[see eq. (19.280)] or, equivalently, $S_h(f)$ in eq. (19.284) is independent of the frequency (recall that $k = 2\pi f$). Then, from eq. (19.288) or from eq. (7.202) of Vol. 1, we get eq. (22.3).

Thus, for typical gravitons produced at time t_*, we expect $\epsilon_* = O(1)$ or smaller. More precisely, in the absence of any other length-scale in the problem, we can indeed expect that the peak of the spectrum will be at $\epsilon_* = O(1)$, while if other length-scales smaller than H_*^{-1} enter the problem (or small dimensionless parameters such as a speed of sound $c_s^2 < 1$) , ϵ_* will in general be some orders of magnitude smaller than 1.

During RD, $H_*^2 = (8\pi/3) G \rho_{\rm rad}$. Then, from eq. (17.98),

$$H_*^2 = \frac{\pi^2 g_* T_*^4}{90 m_{\rm Pl}^2} \tag{22.5}$$

and, using $2\pi f_* \equiv H_*/\epsilon_*$, eq. (22.1) can be written as

$$f_0 \simeq 2.65 \times 10^{-8} \frac{1}{\epsilon_*} \left(\frac{T_*}{1{\rm GeV}}\right) \left(\frac{g_*}{106.75}\right)^{1/6} \ {\rm Hz} \tag{22.6}$$

The effect of the dynamics on the value of f_0 has therefore been isolated into the parameter ϵ_*[1]

We will typically be interested in production mechanisms taking place at temperatures $T \gtrsim 100{\rm GeV}$. Then the relation between time and temperature during the RD phase can be obtained using eqs. (17.114) and

Table 22.1 The production time t_* and the production temperature T_* for GWs observed today at frequency f_0, if at the time of production they had a reduced wavelength equal to the horizon scale.

f_0(Hz)	t_*(s)	T_*(GeV)
10^{-4}	1.6×10^{-14}	3.8×10^3
1	1.6×10^{-22}	3.8×10^7
10^2	1.6×10^{-26}	3.8×10^9
10^3	1.6×10^{-28}	3.8×10^{10}

(17.138). Inserting the numerical values, this gives

$$t_* \simeq 0.23 \left(\frac{106.75}{g_*}\right)^{1/2} \left(\frac{1{\rm MeV}}{T_*}\right)^2 \ {\rm s} \tag{22.7}$$

This tells us how far back in time we are exploring the Universe, when we observe a graviton produced at temperature T_*. Using eq. (22.6) we see that, by detecting a stochastic background of relic GWs that today has a typical frequency f_0, we are looking back at the Universe at time

$$t_* \simeq 1.6 \times 10^{-22} \frac{1}{\epsilon_*^2} \left(\frac{1\ \text{Hz}}{f_0}\right)^2 \left(\frac{106.75}{g_*}\right)^{1/6}\ \text{s} \tag{22.8}$$

Table 22.1 gives some representative values for the production time t_* and the temperature of the Universe T_* corresponding to frequencies relevant to space-borne and to ground-based interferometers, setting $\epsilon_* = 1$. One should, however, keep in mind that the estimate $\epsilon_* \sim 1$ can sometimes be incorrect, even as an order of magnitude. We will later illustrate this point with some specific examples. The values in Table 22.1 are more correctly interpreted as a starting point for understanding how the result is affected by the dynamics of the specific production mechanism.

The highest possible value of the frequency that a relic GW frequency could have today can be obtained assuming that at the Planck epoch a graviton is produced with energy of order of the reduced Planck mass $m_{\rm Pl}$, and redshifting it to the present epoch, assuming no inflationary phase and only RD followed by MD. This would give

$$f_0 \simeq 64 \left(\frac{g_*}{106}\right)^{1/6} \text{GHz} \quad (T_* \simeq m_{\rm Pl}) \tag{22.9}$$

The dependence on g_* is rather weak because of the power 1/6 in eq. (22.6). For $g_* = 1000$, f_0 only increases by a factor ~ 1.5 relative to $g_* = 100$. More realistically, we have by now very convincing evidence for the existence of an inflationary phase in the early Universe, so the maximum production temperature for primordial gravitons should rather be determined by the reheating temperature. Indeed, a stochastic background produced by classical mechanisms before inflation would be diluted and completely washed away by the subsequent inflationary phase. As we already mentioned in Chapter 21, the reheating temperature is at most equal to the inflationary scale (if reheating is almost instantaneous and all the energy density in the inflaton field is immediately converted into radiation), and in typical reheating models is several orders of magnitude lower. With the existing bounds on the scale of inflation discussed in Section 21.3.6, the highest possible frequency today of a relic GW background will rather be $O(0.1)$GHz [see eq. (21.361)], and possibly much smaller, depending on the reheating temperature.

16.1.1 22.2 GW production by classical fields

16.1.2 22.2.1 General formalism

We now study the GWs produced by a classical field. The subject is conceptually interesting by itself, and covers a broad range of situations. [2]

We consider a generic inhomogeneous, time-evolving scalar field in an expanding universe. Such a field has a time-varying and spatially dependent energy-momentum tensor and is therefore in general a source for a stochastic background of GWs. The simplest approach for computing the resulting GW production, which indeed was used in early papers on the subject, is to use the formalism reviewed in Section 3.1 of Vol. 1 [and originally developed in Section 10.4 of the textbook by Weinberg (1972)]. In particular, in this formalism the GW amplitude in the far zone and the GW energy spectrum are given, in terms of the energy-momentum tensor of the source, by eqs. (3.14) and (3.16), which we recall here, [3]

$$h_{ij}^{\rm TT}(t,\mathbf{x}) = \frac{4G}{r}\Lambda_{ij,kl}(\hat{\mathbf{n}}) \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \tilde{T}_{kl}(\omega, \omega\hat{\mathbf{n}}) e^{-i\omega(t-r)} \tag{22.10}$$

$$\frac{dE_{\rm gw}}{d\Omega} = \frac{G}{2\pi^2}\Lambda_{ij,kl}(\hat{\mathbf{n}}) \int_{0}^{\infty} d\omega \omega^2 \tilde{T}_{ij}(\omega, \omega\hat{\mathbf{n}}) \tilde{T}^*_{kl}(\omega, \omega\hat{\mathbf{n}}) \tag{22.11}$$

where $\Lambda_{ij,kl}$ is the Lambda tensor that projects onto the TT gauge, defined in eq. (1.36), and $\hat{\mathbf{n}}$ is the propagation direction of the wave. One could then take an ensemble average of $\left\langle \tilde{T}_{ij}(\omega, \omega\hat{\mathbf{n}}) \tilde{T}^*_{kl}(\omega, \omega\hat{\mathbf{n}}) \right\rangle$, to compute the resulting stochastic background. However, in the present cosmological setting such an approach would be inadequate for several reasons. First of all, in this case the source is not localized and therefore there is no wave zone where eq. (22.10) could hold. Second, this expression has been obtained by expanding the metric perturbation around flat space, rather than around the FRW background. The correct approach is to use cosmological perturbation theory, as we have developed in Chapter 19. In particular, the dynamics of the tensor perturbations is given by eq. (19.220), where the source term σ_A is defined by eqs. (18.114), (18.120), (18.123) and (19.215). In Section 19.5 we were mostly interested in the evolution of the GWs, given some initial value $h_{A,\,\rm in}$, so in eq. (19.220) we set the source term to zero (except when we studied the effect of neutrino free-streaming). Here we are considering the production mechanism, so the role of σ_A is crucial.

Lowering the upper index in eq. (18.114), the helicity decomposition of a generic energy-momentum tensor T_{ij} reads

$$T_{ij} = a^2 \left(p\delta_{ij} + \Sigma_{ij} \right) \tag{22.12}$$

where Σ_{ij} is given in eqs. (18.120)-(18.123). The tensor $T_{ij}^{\rm TT} = a^2 \sigma_{ij}^{\rm TT}$ is the TT part of T_{ij}. It follows from the discussion in Section 1.2 that it can be extracted from T_{ij} by taking the Fourier transform and multiplying the momentum mode $\tilde{T}_{ij}(\mathbf{k})$ by the Lambda tensor $\Lambda_{ij,kl}(\hat{\mathbf{k}})$.

It is convenient to introduce the field $\mathrm{h}_{ij} = a h_{ij}$ and, correspondingly, $\mathrm{h}_A = a h_A$, similarly to what we did in Section 21.3.4 [where we introduced φ_A from eq. (21.309) and, as in eq. (21.177), $\chi_A = a\varphi_A$]. Then eq. (19.220) becomes

$$\tilde{\mathrm{h}}_A''(\mathbf{k}) + \left(k^2 - \frac{a''}{a} \right) \tilde{\mathrm{h}}_A(\mathbf{k}) = 16\pi G a^3 \tilde{\sigma}_A^{\rm TT}(\mathbf{k}) \tag{22.13}$$

This equation further simplifies if the typical momenta $\mathbf{k}$ for which the source term is important are well inside the horizon. For such momenta $k^2 \gg a''/a \sim 1/\eta^2$, and we can solve eq. (22.13) using the retarded Green's function of the operator $\partial_\eta^2 + k^2$. This is just the well-known retarded Green's function of the harmonic oscillator,

$$G\left(\eta - \eta'\right) = \frac{\sin\left[k\left(\eta - \eta'\right)\right]}{k} \theta\left(\eta - \eta'\right) \tag{22.14}$$

In particular, if we assume that before an initial conformal time η_i the source was not active, and as initial condition we set $\tilde{\mathrm{h}}_A\left(\eta_i, \mathbf{k}\right) = \tilde{\mathrm{h}}_A'\left(\eta_i, \mathbf{k}\right) = 0$, the solution is

$$\tilde{\mathrm{h}}_A(\eta,\mathbf{k}) = 16\pi G \int_{-\infty}^{\infty} d\eta' G\left(\eta-\eta'\right) a^3\left(\eta'\right) \tilde{\sigma}_A^{\mathrm{TT}}\left(\eta',\mathbf{k}\right)$$
$$= \frac{16\pi G}{k} \int_{\eta_i}^{\eta} d\eta' \sin\left[k\left(\eta-\eta'\right)\right] a^3\left(\eta'\right) \tilde{\sigma}_A^{\mathrm{TT}}\left(\eta',\mathbf{k}\right) \tag{22.15}$$

If the source turns off at some final conformal time η_f, then

$$\tilde{\mathrm{h}}_A\left(\eta_f,\mathbf{k}\right) = \frac{16\pi G}{k} \int_{\eta_i}^{\eta_f} d\eta' \sin\left[k\left(\eta_f-\eta'\right)\right] a^3\left(\eta'\right) \tilde{\sigma}_A^{\mathrm{TT}}\left(\eta',\mathbf{k}\right) \tag{22.16}$$

is the initial value of the subsequent free evolution. During RD and MD, modes that were generated as sub-horizon remain inside the horizon, so their subsequent evolution is determined by eq. (22.13) with the source term set to zero and again $|a''/a| \ll k^2$, i.e. by the free wave equation

$$\tilde{\mathrm{h}}_A'' + k^2 \tilde{\mathrm{h}}_A \simeq 0 \tag{22.17}$$

whose solution is simply

$$\tilde{\mathrm{h}}_A(\eta,\mathbf{k}) = \mathcal{A}_A(\mathbf{k}) \sin\left[k\left(\eta-\eta_f\right)\right] + \mathcal{B}_A(\mathbf{k}) \cos\left[k\left(\eta-\eta_f\right)\right] \quad (\eta \geqslant \eta_f) \tag{22.18}$$

Matching at $\eta = \eta_f$ the values of $\tilde{\mathrm{h}}_A$ and $\tilde{\mathrm{h}}_A'$ obtained from eqs. (22.16) and (22.18) one finds

$$\mathcal{A}_A(\mathbf{k}) = \frac{16\pi G}{k} \int_{\eta_i}^{\eta_f} d\eta' \cos\left[k\left(\eta_f-\eta'\right)\right] a^3\left(\eta'\right) \tilde{\sigma}_A^{\mathrm{TT}}\left(\eta',\mathbf{k}\right) \tag{22.19}$$

$$\mathcal{B}_A(\mathbf{k}) = \frac{16\pi G}{k} \int_{\eta_i}^{\eta_f} d\eta' \sin\left[k\left(\eta_f-\eta'\right)\right] a^3\left(\eta'\right) \tilde{\sigma}_A^{\mathrm{TT}}\left(\eta',\mathbf{k}\right) \tag{22.20}$$

The corresponding GW energy density is given by eq. (1.136). Expressed in terms of h and of conformal time, it reads

$$\rho_{\mathrm{gw}}(\eta) = \frac{1}{16\pi G a^4} \sum_{A=+,\times} \left\langle \left(\mathrm{h}_A' - \mathcal{H}\mathrm{h}_A\right)^2(\eta,\mathbf{x}) \right\rangle \tag{22.21}$$

where $\langle \ldots \rangle$ represents a spatial average over several wavelengths and, as usual, $f' \equiv df/d\eta$. For sub-horizon modes, $\mathrm{h}_A' \sim k\ \mathrm{h}_A \gg \mathcal{H}\mathrm{h}_A$, so the second term in parentheses can be neglected, and, in this limit,

$$\rho_{\mathrm{gw}}(\eta) = \frac{1}{16\pi G a^4} \sum_{A=+,\times} \left\langle \mathrm{h}_A'^2(\eta,\mathbf{x}) \right\rangle \tag{22.22}$$

We now expand $\mathbf{h}_A'(\eta,\mathbf{x})$ in Fourier modes. Then

$$\left\langle \mathbf{h}_A'^2(\eta,\mathbf{x}) \right\rangle = \int \frac{d^3k}{(2\pi)^3} \frac{d^3k'}{(2\pi)^3} \mathbf{h}_A'(\eta,\mathbf{k}) \mathbf{h}_A'^*\left(\eta,\mathbf{k}'\right) \left\langle e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} \right\rangle \tag{22.23}$$

The spatial average over a volume V of size much larger than the reduced wavelengths $1/k$ and $1/k'$ gives

$$\left\langle e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}}\right\rangle = \frac{(2\pi)^3}{V}\delta^{(3)}\left(\mathbf{k}-\mathbf{k}'\right) \tag{22.24}$$

so in the end we get

$$\rho_{\rm gw}(\eta) = \frac{1}{16\pi G a^4}\frac{1}{V}\sum_{A=+,\times}\int\frac{d^3k}{(2\pi)^3}\mid \mathrm{h}'_A(\eta,\mathbf{k})\mid^2 \tag{22.25}$$

where $\mathrm{h}_A(\eta,\mathbf{k})$ is given by eq. (22.15) for $\eta\leqslant\eta_f$, and by eq. (22.18) for $\eta\geqslant\eta_f$. Observe that this formalism allows us to follow the time evolution of the energy density in the tensor perturbations, not only during the period in which the source is no longer active and the tensor perturbations evolve freely in a FRW background, but even in the earlier epoch when the source was active and the tensor perturbations were building up. For $\eta\geqslant\eta_f$ eq. (22.25) can be further simplified if we are not interested in the details of the time evolution of the signal on time-scales of the order of $1/k$ or smaller. In this case we can further perform a temporal average of $\rho_{\rm gw}$ over several periods, retaining only the coarser features of the temporal evolution. Then, using eq. (22.18), together with $\left\langle\sin^2\left[k\left(\eta-\eta_f\right)\right]\right\rangle = \left\langle\cos^2\left[k\left(\eta-\eta_f\right)\right]\right\rangle = 1/2$ and $\langle\sin[k(\eta-\eta_f)]\cos\left[k\left(\eta-\eta_f\right)\right]\rangle = 0$, we get[4]

$$\begin{aligned}\rho_{\rm gw}\left(\eta\geqslant\eta_f\right) &= \frac{1}{32\pi G a^4(\eta)V}\sum_{A=+,\times}\int\frac{d^3k}{(2\pi)^3}k^2\left[\left|\mathcal{A}_A(\mathbf{k})\right|^2+\left|\mathcal{B}_A(\mathbf{k})\right|^2\right]\\ &= \frac{4\pi G}{a^4(\eta)V}\sum_{ij}\int\frac{d^3k}{(2\pi)^3}\left[\left|\int_{\eta_i}^{\eta_f}d\eta'\cos\left(k\eta'\right)a\left(\eta'\right)\tilde{T}_{ij}^{\rm TT}\left(\eta',\mathbf{k}\right)\right|^2\right.\\ &\qquad\left.+\left|\int_{\eta_i}^{\eta_f}d\eta'\sin\left(k\eta'\right)a\left(\eta'\right)\tilde{T}_{ij}^{\rm TT}\left(\eta',\mathbf{k}\right)\right|^2\right]\end{aligned}$$

[4] See eq. (26) of Dufaux, Bergman, Felder, Kofman and Uzan (2007). The different factor $(2\pi)^3$ follows from our different conventions for the Fourier transform.

$$= \frac{4\pi G}{a^4(\eta)V}\sum_{ij}\int\frac{d^3k}{(2\pi)^3}\int_{\eta_i}^{\eta_f}d\eta' d\eta'' a\left(\eta'\right)a\left(\eta''\right)\cos\left[k\left(\eta'-\eta''\right)\right]\quad\tilde{T}_{ij}^{\rm TT}\left(\eta',\mathbf{k}\right)\tilde{T}_{ij}^{*,\rm TT}\left(\eta'',\mathbf{k}\right) \tag{22.26}$$

where in the second equality we have expanded $\cos\left[k\left(\eta_f-\eta'\right)\right]$ and $\sin\left[k\left(\eta_f-\eta'\right)\right]$ in terms of sin and cos of $k\eta_f$ and $k\eta'$ and recombined the various terms. We have also made use of

$$\sum_A\left|\tilde{\sigma}_A\right|^2 = \frac{1}{2}\sum_{ij}\left|\tilde{\sigma}_{ij}\right|^2 \tag{22.27}$$

which follows from eqs. (19.215) and (19.218), and we have used $T_{ij}^{\rm TT} = a^2\sigma_{ij}^{\rm TT}$ instead of $\sigma_{ij}^{\rm TT}$. The overall factor $1/a^4$ correctly gives the cosmological redshift of the energy density of the GWs, as it should for sub-horizon modes; compare with the discussion in Section 19.5.1.

These results allow us to compute the energy density of the GWs generated by a given classical source in a FRW background. In the special case of a localized source in flat space, we must therefore be able to recover eq. (22.11). Observe that eq. (22.11) gives the total energy radiated by the source, up to time $+\infty$. To recover it, we therefore set $a(\eta)=1$ in eq. (22.26)

(so η becomes the same as t) and we extend the integration up to $\eta_f = \infty$. Recall that the lower limit of integration was actually $\eta = -\infty$, which we replaced by η_i because the source was not active before η_i; see eq. (22.15). So, the total GW energy density produced by the source in the flat-space limit, as predicted by eq. (22.26), is

$$\rho_{\rm gw} = \frac{2\pi G}{V}\sum_{ij}\int\frac{d^3k}{(2\pi)^3}\int_{-\infty}^{+\infty}dt'dt''\left[e^{ik(t'-t'')}\tilde{T}_{ij}^{\rm TT}\left(t',\mathbf{k}\right)\tilde{T}_{ij}^{*,\rm TT}\left(t'',\mathbf{k}\right)+e^{-ik(t'-t'')}\tilde{T}_{ij}^{\rm TT}\left(t',\mathbf{k}\right)\tilde{T}_{ij}^{*,\rm TT}\left(t'',\mathbf{k}\right)\right]$$
$$= \frac{4\pi G}{V}\sum_{ij}\int\frac{d^3k}{(2\pi)^3}\tilde{T}_{ij}^{\rm TT}(k,\mathbf{k})\tilde{T}_{ij}^{*,\rm TT}(k,\mathbf{k}) \tag{22.28}$$

where, in the last line, $\tilde{T}_{ij}^{\rm TT}(\omega,\mathbf{k})$ (with $\omega = k$) is the Fourier transform of $\tilde{T}_{ij}^{\rm TT}(t,\mathbf{k})$ with respect to the time variable. We next write (with the summation over repeated indices understood)

$$\begin{aligned}\tilde{T}_{ij}^{\rm TT}(k,\mathbf{k})\tilde{T}_{ij}^{*,\rm TT}(k,\mathbf{k}) &= \Lambda_{ij,kl}(\hat{\mathbf{k}})\tilde{T}_{kl}(k,\mathbf{k})\Lambda_{ij,mn}(\hat{\mathbf{k}})\tilde{T}_{mn}^{*}(k,\mathbf{k})\\ &= \Lambda_{kl,mn}(\hat{\mathbf{k}})\tilde{T}_{kl}(k,\mathbf{k})\tilde{T}_{mn}^{*}(k,\mathbf{k})\end{aligned} \tag{22.29}$$

where we have used $\Lambda_{ij,kl}(\hat{\mathbf{k}})\Lambda_{ij,mn}(\hat{\mathbf{k}}) = \Lambda_{kl,mn}(\hat{\mathbf{k}})$; see eq. (1.37). Then the total GW energy $E_{\rm gw} = V\rho_{\rm gw}$ is given by

$$\begin{aligned}E_{\rm gw} &= 4\pi G\Lambda_{ij,kl}\int\frac{d^3k}{(2\pi)^3}\tilde{T}_{ij}(k,\mathbf{k})\tilde{T}_{kl}^{*}(k,\mathbf{k})\\ &= \frac{G}{2\pi^2}\Lambda_{ij,kl}\int_0^\infty k^2dk\int d\Omega\tilde{T}_{ij}(k,\mathbf{k})\tilde{T}_{kl}^{*}(k,\mathbf{k})\end{aligned} \tag{22.30}$$

Therefore, in the flat-space limit, $dE_{\rm gw}/d\Omega$ is indeed given by eq. (22.11).

16.1.3 22.2.2 GW generation by a stochastic scalar field

If the source is given by random fields, a stochastic average over the field configurations is further required. Thus, the energy density becomes

$$\begin{aligned}\rho_{\rm gw}(\eta) =& \frac{4\pi G}{a^4(\eta)V}\int\frac{d^3k}{(2\pi)^3}\int_{\eta_i}^{\eta_f}d\eta'd\eta''a\left(\eta'\right)a\left(\eta''\right)\cos\left[k\left(\eta'-\eta''\right)\right]\\ &\times\sum_{ij}\left\langle\tilde{T}_{ij}^{\rm TT}\left(\eta',\mathbf{k}\right)\tilde{T}_{ij}^{*,\rm TT}\left(\eta'',\mathbf{k}\right)\right\rangle\end{aligned} \tag{22.31}$$

where now $\langle\ldots\rangle$ denotes the ensemble average over the field configurations. We consider for illustration a single scalar field ϕ with action (21.22). The energy-momentum tensor is then given by eq. (21.24), so in particular

$$T_{ij} = \partial_i\phi\partial_j\phi - g_{ij}\left[\frac{1}{2}g^{\rho\sigma}\partial_\rho\phi\partial_\sigma\phi + V(\phi)\right] \tag{22.32}$$

In eq. (22.32), writing $g_{ij} = a^2(\delta_{ij} + h_{ij})$, the term h_{ij} gives a secondorder contribution, in which the GWs produced at first order become themselves sources for further GW production. To first order, we can therefore replace g_{ij} by $a^2\delta_{ij}$. Then, recalling that $\Lambda_{ij,kl}\delta_{kl} = 0$, to first order the term in square brackets in eq. (22.32) does not contribute to the TT part, and we obtain

$$\begin{aligned}\tilde{T}^{\rm TT}_{ij}(\eta,\mathbf{k}) &= \Lambda_{ij,kl}(\hat{\mathbf{k}})\tilde{T}_{kl}(\eta,\mathbf{k}) \\ &= \Lambda_{ij,kl}(\hat{\mathbf{k}})\int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}}\partial_k\phi\partial_l\phi \\ &= \Lambda_{ij,kl}(\hat{\mathbf{k}})\int \frac{d^3p}{(2\pi)^3}p_k(p-k)_l\tilde{\phi}(\eta,\mathbf{p})\tilde{\phi}(\eta,\mathbf{k}-\mathbf{p})\end{aligned} \tag{22.33}$$

Recalling further from Section 1.2 that $\Lambda_{ij,kl}(\hat{\mathbf{k}})k_l = 0$, we get

$$\tilde{T}^{\rm TT}_{ij}(\eta,\mathbf{k}) = \Lambda_{ij,kl}(\hat{\mathbf{k}})\int \frac{d^3p}{(2\pi)^3}p_kp_l\tilde{\phi}(\eta,\mathbf{p})\tilde{\phi}(\eta,\mathbf{k}-\mathbf{p}) \tag{22.34}$$

Then

$$\begin{aligned}\sum_{ij}\left\langle \tilde{T}^{\rm TT}_{ij}(\eta',\mathbf{k})\,\tilde{T}^{*,\rm TT}_{ij}(\eta'',\mathbf{k})\right\rangle &= \Lambda_{kl,mn}(\hat{\mathbf{k}})\int\frac{d^3p}{(2\pi)^3}\frac{d^3p'}{(2\pi)^3} \\ \times p_kp_lp'_mp'_n\left\langle\tilde{\phi}(\eta',\mathbf{p})\,\tilde{\phi}(\eta',\mathbf{k}-\mathbf{p})\,\tilde{\phi}^*(\eta'',\mathbf{p}')\,\tilde{\phi}^*(\eta'',\mathbf{k}-\mathbf{p}')\right\rangle\end{aligned} \tag{22.35}$$

where we have again used $\Lambda_{ij,kl}(\hat{\mathbf{k}})\Lambda_{ij,mn}(\hat{\mathbf{k}}) = \Lambda_{kl,mn}(\hat{\mathbf{k}})$. If the random field ϕ obeys Gaussian statistics, [5] the four-point correlator can be computed in terms of the two-point functions using Wick's theorem,

$$\begin{aligned}&\left\langle\tilde{\phi}(\eta',\mathbf{p})\,\tilde{\phi}\,(\eta',\mathbf{k}-\mathbf{p})\,\tilde{\phi}^*(\eta'',\mathbf{p}')\,\tilde{\phi}^*(\eta'',\mathbf{k}-\mathbf{p}')\right\rangle \\ &\qquad= \left\langle\tilde{\phi}(\eta',\mathbf{p})\,\tilde{\phi}^*(\eta',\mathbf{p}-\mathbf{k})\right\rangle\left\langle\tilde{\phi}(\eta'',-\mathbf{p}')\,\tilde{\phi}^*(\eta'',\mathbf{k}-\mathbf{p}')\right\rangle \\ &\qquad+ \left\langle\tilde{\phi}(\eta',\mathbf{p})\,\tilde{\phi}^*(\eta'',\mathbf{p}')\right\rangle\left\langle\tilde{\phi}(\eta',\mathbf{k}-\mathbf{p})\,\tilde{\phi}^*(\eta'',\mathbf{k}-\mathbf{p}')\right\rangle \\ &\qquad+ \left\langle\tilde{\phi}(\eta',\mathbf{p})\,\tilde{\phi}^*(\eta'',\mathbf{k}-\mathbf{p}')\right\rangle\left\langle\tilde{\phi}(\eta',\mathbf{k}-\mathbf{p})\,\tilde{\phi}^*(\eta'',\mathbf{p}')\right\rangle\end{aligned} \tag{22.36}$$

[5] In the case of preheating that we will discuss in Section 22.3, this is only true in the initial linear phase. At the end of the preheating phase the dynamics becomes highly non-linear, and the fields strongly non-Gaussian.

[6] Observe that in all computations in this section we use the normalization (17.39) of the Fourier transform, rather than the normalization (19.161). Therefore there is no explicit factor $1/V$ in eq. (22.37), in contrast to eq. (19.164), where the normalization (19.161) was used.

[7] Observe that this contraction corresponds to a disconnected diagram, in which the two fields at time η are contracted among themselves, and also the two fields at time η' are contracted among them. So, the disconnected diagram does not contribute, as we know from experience with quantum field theory. Technically, disconnected diagrams can just be reabsorbed into a redefinition of $\tilde{T}^{\rm TT}_{ij}(\eta,\mathbf{k})$, which is equivalent to a renormalization of the corresponding operator in the fieldtheoretical language. where we have used $\tilde{\phi}(\eta,-\mathbf{k}) = \tilde{\phi}^*(\eta,\mathbf{k})$ for real fields. Assuming invariance under spatial translations and rotations, the two-point function has the general form

$$\left\langle \tilde{\phi}\left(\eta', \mathbf{p}\right) \tilde{\phi}^{*}\left(\eta'', \mathbf{p}'\right) \right\rangle = (2\pi)^3 \delta^{(3)}\left(\mathbf{p}-\mathbf{p}'\right) F\left(\eta', \eta''; p\right) \tag{22.37}$$

where the proportionality to $\delta^{(3)}\left(\mathbf{p}-\mathbf{p}'\right)$ follows from invariance under spatial translations, and the dependence of the function $F\left(\eta', \eta''; p\right)$ on $p = |\mathbf{p}|$ follows from invariance under rotations. [6]

Using eq. (22.37) we see that the first contraction on the right-hand side of eq. (22.36) is proportional to $\delta^{(3)}(\mathbf{k})$. After integration over $d^3k = k^2 dk d\Omega$ in eq. (22.31), it therefore gives a vanishing contribution. [7] Using eq. (22.37) to compute the other two terms in eq. (22.36) and recalling that $(2\pi)^3\delta^{(3)}(\mathbf{p}=0) = V$, we get

$$\sum_{ij} \left\langle \tilde{T}_{ij}^{\mathrm{TT}}\left(\eta', \mathbf{k}\right) \tilde{T}_{ij}^{*,\mathrm{TT}}\left(\eta'', \mathbf{k}\right) \right\rangle = V \Lambda_{kl,mn}(\hat{\mathbf{k}}) \tag{22.38}$$
$$\times \int \frac{d^3p}{(2\pi)^3} p_k p_l \left[p_m p_n + (k-p)_m (k-p)_n\right] F\left(\eta', \eta''; p\right) F\left(\eta', \eta''; |\mathbf{k}-\mathbf{p}|\right)$$

The terms containing k_m or k_n give zero when contracted with $\Lambda_{kl,mn}(\hat{\mathbf{k}})$, while, using the explicit expression (1.39), we get

$$\Lambda_{kl,mn}(\hat{\mathbf{k}}) p_k p_l p_m p_n = \frac{1}{2} p^4 \sin^4\theta(\hat{\mathbf{k}}, \hat{\mathbf{p}}) \tag{22.39}$$

where $\theta(\hat{\mathbf{k}}, \hat{\mathbf{p}})$ denotes the angle between $\hat{\mathbf{k}}$ and $\hat{\mathbf{p}}$ [compare with the computation leading to eq. (4.346) in Vol. 1]. Plugging this result into eq. (22.31) we finally obtain

$$\rho_{\mathrm{gw}}(\eta \geqslant \eta_f) = \frac{4\pi G}{a^4(\eta)} \int \frac{d^3k}{(2\pi)^3} \int_{\eta_i}^{\eta_f} d\eta' d\eta'' a\left(\eta'\right) a\left(\eta''\right) \cos\left[k\left(\eta'-\eta''\right)\right]$$
$$\times \int \frac{d^3p}{(2\pi)^3} p^4 \sin^4\theta(\hat{\mathbf{k}}, \hat{\mathbf{p}}) F\left(\eta', \eta''; p\right) F\left(\eta', \eta''; |\mathbf{k}-\mathbf{p}|\right) \cdot (22.4 \tag{22.40}$$

The integral over d^3p in the second line depends on $\mathbf{k}$, through $|\mathbf{k}-\mathbf{p}|$ and $\theta(\hat{\mathbf{k}}, \hat{\mathbf{p}})$. However, since it is a scalar, and there is no other vector with which we can contract k_i, it can depend on $\mathbf{k}$ only through the modulus k, and not on the direction $\hat{\mathbf{k}}$. This means that the integral over d^3k can be written as $d^3k = 4\pi k^2 dk$. As in eq. (7.197), we write

$$\rho_{\mathrm{gw}}(\eta) = \int_{k=0}^{k=\infty} d(\log k) \frac{d\rho_{\mathrm{gw}}}{d\log k}(\eta, k) \tag{22.41}$$

Then, for $\eta \geqslant \eta_f$,

$$\frac{d\rho_{\mathrm{gw}}}{d\log k}(\eta, k) = \frac{2G}{\pi a^4(\eta)} k^3 \int_{\eta_i}^{\eta_f} d\eta' d\eta'' a\left(\eta'\right) a\left(\eta''\right) \cos\left[k\left(\eta'-\eta''\right)\right]$$
$$\times \int \frac{d^3p}{(2\pi)^3} p^4 \sin^4\theta(\hat{\mathbf{k}}, \hat{\mathbf{p}}) F\left(\eta', \eta''; p\right) F\left(\eta', \eta''; |\mathbf{k}-\mathbf{p}|\right)$$

This is the final result of this section, and gives the energy spectrum of the stochastic background of GWs generated by a random Gaussian scalar field, for $\eta \geqslant \eta_f$, i.e. after the source has switched off, in terms of its two-point function $F\left(\eta', \eta''; p\right)$, defined by eq. (22.37).

16.1.4 22.3 GWs from preheating after inflation

If a thermalized hot plasma ever existed before inflation, at the end of inflation its temperature would be redshifted away by a factor $e^{\Delta N}$, where the number of e-folds during inflation, ΔN, is larger than a minimum value $(\Delta N)_{\rm min}$ typically of order $50-60$, as we saw in Section 21.1. In order to generate the hot thermalized relativistic plasma that we identify as the initial state of the hot big-bang theory, the Universe must therefore be reheated. For single-field slow-roll inflation the first successful reheating mechanism that was studied is the perturbative decay of the inflaton field into other matter fields after the end of the slow-roll phase. It was later realized that the energy transfer from the inflaton to normal matter at the end of inflation can be much more efficient, involving non-linear physics and explosive production of particles, which of course is also potentially interesting from the point of view of GW generation. This non-linear mechanism is referred to as "preheating". In particular, two interesting preheating mechanisms have emerged, namely parametric resonance in single-field inflation and tachyonic preheating in hybrid inflation. We discuss them in the following subsections.

16.1.5 22.3.1 Parametric resonance in single-field inflation

We consider a single-field slow-roll inflationary model, of the type discussed in Section 21.1.4, with an inflaton field ϕ and an inflationary potential $V(\phi)$, and we couple it for simplicity to a single scalar matter field χ with a coupling $\chi^2\phi^2$,

$$S=\int d^4x\sqrt{-g}\left[\frac{1}{16\pi G}R-\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi-V(\phi)-\frac{1}{2}g^{\mu\nu}\partial_\mu\chi\partial_\nu\chi-\frac{1}{2}g^2\chi^2\phi^2\right] \tag{22.43}$$

where g is a coupling constant. The background solution is $\phi=\phi(t)$, $\chi=0$ so, including perturbations around it, we write

$$\phi(t,\mathbf{x})=\phi(t)+\delta\phi(t,\mathbf{x}) \tag{22.44}$$
$$\chi(t,\mathbf{x})=\delta\chi(t,\mathbf{x}) \tag{22.45}$$

At the end of inflation the field ϕ performs oscillations around the minimum of $V(\phi)$ with a frequency given by $m_\phi=\sqrt{d^2V/d\phi^2}$, with the second derivative evaluated at the minimum of $V(\phi)$. Therefore the solution of the equation of motion for $\phi(t)$ is approximately

$$\phi(t)=\Phi(t)\sin m_\phi t \tag{22.46}$$

where $\Phi(t)$ is a function that changes slowly on the time-scale m_ϕ^{-1}, and which we will approximate just by a constant. The equation for the perturbations is [compare with eq. (21.175)]

$$\begin{aligned}\delta\ddot{\phi}_{\mathbf{k}}+3H\delta\dot{\phi}_{\mathbf{k}}+\left(\tfrac{k^2}{a^2}+m_\phi^2\right)\delta\phi_{\mathbf{k}}=0\\ \delta\ddot{\chi}_{\mathbf{k}}+3H\delta\dot{\chi}_{\mathbf{k}}+\left(\tfrac{k^2}{a^2}+g^2\phi^2(t)\right)\delta\chi_{\mathbf{k}}=0\end{aligned} \tag{22.48}$$

The equation for $\delta\phi_{\mathbf{k}}$ is the usual equation for the perturbations of a scalar field in FRW, damped by the friction term due to the expansion of the Universe, and furthermore has a mass term with the "correct", non-tachyonic sign. This equation has no unstable mode, as we studied in Section 19.3. In contrast, the equation for $\delta\chi_{\mathbf{k}}$ has a time-dependent frequency

$$\begin{aligned}\omega_k^2(t)&=k^2+g^2a^2(t)\phi^2(t)\\&=k^2+g^2\Phi^2a^2(t)\sin m_\phi t\end{aligned} \tag{22.49}$$

We can rewrite eq. (22.48) in the form

$$\frac{d^2}{dz^2}(\delta\chi_k) + 3\frac{H}{m_\phi}\frac{d}{dz}(\delta\chi_k) + [A(k) - 2q\cos(2z)]\delta\chi_k \tag{22.50}$$

where $z = m_\phi t$,

$$q = \frac{g^2\Phi^2}{4m_\phi^2} \tag{22.51}$$

is called the resonance parameter and

$$A(k) = \frac{k^2}{m_\phi^2 a^2} + 2q \tag{22.52}$$

Equation (22.50), with the friction term proportional to H/m_ϕ neglected (which is a good approximation during the rapid phase of parametric resonance that we will find below), is a Mathieu equation. Depending on the values of $A(k)$ and of q, it can have either oscillatory or exponential solutions, with a band structure corresponding to sets of stable or unstable modes. In general, if q is large, a broad range of values of k can be in resonance, and one is then in the broad-resonance regime. In this situation energy is efficiently pumped from the inflaton field into the resonant modes $\delta\chi_k$, leading to large time-dependent and inhomogeneous field configurations. Once the resonant modes reach an amplitude $\chi^2 \sim \Phi^2$, the perturbative approach in which we expanded around $\chi = 0$ breaks down, and preheating ends. At this point the field χ dissipates its energy via interaction with the other fields of the theory, eventually leading to thermalization.

The scalar field dynamics can be computed numerically, solving the equation of motion of the interacting scalar fields in FRW. The numerical study shows that there is an initial linear phase in which the field χ
grows and its statistics is still Gaussian. In this regime some analytic checks of the numerical results for GW production can be made using the results of Section 22.2.2. This phase evolves into a highly nonlinear phase, where the statistics of the field χ is highly non-Gaussian, characterized by the formation, expansion and collision of bubble-like field inhomogeneities. This is followed by a stage of turbulence and, finally, thermalization. Most of the GW production takes place in the violent bubble-like phase. The resulting GW production can then be computed from eq. (22.13), either dropping the term a''/a and using eq. (22.26), or directly integrating eq. (22.13) numerically. The fact that most of the GWs are produced in the bubble collisions makes the GW production mechanism somewhat similar to that operating in strongly first-order phase transitions, which we will discuss in Section 22.4.

The details of the numerical results depend of course on the specific inflationary model used. Most numerical results have been obtained for the inflationary potential $V(\phi) = (1/4)\lambda\phi^4$, which, as we have seen in Section 21.3.5, is now ruled out, or for $V(\phi) = (1/2)m^2\phi^2$, which is presently very close to being ruled out. The result depends of course also on the coupling g between the inflaton and the field χ or, more generally, on the form assumed for the interaction between ϕ and χ. Thus, the predictions for the GW spectrum are necessarily strongly dependent on the details of the model used.

The typical frequency of the GWs produced can, however, be estimated as in Section 22.1. The characteristic scale H_* in this case is provided by the Hubble parameter at the peak of GW production. In a first approximation, we can identify H_* with the value of the Hubble parameter at the end of the reheating phase and the beginning of RD. In this approximation we can use eq. (22.5) with T_* equal to the reheating temperature. However, the numerical studies typically show that the spectrum peaks at wavelengths one or two orders of magnitude smaller than H_*^{-1}, so the peak frequency is shifted toward correspondingly larger values. We see from

eq. (22.6) (taking ϵ_* of order, say, 0.1) that, to have a signal at $f \sim 200 - 300$ Hz, near the best sensitivity region of ground-based interferometers such as LIGO and VIRGO, one should have $T_* \sim 10^9$GeV. The spectrum is quite peaked around the maximum; on the low-frequency side its tails must decrease as f^3 by causality [see eq. (22.3)], while at frequencies higher than the peak frequency it declines even more rapidly, corresponding to the fact that modes with larger values of k are not excited by parametric resonance. The numerical studies also indicate that, for inflationary models of the type $V(\phi) = (1/2)m^2\phi^2$, the typical peak value of $\Omega_{\rm gw}$ today is roughly independent of the inflationary scale, and of order

$$h_0^2\Omega_{\rm gw} \sim 10^{-11} \tag{22.53}$$

This is considerably higher than the value obtained in inflation from the amplification of vacuum fluctuations which, as we saw in eq. (21.368), with the current bounds on the tensor-to-scalar tensor ratio r cannot exceed a value of order 10^{-16}.

16.1.6 22.3.2 Tachyonic preheating in hybrid inflation

In hybrid inflationary models, discussed in Section 21.1.4, preheating occurs when the inflaton field ϕ decreases below the critical value ϕ_c for which the Higgs-like field Φ acquires a tachyonic mass; see the discussion following eq. (21.84).

Naively, one might think that the process in which a Higgs field Φ evolves from the local maximum $\Phi = 0$ of its Mexican hat potential, toward a spontaneous symmetry-breaking minimum, could be described in terms of the evolution of a spatially homogeneous component of the Higgs field [or, equivalently, in terms of a spatially homogeneous configuration of the field $\sigma \equiv \left(\Phi^\dagger\Phi\right)^{1/2}$, see the definitions after eq. (21.82)], which rolls down from the top of the potential to the symmetry-breaking minimum (which, for hybrid inflation, is at $\phi = 0$ and $\sigma = v$). The field should then starts performing oscillations around this minimum. Initially the oscillations would have an amplitude comparable to the vacuum expectation value itself, and would then be damped by the production of particles coupled to σ, so that σ eventually settles into the minimum of the potential, with amplitude of oscillations much smaller than v. This can take place after a large number of oscillations.

However, numerical investigations have revealed that the actual dynamical evolution leading to the final spontaneous symmetry-breaking state is more complex and violent, and in particular the approximation of homogenous field is inadequate. Indeed, when the field σ acquires a tachyonic mass $-(1/2)m^2\sigma^2$, all momentum modes with $k < m$ become unstable, and start to grow exponentially as

$$\tilde{\sigma}_k(t) \propto e^{t\sqrt{m^2-k^2}} \tag{22.54}$$

When these fluctuations become large they correspond to quantum states with large occupation numbers and can then be described as classical waves. As soon as the oscillations in the various momentum modes become large, the process becomes very complicated and non-linear, with these classical waves scattering among them because of the $\lambda\sigma^4$ interaction in eq. (21.83). The full dynamics can, however, be studied numerically by solving relativistic wave equations on a lattice, and one finds that the field sets into its minimum after only $O(1)$ oscillations. The rapid transfer of energy from the potential $V(0)$ into the amplitude of the inhomogeneous oscillations due to the tachyonic instability has been called tachyonic preheating, and is substantially different from the parametric resonance phenomenon that drives preheating in single-field inflation.

Again, the process of tachyonic preheating can efficiently create GWs. The specific form of the potential $U(\phi)$ in eq. (21.83) mostly affects the dynamics during preheating by setting the

velocity with which ϕ reaches the critical value ϕ_c, so we can just use $\dot{\phi}_c$ as one of the parameters of the model or, equivalently, we can use the dimensionless quantity $V_c = g\dot{\phi}_c/(\lambda v^2)$ [not to be confused with the potential $V(\phi)$ defined in eq. (21.84)]. The parameters of the model are therefore $\{\lambda, g, v, V_c\}$. The
dependence on v can be extracted analytically, since by passing to the dimensionless variables $\phi/v, \sigma/v$ and vx^μ (and neglecting the expansion of the Universe during the short preheating era) the dependence on v drops from the equations of motion derived from eq. (21.83). In particular, one finds that the peak frequency is independent of v while the value of $h_0^2\Omega_{\rm gw}$ at the peak scales as v^2. A detailed analysis of the dependence of the results on the model parameters can be found in Dufaux, Felder, Kofman and Navros (2009). For instance analytic estimates (confirmed within one order of magnitude by the numerical simulations) indicate that, when V_c is not too low, so that preheating is actually initiated by the classical rolling of the inflaton, the peak frequency today, f_0, is given by

$$f_0 \sim 7 \times 10^{10} \lambda^{1/4} V_c^{1/3} \text{ Hz} \tag{22.55}$$

(except if $g^2/\lambda \ll 1$, in which case, after the first tachyonic growth, the typical momenta are significantly shifted toward lower frequencies), while the corresponding peak value of the GW energy density is given by

$$h_0^2\Omega_{\rm gw}(f_0) \sim 10^{-6} V_c^{-2/3} \left(\frac{v}{M_{\rm Pl}}\right)^2 \tag{22.56}$$

As an example, in Fig. 22.1 we show the predictions for $h_0^2\Omega_{\rm gw}(f)$ from tachyonic preheating for $\lambda = 10^{-5}$ and different values of V_c, while in Fig. 22.2 we show the result for fixed $V_c = 10^{-3}$, two different values $\lambda = 10^{-14}$ (left) and 10^{-5} (right), and several different values of λ/g^2. All spectra are shown for $v = 10^{-3} M_{\rm Pl}$. Lowering v leaves the shape of the spectra unchanged but reduces their amplitude proportionally to v^2, as we see from eq. (22.56). Observe from eq. (22.55) that for sufficiently small values of λ the peak of the spectrum can be in the region accessible to ground-based or even space-borne interferometers.

16.2 22.4 GWs from first-order phase transitions

16.2.1 22.4.1 Crossovers and phase transitions

During its cosmological evolution the Universe has undergone a number of changes of regime that, depending on the precise dynamics of the model, can be either phase transitions or smooth crossovers. The most recent one took place at the QCD scale, at a critical temperature T_c whose value is estimated from lattice QCD simulations to be about 150 MeV. Above this temperature quarks and gluons are deconfined and form a quark-gluon plasma, while below they are confined into hadrons. At earlier times and higher temperature, $T_c \sim 100$GeV, there takes place the electroweak phase transition (or, as we will see, crossover). As the Universe cools and T decreases below T_c, the Higgs field acquires a non-vanishing expectation value, breaking $SU(2) \times U(1)$ to $U(1)_{\rm em}$, and

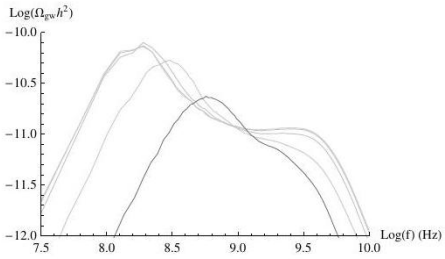


Fig. 22.1 Gravity wave spectra after preheating for $\lambda = 10^{-5}$ and different values of V_c. From Dufaux, Felder, Kofman and Navros (2009). (C)SISSA Medialab Srl. Reproduced by permission of IOP Publishing. All rights reserved.

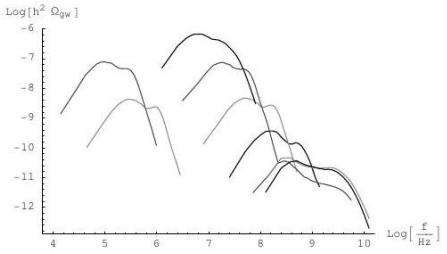


Fig. 22.2 $h_0^2\Omega_{\rm gw}(f)$ for fixed $V_c = 10^{-3}$, two different values $\lambda = 10^{-14}$ (left) and 10^{-5} (right), and several different values of λ/g^2. From Dufaux, Felder, Kofman and Navros (2009). CSISSA Medialab Srl. Reproduced by permission of IOP Publishing. All rights reserved.
the W and Z bosons become massive. Further phase transitions could have occurred earlier, for instance at the grand unification scale.

From the point of view of GW production, the interesting situation is when there is a first-order phase transition. A first-order phase transition is a violent event, and can be a significant source of GWs, as we will see in this section. In contrast, if the transition is second-order (or just a smooth crossover), no significant GW production is expected to take place.

Crossover/phase transition at the electroweak scale

The order of the electroweak phase transition depends on the value of the parameters of the underlying particle physics model, and in particular on the Higgs mass. Unfortunately, in the Standard Model, with the observed value of the Higgs mass $m_H = 125.09 \pm 0.21$ (stat.) $\pm$ 0.11 (syst.) GeV, at the electroweak scale there is no first-order phase transition. Rather, there is just a smooth crossover at a temperature $T = 150.5 \pm 1.5$GeV (a first-order phase transition would have taken place if the Higgs field had been much lighter, $m_H \lesssim 80$GeV) and no GW production. This is unfortunate since, as we will see in more detail in eq. (22.135) (and we already understand from the approximate estimate in Table 22.1 on page 646), the GW signal from a first-order phase transition at the electroweak scale can fall within the LISA frequency band.

However, a strong first-order phase transition at the electroweak scale would also be welcome because it would make possible electroweak baryogenesis. Thus, the search for extensions of the Standard Model which feature a strong first-order phase transition is well motivated. In the minimal supersymmetric Standard Model (MSSM), a strong first-order phase transition requires $m_H < 115$GeV, so again with the actual Higgs mass there is no appreciable GW signal. One can, however, construct other extensions of the Standard Model where the phase transition is first-order. In particular, the addition of an extra real singlet field or of a second Higgs doublet can induce a strong first-order phase transition (in the latter case, if the extra Higgs states are heavier than about 300GeV). Another example of a model that leads to a first-order phase transition is obtained by adding to the Standard Model a dimension-6 operator, which could emerge as an effective low-energy description of some strongly-coupled dynamics at the TeV scale.

Crossover/phase transition at the QCD scale

For the QCD deconfinement transition the situation is somewhat similar. On the one hand, for a first-order phase transition the GW signal could be in the frequency range accessible to pulsar timing arrays [see eq. (22.135)], so it is potentially very interesting. On the other hand,

after many years of studies, lattice QCD simulations with three dynamical quark flavors have conclusively shown that in QCD there is no phase transition but just a crossover.

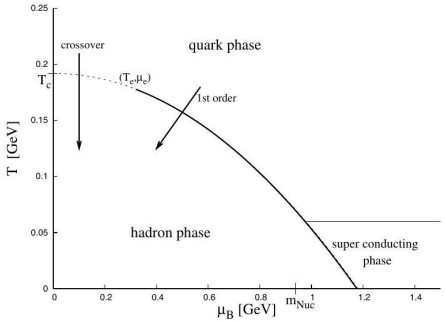


Fig. 22.3 A conjectured form of the phase diagram of QCD in the (μ_B, T) plane. Solid lines indicate a first-order phase transition, while the dashed line indicates a crossover. From Schwarz and Stuke (2009). (c)SISSA Medialab Srl. Reproduced by permission of IOP Publishing. All rights reserved. [Observe that the exact phase diagram, including the existence of the critical endpoint (μ_e, T_e), and the value of the critical temperature, are still under debate. In particular, more recent lattice QCD calculations give a critical temperature $T_c \simeq 155$MeV].

These studies have, however, been performed assuming that the baryonic chemical potential μ_B (i.e. the energy needed to add a baryon to a thermalized state at fixed volume and entropy) is small, $\mu_B \ll T$. The situation for larger values of μ_B is still uncertain. Figure 22.3 shows the conjectured QCD phase diagram in the (μ_B, T) plane, using results from perturbative computations, lattice QCD and nuclear physics. The solid line indicates a first-order phase transition, which terminates at a critical endpoint (μ_e, T_e). The correctness of this picture is still under debate. In the standard cosmological setting, μ_B/T is very small. Indeed, CMB and nucleosynthesis data show that the baryon-to-photon ratio

$$\frac{n_B}{n_\gamma} \sim \frac{\mu_B}{T} \sim 10^{-9} \tag{22.57}$$

This quantity is conserved after baryogenesis and the last first-order phase transition, when the evolution conserves entropy. Thus, in the standard scenario, at the QCD scale $\mu_B/T \sim 10^{-9}$, and in this case the cosmological transition at the QCD scale is a crossover and does not generate GWs.

Non-standard scenarios have been investigated that could lead to a large value of μ_B/T and then, possibly, to a first-order phase transition. A possibility is that, at temperatures above the QCD phase transition, the value of μ_B/T was $O(1)$ thanks to some mechanism that generates a high baryon number. Then the QCD phase transition could be of first order. At the same time, during this first-order phase transition, the QCD vacuum could become trapped in a false vacuum state, generating a short period of inflation that dilutes μ_B/T to the current value. A different attempt at obtaining a large value of μ_B/T is based on the observation that μ_B depends both on the specific baryon asymmetry b (i.e. the baryon asymmetry per entropy) and on the specific lepton asymmetry l. The specific baryon asymmetry is fixed by cosmological observations and is very small, $b = (8.85 \pm 0.24) \times 10^{-11}$. Direct measurement of the specific lepton asymmetry l are more difficult, because the cosmic neutrino background escapes direct detection and is only indirectly constrained. The combined bounds from CMB and nucleosynthesis are only of order $|l| \lesssim 0.02$. The standard cosmological scenario predicts that the lepton asymmetry should be of the same order as the baryon asymmetry. This is due to the fact that after inflation any preexisting baryon or lepton asymmetry has been washed away. They must therefore be re-generated after inflation, by baryon- and lepton-numberviolating processes. In the Standard Model this can happen somewhat above the electroweak scale,

due to sphaleron processes. In this case in each single sphaleron process the baryon-number violation is equal to the lepton-number violation, $\Delta(B-L)=0$. However, there are extensions of the Standard Model, for example with heavy (s)neutrinos, where a large lepton asymmetry can be produced. For $|l| \gg b, \mu_B$ grows linearly with l, reaching values of order 0.2GeV for $|l| \sim 0.02$, so not far from the critical point (μ_e, T_e) in Fig. 22.3. Thus, considering also the theoretical uncertainties in Fig. 22.3, it is not yet completely excluded that in extensions of the Standard Model leading to a large lepton asymmetry the QCD transition could be first-order.

The bottom-line is that, both at the electroweak scale and at the QCD scale, using the Standard Model (and, for the QCD scale, standard scenarios for the baryonic chemical potential) there is no first-order phase transition, so no GW production. We also see from eq. (22.6) that a phase transition at the grand-unification scale would give a signal at frequencies to high to be observed, even at ground-based interferometers. There is, however, at present still room for extensions of the Standard Model, or of standard scenarios, that could lead to a first-order phase transition, either at the electroweak or at the QCD scale. This is an example of the fact that the predictions for cosmological stochastic backgrounds suffer from uncertainties in the particle physics model at high energies, as well as from uncertainties in the cosmological model at early times. On the other hand, this means that the detection of a primordial stochastic background due to a first-order cosmological phase transition would reveal features of high-energy particle physics that would hardly be accessible otherwise, and would therefore be a very remarkable result. In Section 22.4.2 we will therefore study the GW production in a first-order phase transition in the early Universe. We will see that this study is also quite interesting from a conceptual point of view.

16.2.2 22.4.2 First-order phase transitions in cosmology

The dynamics of a phase transition is governed by the finite-temperature effective potential. [8] A typical example of a finite-temperature effective potential $V(\phi, T)$ inducing a first-order phase transition is given by

$$V(\phi, T) = \frac{1}{2}M^2(T)\phi^2 - \frac{1}{3}\delta(T)\phi^3 + \frac{1}{4}\lambda\phi^4 \tag{22.58}$$

where $\phi \geqslant 0$ (for example ϕ is the modulus of a field) and, at least in some range of temperatures near a value T_0,

$$M^2(T) = \gamma\left(T^2 - T_0^2\right) \tag{22.59}$$

$$\delta(T) = AT \tag{22.60}$$

where the parameters γ, A and λ are strictly positive. The potential $V(\phi, T)$ is shown in Fig. 22.4 for different values of T, in the case $A^2 < 4\lambda\gamma$. For $T > T_0$ the field has a minimum at $\phi = 0$, where $V = 0$. If ϕ is the order parameter of a symmetry, then at this minimum the symmetry is unbroken. At $T \gg T_0$ this is the only minimum of the potential. If $A^2 < 4\lambda\gamma$, then, as we lower the temperature, a second extremum of $V(\phi)$ appears, at a temperature T_1 given by

$$T_1^2 = \frac{T_0^2}{1 - A^2/(4\lambda\gamma)} \tag{22.61}$$

This extremum is at $\phi = \phi_1 = AT_1/(2\lambda)$, and $V(\phi_1) > 0$. At $T < T_1$ this extremum splits into a local maximum and a local minimum. Lowering the temperature further, the value of the potential at the local minimum gets lower and lower, until is becomes degenerate with the minimum at $\phi = 0$, at a critical temperature T_c given by

$$T_c^2 = \frac{T_0^2}{1 - 2A^2/(9\lambda\gamma)} \tag{22.62}$$

At $T < T_c$ this minimum becomes the true minimum of the potential. Finally, as the temperature is lowered further to the value T_0, the extremum in $\phi = 0$ becomes a local maximum.

Before entering into a more technical analysis, let us first understand qualitatively, from Fig. 22.4, how a first-order phase transition takes place in an expanding Universe. At large temperatures the effective potential $V(\phi, T)$ has only the minimum at $\phi = 0$, and the field sits at this minimum. As the Universe cools down and its temperature becomes lower than T_c, this minimum becomes metastable, so the Universe is now trapped in a false vacuum. The transition to the true vacuum can take place via either thermal or quantum fluctuations, which nucleate bubbles of true vacuum. At temperature T, the nucleation of a vacuum bubble results in an energy gain equal to $(4\pi/3)r^3\Delta\rho_{\rm vac}\,(T)$, where r is the bubble radius and $\Delta\rho_{\rm vac}(T)$ is the energy density difference between the metastable state and the true vacuum. On the other hand, one loses an energy equal to $4\pi r^2\sigma$, where σ is the surface tension of the bubble [8] For a review of the definition of the finite-temperature effective potential and the relevant computational techniques in thermal field theory see for example Quiros (1999).

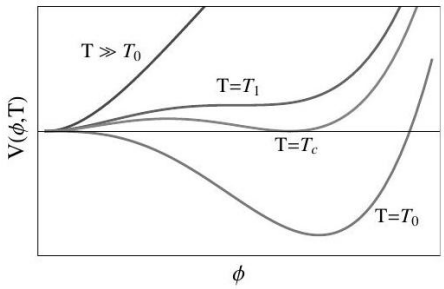


Fig. 22.4 The potential $V(\phi, T)$ for a first-order phase transition, for different values of T. We have set $A^2/(4\lambda\gamma) = 0.5$.

[9] A more precise quantitative estimate will be given in eq. (22.124). (which is calculable given the effective potential, as we will see). For a given value of $T < T_c$, corresponding to a given value of $\Delta\rho_{\rm vac}(T)$, there is therefore a critical radius $r_c(T) = 3\sigma/\Delta\rho_{\rm vac}(T)$ (a more precise computation, using the exact profile of the scalar field, will be given below) such that bubbles nucleated with $r < r_c(T)$ shrink and disappear because of the surface tension, while bubbles with $r > r_c(T)$ survive and expand.

Of course, the probability of nucleating a large bubble is small. In standard equilibrium thermodynamics, if the system is in a metastable state and we wait long enough, it will eventually undergo the transition to the true vacuum state, even if the transition rate is small. In the cosmological setting, however, there is an explicit time-scale due to the expansion of the Universe, and we must compare the transition rate from the false to the true vacuum with the Hubble expansion rate. If, at a given temperature, the nucleation rate $\Gamma(T)$ of supercritical bubbles, i.e. bubbles with $r > r_c(T)$, is much smaller than the Hubble parameter $H(T)$, the Universe will have cooled down significantly further before nucleation of supercritical bubbles has a significant chance of taking place. When T is smaller than T_c but close to it, $\Delta\rho_{\rm vac}\,(T)$ is small, so $r_c(T)$ is large and the nucleation rate of bubbles with $r > r_c(T)$ is small. As the Universe cools down further, $\Delta\rho_{\rm vac}(T)$ increases, $r_c(T)$ decreases, and the nucleation rate $\Gamma(T)$ of supercritical bubbles increases. A firstorder phase transition will take place if, at some temperature T_* in the range $T_0 < T_* < T_c, \Gamma\,(T_*)$ becomes as large as $H\,(T_*)$.[9] If this never happens, then, once the system reaches the temperature T_0, the field ϕ smoothly rolls down for $\phi = 0$, which is now a maximum, toward the true minimum, and we have a second-order phase transition.

Denoting by $\eta(T)$ the vacuum expectation value of the scalar field at the true vacuum at temperature T and observing that the effective potential $V(\phi, T)$ is actually the free energy density of the system, the vacuum energy density associated with the transition is

$$\rho_* = \left[-V[\eta(T), T] + T\frac{d}{dT}V[\eta(T), T]\right]_{T=T_*} \tag{22.63}$$

Since we are interested in transitions taking place during RD, it is convenient to normalize this quantity to the radiation energy density (17.98), defining

$$\alpha = \frac{30\rho_*}{\pi^2 g_*\left(T_*\right) T_*^4} \tag{22.64}$$

Another important parameter is the bubble nucleation rate at the transition temperature, which is modeled as an exponential,

$$\Gamma(t) = \Gamma_* e^{\beta(t-t_*)} \tag{22.65}$$

where t_* is the time at which the transition takes place [defined more quantitatively in eq. (22.123)], and at which the rate has a value Γ_*. We will see later how this exponential form emerges, and how to compute β from the effective potential. Since β has dimensions of inverse time, the relevant dimensionless quantity is actually the ratio β/H_*, where $H_* = H\left(T_*\right)$, which indeed will appear in the computation of GW production.

16.2.3 22.4.3 Thermal tunneling theory

We now discuss more quantitatively how bubbles of true vacuum are nucleated and how their nucleation rate can be computed, first reviewing the bounce computation in quantum mechanics, then studying QFT at zero temperature and finally turning to the relevant case of tunneling in QFT at finite temperature. Apart from the application to GW production in first-order phase transitions, the theory has several applications in QFT, and is of great elegance in itself.

Tunneling and the bounce solution

Let us consider a quantum-mechanical system with one degree of freedom $q(t)$ and Hamiltonian $H = p^2/2 + V(q)$, where

$$V(q) = \frac{1}{2}\omega^2 q^2 + \frac{1}{2}gq^4 \tag{22.66}$$

We have set the particle mass $m = 1$, and g is a coupling constant. For $g > 0$ the potential is bounded from below and the Hamiltonian H is positive definite. The partition function of the system at temperature $T = 1/\beta$ can be computed using the basic result relating the operator formalism of quantum mechanics to the Euclidean path-integral formulation

$$\operatorname{tr} e^{-\beta H/\hbar} = \int\limits_{q(-\beta/2)=q(\beta/2)} [dq] e^{-S_E/\hbar} \tag{22.67}$$

where the trace is over the Hilbert space of the system and $[dq]$ denotes the integration measure in the path integral. [10] Since we will use a semiclassical approximation, which is formally an expansion in $\hbar$, it is convenient for the moment to keep $\hbar$ explicit rather than setting $\hbar = 1$. The path integral in eq. (22.67) is performed over trajectories $q(t)$ with $-\beta/2 \leqslant t \leqslant \beta/2$, and periodic boundary conditions. In eq. (22.67) we have performed the rotation to Euclidean space, and the Euclidean action is

$$S_E = \int_{-\beta/2}^{\beta/2} dt \left[\frac{1}{2}\dot{q}^2 + V(q)\right] \tag{22.68}$$

In the limit $\beta \to \infty$ the left-hand side of eq. (22.67) is dominated by the contribution of the ground state, whose energy we denote by E_0, so $\mathrm{tr}\, e^{-\beta H/\hbar} \simeq e^{-\beta E_0/\hbar}$, while on the right-hand side the only trajectories that contribute are those that satisfy $q(-\beta/2) = q(\beta/2) = 0$, since otherwise the Euclidean action diverges as $\beta \to \infty$. Thus, the groundstate energy can be computed from

$$E_0(g) = -\lim_{\beta\to\infty} \frac{\hbar}{\beta} \log \int_{q(-\beta/2)=q(\beta/2)=0} [dq] e^{-S_E/\hbar} \tag{22.69}$$

As long as $g > 0$, this allows us to compute the energy $E_0(g)$, which will have the form $E_0(g) = (1/2)\hbar\omega[1 + O(g)]$, corresponding to a harmonic [10] See for example Section 9.6 of Maggiore (2005).

[11] For a detailed discussion of how analytic continuation works in this problem, see Chapter 36 of Zinn-Justin (2002).

[12] Observe that for $g < 0$ the Hamiltonian does not have a real spectrum, because it is unbounded from below and the boundary conditions on the wavefunction at spatial infinity are not real.

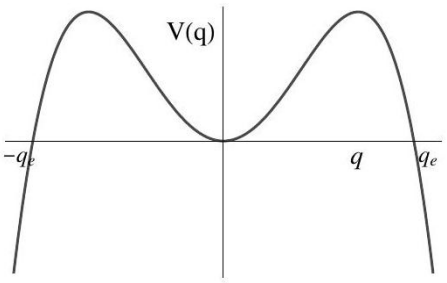


Fig. 22.5 The potential $V(q)$ for $g < 0$. The points $q = \pm q_e$ where $V(q) = 0$ correspond to the escape points of the tunneling process in Minkowski space. oscillator with frequency ω, plus anharmonic corrections parametrized by g. Observe that, in the limit $\beta \to \infty$, rescaling $q \to q/\sqrt{g}$,

$$\frac{1}{\hbar}\int_{-\infty}^{\infty} dt \left[\frac{1}{2}\dot{q}^2 + \frac{1}{2}\omega^2 q^2 + \frac{1}{2}g q^4\right] \to \frac{1}{g\hbar}\int_{-\infty}^{\infty} dt \left[\frac{1}{2}\dot{q}^2 + \frac{1}{2}\omega^2 q^2 + \frac{1}{2}q^4\right] \tag{22.70}$$

so the perturbative expansion in powers of g is the same as the semiclassical expansion in powers of $\hbar$.

We now want to consider the situation $g < 0$. The corresponding potential is shown in Fig. 22.5. Now the potential is unbounded from below, and $E_0(g)$ can be obtained by analytic continuation from the expression valid for $g > 0$, given in eq. (22.69). [11] As is physically clear, and as we will show below, this analytic continuation generates an imaginary part for E_0, so that now the wavefunction $\psi(t)$ of the corresponding eigenstate evolves as

$$\psi(t) \propto e^{-i(\mathrm{Re}E_0)t - |\mathrm{Im}\, E_0|t} \tag{22.71}$$

corresponding to the fact that a state localized near the minimum at $q = 0$ eventually decays by tunneling. [12] Thus, to find E_0 we still have to compute the path-integral on the right-hand

side of eq. (22.69), except that now $g < 0$ so the potential is the one given shown in Fig. 22.5. Observe also that, from the procedure of analytic continuation, we inherit the boundary conditions $q(-\beta/2) = q(\beta/2) = 0$.

We compute this path integral in the semiclassical approximation. Thus, we look for the classical solutions $q_{\rm cl}(t)$ of the equation of motion derived from the action S_E, that satisfy the boundary conditions $q(-\beta/2) = q(\beta/2) = 0$, with $\beta \to \infty$. We then write

$$q(t) = q_{\rm cl}(t) + \hbar^{1/2}\varphi(t) \tag{22.72}$$

where φ are the quantum fluctuations, and we integrate over φ. Then, to quadratic order in φ

$$\begin{aligned} &\int\limits_{q(-\beta/2)=q(\beta/2)=0} [dq] e^{-S_E/\hbar} = \\ &= e^{-S[q_{\rm cl}]/\hbar} \int [d\varphi] e^{-\frac{1}{2}\int\limits_{-\infty}^{\infty} dt_1 dt_2 \varphi(t_1) M(t_1,t_2) \varphi(t_2)} \end{aligned} \tag{22.73}$$

where the operator $M(t_1, t_2)$ is given by

$$\begin{aligned} M(t_1,t_2) &= \left.\frac{\delta^2 S}{\delta q(t_1)\,\delta q(t_2)}\right|_{q=q_{\rm cl}} \\ &= \left[-\frac{d^2}{dt_1^2} + V''[q_{\rm cl}(t_1)]\right] \delta(t_1 - t_2) \end{aligned} \tag{22.74}$$

Let $\varphi_n(t)$ be a complete set of (in general complex) normal modes of the operator M,

$$\int\limits_{-\infty}^{\infty} dt' M(t,t')\,\varphi_n(t') = \lambda_n \varphi_n(t) \tag{22.75}$$

We take these normal modes to be orthogonal with respect to the scalar product

$$\langle f|g\rangle = \int\limits_{-\infty}^{\infty} dt f^*(t) g(t) \tag{22.76}$$

but we find it convenient not to require them to be orthonormal, so

$$\langle \varphi_n|\varphi_m\rangle = \delta_{nm} u_n \tag{22.77}$$

where the u_n give the norms of the modes. We then expand

$$\varphi(t) = \sum_n \xi_n \varphi_n(t) \tag{22.78}$$

where the ξ_n are real coefficients, ranging from $-\infty$ to $+\infty$. The measure of the path integral is defined as

$$[d\varphi] = \mathcal{N} \prod_n \sqrt{\frac{u_n}{2\pi\hbar}} d\xi_n \tag{22.79}$$

where $\mathcal{N}$ is a normalization constant that we will fix later. We then have

$$\int_{-\infty}^{\infty} dt_1 dt_2 \varphi(t_1) M(t_1, t_2) \varphi(t_2) = \sum_n \lambda_n u_n \xi_n^2 \tag{22.80}$$

The integration over the modes with non-zero eigenvalues λ_n is Gaussian and gives a result proportional to

$$e^{-S[q_{c1}]/\hbar} (\det' M)^{1/2} \tag{22.81}$$

where the prime means that we have excluded the eigenvectors with $\lambda_n = 0$.[13]We will discuss later the treatment of the zero modes. If $S[q_{cl}]$ is non-zero, the factor $e^{-S[q_{cl}]/\hbar}$ is non-perturbative in $\hbar$. If the non-zero eigenvalues of M are all positive, or more generally if there are an even number of negative eigenvalues, $\det' M > 0$ and the contribution (22.81) is real, so it contributes to the real part of E_0. However, in this case we have no right to retain it, since this term is exponentially small as $\hbar \to 0^+$. It is then negligible compared with the terms obtained by performing perturbation theory around the trivial solution $q_{cl}(t) = 0$, since these contributions are given by an expansion in powers of the coupling g, which, in view of the comment after eq. (22.70), is the same as an expansion in powers of $\hbar$. However, if there is one or, more generally, an odd number of negative eigenvalues, then $\det' M < 0$ and the contribution (22.81) is imaginary. In this case, it gives the leading correction to $\mathrm{Im}\, E_0$.

The equation of motion obtained from the Euclidean action (22.68) is

$$\ddot{q} = V' \tag{22.82}$$

which is formally the same as that obtained from a Minkowskian action with a potential $-V(q)$, shown in Fig. 22.6. One obvious solution satisfying the boundary conditions $q(-\beta/2) = q(\beta/2) = 0$ is the trivial [13] The integration over a mode with $\lambda_n < 0$ is defined by analytic continuation and again gives a factor proportional to $\lambda_n^{1/2}$, so eq. (22.81) holds for both positive and negative eigenvalues. For negative eigenvalues, however, the proportionality constant differs by a factor 1/2 because of the procedure of analytic continuation; see the detailed discussion in Callan and Coleman (1977), after their eq. (2.22). In the following we will not need the explicit proportionality factors, and we will not keep track of them, limiting ourselves to quoting the final result in eq. (22.98). These factors are accurately computed in Coleman (1977a,b) and Callan and Coleman (1977).

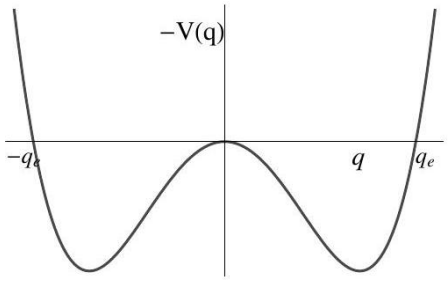


Fig. 22.6 The potential $-V(q)$ for $g < 0$.

one, $q_{cl}(t) = 0$. However, in this case $V''(q_{cl}) = V''(0) \equiv \omega^2$ and the operator M is the one corresponding to the harmonic oscillator, with only positive eigenvalues. So, it does not contribute to the imaginary part. Thus, we have to look for non-trivial solutions. The first integral of eq. (22.82) is

$$\frac{1}{2}\dot{q}_{cl}^2 - V(q_{cl}) = E \tag{22.83}$$

The solutions with $E < 0$ corresponds to oscillations around a minimum of $-V$, say the minimum at $q > 0$, with inversion points q_1 and q_2. From eq. (22.83), the period β of this solution is given by

$$\begin{aligned}\beta &= 2\int_{q_1}^{q_2} dq \frac{1}{\sqrt{2(E+V)}} \\ &= 2\int_{q_1}^{q_2} dq \frac{1}{\sqrt{2E+\omega^2 q^2 + (1/2)gq^4}}\end{aligned} \tag{22.84}$$

In the limit $\beta \to \infty$ the integral on the right-hand side must diverge. For $E < 0$ but close to zero, the inversion point q_1 is at $q_1 \simeq 2|E|/\omega$, and the integrand near the extremum $q = q_1$ is approximately proportional to $1/\sqrt{q^2 - q_1^2}$. For $q_1 \neq 0$, $\sqrt{q^2 - q_1^2} \simeq \sqrt{(q - q_1)\, 2q_1}$ and this singularity is integrable. In contrast, for $q_1 = 0$, we have $1/\sqrt{q^2 - q_1^2} = 1/q$, whose integral diverges at $q = 0$. This means that the solution whose period $\beta \to \infty$ is the one whose inversion point $q_1 \to 0^+$, and therefore with $E \to 0$. Thus, in the limit $\beta \to \infty$, on the solution satisfying the boundary conditions $q(-\infty) = q(\infty) = 0$ the first integral of eq. (22.82) is

$$\frac{1}{2}\dot{q}_{\rm cl}^2 - V\left(q_{\rm cl}\right) = 0 \tag{22.85}$$

In the limit $\beta \to \infty$ we understand from Fig. 22.6 that this solution starts from $q \to 0^+$at $t \to -\infty$, goes to the point $q = q_e$ where it bounces, and then goes back to $q = 0$ as $t \to \infty$. This solution is called the bounce. Actually, in the limit $\beta \to \infty$ the system is invariant under time translations, so if $q_{\rm cl}(t)$ is a solutions of the equations of motion with these boundary conditions, then $q_{\rm cl}\left(t; t_0\right) = q_{\rm cl}\left(t - t_0\right)$ is also a solution, for an arbitrary value of the parameter t_0. Therefore, there is a whole family of classical bounce solutions around which we must expand the path integral, parametrized by t_0, which is called a "collective coordinate". Using eq. (22.85), we see that the action (22.68) evaluated on this solution is positive-definite and, in the limit $\beta \to \infty$, has the value

$$S_{\rm cl} = \int_{-\infty}^{\infty} dt \dot{q}_{\rm cl}^2 \tag{22.86}$$

independent of t_0. We next consider the fluctuations around a bounce solution. It is easy to see that there is a zero mode, i.e. an eigenvector with zero eigenvalue, given by

$$\varphi_0(t) = \dot{q}_{\rm cl}(t) \tag{22.87}$$

Indeed, differentiating eq. (22.82) with respect to time, we get

$$\left[-\frac{d^2}{dt^2} + V''\left[q_{\rm cl}(t)\right]\right]\dot{q}_{\rm cl}(t) = 0 \tag{22.88}$$

which shows that $\dot{q}_{\rm cl}(t)$ is an eigenvector of the operator M, with zero eigenvalue. More generally we can observe that, whenever the classical solution $q_{\rm cl}(t;\gamma)$ depends on a set of collective coordinates γ_i, the functions $\partial q_{\rm cl}/\partial\gamma_i$ are zero modes. This can be shown starting from the classical equation of motion, $(\delta S/\delta q)_{q=q_{\rm cl}} = 0$, and taking its derivative with respect to γ_i,

$$0 = \frac{\partial}{\partial \gamma_i}\left(\frac{\delta S}{\delta q(t)}\right)_{q=q_{\rm cl}} = \int dt' \left(\frac{\delta^2 S}{\delta q(t)\delta q\left(t'\right)}\right)_{q=q_{\rm cl}} \frac{\partial q_{\rm cl}\left(t';\gamma\right)}{\partial \gamma_i} \tag{22.89}$$

The operator $\left[\delta^2 S/\delta q(t)\delta q\left(t'\right)\right]_{q=q_{\rm cl}}$ is just $M\left(t,t'\right)$, and we see that $\partial q_{\rm cl}/\partial\gamma_i$ are zero modes. In our case the collective coordinate is t_0 and $q_{\rm cl}$ is a function of $t-t_0$, so $\partial q_{\rm cl}/\partial t_0 = -\partial q_{\rm cl}/\partial t$, and hence $\dot{q}_{\rm cl}(t)$ is a zero mode.

From the fact that $\dot{q}_{\rm cl}(t)$ is a zero mode it also follows that the operator M in eq. (22.74) has one and only one negative eigenvalue. In fact, M has the form of a Hermitian Hamiltonian in a one-dimensional quantummechanics problem. As we know from quantum mechanics, its spectrum is such that the eigenfunction with the lowest eigenvalue (i.e. the ground state in the Hamiltonian analogy) has no node, the first excited state has one node, etc. Since the zero mode $\dot{q}_{\rm cl}(t)$ has one node, corresponding to the point $q=q_e$ where the solution bounces back, it follows that the zero mode is the first excited state of the operator M, and M has one and only one eigenvector with lower eigenvalue, i.e. with $\lambda<0$. Thus, $\left(\det' M\right)^{1/2} = i\left|\det' M\right|^{1/2}$, and we get a contribution to $\operatorname{Im} E_0$.

We must next understand how to treat the zero mode, i.e. the integration over the variable ξ_0 associated with the mode $\varphi_0(t)=\dot{q}_{\rm cl}(t)$. In the path integral the integration over ξ_0 runs from $-\infty$ to $+\infty$, and nothing else depends on ξ_0, so this is a formally divergent term. However, we must recall that we are working in the limit $\beta\to\infty$, and that there is already a factor of $1/\beta$ on the right-hand side of eq. (22.69), which eventually must cancel. So, while for the other occurrences of β in our computation to lowest order it was legitimate to set directly $\beta=\infty$ [for example in the integration limits in eq. (22.86)], here we have to be more careful. The infinitesimal fluctuation in the direction parametrized by ξ_0 actually corresponds to moving in the space of classical solutions. Indeed, consider an infinitesimal fluctuation $\delta\xi_0\varphi_0$ around the classical solution $q_{\rm cl}\left(t;t_0\right)$. Using the fact that $\varphi_0(t) = (\partial/\partial t) q_{\rm cl}\left(t;t_0\right) = -\left(\partial/\partial t_0\right) q_{\rm cl}\left(t;t_0\right)$, we see that

$$\begin{aligned} q_{\rm cl}\left(t;t_0\right) + \delta\xi_0\varphi_0(t) &= q_{\rm cl}\left(t;t_0\right) - \delta\xi_0 \frac{\partial}{\partial t_0} q_{\rm cl}\left(t;t_0\right) \\ &= q_{\rm cl}\left(t;t_0-\delta\xi_0\right) \end{aligned} \tag{22.90}$$

[14] More precisely, we can transform the integral over $d\xi_0$ into an integral over dt_0 using a simpler version of the Faddeev-Popov trick used in the quantization of gauge theories. First of all observe that, given a classical configuration $q_{\rm cl}\left(t;t_0\right)$, we define the fluctuations around it as

$$\varphi\left(t;t_0\right) \equiv q(t) - q_{\rm cl}\left(t;t_0\right) \tag{22.91}$$

The generic field configuration $q(t)$ knows nothing about t_0, so φ depends on both t and t_0. We start from the identity

$$1 = \int dt_0 \delta\left[f\left(t_0\right)\right] \left|\frac{\partial f}{\partial t_0}\right| \tag{22.92}$$

which is valid for any (differentiable) function $f\left(t_0\right)$, and we chose

$$f\left(t_0\right) = \int dt \varphi\left(t;t_0\right) \dot{q}_{\rm cl}\left(t;t_0\right) \tag{22.93}$$

i.e. the projection of the quantum fluctuations $\varphi(t;t_0)$ onto the zero mode. Using eq. (22.78), $f(t_0) = \xi_0(t_0)\, u_0$ while, using eq. (22.91),

$$\frac{\partial f}{\partial t_0} = \int dt \left[-\dot{q}_{\rm cl}(t;t_0) \frac{\partial}{\partial t_0} q_{\rm cl}(t;t_0) + \varphi(t;t_0) \frac{\partial}{\partial t_0} \dot{q}_{\rm cl}(t;t_0) \right] \tag{22.94}$$

In the first term we use $-\partial q_{\rm cl}/\partial t_0 = +\dot{q}_{\rm cl}$, where the dot is as usual the derivative with respect to t, so the first term is just u_0. Then we get

$$1 = \int dt_0 \delta(\xi_0) \left[1 - \frac{1}{u_0} \varphi(t;t_0)\, \ddot{q}_{\rm cl} \right] \tag{22.95}$$

The second term depends on the quantum field φ. Exponentiating the square bracket in the form $\exp\{-(1/u_0)\,\varphi(t;t_0)\,\ddot{q}_{\rm cl}\}$, this term can be reabsorbed into the action. However, the action appears in the exponent in the form $S_E/\hbar$, while this term has no factor of $\hbar$, and is therefore a loop correction (actually, it can be shown that it only contributes at the two-loop level). To lowest order in the semiclassical expansion we can therefore neglect it, and write eq. (22.95) as

$$1 \simeq \int dt_0 \delta(\xi_0) \tag{22.96}$$

Inserting this into eq. (22.79) eliminates the integral over ξ_0 and replaces it with an integral over the collective coordinate t_0, with a Jacobian equal to 1.
Thus, moving by $\delta\xi_0$ in the direction of the zero mode $\varphi_0(t)$ is equivalent to changing the classical solution around which we are expanding, i.e. it is equivalent to performing a shift of the collective coordinate $t_0 \to t_0 + \delta t_0$, with $\delta t_0 = -\delta\xi_0$. Thus, the integration over ξ_0 is equivalent to the integration over the collective coordinate t_0, and corresponds to summing, in the path integral, over all possible bounce solutions. [14] For finite β, t_0 ranges between $-\beta/2$ and $\beta/2$, so this finally give an overall factor of β, which is very welcome since it matches the factor $1/\beta$ on the right-hand side of eq. (22.69).

The final step of this (long!) computation is to observe that, besides the bounce solutions, we have multi-bounces, corresponding to solutions bouncing back and forth several times between $q = 0$ and $q = q_e$. Their effect is to exponentiate the one-bounce contribution, which compensates for the log in eq. (22.69).

Thus, writing $\mathrm{Im}\, E_0 = \Gamma/2$, the tunneling rate Γ is given by

$$\Gamma = A e^{-S_b} \tag{22.97}$$

where S_b is the value of the Euclidean action, evaluated on the bounce solution, and A is a proportionality constant, which is given by

$$A = \hbar \left(\frac{S_b}{2\pi\hbar} \right)^{1/2} \left| \frac{\det'\left[-\partial_t^2 + V''(\phi_b)\right]}{\det\left[-\partial_t^2 + V''(0)\right]} \right|^{-1/2} \tag{22.98}$$

The factor $(S_b/2\pi\hbar)^{1/2}$ is related to the integration over the zero mode, and comes from the factor $(u_0/2\pi\hbar)^{1/2}$ in eq. (22.79) on observing, from eq. (22.86), that the norm u_0 of the zero mode $\varphi_0 = \dot{q}_{\rm cl}$ is just S_b. The factor $\mathcal{N}$ in eq. (22.79) is fixed by requiring that the path integral computed with the integration measure (22.79) reproduces the standard result for the transition amplitude of the harmonic oscillator, and provides the factor $\det^{1/2}\left[-\partial_t^2 + V''(0)\right]$ in

eq. (22.98). Finally, the factor of two from $\Gamma = 2\,\mathrm{Im}\,E_0$ is compensated by an extra factor 1/2 from the integration over the negative eigenvalue; see Note 13.

The same computation can be performed for a potential $V(q)$ such as that shown in Fig. 22.7. Now the system has a "false vacuum" in $q = 0$ and a true vacuum in $q = q_0$. The equation of motion (22.82) derived from the Euclidean action is again formally the same as that of a particle moving in the potential $-V(q)$, which is shown in Fig. 22.8, and again has a bounce solution, from $q = 0$ to $q = q_e$ and back. Observe, however, that, in this case, what we are computing in this way is not the imaginary part of the energy of a state in the spectrum of the Hamiltonian. The potential shown in Fig. 22.7 is bounded from below, and the corresponding Hamiltonian has a spectrum of purely real eigenvalues. The rate Γ must simply interpreted as the tunneling rate of a state initially localized near the false vacuum.

Tunneling in QFT at zero temperature

We next consider a scalar field theory in flat space, with Minkowskian action

$$S_M = \int d^4x \left[-\frac{1}{2}\eta^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi) \right] \tag{22.99}$$

[recall that our signature is $\eta_{\mu\nu} = (-,+,+,+)$], and we consider first QFT at zero temperature. We chose a potential $V(\phi)$ that has a false vacuum in $\phi = 0$, chosing the zero of the potential such that $V(\phi) = 0$, and a true vacuum at $\phi = \phi_0 > 0$, with $V(\phi_0) < 0$, as in Fig. 22.7 (with $q \to \phi, q_0 \to \phi_0$).[15] The rotation to Euclidean signature $(+,+,+,+)$ is obtained by performing the Wick rotation, i.e. introducing Euclidean time τ from $\tau = it$.[16] Then $d^4x = -id\tau d^3x \equiv -i\,(d^4x)_E$. Defining the Euclidean action S_E from $e^{iS_M} = e^{-S_E}$, i.e. $S_E = -iS_M$, we therefore find

$$S_E = \int d^4x_E \left[\frac{1}{2}\partial_\mu\phi\partial_\mu\phi + V(\phi) \right] \tag{22.100}$$

where the sum over repeated lower indices is performed with the Euclidean metric, $\partial_\mu\phi\partial_\mu\phi \equiv (\partial\phi/\partial\tau)^2 + (\partial_i\phi)^2$. We can now repeat the computation performed in the quantum-mechanical case, with minor modifications. Namely, we look for solutions of the equations of motion derived from the Euclidean action, subject to the boundary condition

$$\lim_{\tau\to\pm\infty} \phi(\tau, \mathbf{x}) = 0 \tag{22.101}$$

Furthermore, we must now also impose

$$\lim_{|\mathbf{x}|\to\infty} \phi(\tau, \mathbf{x}) = 0 \tag{22.102}$$

otherwise the action evaluated on the solution would be infinite. Again, the fluctuations around the trivial solution $\phi(\tau, \mathbf{x}) = 0$ have no negative eigenvalue and do not contribute to the imaginary part, so we must rather look for a non-trivial classical solution. In the zero-temperature case that we are considering here, it is reasonable to look for a solution that respects the $O(4)$ symmetry of Euclidean space. Namely, we chose an arbitrary point of four-dimensional Euclidean space, and we use it as the origin for the coordinate system. With respect to this origin, we define a radial coordinate $\rho = (\tau^2 + \mathbf{x}^2)^{1/2}$, and we look for a solution $\phi = \phi(\rho)$. Of course, the origin of the coordinate system can be chosen arbitrarily, which means that now the solution has a set of collective coordinates $\{\tau_0, \mathbf{x}_0\}$ corresponding to the choice of the origin, reflecting the invariance under (Euclidean) time translation as well as spatial translation. As in the quantum-mechanical tunneling computation of the previous subsection, we let τ run over the finite interval $-\beta/2 < \tau < \beta$, letting $\beta \to \infty$ at the end, and similarly we

work in a large but finite spatial volume V. Then, the integration over the collective coordinate in the path integral gives a factor βV. The factor β is canceled by the factor $1/\beta$ in eq. (22.69), as in the quantum-mechanical case. The result [15] More precisely, we take $V(\phi)$ to be the effective potential, which also takes into account the loop corrections. As first shown in a classic paper by Coleman and Weinberg (1973), radiative corrections can change the qualitative shape of the potential, and develop a true minimum at $\phi = \phi_0 > 0$, thus inducing spontaneous symmetry breaking.

[16] The plus sign in the definition $\tau = +it$ is fixed by the condition that the pole structure of the propagator obtained from the Euclidean action is the same as that of the $-i\epsilon$ prescription of the Feynman propagator in Minkowski space.

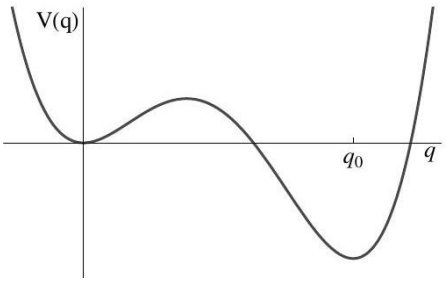


Fig. 22.7 The potential $V(q)$.

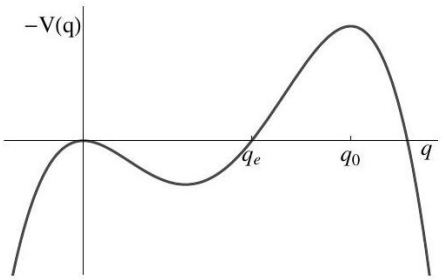


Fig. 22.8 The function $-V(q)$.

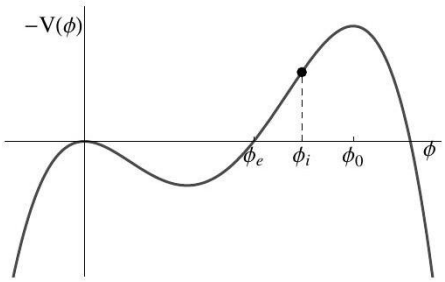


Fig. 22.9 The function $-V(\phi)$, and the initial position ϕ_i.

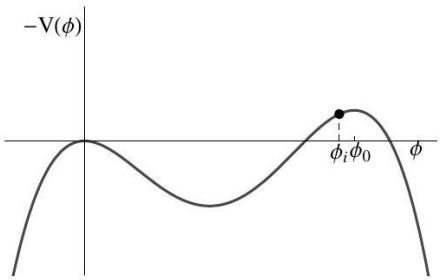


Fig. 22.10 As in Fig. 22.9, with a smaller energy difference between the true and the false vacuua is therefore that in QFT the rate Γ is proportional to the volume V, and the quantity that is finite is the rate per unit volume, Γ/V. This corresponds to the fact that the nucleation of bubbles of true vacuum can take place anywhere in space.

On a function $\phi = \phi(\rho)$, the action (22.100) reads

$$S_E = 2\pi^2 \int_0^\infty d\rho \rho^3 \left[\frac{1}{2} \left(\frac{d\phi}{d\rho} \right)^2 + V(\phi) \right] \tag{22.103}$$

(where $2\pi^2$ is the solid angle in four dimensions), so the equation of motion becomes

$$\frac{d^2\phi}{d\rho^2} + \frac{3}{\rho} \frac{d\phi}{d\rho} = V'(\phi) \tag{22.104}$$

On the $O(4)$-symmetric solution, the two boundary conditions (22.101) are unified into a single boundary condition

$$\lim_{\rho \to \infty} \phi(\rho) = 0 \tag{22.105}$$

Furthermore, requiring that the solution be regular at the origin of the polar coordinate system gives

$$\left. \frac{d\phi}{d\rho} \right|_{\rho=0} = 0 \tag{22.106}$$

Equation (22.104) can be formally interpreted as the motion of a particle of unit mass in one dimension, described by the coordinate ϕ, with ρ playing the role of time, moving in the potential $-V$ shown in Fig. 22.8, and further subject to a damping force $-\eta(\rho)v$, where $v = d\phi/d\rho$ is the velocity of this fictitious particle and $\eta(\rho) = 3/\rho$ a rather peculiar friction coefficient, depending on "time" ρ. This fictitious particle is subject to the boundary conditions (22.105) and (22.106). This mechanical analogy allows us to understand the qualitative features of the solution. The "particle" must start at "time" $\rho = 0$ with zero velocity [see eq. (22.106)], from some initial position ϕ_i chosen so that, at time $\rho \to \infty$, it reaches the position $\phi = 0$. Because of the friction term, its initial energy must therefore be greater than zero, so the initial position must be in the region $\phi_e < \phi_i < \phi_0$, see Fig. 22.9. Its precise value can be determined by integrating eq. (22.104) numerically, assigning a value of $\phi(\rho = 0)$ in the range $\phi_e < \phi(\rho = 0) < \phi_0$, setting $d\phi/\, d\rho|_{\rho=0} = 0$ and using an under-shooting/over-shooting procedure. If we chose a value of $\phi(\rho = 0)$ smaller than ϕ_i, then, because of the friction term, the solution will not have enough energy to reach $\phi = 0$, and will bounce back and perform oscillations around the minimum of $-V$. If we choose a value of $\phi(\rho = 0)$ such that $\phi_i < \phi(\rho = 0) < \phi_0$, then the solution over-shoots, and reaches the region $\phi < 0$, so eventually $\phi \to -\infty$, as we see from the form of $-V$ in Fig. 22.9 (observe that the friction term goes to zero as $\rho \to +\infty$, so it does not significantly damp the evolution at large values of "time" ρ).

In this way we can determine the solution, $\phi_b(\rho)$, numerically. We still call this solution a bounce since, if we follow its evolution as a
function of Euclidean time τ at $\mathbf{x} = 0$, it corresponds to a solution that for $\tau \to -\infty$ (which means $\rho \to +\infty$) goes to zero, increases to the maximum value ϕ_i, which is reached at $\tau = 0$, and then for positive τ bounces back, reaching $\phi = 0$ again as $\tau \to +\infty$. When we continue back to Minkowski space, the solution $\phi = \phi_b(\rho)$ becomes $\phi = \phi_b\left(\sqrt{-t^2 + r^2}\right)$, where $r = |\mathbf{x}|$. If we look at it as a function of r for fixed t, for example at $t = 0$, we see that at small r it is near the value $\phi = \phi_i$, and thus close to the true vacuum ϕ_0. As r increases, ϕ evolves toward the value $\phi = 0$, corresponding to the false vacuum. It therefore represents a "bubble"

in which the field configuration in the inner region is close to the true vacuum, and the field in the exterior is close to the false vacuum. Since the full dependence of the solution on t and r is through the combination $-t^2 + r^2$, with time this bubble of true vacuum expands into the false vacuum at the speed of light. [17] Actually, the description in terms of a bubble is especially appropriate if the transition between the inner and outer region is sharp. This is the case when the energy difference between the true and the false vacuum is small, which is the situation depicted in Fig. 22.10. In this case we see from Fig. 22.10 that the value ϕ_i must be quite close to the maximum ϕ_i of the function $-V$. From the mechanical analogy we see that in this case, in the Euclidean solution, the "particle" with coordinate ϕ will stay for a long "time" ρ near the value ϕ_i, since here the potential is close to flat, up to a "time" $\rho = R$ when this slow-roll regime terminates and the particle quickly rolls toward the "position" $\phi = 0$. Translated in terms of the Minkowskian solution, at fixed time $t = 0$, this describes a field configuration in which, for $r < R, \phi$ is very close to the value $\phi = \phi_0$ corresponding to the true vacuum; it remains almost constant up to $r = R$, and then it quickly goes to the false vacuum value $\phi = 0$. The transition between the region where $\phi \simeq \phi_0$ and the region where $\phi \simeq 0$ takes place in a layer of width $\delta r \ll R$. This corresponds to the thin-wall solution, in which a bubble of true vacuum is separated by a thin wall from the surrounding false vacuum. Again, because of the dependence on $-t^2 + r^2$, with time this bubble expands into the false vacuum at the speed of light.

We now have all the elements to readily adapt the computation of the transition rate that we performed in the quantum-mechanical case. [18] The result is

$$\frac{\Gamma}{V} = Ae^{-S_b/\hbar} \tag{22.107}$$

where S_b is the Euclidean action (22.103) evaluated on the bounce solution $\phi_b(\rho)$. [19] As in the quantum-mechanical computation, the prefactor A in eq. (22.107) can be computed by keeping track of the normalization factors u_n for the four zero modes [which give a factor $(S_b/2\pi)^{1/2}$ each], as well as of the normalization factor $\mathcal{N}$ in the measure (22.79). This gives [20]

$$A = \left(\frac{S_b}{2\pi}\right)^2 \left|\frac{\det'\left[-\partial_\mu\partial_\mu + V''\left(\phi_b\right)\right]}{\det\left[-\partial_\mu\partial_\mu + V''(0)\right]}\right|^{-1/2} \tag{22.111}$$

[17] Observe, however, that this Minkowskian description in terms of an expanding bubble only holds after the bubble has been nucleated, and therefore only for the expanding branch of the hyperboloid $r^2(t) = R^2 + t^2$. The situation is analogous to the description of barrier penetration by a particle in quantum mechanics, in which the WKB approximation is used to compute the tunneling probability, and only after the particle escapes can it be described as freely propagating.

[18] It can again be proved that the fluctuations around the bounce solution have one and only one negative eigenvalue; see Section 6.5 of Coleman (1977b). 19 Observe that, even if V is negative for some values of the field, it can be shown that $S_b > 0$. Indeed, consider the family of functions $\phi_\lambda(x) = \phi_b(x/\lambda)$. Rescaling the integration variable $x \to \lambda x$ in the action, we get $S\left[\phi_\lambda\right] = \int d^4x \left[\frac{\lambda^2}{2}\left(\partial_\mu\phi_b\right)^2 + \lambda^4 V\left(\phi_b\right)\right]$.

Since ϕ_b is a solution of the equations of motion, it is an extremum in the space of all functions, so in particular within the family of function $\phi_\lambda(x)$. Thus, the action (22.108) must be stationary at $\lambda = 1$. This gives

$\int d^4x V(\phi_b) = -(1/4)\int d^4x \left(\partial_\mu \phi_b\right)^2$,
and therefore

$$S_b = \frac{1}{4}\int d^4x \left(\partial_\mu \phi_b\right)^2 > 0 \tag{22.109}$$

see Section 6.2 of Coleman (1977b). Observe that this is a generalization of the result (22.85), (22.86) valid in the quantum-mechanical case.

[20] See Coleman (1977b), eq. (6.47).

[21] More precisely, at a given T one can have both solutions of the equations of motion with $O(4)$ symmetry and solutions with $O(3)$ symmetry. The solution with the lowest action dominates the path integral. At low T the $O(4)$ symmetric solution will dominate, and will reproduce the result of the previous subsection, while above a critical temperature the $O(3)$-symmetric solution will dominate.

Tunneling in QFT at finite temperature

We can finally generalize this formalism to finite-temperature QFT, which is the situation relevant to a phase transition in the early Universe. A first difference, with respect to the previous computation, is that the potential $V(\phi)$ must now be replaced by the effective potential at finite temperature, $V(\phi, T)$. Furthermore, we should not take the limit $\beta \to \infty$ in eq. (22.67), or in its field-theoretical generalization, since $\beta = 1/T$ is the temperature of the system. Thus, the tunneling rate is now obtained from the finite-temperature path integral

$$\int \mathcal{D}\phi e^{-S_4} \tag{22.112}$$

where

$$S_4[\phi] = \int_0^{1/T} d\tau \int d^3x \left[\frac{1}{2}\partial_\mu \phi \partial_\mu \phi + V(\phi, T)\right] \tag{22.113}$$

The boundary conditions in the path integral are $\phi(0, \mathbf{x}) = \phi(1/T, \mathbf{x})$, and we have finally set $\hbar = 1$. In the limit $T \to 0$ we recover the results of the previous subsection. Here we are rather interested in the opposite limit, in which T is large compared with $1/R$, where R is the typical bubble radius. In this limit the dependence of ϕ on Euclidean time becomes negligible, since the overall range of Euclidean times becomes small, and we can write

$$S_4[\phi] \simeq \frac{1}{T} S_3[\phi] \tag{22.114}$$

where

$$S_3[\phi] = \int d^3x \left[\frac{1}{2}\left(\partial_i \phi\right)^2 + V(\phi, T)\right] \tag{22.115}$$

and $\phi = \phi(\mathbf{x})$. In this limit it is natural to look for a solution of the equation of motion with $O(3)$ symmetry, $\phi = \phi_b(r)$, where $r = |\mathbf{x}|$[21] For an $O(3)$-symmetric solution $\phi(r)$, the action S_3 reduces to

$$S_3 = 4\pi \int_0^\infty dr r^2 \left[\frac{1}{2}\left(\frac{d\phi}{dr}\right)^2 + V(\phi, T)\right] \tag{22.116}$$

and the corresponding equation of motion is

$$\frac{d^2\phi}{dr^2} + \frac{2}{r}\frac{d\phi}{dr} = V'(\phi, T) \tag{22.117}$$

The boundary conditions are now

$$\lim_{r\to\infty} \phi(r) = 0 \tag{22.118}$$

in order to have a finite action, and

$$\left.\frac{d\phi}{dr}\right|_{r=0} = 0 \tag{22.119}$$

in order to ensure regularity at the origin. There are now only three zero modes, corresponding to the invariance under spatial translation, so Γ is proportional to V. The factor $1/\beta = T$ in eq. (22.69) is no longer canceled, but this is not a problem since now T is finite. Furthermore, each of the three zero modes contributes to the prefactor with a factor $[S_4/(2\pi)]^{1/2} = [S_3/(2\pi T)]^{1/2}$, so overall they give a factor $[S_3/(2\pi T)]^{3/2}$. Thus, the tunneling rate Γ is given by

$$\frac{\Gamma}{V} = Ae^{-S_{3,b}(T)/T} \tag{22.120}$$

where $S_{3,b}(T)$ is the action $S_3[\phi, T]$ evaluated on the bounce solution $\phi_b(r)$, and the prefactor A is given by

$$A = T\left(\frac{S_3(T)}{2\pi T}\right)^{3/2} \left|\frac{\det'\left[-\nabla^2 + V''\left(\phi_b\right)\right]}{\det\left[-\nabla^2 + V''(0)\right]}\right|^{-1/2} \tag{22.121}$$

A precise evaluation of the determinants in eq. (22.121) is in general difficult, even numerically. However, for most purposes it is sufficient to observe that (in units $\hbar = c = 1$) A has dimensions of (mass) ^{4}, as is clear from the fact that $\Gamma/V \sim (\text{ mass })^4$.[22] In the high-temperature limit the only relevant mass scale becomes T, so we must have $A = cT^4$, with c a number of order 1. Thus, finally,

$$\frac{\Gamma}{V} \simeq cT^4 e^{-S_{3,b}(T)/T} \tag{22.122}$$

At time t the rate of bubble nucleation in a Hubble volume $H^{-3}(t)$ is $(\Gamma/V) \times H^{-3}(t)$. The time t_* at which the phase transition takes place can then be obtained by requiring that the total number of bubbles nucleated from time $t = 0$ to $t = t_*$ is of order 1,

$$\int_0^{t_*} dt \frac{\Gamma}{VH^3(t)} = O(1) \tag{22.123}$$

We can transform the integral with respect to time into an integral with respect to temperature using $T \propto 1/a$, which gives $dT/T = -Hdt$. For transitions taking place in RD, which is the case of interest, using eq. (17.112) for $H^2\left(t_*\right)$ and setting $c = O(1)$, we find that the temperature T_* at which the phase transition takes place is determined by the condition

$$\int_{T_*}^{\infty} \frac{dT}{T}\left(\frac{45}{4\pi^3 g_*(T)}\right)^2 \left(\frac{M_{\rm Pl}}{T}\right)^4 e^{-S_3(T)/T} = O(1) \tag{22.124}$$

Once T_* has been determined in this way, the parameter α is given by eqs. (22.63) and (22.64). The information on the transition rate is in principle fully contained in eq. (22.122).

However, it can be useful to [22] Equivalently, we can directly observe that, in eq. (22.121), det' differs from det by the omission of three eigenvalues corresponding to the three zero modes, so

$$\frac{\det'\left[-\nabla^2+V''\left(\phi_b\right)\right]}{\det\left[-\nabla^2+V''(0)\right]}$$

has dimensions $1/\lambda^3$, where λ is the dimension of the eigenvalues. The latter is the same as the dimension of the operator ∇^2, i.e. (mass) ^{2}, so

$\left|\frac{\det'\left[-\nabla^2+V''(\phi_b)\right]}{\det[-\nabla^2+V''(0)]}\right|^{-1/2} \sim (\text{ mass })^3$.

[23] It should also be stressed that the whole formalism that we have discussed computes the nucleation rate assuming that quantum fluctuations in the thermal environment are responsible for the bubble nucleation. It is in principle possible that different mechanisms are at work, in which bubbles are nucleated by various forms of impurities and inhomogeneities of non-thermal origin. For instance, cosmic strings could act as nucleation sites, as could inhomogeneities due to some pre-existing turbulence; see Witten (1984). In this case β could be very different, and very difficult to predict.

Taylor-expand this expression around the transition time t_*, expressing T as a function of time, $T = T(t)$, and writing

$$\frac{S_{3,b}(T)}{T} \simeq \left(\frac{S_{3,b}(T)}{T}\right)_{t=t_*} + (t-t_*)\frac{d}{dt}\left(\frac{S_{3,b}(T)}{T}\right)_{t=t_*} \tag{22.125}$$

We then define

$$\beta = -\frac{d}{dt}\left(\frac{S_{3,b}(T)}{T}\right)_{t=t_*} \tag{22.126}$$

so, in a small interval of times near $t = t_*$, eq. (22.122) becomes

$$\Gamma(t) \simeq \Gamma_* e^{\beta(t-t_*)} \tag{22.127}$$

Again using $dT/T = -Hdt$, eq. (22.126) becomes

$$\frac{\beta}{H_*} = T_*\frac{d}{dT}\left(\frac{S_{3,b}(T)}{T}\right)_{T=T_*} \tag{22.128}$$

Notice that, in order to compute α, it is sufficient to know the transition temperature T_* and the potential $V(\phi, T)$, while to compute β one must also know how the action $S_{3,b}(T)$ depends on T in a neighborhood of T_* [23]

Equations (22.124) and (22.128) allows one in principle to perform an accurate numerical computation of T_* and of β. We can, however, also obtain an analytic understanding of typical orders of magnitude that one would obtain for β. The integral in eq. (22.124) will in general be dominated by its value at $T = T_*$, so in a first approximation

$$\left(\frac{45}{4\pi^3 g_*\left(T_*\right)}\right)^2\left(\frac{M_{\rm Pl}}{T_*}\right)^4 e^{-S_3(T_*)/T_*} = O(1) \tag{22.129}$$

Setting $g_*\left(T_*\right)$ to its high-temperature value in the Standard Model, $g_* \simeq 106.75$, this gives

$$\frac{S_3\left(T_*\right)}{T_*} \simeq 4\log\frac{M_{\rm Pl}}{T_*} - 11.4 \tag{22.130}$$

For a transition at the electroweak scale, $T_* = O(100\text{GeV})$ and we get $S_3\left(T_*\right)/T_* \simeq 145$, while, for the QCD scale, $S_3\left(T_*\right)/T_* \simeq 173$. In general, barring fine tuning, we expect that

$T_* d/dT \left[S_3(T)/T\right]_{T_*}$ will be of the same order of magnitude as $S_3\left(T_*\right)/T_*$, so as a rough estimate we expect $\beta/H_* = O(100)$, both at the electroweak scale and at the QCD scale. This estimate is roughly confirmed by the explicit numerical computations. Of course, precise numbers can only obtained with an accurate numerical evaluation of eq. (22.128), given the potential $V(\phi, T)$, and values of β/H_* as low as $O(10)$ or as high as $O\left(10^3\right)$ can be obtained in different models. In general α, β and T_* are correlated and, in typical models, as α increases, so that the energy released in the transition increases, β and T_* decrease.

Finally we observe that, in the thin-wall limit, eq. (22.116) allows us to formalize the physical intuition, discussed at the beginning of this section, that the dynamics of true vacuum bubbles is determined by a competition between the volume energy gained by tunneling to the true vacuum, and the effect of surface tension, which tends to shrink the bubbles to zero. Indeed, in the thin-wall limit the bubble configuration has $\phi \simeq \phi_0$ for $r < R$, and $\phi \simeq 0$ and $V(\phi) \simeq 0$ for $r > R$, with the transition taking place in a small range δr near $r = R$. Then the term proportional to the potential in eq. (22.116) is negative and proportional to the bubble volume,

$$4\pi \int_0^\infty dr r^2 V(\phi, T) \simeq 4\pi \int_0^R dr r^2 V(\phi, T) = -|\bar{V}(T)| \frac{4\pi R^3}{3} \tag{22.131}$$

where $\bar{V}(T) < 0$ is the average of the potential inside the bubble volume, which is of order $V\left(\phi_0(T), T\right)$. The term proportional to $(d\phi/dr)^2$ in eq. (22.116) can instead be simplified by observing that, in the thin-wall limit, $(d\phi/dr)$ is sharply peaked around $r = R$, so

$$\begin{aligned} 4\pi \int_0^\infty dr r^2 \frac{1}{2} \left(\frac{d\phi}{dr}\right)^2 &\simeq 4\pi R^2 \int_0^\infty dr \frac{1}{2} \left(\frac{d\phi}{dr}\right)^2 \\ &\equiv 4\pi R^2 S_1(T) \end{aligned} \tag{22.132}$$

So, this term is positive and proportional to the bubble surface, and we recognize the contribution from the surface tension. Thus, in the thin-wall approximation, on a bubble of radius R,

$$S_3(R) = -|\bar{V}(T)| \frac{4\pi R^3}{3} + 4\pi R^2 S_1(T) \tag{22.133}$$

The bounce solution is an extremum of this action in the whole space of field configurations, so in particular with respect to the parameter R. The bubble radius R in the thin-wall approximation can therefore be determined by minimizing eq. (22.133) with respect to R, which gives

$$R(T) = \frac{2 S_1(T)}{|\bar{V}(T)|} \tag{22.134}$$

16.2.4 22.4.4 Bubble dynamics and GW production

In a first-order phase transition, once these supercritical bubbles have nucleated, their subsequent dynamics depends on their environment. Conceptually, the simplest situation is that of vacuum bubbles, i.e. bubbles that are assumed to expand in the vacuum, neglecting their interaction with the cosmic fluid. In that case, the bubbles expand until their walls move at a speed close to the speed of light. The energy gained in the transition from the metastable state

to the ground state is transferred to the kinetic energy of the bubble wall. As the bubble expands, more and more regions of space convert to the ground state, and the wall becomes more and more energetic. At the same time, it also becomes thinner, and therefore the energy density stored in the wall increases very rapidly. Of course, as long as we have a single spherical bubble, this large energy cannot be converted to GWs. However, when two bubbles collide, spherical symmetry is broken and GW production can take place.

More realistically, in the cosmological setting in which we are eventually interested, the bubbles interact with the cosmic fluid. This interaction can in principle be quite complex, and results in friction over the bubble wall, which slows down and changes the bubble evolution, while at the same time the cosmic fluid is stirred by the expanding bubbles, with the generation of turbulent motions and acoustic waves. Indeed, three main mechanisms for GW production have been identified, namely GWs from bubble collisions, from turbulence and from colliding acoustic waves in the fluid. We examine them in turn.

GWs from bubble collisions

Before discussing the results obtained in the literature from numerical computations, let us first estimate the orders of magnitude involved, for the peak frequency and the amplitude of the stochastic background produced by bubble collisions.

At the time of production, the characteristic frequency scale of the process is given by the parameter β that appears in eq. (22.127); thus the typical frequency f_* at the time of production can be estimated as $2\pi f_* \sim \beta$.[24] Inserting this value into eq. (22.6) and setting $g_* \simeq 106.75$, as is appropriate at the electroweak scale, the frequency of the GWs produced, redshifted to the present epoch, can then be estimated to be peaked at

$$f_{\rm peak} \;\sim\; \text{a few} \;\times 10^{-6} \left(\frac{\beta}{H_*}\right) \left(\frac{T_*}{100\text{GeV}}\right) \text{Hz} \tag{22.135}$$

Thus, if there is a first-order phase transition at the electroweak scale $T_* = O(100)$GeV, for values of β/H_* in the typical range $10 - 10^3$ we find that f_0 is in the range $10^{-4} - 10^{-2}$ Hz, which is just in the frequency band of the LISA space-interferometer.

Similarly, if there is a first-order phase transition in QCD at $T_* \simeq 150$MeV, we get

$$f_{\rm peak} \;\sim\; \text{a few} \;\times 10^{-9} \left(\frac{\beta}{H_*}\right) \left(\frac{T_*}{150\text{MeV}}\right) \text{Hz} \tag{22.136}$$

If β/H_* is not too large, this could potentially be within the frequency window of pulsar timing arrays, as we will see in Chapter 23.

To get a first rough order-of-magnitude estimate of the value of $\Omega_{\rm gw}(f)$ at the peak frequency, we can observe that the typical wavelength λ_* of the radiation, at the time of production, is of the order of the typical radius R_* of the bubbles when they collide. Then, the energy liberated in

[24] Actually, when the bubble velocity v_b is not close to the speed of light, there are two different time- or lengthscales in the problem. One is the typical time-scale of the phase transition, β^{-1}, which can also identified with the duration of the phase transition. The other is the typical size of the bubbles near the end of the phase transition, when most bubble collisions take place, which is rather given by $R_* = v_b\beta^{-1}$. One should then ask which of these two scales is more appropriate for estimating the peak frequency through eq. (22.6). The answer is quite subtle, and depends on the detailed time evolution of the energy-momentum tensor of the source. It turns out that, in a first approximation, β^{-1} is appropriate for GWs from bubble collisions, while $v_b\beta^{-1}$ is appropriate for turbulence; see the Further Reading. A fit to the dependence on v_b for bubble collisions will be given in eqs. (22.144) and (22.145).

GWs in the collision of two bubbles is of order $E_{GW} \sim GM_B^2/\lambda_*$, where $M_B \sim \rho_{\rm vac}\lambda_*^3$ is the energy of a typical bubble and $\rho_{\rm vac}$ is the difference in energy density from the metastable state to the ground state. The fraction of the false-vacuum energy that goes into GWs instead of going into radiation is given by

$$\frac{\Omega_{\rm gw}}{\Omega_R} \sim \frac{E_{\rm GW}}{M_B} \tag{22.137}$$

This is, however, valid only for bubble collisions in vacuum, where the energy gained in the transition to the true vacuum can only go into the kinetic energy of the bubble walls, i.e., at the fundamental level, into the kinetic energy of the scalar field. For bubbles evolving in a primordial plasma, the underlying physics is quite different. Indeed, since the energy gained in the transition grows with the bubble radius as R^3, while the energy stored into the bubble walls only grows as R^2, in the presence of other channels only a tiny fraction of the vacuum energy will go into the scalar field. Most of the energy goes into the primordial fluid, where part of it is dissipated. We must therefore insert an efficiency factor κ, which gives the fraction of the available energy that actually contributes to GW generation, and replace M_B by κM_B. Then

$$E_{GW} \sim \kappa^2 \frac{GM_B^2}{\lambda_*} \tag{22.138}$$

Therefore

$$\frac{E_{\rm GW}}{M_B} \sim \kappa^2 \frac{GM_B}{\lambda_*} \sim \kappa^2 G\rho_{\rm vac}\lambda_*^2 \tag{22.139}$$

Using $G\rho_{\rm vac} \sim H_*^2$, we get, at the time of production

$$\Omega_{\rm gw} \sim \kappa^2 \lambda_*^2 H_*^2 \Omega_R \tag{22.140}$$

For a phase transition with characteristic time-scale β we have $\lambda_* = 1/f_* \sim 1/\beta$, and therefore we can also rewrite it as

$$\Omega_{\rm gw} \sim \kappa^2 \left(\frac{H_*}{\beta}\right)^2 \Omega_R \tag{22.141}$$

Despite the crudeness of the approximation, this relation allows us to get a first feeling for the numbers involved. Equation (22.141) suggests that, even in the most optimistic case $\kappa = 1$ (which, as we will see, is far from being realized in realistic situations), for typical values of β/H_* in the range range $10 - 10^3, \Omega_{\rm gw}$ at the peak ranges from a (very optimistic) $10^{-2}\Omega_R \sim 10^{-7}$ to $10^{-6}\Omega_R \sim 10^{-11}$. As we will see, however, the efficiency factor κ and other effects can significantly lower these estimates. It is also interesting to observe that the dependence on $\epsilon_* = \lambda_* H_*$ in eqs. (22.6) and (22.140) is such that, as ϵ_* decreases, the peak frequency increases as $1/\epsilon_*$ while the amplitude at the peak decreases as ϵ_*^2.

To compute the GW spectrum more quantitatively, one can begin by considering the collision of two vacuum bubbles. This can be simulated

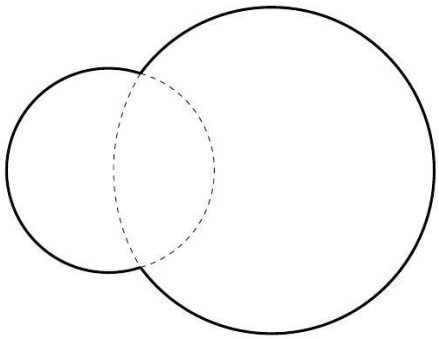


Fig. 22.11 The envelope approximation for bubble collision.

numerically, studying the dynamics of the underlying scalar field and computing the GW production from the time-evolving energy-momentum tensor of the scalar fields. For bubbles evolving in vacuum, it turns out that the numerical results are well reproduced, with an accuracy of about 20%, by the so-called envelope approximation. This is an analytic approximation in which all the energy is assumed to be stored in the bubble walls, assumed to be thin, and, as the two bubbles collide, one uses as a source for GW production the energy-momentum tensor of the uncollided part of the bubble walls, represented by the thick lines in Fig. 22.11, while the complicated overlap region between the colliding bubbles is simply neglected. This approximation can then be used to study the GWs produced by nucleating at random points in space an ensemble of bubbles. The number of bubbles nucleated in a time step δt is drawn from a Poisson distribution whose mean is determined by the bubble nucleation rate $\Gamma(t)$ computed in Section 22.4.3, and the results are then averaged over several realization of the nucleation process. The envelope approximation allows one to compute the cumulative GWs produced by collisions in a system with hundreds of bubbles, which would be unfeasible if one had to evolve numerically the profile of the scalar field of each bubble. With this combination of analytic and numerical techniques, Kamionkowski, Kosowsky and Turner (1994) found that the peak value of $h_0^2\Omega_{\rm gw}$ is well fitted by

$$h_0^2\Omega_{\rm gw}\left(f_{\rm peak}\right) \simeq 1.1\times 10^{-6}\kappa^2\left(\frac{H_*}{\beta}\right)^2\left(\frac{\alpha}{1+\alpha}\right)^2\left(\frac{v_b^3}{0.24+v_b^3}\right)\left(\frac{100}{g_*}\right)^{1/3} \tag{22.142}$$

where α, defined in eq. (22.64), determines the strength of the phase transition, and v_b is the average velocity of the bubble walls at collision. Comparing with eq. (22.141) we see that the crude estimate leading to eq. (22.141) correctly reproduces the dependence on κ and on H_*/β, and even gives the correct order of magnitude of the overall numerical factor, in the limit of strong first-order phase transitions, $\alpha \gg 1$, and relativistic bubble velocities. If α or v_b are small, however, there are further suppression factors. Equation (22.142) has been widely used in the literature. Later, numerical work with increased accuracy [25] has confirmed the basic features of this result and provided a more accurate fitting formula,

$$h_0^2\Omega_{\rm gw}\left(f_{\rm peak}\right) \simeq 1.84\times 10^{-6}\kappa^2\left(\frac{H_*}{\beta}\right)^2\left(\frac{\alpha}{1+\alpha}\right)^2\left(\frac{v_b^3}{0.42+v_b^2}\right)\left(\frac{100}{g_*}\right)^{1/3} \tag{22.143}$$

while the peak frequency is well fitted by

$$f_{\rm peak} \simeq 3.83\times 10^{-6}\ {\rm Hz}\left(\frac{\beta}{H_*}\right)\left(\frac{T}{100{\rm GeV}}\right)\left(\frac{g_*}{106.75}\right)^{1/6} g\left(v_b\right) \tag{22.144}$$

[12525] See Huber and Konstandin (2008a).

where

$$g\left(v_b\right) \simeq \frac{2.7}{1.8 - 0.1 v_b + v_b^2} \tag{22.145}$$

is a function of v_b, which monotonically decreases from $g(0) = 1.5$ to $g(1) = 1$. We see that the approximate estimate (22.135) is consistent with the numerical result (22.144). [26]

For QCD we rather use as a reference value $g_* = 64.75$, which takes into account all degrees of freedom that are relativistic just above the transition (including the strange quark), [27] and setting $T_* = 150\text{MeV}$. Then eq. (22.144) reads

$$f_{\text{peak}} \simeq 5.28 \times 10^{-9}\ \text{Hz} \left(\frac{\beta}{H_*}\right) \left(\frac{T}{150\text{MeV}}\right) \left(\frac{g_*}{64.75}\right)^{1/6} g\left(v_b\right) \tag{22.146}$$

The latter numerical study also allowed determination of the frequency spectrum in the region near the peak. In general, at sufficiently low frequency, such that the corresponding wavelength is well outside the horizon, causality requires that $h_0^2 \Omega_{\text{gw}}(f) \propto f^3$; see eq. (22.3) and the discussion after it. Existing numerical studies of GW production in bubble collisions in the envelope approximation only explore the frequency region not very far from the peak, and do not reach the regime of superhorizon modes. For frequencies lower than the peak, but not yet in the super-horizon regime, the frequency dependence is well fitted by a power $h_0^2 \Omega_{\text{gw}}(f) \propto f^{2.8}$, not far from the f^3 behavior expected for the even lower frequencies corresponding to super-horizon modes. At frequencies higher than the peak, in contrast, the numerical study shows that $h_0^2 \Omega_{\text{gw}}(f)$ decays only as $1/f$. This relatively slow decay can be quite interesting from the observational point of view, since it means that a sizable signal could still be present in the bandwidth of a detectors, such as LISA, even when the peak of the spectrum lies at a lower frequency.

The next step is to compute κ and v_b in terms of the parameters of the phase transition. In the physically relevant case of bubbles evolving in the cosmic fluid this is a complicated problem, and the result depends on the detailed dynamics of the bubbles. In general, there are two possible "combustion" modes in the collision of two bubbles: detonation, which basically takes place when the boundaries propagate faster than the speed of sound, and deflagration, when instead they move slower. Only in the first case can we expect a large production of GWs. In deflagration there will be no large concentration of kinetic energy near the bubble walls, so GW production is suppressed. The bubble velocity has a relatively simple form in the case of Chapman-Jouguet detonation (i.e. when the detonation propagates at a speed such that, in the frame of the shock wave, the fluid velocity is exactly equal to the speed of sound). In this case one can show that v_b is fixed in terms of α by

$$v_b(\alpha) = \frac{\sqrt{1/3} + \sqrt{\alpha^2 + 2\alpha/3}}{1 + \alpha} \tag{22.147}$$

[28] See again Kamionkowski, Kosowsky and Turner (1994)

[29] See Espinosa, Konstandin, No and Servant (2010) for more recent estimates of $\kappa(\alpha)$ in different regimes.

Observe also that, while for bubbles propagating and colliding in vacuum the source of GW production is the energy-momentum tensor of the scalar field, for bubbles propagating in the cosmic fluid by far the dominant contribution is given by the energy-momentum tensor $T_{ij} = (\rho + p) v_i v_j$ of the fluid displaced and stirred by the bubbles. In the envelope approximation one

[25][26] See also Jinno and Takimoto (2017) for related analytic results in the thinwall and envelope approximation.

[27] A listing of the degrees of freedom as a function of mass can be found in Table I of Watanabe and Komatsu (2006). As we saw in eq. (17.99), the contribution of fermions must be multiplied by a factor 7/8.

assumes again that only the fluid at the envelope contributes (an assumption to which we will return later). Assuming Jouguet detonation and the envelope approximation, one can estimate [28] that

$$\kappa(\alpha) = \frac{1}{1+A\alpha}\left(A\alpha + \frac{4}{27}\sqrt{\frac{3\alpha}{2}}\right) \tag{22.148}$$

where $A \simeq 0.715$. Equations (22.147) and (22.148) have been widely used in the literature, because of their relative simplicity. However, to model bubble evolution in a plasma more realistically it is also necessary to include terms representing friction between the bubble walls and the plasma. These have been parametrized phenomenologically by adding to the equation of motion of the scalar field a term proportional to $\eta u^\mu \partial_\mu \phi$, where u^μ is the plasma four-velocity and η a friction coefficient. Then, κ and v_b are also functions of η. In general, only for a limited range of values of the parameters α and η does the solution correspond to detonation, while the most probable outcome is either deflagration or a runaway solution where the walls continuously accelerate. [29]

Finally, the values of α and β should in principle be derived from the fundamental theory under consideration, using the nucleation theory discussed in Section 22.4.3. This has been done for some selected models that could give a first-order phase transition at the electroweak scale, such as non-minimal supersymmetric extensions of the Standard Model; see the Further Reading. It is clear that, in the end, the computation of GW production from bubble collisions suffers from many uncertainties, first of all because at present these extensions of the Standard Model are unavoidably somewhat speculative and, on a more technical side, because of the uncertainties regarding bubble evolution in a realistic thermal environment. In any case, existing predictions at the electroweak scale in general do not exceed $h_0^2\Omega_{\rm gw} \sim 10^{-12}$ at the peak frequency. Figure 22.12 shows the signal obtained for different typical values of the parameters α and β/H_*, comparing the result from the collision of two bubbles (left panel) with the result for the nucleation and collision of an ensemble of bubbles (right panel). The slower decay with frequency in the latter case is due to the cumulative effect from the late nucleation of small bubbles, which is absent in the two-bubble case. Observe that the signal is too small for a space interferometer such as LISA but, thanks to the slow $1/f$ decrease at high frequencies, for some values of the parameters it could be detectable at an advanced space interferometer such as DECIGO or BBO.

It is important to observe that, when the envelope approximation is applied to bubbles evolving in the primordial plasma, rather than in vacuum, we are actually assuming that the main source of GW gener-

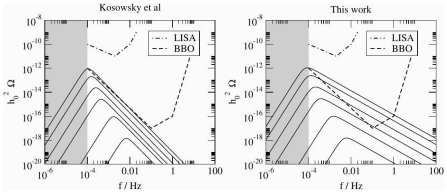


Fig. 22.12 $h_0^2\Omega_{\rm gw}(f)$ for several values of α and β/H_*. Left panel, using the numerical results from the collision of two bubbles studied in Kosowsky, Turner and Watkins (1992a, 1992b). Right panel, using the envelope approximation for an ensemble of bubbles, from the numerical study of Huber and Konstandin (2008b). The curves from top to bottom correspond to values of $(\alpha, \beta/H_*)$ equal to $(0.20, 30), (0.15, 40), (0.10, 60), (0.07, 100), (0.05, 300)$ and $(0.03, 1000)$, respectively. The corresponding transition temperature decreases from 130GeV (top) to 70GeV (bottom). From Huber and Konstandin (2008b). CCSISSA Medialab Srl. Reproduced by permission of IOP Publishing. All rights reserved.

ation is the energy-momentum tensor of the fluid near the envelope of the colliding bubbles.

This is an assumption that can actually miss a potentially dominant contribution to GW production, as we will discuss later.

GWs from turbulence in the cosmic fluid

Collision of bubbles injects a large amount of energy into the cosmic fluid. At the electroweak epoch this fluid has a very high Reynolds number, of order 10^{13}, and therefore turbulent motion rapidly sets in. Furthermore, even the magnetic Reynolds number is very large, so magnetic fields generated during the phase transition are amplified and full magnetohydrodynamic (MHD) turbulence develops. When the phase transition terminates, so does the energy injection. However, because of the very high Reynolds number, MHD turbulence decays quite slowly, over several Hubble times, and therefore provides a long-lasting source of GWs. The peak frequency $f_{\rm peak}$ turns out to be determined by the size R_* of the bubbles at collisions, $k_{\rm peak} = 2\pi f_{\rm peak} \simeq 1/R_*$; see the discussion in Note 24 on page 676. At low frequencies one has a spectrum $h_0^2\Omega_{\rm gw}(k) \propto k^3$, as required by causality (at least in the limit of superhorizon modes). On the high-frequency side the GW spectrum decays as k^{-n}, but the value of n depends strongly on how turbulence is modeled, in particular on the power spectrum of the turbulent fluid motion and on whether the source is modeled as short-lasting or long-lasting. In the more appropriate case of a long-lasting source analytical estimates suggest $n = 5/3$ if the turbulent fluid motion has a Kolmogorov spectrum,
[30] See Caprini, Durrer and Siemens (2010).

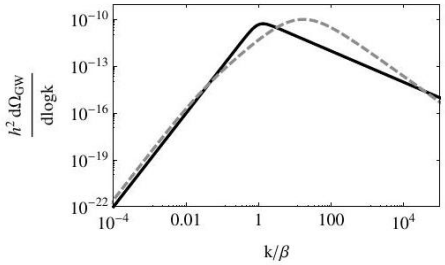


Fig. 22.13 The GW spectra from bubble collisions (solid) and MHD turbulence (dashed) for $v_b = 0.7$, $\kappa\alpha/(1+\alpha) = 0.1, \beta/H_* = 10$ and $T_* = 100$MeV (as appropriate for a phase transition at the QCD scale). From Caprini, Durrer and Siemens (2010).

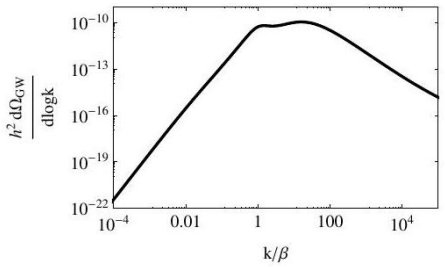


Fig. 22.14 The total GW spectrum resulting from the sum of the two contributions in Fig. 22.13. From Caprini, Durrer and Siemens (2010). and $n = 3/2$ for a Iroshnikov-Kraichnan spectrum. Probably, only full MHD numerical simulations can address the issue precisely. A fit to the numerical results for MHD turbulence is given by [30]

$$h_0^2\Omega_{\rm gw}\left(f_{\rm peak}\right)\big|_{\rm turb} \simeq 3.3\times 10^{-5}\left(\frac{\kappa\alpha}{1+\alpha}\right)^{3/2}\left(\frac{H_*}{\beta}\right)v_b \times \frac{\left(f/f_{\rm peak}\right)^3}{\left(\left[1+\left(f/f_{\rm peak}\right)\right]^{11/3}\left(1+4k_*/H_*\right)\right.} \tag{22.149}$$

where k_* is the physical wavenumber at the time of production, and

$$f_{\rm peak} \simeq 2.7 \times 10^{-5}\ {\rm Hz} \frac{1}{v_b} \left(\frac{\beta}{H_*}\right) \left(\frac{T_*}{100{\rm GeV}}\right) \left(\frac{g_*}{100}\right)^{1/6} \tag{22.150}$$

These formulas include both the contribution to GW production from the turbulent kinetic motion of the fluid, and that due to the magnetic field, which are generated by the turbulent motion itself, and, for simplicity, assumes equipartition, i.e. the same amount of energy density in the turbulent kinetic motion and in magnetic fields.

In general, the GW spectrum produced by turbulence peaks at a higher frequency than that due to bubble collisions, because of the factor $1/v_b$ in eq. (22.150), and the overall amplitude can be somewhat larger. The two contributions are shown separately in Fig. 22.13, and their sum is shown in Fig. 22.14. We see that the lower peak from bubble collisions causes the hump on the left of the true peak of the spectrum, which could be a distinctive feature of this signal.

GWs from sound waves in the cosmic fluid

The picture discussed in the preceding subsections basically assumes that the primordial plasma can be described as an incompressible and turbulent fluid. Recently, it has become possible to test this assumption using fully three-dimensional simulations of bubble nucleation in the coupled system of the scalar field and the fluid, and to compute the resulting GW production. The numerical simulations indicate that sound waves in the fluid, created by the bubble collisions, are important. These are completely missed in the envelope approximation, where the fluid on the colliding part of the bubble walls (the dashed lines in Fig. 22.11) is simply neglected. These sound waves, which describe coherent motions of the fluid, are, however, a potentially important source of GWs. Furthermore, they last for a long time after the phase transition has been completed. This boosts their GW signal by the ratio of their lifetime, which can be of the order of the Hubble time, to the duration of the transition, i.e. by a factor H_*^{-1}/β^{-1}. For a typical value $\beta/H_* \sim 10^2$, the resulting GW signal can then be two orders of magnitudes larger than that due to bubble collisions.

To describe the coupled field-fluid system one starts from the energy-momentum tensor of the total system,

$$T^{\mu\nu}_{\rm tot} = T^{\mu\nu}_{\rm field} + T^{\mu\nu}_{\rm fluid} \tag{22.151}$$

where

$$T^{\mu\nu}_{\rm field} = \partial^\mu\phi\partial^\nu\phi - \frac{1}{2}\eta^{\mu\nu}(\partial\phi)^2 \tag{22.152}$$

and

$$T^{\mu\nu}_{\rm fluid} = (\rho + p)u^\mu u^\nu + \eta^{\mu\nu}p \tag{22.153}$$

where p is the pressure in the rest frame, ρ is the energy density, and u^μ the four-velocity of the fluid; compare with eq. (10.30). Over the relatively short time-scale of the phase transition one can in a first approximation neglect the expansion of the Universe and use the flat-space metric. The fundamental interaction between the field ϕ responsible for the bubble profile and the matter fields composing the primordial plasma results in an effective friction on the bubble. To describe this fieldfluid interaction phenomenologically it is common to split the energymomentum conservation equation $\partial_\mu T^{\mu\nu}_{\rm tot} = 0$ as

$$\partial_\mu T^{\mu\nu}_{\rm field} = \delta^\nu \tag{22.154}$$

$$\partial_\mu T^{\mu\nu}_{\rm fluid} = -\delta^\nu \tag{22.155}$$

where

$$\delta^\nu = \eta u^\mu \partial_\mu \phi \partial^\nu \phi \tag{22.156}$$

and η is an adjustable friction parameter. Of course this splitting, and the precise choice of δ^ν, are somewhat arbitrary, but this latter term, proportional to the fluid four-velocity, is believed to give a good phenomenological description of the friction-induced interaction between the field and the fluid. Starting from a system in the symmetric phase, one models as usual bubble nucleation with the rate given in eq. (22.65), and from this one generates a set of nucleation times and locations, at each of which one inserts a static bubble with a Gaussian profile for the scalar field. One can then evolve the system numerically on a lattice, and compute the time evolution of the energy-momentum tensor, as well as the resulting GW production. In this way one finds [31] that the GW energy $\Omega_{\rm GW}$ from acoustic waves is related to the value $\Omega_{\rm GW}^{\rm ea}$ obtained in the envelope approximation by

$$\frac{\Omega_{\rm GW}}{\Omega_{\rm GW}^{\rm ea}} \gtrsim 60 C_{\rm GW} \left(\frac{\beta}{H_*}\right) \tag{22.157}$$

where $C_{\rm GW}$ is a dimensionless parameter. The numerical simulation gives $C_{\rm GW} \sim 0.04$. Since β/H_* can easily be of order 10^2, this "acoustic" production mechanism can give a signal orders of magnitude larger than the standard estimates based on the envelope approximation. The peak frequency is estimated to be comparable to the value (22.150) obtained from turbulence. One should keep in mind, however, that these results were obtained in simulations with values of $\alpha \lesssim 0.1$ and small fluid velocities. The extrapolation to higher values of α might be nontrivial, with the possible conversion of part of this acoustic signal into the turbulent one discussed earlier. [32] [31] See Hindmarsh, Huber, Rummukainen and Weir (2014, 2015).

[32] See Section 2.1.2 of Caprini et al (2016).

[33] More formally, the existence of this topologically non-trivial configuration is due to the fact that the first homotopy group of the vacuum manifold, $\pi_1(\mathcal{M})$, is non-trivial. The first homotopy group classifies the mappings $S^1 \to \mathcal{M}$ from a circle to the vacuum manifold $\mathcal{M}$, which for a $U(1)$ group is itself a circle S^1.

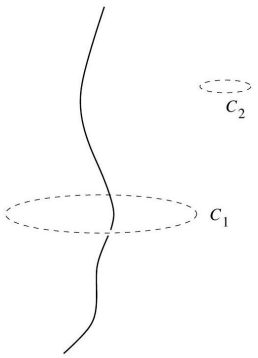


Fig. 22.15 A infinite string caught inside the circle $\mathcal{C}_1$.

16.3 22.5 Cosmic strings

16.3.1 22.5.1 Global and local strings

Cosmic strings are configurations of scalar and/or gauge fields whose energy density is concentrated along a one-dimensional line, either an infinite line or a closed loop. Such objects might have formed in the early Universe during a symmetry-breaking phase transitions (independently of the order of the phase transition) in extensions of the Standard Model. The

prototype example is given by the breaking of a $U(1)$ symmetry, due to a complex scalar field Φ whose potential, in the broken phase, has the standard Mexican hat potential,

$$V(\Phi) = \frac{1}{4}\lambda\left(|\Phi|^2 - \eta^2\right)^2 \tag{22.158}$$

If we write $\Phi = |\Phi|e^{i\psi}$, then at the minimum of the potential we have $|\Phi| = \eta$, but the phase ψ is undefined. Therefore there is not just a single vacuum state, but a whole manifold $\mathcal{M}$ of vacua, which in this case is just a circle, parametrized by an angle ψ, with $0 \leqslant \psi \leqslant 2\pi$. Since Φ is single-valued, as we go around a closed path in physical space, the phase ψ must change by $\Delta\psi = 2\pi n$, where $n = 0, \pm 1, \pm 2, \ldots$ If $n \neq 0$, as we go around the circle in physical space, the phase ψ wraps one or more times around this circle in the field space. We then have a configuration that is topologically non-trivial, and cannot be deformed continuously into the topologically trivial configuration $\psi = 0$.[33] Consider now a circle $\mathcal{C}_1$ in physical space, such that along it $\Delta\psi = 2\pi$, and study what happens if we shrink it to a point or, more precisely, to an infinitesimally small circle $\mathcal{C}_2$ with $\Delta\psi = 0$. In order for n to change discontinuously from $n = 1$ to $n = 0$ it is necessary that, in the process, a singularity be encountered, i.e. we must encounter at least a point where the phase ψ is undefined, which happens when the modulus $|\Phi| = 0$. This means that at least one string of false vacuum must be caught inside a loop with $\Delta\psi = 2\pi$; see Fig. 22.15. Furthermore, the string must either be infinitely long or be a loop that winds around the circle $\mathcal{C}_1$, otherwise we would be able to unwind the circle $\mathcal{C}_1$ from this line of singularities. Of course, in a cosmological setting, an infinite loop should rather be seen as an idealization of a closed loop with size much larger than the horizon. At the string location we have $\Phi = 0$, so we are in the false vacuum, and there is an associated energy density. Thus, we have a concentration of energy density along a string.

To be more quantitative, we have to specify the Lagrangian, and in particular we will see that crucial differences emerge depending on whether the $U(1)$ symmetry that is broken is a global or a local symmetry. The corresponding strings are called global or local strings, respectively. Consider first a theory with a global $U(1)$ symmetry, with just a complex scalar field Φ and Lagrangian density

$$\mathcal{L} = -\partial_\mu\Phi\partial^\mu\Phi^* - V(\Phi) \tag{22.159}$$

with $V(\Phi)$ given by eq. (22.158) (we limit ourselves for the moment to flat space). For $\eta \neq 0$ we can write

$$\Phi(x) = \left(\eta + \frac{\varphi(x)}{\sqrt{2}}\right)e^{i\psi(x)} \tag{22.160}$$

and we get

$$\begin{aligned}\mathcal{L} &= -\frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - \frac{\lambda}{4}\left(\sqrt{2}\eta\varphi + \frac{1}{2}\varphi^2\right)^2 - \left(\eta + \frac{\varphi}{\sqrt{2}}\right)^2\partial_\mu\psi\partial^\mu\psi \\ &= \mathcal{L}_2 + \mathcal{L}_{\rm int}\end{aligned} \tag{22.161}$$

where the quadratic term is

$$\mathcal{L}_2 = -\frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - \frac{1}{2}m_s^2\varphi^2 - \eta^2\partial_\mu\psi\partial^\mu\psi \tag{22.162}$$

with $m_s^2 = \lambda\eta^2$, and $\mathcal{L}_{\rm int}$ contains the interaction terms. We therefore have a massless scalar field ψ, which is the Goldstone boson associated with the breaking of the global $U(1)$ symmetry, and a massive scalar field φ with mass m_s.

The equation of motion derived from the Lagangian (22.159) is

$$\Box\Phi = \frac{\lambda}{2}\left(|\Phi|^2 - \eta^2\right)\Phi \tag{22.163}$$

We can look for a static solution representing a string along the z axis, of the form

$$\Phi(\rho,\theta) = \eta f\left(m_s\rho\right)e^{in\theta} \tag{22.164}$$

where (ρ,θ) are polar coordinates in the (x,y) plane, and the integer n is the winding number of the string. Observe that, in terms of the phase ψ of the field Φ, this ansatz means $\psi(\rho,\theta) = n\theta$. Substituting into the equations of motion, one finds that near the origin $f\left(m_s\rho\right) \propto \rho^{|n|}$ while, for $m_s\rho \gg 1$,

$$f\left(m_s\rho\right) \simeq 1 - O\left(\frac{1}{m_s^2\rho^2}\right) \tag{22.165}$$

The energy density of this configuration is

$$\begin{aligned}\mathcal{E} &= \partial_i\Phi\partial_i\Phi^* + V(\Phi)\\ &= \left|\partial_\rho\Phi\right|^2 + \frac{1}{\rho^2}\left|\partial_\theta\Phi\right|^2 + V(\Phi)\end{aligned} \tag{22.166}$$

Using the asymptotic form of $f\left(m_s\rho\right)$ we see that the energy density is peaked around $\rho \sim m_s^{-1}$, and vanishes at large distances as $1/\rho^2$. The $1/\rho^2$ behavior at infinity is due to the $\partial_\theta\Phi$ term in eq. (22.166). Indeed, $\partial_\theta\Phi = in\eta f\left(m_s\rho\right)e^{in\theta}$, so at large distances $|\partial_\theta\Phi| \to n\eta$. Then, as $\rho\to\infty$

$$\mathcal{E}(\rho,\theta) \to \frac{n^2\eta^2}{\rho^2} \tag{22.167}$$

Since the energy density is localized in a region of width $\sim m_s^{-1}$ (which is the Compton wavelength of the massive state described by φ) and
goes to zero as $1/\rho^2$ at infinity, this configuration describes a localized string. However, when we integrate $\mathcal{E}$ over $dxdy = \rho d\rho d\theta$ to get the energy of the configuration per unit length, the $1/\rho^2$ behavior is not integrable, and the energy per unit length diverges logarithmically. The string tension μ, i.e. the energy per unit length, is given by

$$\mu \simeq 2\pi n^2\eta^2\log\left(m_sR\right) \tag{22.168}$$

where R is the infrared cutoff, and we have taken m_s^{-1} as the order of magnitude of the distance where the asymptotic behavior is no longer valid and the behavior $f\left(m_s\rho\right) \propto \rho^{|n|}$ sets in. This infrared divergence is a reflection of the existence in the spectrum of the Goldstone boson θ, which, being massless, induces long-range interactions between different parts of the string. In a condensed matter system the infrared divergence is cut off by the finite volume of the system. Indeed, global strings of this type are observed in condensed matter, in particular as quantized vortex lines in liquid helium. More generally, in a system with several global strings the cutoff will be provided by the average distance between strings, and for a string forming a loop this will the loop size itself.

Consider now a model with spontaneous breaking of a local $U(1)$ symmetry. The prototype example is the Abelian Higgs model, with Lagrangian density

$$\mathcal{L} = -\left|D_\mu\Phi\right|^2 - V(\Phi) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{22.169}$$

where again $V(\Phi)$ is given by eq. (22.158) and the covariant derivative is given by $D_\mu\Phi = (\partial_\mu - igA_\mu)\,\Phi$. The particle spectrum now consists of a massive Higgs field with mass $m_s = \lambda^{1/2}\eta$, as before, plus a massive gauge boson with mass $m_v = g\eta$. As usual in the Higgs mechanism, the Goldstone boson ψ that appears in the model with global symmetry breaking has been turned into the longitudinal component of the massive gauge field.

This system again has static string-like solutions, which have the form

$$\Phi(\rho,\theta) = \eta f(\rho) e^{in\theta} \tag{22.170}$$

$$A_A = -\epsilon_{AB} x_B \frac{n}{g\rho^2}\alpha(\rho) \tag{22.171}$$

where $A, B = 1, 2$ correspond to the x and y directions, (ρ, θ) are again the polar coordinates in the (x, y) plane and $f(\rho)$ and $\alpha(\rho)$ are two functions. Now there are two mass scales in the problem, m_s and m_v, so $f(\rho)$ no longer depends on the single combination $m_s\rho$, as in the global case. Inserting this ansatz into the equations of motion one finds that at large distances the field Φ tends asymptotically to the value

$$\Phi \to \eta e^{in\theta} \tag{22.172}$$

as in the global case, while, at the same time,

$$A_\mu \to \frac{n}{g}\partial_\mu\theta \tag{22.173}$$

Observe that this configuration carries a quantized magnetic flux,

$$\int_S d\mathbf{S}\cdot\mathbf{B} = \int_C d\mathbf{l}\cdot\mathbf{A} = \frac{2\pi n}{g} \tag{22.174}$$

where C is a closed curve bounding the surface S. This configuration is called the Nielsen-Olesen vortex and is the prototype example of a local string, i.e. a string-like solution in a theory with a spontaneously broken local symmetry. As $\rho \to \infty$ we have $D_\mu\Phi \to 0$, since the gauge field contribution cancels the scalar field contribution. Thus, the term $\partial_\theta\Phi$, which in the global case goes to a constant, leading to a $1/\rho^2$ behavior in the energy density and to a logarithmic divergence in the string tension, is now replaced by $D_\theta\Phi$, which goes to zero as $\rho \to \infty$. Furthermore, also $F_{\mu\nu} \to 0$ as $\rho \to \infty$. Thus, the energy per unit length of a local string is finite and, on setting $n = 1$, is given by

$$\mu = 2\pi\eta^2 g(\beta) \tag{22.175}$$

where $\beta = (m_s/m_v)^2 = \lambda/g^2$ and $g(\beta)$ is a slowly varying function such that $g(1) = 1$.

Beside the fact that local strings have a finite string tension, while a single straight global string has a logarithmically divergent string tension, the other crucial difference between global and local strings is in the mechanisms by which they can radiate away energy. Both global and local strings can in principle emit gravitational waves. We will discuss in section 22.5.4 the GW emission by loops of cosmic strings, and we will see that the power radiated in GWs is $P_{\rm GW} = \Gamma G\mu^2$, where Γ is a numerical coefficient that depends on the loop shape, with typically $\Gamma \sim 50 - 100$. For a global string there is, however, another channel of massless particles available, i.e. it can emit the Goldstone particles associated with the broken global symmetry. In this case a computation similar to the one that we will discuss for the GW emission gives a power emitted in scalar radiation $P_S = \Gamma_S\eta^2$, where again $\Gamma_S = O(100)$. Apart from the logarithmic enhancement in eq. (22.168), which, as we have discussed, for a loop is actually cut off by the loop size, even for a global string we have $\mu \sim \eta^2$, so $G\mu^2 \sim (\eta/M_{\rm Pl})^2\,\eta^2$, and the

radiation in GWs is suppressed, compared with the radiation in Goldstone bosons, by a factor $(\eta/M_{\rm Pl})^2$, which even for a GUT theory is at least of order 10^{-8}. Thus, a global string loop mainly decays by emission of Goldstone bosons. This channel does not exist for local strings, which rather radiate GWs. In the following, we will therefore restrict ourselves to the study of local strings.

Finally, we observe that for quite some time cosmic strings have been considered as an alternative to inflation as a mechanism for producing the initial density perturbations from which cosmological structure formation evolved. However, as discussed in Section 19.2.1, they would seed isocurvature perturbations and, as the main mechanism for structure formation, they have been ruled out by cosmological observations, in particular those of the CMB. The spectrum of the C_l predicted by cosmic strings has only a single broad peak, covering values of l in the range

[34] See [Planck Collaboration], Ade et al. (2016a), "Planck 2015 results. XIII. Cosmological parameters".

$\sim 150 - 400$, rather than the set of much sharper acoustic peaks beautifully observed by CMB experiments. As we have already mentioned in Section 19.2.1, the present cosmological data are entirely consistent with adiabatic perturbations, and any admixture of CDM isocurvature perturbations in the initial conditions, if present at all, is below a few percent. As we will discuss later, the abundance of strings is fixed by their string tension. Writing

$$G\mu = A \times 10^{-7} \tag{22.176}$$

and requiring that cosmic strings do not spoil the agreement with CMB data, gives a bound on the string tension. The most recent analysis [34] gives (at 95% c.l.) $A < 2.4$ for strings described by the Abelian Higgs model, and $A < 1.3$ for strings modeled with the Nambu-Goto action, which we will describe in the next subsection.

16.3.2 22.5.2 Effective description and Nambu-Goto action

At a fundamental level the dynamics of Nielsen-Olesen vortices, i.e. their motion in flat or curved space as well as their interaction with other vortices, is determined by the dynamics of the underlying scalar and gauge fields. However, if we are only interested in the dynamics on scales much larger than the characteristic thickness of the vortex (which is determined by whichever is the largest between the Compton wavelength of the scalar, m_s^{-1}, and that of the massive vector field, m_v^{-1}), we can use an effective description in which we neglect the internal structure and regard the string as a one-dimensional object in space (defined by the line of zeros of the Higgs field), which spans a twodimensional surface in space-time.

The dynamics of such a string is a generalization of the dynamics of a relativistic point particle, which may be useful to recall here (see also the discussion in Section 20.2). In Newtonian mechanics a point particle moving in three-dimensional space is described by three functions $X^i(t)$ with $i = 1, 2, 3$. This description is not convenient in the relativistic case, since it breaks Lorentz invariance. Rather, it is convenient to introduce a variable τ that parametrizes the world-line swept by the particle, and describe the particle using four functions $X^\mu(\tau)$, with $\mu = 0, \ldots, 3$. Apparently, we have introduced one extra degree of freedom, $X^0(\tau)$. However, the way we decide to parametrize the world-line is arbitrary, so we also impose invariance under reparametrizations: if we transform $\tau \to \tilde{\tau}(\tau)$, the four functions $X^\mu(\tau)$ transform into four new functions $\tilde{X}^\mu(\tilde{\tau})$, such that $\tilde{X}^\mu(\tilde{\tau}) = X^\mu(\tau)$, and we require that the action governing $X^\mu(\tau)$ be invariant under such a transformation. In field-theoretical language, we can view the four functions $X^\mu(\tau)$ as four fields living in a $(0+1)$-dimensional space-time, scalar under reparametrizations of the world-line, i.e. under one-dimensional diffeomorphisms. With respect to the description in terms of $X^i(t)$ we have therefore added one more field, X^0, but we also

have one "gauge" symmetry, so the number of
degrees of freedom is the same. We could for instance chose the gauge $X^0(\tau) = \tau$, and we would be back to the description in terms of $X^i(t)$, with $t \equiv X^0(\tau) = \tau$. However, it is in general more useful to keep the redundant description in terms of $X^\mu(\tau)$, which explicitly preserves Lorentz invariance. The action of a point-like massive particle in a curved space with metric $g_{\mu\nu}$ is given by

$$S_{\rm pp} = -m \int ds \tag{22.177}$$

where $ds^2 = -g_{\mu\nu}(X) dX^\mu(\tau) dX^\nu(\tau)$. This action is explicitly invariant under reparametrizations, since ds is the invariant length. Observe that, for a time-like path, $ds^2 > 0$. Using $dX^\mu(\tau) = \dot{X}^\mu d\tau$, where the dot denotes the derivative with respect to τ, we can rewrite eq. (22.177) as

$$S_{\rm pp} = -m \int d\tau \left(-g_{\mu\nu} \dot{X}^\mu \dot{X}^\nu \right)^{1/2} \tag{22.178}$$

see also eq. (20.18). As we have already discussed [see eqs. (20.21)(20.23)], we can get rid of the square root by introducing an auxiliary field $e(\tau)$ and writing

$$S_{\rm pp} = \frac{1}{2} \int d\tau \left(e^{-1} g_{\mu\nu} \dot{X}^\mu \dot{X}^\nu - e m^2 \right) \tag{22.179}$$

Indeed, the variation with respect to $e(\tau)$ gives

$$e(\tau) = \frac{1}{m} \left(-g_{\mu\nu} \dot{X}^\mu \dot{X}^\nu \right)^{1/2} \tag{22.180}$$

which, inserted into eq. (22.179), gives back eq. (22.178). The action (22.179) is invariant under the reparametrization $\tau \to \tilde{\tau}(\tau)$, with $X^\mu(\tau) \to \tilde{X}^\mu(\tilde{\tau})$ and $e(\tau) \to \tilde{e}(\tilde{\tau})$ such that

$$\begin{aligned} \tilde{X}^\mu(\tilde{\tau}) &= X^\mu(\tau) \\ \tilde{e}(\tilde{\tau}) d\tilde{\tau} &= e(\tau) d\tau \end{aligned} \tag{22.181}$$

Observe also that eq. (22.179) has a well-defined massless limit, which is not the case for the action (22.177).

The generalization to the string dynamics is obtained by parametrizing the world-sheet of the string by a time-like parameter τ (analogous to the parameter τ for the world-line of a point-like particle) and a spacelike parameter σ. The position of the string in space-time (whose coordinates we denote by x^μ) is then described by the four functions $X^\mu(\tau, \sigma)$, so that the position in space-time of the point of the world-sheet labeled by (τ, σ) is given by $x^\mu = X^\mu(\tau, \sigma)$. For cosmic strings we will consider a four-dimensional space-time, while the theory of a fundamental bosonic string requires for consistency $D = 26$ space-time dimensions and superstrings live in 10 dimensions (or, non-perturbatively, in 11 dimensions). The generalization of the point-particle action (22.178) is given by the Nambu - Goto action

$$S_{\rm NG} = -\mu \int d^2\zeta \sqrt{-h} \tag{22.182}$$

[35] See Section 1.2 of Polchinski (1998), which we are following here. where $\zeta^a = (\tau, \sigma)$, the indices a, b take the values 1 and 2 corresponding to τ and σ, and μ is the string tension. We have also introduced $h = \det h_{ab}$, where

$$h_{ab} = g_{\mu\nu}(X) \partial_a X^\mu \partial_b X^\nu \tag{22.183}$$

is the induced metric, i.e. the metric induced on the world-sheet by its embedding into a space-time with metric $g_{\mu\nu}$, through the functions $X^\mu(\tau,\sigma)$. Under diffeomorphisms of the world-sheet, $\zeta \to \tilde{\zeta}(\zeta)$, the four fields X^μ transform as scalars,

$$X^\mu(\zeta) \to \tilde{X}^\mu(\tilde{\zeta}) = X^\mu(\zeta) \tag{22.184}$$

Then h_{ab} transforms as a metric tensor on the world-sheet, and the Nambu-Goto action is invariant. Just as in the point-particle case, we can get rid of the square root by introducing an auxiliary field, which in this case is a tensor γ_{ab}, with Lorentzian signature $(-,+)$. The action (22.182) is in fact equivalent to the Polyakov action

$$S_P = -\frac{\mu}{2}\int d^2\zeta\sqrt{-\gamma}\gamma^{ab}g_{\mu\nu}(X)\partial_a X^\mu \partial_b X^\nu \tag{22.185}$$

where $\gamma = \det\gamma_{ab}$. This can be shown by taking the variation with respect to γ_{ab}, which gives [35] $\gamma_{ab} = \kappa h_{ab}$, where κ is an arbitrary constant. Inserting this back into eq. (22.185), κ cancels between $\sqrt{-\gamma}$ and γ^{ab}, while $h^{ab}g_{\mu\nu}\partial_a X^\mu \partial_b X^\nu = h^{ab}h_{ab} = 2$, and we get back eq. (22.182). In this formulation, the theory is invariant under two-dimensional diffeomorphisms, under which $X^\mu(\zeta)$ transforms as in eq. (22.184), while γ_{ab} transforms as a metric,

$$\gamma_{ab}(\zeta) \to \tilde{\gamma}_{ab}(\tilde{\zeta}) = \gamma_{cd}(\zeta)\frac{\partial\zeta^c}{\partial\tilde{\zeta}^a}\frac{\partial\zeta^d}{\partial\tilde{\zeta}^b} \tag{22.186}$$

In the Polyakov formulation we further have an invariance under Weyl transformation, in which the world-sheet coordinates ζ^a, as well as the fields X^μ, are not changed, while

$$\gamma_{ab} \to e^{2\omega(\zeta)}\gamma_{ab} \tag{22.187}$$

The Nambu-Goto action can indeed be obtained from the NielsenOlesen vortex solution, as follows. The solution (22.170), (22.171) was obtained considering a static string along the z axis. Consider now a more general string solution, in general neither static nor a straight line. The location of the string, identified by the line of points where the Higgs field vanishes, defines a world-sheet in space-time $X^\mu(\zeta)$, so that a point labeled by ζ^a on the world-sheet has space-time coordinates $x^\mu = X^\mu(\zeta)$. At each point on the world-sheet we can define two space-like vectors $n^\mu_A(\zeta), A = 1,2$, which are orthogonal to the world-sheet, i.e.

$$g_{\mu\nu}n^\mu_A(\zeta)\partial_a X^\nu(\zeta) = 0 \tag{22.188}$$

for $a = 1,2$. We also normalize them so that

$$g_{\mu\nu}n^\mu_A n^\nu_B = \delta_{AB} \tag{22.189}$$

Then, at least infinitesimally close to the world-sheet, a generic spacetime point x^μ can be written as

$$x^\mu(\xi) \equiv x^\mu\left(\zeta^a,\rho^A\right) = X^\mu(\zeta) + \rho^A n^\mu_A(\zeta) \tag{22.190}$$

and we can therefore trade the four coordinates x^μ for the four coordinates $\xi \equiv \left(\zeta^a,\rho^A\right)$.[36]

Sufficiently close to the world-sheet (i.e. at spatial distances much smaller than its typical radius of curvature) we can still approximate the solution as a straight line in space, so the Nielsen-Olesen vortex solution at the space-time point $x^\mu(\xi)$ is still given by eqs. (22.170) and (22.171), in which now the transverse plane is the one spanned by the (n^μ_1, n^μ_2) vectors, and the role of the (x,y) coordinates is taken by (ρ_1,ρ_2). Thus, sufficiently close to the world-sheet, Φ is still given by (22.170), where now (ρ,θ) are the polar coordinates corresponding to the Cartesian coordinates (ρ_1,ρ_2), i.e.

$$\Phi[x(\xi)] = \eta f(\rho) e^{in\theta} \tag{22.191}$$

while the gauge field can be written as

$$A^\mu[x(\xi)] = -n^\mu_A(\zeta)\epsilon_{AB}\rho_B \frac{n}{g\rho^2}\alpha(\rho) \tag{22.192}$$

We can now reduce the action of the Abelian Higgs model,

$$S = -\int d^4x\sqrt{-g}\left[|D_\mu\Phi|^2 + V(\Phi) + \frac{1}{4}F_{\mu\nu}F^{\mu\nu}\right] \tag{22.193}$$

to an effective action for the variables $X^\mu(\zeta)$, observing that

$$d^4x\sqrt{-g} = d^4\xi\sqrt{-M} \tag{22.194}$$

where $d^4\xi = d^2\zeta d^2\rho$ and $M = \det M_{\alpha\beta}$, with

$$M_{\alpha\beta} = g_{\mu\nu}\frac{\partial x^\mu}{\partial\xi^\alpha}\frac{\partial x^\nu}{\partial\xi^\beta} \tag{22.195}$$

and α runs over the four values $a = 1, 2$ and $A = 1, 2$. Sufficiently close to the string world-sheet we can use eq. (22.190). Then, for the matrix elements where $\alpha = a$ and $\beta = b$ we get

$$\begin{aligned} M_{ab} &\simeq g_{\mu\nu}\left(\partial_a X^\mu + \rho^A\partial_a n^\mu_A\right)\left(\partial_b X^\nu + \rho^B\partial_b n^\nu_B\right) \\ &= h_{ab} + O(\rho/R) \end{aligned} \tag{22.196}$$

where R is the typical radius of curvature of the string. Similarly,

$$M_{AB} \simeq g_{\mu\nu}n^\mu_A n^\nu_B = \delta_{AB} \tag{22.197}$$

$$M_{Aa} \simeq g_{\mu\nu}n^\mu_A\left(\partial_a X^\nu + \rho^A\partial_a n^\nu_A\right) = O(\rho/R) \tag{22.198}$$

where we have used eqs. (22.188) and (22.189). Therefore, to lowest order in ρ/R, $\det M_{\alpha\beta} \simeq \det h_{ab}$, and the action (22.193) becomes

$$S \simeq -\int d^2\zeta\sqrt{-h}\int d^2\rho\left[|D_\mu\Phi|^2 + V(\Phi) + \frac{1}{4}F_{\mu\nu}F^{\mu\nu}\right](\zeta,\rho) \tag{22.199}$$

[12536] Actually, in a curved space-time with metric $g_{\mu\nu}$, to make the construction generally covariant we must pass to a system of Riemann normal coordinates, as discussed in Section 1.3.2 of Vol. 1, based on the geodesics leaving from the point $X^\mu(\zeta)$ with a given tangent vector n^μ. However, in a sufficiently small region close to the worldsheet, the geodesics in the direction n^μ_A parametrized by a proper-time parameter ρ_A becomes the same as the straight line $\rho^A n^\mu_A$. Furthermore, in the cosmological context in which we are eventually interested, the metric $g_{\mu\nu}$ only changes on a cosmological scale, while the string thickness is a microscopic quantity.

[125]

[37] Observe that the derivation that we have given does not go through for global strings, as signaled by the fact that their energy per unit length diverges. The reason is that in this case the theory also has a Goldstone particle, which, being massless, cannot be integrated out, but must be kept in the low-energy effective theory. The corresponding effective string action in this case is the Kalb-Ramond action, which besides the Nambu-Goto term also involves an antisymmetric tensor field, whose dual is related to the Goldstone boson; see Section 4.4.2 of Vilenkin and Shellard (2000).

The expression in square brackets is evaluated over the fields (22.191) and (22.192). On this static configuration, this expression coincides with the component $-T^0_0$ of the energy-momentum tensor derived from the Lagrangian (22.169), and its integral over $d^2\rho$ is therefore the energy per unit length of the string, i.e. the string tension μ. We have therefore recovered the Nambu-Goto action (22.182). Keeping also higher-order terms in ρ/R gives a further contribution, which depends for instance on the extrinsic curvatures $K^A_{ab} = -g_{\mu\nu}\partial_a n^\mu_A \partial_b X^\nu$ associated with the directions n^μ_A. These terms, in the language of effective actions, are higher-derivative operators, and are suppressed at scales $r \ll R$ with respect to the Nambu-Goto term. Their presence, however, is a signal of the fact that the string description is here only an effective one. A fundamental (bosonic) string, or the bosonic sector of superstring theory, taken as a fundamental theory, has only the Nambu-Goto term. [37]

16.3.3 22.5.3 String dynamics. Cusps and kinks

The use of the Nambu-Goto action allows us to study the dynamics of a local string much more easily, compared with what could be done by studying the dynamics of the vortex solution directly in the Abelian Higgs model. Furthermore, this effective action approach makes it clear that the dynamics on scales much larger than the thickness of the string is independent of the details of the underlying fundamental theory. We consider for simplicity the motion of strings in flat space-time, $g_{\mu\nu} = \eta_{\mu\nu}$, and we use the Polyakov action (22.185). As we have already mentioned, the variation with respect to γ_{ab} gives $\gamma_{ab} = \kappa h_{ab}$, where κ is an arbitrary constant, and h_{ab} is defined in eq. (22.183). Using the Weyl symmetry (22.187) we can always set $\kappa = 1$. Setting $\gamma_{ab} = h_{ab}$, the equation of motion derived by performing the variation of the action (22.185) with respect to X^μ is

$$\partial_a \left(\sqrt{-h} h^{ab} \partial_b X^\mu \right) = 0 \tag{22.200}$$

This is the same as $\Box X^\mu = 0$, where $\Box = (-h)^{-1/2}\partial_a \left(\sqrt{-h}h^{ab}\partial_b\right)$ is the two-dimensional d'Alembertian operator with respect to the metric h^{ab}, and acting on fields, such as X^μ, that are scalars under diffeomorphisms of the world-sheet. Of course, one obtains the same equation of motion by directly taking the variation with respect to X^μ of the Nambu-Goto action (22.182), using $\delta h = h h^{ab}\delta h_{ab}$.

In general, eq. (22.200) has a highly non-linear dependence on X^μ, since h_{ab} is itself a functional of X^μ given by eq. (22.183). However, we can now make use of the two-dimensional diffeomorphism invariance to fix the gauge to simplify this equation. The two-dimensional diffeomorphisms allow us to fix two conditions. It is convenient to require

$$h_{01} = 0 \tag{22.201}$$

and

$$h_{00} + h_{11} = 0 \tag{22.202}$$

so that h_{ab} can be written in the form

$$h_{ab} = e^\phi \eta_{ab} \tag{22.203}$$

i.e. the world-sheet metric becomes conformally flat. [38] This gauge is therefore called the conformal gauge. The advantage of this gauge choice is that $-h = e^{2\phi}$ while $h^{ab} = e^{-\phi}\eta^{ab}$, so in the combination $\sqrt{-h}h^{ab}$ that appears in eq. (22.200) the conformal factor cancels, and we are left with a simple linear equation

$$\left(-\partial_\tau^2 + \partial_\sigma^2\right) X^\mu = 0 \tag{22.204}$$

which is just the Klein-Gordon wave equation in two-dimensional flat space. Furthermore, using eq. (22.183) with $g_{\mu\nu} = \eta_{\mu\nu}$, the gauge conditions (22.201) and (22.202) give, respectively,

$$\eta_{\mu\nu}\partial_\sigma X^\mu \partial_\tau X^\nu = 0 \tag{22.205}$$

$$\eta_{\mu\nu}\left(\partial_\sigma X^\mu \partial_\sigma X^\nu + \partial_\tau X^\mu \partial_\tau X^\nu\right) = 0 \tag{22.206}$$

These two equations do not involve second derivative with respect to τ, and are therefore constraints on the initial values of $\partial_\tau X^\mu$ and $\partial_\sigma X^\mu$, which are preserved by the time evolution.

The dynamics of the string can be further simplified as follows. [39] The conditions (22.201) and (22.202) do not remove the gauge freedom completely. Indeed, consider a diffeomorphism $\zeta^a \to \tilde{\zeta}^a(\zeta)$, i.e. $\tau \to \tilde{\tau}(\tau,\sigma)$ and $\sigma \to \tilde{\sigma}(\tau,\sigma)$. The metric h_{ab} transforms in the same way as γ_{ab} in eq. (22.186). Using eqs. (22.201) and (22.202) and requiring that $\hat{h}_{01} = 0$ gives

$$\partial_\tau\tilde{\tau}\partial_\sigma\tilde{\tau} = \partial_\tau\tilde{\sigma}\partial_\sigma\tilde{\sigma} \tag{22.207}$$

while requiring $\tilde{h}_{00} + \tilde{h}_{11} = 0$ gives

$$\left(\partial_\tau\tilde{\tau}\right)^2 + \left(\partial_\sigma\tilde{\tau}\right)^2 = \left(\partial_\tau\tilde{\sigma}\right)^2 + \left(\partial_\sigma\tilde{\sigma}\right)^2 \tag{22.208}$$

We see by inspection that a solution of eqs. (22.207) and (22.208) is obtained using functions $\tilde{\sigma}(\tau,\sigma)$ and $\tilde{\tau}(\tau,\sigma)$ that satisfy the two conditions

$$\partial_\tau\tilde{\sigma} = \partial_\sigma\tilde{\tau} \tag{22.209}$$

$$\partial_\sigma\tilde{\sigma} = \partial_\tau\tilde{\tau} \tag{22.210}$$

Applying ∂_σ to the first equation, and using the second one, we also get

$$\left(-\partial_\tau^2 + \partial_\sigma^2\right)\tilde{\tau} = 0 \tag{22.211}$$

and similarly $\left(-\partial_\tau^2 + \partial_\sigma^2\right)\tilde{\sigma} = 0$. Having at our disposal a function $\tilde{\tau}$ that satisfies eq. (22.211) but is otherwise arbitrary, we can use it to fix the gauge $X^0(\tau,\sigma) = \tau$, since X^0 also satisfies eq. (22.204). Once $\tilde{\tau}(\tau,\sigma)$ has been fixed in this way, the derivatives of the function $\tilde{\sigma}$ are also fixed, so the gauge fixing is now complete, apart from the irrelevant possibility of shifting σ by a constant. In this gauge, therefore, X^0 drops from the set of dynamical variables and, using the notation $X^0 \equiv t$, we [38] Locally, i.e. in an open neighborhood of a point on the world-sheet, it is always possible to impose these conditions. The same is true globally, if the world-sheet has the topology of a plane or, for closed string loops, of a cylinder. These are the cases of relevance for the effective string that we are discussing here.

[39] We follow Section 6.2 of Vilenkin and Shellard (2000), to which we refer for more details. have $\tau = t$. Thus, the string is now described by a vector $\mathbf{X}(t,\sigma)$, and eqs. (22.204) $-$(22.206) become

$$\ddot{\mathbf{X}} - \mathbf{X}'' = 0 \tag{22.212}$$

$$\dot{\mathbf{X}} \cdot \mathbf{X}' = 0 \tag{22.213}$$

$$\dot{\mathbf{X}}^2 + \mathbf{X}'^2 = 1 \tag{22.214}$$

where the dot denotes the derivative with respect to t and the prime that with respect to σ. The solution of eq. (22.212) is a superposition of generic left-moving and right-moving waves,

$$\mathbf{X}(t,\sigma) = \frac{1}{2}[\mathbf{a}(\sigma - t) + \mathbf{b}(\sigma + t)] \tag{22.215}$$

with $\mathbf{a}$ and $\mathbf{b}$ arbitrary vector functions of one variable. The constraints (22.213) and (22.214) then give

$$\mathbf{a}'^2 = \mathbf{b}'^2 = 1 \tag{22.216}$$

The energy-momentum tensor of the string is obtained from eq. (22.185),

$$\begin{aligned} T^{\mu\nu}(x) &= \frac{2}{\sqrt{-g}} \frac{\delta S}{\delta g_{\mu\nu}(x)} \\ &= -\mu \int d^2\zeta \sqrt{-h} h^{ab} \frac{1}{\sqrt{-g}} \delta^{(4)}(x - X) \partial_a X^\mu \partial_b X^\nu \end{aligned} \tag{22.217}$$

where, after taking the variation with respect to $g_{\mu\nu}$, we have used the equation of motion $\gamma^{ab} = h^{ab}$. In flat space, and using the conformal gauge (22.203), we get

$$T^{\mu\nu}(x) = \mu \int d^2\zeta \delta^{(4)}(x - X) \left[\partial_\tau X^\mu \partial_\tau X^\nu - \partial_\sigma X^\mu \partial_\sigma X^\nu\right] \tag{22.218}$$

If we fix the residual gauge freedom by further imposing $X^0(\tau, \sigma) = \tau$, we finally get for the energy density $\rho(x) = T^{00}$,

$$\begin{aligned} \rho(t, \mathbf{x}) &= \mu \int d\sigma d\tau \delta^{(3)}[\mathbf{x} - \mathbf{X}(\tau, \sigma)] \delta(t - \tau) \\ &= \mu \int d\sigma \delta^{(3)}[\mathbf{x} - \mathbf{X}(t, \sigma)] \end{aligned} \tag{22.219}$$

The total energy is therefore

$$\begin{aligned} E &= \int d^3x \rho(x) \\ &= \mu \int d\sigma \end{aligned} \tag{22.220}$$

Thus, in this gauge σ provides a measure of length on the string, given by $d\sigma = dE/\mu$. Similarly, for the momentum we get

$$\begin{aligned} P^i &= \int d^3x T^{0i} \\ &= \mu \int d\sigma \dot{X}^i \end{aligned} \tag{22.221}$$

Consider now a closed loop. According to eq. (22.220), the parameter σ can be taken to be in the range $0 \leqslant \sigma \leqslant L$, where $L = E/\mu$ is called the invariant length of the loop. For a closed loop we require periodic boundary conditions $\mathbf{X}(t, \sigma + L) = \mathbf{X}(t, \sigma)$. In terms of the functions a and $\mathbf{b}$ introduced in eq. (22.215), this implies

$$\begin{aligned} -\mathbf{a}(\sigma + L - t) + \mathbf{a}(\sigma - t) &= \mathbf{b}(\sigma + L + t) - \mathbf{b}(\sigma + t) \\ &\equiv \boldsymbol{\Delta} \end{aligned} \tag{22.222}$$

and, because of invariance under translations in the world-sheet, the vector $\mathbf{\Delta}$ is independent of σ. From eq. (22.221), the momentum $\mathbf{P}$ is given by

$$\begin{aligned}\mathbf{P} &= \frac{\mu}{2}\int_0^L d\sigma\,[\partial_t\mathbf{a}(\sigma-t)+\partial_t\mathbf{b}(\sigma+t)]\\ &= \frac{\mu}{2}\int_0^L d\sigma\,[-\partial_\sigma\mathbf{a}(\sigma-t)+\partial_\sigma\mathbf{b}(\sigma+t)]\\ &= \frac{\mu}{2}[-\mathbf{a}(L-t)+\mathbf{a}(-t)+\mathbf{b}(L+t)-\mathbf{b}(t)]\\ &= \mu\mathbf{\Delta}\end{aligned} \tag{22.223}$$

In the center-of-mass frame of the string, which is defined by $\mathbf{P}=0$, we therefore have $\mathbf{\Delta}=0$, and eq. (22.222) shows that $\mathbf{a}$ and $\mathbf{b}$ are separately periodic,

$$\mathbf{a}(\sigma+L-t)=\mathbf{a}(\sigma-t) \tag{22.224}$$
$$\mathbf{b}(\sigma+L+t)=\mathbf{b}(\sigma+t) \tag{22.225}$$

With this formalism it is easy to understand the emergence of cusps in the generic motion of a string. Indeed, from eq. (22.215) we have

$$(\partial_t\mathbf{X})^2=\frac{1}{4}\left(-\mathbf{a}'+\mathbf{b}'\right)^2 \tag{22.226}$$

Thus, using eq. (22.216), we see that at a point of the world-sheet where $\mathbf{a}'=-\mathbf{b}'$, the world-sheet moves at the speed of light. Equation (22.216) shows that, as σ runs from zero to L, the functions $\mathbf{a}'(\sigma-t)$ and $\mathbf{b}'(\sigma+t)$ sweep curves on the unit sphere. Because of eqs. (22.224) and (22.225), these are actually closed curves on the sphere. The periodic boundary conditions on these vector functions also imply that

$$\int_0^L d\sigma\mathbf{a}'=\int_0^L d\sigma\mathbf{b}'=0 \tag{22.227}$$

so these curves cannot be confined to a single hemisphere. This means that, as σ runs from zero to L, the curves swept by $\mathbf{a}'(\sigma-t)$ and by $-\mathbf{b}'(\sigma+t)$ will generically intersect, and therefore in a typical motion of the string there will be points on the world-sheet that travel at the speed of light. Since this happens just at the intersection point, the corresponding point of the string travels at the speed of light only at

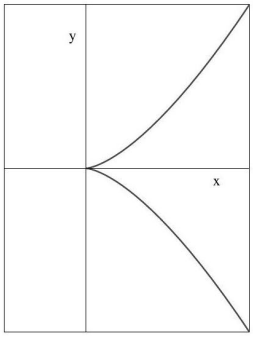


Fig. 22.16 The cusp developed by a cosmic string. some instant of time. When this happens, the spatial configuration of the string takes the form of a cusp. Indeed, let us choose the origin of the world-sheet coordinates so that the point of the world-sheet travelling at the speed of light corresponds to $\tau = \sigma = 0$. We also choose the origin of the spatial coordinates so that, at $\tau = 0$, this point is at $\mathbf{x} = 0$, and $\mathbf{a} = \mathbf{b} = 0$, and we again use the gauge $t = \tau$. Expanding $\mathbf{a}(\sigma_-)$ and $\mathbf{b}(\sigma_+)$in powers of $\sigma_\pm \equiv \sigma \pm t$ near the point $t = \sigma = 0$, i.e. $\sigma_+ = \sigma_- = 0$, we get

$$\mathbf{a}(\sigma_-) = \mathbf{a}_0'\sigma_- + \frac{1}{2}\mathbf{a}_0''\sigma_-^2 + \frac{1}{6}\mathbf{a}_0'''\sigma_-^3 + \dots \tag{22.228}$$

$$\mathbf{b}(\sigma_+) = \mathbf{b}_0'\sigma_+ + \frac{1}{2}\mathbf{b}_0''\sigma_+^2 + \frac{1}{6}\mathbf{b}_0'''\sigma_+^3 + \dots \tag{22.229}$$

so that

$$\mathbf{a}'(\sigma_-) = \mathbf{a}_0' + \mathbf{a}_0''\sigma_- + \dots \tag{22.230}$$

$$\mathbf{b}'(\sigma_+) = \mathbf{b}_0' + \mathbf{b}_0''\sigma_+ + \dots \tag{22.231}$$

The condition that the curves spanned by $\mathbf{a}'(\sigma - t)$ and $-\mathbf{b}'(\sigma + t)$ intersect at $\sigma = t = 0$ gives $\mathbf{a}_0' = -\mathbf{b}_0'$. Thus, from eq. (22.215), at $t = 0$ we have

$$\mathbf{X}(0,\sigma) = \frac{1}{4}\left(\mathbf{a}_0'' + \mathbf{b}_0''\right)\sigma^2 + \frac{1}{6}\left(\mathbf{a}_0''' + \mathbf{b}_0'''\right)\sigma^3 + \dots \tag{22.232}$$

If $(\mathbf{a}_0'' + \mathbf{b}_0'') \neq 0$, we can chose the axes so that $(\mathbf{a}_0'' + \mathbf{b}_0'')$ points, say, along the x axis. Then, in the (x, y) plane, we have

$$X_x(0,\sigma) \simeq c_x\sigma^2 \tag{22.233}$$

$$X_y(0,\sigma) \simeq c_y\sigma^3 \tag{22.234}$$

where $c_x = (1/4)\,|\mathbf{a}_0'' + \mathbf{b}_0''|$ and $c_y = (1/6)\left(\mathbf{a}_0''' + \mathbf{b}_0'''\right)_y$ (assuming that $c_y \neq 0$). This is the parametric description of a cusp, as shown in Fig. 22.16 and, near $\sigma = 0$, the shape of the string is given by $X_x^3(0,\sigma) \propto X_y^2(0,\sigma)$. The velocity of the string at this point is orthogonal to the direction of the cusp. This can be shown by inserting eqs. (22.230) and (22.231) into eq. (22.216) and requiring that it be satisfied order by order in σ, which gives

$$|\mathbf{a}_0'| = |\mathbf{b}_0'| = 1 \tag{22.235}$$

$$\mathbf{a}_0' \cdot \mathbf{a}_0'' = \mathbf{b}_0' \cdot \mathbf{b}_0'' = 0 \tag{22.236}$$

Thus the direction of the cusp, which is given by $(\mathbf{a}_0'' + \mathbf{b}_0'')$, is orthogonal to the direction of the velocity, which is given by $(1/2)\left(-\mathbf{a}_0' + \mathbf{b}_0'\right) = \mathbf{b}_0'$, since $\mathbf{a}_0' = -\mathbf{b}_0'$.

Of course, the cusp singularity is an artifact of the zero-thickness approximation implicit in the use of the Nambu-Goto action. For a Nielsen-Olesen vortex the singularity will be smoothed by its inner structure, on the scales m_s^{-1} or m_v^{-1}, whichever is larger, and the velocity at the cusp will be strictly smaller than the speed of light. However, this
smoothing occurs on scales that are much smaller than the horizon-scale length typical of the cosmic strings in which we are interested, and the idealized cusp obtained in the Nambu-Goto approximation is an exceedingly good description of the dynamics. Clearly, the presence of such a singularity where formally the string reaches the speed of light is very interesting for GW production, as we will discuss further in Section 22.5.4.

Another significant structure that can develop on cosmic strings is a kink, i.e. a point where the functions $\mathbf{a}'$ and $\mathbf{b}'$ are discontinuous, so that there is a discontinuity in the vector $\mathbf{X}'$ tangent to the string. While cusps, as we have seen, are a rather generic product of the evolution of a single string, kinks are formed when two strings interact with each other, exchanging parts as in Fig. 22.17, or when a single string self-intersects, as in Fig. 22.18. In the latter case there is both formation of a kink and emission of a smaller string loop. Numerical simulations for the Abelian Higgs model show that such processes, also referred as reconnections, are quite frequent. A kink implies that both $\mathbf{a}'$ and $\mathbf{b}'$ are discontinuous. Since $\mathbf{a}$ is a function of $\sigma - t$ and $\mathbf{b}$ is a function of $\sigma + t$, after the kink has been generated it splits into two kinks, moving in opposite directions along the string, at the speed of light. The argument that we have presented regarding the generic intersection of the curves spanned by $\mathbf{a}'$ and $\mathbf{b}'$ on the unit sphere indicates that cusps are generically expected to appear on smooth strings a few times per oscillation period. However, if a string develops kinks, the corresponding curve on the unit sphere has discontinuities, and in this case there are more chances that the two curves do not intersect. In that case the string has kinks, traveling on it at the speed of light, but does not develop cusps.

16.3.4 22.5.4 Gravitational radiation from cosmic strings

In the Standard Model these is no spontaneously broken local $U(1)$ symmetry that could give rise to cosmic strings. [40] However, cosmic strings could emerge in extensions of the Standard Model based on larger grandunified groups. If cosmic strings exist, they would be very remarkable objects. For $G\mu \simeq 10^{-7}$, which is about the largest value consistent with the CMB [see eq. (22.176)], we have $\mu^{1/2} \simeq 4 \times 10^{15}$GeV, in our units $\hbar = c = 1$. Reinstating c, a mass per unit length μ such that $G\mu/c^2 = 10^{-7}$ has the huge numerical value $\mu \simeq 10^{21}$ g/cm. A cosmic string loop with this tension and a length equal to one astronomical unit (and thus much smaller than the cosmological scale that most cosmic string loops would be expected to have, as we shall see) would already have a mass $\sim 10 M_\odot$. This energy is concentrated in a tube of width $\sim m_s^{-1}$ or $\sim m_v^{-1}$. As we saw in Section 22.5.1, $m_s = \lambda^{1/2}\eta$ and $m_v = g\eta$, so their value is basically determined by the Higgs vacuum expectation value η. For a phase transition at a GUT scale, and any reasonable values of the couplings λ and g, the thickness of the string is much smaller than, say, the proton radius. Thus, we have a huge energy per unit length, concentrated in an extremely narrow tube. Because of their huge tension, these objects oscillate relativistically and we have

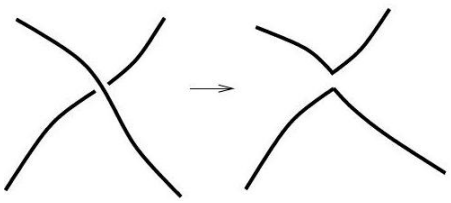


Fig. 22.17 The kinks generated by the reconnection of two cosmic strings.

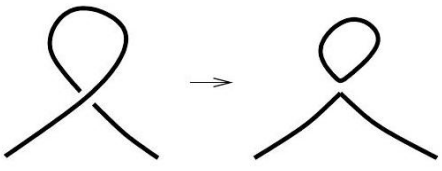


Fig. 22.18 The kink and the closed loop generated by the self intersection of a cosmic string.

even seen that, during their evolutions, cusps are generically formed, where a point of the string instantaneously moves at the speed of light. Cosmic strings are therefore potentially very interesting sources of GWs.

Stochastic GW background from a cosmic string network

To study GW generation by a network of cosmic strings, we must address two questions. First, what is the radiation generated by a cosmic string loop and, second, what are the statistical properties of a cosmic string network. Let us begin with the radiation from a single loop. A first naive estimate can be obtained using the quadrupole formula (3.75). The mass of a string loop of length L is $M \sim \mu L$, so its quadrupole moment is $Q \sim \mu L^3$. The typical oscillation frequency of a loop of length L oscillating relativistically is $\omega \sim 1/L$, so $\dddot{Q} \sim Q\omega^3 \sim \mu$, and eq. (3.75) gives $P_{\rm quad} \sim G\mu^2$. However, the use of the quadrupole formula is obviously inappropriate, given that the loop moves relativistically. One must rather use the formalism discussed in Section 3.1 of Vol. 1, valid for (weak-field) sources with arbitrary velocity. An oscillating string loop of length L emits at a discrete set of frequencies ω_n. Observe that, for a closed loop of length L in its rest frame, the functions $\mathbf{a}$ and $\mathbf{b}$ are periodic with respect to σ period L; see eqs. (22.224) and (22.225). From eq. (22.215) this means that

$$\mathbf{X}\left(\sigma + \frac{L}{2}, t + \frac{L}{2}\right) = \mathbf{X}(\sigma, t) \tag{22.237}$$

so the string configuration is periodic in time with period $L/2$, modulo a shift in the parameter σ. Thus, the oscillation frequencies of a cosmic string loop of length L are $\omega_n = 2\pi f_n$, with $f_n = n/(L/2) = 2n/L$. Adapting to this case eqs. (3.16)-(3.20), which were written for a simply periodic source, we see that the energy radiated per unit solid angle is still given by eq. (3.16), where the Fourier transform of the energymomentum tensor of the source is

$$\tilde{T}_{ij}(\omega, \mathbf{k}) = \sum_n \theta_{ij}\left(\omega_n, \mathbf{k}\right) 2\pi\delta\left(\omega - \omega_n\right) \tag{22.238}$$

and the radiated power is

$$P = \sum_n P_n \tag{22.239}$$

where (in our present units $c = 1$)

$$\frac{dP_n}{d\Omega} = \frac{G\omega_n^2}{\pi}\Lambda_{ij,kl}(\hat{\mathbf{n}})\theta_{ij}\left(\omega_n, \omega_n\hat{\mathbf{n}}\right)\theta^*_{kl}\left(\omega_n, \omega_n\hat{\mathbf{n}}\right) \tag{22.240}$$

One can now consider some explicit exact solutions for the motion of the loop, and compute numerically the power radiated in the various harmonics, and hence the total power. The result is of the form

$$P = \Gamma G\mu^2 \tag{22.241}$$

where the proportionality to $G\mu^2$ is the same as in the quadrupole formula, and is fixed by the facts that eq. (22.240) is proportional to G and that each factor $\theta_{ij}\left(\omega_n, \omega_n\hat{\mathbf{n}}\right)$ is proportional to μ. The numerical factor Γ depends on the shape of the loop, and turns out to be a large number, $\Gamma \sim 50 - 100$, because many harmonics contribute. A value of Γ of this order is obtained for loops with cusps, for loops with kinks, and also for loops with neither cusps nor kinks. However, the dependence of P_n on n is different in these cases. In particular, cusps or

[12540] The spontaneous breaking of the $SU(2) \times U(1)$ gauge symmetry of the Standard Model does not give rise to stable string configurations; see Section 4.2.6 of Vilenkin and Shellard (2000).

kinks correspond to very small-scale structures on the string, and contribute to harmonics with very large n. Thus, for loops with cusps or with kinks, P_n decreases quite slowly with n. For loops with cusps, at large n, one has

$$P_n^{\rm cusps} \sim G\mu^2 n^{-4/3} \tag{22.242}$$

while for loops with kinks

$$P_n^{\rm kinks} \sim G\mu^2 n^{-5/3} \tag{22.243}$$

The energy radiated in GWs is taken from the energy $E = \mu L$ of the loop. Thus, as the loop loses energy, it must shrink and eventually disappear. From eq. (22.241) we get we have

$$\mu \frac{dL}{dt} = -\Gamma G\mu^2 \tag{22.244}$$

Since Γ, to a first approximation, is independent of the loop size, a loop created initially with a length L_0 at time t_0 evolves as

$$L(t) = L_0 - \Gamma G\mu\,(t - t_0) \tag{22.245}$$

Thus, in the absence of interaction with other strings, the loop lifetime is due to GW emission and is given by $T = L_0/(\Gamma G\mu)$. For $G\mu \sim 10^{-7}$ and $\Gamma \sim 10^2$ we have $T \sim 10^5 L_0$. Therefore, the loop is long-lived compared with the natural length-scale given by its size, and performs $O\,(10^5)$ oscillations before disappearing.

The other crucial aspect for computing the GW production from an ensemble of cosmic strings is the distribution of loop sizes. Several analytical and numerical investigations have shown that a cosmic string network evolves toward an attractor solution characterized by a scaling regime, in which the average distance between strings remains a fixed fraction of the Hubble length $H^{-1}(t)$, i.e. it scales linearly with cosmic time t, and the energy density of the network remains a constant fraction of the total energy density of the Universe. If cosmic strings were absolutely stable, the energy of a single cosmic string $E = \mu L$ would scale with the expansion as $a(t)$, since $L \propto a(t)$. A physical volume V scales as a^3, so the energy density of cosmic strings would scale only as $1/a^2$, i.e. much more slowly than radiation, so during RD the ratio of the energy density in cosmic strings to the energy density in radiation would increase as a^2. If this were the case, cosmic strings would quickly reach an unacceptably large density and would be dangerous relics to

[41] See Blanco-Pillado, Olum and Shlaer (2014).

be washed out by an inflationary epoch, just like monopoles or domain walls. However, cosmic strings lose energy by GW emission, and this saves them from becoming dangerous relics. What happens in a cosmic string network is that long strings cross each other and produce smaller loops, as in Fig. 22.18, which then shrink and eventually disappear by emitting GWs. Numerical simulations give, for the number density (per physical volume) of loops of length L, [41]

$$\frac{n(L,t)}{a^3(t)} \simeq \frac{0.18}{t^{3/2}(L + \Gamma G\mu t)^{5/2}} \tag{RD}$$

which is valid for L smaller than a cutoff value $L_{\rm max} \simeq 0.1t$, and

$$\frac{n(L,t)}{a^3(t)} \simeq \frac{0.27 - 0.45(L/t)^{0.31}}{t^2(L + \Gamma G\mu t)^2} \tag{MD}$$

with a cutoff $L_{\rm max} \simeq 0.18t$. This describes a population of loops with a broad range of values of L, from $L = 0$ up to values not much smaller than the horizon radius. Writing $L = \alpha t$, in this distribution there are loops up to $\alpha \sim 0.1$. It is still under debate whether there exists also

a separate population of small loops, with typical values $\alpha \sim \Gamma G\mu$, as was assumed in earlier works. The scaling property of these distributions is reflected in the fact that they have the form $n(L,t)/a^3(t) = t^{-4}N(L/t)$, for some function N.

Given the energy radiated by a loop and the loop distribution, the resulting GW background can be computed as follows. The power radiated by a single loop of length L can be written as in eq. (22.239). It is convenient to extract explicitly the factor $G\mu^2$, writing $P_n = G\mu^2 p_n$, so

$$\begin{aligned}\frac{dE_{\rm GW}}{dt} &= G\mu^2 \sum_n p_n \\ &= G\mu^2 L \int_0^\infty df_e S\left(Lf_e\right)\end{aligned} \tag{22.248}$$

where

$$S(x) \equiv \sum_n p_n \delta\left(x - x_n\right) \tag{22.249}$$

The subscript e in f_e is a reminder that this is the frequency at the time of emission, before any redshifting, $x = Lf_e, x_n = Lf_{e,n}$, and $f_{e,n} = 2n/L$ are the frequencies at which the loop of length L emits. Comparing with eq. (22.241) we have

$$\int_0^\infty dx S(x) = \Gamma \tag{22.250}$$

This continuous notation is useful because is allows us to write the frequency spectrum of the GWs emitted by a single loop, in the interval between the emission time t_e and $t_e + dt_e$, as

$$\frac{dE_{\rm GW}\left(t_e\right)}{df_e} = G\mu^2 LS\left(Lf_e\right) dt_e \tag{22.251}$$

The energy density of the radiation emitted by a network of loops with number density $n(L,t)dL$ in the interval between time t_e and $t_e + dt_e$ is therefore

$$\frac{d\rho_{\rm GW}\left(t_e\right)}{df_e} = G\mu^2 dt_e \int_0^{\alpha t_e} dL n\left(L,t_e\right) LS\left(Lf_e\right) \tag{22.252}$$

where we have taken into account the fact that, at time t_e, the maximum length of the loop is αt_e. We now observe that $\rho_{\rm GW}$ redshifts as $1/a^4$, so $\rho_{\rm GW}\left(t_e\right) = a^{-4}\left(t_e\right)\rho_{\rm GW}\left(t_0\right)$, where t_0 is the present value of cosmic time and we have set $a\left(t_0\right) = 1$. The emission frequency is related to the frequency $f_{\rm obs}$ observed today by $f_e = f_{\rm obs}/a\left(t_e\right)$, so

$$\frac{d\rho_{\rm GW}\left(t_e\right)}{df_e} = \frac{1}{a^3\left(t_e\right)} \frac{d\rho_{\rm GW}\left(t_0\right)}{df_{\rm obs}} \tag{22.253}$$

Integrating over the emission time t_e from the primordial epoch $t_e \simeq t_*$ when cosmic strings first formed to the present time t_0, we get

$$\left(\frac{d\rho_{\rm GW}}{df}\right)_0 = G\mu^2 \int_{t_*}^{t_0} dt a^3(t) \int_0^{\alpha t} dL n(L,t) LS\left(\frac{Lf}{a(t)}\right) \tag{22.254}$$

where we henceforth use simply f to denote $f_{\rm obs}$, and we have denoted the integration variable t_e simply by t. Therefore, using the definition (7.198), $\Omega_{\rm gw}(f) = \rho_c^{-1} f d\rho_{\rm GW}/df$, together with $\rho_c = 3H_0^2/(8\pi G)$, we get

$$h_0^2 \Omega_{\rm gw}(f) = \frac{8\pi}{3} \left(\frac{h_0}{H_0}\right)^2 (G\mu)^2 f \int_{t_*}^{t_0} dt a^3(t) \int_0^{\alpha t} dL n(L,t) L S\left(\frac{Lf}{a(t)}\right) \tag{22.255}$$

Given the function $S(x)$ that models the loop emission and the number density of loops $n(L,t)$, this integral can then be computed numerically. Currently the typical value of the power P_n in a realistic string network is not known. One can, however, study various limiting cases. In particular, as a model for the loop radiation we can use the radiation from loops with cusps, as given in eq. (22.242), or that for loops with kinks, as given in eq. (22.243). For the number density we can use the distribution for large loops given in eqs. (22.246) and (22.247), which is the case favored by the most recent numerical studies. Alternatively, one can study the contribution from a population of small loops, with $\alpha \sim \Gamma G\mu$, which was suggested by earlier analytic and numerical studies that assumed that the loop size is determined by the smallest lengthscale on which gravitational back-reaction washes out inhomogeneities on the string. Actually, more recent studies have suggested that such small-scale structures should be even smaller, with $\alpha \sim (\Gamma G\mu)^n$, with $n = 3/2$ in RD and $n = 5/2$ in MD. It is conventional to parametrize this uncertainty by introducing $\varepsilon \equiv \alpha/(\Gamma G\mu)$, so $\varepsilon = 1$ corresponds to the "old" scenario, while $\varepsilon \ll 1$ in the scenario with $\alpha \sim (\Gamma G\mu)^n$ with $n > 1$; for instance, if $\Gamma = 50, G\mu = 10^{-7}$ and $n = 5/2$ we get $\varepsilon \sim 10^{-8}$.

A final parameter that must be taken into account is the recombination probability p, which measures the probability that two strings

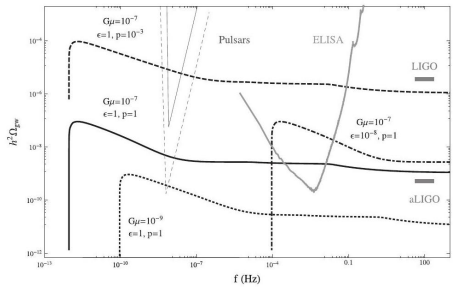


Fig. 22.19 $h_0^2 \Omega_{\rm gw}(f)$ for cosmic strings with cusps, with a small initial loop size $\varepsilon \equiv \alpha/(\Gamma G\mu)$, for different choices of $\varepsilon, G\mu$ and the reconnection probability p, assuming $\Gamma = 50$. From Binetruy, Bohe, Caprini and Dufaux (2012). (C)SISSA Medialab Srl. Reproduced by permission of IOP Publishing. All rights reserved.

crossing each other exchange their segments, as in Fig. 22.17. For the effective Nambu-Goto strings that describe Nielsen-Olesen vortices one basically has $p = 1$. However, it has also been suggested that fundamental superstrings can expand to cosmological scales. For fundamental strings in typical string models, where the string scale is close to the Planck scale, this is quite unlikely, since such strings would be unstable to break-up into smaller Planckian-size strings, and even if they were stable, they would anyhow be removed by inflation. However, in string models with a lower scale, based for example on large extra dimensions, fundamental strings (F-strings) or one-dimensional D-branes (D-strings) of cosmological size could exist, and could be produced after inflation. For such strings the reconnection probability p is suppressed by a factor of order g_s^2, where g_s is the string coupling. Furthermore, strings moving in a higher-dimensional space-time can avoid intersections more easily. As a result, one estimates $10^{-3} \lesssim p \lesssim 1$ for F-strings and $0.1 \lesssim p \lesssim 1$ for D-strings. In turn, this gives a larger number density of long strings, since a small reconnection probability means that processes such as

those in Fig. 22.18 are suppressed, and long strings are much more abundant. Indeed, $n(L,t)$ scales possibly as $1/p$ (although the exact dependence is still debated) and the values in eqs. (22.246) and (22.247) correspond to $p = 1$.

When computing the stochastic background of GWs one must also take into account that, depending on the detector sensitivity, some of the strongest bursts coming from cusps could be detected individually (a point that we will discuss in more detail later), and therefore must be subtracted from the contribution of the stochastic background.

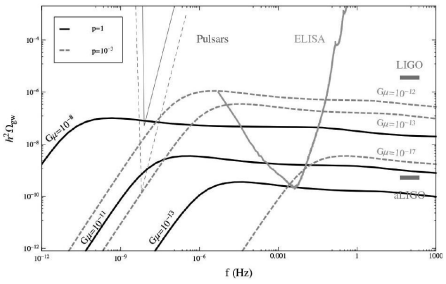


Fig. 22.20 $h_0^2\Omega_{\rm gw}(f)$ for cosmic strings with cusps, with a large initial loop size ($\alpha = 0.1$) and different choices of $G\mu$ and the reconnection probability p, assuming $\Gamma = 50$. From Binetruy, Bohe, Caprini and Dufaux (2012). (C)SISSA Medialab Srl. Reproduced by permission of IOP Publishing. All rights reserved.

Figure 22.19 shows $h_0^2\Omega_{\rm gw}(f)$ generated by strings with cusps, taking for $n(L,t)$ a distribution of small loops with a fixed value $\varepsilon \equiv \alpha/(\Gamma G\mu)$, for different choices of $\varepsilon, G\mu$, and the reconnection probability p, assuming $\Gamma = 50$. The theoretical prediction is compared with the limit set by the LIGO S5 run, with the design sensitivity of advanced LIGO, with the sensitivity expected for LISA and with pulsar timing experiments. Pulsar timing arrays (PTAs) will be discussed in Chapter 23, and provide very interesting limits in the frequency range between 10^{-9} and 10^{-7} Hz. The solid wedge-shaped line corresponds to the current sensitivities, and the dashed line to expected future sensitivities.

Figure 22.20 shows $h_0^2\Omega_{\rm gw}(f)$ generated again by strings with cusps, but taking for $n(L,t)$ a distribution of large loops with $\alpha = 0.1$ and various values of $G\mu$ and p. For large loop size and recombination parameter $p = 1$, pulsar timing arrays impose a very strong bound. The nine-year data set released in 2015 by the pulsar timing array NANOGrav provides a limit

$$G\mu < 3.3 \times 10^{-8} \tag{22.256}$$

at 95% c.l., while the European Pulsar Timing Array (EPTA) currently provides a limit $G\mu < 1.3 \times 10^{-7}$. These bounds are stronger than the CMB limit (22.176), although they are more model-dependent since various parameters, such as Γ and α, enter the computation. The space interferometer LISA is expected to reach values as small as $G\mu \sim 10^{-13}$, for $p = 1$ and large loops.

GW bursts from cusps and kinks

Beside generating a stochastic background, cusps and kinks on strings can also generate GW bursts so strong that, potentially, they could be detectable individually. To compute the burst generated by a cusp one can use the expansion (22.228), (22.229) for the time evolution of the cosmic string near the cusp, and plug it into the energy-momentum tensor of the string, eq. (22.218). The resulting waveform is obtained using the formalism of Section 3.1 of Vol. 1, valid for weak-field sources with arbitrary velocity. This gives the result in the local wave zone, i.e. at distances large compared with the size of the source, but small compared with cosmological scales. One must then evolve the GWs to the present time through a FRW metric.

The computation has been performed in detail in Damour and Vilenkin (2000, 2001). The result is that the amplitude in Fourier space is

$$\tilde{h}(f) \sim \frac{G\mu L}{[(1+z)fL]^{1/3}} \frac{1+z}{t_0 z} \tag{22.257}$$

where t_0 is the present value of cosmic time and z the redshift of the source. The Fourier mode $\tilde{h}(f)$ is focused in a narrow cone in the direction of the cusp velocity, with an opening angle

$$\theta_m \simeq [(1+z)fL/2]^{-1/3} \tag{22.258}$$

Writing the loop length as $L = \alpha t$, and expressing t in terms of the redshift, one finds

$$\tilde{h}(f) \sim G\mu\alpha^{2/3} (ft_0)^{-1/3} \phi_h(z) \tag{22.259}$$

where

$$\phi_h(z) = z^{-1}(1+z)^{-1/3} (1+z/z_{\rm eq})^{-1/3} \tag{22.260}$$

The rate of bursts due to cusps can be estimated assuming that C cusps are formed per oscillation cycle of the string, with $C = O(1)$. A string with length L has on oscillation period $T(L) = L/2$ [see eq. (22.237)], so the number of cusp events per unit space-time volume is

$$\begin{aligned} \nu(t) &\simeq \frac{Cn(L,t)}{T(L)} \\ &= \frac{2Cn(L,t)}{L} \end{aligned} \tag{22.261}$$

Then, the rate of GW bursts in the direction of the observer due to cusps in the redshift interval dz, observed at frequency f, is

$$d\dot{N} \sim \frac{\theta_m^2}{4}(1+z)^{-1}\nu(z)dV(z) \tag{22.262}$$

where the factor $\theta_m^2/4$ is the beaming angle of the burst, the factor $(1+z)^{-1}$ comes from the cosmological time dilatation between the time $t_{\rm obs}$ measured by the observer and the time t of the source, $dt_{\rm obs} =$
$(1+z)dt$, and $dV(z)$ is the proper volume between redshifts z and $z+dz$. In MD

$$dV(z) = 54\pi t_0^3 \left[(1+z)^{1/2} - 1\right]^2 (1+z)^{-11/2} dz \tag{22.263}$$

while in RD

$$dV(z) = 72\pi t_0^3 (1+z_{\rm eq})^{1/2} (1+z)^{-5} dz \tag{22.264}$$

Given the amplitude of a single burst (22.257) and the burst rate (22.262) as a function of redshift, one can compute the typical amplitude of a burst generated by cusps over a period, say, of one year, and compare with the experimental sensitivities. A similar analysis can be done for kinks. Similarly to the stochastic background case discussed earlier, the result depends on the population of loops considered, in particular on whether one considers loops with typical length $L = \alpha t$ with $\alpha \sim \Gamma G\mu$, which was once considered the standard scenario, or whether $\alpha \sim 0.1$, which is favored by the most recent simulations, or even values of $\varepsilon = \alpha/(\Gamma G\mu) \ll 1$, as discussed following eq. (22.255). As we see from eq. (22.259), the signal is not sharply peaked

at a specific frequency, but rather has a smooth power-law behavior, so it can be relevant both for ground-based and for space-borne interferometers.

The result for the scenario with $\varepsilon = 1$ is shown in Fig. 22.21, in the frequency range of LIGO/Virgo, as a function of $\alpha = 50G\mu$. We see that, for some values of the parameters, the signal is potentially detectable.

16.4 22.6 Alternatives to inflation

We have seen in Chapter 21 that inflation provides an elegant solution to the homogeneity and flatness problems of FRW cosmology (as well as diluting unwanted relics such as monopoles that could appear in extensions of the Standard Model and would otherwise dominate over radiation since their density scales slower than $1/a^4$), and furthermore provides the initial adiabatic perturbations that seed structure formation. As such, it is a very successful paradigm. The simplest scenario for obtaining inflation is the single-field slow-roll inflation that we discussed a length in Section 21.1, or relatively simple modifications of it, such as the hybrid inflationary model [see eq. (21.82)]. There have, however, been several attempts at incorporating inflation into a more fundamental theory, such as string theory, that could at the same time address fundamental problems such as the big-bang singularity. Such models have in general distinctive predictions for the stochastic GW background, which differ from that of standard slow-roll inflation computed in Section 21.3. [42] In the following we will briefly discuss some of these models, just to give a flavor of how the prediction of standard slow-roll inflation can be modified in alternative theories of inflation. Of course, all these models suffer from uncertainties due to the difficulty of dealing with the regime of Planckian curvatures, so the results must be taken only as examples of possible scenarios.

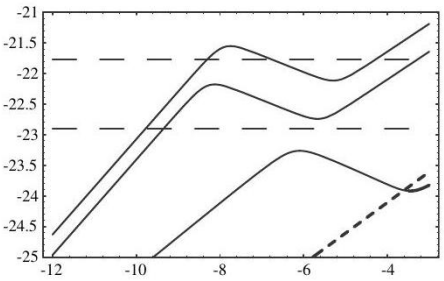


Fig. 22.21 GW amplitude of bursts emitted by cosmic string cusps (upper and middle curves) and kinks (lower curve), in the LIGO/Virgo frequency band, as a function of $\alpha = 50G\mu$, in a base- $10\log - \log$ plot. The upper curve assumes that the average number of cusps per loop oscillation is $C = 1$. The middle curve assumes $C = 0.1$. The lower curve gives the kink signal (assuming only one kink per loop). The horizontal dashed lines indicate the 1-sigma noise levels (after optimal filtering) of initial LIGO and advanced LIGO. The short-dashed line indicates the "confusion" amplitude noise of the stochastic GW background. From Damour and Vilenkin (2001).

As an example of such alternatives to standard inflation, we consider the pre-big-bang model. The starting point is the low-energy effective action from string theory, which, besides

[12542] It should also be observed that in principle these models could also be joined, at sub-Planckian curvature, to standard slow-roll inflation, providing a way of addressing the initial singularity problem while still matching later to standard inflationary cosmology.
[125]

[43] The low-energy effective action of string theory also includes an antisymmetric tensor, whose effect has also been studied in this context. Here, for simplicity, we limit ourselves to the metric-dilaton sector.

the metric, also includes a scalar field, the dilaton ϕ.[43] The effective action in the metric-dilaton sector, to lowest order in the derivatives and in the coupling e^{ϕ}, is given by

$$S_{\rm eff} \sim \int d^D x \sqrt{-g} \left[e^{-\phi} \left(R - \partial_\mu \phi \partial^\mu \phi \right) \right] \tag{22.265}$$

where ϕ is the dilaton field. This action receives two kinds of perturbative corrections: corrections parametrized by e^{ϕ}, which are just higher loops in the field-theory sense, and corrections proportional to α', where $\lambda_s = \sqrt{2\alpha'}$ is the string length. The latter are genuinely string effects, and take into account the finite size of the string. Since α' has dimensions of (length) ^{2}, it is associated with higher-derivative operators such as the square of the Riemann tensor, $R^2_{\mu\nu\rho\sigma}$. Remarkably, there exists a regime in the early Universe evolution of this model where the perturbative approach is well justified, and in a first approximation one can use eq. (22.265), neglecting the corrections. This regime occurs if we take as initial condition a Universe at weak coupling and low curvatures. Then, a generic inhomogeneous string vacuum shows a gravitational instability. Specializing for simplicity to a homogeneous model, the solution of the equations of motion derived from the action (22.265) reads, for generic anisotropic FRW scale factors $a_i(t), i = 1, \ldots, D-1$,

$$\begin{aligned} a_i(t) &= (-t)^{c_i} \\ \phi_i(t) &= \phi_0 + c_0 \log(-t) \end{aligned} \tag{22.266}$$

with

$$\sum_{i=1}^{D-1} c_i^2 = 1, \quad \sum_{i=1}^{D-1} c_i = 1 + c_0 \tag{22.267}$$

This is just a generalization in the presence of the dilaton of the Kasner solution of general relativity. Considering for simplicity an isotropic model, one has in particular a solution with $c_i = -1/\sqrt{D-1}$ and $c_0 = -1 - \sqrt{D-1}$. This corresponds to a super-inflationary evolution: the Hubble parameter $H = \dot{a}/a$ grows until it reaches the string scale $1/\lambda_s$, and eventually formally diverges at a singularity. When $H \sim 1/\lambda_s$, of course, the description based on the lowest-order action is no longer adequate. The inclusion of α' corrections, at least to first order, i.e. terms $O\left(R^2_{\mu\nu\rho\sigma}\right)$, has been shown to turn the unbounded growth of the curvature into a de Sitter phase with a linearly growing dilaton. Of course, in this regime the curvature is anyhow of order one in string units, so a full non-perturbative study would in principle be necessary, but the fact that the $O(\alpha')$ corrections already regularize the big-bang singularity is encouraging.

Since in this new regime the dilaton keeps growing linearly, finally even e^{ϕ} becomes large, even if the evolution started at very weak coupling; at this stage loop corrections are important, and they can trigger the exit from the de Sitter phase. An example of this behavior, using both α' corrections to first order and loop corrections derived from orbifold compactifications of string theory (using modular-invariant loop corrections to the Kähler potential known to all orders, as well as oneloop corrections to the gravitational couplings of the higher-derivative terms), has been studied explicitly, and is shown in Fig. 22.22.

Using this type of evolution as an example of what could be obtained from string theory, the corresponding stochastic GW background can be computed as follows. We consider a model with three phases: first an isotropic superinflationary evolution described by eq. (22.266), with the value of the constants c_i and c_0 determined by eq. (22.267). Then this phase is matched to a de Sitter phase with linearly growing dilaton, which may be considered as representative of a typical solution in the large-curvature regime; the constant values of H and $\dot{\phi}$ in this phase are taken as free parameters, of the order of the string scale. Finally, we match this phase to the

standard RD era. Of course, this model can only be taken as indicative of possible behavior in string cosmology, since what happens in the large-curvature phase is certainly much more complicated, and a full understanding would require a non-perturbative approach. However, we will see that some important characteristics of the GW spectrum depend only on the low-curvature regime, where we can use the lowest-order effective action and which is therefore under good theoretical control.

We specialize for illustration to a isotropic model compactified to four dimensions. Writing the scale factors and the dilaton in terms of conformal time, from eq. (22.266) we find that for $-\infty < \eta < \eta_s$ (with $\eta_s < 0$) we have a dilaton-dominated regime with

$$a(\eta) = -\frac{1}{H_s \eta_s} \left(\frac{\eta - (1-\alpha)\eta_s}{\alpha \eta_s} \right)^{-\alpha} \tag{22.268}$$

$$\phi(\eta) = \phi_s - \gamma \log \frac{\eta - (1-\alpha)\eta_s}{\eta_s} \tag{22.269}$$

and the Kasner conditions eq. (22.267) give $\alpha = 1/(1+\sqrt{3})$ and $\gamma = \sqrt{3}$.

At a value $\eta = \eta_s$ the curvature becomes of the order of the string scale, and there we assume a de Sitter expansion with linearly growing dilaton. In terms of conformal time, this means

$$a(\eta) = -\frac{1}{H_s \eta}, \quad \phi(\eta) = \phi_s - 2\beta \log \frac{\eta}{\eta_s} \tag{22.270}$$

The parameter H_s is of the order of the string scale, while β is a free parameter of the model. Note that $2\beta = \dot{\phi}/H_s$, where $\dot{\phi}$ is the derivative with respect to cosmic time, and in this model $\beta > 0$.

The stringy phase lasts for $\eta_s < \eta < \eta_1$ (again $\eta_1 < 0$), and at η_1 we match this phase to a standard RD era. This gives, for $\eta_1 < \eta < \eta_r$ (with $\eta_r > 0$),

$$a(\eta) = \frac{1}{H_s \eta_1^2} (\eta - 2\eta_1), \quad \phi = \phi_0 \tag{22.271}$$

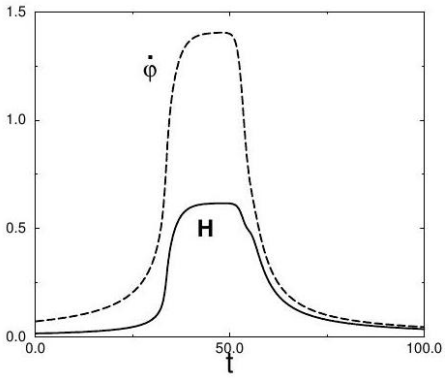


Fig. 22.22 The evolution of the Hubble parameter H and of $\dot{\phi}$ in string cosmology with α' and loop corrections. From Foffa, Maggiore and Sturani (1999).

Then, the standard MD era follows. We have chosen the additive and multiplicative constants in $a(\eta)$ in such a way that $a(\eta)$ and $da/d\eta$ (and therefore also da/dt) are continuous across the transitions. The equation for the tensor perturbations in these three regimes can be solved exactly (see the Further Reading), and matched at the transition points. The matching gives the Bogoliubov coefficients, similarly to the computation in Section 21.3.7. The quantity η_1 can be estimated by observing that, since it marks the end of the de Sitter phase,

$|\eta_1| = 1/\left[H_s a\left(\eta_1\right)\right]$. On the other hand, assuming no intermediate inflationary phase and a direct matching to RD, η_1 also marks the beginning of the RD phase. Therefore, the Friedmann equation (17.117) gives

$$H_s \simeq 0.62 \left(\frac{106.75}{g_*(T)}\right)^{1/6} \frac{H_0 \Omega_R^{1/2}}{a^2\left(\eta_1\right)} \tag{22.272}$$

Then, writing

$$k\left|\eta_1\right| = \frac{2\pi f}{H_s a\left(\eta_1\right)} \tag{22.273}$$

the parameter η_1 can be traded for a parameter f_1 defined by $k\left|\eta_1\right| = f/f_1$. Numerically,

$$f_1 \simeq 2.5 \times 10^{10}\ \text{Hz} \left(\frac{H_s}{0.1 M_{\rm pl}}\right)^{1/2} \left(\frac{106.75}{g_*}\right)^{1/12} \tag{22.274}$$

where we have chosen a reference value H_s of the order of a typical string scale. Similarly, we can introduce a parameter f_s instead of η_s, from $k\left|\eta_s\right| = f/f_s$. However, this parameter depends on the duration of the string phase, and therefore, in contrast to f_1, is totally unknown, even as an order of magnitude. However, since $|\eta_1| < |\eta_s|$, we have $f_s < f_1$.

To summarize, the model has a parameter f_s with dimensions of a frequency, which can have any value in the range $0 < f_s < f_1$, with $f_1 = O(1-10)$GHz, and a dimensionless parameter β. It is useful to further trade the latter for the parameter $\mu = |\beta - (3/2)|$.

Performing the matching between the solutions in the three phases one obtains the Bogoliubov coefficients and then one finally gets the spectrum. The computation can be done fully analytically. From the exact expressions one finds that, in the regime $f \ll f_s$ we have $h_0^2 \Omega_{\rm gw}(f) \propto f^3$, while for $f_s \ll f < f_1$ the slope changes to $h_0^2 \Omega_{\rm gw}(f) \propto f^{3-2\mu}$. Thus, the exact analytic solution obtained from the computation of the Bogoliubov coefficients is well approximated by

$$h_0^2 \Omega_{\rm gw}(f) \simeq h_0^2 \Omega_{\rm gw}\left(f_1\right) \times \begin{cases} \left(\frac{f_1}{f_s}\right)^{2\mu} \left(\frac{f}{f_1}\right)^3 & (f \ll f_s) \\ \left(\frac{f}{f_1}\right)^{3-2\mu} & (f_s \ll f < f_1) \end{cases} \tag{22.275}$$

where the factor $\left(f_1/f_s\right)^{2\mu}$ is determined by continuity in $f = f_s$. The overall normalization is determined by the fact that the value of $h_0^2 \Omega_{\rm gw}(f)$
at the cutoff value f_1 is fixed by the scale H_s reached by the Hubble parameter during the string phase. Writing $H_s = g_1 M_{\rm Pl}$, one finds $h_0^2 \Omega_{\rm gw}\left(f_1\right) \simeq g_1^2 h_0^2 \Omega_R$, where Ω_R is the present fraction of the energy density in radiation. Assuming that $H_s = g_1 M_{\rm Pl}$ is of the order of the string scale, one gets, in typical string models, $g_1 \sim 0.1$. Thus, the value of the peak frequency f_1, as well as that $h_0^2 \Omega_{\rm gw}\left(f_1\right)$, is fixed, at least within some order of magnitude, and assuming typical values for the string mass scale. In contrast, the parameters (μ, f_s) must be taken as free parameters of the model, except for the bounds $\mu \geqslant 0$ and $f_s < f_1$ that follow from their definition. Furthermore we must have $\mu \leqslant 3/2$, otherwise, moving from its value at $f = f_s$ toward lower frequencies, $h_0^2 \Omega_{\rm gw}(f)$ would increase and would eventually violate observational limits, such as the nucleosynthesis bound, that we will discuss in Section 22.7. Depending on the values of the two parameters μ and f_s, a broad range of possibilities for the GW spectrum is possible. If f_s is very close to f_1, or if μ is very close to zero, when moving from its peak at f_1 toward lower frequencies the spectrum decreases as f^3 (just like a black-body spectrum), and is totally unobservable at the frequencies at which ground-based interferometers, space interferometers and pulsar timing array operate. This corresponds to the solid line in Fig. 22.23. In contrast, if $\mu \simeq 3/2$, the spectrum is almost flat in the region $f_s \ll f < f_1$, so, if f_s is

sufficiently small, it can be sizable at the frequencies of interferometers, or even pulsar timing, and in fact possibly even at the very small frequencies probed by the CMB, as shown by the two dashed lines in Fig. 22.23.

The preceding computation only takes into account the matching between the string phase and the RD phase. For waves that enter during RD this is all that is needed. Otherwise, in general, the spectrum is multiplied by the transfer function $T^2\,(f/f_{\rm eq}\,)$, just as we discussed for the slow-roll inflationary spectrum. For $f < f_{\rm eq}$ we then have an extra factor of order $(f_{\rm eq}\,/f)^2$. Thus, just as in slow-roll inflation $h_0^2\Omega_{\rm gw}(f)$ changes from an (almost) flat spectrum at $f \gg f_{\rm eq}$ to a $1/f^2$ behavior at $f < f_{\rm eq}$, in the pre-big-bang scenario $h_0^2\Omega_{\rm gw}(f)$ changes from an f^3 to an f behavior. Of course, this is relevant only if the knee frequency f_s is so low as to be itself comparable to $f_{\rm eq}$, otherwise, at $f < f_{\rm eq}$, $h_0^2\Omega_{\rm gw}(f)$ is anyhow utterly negligible. However, this is the case that would be interesting for detection of this stochastic background from its effect on the CMB. The $\propto f$ behavior of $h_0^2\Omega_{\rm gw}(f)$ at $f \lesssim 10^{-17}$ Hz is visible in the left-most curve of Fig. 22.23.

Despite this wide range of possibilities, the low-frequency behaviors $\propto f$ at $f < f_{\rm eq}$ and $\propto f^3$ at $f_{\rm eq} < f < f_s$ are general properties of this model that depend only on the form of the low-energy effective action, and would therefore be a specific signature of the model. Performing the exact computation one finds that some oscillations are also superimposed on the asymptotic behaviors (22.275), and are visible when the spectrum is plotted on a more narrow range of frequencies; see Fig. 22.24.

The main phenomenological difficulty that must be addressed in this model, to make it a viable competitor with standard slow-roll inflation,

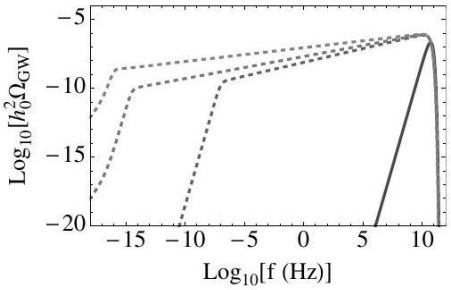


Fig. 22.23 $\Omega_{\rm gw}(f)$ versus f for different values of f_s and μ. From left to right, the curves correspond to $\{f_s = 10^{-16}$ Hz, $\mu = 1.45\}$, $\{f_s = 10^{-14.5}$ Hz, $\mu = 1.42\}$ and $\{f_s = 10^{-7}$ Hz, $\mu = 1.40\}$ (all three dashed), and $f_s = f_1$ (solid line). We set $f_1 = 2.5 \times 10^{10}$ Hz and $h_0^2\Omega_{\rm gw}\,(f_1) = 10^{-6}$.

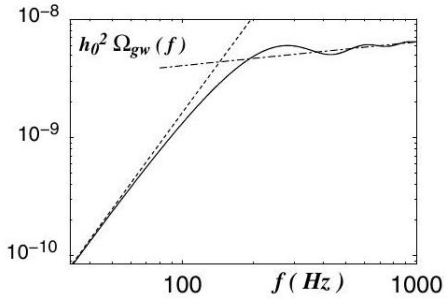


Fig. 22.24$\Omega_{\rm gw}(f)$ versus f for $\mu = 1.4$, $f_s = 100$ Hz and $f_1 = 2.5\times\ 10^{10}$ Hz (solid line). The dashed lines are the the low- and highfrequency limits. From Buonanno, Maggiore and Ungarelli (1997).

[44] There are in principle analogous neutrino asymmetry parameters η_{ν_l}, which, however, have an effect only if they are several orders of magnitude larger than η_B. In standard BBN one assume that this is not the case, for example that η_{ν_l} is of order η_B, in which case they have

a negligible effect, and can be omitted altogether. is the generation of a nearly scale-invariant spectrum of adiabatic perturbations. The simplest mechanism give an unacceptable blue-tilted spectrum of perturbations. However, mechanisms based on the decay of massive axions ("curvaton mechanism") could give a correct spectrum of primordial perturbations; see the Further Reading. In any case, the pre-big-bang model provides an interesting explicit example of how the spectrum of relic GWs can be a signature of alternative early-universe scenarios, as we can appreciate by comparing the pre-big-bang prediction in Fig. 22.23 with the predictions from standard single-field slow-roll inflation in Figs. 21.7 and 21.9.

Several other alternative models addressing the issue of the big-bang singularity have been proposed. Many of them are characterized by an initial contracting phase, leading to Planckian curvatures. At this point the universe bounces back, and the standard phase of expansion follows. Similarly to the pre-big-bang model, the main theoretical problem is the lack of control on the Planckian regime. An example of this family is the ekpyrotic model, which can be described in terms of collision of branes moving in an extra spatial dimension, as well as related cyclic models where the Universe undergoes subsequents phases of contraction, bounce and expansion; see the Further Reading. These models generically predict a strongly blue spectrum, similar to the $f_s = f_1$ case of the pre-big-bang model, shown as a solid line in Fig. 22.23. Once their peak value at high frequency has been normalized so that it does not exceed the nucleosynthesis bound (which we will discuss in Section 22.7), the GW spectrum in these models is therefore completely unobservable at all frequencies at which existing or planned detectors operate.

16.5 22.7 Bounds on primordial GW backgrounds

We next investigate the observational limits on stochastic GW backgrounds of primordial origin that can be obtained from various cosmological observations.

16.5.1 22.7.1 The nucleosynthesis bound

Big-bang nucleosynthesis (BBN) successfully predicts the primordial abundances of deuterium (D), ^{3}He, ^{4}He and ^{7}Li (while heavier elements, up the iron group, are synthesized in stars, and even heavier elements are produced through neutron capture, as we discussed in Section 15.3.4). In the simplest scenario the only cosmological parameter that enters into these predictions is the baryon-to-photon ratio $\eta_B = n_B/n_\gamma$ · [44] The agreement with observation of the prediction for the primordial abundances of these four nuclei (whose abundances span more than nine of orders of magnitude) in terms of a single parameter η_B has long been one of the cornerstones of standard big-bang cosmology. In more recent years, η_B has also been accurately determined by CMB experiments
such as WMAP and Planck, through their determination of the baryon fraction Ω_B. The agreement between the CMB and the BBN determinations of η_B is another remarkable success of the standard cosmological model. In turn, this agreement puts constraints on possible deviations from the standard cosmological model at the BBN epoch. In particular it constraints the existence of extra forms of energy density at the BBN epoch, such as the energy density in GWs or in more exotic components such as primordial black holes, which are not included in the primordial nucleosynthesis computations, and which would spoil the agreement if they were too large.

To physically understand how such a limit comes out, we begin by observing that a crucial quantity in the computations of BBN is the ratio of the number density of neutrons, n_n, to that of protons, n_p. As long as thermal equilibrium is maintained we have (for non-relativistic nucleons, as is appropriate at $T \sim$ MeV, when BBN takes place)

$$\frac{n_n}{n_p} = e^{-Q/T} \tag{22.276}$$

where $Q = m_n - m_p \simeq 1.3\text{MeV}$. Equilibrium is maintained by the process $pe \leftrightarrow n\nu$, as long as its rate $\Gamma_{pe\to n\nu}$ is much larger than the Hubble parameter $H(t)$. When the rate drops below $H(t)$ the process can no longer compete with the expansion of the Universe and the ratio n_n/n_p remains frozen at the value $\exp(-Q/T_\text{f})$, where T_f is the value of the temperature at the time of freeze-out. This number therefore determines the density of neutrons available for nucleosynthesis. Since practically all neutrons available will eventually form ^{4}He, the final primordial abundance of ^{4}He is exponentially sensitive to the freeze-out temperature T_f.

Precise computations of BBN involve detailed numerical studies of the relevant kinetic equations, including a whole series of thermonuclear reactions. However, the main effect of extra contributions to the energy density on T_f can be easily understood analytically by restricting to the process $pe \leftrightarrow n\nu$ and taking for simplicity

$$\Gamma_{pe\to n\nu} \sim G_F^2 T^5 \tag{22.277}$$

(which is really appropriate only in the limit $T \gg Q$). The Hubble parameter is given as usual by the Friedmann equation (17.48), $H^2 = (8\pi/3)G\rho$, where ρ includes all forms of energy density at the time of BBN. BBN takes place deep into the RD phase, but slightly below the MeV scale, when ρ is made up of photons and neutrinos plus any types of extra energy density ρ_extra that might be present, such as, indeed, GWs,

$$\rho = \rho_\gamma + \rho_\nu + \rho_\text{extra} \tag{22.278}$$

The principle of the nucleosynthesis bound is that, since the computation performed with $\rho_\text{extra} = 0$ correctly reproduces the primordial abundances of the light elements, any extra form of energy density present at the time of nucleosynthesis cannot be too large, since otherwise it would
spoil the agreement. To understand the physics involved we must take into account that there are several events that happen at temperatures around an MeV.

First of all, the electron neutrino ν_e decouples from the primordial plasma at $T \simeq 1.9\text{MeV}$ (while ν_μ and ν_τ already decoupled at $T \simeq 3.1\text{MeV}$; see Note 25 on page 383). Subsequently, at $T \simeq 0.3\text{MeV}$, e^+e^-annihilation takes place. As we have discussed in Section 17.5, in a first approximation the entropy of electrons and positrons is transferred uniquely to the photons, whose temperature is thereby raised, compared with the neutrino temperature, by a factor $(11/4)^{1/3}$, as we saw in eq. (17.106). However, neutrino decoupling is not instantaneous, and given that the temperature corresponding to an ideal instantaneous neutrino decoupling is close to the temperature of e^+e^-annihilation, a small fraction of the $e^\pm$entropy actually goes into neutrinos. This, together with a finite-temperature QED effect due to the fact that electromagnetic interactions in the plasma modify the $e^\pm$and γ dispersion relation, actually increases the ratio T_ν/T_γ, compared with the value $(4/11)^{1/3}$. It is customary to take this effect into account by introducing an effective number of neutrino species [see eq. (17.107)], which in the Standard Model with three neutrino families turns out to have the value $N_\text{eff}^{(\nu)} \simeq 3.046$ [see eq. (17.108)].

A third relevant temperature is given by the temperature at which BBN starts. Naively, one would imagine that nucleosynthesis begins, with the formation of deuterium through the process $p + n \to \text{D} + \gamma$, when the temperature drops below the deuterium binding energy, $B_\text{D} \simeq 2.2\text{MeV}$. However, the value of η_B is of order 10^{-9}, so the number density of photons exceeds that of baryons by a factor $\eta_B^{-1} \sim 10^9$. This means that, even at temperatures somewhat below B_D, there are still enough photons in the high-energy tail of the thermal distribution

to destroy the newly formed deuterium nuclei through the inverse process $\mathrm{D} + \gamma \to p + n$. Deuterium production is then delayed (the so-called deuterium bottleneck) until $T \simeq 100\mathrm{keV}$, when the number density of the photons in the energy tail of the thermal distribution has been sufficiently suppressed. Once deuterium begins to form, the whole chain of BBN reactions starts. Thus, because of the deuterium bottleneck, BBN takes place after neutrino decoupling and e^+e^-annihilation, when the neutrino energy density has the form (17.107), with $N_{\rm eff}^{(\nu)} \simeq 3.046$.

To take into account the possible existence of extra energy components, it is convenient to introduce an "effective number of neutrino species" $N_{\rm eff}$, defined so that before e^+e^-annihilation

$$\rho_\nu + \rho_{\rm extra} \equiv \frac{7}{8} N_{\rm eff} \rho_\gamma \tag{22.279}$$

Thus, the addition of an extra component with an energy density equal to that of a massless neutrino would correspond to increasing $N_{\rm eff}$ by one unit, compared with its Standard Model value, but of course in general this is just a way of parametrizing any extra form of energy. To obtain the expression for $\rho_\nu + \rho_{\rm extra}$ just after e^+e^-annihilation we assume
that, similarly to the neutrinos, $\rho_{\rm extra}$ describes a relativistic component, decoupled from the primordial plasma (as would indeed be the case for GWs). Then both ρ_ν and $\rho_{\rm extra}$ scale as $1/a^4$, while the photon temperature evolves according to $g_*^S(T)T^3a^3 = \text{const}$, i.e. constant entropy, so

$$\rho_\gamma \sim T^4 \sim \frac{1}{a^4 \left(g_*^S\right)^{4/3}} \tag{22.280}$$

As we saw in Section 17.5, before e^+e^-annihilation the active degrees of freedoms are the photon (with two helicities), $e^\pm$with two helicities each, and the three neutrinos and three antineutrinos, all with the same temperature. Then, at this epoch, $g_*^S(1\mathrm{MeV}) = 2 + (7/8)(4 + 6) = 43/4$, while, after e^+e^-annihilation, $\left(g_*^S\right)_0 = 43/11$ [see eq. (17.103)], so

$$\frac{\left(g_*^S\right)_0}{g_*^S(1\mathrm{MeV})} = \frac{4}{11} \tag{22.281}$$

Then, just after e^+e^-annihilation (and up to the present epoch, since, neglecting neutrino masses, no other particle becomes non-relativistic in the subsequent evolution),

$$\rho_\nu + \rho_{\rm extra} \equiv N_{\rm eff} \frac{7}{8} \left(\frac{4}{11}\right)^{4/3} \rho_\gamma \tag{22.282}$$

We can then write, after e^+e^-annihilation,

$$\begin{aligned} \rho_{\rm extra} &= \left(N_{\rm eff} - N_{\rm eff}^{(\nu)}\right) \frac{7}{8} \left(\frac{4}{11}\right)^{4/3} \rho_\gamma \\ &\simeq \left(N_{\rm eff} - 3.046\right) \frac{7}{8} \left(\frac{4}{11}\right)^{4/3} \rho_\gamma \end{aligned} \tag{22.283}$$

The Hubble parameter during BBN can be written as

$$\begin{aligned} H^2 &= \frac{8\pi G}{3} \left[1 + N_{\rm eff} \frac{7}{8} \left(\frac{4}{11}\right)^{4/3}\right] \rho_\gamma \\ &= \frac{8\pi G}{3} \left[1 + N_{\rm eff} \frac{7}{8} \left(\frac{4}{11}\right)^{4/3}\right] \frac{\pi^2}{15} T^4 \end{aligned} \tag{22.284}$$

The freeze-out temperature $T_{\rm f}$ is determined by the condition $\Gamma \sim H$, i.e.

$$G_F^2 T_{\rm f}^5 \sim \left(\frac{8\pi^3}{45}\right)^{1/2} \left[1 + N_{\rm eff}\frac{7}{8}\left(\frac{4}{11}\right)^{4/3}\right]^{1/2} \frac{T_{\rm f}^2}{M_{\rm Pl}} \tag{22.285}$$

which gives

$$T_{\rm f} \sim \left(\frac{8\pi^3}{45}\right)^{1/6} \left[1 + N_{\rm eff}\frac{7}{8}\left(\frac{4}{11}\right)^{4/3}\right]^{1/6} \frac{1}{\left(G_F^2 M_{\rm Pl}\right)^{1/3}} \tag{22.286}$$

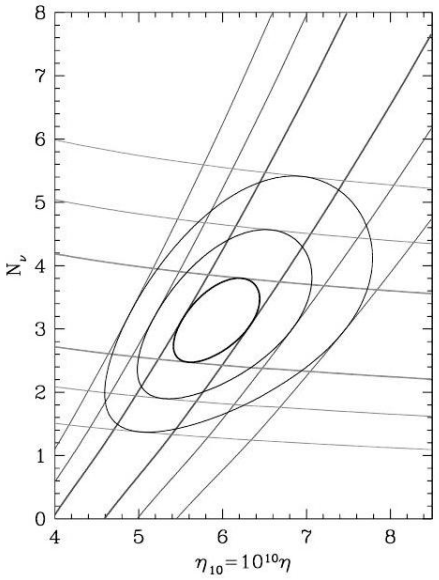


Fig. 22.25 BBN-only constraints on η_B and $N_{\rm eff}$. The nearly vertical curves are the $1\sigma, 2\sigma$ and 3σ limits from D/H, while the nearly horizontal curves are from ^{4}He, and the closed contours combine both. From Cyburt, Fields, Olive and Skillman (2005).

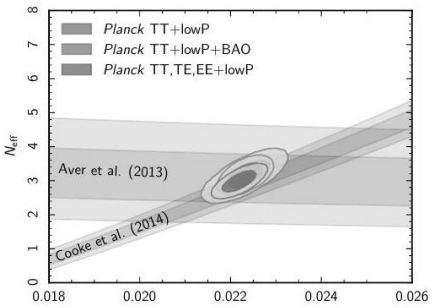


Fig. 22.26 The combined limit in the plane $(N_{\rm eff}\,, \omega_b = h_0^2\Omega_B)$ from Planck+BAO data and BBN. From [Planck Collaboration], Ade et al. (2016a). "Planck 2015 results. XIII. Cosmological parameters". Reproduced with permission (CESO.

We see that any extra form of energy, and in particular an energy density in relic gravitons, results in a larger freeze-out temperature, so more available neutrons and then overproduction of ^{4}He. However, since the density of ^{4}He also increases with the baryon-to-photon ratio η_B, we could compensate for an increase in $N_{\rm eff}$ with a decrease in η_B. Thus, the comparison with ^{4}He alone is not sufficient to put a bound on $N_{\rm eff}$. However, lowering the value of η_B results in an increase in the abundance of D, so this provides a lower limit on η_B. Figure 22.25 shows in particular that the measurements of the abundances of ^{4}He and D/H are nicely complementary for the determination of η_B and $N_{\rm eff}$. Indeed, the ^{4}He fraction has only a very weak (actually,

logarithmic) sensitivity to η_B, but is very sensitive to $N_{\rm eff}$, while for D/H the opposite is the case. Combined together, they already provide a useful bound $N_{\rm eff} < 4.6$ at 95% c.l.

The limit is further improved by adding the value of η_B determined directly by the CMB, as was first done by WMAP and then, with improved accuracy, by Planck. In particular, combining Planck 2015 data from the TT, TE and EE correlators and low-multipole polarization (lowP) data, together with BBN computations of the deuterium abundance, gives

$$N_{\rm eff} = 2.91 \pm 0.37 \tag{22.287}$$

at 95% c.l. (while using helium abundance give a similar value, $N_{\rm eff} = 2.99 \pm 0.39$). The combined contours are shown in Fig. 22.26.

If we use as upper limit the value $N_{\rm eff} = 2.91 + 0.37 = 3.28$, we have $N_{\rm eff} - 3.046 = 0.234$, which we use as the reference value for this quantity. Then, eq. (22.283) reads

$$\rho_{\rm extra}\left(t_{\rm BBN}\right) < 0.053 \left(\frac{N_{\rm eff} - 3.046}{0.234}\right) \rho_\gamma\left(t_{\rm BBN}\right) \tag{22.288}$$

and the energy density in GWs at time of BBN is bounded by $\rho_{\rm gw}\left(t_{\rm BBN}\right) \leqslant \rho_{\rm extra}\left(t_{\rm BBN}\right)$. To obtain a bound on the GW energy density at the present time we observe that, since we have defined $N_{\rm eff}$ in eq. (22.282) in terms of the energy density of neutrinos after e^+e^-annihilation, from this epoch to the present time there is no further change in the effective number of degrees of freedom $g_*(T)$, so the GW energy density scales exactly as the photon energy density. Thus, we get the nucleosynthesis bound at the present time,

$$\frac{\rho_{\rm gw}}{\rho_\gamma} < 0.053 \left(\frac{N_{\rm eff} - 3.046}{0.234}\right) \tag{22.289}$$

Of course this bound holds only for GWs that were already produced at the time of nucleosynthesis, $T \sim 0.1$MeV. It does not apply to stochastic backgrounds produced later, such as backgrounds of astrophysical origin.

Note that this is a bound on the total energy density in GWs, integrated over all frequencies. However, only the frequencies that were inside the horizon at the time of BBN contribute to the expansion of the Universe. The corresponding GW frequency today is obtained from eq. (22.6). We take for definiteness $T \sim 0.1$MeV (corresponding to the onset of BBN after the deuterium bottleneck), so $g_*^S\left(T_*\right) = 43/11$, and we set for definiteness $\epsilon_* = 0.1$, to require that the (reduced) wavelengths are well inside the horizon scale at production. [45] Then this corresponds to GWs that today have a frequency

$$f_{\rm BBN} \simeq 1.5 \times 10^{-11}\ {\rm Hz} \tag{22.290}$$

Writing, as in eq. (7.198) of Vol. 1,

$$\Omega_{\rm gw}(f) = \frac{1}{\rho_c}\frac{d\rho_{\rm gw}}{d\log f} \tag{22.291}$$

and inserting the numerical value $h_0^2\rho_\gamma/\rho_c \simeq 2.473 \times 10^{-5}$ [see eq. (17.110)], eq. (22.289) gives [46]

$$\int_{f=f_{\rm BBN}}^{f=\infty} d(\log f) h_0^2 \Omega_{\rm gw}(f) < 1.3 \times 10^{-6} \left(\frac{N_{\rm eff} - 3.046}{0.234}\right) \tag{22.292}$$

where $\log f$ denotes as usual the natural logarithm, and we have restricted the integral over f to $f > f_{\rm BBN}$. The BBN bound is therefore a bound on the integral of $h_0^2\Omega_{\rm gw}(f)$ for $f > f_{\rm BBN}$, and not directly on $h_0^2\Omega_{\rm gw}(f)$. From the mathematical point of view one could of course imagine

a function $h_0^2\Omega_{\rm gw}(f)$ with a very narrow peak at some frequency f, with a peak value larger than the right-hand side of eq. (22.292), but sufficiently narrow that its contribution to the integral was small enough. However, all cosmological mechanisms that we have discussed rather give a spectrum that, even in the most peaked cases, such as phase transitions, still covers at least one decade in frequency. For such spectra, we can transform the bound in eq. (22.292) into a bound on $h_0^2\Omega_{\rm gw}(f)$, requiring that, if the integral cannot exceed a given value, even its positive-definite integrand $h_0^2\Omega_{\rm gw}(f)$ cannot exceed it over an appreciable interval of frequencies $\Delta \log f \sim 1$, so we can impose

$$h_0^2\Omega_{\rm gw}(f) < 1.3 \times 10^{-6} \left(\frac{N_{\rm eff} - 3.046}{0.234} \right) \quad (\text{ for } f > f_{\rm BBN}) \tag{22.293}$$

This is the meaning of the bound marked as "BBN" in Fig. 22.27 (where we set $N_{\rm eff} - 3.046 = 0.234$). Of course, the precise bound on the spectrum depends on the actual shape of $h_0^2\Omega_{\rm gw}(f)$. For instance if $h_0^2\Omega_{\rm gw}(f)$ has the form

$$h_0^2\Omega_{\rm gw}(f) = \left(h_0^2\Omega_{\rm gw}\right)_* \left(\frac{f}{f_*} \right)^\alpha \tag{22.294}$$

with $\alpha > 0$, for $f_{\rm min} < f < f_{\rm max} \equiv f_*$, and vanishes beyond the cutoff $f_{\rm max}$, then the bound on the peak value $\left(h_0^2\Omega_{\rm gw}\right)_*$ of the spectrum is

$$\begin{aligned} \left(h_0^2\Omega_{\rm gw}\right)_* &< 1.3 \times 10^{-6} \alpha \left[1 - \left(\frac{f_{\rm min}}{f_{\rm max}} \right)^\alpha \right]^{-1} \left(\frac{N_{\rm eff} - 3.046}{0.234} \right) \\ &\simeq 1.3 \times 10^{-6} \alpha \left(\frac{N_{\rm eff} - 3.046}{0.234} \right) \end{aligned} \tag{22.295}$$

[45] There is of course some arbitrariness in the choice $\epsilon_* = 0.1$. In any case ϵ_* must be sufficiently smaller than 1 in order for the bound to apply. Recall in fact from Section 1.4 of Vol. 1 that the energy of GWs is defined only as an average over a volume larger than several wavelengths, so the wavelength must be sufficiently small that the volume used for averaging is at least well inside the Hubble volume. Indeed, we saw in Section 19.5.1 that only when the modes are well inside the horizon can we associate with them an energy density that scales as $1/a^4$.

There is also some arbitrariness in the choice $T \sim 0.1$MeV. BBN is an extended process that lasts from temperatures of a few MeV until temperatures of about 10keV. Certainly GWs that were still outside the horizon at $T = 10$keV, corresponding to $f \simeq 10^{-12}$ Hz, are not bounded by BBN, and those that entered slightly earlier have less of an effect on BBN. So, rather than a sharp cutoff at $f \simeq 10^{-11}$ Hz, one has more precisely a smooth decreases of the effect of GWs, and therefore a less and less stringent bound, as f decreases from $\sim 10^{-10}$ to $\sim 10^{-12}$ Hz.

[46] Observe, from the definition

$$h_0^2\Omega_{\rm gw} = \frac{h_0^2}{\rho_c} \frac{d\rho_{\rm gw}}{d \log f}$$

that $h_0^2\Omega_{\rm gw}$ is independent of h_0, since the explicit factor h_0^2 cancels with the factor h_0^2 that is present in ρ_c. For this reason we prefer to give the bounds in terms of $h_0^2\Omega_{\rm gw}$ rather than $\Omega_{\rm gw}$. This avoids the need to rescale the values of the bounds whenever a more accurate value of h_0 is determined. Furthermore, the determination of h_0 from CMB data assumes the validity of the Λ CDM cosmological model; see Note 16 on page 342. Thus, quoting a result for $\Omega_{\rm gw}(f)$ assuming a value of h_0, such as $h_0 = 0.68$, rather than giving $h_0^2\Omega_{\rm gw}(f)$, introduces a spurious dependence on the cosmological model that, even if numerically small, has nothing to do with the measurement being performed.

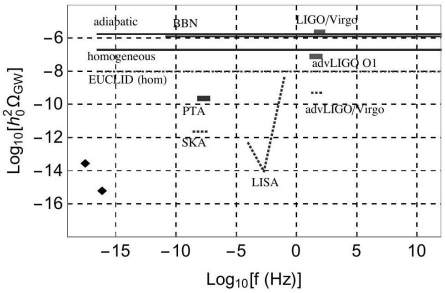


Fig. 22.27 The limits on $h_0^2\Omega_{\rm gw}(f)$ discussed in the text. The solid lines refer to existing (2017) limits, while dashed lines are forecasts for future experiments. The BBN limit is the solid line at $f \gtrsim 1.5 \times 10^{-11}$ Hz. The bound on extra radiation in the CMB from Planck 2015 plus structure formation, for adiabatic initial conditions, is almost superimposed but holds for $f \gtrsim 3 \times 10^{-17}$ Hz, and is labeled "adiabatic". The same bound for homogeneous initial conditions (extrapolated from the bound on adiabatic initial conditions as discussed in Note 47) is labeled "homogeneous". Expectations for EUCLID for homogeneous initial conditions are shown dotted, assuming an improvement by a factor 20 compared with Planck. The CMB limits on r at the pivot scales $k_* = 0.002{\rm Mpc}^{-1}$ and $k_* = 0.05{\rm Mpc}^{-1}$ (corresponding to $f_* = 3.09 \times 10^{-18}$ Hz and 7.73×10^{-17} Hz, respectively) are shown as filled diamonds. The short segments at $f = O(10-100)$Hz show the limit set by initial LIGO/Virgo and the limit set by advanced LIGO in the O1 run. The expectations for the final design of advanced LIGO/Virgo and for LISA are also shown. The present PTA limit (discussed in Section 23.5) is compared with future expectations for the SKA.

where the second equality holds in the limit $(f_{\rm min}/f_{\rm max})^\alpha \ll 1$. Thus the bound on the peak value of the spectrum can be stronger or weaker than eq. (22.293), depending on whether $\alpha < 1$ or $\alpha > 1$, respectively. If $h_0^2\Omega_{\rm gw}(f)$ has a constant value $(h_0^2\Omega_{\rm gw})_*$ between the frequencies $f_{\rm min}$ and $f_{\rm max}$, and is small outside this interval, then eq. (22.292) rather gives

$$\left(h_0^2\Omega_{\rm gw}\right)_* < \frac{1.3\times10^{-6}}{\log\left(f_{\rm max}/f_{\rm min}\right)}\left(\frac{N_{\rm eff}-3.046}{0.234}\right) \tag{22.296}$$

As we have seen in this chapter, the spectrum of some cosmological signals, such as those due to cosmic strings, indeed extends over many decades in frequencies, and in this case the BBN bound is correspondingly more stringent.

16.5.2 22.7.2 Bounds on extra radiation from the CMB

CMB observations provide another sensitive probe of the presence of extra relativistic components, in this case at the time of last scattering. Indeed, extra radiation, such as that due to a fourth relativistic neutrino or to GWs, affects the rate of expansion of the Universe and the epoch of matter-radiation equilibrium, shifting it to a later time, and as a consequence it shifts the position of the acoustic peaks of the CMB. Furthermore, since structure formation can only start in MD, extra radiation delays the start of structure formation. Then, the horizon is larger when structures begin to collapse, and this shifts the peak of the matter power spectrum to larger scales, which can be tested using probes of structure formation, such as galaxy surveys and the Lyman- α forest.

If the GW perturbations are adiabatic, their effect on the CMB is indistinguishable from that of extra neutrinos, and one can therefore immediately apply the bounds on extra massless neutrino families obtained from the CMB. In this case the Planck 2015 data (combining TT, TE, EE and lowP data) together with BAO give $N_{\rm eff} < 3.5$ at 95% c.l. Proceeding as in the derivation of eq. (22.292) we then obtain

$$\int_{f=f_{\rm CMB}}^{f=\infty} d(\log f) h_0^2 \Omega_{\rm gw}(f) < 2.5 \times 10^{-6} \tag{22.297}$$

Combining Planck data with BAO, lensing and the deuterium BBN abundance tightens the bound to

$$\int_{f=f_{\rm CMB}}^{f=\infty} d(\log f) h_0^2 \Omega_{\rm gw}(f) < 1.7 \times 10^{-6} \tag{22.298}$$

which is the value that we plot in Fig. 22.27. The numerical value is therefore of the same order as that from the BBN bound (22.292). However, now the lower limit of the integral is lower, since all modes that were inside the horizon at the epoch of CMB decoupling contribute to the expansion. Putting into eq. (22.6) the value $T_{\rm dec} \simeq 0.26$eV from eq. (17.93) and the corresponding value of g_*, and again setting for definiteness $\epsilon_* = 0.1$, we get

$$f_{\rm CMB} \simeq 3 \times 10^{-17}\ {\rm Hz} \tag{22.299}$$

The bound therefore extends to much lower frequencies, and is shown as the dashed line (marked "adiabatic") in Fig. 22.27. Just as in the case of the BBN bound, it must be recalled that this is actually an integrated bound, and the corresponding bound on $h_0^2 \Omega_{\rm gw}(f)$ becomes stronger for a spectrum that is spread over several decades in frequency, as in eq. (22.296).

As discussed in Section 19.2.1, adiabatic perturbations are defined by the fact that, in the conformal Newtonian gauge, we only have temperature perturbations, so the perturbations in radiation (and therefore, in this case, the perturbations in GWs) are generated by the same temperature fluctuations responsible for matter perturbation. This is not the

[47] See Smith, Pierpaoli and Kamionkowski (2006) and Sendra and Smith (2012). More precisely, using WMAP data the limit is 8.7×10^{-6} for adiabatic perturbations and 1.0×10^{-6} for homogeneous perturbations. With Planck 2015 data the limit on adiabatic perturbations becomes the one in eq. (22.297), while the limit on homogeneous perturbations has not been recomputed. We have drawn the line in Fig. 22.27 assuming that the same factor $\simeq 8.7$ between adiabatic and homogeneous perturbations remains with Planck 2015 data, so this line should only be taken as indicative.

case for any of the GW production mechanism that we have discussed, such as inflation, phase transitions, cosmic strings or pre-big-bang cosmology. For instance, in inflation the initial matter perturbations originate from the decay of the inflaton field at reheating after inflation, while GWs are generated by the amplification of quantum fluctuations during inflation, as discussed in Section 21.3, so the two kind of perturbations are totally unrelated. The same holds for example for the radiation emitted by cosmic string cusps, for which again there is no reason why δ_M should be equal to $(3/4)\delta_{\rm gw}$, or why $\delta_{\rm gw}$ should be equal to δ_γ. As discussed in Section 19.2.1, a second independent type of initial perturbation are isocurvature perturbations, defined by the fact that, in the conformal Newtonian gauge, the total density contrast vanishes. Alternatively, as far as GW perturbations are concerned, we can consider as a second independent component the perturbations whose initial value $(\delta_{\rm gw})_{\rm in} = 0$, i.e. homogeneous initial conditions in the GW field. Because of the source term in eq. (19.39), non-vanishing GW perturbations will anyhow be generated at later times. For such initial conditions the degeneracy with the effect of extra massless neutrinos is broken and, since the CMB is fully consistent with adiabatic initial condition, the resulting bound for homogeneous initial conditions is stronger than that for adiabatic initial conditions, by almost an order of magnitude; see the line labeled "homogeneous" in Fig. 22.27.47 Future CMB experiments, such as EUCLID, could improve these limits by about a factor of 20.

16.5.3 22.7.3 Bounds from the CMB at large angles

We have seen in Chapter 20 that the low- l multipoles of the temperature and polarization anisotropies of the CMB provide an upper bound on the tensor-to-scalar ratio r at a given pivot scale $k_* = 2\pi f_*$. Using eq. (19.290) this translates into an upper bound on $h_0^2\Omega_{\rm gw}(f_*)$. In particular, we saw that CMB experiments typically express their results in terms of an upper bound on $r(k_*)$ at a pivot scale $k_* = 0.05{\rm Mpc}^{-1}$, corresponding to $f_* \simeq 7.73\times 10^{-17}$ Hz; see eq. (21.354). This is slightly larger than $f_{\rm eq} \simeq 3\times 10^{-17}$ Hz [see eq. (19.254)], and we can use the approximation to the transfer function leading to eq. (19.291), apart from a numerical factor close to 1.

At the pivot scale the tensor primordial spectrum (19.281) is simply given by $\mathcal{P}_{T,\,\rm in}(f_*) = A_T(f_*)$, and we write as usual $A_T(f_*) = r(f_*)\,A_{\mathcal{R}}(f_*)$. Then we get

$$h_0^2\Omega_{\rm gw}\left(f = 7.73\times 10^{-17}\ {\rm Hz}\right) \simeq 4\times 10^{-16}\left(\frac{r_{0.05}}{0.1}\right)\left(\frac{A_{\mathcal{R}}(k_{0.05})}{2.14\times 10^{-9}}\right) \tag{22.300}$$

apart from a numerical factor close to 1, which can be obtained from a numerical evaluation of the transfer function. As we saw in eq. (20.332), the current limit on $r_{0.05}$ is $r_{0.05} < 0.12$.

Alternatively, we can use the bound on $r(k_*)$ obtained at the pivot scale $k_* = 0.002{\rm Mpc}^{-1}$, corresponding to $f_* \simeq 3.09\times 10^{-18}$ Hz; see
eq. (21.354). This pivot scale is smaller than $f_{\rm eq}$, so we can use the approximation to the transfer function leading to eq. (19.295), which gives

$$h_0^2\Omega_{\rm gw}(f_*) \simeq 1.87\times 10^{-16}\left(\frac{f_{\rm eq}}{f_*}\right)^2\left(\frac{r(k_*)}{0.1}\right)\left(\frac{A_{\mathcal{R}}(k_*)}{2.14\times 10^{-9}}\right) \tag{22.301}$$

compare with eq. (21.358). In particular, inserting the numerical value $f_* \simeq 3.09\times 10^{-18}$ Hz, we get[48]

$$\begin{aligned} h_0^2\Omega_{\rm gw}\left(f = 3.09\times 10^{-18}\ {\rm Hz}\right) \simeq & 2.4\times 10^{-14}\left(\frac{r_{0.002}}{0.1}\right) \\ & \times\left(\frac{h_0}{0.7}\right)^4\left(\frac{\Omega_M}{0.3}\right)^2\left(\frac{A_{\mathcal{R}}(k_{0.002})}{2.37\times 10^{-9}}\right) \end{aligned} \tag{22.302}$$

The current Planck limit from temperature anisotropies is $r_{0.002} < 0.11$ at 95% c.l. [see eq. (20.331)], while the combined BICEP2/Keck Array/Planck limit from polarization is $r_{0.002} < 0.09$ [see eq. (20.333)].

Observe that the bound on $h_0^2\Omega_{\rm gw}(f)$ at $f = 3.09\times 10^{-18}$ Hz is much weaker than the bound at $f = 7.73\times 10^{-17}$ Hz, despite the fact that the bounds on $r_{0.002}$ and on $r_{0.05}$ are very similar. This is due to the factor $(f_{\rm eq}/f)^2$ that appears in the transfer function at $f \lesssim f_{\rm eq}$. These limits are marked as filled diamonds in Fig. 22.27.

16.5.4 22.7.4 Limits on stochastic backgrounds from interferometers

Correlating the output of different detectors, as discussed in detail in Section 7.8, upper limits on stochastic GW background have been put by the initial LIGO and Virgo interferometers, and have been further improved by advanced LIGO. Assuming a flat spectrum for $\Omega_{\rm gw}(f)$, the limit set in the first advanced LIGO run, in the frequency band $20 - 86$ Hz, is

$$h_0^2\Omega_{\rm gw} < 7.9\times 10^{-8} \quad (95\%\ {\rm c.l}\ ) \tag{22.303}$$

In this frequency range this is better than the bounds from BBN and from the CMB plus structure formation, and also has a different meaning, since it comes from a direct search

for GWs. Thus, it applies not only to stochastic backgrounds that were already present at cosmological epochs, such as those probed by CMB or by BBN, but even to stochastic backgrounds of more recent astrophysical origin, such as those generated by a large number of unresolved BH − BH binaries. Advanced LIGO/Virgo are expected to probe further down to $h_0^2\Omega_{\rm gw} \simeq 5 \times 10^{-9}$ (for a flat spectrum), on a band centered around 30 Hz.

The space-borne interferometer LISA can search for stochastic backgrounds which are not flat in frequency, but rather a broken power-law, such as those due to first-order phase transitions or cosmic strings. For

such spectra, LISA aims at reaching, [49] [48] Recall, from eq. (19.185) and Note 33 on page 624, that $A_{\mathcal{R}} \simeq 2.14 \times 10^{-9}$ is the value measured at the pivot scale $k_* = 0.05{\rm Mpc}^{-1}$. Using the spectral tilt $n_s = 0.9677$ one finds that, at $k_* = 0.002{\rm Mpc}^{-1}$, $A_{\mathcal{R}}(k_*) \simeq 2.37 \times 10^{-9}$. Therefore, in eq. (22.302) we use this as a reference value for $A_{\mathcal{R}}(k_*)$. [49] See the LISA proposal, Audley et al. (2017).

$$h_0^2\Omega_{\rm gw}(f) \simeq 1.3 \times 10^{-11} \left(\frac{10^{-4}\ {\rm Hz}}{f}\right), \quad 0.1{\rm mHz} < f < 20{\rm mHz} \tag{22.304}$$

and

$$h_0^2\Omega_{\rm gw}(f) \simeq 4.5 \times 10^{-12} \left(\frac{f}{10^{-2}\ {\rm Hz}}\right)^3, \quad 20{\rm mHz} < f < 0.2\ {\rm Hz} \tag{22.305}$$

These existing and forecast limits are shown in Fig. 22.27. Finally, we have already mentioned limits on stochastic backgrounds of GWs that can be obtained by pulsar timing arrays (PTAs). We will discuss them in more detail in the next chapter, after having discussed the basic principles of PTAs.

16.6 Further reading

- For reviews of stochastic backgrounds of GWs see Allen (1997) and Maggiore (2000). Reviews of cosmological backgrounds focused on LISA are Binetruy, Bohe, Caprini and Dufaux (2012) and, for cosmological phase transitions, Caprini et al. (2016).

- Parametric resonance and out-of-equilibrium particle production at the end of inflation are discussed by Traschen and Brandenberger (1990), Kofman, Linde and Starobinsky (1994, 1997), Shtanov, Traschen and Brandenberger (1995) and Boyanovsky, de Vega, Holman, Lee and Singh (1995). GW production from preheating in chaotic inflation was first studied by Khlebnikov and Tkachev (1997a, 1997b) and considered again by Easther and Lim (2006). The general formalism for GW production by inhomogeneous random fields and accurate numerical studies of GW production during preheating are presented in Dufaux, Bergman, Felder, Kofman and Uzan (2007), which we have followed in our presentation. Further works include Felder and Kofman (2007) and Easther, Giblin and Lim (2007). Tachyonic preheating in hybrid inflation has been studied in Felder et al. (2001) and Felder, Kofman and Linde (2001). GW production from preheating in hybrid inflation has been studied in Garcia-Bellido and Figueroa (2007), GarciaBellido, Figueroa and Sastre (2008) and Dufaux, Felder, Kofman and Navros (2009), which also contains a useful critical review of earlier literature and a detailed study of parameter space. The pro- duction of GWs from gauge fields at preheating is discussed in Dufaux, Figueroa and Garcia-Bellido (2010).

- The (lack of) electroweak phase transition in the Standard Model for realistic values of the Higgs mass is studied in Kajantie, Laine, Rummukainen and Shaposhnikov (1996a, 1996b). The corresponding crossover temperature is accurately determined, with lattice

simulations, in D'Onofrio and Rummukainen (2016). The addition of a second Higgs doublet or of an extra real scalar singlet can, however, lead to a strong first-order phase transition; see Fromme, Huber and Seniuch (2006) and Barger, Langacker, McCaskey, Ramsey-Musolf and Shaughnessy (2008), respectively. The crossover nature of the QCD phase transition with dynamical quarks (and zero baryon chemical potential) is determined, using lattice QCD, in Aoki, Endrodi, Fodor, Katz and Szabo (2006). The possibility that a large lepton asymmetry could induce a first-order QCD phase transition is discussed in Schwarz and Stuke (2009). The possibility of having a firstorder QCD phase transition thanks to a large preexisting value of μ_B/T, which is later diluted by a short inflationary period at the QCD scale, is discussed in Boeckel and Schaffner-Bielich (2010). Schwaller (2015) discusses the GW production in a first-order phase transition in extensions of the Standard Model with a confining dark sector.

- The theory of decay of false vacuum at finite temperature was developed in classic papers by Cole-
man (1977a) and Callan and Coleman (1977), elaborating on previous work by Voloshin, Kobzarev and Okun (1974), and is reviewed in Coleman (1977b). Instantons and bounces are also discussed in detail in the textbook by Zinn-Justin (2002) and in the review by Vandoren and van Nieuwenhuizen (2008). The extension to finite temperature and the cosmological consequences are studied in particular by Linde (1977, 1981, 1983b). The theory of relativistic bubble combustions is developed in Steinhardt (1982). Bubble nucleation in first-order cosmological phase transitions is studied in Hogan (1983), Turner, Weinberg and Widrow (1992), Anderson and Hall (1992), Enqvist, Ignatius, Kajantie and Rummukainen (1992) and Ignatius, Kajantie, Kurki-Suonio and Laine (1994), and is reviewed in Quiros (1999).

- The fact that a first-order cosmological phase transition can be a powerful source of GWs was first pointed out by Witten (1984) in the context of the QCD phase transition, and was further investigated by Hogan (1986), while Turner and Wilczek (1990) estimated GW production in cosmological bubble collisions in the context of inflation.

The theory of GWs production from bubble collisions in a first-order phase transition was developed in a series of papers by Kosowsky, Turner and Watkins (1992a, 1992b), Kosowsky and Turner (1993) and Kamionkowski, Kosowsky and Turner (1994), using a combination of analytic and numerical techniques. In particular the envelope approximation was proposed in Kosowsky and Turner (1993). The numerical results in the envelope approximation were subsequently refined by Huber and Konstandin (2008b), who also determined the shape of the spectrum near the peak. Jinno and Takimoto (2017) show how to obtain analytic results in the envelope approximation by computing the two-point correlator of the energy-momentum tensor, using the formalism that we have discussed in Section 22.2. Different regimes of bubble evolution are studied in Espinosa, Konstandin, No and Servant (2010). The determination of the parameters α and β in supersymmetric extensions of the Standard Model that could give a first-order phase transition is discussed in Apreda, Maggiore, Nicolis and Riotto (2002) and Huber and Konstandin (2008a, 2008b). The dependence of the peak frequency on the bubble velocity v_b is discussed in Caprini, Durrer, Konstandin and Servant (2009), where it is also pointed out that causality implies $h_0^2\Omega_{\rm gw}(f) \propto f^3$ for super-horizon modes.
In Caprini, Durrer and Servant (2009) it is found that the relevant time-scale is β^{-1} for bubble collisions and $v_b\beta^{-1}$ for turbulence. GW production from bubble collisions and MHD turbulence at the QCD phase transition has been studied in Caprini, Durrer and Siemens (2010). Detailed reviews of the literature are given in Binetruy, Bohe, Caprini and Dufaux (2012), Caprini et al. (2016) and Weir (2017).

- GW production from turbulence was discussed in Kamionkowski, Kosowsky and Turner (1994) and subsequently studied in Kosowsky, Mack and Kahniashvili (2002), Dolgov, Grasso and Nicolis (2002), Caprini and Durrer (2006), Caprini, Durrer and Servant (2009) and Caprini, Durrer and Siemens (2010). GW production from sound waves generated in first-order phase transitions is studied in Hindmarsh, Huber, Rummukainen and Weir (2014, 2015) and Giblin and Mertens (2014).

- Cosmic strings are reviewed in Vilenkin (1985), Hindmarsh and Kibble (1995) and the textbook by Vilenkin and Shellard (2000). The relation of topological defects (such as domain walls and strings) to the topology of the vacuum manifold was discussed in a classic paper by Kibble (1976). The idea that cosmic strings could seed structure formation was proposed by Zeldovich (1980) and Vilenkin (1981a). Our discussion of the string dynamics and NambuGoto action follows Polchinski (1998) and Vilenkin and Shellard (2000). The development of cusps is discussed in Turok (1984). The fact that oscillating loops emit GWs efficiently was first pointed out by Vilenkin (1981b), and has been much studied in the literature; see for example Hogan and Rees (1984), Vachaspati and Vilenkin (1985), Bennett and Bouchet (1988), Caldwell and Allen (1992) and Caldwell, Battye and Shellard (1996). Early work on GWs from cosmic strings is reviewed in Allen (1997). A detailed review of the stochastic GW background from cosmic strings is given in Binetruy, Bohe, Caprini and Dufaux (2012). Numerical simulations of the cosmic string network are discussed in Blanco-Pillado, Olum and Shlaer (2014), where it is found that the network is dominated by large loops with $\alpha \sim 0.1$. The emission of GW bursts from cosmic string cusps was suggested in Berezinsky, Hnatyk and Vilenkin (2000) and studied in detail by Damour and Vilenkin (2000, 2001, 2005). For comparison with experimental sensitivities and existing bounds see also Siemens, Mandic and Creighton (2007) and Olmez, Mandic and Siemens (2010).

- The possibility of production of fundamental strings and D-strings of cosmological size during brane inflation is discussed in Majumdar and Christine-Davis (2002), Jones, Stoica and Tye (2002, 2003), Sarangi and Tye (2002) and Dvali and Vilenkin (2004). Interactions and reconnections between such cosmological-size F- and D-strings is studied in Jackson, Jones and Polchinski (2005).

- The pre-big-bang model was developed in Veneziano (1991), Gasperini and Veneziano (1993a, 1993b, 1994) and Brustein and Veneziano (1994) and is reviewed in Gasperini and Veneziano (2003) and Gasperini (2007). The possibility of obtaining a smooth evolution from the pre-big-bang phase to the post-big-bang era using the α' correction of string theory as well as the string loop corrections is discussed in Gasperini, Maggiore and Veneziano (1997), Brustein and Madden (1998) and Foffa, Maggiore and Sturani (1999). The generation of adiabatic density perturbations in this model can take place through the decay of massive axions, leading to the "curvaton" mechanism, as discussed in Enqvist and Sloth (2002), Lyth and Wands (2002) and Lyth, Ungarelli and Wands (2003); see also Mollerach (1990) and Copeland, Easther and Wands (1997). The spectrum of GWs is computed in Brustein, Gasperini, Giovannini and Veneziano (1995), Brustein, Gasperini and Veneziano (1997) and Buonanno, Maggiore and Ungarelli (1997). Our discussion of the computation of the GW background closely follows Maggiore (2000).

- The ekpyrotic model was proposed in Khoury, Ovrut, Steinhardt and Turok (2001) and Khoury, Ovrut, Seiberg, Steinhardt and Turok (2002). Its cycic extension is discussed in Steinhardt and Turok (2002). The GW spectrum is computed in Khoury, Ovrut, Steinhardt and Turok (2001) and in Boyle, Steinhardt and Turok (2004). For reviews see

Lehners (2008) and Appendix B of Baumann et al. [CMBPol Study Team Collaboration] (2009).

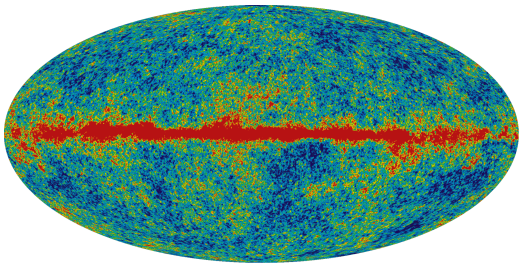
- For reviews of nucleosynthesis and discussions of the BBN bound see Sarkar (1996), Mangano, Miele, Pastor and Peloso (2002), Cyburt, Fields, Olive and Skillman (2005) and Iocco, Mangano, Miele, Pisanti and Serpico (2009). The bound on N_{eff} from Planck is given in Ade et al., [Planck Collaboration] (2016a). The limit of N_{eff} from the effect of extra radiation on the CMB was proposed in Smith, Pierpaoli and Kamionkowski (2006) and updated in Sendra and Smith (2012) and Pagano, Salvati and Melchiorri (2016).
- Limits on stochastic GW backgrounds from the initial LIGO/Virgo interferometers are presented in Aasi et al. [LIGO Scientific and VIRGO Collaborations] (2014). The limit from the O1 run of advanced LIGO is presented in [LIGO Scientific Collaboration and Virgo Collaboration], Abbott et al. (2017a).
- For reviews of stochastic GW background of astrophysical origin see Section 11 of Maggiore (2000) and references therein and, for a more recent discussion, Rosado (2011).

17 Special Models of Cosmology

(пока все остальное тут, потом посмотрим.)

17.1 Cosmic Web

(пока тут, потом мб подниму в приоритете)
(Мб Зельдовича почитаю потом, он вроде про это начал думать.)
(!!!! очень важное применение теории катастроф, кстати, не сложно дожать его!!!)

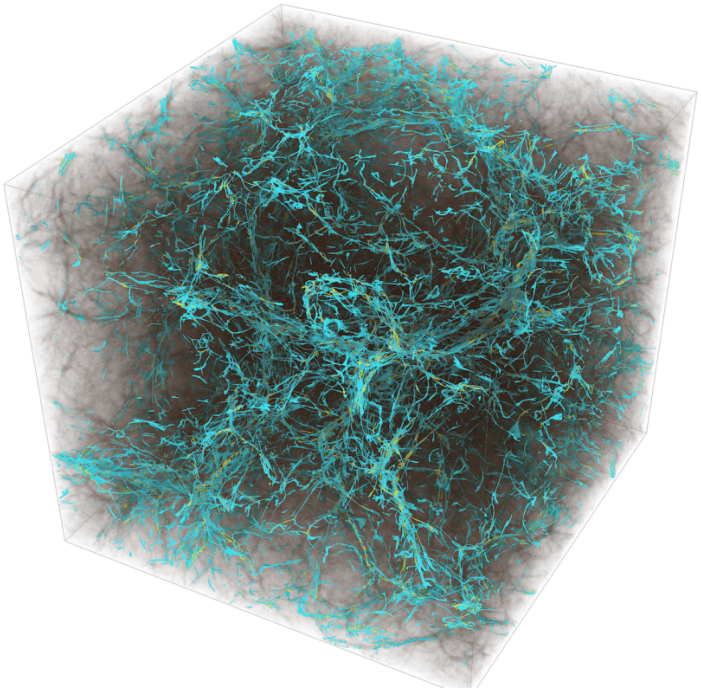


Cosmic web by Feldbrugge

(мб добавлю его потом статью)
(пока хз, насколько это полезно знать, просто прикольные картинки)

Суть (???)

см. Caustic Skeleton & Cosmic Web Job Feldbruggea Rien van de Weygaertb Johan Hiddingb,c Joost Feldbrugged



О применениях (??)

(хз, пока не знаю, если узнаю, зачем это нужно, то напишу сюда потом.)

17.2 Special Models of Cosmology from Landau's Field Theory (??)

(??? I'll add them later, they are not very important)

17.3 Other Special Models

17.3.1 Press–Schechter formalism (???)

пока из Wikipedia, потом подумаю, интересная очень тема, не до нее просто

The Press–Schechter formalism is a mathematical model for predicting the number of objects (such as galaxies, galaxy clusters or dark matter halos[1]) of a certain mass within a given volume of the Universe. It was described in an academic paper by William H. Press and Paul Schechter in 1974.[2]

Background In the context of cold dark matter cosmological models, perturbations on all scales are imprinted on the universe at very early times, for example by quantum fluctuations during an inflationary era. Later, as radiation redshifts away, these become mass perturbations, and they start to grow linearly. Only long after that, starting with small mass scales and advancing over time to larger mass scales, do the perturbations actually collapse to form (for example) galaxies or clusters of galaxies, in so-called hierarchical structure formation (see Physical cosmology).

Press and Schechter observed that the fraction of mass in collapsed objects more massive than some mass M is related to the fraction of volume samples in which the smoothed initial density fluctuations are above some density threshold. This yields a formula for the mass function (distribution of masses) of objects at any given time.

17.3.2 Lyman-alpha forest (???)

In astronomical spectroscopy, the Lyman-alpha forest is a series of absorption lines in the spectra of distant galaxies and quasars arising from the Lyman-alpha electron transition of the neutral hydrogen atom. As the light travels through multiple gas clouds with different redshifts, multiple absorption lines are formed.

The Lyman-alpha forest was first discovered in 1970 by astronomer Roger Lynds in an observation of the quasar 4C 05.34.[1] Quasar 4C 05.34 was the farthest object observed to that date, and Lynds noted an unusually large number of absorption lines in its spectrum and suggested that most of the absorption lines were all due to the same Lyman-alpha transition.[2] Follow-up observations by John Bahcall and Samuel Goldsmith confirmed the presence of the unusual absorption lines, though they were less conclusive about the origin of the lines.[3] Subsequently, the spectra of many other high-redshift quasars were observed to have the same system of narrow absorption lines. Lynds was the first to describe them as the "Lyman-alpha forest".[4] Jan Oort argued that the absorption features are due not to any physical interactions within the quasars themselves, but to absorption inside clouds of intergalactic gas in superclusters.[5]

18 Dark Matter Candidates

(all discussion of this question)

19 Primordial Gravitational Waves

(a special section for them)

20 Black Holes in Cosmology

(a special section for them)

21 Quantum Cosmology: Basics

(see a special note on quantum cosmology for now, there are a lot of theories. I'll write later most basic topics here also)

22 Cosmology from Philosophical Point of View

(general abstract ideas will be here. also, I'll write about ideas which were long time ago about space and Universe)

22.0.1 Other, Exotic Questions

Что на краю Вселенной?

Никто не знает.

Даже можно сказать, что там Чебурашка, потому что никто не сможет это опровергнуть.

(?? спрошу потом еще у умных людей)

23 Unfinished and Wrong Cosmological Theories

(все остальное и во многом бредовое тут. в том числе идеи в истории)

23.1 Theories that Living Creatures Influence Universe (???)

(??? I'll discuss, how much sense is in them?)

23.2 Astrology (???)

(??? I'll discuss, how much sense is in it?)

23.3 What Could be in the Universe if Something Was Different?

(I'll think about some of them, maybe)

24 Cosmological constant by Jérôme Martin

Abstract

This article aims at discussing the cosmological constant problem at a pedagogical but fully technical level. We review how the vacuum energy can be regularized in flat and curved space-time and how it can be understood in terms of Feynman bubble diagrams. In particular, we show that the properly renormalized value of the zero-point energy density today (for a free theory) is in fact far from being 122 orders of magnitude larger than the critical energy density, as often quoted in the literature. We mainly consider the case of scalar fields but also treat the cases of fermions and gauge bosons which allows us to discuss the question of vacuum energy in super-symmetry. Then, we discuss how the cosmological constant can be measured in cosmology and constrained with experiments such as measurements of planet orbits in our solar system or atomic spectra. We also review why the Lamb shift and the Casimir effect seem to indicate that the quantum zero-point fluctuations are not an artifact of the quantum field theory formalism. We investigate how experiments on the universality of free fall can constrain the gravitational properties of vacuum energy and we discuss the status of the weak equivalence principle in quantum mechanics, in particular the Collela, Overhausser and Werner experiment and the quantum Galileo experiment performed with a Salecker-Wigner-Peres clock. Finally, we briefly conclude with a discussion on the solutions to the cosmological constant problem that have been proposed so far.

Everything You Always Wanted To Know About The Cosmological Constant Problem (But Were Afraid To Ask)

24.1 Introduction

This review article aims at discussing the cosmological constant problem [1819, 451, 3, 4, 5, 6, 449, 8, 1393, 10, 11, 12, 508, 14, 15]. This question is central in modern physics because its resolution would certainly mean a very important step forward in our attempts to understand physics beyond the current standard model.

The history of the cosmological constant problem is a long and rich one [5]. Its premises were in fact already present immediately after the birth of quantum field theory. Indeed, from the Heisenberg uncertainty principle, we know that the ground state energy of the quantum mechanical oscillator cannot be zero because the potential and the kinetic energies cannot

vanish at the same time. Since a (free) quantum field can be viewed as an infinite collection of harmonic oscillators, it immediately comes that its ground state energy must be infinite. Of course, this is not yet the cosmological constant problem because gravity does not enter the stage. But, clearly, potentially, a very severe problem is already present.

The infinity mentioned above is the first infinity that one encounters in quantum field theory [16, 17, 18, 19, 20, 21, 22]. However, the presence of this infinity did not prevent the founding fathers of quantum field theory to develop the theory since, in absence of gravity (a working assumption of quantum field theory justified by the weakness of gravity), only differences in energies are observable. Therefore, in this context, this infinity can be merely ignored and the rest of the formalism can be worked out without any problem. As is well-known, other types of infinities appear and, this time, there is no way to get around them. Treating and taming them is the goal of renormalization. The very impressive agreement between high precision measurements in accelerators and the predictions of quantum field theory in presence of radiative corrections is the proof that renormalization is able to correctly regulate these infinities.

Therefore, we see that these two types of infinities are treated very differently. The problem of the zero-point energy is just avoided while the problem of the radiative corrections is directly and explicitly addressed. Clearly, one cannot help thinking that the first problem is in fact swept under the carpet. And, indeed, as soon as gravity is turned on, it badly strikes back.

At this point, it is worth noticing the following. One should not get the impression that the zero-point energy cannot be renormalized. As a matter of fact, as will be discussed in this review, it can be made perfectly finite. However, this finite, renormalized, value of the zero-point energy seems to be too large to be compatible with the observations. Again, this is very different from the usual case of quantum field theory where the finite part extracted from a divergent expression always leads to a good agreement with the experiments. At first sight, the cosmological constant problem is therefore neither the presence of a new infinity nor our inability to regularize it but rather the apparent failure of the renormalization scheme to produce, at the quantitative level, a finite vacuum energy compatible with the observational data.

In fact, a failure of renormalization is not the only logical possibility. The vacuum fluctuations could also be a mathematical artifact of the quantum field theory framework and/or their gravitational properties could be abnormal. However, when one tries to test (theoretically or experimentally) these ideas, it seems that one never encounters a problem of the kind mentioned above. This fact must also be considered as a part of the cosmological constant problem which, therefore, appears to have many different ramifications. We see that this problem is in fact a very deep problem which lies at the cross roads between different branches of physics (gravitational physics, quantum field theory, cosmology etc ...). In brief, it has to do with the gravitational properties of the quantum vacuum.

The importance of this question has recently been reinforced by the discovery that the expansion of our Universe is accelerated [1445, 1524]. According to the standard lore, this could be the first observational evidence that the quantum vacuum is able to curve space-time. On the other hand, the effect of a cosmological constant is visible only on large scales and, therefore, it seems to be problematic to check this result elsewhere than in cosmology. But, if true, the cosmological constant problem represents a unique situation where some aspects of quantum gravity are at play and where, at the same time, corresponding observational signatures are not hopelessly beyond our technical capabilities. This makes it a valuable opportunity to go beyond our current understanding of theoretical physics. The only other situation where the three above mentioned aspects (quantum mechanics, gravity and the possibility to detect sizable observational effects) are mixed is the theory of cosmological inflation [260, 119, 205, 170, 206, 261, 121, 130, 15, 171] (for a review, see Refs. [35, 36, 37]) where the Cosmic Microwave

Background (CMB) anisotropies, first detected by the COBE (COsmic Background Explorer) satellite [38], are supposed to originate from the quantum fluctuations of the inflation and gravitational fields (for the observational status of inflation, see Ref. [39]).

In this review article, we go through all the issues mentioned before. Our prejudice is to present the question at the technical level, in full details, in order for the paper to be reasonably self-contained. The price to pay for this approach is that, sometimes, we are led to consider problems that are well treated in standard textbooks and/or are not, strictly speaking, directly concerned with the cosmological constant problem. On the other hand, this is also an opportunity to review many different subjects belonging to different types of physics.

The outline of the article is the following one. In Sec. 24.2, we discuss how the cosmological constant Λ is introduced in the Einstein field equations and we explain why the vacuum energy also participates in the value of Λ. In Sec. 24.3, we treat the classical cosmological constant problem, *i.e.* we show that, in the presence of gravity, it is no longer true that only differences of energy are observable. We illustrate this discussion by means of the electro-weak phase transition. In Sec. 24.4, we demonstrate that, even if the classical ground state is tuned to zero, the quantum vacuum fluctuations still give a contribution to the cosmological constant. We discuss how the corresponding quantum vacuum energy can be calculated and argue that the method often used in the literature and which consists in introducing a cut-off, is not appropriate. We also investigate how the quantum vacuum energy can be expressed in terms of Feynman diagrams (bubble diagrams) and investigate the case where interactions are present. In Sec. 24.5, we focus on the so-called bubble diagrams. We discuss them in the context of quantum field theory but also in ordinary quantum mechanics. We then show that they have very different properties than the other loop diagrams. In Sec. 24.6, we use the Gaussian effective potential approach to calculate the vacuum energy in a situation where non perturbative effects are present. Since all the calculations are carried out in the case of a scalar field, we treat in Sec. 24.7 the case of other types of fields (spinor and vector fields). This also gives us the opportunity to discuss the idea of super-symmetry. In Sec. 24.8, we present the cosmological constant problem in a more rigorous way. In particular, we show that a flat space-time calculation is a good approximation although a fully consistent approach should be formulated in curved space-time. One concludes the first part of this review article with Sec. 24.9 where we estimate the vacuum energy and find a value very far from the often quoted "122 orders of magnitude". In Sec. 24.10, we explain, from a theoretical point of view, how the value of the cosmological constant was recently measured in cosmology. In particular, we discuss the hypothesis that are implicitly assumed in order to obtain this result. In Sec. 24.11, we discuss whether the cosmological constant can be measured elsewhere than in cosmology and argue that no other experimental context can compete with cosmology. However, we also show that other experiments can put constraints on the vacuum energy that are interesting from the point of view of the cosmological constant problem. In Sec. 24.12, we review the experiments (Lamb shift, Casimir effect) that seem to prove the existence of the vacuum fluctuations. In Sec. 24.13, we investigate whether there are observations that can probe their gravitational properties. In Sec. 24.14, we discuss whether the weak equivalence principle still holds in quantum mechanics since this has obvious implications for the question of the vacuum weight. Finally, in Sec. 28.11, we present our conclusions and very briefly review the solutions to the cosmological constant problem that have been proposed so far. Let us mention that this article is part of a more general review on the dark energy question, see Refs. [40, 41, 42, 43].

Before concluding this introduction, we would like to make the following remark. This article aims at discussing what the cosmological constant problem is. This article does not aim at discussing what solutions to this problem have been proposed even if, as already mentioned, in Sec. 28.11, we say a few words on this topic (and choose, in a totally arbitrary way which only reflects the author prejudices, to focus more on some of them). The justification for the

above mentioned choice is that the question of the gravitational properties of vacuum zero-point fluctuations is already a highly non-trivial one as the variety of the subjects listed above demonstrate. Therefore, it seems to us better to deeply understand what is at play before trying to propose a way out. Of course, it is because this problem is very difficult and has many different ramifications in many different branches of physics that it is so interesting.

24.2 The Cosmological Constant

In this first section, we introduce the cosmological constant as a free parameter in the classical action of the gravitational field. This action, together with the action describing matter, can be expressed as

$$S = \frac{1}{2\kappa}\int \mathrm{d}^4x\sqrt{-g}\,(R-2\Lambda_{\mathrm{B}}) + S_{\mathrm{matter}}\left[g_{\mu\nu},\Psi\right], \tag{297}$$

where $\kappa \equiv 8\pi G/c^4 \equiv 8\pi/m_{\mathrm{Pl}}^2 \equiv 1/M_{\mathrm{pl}}^2$, m_{Pl} and M_{pl} being the Planck mass and the reduced Planck mass respectively. The first term is the standard Einstein-Hilbert action and represents the gravitational part of the total action given by Eq. (297). The cosmological constant appears in the second term of the above expression. As announced, at this level, it is merely a new parameter of the total action and for this reason we write it as Λ_{B} for "bare cosmological constant". It has the dimension of the inverse of a square length. It is compatible with general covariance and is of course compatible with a conserved energy momentum tensor since $\nabla^\mu g_{\mu\nu} = 0$. Therefore, this term appears to be totally natural from the relativistic point of view and there is a priori no reason to discard it. Since, according to the standard lore of field theory, everything which is not forbidden should be considered, the cosmological constant should clearly be included in our description of the gravitational field. Finally, the third term in the above equation denotes the matter action where Ψ represents a generic matter field that we do not need to specify at this stage. Variation of the total action with respect to the metric tensor leads to the Einstein equations of motion which read

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda_{\mathrm{B}}g_{\mu\nu} = \kappa T_{\mu\nu}\,, \tag{298}$$

where the stress-energy tensor is defined by

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S_{\mathrm{matter}}}{\delta g^{\mu\nu}}\,. \tag{299}$$

As already mentioned, Λ_{B} appears to be just a parameter of the model. At this level, the only thing that can be done is to try to constrain it using various observations. But, clearly, in the present framework, there is no way to calculate its value from more fundamental considerations.

However, as originally shown by Sakharov [1563], the nature of the discussion is crucially changed when one takes into account quantum field theory. The point is that the stress energy tensor of a field placed in the vacuum state must be given by

$$\langle 0|T_{\mu\nu}|0\rangle = -\rho_{\mathrm{vac}}g_{\mu\nu}, \tag{300}$$

where ρ_{vac} is the constant energy density of the vacuum. This equation is valid for all the fields present in the Universe, as will be shown below. There are several ways to prove the above result. Firstly, one can use the fact that, in flat space-time (in Minkowski space-time), the only invariant tensor is $\eta_{\mu\nu}$. Since the vacuum state must be the same for all observers, one necessarily has $\langle T_{\mu\nu}\rangle \propto \eta_{\mu\nu}$. In curved space-time, this means that

$$\langle T_{\mu\nu}\rangle = -\rho_{\mathrm{vac}}(t,\mathbf{x})g_{\mu\nu}, \tag{301}$$

and from the fact that the stress-energy tensor must be conserved, we reach the conclusion that ρ_{vac} must be a constant. Therefore, one obtains the formula shown above.

Another way to obtain the same result is to consider a specific example, for instance a scalar field Φ. The corresponding action reads

$$S_{\Phi} = -\int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2}g^{\mu\nu}\partial_\mu\Phi\partial_\nu\Phi + V(\Phi)\right]. \tag{302}$$

where $V(\Phi)$ is the potential. Using the definition (299), the corresponding stress-energy tensor can be written explicitly as

$$T_{\mu\nu} = \partial_\mu\Phi\partial_\nu\Phi - g_{\mu\nu}\left[\frac{1}{2}g^{\alpha\beta}\partial_\alpha\Phi\partial_\beta\Phi + V(\Phi)\right]. \tag{303}$$

From this expression, one sees that the scalar field is in fact a perfect fluid. Now, the vacuum state is the minimum energy state. Clearly, in order to minimize the energy we have to consider a situation where the kinetic energy vanishes and where the field sits at the minimum of its potential. In this case, the stress energy tensor reduces to

$$\langle T_{\mu\nu}\rangle = -V(\Phi_{\text{min}})g_{\mu\nu}, \tag{304}$$

which has exactly the expected form with $\rho_{\text{vac}} = V(\Phi_{\text{min}})$.

In fact what was shown before is that there are at least two sources for the vacuum energy. There is a "classical" contribution given by Eq. (304) which originates from the value of the potential at its minimum. The corresponding vacuum energy will be calculated in Sec. 24.3. There is also a "quantum-mechanical" source given by Eq. (301) which originates from the zero point fluctuations of the ground state. This problem will be treated in Sec. 24.4.

Having established the form of the stress energy momentum tensor in the vacuum, one can now proceed with the Sakharov argument. The next step consists in assuming that the equivalence principle applies to the zero-point fluctuations (here, and from now on, we mean to the "classical" and "quantum-mechanical" contributions as discussed in the last paragraph), that is to say that the zero-point fluctuations gravitate. Clearly, this is not a trivial step but, after all, the vacuum fluctuations are just a specific type of energy and, in general relativity, all forms of energy gravitate. Therefore, a consistent way of writing the Einstein equations when quantum field theory is taken into account seems to be

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda_{\text{B}}g_{\mu\nu} = \kappa T^{\text{matter}}_{\mu\nu} + \kappa\langle T_{\mu\nu}\rangle\,, \tag{305}$$

where on the right hand side the first contribution comes from ordinary matter while the second one represents the contribution originating from the vacuum (of course, it is to be understood that we sum up the contributions coming form all the fields present in the universe). Using the form of $\langle T_{\mu\nu}\rangle$ established before, one arrives at

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda_{\text{eff}}g_{\mu\nu} = \kappa T^{\text{matter}}_{\mu\nu}\,, \tag{306}$$

where

$$\Lambda_{\text{eff}} = \Lambda_{\text{B}} + \kappa\rho_{\text{vac}}. \tag{307}$$

Therefore, we conclude that the effective cosmological constant is the sum of the bare cosmological constant and of a contribution originating from the vacuum fluctuations. The effective cosmological constant Λ_{eff} is the quantity that one can observe and constrain when tests of the Einstein equations are carried out.

As we now discuss, the problem is that $\kappa\rho_{\text{vac}}$ is made of several terms which are all huge in comparison with the observed value of Λ_{eff}.

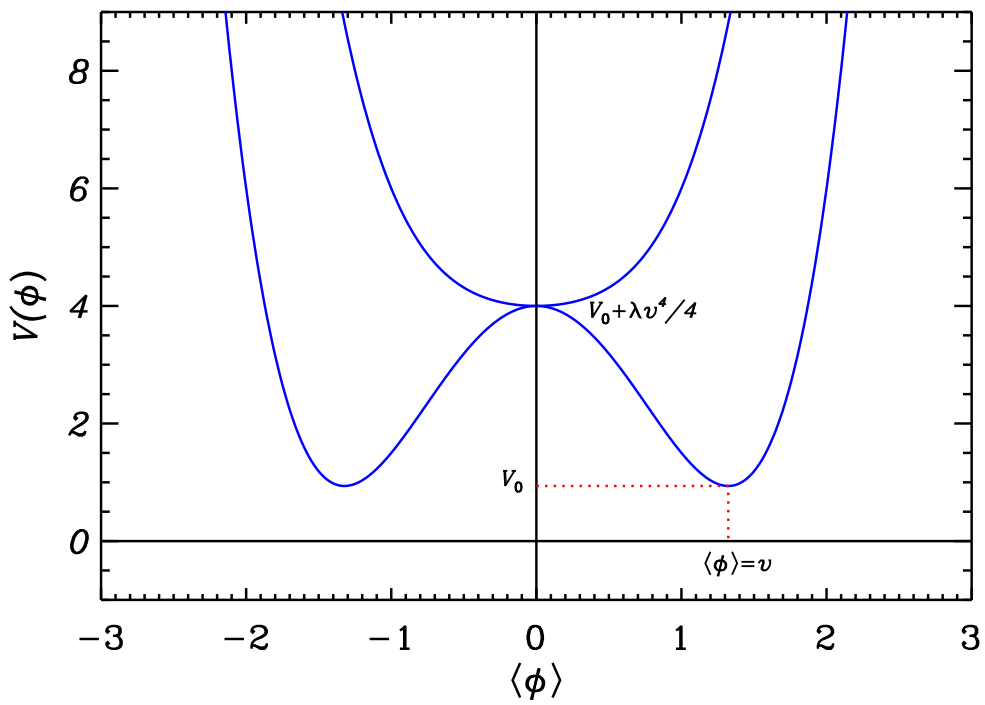


Figure 153: The effective potential given by Eq. (310). Before the transition, for $T > T_{\rm cri}$, the minimum of the potential is located at the origin and the vacuum energy is given by $V_0 + \lambda v^4/4$. After the transition, for $T < T_{\rm cri}$, the minimum is located at $\Phi = v$ and the corresponding vacuum energy has changed and now equals V_0. It is clear that V_0 can always be chosen such that the vacuum energy vanishes either before or after the transition. It is equally clear that one cannot choose the parameters of the potential such that $\rho_{\rm vac}$ is zero before and after the phase transition.

24.3 The Classical Cosmological Constant Problem

The cosmological constant problem is in fact a multi facets question as the quantity $\Lambda_{\rm eff}$ receives contributions from different origins. As was demonstrated in the last section on the example of a scalar field, a classical contribution to $\Lambda_{\rm eff}$ comes from the value of the potential at its minimum. If this is not zero, then it affects the value of the effective cosmological constant. This point is discussed in this section and is exemplified with the electroweak transition. Another contribution comes from the quantum zero point fluctuations but this will be discussed in the next sections.

24.3.1 Phase Transition

It is possible to study the dynamics of a phase transition with the help of the following simple model. Let us consider a scalar field Φ in interaction with another scalar field Ψ such that

$$V(\Phi,\Psi) = V(\Phi) + \frac{\bar{g}}{2}\Phi^2\Psi^2, \tag{308}$$

where $\bar{g}$ is a dimensionless coupling constant and where the self-interacting potential is given by

$$V(\Phi) = V_0 + \frac{\lambda}{4}\left(\Phi^2 - v^2\right)^2, \tag{309}$$

λ being a coupling constant describing the self-interaction and v the value of the scalar field at the minimum of its potential in absence of an interaction with Ψ. The quantity V_0 denotes the classical off-set. If the field Ψ is in thermal equilibrium, then one is entitled to replace Ψ^2 with $\langle\Psi^2\rangle_T$ where the average is taken in a thermal state with temperature T. Since $\langle\Psi^2\rangle_T \propto T^2$ [45], the effective potential becomes

$$V_{\rm eff}(\Phi) = V_0 + \frac{\lambda}{4}\left(\Phi^2 - v^2\right)^2 + \frac{\bar{g}}{2}T^2\Phi^2, \tag{310}$$

where we have slightly redefined the coupling constant $\bar{g}$ in order to take into account the proportionality constant between the thermal average and the temperature. This potential can also be expressed as

$$V_{\rm eff}(\Phi) = V_0 + \frac{\lambda v^4}{4} + \frac{\lambda v^2}{2}\left(\frac{T^2}{T_{\rm cri}^2} - 1\right)\Phi^2 + \frac{\lambda}{4}\Phi^4, \tag{311}$$

where we have defined $T_{\rm cri} = v\sqrt{\lambda/\bar{g}}$. One sees on the above equation that the interaction with the field Ψ gives a temperature dependence to the effective mass of the Φ field, namely

$$m_{\rm eff}^2(T) \equiv \lambda v^2\left(\frac{T^2}{T_{\rm cri}^2} - 1\right). \tag{312}$$

As a consequence, when $T > T_{\rm cri}$, that is to say before the transition, the square of the effective mass is positive while, after the transition, when $T < T_{\rm cri}$, it becomes negative. In the first situation, the minimum is located at $\Phi = 0$ and the corresponding value of the vacuum energy is $V_0 + \lambda v^4/4$, see Fig. 153. In the second situation, the minimum is located at $\Phi = v$ and the off-set is simply given by V_0, see Fig. 153. We now clearly see the problem. If we require the vacuum energy to vanish before the transition, we must choose $V_0 = -\lambda v^4/4$. But, then, after the transition, the vacuum energy is no longer zero and is given by the negative value $\rho_{\rm vac} = -\lambda v^4/4$. On the other hand, one can choose $\rho_{\rm vac}$ to vanish after the transition. In this case, one must choose $V_0 = 0$. But this means that it was not zero before the transition, $\rho_{\rm vac} = \lambda v^4/4$. This last option can maybe viewed as the preferred solution since we do not have direct observational constraints on the matter content of the universe prior to Big Bang Nucleosynthesis (BBN). Anyhow, we see that the choice of a parameter in the potential allows us to change the vacuum energy at the classical level, hence the name "classical cosmological constant problem". But, clearly, the off-set cannot be zero before and after the transition. Of course, in order to really establish that there is a problem, the question remains to estimate this vacuum energy for realistic phase transition and to compare it with the observational constraints. In the next subsection, we turn to the first issue and consider the case of the electroweak phase transition.

24.3.2 The Electro-weak Phase Transition

In this section, we calculate the vacuum energy "induced" by the electroweak phase transition. In order to achieve this goal, we first must recall some basics facts about the standard model of particle physics, see Refs. [16, 17, 18, 19, 20, 21, 22].

The Higgs field is a doublet of complex scalar fields. It is charged under the group $\mathrm{U(1)_Y} \times \mathrm{SU(2)_L}$ characterized by two coupling constant g' and g. The Higgs field Lagrangian reads

$$\mathcal{L}_{\rm Higgs} = -\left(D_\mu\Sigma\right)^\dagger D^\mu\Sigma - V\left(\Sigma, \Sigma^\dagger\right), \tag{313}$$

where the covariant derivative can be expressed as

$$D_\mu\Sigma = \partial_\mu\Sigma + i\frac{g'}{2}Y_{\rm H}B_\mu\Sigma + igT_aW_\mu^a\Sigma, \tag{314}$$

where a runs from one to three. The quantities B_μ and W_μ^a are the gauge bosons, Y_{H} is the Higgs weak hyper-charge and T_a are the generators of the SU(2) group (see below for more details about these quantities). The potential is chosen to be

$$V\left(\Sigma,\Sigma^\dagger\right)=\frac{m^2}{2}\Sigma^\dagger\Sigma+\frac{\lambda}{4}\left(\Sigma^\dagger\Sigma\right)^2, \tag{315}$$

and clearly resembles Eqs. (310) and (311) for $T<T_{\text{cri}}$. Indeed, in this expression, the quantity m^2 is negative (as explained before, m^2 is in fact the square of the effective mass). This also means that the minimum of the potential is not located at $\langle\Sigma\rangle=0$ but at

$$\langle\Sigma\rangle=v=\sqrt{-\frac{m^2}{\lambda}}, \tag{316}$$

obtained from the condition $\partial V/\partial\left(\Sigma^\dagger\Sigma\right)=0$. The quantity v is in fact the new vacuum expectation value of the Higgs after spontaneous symmetry breaking. In particle physics, one studies the theory after the transition, when the Higgs field has stabilized at its new minimum. But, in reality, as explained before, we must consider that the transition is a dynamical process. Before the transition we had $m^2>0$ and the minimum was located at $\langle\Sigma\rangle=0$ and after the transition, m^2 has become negative and, consequently, the new vacuum is given by v. Returning to the minimum after the transition, it is easy to show that the value of the potential at $\langle\Sigma\rangle=v$ is given by

$$V\left(\langle\Sigma\rangle=v\right)=-\frac{m^4}{4\lambda}, \tag{317}$$

and is negative. Clearly, this is because we have chosen $V_0=0$ in Eq. (315). As already explained in Sec. 24.3.1, this means that the vacuum energy vanishes before the transition and becomes negative after, see also Fig. 153. This means that we have considered the situation corresponding to the left panel in Fig. 154. Of course, this is arbitrary and we could also have considered the situation corresponding to the right panel in Fig. 154. Again, as explained in the previous subsection, in this case, the vacuum energy would vanish after the transition but not before.

The numerical value of the vacuum energy given by Eq. (317) is fixed by the electroweak physics in particular by the parameters m and λ. Therefore, in order to compute ρ_{vac} explicitly, we must explain how m and λ are evaluated. For this purpose, we now briefly recall the main features of the standard model of particle physics. Let us come back to the Higgs Lagrangian given at the beginning of this section, see Eq. (313). In the expression of $\mathcal{L}_{\text{Higgs}}$, the quantities T_a are the generator of $\text{SU}(2)_{\text{L}}$ and are given by $T_a=\sigma_a/2$ where σ_a are the Pauli matrices

$$\sigma_1=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \sigma_2=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\quad \sigma_3=\begin{pmatrix}1&0\\0&-1\end{pmatrix}. \tag{318}$$

Using the gauge invariance, one can always rewrite the complex doublet as

$$\Sigma=\left(v+\frac{H}{\sqrt{2}}\right)\begin{pmatrix}0\\1\end{pmatrix}, \tag{319}$$

where H is a real scalar field, the "Higgs boson". Therefore, the covariant derivative (314) can be expressed as

$$D_\mu\Sigma=\begin{pmatrix}i\dfrac{g}{2}\left(v+\dfrac{H}{\sqrt{2}}\right)\left(W_\mu^1-iW_\mu^2\right)\\[2ex]\dfrac{1}{\sqrt{2}}\partial_\mu H+i\left(v+\dfrac{H}{\sqrt{2}}\right)\left(\dfrac{g'}{2}Y_{\text{H}}B_\mu-\dfrac{g}{2}W_\mu^3\right)\end{pmatrix}, \tag{320}$$

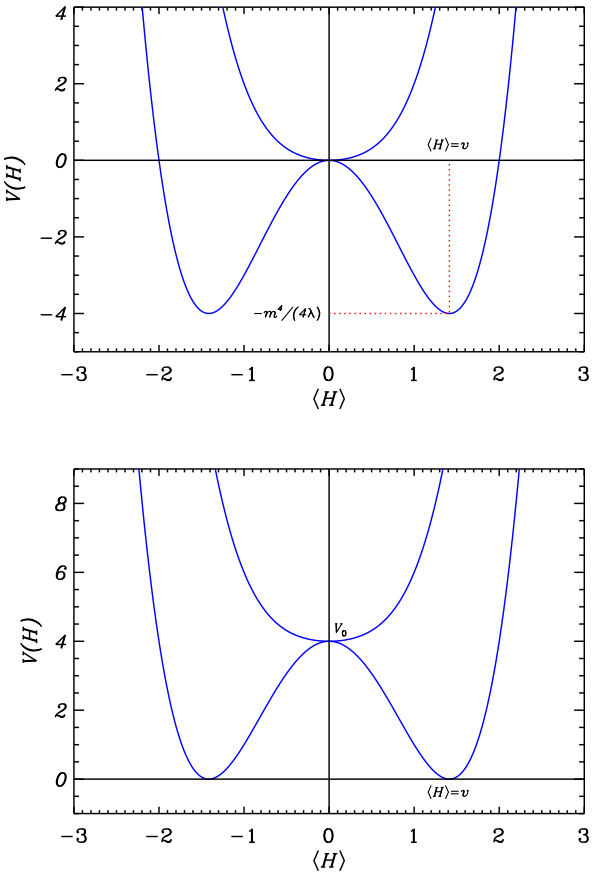


Figure 154: Effective potential of the Higgs boson before and after the electroweak phase transition. The left panel corresponds to a situation where the vacuum energy vanishes at high temperature. As a consequence $\rho_{\rm vac}$ is negative at temperature smaller than the critical temperature. This is the situation treated in the text where the quantity $-m^4/(4\lambda)$ is explicitly calculated. On the right panel, the off-set parameter V_0 is chosen such that the vacuum energy is zero after the transition. As a consequence, it does not vanish at high temperatures.

and the Higgs Lagrangian (313) becomes

$$\begin{aligned}\mathcal{L}_{\text{Higgs}} &= -\frac{1}{2}\partial_\mu H\partial^\mu H - V\left(\Sigma,\Sigma^\dagger\right) - \left(v^2+\sqrt{2}vH+\frac{1}{2}H^2\right)\left(\frac{g^2}{4}W^1_\mu W^{1\mu}+\frac{g^2}{4}W^2_\mu W^{2\mu}+\frac{g^2}{4}W^3_\mu W^{3\mu}\right.\\ &\left.+\frac{g'^2}{4}Y_{\text{H}}^2 B_\mu B^\mu - \frac{gg'}{2}Y_{\text{H}} B_\mu W^{3\mu}\right),\end{aligned} \tag{321}$$

where the scalar potential can be written as

$$V\left(\Sigma,\Sigma^\dagger\right) = -\frac{\lambda v^4}{4}+\frac{1}{2}\lambda v^2 H^2+\frac{\lambda}{2}\frac{v}{\sqrt{2}}H^3+\frac{\lambda}{16}H^4. \tag{322}$$

We now introduce the physical gauge bosons. As is well-known, they are defined by

$$B_\mu = -\sin\theta Z_\mu+\cos\theta A_\mu\,,\quad W^3_\mu = \cos\theta Z_\mu+\sin\theta A_\mu\,,\quad W^2_\mu = \frac{W^+_\mu - W^-_\mu}{i\sqrt{2}}\,,\quad W^1_\mu = \frac{W^+_\mu+W^-_\mu}{\sqrt{2}}\,, \tag{323}$$

where θ is the Weinberg angle, A_μ represents the photon and Z_μ, $W^\pm_\mu$ the massive gauge bosons. Expressing the Higgs Lagrangian (321) in terms of these new fields, one obtains the following expression

$$\begin{aligned}\mathcal{L}_{\text{Higgs}} &= -\frac{1}{2}\partial_\mu H\partial^\mu H - V\left(\Sigma,\Sigma^\dagger\right) - \left(v^2+\sqrt{2}vH+\frac{1}{2}H^2\right)\left[\frac{g^2}{4}W^+_\mu W^{-\mu}+\frac{g^2}{4}W^-_\mu W^{+\mu}\right.\\ &+Z_\mu Z^\mu\left(\frac{g^2}{4}\cos^2\theta+\frac{g'^2}{4}Y_{\text{H}}^2\sin^2\theta+\frac{gg'}{2}Y_{\text{H}}\sin\theta\cos\theta\right)+A_\mu A^\mu\left(\frac{g^2}{4}\sin^2\theta+\frac{g'^2}{4}Y_{\text{H}}^2\cos^2\theta-\frac{gg'}{2}Y_{\text{H}}\sin\theta\cos\theta\right)\\ &\left.+Z_\mu A^\mu\left(\frac{g^2}{2}\cos\theta\sin\theta-\frac{g'^2}{2}Y_{\text{H}}^2\cos\theta\sin\theta-\frac{gg'}{2}Y_{\text{H}}\cos^2\theta+\frac{gg'}{2}Y_{\text{H}}\sin^2\theta\right)\right].\end{aligned} \tag{324}$$

Of course $\mathcal{L}_{\text{Higgs}}$ is not the only term in the standard model Lagrangian. There are also the terms describing the kinetic terms of the gauge fields

$$\mathcal{L}_{\text{K-Gauge}} = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu}-\frac{1}{4}W_{a\mu\nu}W^{a\mu\nu} = -\frac{1}{4}A_{\mu\nu}A^{\mu\nu}-\frac{1}{4}Z_{\mu\nu}Z^{\mu\nu}-\frac{1}{4}W_{+\mu\nu}W^{+\mu\nu}-\frac{1}{4}W_{-\mu\nu}W^{-\mu\nu}+\cdots\,, \tag{325}$$

where $B_{\mu\nu}$ is the B field strength, namely $B_{\mu\nu}\equiv\partial_\mu B_\nu-\partial_\nu B_\mu$ and where a similar definition applies to the other gauge fields.

Finally, one must also specify how leptons are described. In the standard model of particle physics, a fermion is represented by a four-component Dirac spinor Ψ, see also Sec. 24.7.1 where the zero-point fluctuations of spinors are calculated. Let us now recall what chirality is. For this purpose, we introduce the matrix γ^5 such that $\left(\gamma^5\right)^2=\mathbb{I}$ and $\left(\gamma^5\right)^\dagger=\gamma^5$. This allows us to define two chiral projectors by

$$P_{\text{R}}\equiv\frac{\mathbb{I}_4+\gamma_5}{2},\quad P_{\text{L}}=\frac{\mathbb{I}_4-\gamma_5}{2}, \tag{326}$$

where $\mathbb{I}_4$ denotes the identity four by four matrix. We now define the Dirac matrices, see Sec. 24.7.1 for a more general approach. In the chiral representation, they are given by

$$\begin{aligned}\gamma^0 &= \begin{pmatrix}0 & -\mathbb{I}_2\\ -\mathbb{I}_2 & 0\end{pmatrix},\quad \gamma^i=\begin{pmatrix}0 & \sigma^i\\ -\sigma^i & 0\end{pmatrix},\\ \gamma^5 &= \begin{pmatrix}\mathbb{I}_2 & 0\\ 0 & -\mathbb{I}_2\end{pmatrix},\end{aligned} \tag{327}$$

and the two projectors can be written as

$$P_{\text{R}} = \begin{pmatrix}\mathbb{I}_2 & 0\\ 0 & 0\end{pmatrix},\quad P_{\text{L}}=\begin{pmatrix}0 & 0\\ 0 & \mathbb{I}_2\end{pmatrix}. \tag{328}$$

Given a Dirac spinor Ψ (we denote the four-component spinors by a capital Greek letter and the two-component spinors by an ordinary Greek letter)

$$\Psi=\begin{pmatrix}\psi_{\text{R}}\\ \psi_{\text{L}}\end{pmatrix}, \tag{329}$$

one can thus define the two states of chirality (right and left handed) by means of the two following expressions

$$\Psi_{\rm R} \equiv P_{\rm R}\Psi = \begin{pmatrix}\psi_{\rm R}\\ 0\end{pmatrix}, \quad \Psi_{\rm L} \equiv P_{\rm L}\Psi \begin{pmatrix}0\\ \psi_{\rm L}\end{pmatrix}. \tag{330}$$

The quantities $\psi_{\rm R,L}$ are two-component spinors known as Weyl spinors. They play a very important role in super-symmetry, see Sec. 24.7.4. Notice that we also have $\overline{\Psi} \equiv \Psi^\dagger\gamma^0 = -\left(\psi_{\rm L}^\dagger, \psi_{\rm R}^\dagger\right)$. This implies that the (mass-less) Dirac Lagrangian can be expressed as

$$\mathcal{L}_{\rm Dirac} = -i\overline{\Psi}\gamma^\mu\partial_\mu\Psi = -i\psi_{\rm L}^\dagger\overline{\sigma}^\mu\partial_\mu\psi_{\rm L} - i\psi_{\rm R}^\dagger\sigma^\mu\partial_\mu\psi_{\rm R}, \tag{331}$$

where we have defined the covariant Pauli matrices $\sigma^\mu = (\mathbb{I}_2, \sigma^i)$ and $\overline{\sigma}^\mu = (\mathbb{I}_2, -\sigma^i)$.

After these brief reminders, we now consider the first family of the standard model leptons (the other two families can be treated in the same fashion and the quarks must be considered separately), *i.e.* the electron and the neutrino. Following the usual convention, we denote by $e_{\rm R}$ the right-handed electron Weyl spinor and by $e_{\rm L}$ the left-handed one. With regards to the neutrino, only the left-handed particle is present in the standard model and will be denoted as $\nu_{\rm L}$. Then, we have the following properties. The particles $e_{\rm L}$ and $\nu_{\rm L}$ are charged under $\mathrm{U}(1)_{\rm Y}$ but also under $\mathrm{SU}(2)_{\rm L}$, *i.e.* they are considered as a complex doublet of two-component spinors

$$L_e = \begin{pmatrix}\nu_{\rm L}\\ e_{\rm L}\end{pmatrix}, \tag{332}$$

On the other hand, $e_{\rm R}$ is charged under $\mathrm{U}(1)_{\rm Y}$ but is postulated to be a $\mathrm{SU}(2)_{\rm L}$ singlet. Therefore, concretely, the Lagrangian can be written as

$$\begin{aligned}\mathcal{L}_{\rm Leptons} &= -iL_e^\dagger\overline{\sigma}^\mu\left(\partial_\mu + i\frac{g'}{2}Y_{\rm L}B_\mu + igT_aW_\mu^a\right)L_e\\ &\quad -ie_{\rm R}^\dagger\bar{\sigma}^\mu\left(\partial_\mu + i\frac{g'}{2}Y_{\rm R}B_\mu\right)e_{\rm R},\end{aligned} \tag{333}$$

where $L_e^\dagger$ means $\left(\nu_{\rm L}^\dagger, e_{\rm L}^\dagger\right)$. Notice the factor $1/2$ in front of the gauge boson B_μ which is the usual convention [this factor was already present in Eq. (314)]. The quantities $Y_{\rm L}$ and $Y_{\rm R}$ are the leptonic weak hyper-charges. Let us focus on the first term in the above expression. The term between the parenthesis is a two by two matrix that can be written in terms of the physical gauge bosons, using the definitions (323). Similarly, the second term can be expressed in terms of the fields A_μ and Z_μ. Then, the Lagrangian for the leptons can be rewritten as

$$\begin{aligned}\mathcal{L}_{\rm Leptons} &= -i\nu_{\rm L}^\dagger\overline{\sigma}_\mu\left[\partial_\mu - i\left(\frac{g'}{2}Y_{\rm L}\sin\theta - \frac{g}{2}\cos\theta\right)Z_\mu + i\left(\frac{g'}{2}Y_{\rm L}\cos\theta + \frac{g}{2}\sin\theta\right)A_\mu\right]\nu_{\rm L}\\ &\quad -ie_{\rm L}^\dagger\overline{\sigma}_\mu\left[\partial_\mu - i\left(\frac{g'}{2}Y_{\rm L}\sin\theta + \frac{g}{2}\cos\theta\right)Z_\mu + i\left(\frac{g'}{2}Y_{\rm L}\cos\theta - \frac{g}{2}\sin\theta\right)A_\mu\right]e_{\rm L}\\ &\quad +\frac{g}{\sqrt{2}}\nu_{\rm L}^\dagger\overline{\sigma}^\mu W_\mu^- e_{\rm L} + \frac{g}{\sqrt{2}}e_{\rm L}^\dagger\overline{\sigma}^\mu W_\mu^+\nu_{\rm L} - ie_{\rm R}^\dagger\overline{\sigma}^\mu\left(\partial_\mu - i\frac{g'}{2}Y_{\rm R}\sin\theta Z_\mu + i\frac{g'}{2}Y_{\rm R}\cos\theta A_\mu\right)e_{\rm R}.\end{aligned} \tag{334}$$

Therefore, the total Lagrangian is now given by $\mathcal{L} = \mathcal{L}_{\rm Higgs} + \mathcal{L}_{\rm K-Gauge} + \mathcal{L}_{\rm Leptons}$. This Lagrangian depends on various free parameters that we now discuss and choose. Firstly, the neutrino is not electromagnetically charged and from Eq. (334) one has that

$$g'Y_{\rm L}\cos\theta + g\sin\theta = 0. \tag{335}$$

Secondly, the electron carries the charge e. This means that one must recover the following Lagrangian,

$$\mathcal{L}_{\rm EM} = -i\overline{\Psi}\gamma^\mu\left(\partial_\mu + iQA_\mu\right)\Psi, \tag{336}$$

where Q is the electromagnetic charge, $Q = e$ in our case. Working out the interaction part of the above Lagrangian, one obtains

$$\mathcal{L}_{\rm EM}^{\rm int} = -Qe_{\rm L}^\dagger\overline{\sigma}_\mu A_\mu e_{\rm L} - Qe_{\rm R}^\dagger\sigma_\mu A_\mu e_{\rm R}, \tag{337}$$

and, looking at Eq. (334), one deduces that

$$\frac{g'}{2}Y_{\rm L}\cos\theta - \frac{g}{2}\sin\theta \;=\; \frac{g'}{2}Y_{\rm R}\cos\theta = e. \tag{338}$$

As a consequence, comparing Eq. (335) and (338), one can establish that $2Y_{\rm L} = Y_{\rm R}$. It is conventional to choose $Y_{\rm L} = -1$ and therefore $Y_{\rm R} = -2$. Then, Eq. (335) leads to $g' = g\tan\theta$, from which one can write $\cos\theta = g/\sqrt{g^2+g'^2}$ and $\sin\theta = g'/\sqrt{g^2+g'^2}$. Finally, Eq. (338) implies that $g\sin\theta = e$. Everything can be summarized by the Gell-Mann formula

$$Q = e\left(I_3 + \frac{Y}{2}\right), \tag{339}$$

if the leptonic isospin doublet is assigned a "weak isospin" $I = 1/2$ such that the neutrino has a third component $I_3 = 1/2$ while the electron is such that $I_3 = -1/2$. Of course the right handed particles are such that $I = 0$. Returning to Eq. (324) and using the above equations, we see that the photons remains mass-less if $Y_{\rm H} = 1$. This also allows us to determine the mass of the gauge bosons and one arrives at

$$m^2_{W^\pm} \;=\; \frac{g^2v^2}{2}, \tag{340}$$

$$m^2_Z \;=\; 2v^2\left(\frac{g}{2}\cos\theta + \frac{g'}{2}\sin\theta\right)^2 = \frac{m^2_{W^\pm}}{\cos^2\theta}. \tag{341}$$

One also easily proves that the coefficient of the cross-term $Z_\mu A^\mu$ vanishes. Finally, looking at Eq. (322), one establishes that the Higgs mass is given by

$$m^2_{\rm H} = \lambda v^2, \tag{342}$$

and, from the expression of the potential (322), one finds the expression of its value at the minimum

$$V(\langle\Sigma\rangle = v) = -\lambda\frac{v^4}{4} = -\frac{m^4_{\rm H}}{4\lambda}, \tag{343}$$

in agreement with what we had before. Using the expression of the Higgs mass, one can re-write this expression as

$$\rho_{\rm vac} = V(\langle\Sigma\rangle = v) = -\frac{1}{4}m^2_{\rm H}v^2. \tag{344}$$

We see that in order to calculate the vacuum energy, we need to know v and the Higgs mass. Let us first discuss how the vacuum expectation value of the Higgs can be obtained. For this purpose, we focus on the term in the electroweak Lagrangian describing the interaction between leptons and charged gauge bosons (the so-called charged currents). From the above considerations, see Eq. (334), they are given by

$$\mathcal{L}_{\rm intL-W} = \frac{g}{\sqrt{2}}\nu^\dagger_{\rm L}\overline{\sigma}^\mu W^-_\mu e_{\rm L} + \frac{g}{\sqrt{2}}e^\dagger_{\rm L}\overline{\sigma}^\mu W^+_\mu \nu_{\rm L}. \tag{345}$$

It is easy to verify by an explicit calculation that it can be re-expressed in terms of Dirac spinors[126] as

$$\begin{aligned}\mathcal{L}_{\rm intL-W} \;&=\; \frac{g}{\sqrt{2}}\overline{\Psi}_\nu\gamma^\mu\frac{1-\gamma_5}{2}W^-_\mu\Psi_e \\ &+\; \frac{g}{\sqrt{2}}\overline{\Psi}_e\gamma^\mu\frac{1-\gamma_5}{2}W^+_\mu\Psi_\nu,\end{aligned} \tag{347}$$

[126] Let us recall that they are defined by the following expression

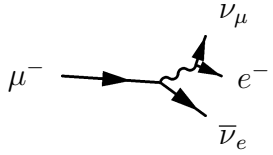
$$\Psi_e = \begin{pmatrix} e_{\rm L} \\ e_{\rm R}\end{pmatrix}, \tag{346}$$

and a similar expression for Ψ_ν.

which is similar to the Fermi theory for the weak interaction described by the following Lagrangian

$$\mathcal{L}_{\text{Fermi}} = \frac{G_{\text{F}}}{\sqrt{2}} J_\lambda J^{\lambda\dagger}, \tag{348}$$

with $J_\lambda = J_\lambda^{(e)} + J_\lambda^{(\mu)}$ (*i.e.* an electronic and a muonic part) where the current is given by $J_\lambda^{(e)} = \overline{\Psi}_e \gamma_\lambda (1-\gamma_5)\Psi_{\nu_e}$ and a similar definition for $J_\lambda^{(\nu)}$. If, for instance we consider the decay of muon into an electron, a muonic neutrino and an anti electronic neutrino, described by the graph



(349)

the two theories will lead to the same leading contribution at small energies if the following identification is made

$$\frac{g^2}{8m_W^2} = \frac{G_{\text{F}}}{\sqrt{2}}. \tag{350}$$

Indeed, the appearance of the gauge boson mass comes from the propagators of W_μ. At small momenta compared to the mass, only m_W remains and the above graph is in fact equivalent to the graph corresponding to the Fermi theory. Then, using the expression of the W-bosons mass established before, $m_W^2 = g^2 v^2/2$, one arrives at

$$v^2 = \frac{\sqrt{2}}{4G_{\text{F}}^2}. \tag{351}$$

Notice that there is a factor $\sqrt{2}$ difference with respect to the textbook [20] because we have defined the Higgs as $v + H/\sqrt{2}$ instead of $(v+H)/\sqrt{2}$. The Fermi constant is given by $G_{\text{F}} \simeq 1.16 \times 10^{-5}\ (\text{GeV})^{-2}$. Therefore,

$$v \simeq 175\,\text{GeV}. \tag{352}$$

or $\sqrt{2}v \simeq 246\,\text{GeV}$. We see that we have been able to calculate the Higgs vacuum expectation value v. As a consequence, Eq. (344) reads

$$\rho_{\text{vac}} = -\frac{\sqrt{2}}{16}\frac{m_{\text{H}}^2}{G_{\text{F}}^2}. \tag{353}$$

The mass of the Higgs boson is not known although, at the time of writing, there are reasons to believe that $m_{\text{H}} < 129\,\text{GeV}$ at 95% CL [46] (there are even reasons to believe that $m_{\text{H}} \simeq 125\,\text{GeV}$ but this is not yet established at a sufficient statistical level at the time of writing). Therefore, we find that

$$\rho_{\text{vac}}^{\text{EW}} \simeq -1.2 \times 10^8\,\text{GeV}^4 \simeq -10^{55}\rho_{\text{cri}}, \tag{354}$$

where $\rho_{\text{cri}} \simeq 10^{-47}\,\text{GeV}^4$ is the critical energy density today. As will be discussed in the following, this result is in fact in contradiction with various observations.

To conclude this section, let us recall that one can always adjust the vacuum energy today to zero by tuning the parameter V_0. This is clearly not a very satisfactory method and this implies that the vacuum energy density was huge prior to the electroweak phase transition. It is also worth noticing that this problem is not specific to the electroweak transition. Exactly the same discussion could be presented for the Quantum Chromo Dynamics (QCD) transition [3, 8] (for instance) and would lead to the same problem since

$$\rho_{\text{vac}}^{\text{QCD}} \simeq 10^{-2}\,\text{GeV}^4 \simeq 10^{45}\rho_{\text{cri}}. \tag{355}$$

Moreover, as we are now going to discuss in detail, even if we "solve" the classical cosmological constant by tuning the vacuum energy to zero, see Eq. (304), the problem reappears at the quantum level. Indeed, we have seen that the zero-point quantum fluctuations also give a contribution to the cosmological constant, see Eq. (301). The goal of the next section is to estimate the corresponding energy density.

24.4 The Quantum-Mechanical Cosmological Constant Problem

We have seen that, even if we solve the classical cosmological constant problem and find a convincing reason to put the minimum of the potential to zero, $\rho_{\rm vac}$ is a quantity which also receives contributions from the zero-point fluctuations of all the quantum fields present in the Universe. We now discuss this "quantum-mechanical" cosmological constant problem. In principle, the corresponding vacuum energy density can be estimated from first principles. The goal of this section (and of Sec. 24.6) is to carry out this task. We will first discuss how the Lorentz invariance affects the calculation of the vacuum energy, see the next subsection 24.4.1. Then, in sub-Sec. 24.4.2, we will show how this question can be formulated in terms of Feynman diagrams. In sub-Sec. 24.4.3, we will also evaluate the vacuum energy not only for a free theory as usually done but also in the more realistic case where interactions are present. Finally, in order to understand better the origin of the vacuum energy density, in sub-Secs. 24.5.1 and 24.5.2, we will address the same question but in the context of ordinary quantum mechanics. This will allow us in sub-Sec. 24.5.3 and 24.5.4 to discuss in more detail the properties of the so-called bubble diagrams from which the cosmological constant originates. As mentioned above, all the fields present in the universe participate in the vacuum energy. Therefore, we need to do the calculation for scalar, fermion and vector fields. Here, in a first step, we carry out the calculation for a real scalar field. The other types will be treated later on in Sec. 24.7. For simplicity, we will also consider that the metric is flat and that the fields live in Minkowski space-time. A sloppy justification is that the main contribution to the vacuum energy density comes from modes with very high momenta, corresponding to scales at which the curvature of space-time is negligible. In fact, we will come back to this very important question in more detail in what follows and will try to discuss this point in a more rigorous way, see Sec. 24.8.

24.4.1 The Zero-Point Energy Density

Let us consider a simple real free scalar field with the potential $V(\Phi) = m^2\Phi^2/2$ where m is the mass of the scalar particle. In flat space-time, the equation of motion is nothing but the Klein-Gordon equation, namely

$$-\ddot{\Phi} + \delta^{ij}\partial_i\partial_j\Phi - m^2\Phi = 0. \tag{356}$$

Since the scalar field is free, the equation of motion is linear and, as a consequence, one can Fourier expand $\Phi(t, \boldsymbol{x})$ as

$$\begin{aligned}\Phi\left(t, \boldsymbol{x}\right) &= \frac{1}{\left(2\pi\right)^{3/2}} \int \frac{\mathrm{d}^3\boldsymbol{k}}{\sqrt{2\omega(k)}} \Big(c_{\boldsymbol{k}} \mathrm{e}^{-i\omega t + i\boldsymbol{k}\cdot\boldsymbol{x}} \\ &\quad + c^\dagger_{\boldsymbol{k}} \mathrm{e}^{i\omega t - i\boldsymbol{k}\cdot\boldsymbol{x}} \Big),\end{aligned} \tag{357}$$

with

$$\omega(k) \equiv \sqrt{k^2 + m^2}. \tag{358}$$

The four-dimensional momentum has been written as $k^\mu = (k^0, \boldsymbol{k})$ and the integration is performed over its spatial part only. We have used the notation $|\boldsymbol{k}| = k$. Its time component is given by $k^0 = \omega$. As usual, the field has been quantized by considering $c_{\boldsymbol{k}}$ and $c^\dagger_{\boldsymbol{k}}$ (the so-called annihilation and creation operators) as quantum operators satisfying the following commutation relation

$$\left[c_{\boldsymbol{k}}, c^\dagger_{\boldsymbol{k}'}\right] = \delta^{(3)}\left(\boldsymbol{k} - \boldsymbol{k}'\right). \tag{359}$$

This equation is in fact equivalent to the standard commutation rule between the field operator and its conjugate momentum.

We are now in a position where one can compute $\langle 0|T_{\mu\nu}|0\rangle$. From Eq. (303), we see that we must first evaluate the mean values of various quantities depending on the field operator and

its derivatives. Straightforward calculations lead to

$$\langle 0|\dot{\Phi}^2|0\rangle = \frac{1}{(2\pi)^3}\int \frac{\mathrm{d}^3\boldsymbol{k}}{2\omega(k)}\,\omega^2(k), \tag{360}$$

$$\langle 0|\delta^{ij}\partial_i\Phi\partial_j\Phi|0\rangle = \frac{1}{(2\pi)^3}\int \frac{\mathrm{d}^3\boldsymbol{k}}{2\omega(k)}\,\boldsymbol{k}^2, \tag{361}$$

$$\langle 0|\Phi^2|0\rangle = \frac{1}{(2\pi)^3}\int \frac{\mathrm{d}^3\boldsymbol{k}}{2\omega(k)}. \tag{362}$$

These equations can then be used to determine the energy density and pressure of the vacuum. From the above expression (303) of the stress-energy tensor, one has

$$T_{00} = \mathcal{H} = \frac{1}{2}\dot{\Phi}^2 + \frac{1}{2}\delta^{ij}\partial_i\Phi\partial_j\Phi + \frac{1}{2}m^2\Phi^2, \tag{363}$$

where $\mathcal{H}$ denotes the Hamiltonian density. This leads to the following expression for the energy density $\rho = u^\mu u^\nu T_{\mu\nu}$ measured by a fundamental observer characterized by its four-velocity vector $u^\mu = (1, \boldsymbol{0})$

$$\langle\rho\rangle = \langle 0|u^0u^0T_{00}|0\rangle = \frac{1}{(2\pi)^3}\frac{1}{2}\int \mathrm{d}^3\boldsymbol{k}\,\omega(k). \tag{364}$$

In the same manner, one can easily estimate the pressure. One obtains

$$\begin{aligned}\langle p\rangle &= \left\langle 0\left|\frac{1}{3}\perp^{\mu\nu}T_{\mu\nu}\right|0\right\rangle \\ &= \left\langle 0\left|\frac{1}{2}\dot{\Phi}^2 - \frac{1}{6}\delta^{ij}\partial_i\Phi\partial_j\Phi - \frac{1}{2}m^2\Phi^2\right|0\right\rangle,\end{aligned} \tag{365}$$

where $\perp^{\mu\nu}\equiv g^{\mu\nu} + u^\mu u^\nu$ is a projector. Upon using the expressions (360), (361) and (362), one arrives at the following expression

$$\langle p\rangle = \frac{1}{(2\pi)^3}\frac{1}{6}\int \mathrm{d}^3\boldsymbol{k}\,\frac{k^2}{\omega(k)}. \tag{366}$$

Notice that we also have

$$\langle 0|T_{0i}|0\rangle = -\frac{1}{(2\pi)^3}\frac{1}{2}\int \mathrm{d}^3\boldsymbol{k}\,k_i = 0, \tag{367}$$

and, hence, we do not need to consider to off-diagonal elements. This completes our calculation of the vacuum stress-energy tensor.

From the above considerations, using Eq. (307), we deduce that

$$\Lambda_{\mathrm{eff}} = \Lambda_{\mathrm{B}} + \frac{\kappa}{(2\pi)^3}\int \mathrm{d}^3\boldsymbol{k}\,\frac{1}{2}\omega(k). \tag{368}$$

The obvious problem with the previous calculation is that the integrals expressing the energy density $\langle\rho\rangle$ and the pressure $\langle p\rangle$ blow up in the ultra-violet regime (and, therefore, strictly speaking, the effective cosmological constant is in fact infinite). This is of course a well-known problem in quantum field theory. Usually, this type of divergences is ignored (or sometimes removed by mean of the normal ordering product) because, in absence of gravity, only differences of energy can be detected [16, 17]. However, in the present context, this is clearly no longer the case since the absolute value of the vacuum energy weighs and, therefore, can be measured. This is why regularizing these infinities becomes a crucial issue that we now discuss in some details.

The common method found in the literature is to introduce a cut-off at $k = M$, where the physical interpretation of M is that this is the scale at which the effective theory used before breaks down [451, 4, 5, 449, 1393]. The quantity M is not known: it could be the Planck scale,

the string scale or even the super-symmetric breaking scale. As we will see later on, this is in fact not important because the cosmological constant problem is present regardless of the precise value of the cut-off. In this approach, the energy density becomes

$$\langle\rho\rangle = \frac{1}{4\pi^2}\int_0^M \mathrm{d}k k^2\sqrt{k^2+m^2} \tag{369}$$

$$= \frac{M^4}{16\pi^2}\left[\sqrt{1+\frac{m^2}{M^2}}\left(1+\frac{1}{2}\frac{m^2}{M^2}\right) - \frac{1}{2}\frac{m^4}{M^4}\ln\left(\frac{M}{m}+\frac{M}{m}\sqrt{1+\frac{m^2}{M^2}}\right)\right] \tag{370}$$

$$= \frac{M^4}{16\pi^2}\left(1+\frac{m^2}{M^2}+\cdots\right), \tag{371}$$

where, in the last expression, we have expanded the exact expression in terms of the small parameter m/M. We see that the divergence is quartic in the cut-off scale. Let us now perform the same calculation for the pressure. Using Eq. (366), one obtains

$$\langle p\rangle = \frac{1}{3}\frac{1}{4\pi^2}\int_0^M \mathrm{d}k \frac{k^4}{\sqrt{k^2+m^2}} \tag{372}$$

$$= \frac{1}{3}\frac{M^4}{16\pi^2}\left[\sqrt{1+\frac{m^2}{M^2}}\left(1-\frac{3}{2}\frac{m^2}{M^2}\right) + \frac{3}{2}\frac{m^4}{M^4}\ln\left(\frac{M}{m}+\frac{M}{m}\sqrt{1+\frac{m^2}{M^2}}\right)\right] \tag{373}$$

$$= \frac{1}{3}\frac{M^4}{16\pi^2}\left(1-\frac{m^2}{M^2}+\cdots\right). \tag{374}$$

It is clear from the previous expressions that $\langle p\rangle/\langle\rho\rangle \neq -1$ which indicates that the stress energy tensor is not of the form $\propto -\rho g_{\mu\nu}$. In the limit $m \to 0$, as can be easily shown from Eqs. (371) and (374), the equation of state is in fact $\langle p\rangle/\langle\rho\rangle = 1/3$. This would mean that the zero point fluctuations do not behave like a cosmological constant but rather like radiation. On the other hand, this result seems to be strange since we know from fundamental considerations that the stress-energy tensor of a system in its ground state must be proportional to the metric tensor. Clearly, something has gone wrong is our calculation.

Interestingly enough, let us also note in passing that if one only focuses on the logarithmic terms, namely

$$\langle\rho\rangle^{\log} = -\frac{m^4}{32\pi^2}\ln\left(\frac{M}{m}+\frac{M}{m}\sqrt{1+\frac{m^2}{M^2}}\right), \tag{375}$$

$$\langle p\rangle^{\log} = \frac{m^4}{32\pi^2}\ln\left(\frac{M}{m}+\frac{M}{m}\sqrt{1+\frac{m^2}{M^2}}\right), \tag{376}$$

then the equation of state of the vacuum is correctly reproduced (even if the energy density can be negative). In fact, as we explain below, this is not a coincidence [47].

It is easy to check that the previous considerations are not an artifact of the fact that we have introduced a sharp cut-off. Indeed, if rather than a sharp cut-off in the upper limit of the integrals, one introduces a smooth exponential cut-off, one obtains (for simplicity, one only

considers the case $m = 0$),

$$\langle\rho\rangle = \frac{1}{4\pi^2}\int_0^\infty \mathrm{d}k\, k^3\, \mathrm{e}^{-\alpha k} = \frac{1}{4\pi^2}\frac{\Gamma(4)}{\alpha^4} \tag{377}$$

$$= \frac{1}{4\pi^2} 6M^4, \tag{378}$$

where we have written $\alpha = 1/M$ and where $\Gamma(z)$ is the Euler's integral of the second kind [48, 49]. One recovers our familiar quartic divergence. Doing the same manipulation for the pressure, one obtains again that $\langle p\rangle/\langle\rho\rangle = 1/3$.

In fact, what has gone wrong is that we have used schemes of regularization that do not respect the symmetry of our underlying theory [47, 50, 51]. Indeed, the Lorentz invariance is broken when we impose a cut-off on the spatial momentum only. As exemplified by Eq. (378), it is easy to show that this is a feature of any regularization scheme breaking Lorentz invariance and not a specific property of the particular cut-off scheme used above (and usually used in the literature). In fact, this property is well-known in the context of gauge theories. If one regulates divergent graphs with a method that is not gauge invariant, then one obtains incorrect results [52]. Here, the situation is very similar.

Therefore, in order to meaningfully evaluate the zero-point energy density, one must use a regularization scheme that does not break Lorentz invariance [47, 50, 51]. An obvious choice is dimensional regularization and we now explore this route. In order to use this method, we first reformulate our problem in a d dimensional Minkowski space-time. Obviously, the Klein-Gordon equation (356) remains unchanged except that the indices of the Krönecker symbol now run from 0 to $d-1$. The Fourier expansion of the field becomes [compare with Eq. (357)]

$$\Phi(t,\boldsymbol{x}) = \frac{1}{(2\pi)^{(d-1)/2}}\int \frac{\mathrm{d}^{d-1}\boldsymbol{k}}{\sqrt{2\omega(k)}}\Big(c_{\boldsymbol{k}}\mathrm{e}^{-i\omega t+i\boldsymbol{k}\cdot\boldsymbol{x}} + c^\dagger_{\boldsymbol{k}}\mathrm{e}^{i\omega t-i\boldsymbol{k}\cdot\boldsymbol{x}}\Big), \tag{379}$$

and, as a consequence, the energy density can be expressed as

$$\langle\rho\rangle = \frac{\mu^{4-d}}{(2\pi)^{(d-1)}}\frac{1}{2}\int \mathrm{d}^{d-1}\boldsymbol{k}\,\omega(k) \tag{380}$$

$$= \frac{\mu^{4-d}}{(2\pi)^{(d-1)}}\frac{1}{2}\int_0^\infty \mathrm{d}k k^{d-2}\mathrm{d}^{d-2}\Omega\,\omega(k), \tag{381}$$

where we have introduced a scale μ in order for the equation to be dimensionally correct. This equation should be compared with Eq. (364). Then, using the fact that the angular integrals can be easily performed, namely $\int \mathrm{d}^{d-2}\Omega = 2\pi^{(d-1)/2}/\Gamma(d/2-1/2)$, where, again, $\Gamma(z)$ is the Euler's integral of the second kind [48, 49], one obtains

$$\langle\rho\rangle = \frac{\mu^4}{2\,(4\pi)^{(d-1)/2}}\frac{\Gamma(-d/2)}{\Gamma(-1/2)}\left(\frac{m}{\mu}\right)^d. \tag{382}$$

Performing the same calculation for the pressure, one arrives at a similar expression

$$\langle p\rangle = \frac{\mu^{4-d}}{(2\pi)^{(d-1)}}\frac{1}{2(d-1)}\int \mathrm{d}^{d-1}\boldsymbol{k}\,\frac{k^2}{\omega(k)} \tag{383}$$

$$= \frac{\mu^4}{4\,(4\pi)^{(d-1)/2}}\frac{\Gamma(-d/2)}{\Gamma(1/2)}\left(\frac{m}{\mu}\right)^d. \tag{384}$$

But $\Gamma(-1/2) = -2\Gamma(1/2)$ and, as a consequence,

$$\langle p\rangle = -\langle\rho\rangle. \tag{385}$$

This time, as expected, we have obtained the equation of state of the vacuum. The fact that the regularization method used satisfies the Lorentz symmetry has led us to the correct result.

The consistency of the above result can be checked in a different manner [47]. Using Eqs. (380) and (383) for the energy density and the pressure, one sees that they satisfy the following differential equation

$$\langle\rho\rangle - 2m^2\frac{\mathrm{d}\langle\rho\rangle}{\mathrm{d}m^2} = (d-1)\langle p\rangle. \tag{386}$$

Using the fact that $\langle p\rangle = -\langle\rho\rangle$ one obtains

$$m^2\frac{\mathrm{d}\langle\rho\rangle}{\mathrm{d}m^2} = \frac{d}{2}\langle\rho\rangle, \tag{387}$$

which can be integrated to give $\langle\rho\rangle \propto m^d$ in full agreement with Eq. (382). Therefore, we conclude that the above considerations are all consistent.

Let us now analyze in more details the structure of the divergences. For this purpose one writes $d = 4-\epsilon$, where ϵ is supposed to be a small parameter, and we expand $\langle\rho\rangle$ in terms of ϵ. The Euler function is such that

$$\Gamma\left(-2+\frac{\epsilon}{2}\right) = \frac{1}{-2+\epsilon/2}\frac{1}{-1+\epsilon/2}\frac{1}{\epsilon/2}\Gamma\left(1+\frac{\epsilon}{2}\right), \tag{388}$$

with $\Gamma(1+\epsilon/2) \simeq \Gamma(1)+\epsilon\Gamma'(1)/2$, $\Gamma'(1) = \Psi(1)\Gamma(1)$ where Ψ is the Polygamma function [48, 49] and $\Psi(1) = -\gamma \simeq 0.57772$ the Euler-Mascheroni constant [48, 49]. If, in addition, we use the two following expansions

$$(4\pi)^{-3/2+\epsilon/2} \simeq \frac{1}{(4\pi)^{3/2}}\left[1+\frac{\epsilon}{2}\ln(4\pi)\right], \tag{389}$$

$$\left(\frac{m}{\mu}\right)^{4-\epsilon} \simeq \left(\frac{m}{\mu}\right)^4\left(1-\epsilon\frac{m}{\mu}\right), \tag{390}$$

then one arrives at the following expression for the energy density of the vacuum

$$\langle\rho\rangle \simeq -\frac{m^4}{64\pi^2}\left[\frac{2}{\epsilon}+\frac{3}{2}-\gamma-\ln\left(\frac{m^2}{4\pi\mu^2}\right)\right]+\cdots. \tag{391}$$

Using a $\overline{\mathrm{MS}}$ renormalization scheme convention [$\overline{\mathrm{MS}}$ means "modified minimal subtraction" and simply consists in subtracting the pole $\sim 1/\epsilon$ together with the accompanying terms γ and $\ln(4\pi)$, see chapter 11 of Ref. [16]; in fact, here, we slightly modify the convention such that the term 3/2 is subtracted as well], the regularized, finite, energy density of the vacuum now reads

$$\langle\rho\rangle = \frac{m^4}{64\pi^2}\ln\left(\frac{m^2}{\mu^2}\right). \tag{392}$$

This result is very different from the result obtained by imposing a sharp cut-off. In fact, one recovers the structures of the logarithm terms found in Eqs. (375) and (376) since it is well-known that the dimensional regularization scheme removes the power law terms. As discussed around Eqs. (375) and (376), this is the reason why those terms lead to the correct vacuum equation of state. We see that the result is not proportional to the cut-off scale to the power four, as usually claimed on dimensional grounds, but to the mass of the particle to the power four, which is also dimensionally correct. Physically, for instance, this means that the photon does not contribute to the vacuum energy density contrary to what a sharp cut-off calculation would predict. Therefore, the order of magnitude of the final result can a priori be very different. We also notice that the sign of $\langle\rho\rangle$ can change according to $m > \mu$ or $m < \mu$. Of course, it remains to be seen whether this could solve the cosmological constant problem and we will argue below that, unfortunately, the answer to this question is no. It should also be kept in mind that the above calculation is valid for free fields. It is clear that a more realistic version should take into account the interactions and we will also consider this issue in the following. However, before addressing these various points and in order to gain further physical insights, we would like to investigate again the above issues, this time from the Feynman diagrams point of view. We turn to this question in the next sub-section.

24.4.2 The Vacuum and the Feynman Diagrams

We have seen in the last sub-section that the vacuum contribution to the cosmological constant is given in terms of divergent integrals. A priori, there is no reason to be surprised since divergent integrals are routinely found in quantum field theory. We have the method of renormalization at our disposal to get rid of these infinities and the so accurate predictions of particle physics are here to remind us that this is a very convincing way of taming these divergences. However, we will see that the divergent integrals that we encounter when we calculate the vacuum energy are in fact of a different nature than the divergent integrals met when we calculate, say, a cross section. In order to discuss this point in more detail and to understand better the structure of these divergences, we now briefly return to some textbook considerations [16, 17, 18, 19, 20, 21, 22].

In quantum field theory, the basic object is the generating functional defined by the following path integral (we still restrict our considerations to a simple scalar field)

$$Z\left[J\right]=\mathcal{N}\int\mathcal{D}\Phi\exp\left\{i\int\mathrm{d}^{4}x\left[\mathcal{L}+J(x)\Phi(x)\right]\right\}, \tag{393}$$

where $\mathcal{N}$ is a number chosen such that $Z[J=0]=1$. The quantity $\mathcal{L}$ is the Lagrangian density while J represents a source coupled to the scalar field. From the generating functional, one can compute the various correlation functions

$$\begin{aligned} G^{(n)}(x_1\cdots x_n) &= \left\langle\Omega\left|T\left[\Phi(x_1)\cdots\Phi(x_n)\right]\right|\Omega\right\rangle & (394)\\ &= \left(\frac{1}{i}\right)^{n}\frac{\delta^{n}Z\left[J\right]}{\delta J(x_1)\cdots\delta J(x_n)}\bigg|_{J=0}, & (395)\end{aligned}$$

where $|\Omega\rangle$ denotes the vacuum state of the theory. We distinguish it from $|0\rangle$ which denotes the vacuum state of the theory without interaction. In this last case, the generating functional can be computed exactly. It reads (notice that we have $Z_0[0]=1$ and that the "0" subscript indicates that we refer to the free theory)

$$Z_0[J]=\exp\left[-\frac{1}{2}\int\mathrm{d}^{4}x\mathrm{d}^{4}yJ(x)D_{\mathrm{F}}(x-y)J(y)\right], \tag{396}$$

where D_{F} is the Feynman propagator defined by the following expression

$$D_{\mathrm{F}}\left(x_1-x_2\right)=\left\langle\Omega\left|T\left[\Phi(x_1)\Phi(x_2)\right]\right|\Omega\right\rangle, \tag{397}$$

T denoting the time ordered product. As is well-known, the propagator can be conveniently expressed in Fourier space as

$$D_{\mathrm{F}}\left(x_1-x_2\right)=\frac{i}{(2\pi)^{4}}\int\frac{\mathrm{d}^{4}k}{k^{2}+m^{2}}\mathrm{e}^{ik_{\mu}\cdot\left(x_1^{\mu}-x_2^{\mu}\right)}. \tag{398}$$

In presence of interactions, one can write $\mathcal{L}=\mathcal{L}_0+\mathcal{L}_{\mathrm{int}}$ where $\mathcal{L}_{\mathrm{int}}$ represents the self-interacting part of the Lagrangian density. In this case the generating functional can be written as

$$Z[J]=\mathcal{N}\exp\left\{i\int\mathrm{d}^{4}x\,\mathcal{L}_{\mathrm{int}}\left[\frac{1}{i}\frac{\delta}{\delta J(x)}\right]\right\}Z_0[J], \tag{399}$$

where, in order to have the correct normalization, the factor $\mathcal{N}$ must be given by

$$\mathcal{N}^{-1}\equiv\exp\left\{i\int\mathrm{d}^{4}x\,\mathcal{L}_{\mathrm{int}}\left[\frac{1}{i}\frac{\delta}{\delta J(x)}\right]\right\}Z_0[J]\bigg|_{J=0}. \tag{400}$$

As explained in the standard quantum field theory textbooks, the above expressions (399) and (400) are particularly well suited for a perturbative expansion. For instance, if we want to

calculate the generating functional for the case of a quartic theory $\mathcal{L}_{\rm int} = \lambda\Phi^4/4!$, where λ is the coupling constant, we are led to the following expansion in λ

$$Z[J] = \mathcal{N}\left[1 - i\frac{\lambda}{4!}\int \mathrm{d}^4x \frac{1}{i^4}\frac{\delta^4 Z_0[J]}{\delta J(x)^4} + \cdots\right]. \tag{401}$$

As usual, the various terms that appear in this expansion can be graphically represented by Feynman diagrams. The Feynman rules of the theory in real space are given by

$$D(x_1 - x_2) \equiv x_1 \bullet\!\!-\!\!-\!\!\bullet\ x_2, \tag{402}$$

$$-i\lambda\int \mathrm{d}^4x \equiv \quad , \tag{403}$$

$$i\int \mathrm{d}^4x J(x) \equiv \circ\!\!-\!\!-\ , \tag{404}$$

and a loop for a propagator evaluated at the same point in space-time. This completes our quick reminder of the basics of field theory that we need in the following.

Now, we compute the energy density of a free scalar field (the interacting case is treated in the next subsection) using the tools that we have just presented. In fact, it is interesting to calculate the trace of the stress-energy tensor. Upon using Eq. (303), one obtains

$$\langle T\rangle = \langle \eta^{\mu\nu}T_{\mu\nu}\rangle = -\langle\rho\rangle + (d-1)\langle p\rangle. \tag{405}$$

Then, since the energy density and the pressure are given by Eqs. (380) and (383), the above relation can be re-expressed as

$$\langle T\rangle = -\mu^{4-d}\int \frac{\mathrm{d}^{d-1}\boldsymbol{k}}{(2\pi)^{d-1}}\frac{m^2}{2\omega(k)} \tag{406}$$

But, from Eqs. (397) and (398), and now specifying to the case $d = 4$, one has

$$D_{\rm F}(0) = \frac{i}{(2\pi)^4}\int \frac{\mathrm{d}^4k}{k^2+m^2} = \frac{i}{(2\pi)^4}\int \frac{\mathrm{d}k^0\mathrm{d}^3\boldsymbol{k}}{-(k^0)^2+\omega^2}, \tag{407}$$

where we have distinguished the integration over the time and space components of the momentum. Using the identity $\int \mathrm{d}k^0/[-(k^0)^2+\omega^2] = -\pi/(i\omega)$, see also Eq. (451), this leads to

$$D_{\rm F}(0) = -\int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3 2\omega}. \tag{408}$$

or, in other words,

$$\langle T\rangle = m^2 D_{\rm F}(0). \tag{409}$$

But, if we now use the fact that $\langle p\rangle = -\langle\rho\rangle$, Eq. (405) can also be written as $\langle T\rangle = -d\langle\rho\rangle$. As a consequence, Eq. (409) implies that

$$\langle\rho\rangle = -\frac{m^2}{4}D_{\rm F}(0). \tag{410}$$

This equation can also be obtained by a direct calculation. Indeed one has

$$D_{\rm F}(0) = \frac{i\mu^{4-d}}{(2\pi)^d}\int \frac{\mathrm{d}k^0\mathrm{d}^{d-1}\boldsymbol{k}}{-(k^0)^2+\omega^2} \tag{411}$$

$$= -\frac{d}{m^2}\frac{\mu^4}{2\,(4\pi)^{(d-1)/2}}\frac{\Gamma(-d/2)}{\Gamma(-1/2)}\left(\frac{m}{\mu}\right)^d \tag{412}$$

$$= -\frac{d}{m^2}\langle\rho\rangle, \tag{413}$$

where, in the second step of the calculation, we have used dimensional regularization. Now, using Eq. (307) and the Feynman rule stipulating that a propagator evaluated at the same point is represented by a loop, one can re-write the cosmological constant as

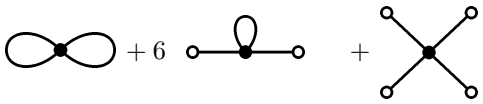
$$\Lambda_{\rm eff} = \Lambda_{\rm B} - \kappa \frac{m^2}{4} \quad . \tag{414}$$

In other words, the cosmological constant is given by a "bubble diagram", *i.e.* a diagram that has no external leg.

We have just established that the vacuum energy density, or equivalently the cosmological constant, can be expressed in term of a bubble diagram. This type of diagram plays a special role in quantum field theory and we will discuss this point in more detail in the following next sub-sections. In order to develop our intuition, let us first calculate explicitly the generating function in the case of a self-interacting scalar field. The calculation is straightforward and upon using Eq. (396), this leads to

$$\begin{aligned} -i\frac{\lambda}{4!}\int \mathrm{d}^4x \frac{1}{i^4}\frac{\delta^4 Z_0[J]}{\delta J(x)^4} &= -i\frac{\lambda}{4!}\Bigg\{3\int \mathrm{d}^4x D^2(0) - 6\int \mathrm{d}^4x D(0)\left[\int \mathrm{d}^4y D(x-y)J(y)\right]^2 \\ &\quad + \int \mathrm{d}^4x \left[\int \mathrm{d}^4y D(x-y)J(y)\right]^4\Bigg\} Z_0[J] + \cdots \end{aligned} \tag{415}$$

$$= \frac{1}{4!}\Bigg(3 \quad + 6 \quad + \quad + \cdots \Bigg) Z_0[J]. \tag{416}$$

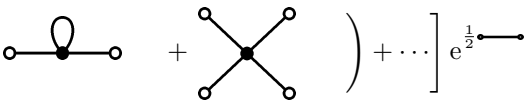


One notices the appearance of a "bubble diagram" (the first one), similar (*i.e.* with no external leg) but not identical to the one encountered in the calculation of the cosmological constant. We also have a diagram with a loop (the second one) but with external legs and a diagram with no loop (the last one). All the diagram with loops (that is to say the two first ones) are divergent in quantum field theory. Nevertheless, the bubble diagram and the loop diagram play in fact a very different role as we are going to see. Let us now calculate the normalization $\mathcal{N}$, see Eq. (400). It uses the same expression as before but with $J = 0$ which means that all the diagram containing a source disappear. As a consequence, we have

$$\mathcal{N}^{-1} = 1 + \frac{3}{4!} \tag{417}$$

We can now gather everything and compute the generating functional. Upon using the expression (401) and the results (416) and (417), one arrives at

$$Z[J] = \Bigg[1 + \frac{1}{4!}\Bigg(6 \quad + \quad \Bigg) + \cdots\Bigg] \mathrm{e}^{\frac{1}{2}} \quad . \tag{418}$$



We see that the bubble diagram has canceled in the final expression. Since all the predictions (cross-sections etc...) follow from the calculation of the generating functional, we see that the bubble diagrams never contribute, in absence of gravity, to any observables. On the contrary, when the gravitational field is turned on, the bubble diagrams become important and can affect the predictions of the theory. In summary, the cosmological constant problem can be viewed as the appearance of a new type of divergent graphs. As demonstrated with Eq. (392), these graphs can be regularized but, as will be discussed below, the regularized value seems to be incompatible with some astrophysical observations.

24.4.3 The Vacuum in Presence of Interactions

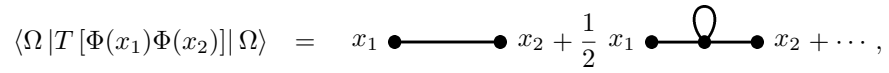
So far we have calculated the vacuum energy density for a free field. It is natural to ask what happens when one considers interaction [50] since we know that a realistic model necessarily contains such terms. For this purpose, let us therefore consider again a simple self-interacting

scalar field with $\mathcal{L}_{\rm int} = \lambda\Phi^4/4!$. As a consequence, this adds the following contribution to the energy density of the vacuum

$$\Delta\rho = \frac{\lambda}{4!}\left\langle\Phi^4\right\rangle = \frac{3\lambda}{4!}\left\langle\Phi^2\right\rangle^2 = \frac{\lambda}{8}D_{\rm F}^2(0) \tag{419}$$

$$= \frac{i}{8\int \mathrm{d}^4x} \quad \text{ } \tag{420}$$

where, on the first line in the above expression, we have used the fact that the vacuum wave-functional is a Gaussian. As expected, this new term is also given by a bubble diagram. However, one should not forget that, when interaction is present, one must also renormalize the mass. As is well-known, at one loop, the renormalization of the mass comes from the tadpole diagram. More precisely, one has to calculate the correction to the propagator. It is given by the following expression

$$\langle\Omega\,|T\left[\Phi(x_1)\Phi(x_2)\right]|\,\Omega\rangle = \quad \text{  } \tag{421}$$

or, in terms of explicit expressions [see Eq. (398)]

$$\langle\Omega\,|T\left[\Phi(x_1)\Phi(x_2)\right]|\,\Omega\rangle = D_{\rm F}(x_1-x_2) - \frac{i\lambda}{2}\int \mathrm{d}^4x D_{\rm F}(x_1-x)D_{\rm F}(0)D_{\rm F}(x-x_2) + \cdots \tag{422}$$

$$= \frac{i}{(2\pi)^4}\int\frac{\mathrm{d}^4p}{p^2+m^2}\mathrm{e}^{ip_\mu\cdot\left(x_1^\mu-x_2^\mu\right)} + \frac{i\lambda}{2}D_{\rm F}(0)\frac{1}{(2\pi)^4}\int\mathrm{d}^4p\frac{\mathrm{e}^{ip_\mu\cdot\left(x_1^\mu-x_2^\mu\right)}}{(p^2+m^2)^2} + \cdots \tag{423}$$

$$= \frac{i}{(2\pi)^4}\int\frac{\mathrm{d}^4p}{p^2+m^2}\mathrm{e}^{ip_\mu\cdot\left(x_1^\mu-x_2^\mu\right)}\left[1+\frac{\lambda}{2}D_{\rm F}(0)\frac{1}{p^2+m^2}\right] + \cdots \tag{424}$$

$$= \frac{i}{(2\pi)^4}\int\frac{\mathrm{d}^4p}{p^2+m^2-\lambda D_{\rm F}(0)/2}\mathrm{e}^{ip_\mu\cdot\left(x_1^\mu-x_2^\mu\right)} + \cdots, \tag{425}$$

where the last equation is equivalent to the previous one at leading order in the coupling constant. We see that this implies the following redefinition (or renormalization) of the mass

$$m_{\rm ren}^2 = m^2 - \frac{\lambda}{2}D_{\rm F}(0) \tag{426}$$

As a consequence, the vacuum energy which is now the sum of the free contribution (410) and of the contribution (419), which originates from the self-interaction of the field, can be re-expressed as

$$\begin{aligned}\langle\rho\rangle &= -\frac{m^2}{4}D_{\rm F}(0) + \frac{\lambda}{8}D_{\rm F}^2(0)\\ &= -\frac{m_{\rm ren}^2}{4}D_{\rm F}(0) - \frac{1}{4}\left[\frac{\lambda}{2}D_{\rm F}(0)\right]D_{\rm F}(0) + \frac{\lambda}{8}D_{\rm F}^2(0)\\ &= -\frac{m_{\rm ren}^2}{4}D_{\rm F}(0),\end{aligned} \tag{427}$$

where, in the second line, we have used Eq. (426). The extra contribution coming from the interacting term is exactly canceled by the renormalization of the mass [53, 50]. In other words, one sees that the presence of the interaction has not modified the expression of the vacuum energy density (410), at least at one loop, provided the final result is expressed in terms of the renormalized mass rather than the bare one.

24.5 The Bubble Diagrams

The previous sections have shown the central role played by the bubble diagrams in the vacuum energy problem. These diagrams do not affect the physical predictions when the gravitational field is turned off and this is why, in standard quantum field theory, there are usually ignored. In this section, we wish to discuss in more detail this type of diagram. For this purpose, we will analyze them in ordinary quantum mechanics where the technical aspects are less complicated and where the interpretation is easier [54, 55, 56].

24.5.1 Quantum Mechanics as a Field Theory

Let us start with the simplest quantum-mechanical system, namely the one-dimensional harmonic oscillator. Let us denote by $x(t)$ the position of the particle evolving into a parabolic potential. Then, the corresponding Lagrangian is given by

$$L(\dot{x}, x) = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}m\omega^2 x^2 \,, \tag{428}$$

where m is the mass of the particle and ω is the effective frequency. Then, let us define the "scalar field" $\Phi(t)$ by

$$\Phi(t) \equiv \sqrt{m}x(t) \,, \tag{429}$$

which implies that Φ has dimension $-1/2$. As a consequence, the Lagrangian now reads

$$L = \frac{\dot{\Phi}^2}{2} - \frac{\omega^2}{2}\Phi^2 \,. \tag{430}$$

This Lagrangian is similar to the one one encounters in field theory for a free scalar field. This means that all the techniques of quantum field theory can in fact be used in ordinary quantum mechanics. Of course, this does not lead to genuine new results since both approaches are equivalent but this can shed a new light of some physical results. This is the strategy that we use here applied to the vacuum energy problem. We now quickly remind how the formalism of quantum field theory can be implemented in quantum mechanics. We will also compare this approach to the standard approach of quantum mechanics. The first step is to define the conjugate momentum. It is given by

$$\Pi(t) = \frac{\partial L}{\partial \dot{\Phi}} = \dot{\Phi}(t) \,. \tag{431}$$

and this leads to the following Hamiltonian

$$H = \frac{\Pi^2}{2} + \frac{\omega^2}{2}\Phi^2 \,. \tag{432}$$

The next step is to introduce the creation and annihilation $c^\dagger$ and c operators. They are defined by the expressions

$$\Phi = \frac{1}{\sqrt{2\omega}}\left(c^\dagger + c\right) \,, \tag{433}$$

$$\Pi = i\sqrt{\frac{\omega}{2}}\left(c^\dagger - c\right) \,. \tag{434}$$

As usual one can check that $[\Phi, \Pi] = i$ implies $\left[c, c^\dagger\right] = 1$. As it is standard in field theory, we now work in the Heisenberg picture where the operators are time dependent. A straightforward calculation shows that the Hamiltonian can be expressed as

$$H = \frac{\omega}{2}\left(cc^\dagger + c^\dagger c\right) \,. \tag{435}$$

This allows us to calculate the time evolution of the operator c. Indeed, the Heisenberg equation reads

$$i\frac{\mathrm{d}c}{\mathrm{d}t} = [c, H] = \omega c \,, \tag{436}$$

from which we deduce that

$$c(t) = c(0)\mathrm{e}^{-i\omega t} = c_0\mathrm{e}^{-i\omega t} \,. \tag{437}$$

As a consequence, the field operator can be written in terms of the creation and annihilation operators in a way which is very similar to Eq. (357), namely

$$\Phi(t) = \frac{1}{\sqrt{2\omega}}\left(c_0\mathrm{e}^{-i\omega t} + c_0^\dagger\mathrm{e}^{i\omega t}\right) \,. \tag{438}$$

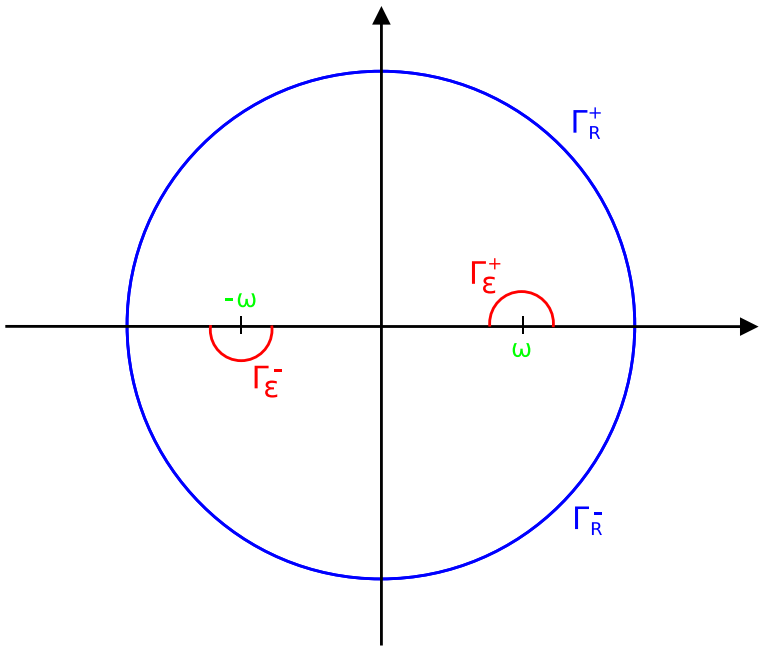


Figure 155: Contours in the complex time plane used to calculate the Feynman propagators of the one-dimensional harmonic oscillator. As usual, there are two poles on the real axis that can be avoided by going along the two small red circles. The final contour is closed by a large blue circle of radius R.

Of course we do not have an integral over the Fourier modes because, in quantum mechanics, we deal with a single harmonic oscillator as opposed to an infinite collection of harmonic oscillators in field theory.

At this level, the theory is known exactly. It is clear that the most interesting part is in fact the treatment of the interactions which, in field theory, is based on the calculation of the Feynman diagrams. Since this calculation is itself based on the calculation of the propagator, one must now evaluate it for our simple system. It is defined by an equation similar to Eq. (397), namely

$$D(t_1,t_2) \equiv \langle 0|T\left[\Phi(t_1)\Phi(t_2)\right]|0\rangle\,, \tag{439}$$

where we recall that the time ordered product is defined by

$$T\left[\Phi(t_1)\Phi(t_2)\right] = \begin{cases}\Phi(t_1)\Phi(t_2), & t_1 > t_2\\ \Phi(t_2)\Phi(t_1), & t_2 > t_1.\end{cases}$$

Then, the calculation proceeds straightforwardly and one obtains for $t_1 > t_2$

$$\begin{aligned} D(t_1,t_2) &= \left\langle 0\left|\frac{1}{\sqrt{2\omega}}\left(c_0 e^{-i\omega t_1} + c_0^\dagger e^{i\omega t_1}\right)\right.\right. \\ &\quad \left.\left.\frac{1}{\sqrt{2\omega}}\left(c_0 e^{-i\omega t_2} + c_0^\dagger e^{i\omega t_2}\right)\right|0\right\rangle \qquad (440)\\ &= \frac{1}{2\omega}e^{-i\omega(t_1-t_2)}\,, \qquad (441)\end{aligned}$$

and, in a similar way if $t_2 > t_1$,

$$\begin{aligned} D(t_1,t_2) &= \left\langle 0\left|\frac{1}{\sqrt{2\omega}}\left(c_0 e^{-i\omega t_2} + c_0^\dagger e^{i\omega t_2}\right)\right.\right. \\ &\quad \left.\left.\frac{1}{\sqrt{2\omega}}\left(c_0 e^{-i\omega t_1} + c_0^\dagger e^{i\omega t_1}\right)\right|0\right\rangle \qquad (442)\\ &= \frac{1}{2\omega}e^{i\omega(t_1-t_2)}\,. \qquad (443)\end{aligned}$$

The two previous equations can be summarized into a single one

$$D(t_1,t_2)=\frac{1}{2\omega}\mathrm{e}^{-i\omega|t_2-t_1|}\,. \tag{444}$$

We see that the system is so simple that we obtain an explicit expression of the propagator in real space (in fact, in field theory, there is also an expression of the propagator in real space but it is not so often used because of its complexity – it is given in terms of ordinary and modified Bessel functions –). Nevertheless, in order to compare with the standard approach and Eq. (398), it is interesting to obtain the expression of the propagator in Fourier space. The Fourier transform of a function $f(t)$ is defined by

$$f(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}f(E)\,\mathrm{e}^{iEt}\,\mathrm{d}E. \tag{445}$$

The calculation is easy and well-known. Here, for convenience, we quickly remind how it proceeds. For that purpose, let us consider the function

$$f(z)\equiv\frac{\mathrm{e}^{-izt}}{z^2-\omega^2}\,. \tag{446}$$

It has two poles at $z=\pm\omega$ on the real axis, see Fig. 155. We assume that $t>0$ and consider the following contour: along the real axis with two small circles $\Gamma_\epsilon^\pm$ around the two poles (the circle around $z=-\omega$ goes into the lower part of the complex plane while the circle around $z=+\omega$ goes into the upper part); and we complete by a big circle Γ_R^- going into the lower part of the complex plane. In this case, only the pole $z=+\omega$ is inside the contour $\Gamma=\mathbb{R}\cup\Gamma_\epsilon^\pm\cup\Gamma_R^-$, see Fig. 155. We have

$$\lim_{|z|\to\infty}|zf(z)|=\lim_{|z|\to\infty}\frac{|z|\mathrm{e}^{\Im(z)t}}{z^2-\omega^2}=0\,, \tag{447}$$

since $\Im(z)<0$ in the lower part of the complex plan. According to the first Jordan lemma, this means that the integral on the big circle vanishes. In the same way, we have $\lim_{z\to0}|zf(z)|=0$ and the second Jordan lemma tells that the integral on $\Gamma_\epsilon^\pm$ is also zero. Then, using the fact that $f(z)\equiv P(z)/Q(z)$ and that, in this case, the residue can be written as $\mathrm{Res}(f,z)=P(z)/Q'(z)$, one obtains

$$\mathrm{Res}(f,z=+\omega)=\frac{\mathrm{e}^{-i\omega t}}{2\omega}. \tag{448}$$

As a consequence

$$\begin{aligned}\lim_{R\to+\infty}\left[\frac{i}{2\pi}\int_\Gamma f(z)\mathrm{d}z\right] &= \frac{i}{2\pi}\int_{-\infty}^{\infty}f(z)\mathrm{d}z\\ &= \frac{i}{2\pi}\times-2i\pi\mathrm{Res}(f,z=+\omega)\\ &= \frac{\mathrm{e}^{-i\omega t}}{2\omega},\end{aligned} \tag{449}$$

where the minus sign in the second line comes from the fact that the contour is clockwise (*i.e.* opposite to the direct direction).

Now if $t<0$, then the big circle is taken in the upper plane. As a consequence, the pole inside the contour is now the one located at $z=-\omega$. Therefore,

$$\begin{aligned}\frac{i}{2\pi}\int_\Gamma f(z)\mathrm{d}z &= \frac{i}{2\pi}\int_{-\infty}^{\infty}f(z)\mathrm{d}z\\ &= \frac{i}{2\pi}\times+2i\pi\mathrm{Res}(f,z=-\omega)\\ &= \frac{\mathrm{e}^{i\omega t}}{2\omega}.\end{aligned} \tag{450}$$

The previous considerations allow us to conclude that

$$\frac{i}{2\pi}\int_{-\infty}^{+\infty}\frac{\mathrm{e}^{-iEt}}{E^2-\omega^2}\mathrm{d}E=\frac{\mathrm{e}^{-i\omega|t|}}{2\omega}=D(t), \tag{451}$$

that is to say we have established the Fourier transform of the propagator. It should be obvious that the above equation is the counterpart of Eq. (398). As in field theory, we will show that it can be used to develop a perturbative method which allows us to take into account interactions.

24.5.2 Perturbations Theory in Quantum Mechanics

Before discussing our main subject, we need to explain how the Feynman diagrams can be used in quantum mechanics [55, 56]. Usually, in quantum mechanics, the perturbation theory is not based on the Feynman diagrams [57]. This gives the impression that the techniques of field theory are very different. In fact, as we now demonstrate, these techniques can also be utilized in ordinary quantum mechanics [54] and, obviously (as could have been guessed from the very beginning), they lead to the same physical predictions. In the following, we will discuss this question and compare the two approaches in some details. We believe that this can improve our physical understanding of the vacuum energy problem.

In order to achieve the above mentioned task, we need to recall how interactions are treated in quantum mechanics. For this purpose, let us consider a system obeying the following Schrödinger equation

$$i\frac{\mathrm{d}|\Psi(t)\rangle_{\mathrm{S}}}{\mathrm{d}t}=H(p_{\mathrm{S}},x_{\mathrm{S}})|\Psi(t)\rangle_{\mathrm{S}}, \tag{452}$$

where $H(p_{\mathrm{S}},x_{\mathrm{S}})$ is the Hamiltonian operator. The above equation is written in the Schrödinger representation, hence the subscript "S". Then, one can define the evolution operator $U(t,t_{\mathrm{i}})$ which relates the state vector at the time t to the state vector at some initial time $t_{\mathrm{i}}<t$ by the following expression

$$|\Psi(t)\rangle_{\mathrm{S}}=U(t,t_{\mathrm{i}})|\Psi(t_{\mathrm{i}})\rangle_{\mathrm{S}}. \tag{453}$$

This operator obeys

$$i\frac{\partial U(t,t_{\mathrm{i}})}{\partial t}=HU(t,t_{\mathrm{i}})\,, \tag{454}$$

and, provided the Hamiltonian is not explicitly time dependent, it can also be written as

$$U(t,t_{\mathrm{i}})=\mathrm{e}^{-iH(t-t_{\mathrm{i}})}. \tag{455}$$

Of course, in this representation, the operator p_{S} and x_{S} are time independent since only the state can change with time.

One the contrary, in the Heisenberg representation, the state vector does not evolve and can be expressed in terms of the Schrödinger state vector evaluated at some reference time $t=t_0$,

$$|\Psi\rangle_{\mathrm{H}}\equiv|\Psi(t_0)\rangle_{\mathrm{S}}. \tag{456}$$

Notice that the reference time is arbitrary and is not necessarily t_{i}. In this picture, the operators are time-dependent and, for instance, $x_{\mathrm{H}}(t)$ is given by

$$x_{\mathrm{H}}(t)\equiv U^{\dagger}(t,t_0)x_{\mathrm{S}}U(t,t_0), \tag{457}$$

such that $x_{\mathrm{H}}(t_0)=x_{\mathrm{S}}$. In this picture, the time evolution of the operators is computed from the Heisenberg equation which reads

$$i\frac{\mathrm{d}x_{\mathrm{H}}}{\mathrm{d}t} = i\frac{\partial U^{\dagger}(t,t_0)}{\partial t}x_{\mathrm{S}}U(t,t_0)+iU^{\dagger}(t,t_0)x_{\mathrm{S}}\frac{\partial U(t,t_0)}{\partial t} \tag{458}$$

$$= -U^{\dagger}Hx_{\mathrm{S}}U+U^{\dagger}x_{\mathrm{S}}HU \tag{459}$$

$$= [x_{\mathrm{S}},H]\,, \tag{460}$$

where in the second line we have used that $-i\partial U^\dagger/\partial t = U^\dagger H$.

Let us now consider the case where interactions are explicitly present. We assume that the corresponding Hamiltonian can be split into two parts, one corresponding to the free theory and one corresponding to the interactions

$$H = H_0 + H_{\rm int}. \tag{461}$$

We also assume that $H_{\rm int}$ is a small perturbation. The free evolution, due to the free Hamiltonian H_0, is described by the evolution operator U_0. This operator is of course different from U which describes the full evolution with the interactions taken into account. Clearly, U_0 can be expressed as

$$U_0(t,t_{\rm i}) = {\rm e}^{-iH_0(t-t_{\rm i})}\,. \tag{462}$$

Then the idea is to factor out the free evolution (which is supposed to be known) and to focus on the time evolution due to the perturbation on top of the free evolution. For this purpose, we define the interaction representation by (again, the representations are supposed to coincide when $t = t_0$)

$$\begin{aligned} x_{\rm I}(t) &= U_0^\dagger(t,t_0)x_{\rm S}U_0(t,t_0) &(463)\\ &= {\rm e}^{iH_0(t-t_0)}x_{\rm S}{\rm e}^{-iH_0(t-t_0)}. &(464)\end{aligned}$$

Clearly $x_{\rm I}(t)$ is different from $x_{\rm H}(t)$ because its time evolution is "generated" only by the free Hamiltonian and not by the total one. Let us now calculate the equation of motion of $x_{\rm I}(t)$. One has

$$\begin{aligned} \frac{{\rm d}x_{\rm I}}{{\rm d}t} &= iH_0x_{\rm I} + {\rm e}^{iH_0(t-t_0)}x_{\rm S}\times -iH_0{\rm e}^{-iH_0(t-t_0)} &(465)\\ &= i\left[H_0, x_{\rm I}\right]. &(466)\end{aligned}$$

Therefore, the time evolution of $x_{\rm I}(t)$ is controlled by the free Hamiltonian. This evolution is supposed to be known. In the same manner, we define the state vector in the interaction picture by

$$|\Psi(t)\rangle_{\rm I} = U_0^\dagger(t,t_0)|\Psi(t)\rangle_{\rm S}. \tag{467}$$

As a consequence, in the interaction picture, both the state vectors and the operators are time dependent. The equation obeyed by the state vector is given by

$$\begin{aligned} i\frac{{\rm d}|\Psi\rangle_{\rm I}}{{\rm d}t} &= -H_0|\Psi\rangle_{\rm I} + {\rm e}^{iH_0(t-t_0)}i\frac{{\rm d}|\Psi\rangle_{\rm S}}{{\rm d}t} &(468)\\ &= -H_0{\rm e}^{iH_0(t-t_0)}|\Psi\rangle_{\rm S}\\ &\quad +{\rm e}^{iH_0(t-t_0)}(H_0+H_{\rm int})|\Psi\rangle_{\rm S} &(469)\\ &= {\rm e}^{iH_0(t-t_0)}H_{\rm int}|\Psi\rangle_{\rm S} &(470)\\ &= {\rm e}^{iH_0(t-t_0)}H_{\rm int}{\rm e}^{-iH_0(t-t_0)}|\Psi\rangle_{\rm I} &(471)\\ &= H_{\rm I}|\Psi\rangle_{\rm I}, &(472)\end{aligned}$$

where $H_{\rm I} = U_0^\dagger(t,t_0)H_{\rm int}U_0(t,t_0)$ in agreement with the definition (463). The time evolution of $|\Psi\rangle_{\rm I}$ is therefore generated only by the interacting Hamiltonian. Then, it is natural to define the evolution operator in the interaction picture. This operator is associated with the Schrödinger equation (472). Its definition reads

$$|\Psi(t)\rangle_{\rm I} = U_{\rm I}(t,t_{\rm i})|\Psi(t_{\rm i})\rangle_{\rm I}. \tag{473}$$

Of course $U_{\rm I}$ is different from U_0 (and from U). A straightforward calculation shows that it can also be expressed as

$$U_{\rm I}(t,t_{\rm i}) = {\rm T}\exp\left[-i\int_{t_{\rm i}}^{t}{\rm d}\tau H_{\rm I}(\tau)\right]. \tag{474}$$

Another expression for $U_{\rm I}$ is based on the following considerations. From Eqs. (473) and (467), we have

$$\begin{aligned} |\Psi(t)\rangle_{\rm I} &= U_{\rm I}(t,t_{\rm i})|\Psi(t_{\rm i})\rangle_{\rm I} && (475)\\ &= U_{\rm I}(t,t_{\rm i})U_0^\dagger(t_{\rm i},t_0)|\Psi(t_{\rm i})\rangle_{\rm S} && (476)\\ &= U_0^\dagger(t,t_0)|\Psi(t)\rangle_{\rm S}. && (477)\end{aligned}$$

Therefore, one obtains

$$\begin{aligned} |\Psi(t)\rangle_{\rm S} &= U_0(t,t_0)U_{\rm I}(t,t_{\rm i})U_0^\dagger(t_{\rm i},t_0)|\Psi(t_{\rm i})\rangle_{\rm S} && (478)\\ &= U(t,t_{\rm i})|\Psi(t_{\rm i})\rangle_{\rm S}, && (479)\end{aligned}$$

from which we deduce that

$$U(t,t_{\rm i}) = U_0(t,t_0)U_{\rm I}(t,t_{\rm i})U_0^\dagger(t_{\rm i},t_0), \tag{480}$$

or

$$\begin{aligned} U_{\rm I}(t,t_{\rm i}) &= U_0^\dagger(t,t_0)U(t,t_{\rm i})U_0^\dagger(t_{\rm i},t_0), && (481)\\ &= {\rm e}^{iH_0(t-t_0)}{\rm e}^{-iH(t-t_{\rm i})}{\rm e}^{-iH_0(t_{\rm i}-t_0)}. && (482)\end{aligned}$$

This expression is very natural since it expresses nothing but the fact that the full evolution of the system is just the combination of the free evolution and of the evolution due to the interacting term. Let us also remark that one can write

$$\begin{aligned} x_{\rm I}(t) &= U_0^\dagger(t,t_0)x_{\rm S}U_0(t,t_0) && (483)\\ &= U_0^\dagger(t,t_0)U(t,t_0)x_{\rm H}U^\dagger(t,t_0)U_0(t,t_0) && (484)\\ &= {\rm e}^{iH_0(t-t_0)}{\rm e}^{-iH(t-t_0)}x_{\rm H}(t){\rm e}^{iH(t-t_0)}{\rm e}^{-iH_0(t-t_0)} && (485)\\ &= U_{\rm I}(t,t_0)x_{\rm H}(t)U_{\rm I}^\dagger(t,t_0). && (486)\end{aligned}$$

Again, this result is very intuitive and expresses the fact that, in the interacting picture, the free evolution of the operators is factorized out.

After these preliminaries, one can now derive the Gell-Man Law equation which is at the heart of the perturbative treatment presented in what follows. Let $|n_{\rm I}\rangle$ be the energy eigenstates of the full Hamiltonian, *i.e.* the free Hamiltonian plus the Hamiltonian describing the interactions. We denote the true vacuum state by $|0_{\rm I}\rangle \equiv |\Omega\rangle$, in agreement with the conventions adopted in the previous subsections. Then, one can always expand the vacuum state of the free theory according to

$$\begin{aligned} |0\rangle &= \sum_{n=0}^{\infty}|n_{\rm I}\rangle\langle n_{\rm I}|0\rangle, && (487)\\ &= |\Omega\rangle\langle\Omega|0\rangle + \sum_{n\neq 0}^{\infty}|n_{\rm I}\rangle\langle n_{\rm I}|0\rangle. && (488)\end{aligned}$$

As a consequence, applying the operator ${\rm e}^{-iH(t+t_0)}$ to the above relation, one obtains

$$\begin{aligned} {\rm e}^{-iH(t+t_0)}|0\rangle &= {\rm e}^{-iE_0(t+t_0)}|\Omega\rangle\langle\Omega|0\rangle \\ &+ \sum_{n\neq 0}^{\infty}{\rm e}^{-iE_n(t+t_0)}|n_{\rm I}\rangle\langle n_{\rm I}|0\rangle. && (489)\end{aligned}$$

Then, we take the limit $t \to \infty(1-i\epsilon)$. This kills all the terms on the right hand side but the first one. Indeed, we have $E_n > E_0$ $(n \neq 0)$ and this means that ${\rm e}^{-iE_0 t}$ goes to zero

less rapidly than all the other exponential terms. Therefore, in this limit, Eq. (489) reads $\mathrm{e}^{-iH(t+t_0)}|0\rangle \simeq \mathrm{e}^{-iE_0(t+t_0)}|\Omega\rangle\langle\Omega|0\rangle$ and, by inverting this formula, this allows us to express the interacting vacuum as

$$|\Omega\rangle = \lim_{t\to\infty(1-i\epsilon)} \frac{\mathrm{e}^{iE_0(t+t_0)}}{\langle\Omega|0\rangle}\mathrm{e}^{-iH(t+t_0)}|0\rangle \tag{490}$$

$$= \lim_{t\to\infty(1-i\epsilon)} \frac{\mathrm{e}^{iE_0(t+t_0)}}{\langle\Omega|0\rangle}\mathrm{e}^{-iH(t+t_0)}\mathrm{e}^{iH_0(t+t_0)}|0\rangle$$

$$= \lim_{t\to\infty(1-i\epsilon)} \frac{\mathrm{e}^{iE_0(t+t_0)}}{\langle\Omega|0\rangle}U_{\mathrm{I}}(t_0,-t)|0\rangle. \tag{491}$$

In the second line, we have used the fact that $H_0|0\rangle = 0$ since we can always define the ground state as the state associated to the zero of energy. This means that $\mathrm{e}^{iH_0(t+t_0)}|0\rangle = |0\rangle$ and explains the appearance of the new factor $\mathrm{e}^{iH_0(t+t_0)}$. In the same fashion, using $\mathrm{e}^{iH(t-t_0)}|\Omega\rangle = \mathrm{e}^{iE_0(t-t_0)}|\Omega\rangle$, one has

$$\langle 0|\mathrm{e}^{-iH(t-t_0)} = \lim_{t\to\infty(1-i\epsilon)} \langle 0|\Omega\rangle\langle\Omega|\mathrm{e}^{-iE_0(t-t_0)}, \tag{492}$$

from which one deduces that

$$\langle\Omega| = \lim_{t\to\infty(1-i\epsilon)} \frac{\mathrm{e}^{iE_0(t-t_0)}}{\langle 0|\Omega\rangle}\langle 0|\mathrm{e}^{-iH(t-t_0)} \tag{493}$$

$$= \lim_{t\to\infty(1-i\epsilon)} \frac{\mathrm{e}^{iE_0(t-t_0)}}{\langle 0|\Omega\rangle}\langle 0|\mathrm{e}^{iH_0(t-t_0)}\mathrm{e}^{-iH(t-t_0)}$$

$$= \lim_{t\to\infty(1-i\epsilon)} \frac{\mathrm{e}^{iE_0(t-t_0)}}{\langle 0|\Omega\rangle}\langle 0|U_{\mathrm{I}}(t,t_0). \tag{494}$$

Therefore, the norm of the vacuum state is given by

$$\langle\Omega|\Omega\rangle = \lim_{t\to\infty(1-i\epsilon)} \frac{\mathrm{e}^{2iE_0t}}{|\langle 0|\Omega\rangle|^2}\langle 0|U_{\mathrm{I}}(t,t_0)U_{\mathrm{I}}(t_0,-t)|0\rangle. \tag{495}$$

In the following, we take $\langle\Omega|\Omega\rangle = 1$ although this choice is not mandatory (but, then, when we compute a correlation function, it would be necessary to divide the corresponding expression by $\langle\Omega|\Omega\rangle$).

We are now in a position where one can calculate the two-point correlation function of the field operator [recall that $x(t)$ is directly proportional to the field operator, see Eq. (429)]. Using the previous results, one obtains

$$\langle\Omega|x_{\mathrm{H}}(t_2)x_{\mathrm{H}}(t_1)|\Omega\rangle = \lim_{t\to\infty(1-i\epsilon)} \frac{\mathrm{e}^{2iE_0t}}{|\langle 0|\Omega\rangle|^2}\langle 0|U_{\mathrm{I}}(t,t_0)U_{\mathrm{I}}^{\dagger}(t_2,t_0)x_{\mathrm{I}}(t_2)U_{\mathrm{I}}(t_2,t_0)U_{\mathrm{I}}^{\dagger}(t_1,t_0)x_{\mathrm{I}}(t_1)U_{\mathrm{I}}(t_1,t_0)U_{\mathrm{I}}(t_0,-t)|0\rangle \tag{496}$$

$$= \lim_{t\to\infty(1-i\epsilon)} \frac{\langle 0|U_{\mathrm{I}}(t,t_0)U_{\mathrm{I}}(t_0,t_2)x_{\mathrm{I}}(t_2)U_{\mathrm{I}}(t_2,t_0)U_{\mathrm{I}}(t_0,t_1)x_{\mathrm{I}}(t_1)U_{\mathrm{I}}(t_1,t_0)U_{\mathrm{I}}(t_0,-t)|0\rangle}{\langle 0|U_{\mathrm{I}}(t,t_0)U_{\mathrm{I}}(t_0,-t)|0\rangle} \tag{497}$$

$$= \lim_{t\to\infty(1-i\epsilon)} \frac{\langle 0|U_{\mathrm{I}}(t,t_2)x_{\mathrm{I}}(t_2)U_{\mathrm{I}}(t_2,t_1)x_{\mathrm{I}}(t_1)U_{\mathrm{I}}(t_1,-t)|0\rangle}{\langle 0|U_{\mathrm{I}}(t,-t)|0\rangle} \tag{498}$$

$$= \lim_{t\to\infty(1-i\epsilon)} \frac{\langle 0|\mathrm{T}\left[x_{\mathrm{I}}(t_2)x_{\mathrm{I}}(t_1)U_{\mathrm{I}}(t,-t)\right]|0\rangle}{\langle 0|U_{\mathrm{I}}(t,-t)|0\rangle} \tag{499}$$

$$= \lim_{t\to\infty(1-i\epsilon)} \frac{\langle 0|\mathrm{T}\left[x_{\mathrm{I}}(t_2)x_{\mathrm{I}}(t_1)\exp\left(-i\int_{-t}^{t}\mathrm{d}\tau H_{\mathrm{I}}\right)\right]|0\rangle}{\langle 0|\exp\left(-i\int_{-t}^{t}\mathrm{d}\tau H_{\mathrm{I}}\right)|0\rangle}, \tag{500}$$

where, in order to go from the first to the second line, we have used Eq. (495) and the normalization $\langle\Omega|\Omega\rangle = 1$ to express the term $|\langle 0|\Omega\rangle|^2$. Eq. (500) is the Gell-Mann Law equation [16]. It will be the basis for our perturbative treatment of quantum mechanics. It plays the same role as Eq. (394) since it allows us to perturbatively evaluate the various correlation functions in terms of Feynman diagrams (the Gell-Mann Law equation can be generalized to higher correlation function if needed).

24.5.3 Quantum Mechanics and Bubble Diagrams

In order to mimic the self-interacting field theory studied before, we now consider a quartic oscillator in quantum mechanics [55, 56], *i.e.* a system where the potential is given by $V(x) \sim x^2 + x^4$. More precisely, this means that the interacting Hamiltonian can be expressed as

$$H_{\rm int} = \frac{\lambda}{4!}\Phi^4, \tag{501}$$

where we recall that Φ is given by Eq. (429). Let us then compute the two-point correlation function. For this purpose, as explained before, we use the Gell-Mann Law equation. The numerator reads

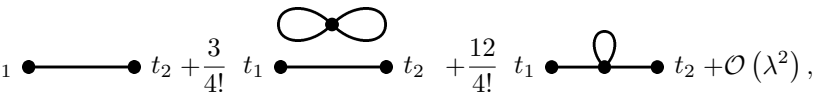
$$\left\langle 0 \left| \mathrm{T}\left[\Phi(t_2)\Phi(t_1)\exp\left(-i\int \mathrm{d}\tau H_{\rm I}\right)\right]\right| 0\right\rangle, \tag{502}$$

where, for simplicity, we do not write the limits (but there will be reestablished when needed). It is also important to remember that the field are written in the interacting picture (and, again for simplicity we do not write the subscript "I", they can be considered as free fields). Obviously, one cannot calculate exactly this quantity so we do it perturbatively by expanding the exponential. At first order in the coupling constant λ, we obtain

$$\begin{aligned}\left\langle 0 \left| \mathrm{T}\left\{\Phi(t_2)\Phi(t_1)\left[1 - i\frac{\lambda}{4!}\int \mathrm{d}\tau\Phi^4(\tau) + \cdots\right]\right\}\right| 0\right\rangle &= \langle 0|\mathrm{T}\left[\Phi(t_2)\Phi(t_1)\right]|0\rangle \\ &\quad -i\frac{\lambda}{4!}\int \mathrm{d}\tau \left\langle 0 \left| \mathrm{T}\left[\Phi(t_2)\Phi(t_1)\Phi^4(\tau)\right]\right| 0\right\rangle + \cdots, \qquad (503)\\ &= D(t_1,t_2) - i\frac{\lambda}{4!}\int \mathrm{d}\tau \left\langle 0 \left| \mathrm{T}\left[\Phi(t_2)\Phi(t_1)\Phi^4(\tau)\right]\right| 0\right\rangle + \cdots. \qquad (504)\end{aligned}$$

It should now be clear that the same expression can also be written in terms on Feynman diagrams. One arrives at the following expression

$$\left\langle 0 \left| \mathrm{T}\left[\Phi(t_2)\Phi(t_1)\exp\left(-i\int \mathrm{d}\tau H_{\rm I}\right)\right]\right| 0\right\rangle =$$



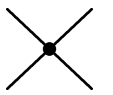
(505)

where we have defined the following Feynman rules [55, 56]

$$D(t_1,t_2) \equiv$$

(506)

$$-i\lambda\int \mathrm{d}\tau \equiv$$

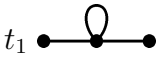


. (507)

Clearly, these Feynman rules are nothing but the one-dimensional version of the standard Feynman rules (402) and (403) of quantum field theory. The only difference is that four dimensional space-time integrations are simply replaced with one-dimensional time integration. In fact, in terms of diagrams, the expressions in field theory or in quantum mechanics are exactly similar. In particular, we see the appearance of diagrams with loops and external legs (the tadpole) and also of diagrams with loops but no external legs, our bubble diagrams. The advantage of our approach now becomes clear. We are in a position where not only one can compute these diagrams in quantum mechanics but we can also investigate their deep meaning by comparing

the result with the standard perturbative theory of quantum mechanics where everything is well under control.

Let us start with the tadpole diagram. Using Eqs. (506) and (507), the corresponding expression can be written as



$$= -i\lambda \int \mathrm{d}\tau D(t_1,\tau)D(\tau,\tau)D(\tau,t_2) \tag{508}$$

$$= -\frac{i\lambda}{8\omega^3}\int \mathrm{d}\tau \mathrm{e}^{-i\omega|t_1-\tau|-i\omega|\tau-t_2|} \tag{509}$$

$$= -\frac{i\lambda}{4\omega^2}D(t_2,t_1)\left(|t_2-t_1|-\frac{i}{\omega}\right). \tag{510}$$

The most striking difference with the case of quantum field theory is that the tadpole diagram is now finite. In some sense, this is the signal that ordinary quantum mechanics does not need renormalization: the perturbative approach leads to finite result. Technically, this is due to the fact that we deal with one-dimensional integrals that have better convergence properties. Physically, this is due to the fact that a single quantum-mechanical oscillator leads to finite predictions while an infinite collections of such systems apparently lead to infinite physical quantities.

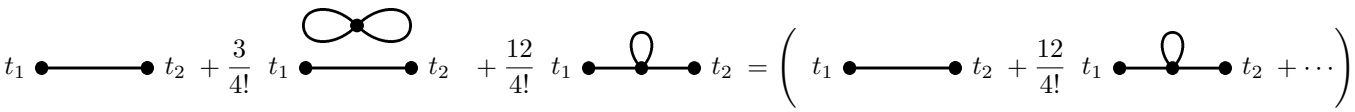
Now let us calculate the "bubble diagram" present in the second term of the above expansion (505). Having established that the tadpole diagram is now finite, it is particularly interesting to see what happens for this type of diagram. Using again the Feynman rules (506) and (507), it is straightforward to establish that

$$= -i\lambda \int D^2(\tau,\tau)\mathrm{d}\tau \tag{511}$$

$$= -i\frac{\lambda}{4\omega^2}\int \mathrm{d}\tau = \infty. \tag{512}$$

Therefore, even in quantum mechanics, the bubble diagram remains divergent. This shows that the divergent nature of the loop and bubble diagrams is, in some sense, different. They are both divergent in quantum field theory but one (loop diagrams with external legs) becomes finite in the limit of quantum mechanics while the other (bubble diagram) remains infinite.

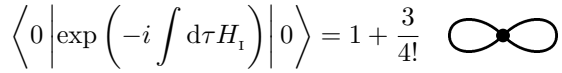
But this now leads to the following question. We know from the standard perturbation theory of quantum mechanics that all the physical predictions are finite, even without renormalization. However, using the other approach based on the Feynman diagrams, we have just seen that it leads to terms that can be divergent. How these two facts can be consistent with each other? The answer is of course that, as in field theory, the bubble diagrams exactly cancel out in the calculation of the correlation functions. This can easily be seen if one notices that



(513)

(514)

and that



(515)

Therefore (as expected) the infinities of the bubble diagram never play a role in ordinary quantum mechanics. As a consequence, the two-point correlation function is given by

$$\langle\Omega|\Phi(t_2)\Phi(t_1)|\Omega\rangle = \quad t_1 \bullet\!\!-\!\!-\!\!\bullet\ t_2 \ + \frac{12}{4!}\ t_1 \bullet\!\!-\!\!\bigcirc\!\!-\!\!\bullet\ t_2 \ + \mathcal{O}\left(\lambda^2\right), \tag{516}$$

which is perfectly finite. Again, in terms of Feynman diagrams the above equation is strictly equivalent to the corresponding one in quantum field theory, see Eq. (421).

24.5.4 Quantum Mechanics and Vacuum Energy

Let us now discuss how the vacuum energy is evaluated in ordinary quantum mechanics. The calculation is straightforward when there is no interaction and, from Eq. (435), one obtains

$$\langle H\rangle = \frac{\omega}{2}. \tag{517}$$

This is of course nothing but the ground state energy of an harmonic oscillator. It cannot be zero because, due the Heisenberg uncertainty principle, the kinetic and potential energy cannot vanish at the same time. It can also be rewritten as

$$\langle H\rangle = \omega^2 D(0), \tag{518}$$

where we have used Eq. (444). This exactly corresponds to Eq. (410) with $m = \omega$ and $4 \to 1$ (since we now work in one dimension rather than in four), except that there is a sign difference which comes from the fact that, in field theory, we have used the signature (- + + +). Of course, the big difference is that $D(0)$ is a finite quantity in quantum mechanics. In some sense, the vacuum energy problem appears to be very simple, almost trivial. The ground state of a quantum-mechanical oscillator is finite but a quantum field is a system which can be described by an infinite collection of harmonic oscillators and, as a consequence, its ground state energy is infinite.

Let us now consider a case where an interaction is present as we did in Sec. 24.4.3. For simplicity, we consider again the case of a quartic oscillator with $V_{\rm int} = \lambda\Phi^4/4!$. Upon using the standard perturbation theory in quantum mechanics, one can calculate the displacement of the energy levels. They are given by the standard formula [57]

$$E_n = E_n^0 + \langle n|V_{\rm int}|n\rangle, \tag{519}$$

where $|n\rangle$ are the unperturbed eigenvectors. It is easy to show that they can be expressed as

$$\Delta E_n \equiv E_n - E_n^0 = \frac{\lambda}{32\omega^2}\left(2n^2+2n+1\right), \tag{520}$$

which means that, for the first two levels, the shift in energy is given by

$$\Delta E_0 \ = \ \frac{\lambda}{32\omega^2}, \quad \Delta E_1 = \frac{5\lambda}{32\omega^2}. \tag{521}$$

We can now use the other method, based on the Feynman diagrams. It is easy to obtain that

$$\begin{aligned}\Delta E_0 \ &= \ \frac{\lambda}{4!}\left\langle\Phi^4\right\rangle = \frac{3\lambda}{4!}\left\langle\Phi^2\right\rangle^2 = \frac{\lambda}{8}D^2(0)\\ &= \ \frac{i}{8\int \mathrm{d}t}\end{aligned} \quad \bigcirc\!\!-\!\!\bullet\!\!-\!\!\bigcirc \tag{522}$$

Clearly, we have just obtained the strict equivalent of Eq. (420), the term $\int \mathrm{d}^4x$ being replaced by $\int \mathrm{d}t$ as could have been guessed in the context of quantum mechanics. Then, using the fact that $D(0) = 1/(2\omega)$, one obtains that

$$\Delta E_0 = \frac{\lambda}{32\omega^2}, \tag{523}$$

that is to say we recover the result established before, see Eq. (521). Therefore, we have shown explicitly that the two approaches lead to the same result. It is also interesting to notice that ΔE_0 is given by a bubble diagram but that this divergent graph is canceled by the infinite term $\int \mathrm{d}t$ so that a finite quantity is left.

The previous considerations also lead to the following question: before we had to deal with the renormalization of the mass. How does this aspect appear in the present context? In this case, we deal with the same diagrams, namely

$$\langle \Omega \,|T\left[\Phi(t_1)\Phi(t_2)\right]|\,\Omega\rangle \quad = \quad x_1 \bullet\!\!-\!\!-\!\!\bullet\, x_2 + \frac{1}{2}\; x_1 \bullet\!\!-\!\!\bigcirc\!\!-\!\!\bullet\, x_2 + \cdots\,, \tag{524}$$

or, in terms of explicit expressions [compare with Eqs. (421) and (422)]

$$\begin{aligned}\langle \Omega \,|T\left[\phi(t_1)\phi(t_2)\right]|\,\Omega\rangle &= D_{\mathrm{F}}(t_1-t_2) - \frac{i\lambda}{2}\int \mathrm{d}\tau D(t_1-\tau)D_{\mathrm{F}}(0)D_{\mathrm{F}}(\tau-t_2) &(525)\\ &= D_{\mathrm{F}}(t_1-t_2) + \frac{i\lambda}{2}D_{\mathrm{F}}(0)\frac{1}{2\pi}\int \mathrm{d}E \frac{\mathrm{e}^{-iE(t_1-t_2)}}{(E^2-\omega^2)^2}, &(526)\end{aligned}$$

and this leads to

$$\omega_{\mathrm{ren}}^2 = \omega^2 + \frac{\lambda}{2}D(0), \tag{527}$$

which is exactly Eq. (426) (again, there is a sign difference which originates from the sign difference in the propagator, see above). In quantum mechanics, the propagator is finite and therefore

$$\omega_{\mathrm{ren}}^2 = \omega^2 + \frac{\lambda}{4\omega}, \tag{528}$$

or

$$\omega_{\mathrm{ren}} = \omega + \frac{\lambda}{8\omega^2}. \tag{529}$$

Then, one can express the spectrum in terms of the renormalized "mass" ω_{ren} as we have done in quantum field theory although this is not necessary here since the result is anyway finite. Straightforward calculations lead to

$$E_n = \omega_{\mathrm{ren}}\left(n+\frac{1}{2}\right) - \frac{\lambda}{32\omega_{\mathrm{ren}}^2}\left(2n^2-2n-1\right). \tag{530}$$

It does not come as a surprised that, as a function of the renormalized mass, the spectrum is a different function of n, see Eq. (520). However, from the above expression one also obtains

$$E_1 - E_0 = \omega_{\mathrm{ren}}. \tag{531}$$

In presence of an interaction, the difference between the first excited state and the fundamental level has been "modified" (or, rather, is expressed differently). But, observationally, we define or measure the spectrum with respect to $E_1 - E_0$. In other words what we define as ω is in fact $E_1 - E_0$. Therefore, renormalizing ω consists in preserving this definition even in presence of an interaction. Obviously, $E_1 - E_0$ is not the bare ω (the one which appears in the Lagrangian) when the an-harmonic term is present. What we do here is to define ω as the observed one, namely as $E_1 - E_0$, and then calculate the spectrum in terms of that observational quantity. What is done in quantum field theory is clearly exactly similar to the above described procedure except of course that the bare quantities are in fact infinite because $D_{\mathrm{F}}(0)$ is infinite. Therefore, this exercise illustrates nicely the deep meaning of renormalization.

This concludes our discussion of the vacuum energy density in terms of Feynman diagrams. We have shown that the cosmological constant is in fact given in terms of very peculiar graphs, the so-called bubble diagrams. These diagrams are, in a sense, more divergent than the usual loop diagrams because they remain infinite even in the limit of quantum mechanics contrary to the last ones. In non gravitational physics, this is not a problem because the bubble diagrams always cancel out in the equations describing observable quantities. When the gravitational field is turned on, we face the tasks of renormalizing these diagrams which is more difficult than in the standard situation because of their bad behavior mentioned before. Given that

astrophysical observations seem to indicate that the vacuum energy is non vanishing (see below), the details of the renormalizing procedure become a crucial issue. Not only we have to extract a finite quantity from a divergent graph but this finite quantity must be in agreement with the observations. As argued before, the technique which consists in imposing a spatial cut-off is certainly not consistent (and, as well-known, gives a very large contribution to the vacuum energy). A more reasonable approach seems to be dimensional regularization since this satisfies Lorentz invariance and is therefore consistent with the vacuum equation of state. As shown before, this typically leads to Eq. (392). In the next section, we present a new argument, based on the Gaussian effective potential, which seems to support this conclusion.

24.6 The Gaussian Effective Potential

The Gaussian effective potential is a non perturbative approach to quantum field theory [58, 59, 60, 61, 62, 63, 64, 65, 66]. Originally developed in the context of quantum mechanics, it is has been generalized to field theory. Much less known is the fact that the Gaussian effective potential method has something to say about the vacuum energy problem and the aim of this section is to present these considerations [59]. In the next subsection, we quickly recall the main idea and how this approach can be applied to ordinary quantum mechanics. Then, we explain how it can be implemented in quantum field theory and, finally, we discuss its implication for the cosmological constant problem.

24.6.1 The Gaussian Effective Potential in Quantum Mechanics

The main idea underlying the Gaussian effective potential approach is the following one [58]. At the classical level, a system is described by a potential. Once this one is specified, one can compute the evolution of the system. When quantum effects are taken into account, the behavior of the system will be modified and, by means of the formalism of quantum mechanics, one can work out the corresponding physical predictions. The idea of the Gaussian effective potential is to find an effective potential which, at the same time, can be used as if it were a "classical" potential (that is to say can be used in the framework of Newtonian dynamics) and takes into account the quantum effects. At first sight, this approach seems equivalent to what is known in the literature as the effective potential [67, 68]. In this method, one requires the wave-function mean value to be centered at a fixed point Φ_0. Then, the effective potential gives the exact ground state energy of the system in presence of quantum corrections. However, in some situation, the result can be quite artificial. Consider for instance a situation where the wave-function possesses two large peaks on each side of Φ_0. In such a case, the effective potential does not necessarily lead to a correct description of what happens in the vicinity of Φ_0 because its shape could just be determined by some average procedure between the two distant peaks. Typically, one would obtain a convex potential centered at Φ_0 while our physical intuition rather indicates that the effective potential should be a double-well potential, the two wells corresponding to the two peaks of the wave-function. Another problem is that the standard effective potential is usually computed perturbatively in $\hbar$. Clearly, when the quantum corrections become important, this approach breaks down. Since the Gaussian effective potential does not suffer from the two above mentioned drawbacks, we will use it in this review. When possible, we will discuss the difference between the two approaches in order to stress their differences and similarities [69].

We now turn to the precise definition of the Gaussian effective potential. This definition is based on a variational principle. Here we consider a one-dimensional system as we have already done in Sec. 24.5.1. In particular, this means that the scalar field is given by Eq. (429). Then, the Gaussian effective potential is defined by [58]

$$V_{\mathrm{G}}\left(\Phi_0\right) = \min_{\Omega} \left\langle \Psi \left| H \right| \Psi \right\rangle , \tag{532}$$

where H is the Hamiltonian of the system and where the wave-function is taken to be a simple Gaussian (hence the name Gaussian effective potential)

$$\Psi\left(\Phi\right) = \left(\frac{\Omega}{\hbar\pi}\right)^{1/4} \mathrm{e}^{-\Omega(\Phi-\Phi_0)^2/(2\hbar)}. \tag{533}$$

The parameter Ω is related to the width of the wave-function. Here the idea is to find the value of Ω which minimizes the energy of the system. In practice, the mean value of the Hamiltonian that appears in Eq. (532) can be expressed as

$$\langle H\rangle = \int_{-\infty}^{\infty} \mathrm{d}\Phi \Psi^{*}(\Phi)\left[-\frac{\hbar^2}{2}\frac{\mathrm{d}^2}{\mathrm{d}\Phi^2} + V(\Phi)\right]\Psi(\Phi). \tag{534}$$

To illustrate how the calculation of the Gaussian effective potential works on a concrete example, we choose the following potential

$$V(\Phi) = \frac{1}{2}m^2\Phi^2 + \lambda\Phi^4, \tag{535}$$

that is to say the potential of an an-harmonic oscillator. Another motivation for this choice is that the comparison with field theory is easy since the prototypical model possesses the same potential. Upon using Eq. (533) in Eq. (534), one arrives at

$$\begin{aligned}\langle H\rangle(\Phi_0,\Omega) &= \frac{\hbar\Omega}{4} + \frac{1}{2}m^2\left(\Phi_0^2 + \frac{\hbar}{2\Omega}\right) \\ &\quad +\lambda\left[\Phi_0^4 + 6\Phi_0^2\frac{\hbar}{2\Omega} + \frac{3\hbar^2}{(2\Omega)^2}\right].\end{aligned} \tag{536}$$

We must now determine the optimal Ω. As explained above, this is achieved by requiring $\partial\langle H\rangle/\partial\Omega = 0$, which leads to the equation

$$\Omega^3 - \left(m^2 + 12\lambda\Phi_0^2\right)\Omega - 6\hbar\lambda = 0. \tag{537}$$

This equation allows us to determine Ω in terms of Φ_0. In the following, we write this result as $\Omega_0 \equiv \Omega(\Phi_0)$. Upon using the above relation in the expression of $\langle H\rangle$, one obtains the following expression for the Gaussian effective potential

$$V_{\mathrm{G}}(\Phi_0) = \frac{1}{2}m^2\Phi_0^2 + \lambda\Phi_0^4 + \frac{\hbar\Omega_0}{2} - \frac{3\hbar^2\lambda}{4\Omega_0}. \tag{538}$$

Of course, as already emphasized before, one should not forget that Ω_0 is a function of Φ_0.

A complete study of the above potential has been provided in Ref. [58] and we will not repeat this analysis. Here, we just want to illustrate that the Gaussian effective potential is an accurate method to estimate the ground state of a system (even in the strong coupling limit) since this is directly relevant for the cosmological constant problem. Let us consider the case where $m^2 > 0$ (the case of the double-well potential, reminiscent of symmetry breaking in field theory has been studied in Ref. [58]). The ground state of the an-harmonic oscillator is given by Eq. (538)

$$V_{\mathrm{G}}(\Phi_0 = 0) = \frac{\hbar\Omega_0}{2}\left(1 - \frac{3}{2}\frac{\hbar\lambda}{\Omega_0^3}\right), \tag{539}$$

where Ω_0 is solution of

$$\Omega^3 - m^2\Omega - 6\hbar\lambda = 0. \tag{540}$$

The discriminant of this cubic equation can be expressed as

$$\Delta(\Phi_0 = 0) = 4m^6\left(1 - \frac{972}{16}\frac{1}{\xi^6}\right), \tag{541}$$

where the quantity ξ in the above relation is given by

$$\xi^2 \equiv \frac{m^2}{(2\lambda)^{2/3}}. \tag{542}$$

The awkward coefficient 972/16 in Eq. (541) originates from our will to work with the definition of ξ used in Ref. [58]. The cubic equation (540) can be explicitly solved. We deal with different branches according to the sign of the discriminant. This sign is determined by $\xi_{\text{lim}} = (972/16)^{1/6} \simeq 1.9827$. If $\xi < \xi_{\text{lim}}$ ($\xi = 0$ represent the quartic oscillator) we are in the strong coupling regime which is of particular interest in order to demonstrate the usefulness of the Gaussian effective potential. Therefore, let us focus on this situation. In this case, the exact solution of Eq. (540) can be written as

$$\begin{aligned}\Omega_0 &= m\left[\frac{3}{2\xi^3} - \frac{1}{27}\sqrt{27\left(\frac{972}{16}\frac{1}{\xi^6} - 1\right)}\right]^{1/3} \\ &\quad + m\left[\frac{3}{2\xi^3} + \frac{1}{27}\sqrt{27\left(\frac{972}{16}\frac{1}{\xi^6} - 1\right)}\right]^{1/3}. \end{aligned} \tag{543}$$

Together with Eq. (539), the above equation gives an excellent approximation for the ground state. For instance, if $m = 1$ and $\lambda = 10$, then $\xi^2 \simeq (1/20)^{2/3} \simeq 0.1357$ and $\Omega_0 = 4$. As a consequence, this gives $V_{\text{G}}(\Phi_0 = 0) \simeq 1.53$, a result that is accurate at $\simeq 1.75\%$ according to Ref. [58].

This simple example has shown that the Gaussian effective potential can be an efficient tool to calculate the ground state of a quantum system even in a non perturbative regime. Therefore, it appears as an interesting approach for the vacuum energy problem. However, before turning to the calculation of the Gaussian effective potential in field theory, it is interesting to compare it with the standard effective potential in more detail. This is the purpose of the next subsection.

24.6.2 Comparison with the One Loop Effective Potential

In order to explain how the effective potential is obtained, we need to quickly return to the basics of quantum field theory. In Eq. (393), we have defined the generating functional $Z[J]$. Let us now define another generating functional $W[J]$ by [16, 17]

$$Z[J] = \mathrm{e}^{iW[J]}. \tag{544}$$

Following the procedure in Eq. (394), one can define new correlation functions $G_{\text{c}}^{(n)}$ according to

$$G_{\text{c}}^{(n)} = \left(\frac{1}{i}\right)^{n-1} \frac{\delta^n W[J]}{\delta J(x_1)\cdots\delta J(x_n)}\bigg|_{J=0}. \tag{545}$$

As is well-known, these functions represent in fact the connected n-points functions. Then, one introduces the classical field Φ_{c}

$$\Phi_{\text{c}} \equiv \frac{\delta W[J]}{\delta J(x)}\bigg|_{J=0}, \tag{546}$$

where we remind that $J(x)$ is a source. This definition can be easily justified. Indeed, from Eqs. (393) and (544), one sees that

$$\Phi_{\text{c}} = \frac{1}{iZ[J]}\frac{\delta Z[J]}{\delta J(x)} \tag{547}$$

$$= \frac{\mathcal{N}\int \mathcal{D}\Phi \exp\left\{i\int \mathrm{d}^4x\left[\mathcal{L} + J(x)\Phi(x)\right]\right\}\Phi(x)}{\mathcal{N}\int \mathcal{D}\Phi \exp\left\{i\int \mathrm{d}^4x\left[\mathcal{L} + J(x)\Phi(x)\right]\right\}}$$

$$= \frac{\langle\Omega\,|\Phi(x)|\,\Omega\rangle}{\langle\Omega|\Omega\rangle}. \tag{548}$$

Therefore, we see that the classical field is in fact the vacuum expectation value of the field operator in the vacuum state of the theory. When the source vanishes, the classical field does not necessarily goes to zero, a typical and well-known example being of course symmetry breaking.

The next step consists in introducing a new generating function, the so-called effective action, defined by

$$\Gamma[\Phi_c] = W[J] - \int \mathrm{d}^4x J(x)\Phi_c(x). \tag{549}$$

Clearly, $\Gamma[\Phi_c]$ does not depend on the source since $\delta\Gamma/\delta J = 0$. Let us now calculate the functional derivative of Γ with respect with the classical field. This gives

$$\begin{aligned}\frac{\delta\Gamma[\Phi_c]}{\delta\Phi_c(x)} &= \frac{\delta W}{\delta\Phi_c(x)} - \int \mathrm{d}^4y \frac{\delta J(y)}{\delta\Phi_c(x)}\Phi_c(y) \\ &\quad - \int \mathrm{d}^4y J(y)\frac{\delta\Phi_c(y)}{\delta\Phi_c(x)} \qquad (550)\\ &= \int \mathrm{d}^4y \frac{\delta W}{\delta J(y)}\frac{\delta J(y)}{\delta\Phi_c(x)} - \int \mathrm{d}^4y \frac{\delta J(y)}{\delta\Phi_c(x)}\Phi_c(y) \\ &\quad - J(x) = -J(x). \qquad (551)\end{aligned}$$

In order to interpret these equations and to better understand their meaning, it is interesting to apply them to the case of a free theory where all the calculations can be carried out explicitly. From Eq. (396), we immediately see that (the subscript zero indicates that the calculation is performed for the free theory)

$$iW_0[J] = -\frac{1}{2}\int \mathrm{d}^4x\,\mathrm{d}^4y J(x)D_{\mathrm{F}}(x-y)J(y). \tag{552}$$

Upon using Eq. (546), one obtains

$$\Phi_c(z) = -\frac{1}{i}\int \mathrm{d}^4x\, D_{\mathrm{F}}(z-x)J(x). \tag{553}$$

But iD_{F} is the Green function of the Klein-Gordon operator. As a consequence, the above equation implies that

$$\left(\eta^{\mu\nu}\partial_\mu\partial_\nu - m^2\right)\Phi_c(x) = J(x), \tag{554}$$

i.e. the classical field obeys the classical equation of motion (hence its name). Then, using Eqs. (552) and (553) into the definition (549), the effective action can be expressed as

$$\Gamma_0[\Phi_c] = \frac{1}{2i}\int \mathrm{d}^4x\mathrm{d}^4y J(x)D_{\mathrm{F}}(x-y)J(y). \tag{555}$$

The next step is to use the equation of motion of the classical field (554) to rewrite the effective action as

$$\begin{aligned}\Gamma_0[\Phi_c] &= \frac{1}{2i}\int \mathrm{d}^4x\,\mathrm{d}^4y D_{\mathrm{F}}(x-y)\left(\eta^{\mu\nu}\partial^x_\mu\partial^x_\nu - m^2\right)\Phi_c(x) \\ &\quad \times\left(\eta^{\mu\nu}\partial^y_\mu\partial^y_\nu - m^2\right)\Phi_c(y), \qquad (556)\end{aligned}$$

where $\partial^{x,y}_\mu$ means a partial derivative with respect to the coordinates at point x and y. Then, integrating by part the x dependent terms in the above integral leads to

$$\begin{aligned}\Gamma_0[\Phi_c] &= \frac{1}{2i}\int \mathrm{d}^4x\mathrm{d}^4y\Phi_c(x)\left(\eta^{\mu\nu}\partial^x_\mu\partial^x_\nu - m^2\right)D_{\mathrm{F}}(x-y) \\ &\quad \times\left(\eta^{\mu\nu}\partial^y_\mu\partial^y_\nu - m^2\right)\Phi_c(y). \qquad (557)\end{aligned}$$

Finally, using the fact that D_{F}/i is the Green function satisfying $-\eta^{\mu\nu}\partial_\mu\partial_\nu D_{\mathrm{F}} + m^2 D_{\mathrm{F}} = \delta$, one obtains

$$\begin{aligned}\Gamma_0[\Phi_c] &= \frac{1}{2}\int \mathrm{d}^4x\Phi_c(x)\left(\eta^{\mu\nu}\partial_\mu\partial_\nu - m^2\right)\Phi_c(x) \\ &= -\frac{1}{2}\int \mathrm{d}^4x\left(\eta^{\mu\nu}\partial_\mu\Phi_c\partial_\nu\Phi_c + m^2\Phi_c\right), \qquad (558)\end{aligned}$$

and one recovers the action of a free scalar field theory. In presence of interactions, $\Gamma[\Phi_c]$ will no longer agree with the classical action. The quantum corrections will transform it into a complicated non local functional. In this situation, one typically expects an expression that can be written as

$$\begin{aligned}\Gamma[\Phi_c] &= -\int \mathrm{d}^4x \Big[\frac{1}{2}A(\Phi_c)\eta^{\mu\nu}\partial_\mu\Phi_c\partial_\nu\Phi_c \\ &+ V_{\text{eff}}(\Phi_c) + B(\Phi_c)\left(\eta^{\mu\nu}\partial_\mu\Phi_c\partial_\nu\Phi_c\right)^4 + \cdots\Big],\end{aligned} \tag{559}$$

where $A(\Phi_c)$ and $B(\Phi_c)$ are functions (not functionals) of Φ_c and the dots represent higher derivative terms. Clearly, the previous considerations justify the name "effective action" for $\Gamma[\Phi_c]$.

Another interesting aspect is that the effective action is also the generating functional of the irreducible correlation functions $\Gamma^{(n)}$ (also known as proper vertices). In other words, one has

$$\Gamma^{(n)}(x_1,\cdots,x_n) = \left.\frac{\delta^n\Gamma[\Phi_c]}{\delta\Phi_c(x_1)\cdots\delta\Phi_c(x_n)}\right|_{\Phi_c=0}. \tag{560}$$

Equivalently, one can express the effective action as a Volterra expansion, namely

$$\begin{aligned}\Gamma[\Phi_c] &= \sum_{n=0}^{\infty}\frac{1}{n!}\int \mathrm{d}^4x_1\cdots\mathrm{d}^4x_n\Gamma^{(n)}(x_1,\cdots,x_n) \\ &\times\Phi_c(x_1)\cdots\Phi_c(x_n),\end{aligned} \tag{561}$$

which is nothing but another way to write the functional (559).

Now, suppose that we are interested in determining the effective action concretely. It is clear that this is a complicated calculation. As a first step, it is worth calculating the effective potential in Eq. (559). It is true that it does not represent all the information contained in the effective action but this would allow us to determine the location of the minimum of the system in presence of quantum corrections. Therefore, this is an interesting quantity. For this purpose, it is sufficient to consider a case where Φ_c is constant. Indeed, in this situation, the effective action reduces to $\Gamma = -VV_{\text{eff}}(\Phi_c)$, where V is the space-time volume. On the other hand, the Volterra expansion can be re-expressed as

$$\Gamma[\Phi_c] = \sum_{n=0}^{\infty}\frac{1}{n!}\int \mathrm{d}^4x_1\cdots\mathrm{d}^4x_n\Gamma^{(n)}(x_1,\cdots,x_n)\Phi_c^n. \tag{562}$$

This expression can be further simplified if one takes the Fourier transform of the proper vertex,

$$\begin{aligned}\Gamma^{(n)}(x_1,\cdots,x_n) &= \int\frac{\mathrm{d}^4k_1}{(2\pi)^4}\cdots\frac{\mathrm{d}^4k_n}{(2\pi)^4}\Gamma^{(n)}(k_1,\cdots,k_n) \\ &\times\mathrm{e}^{-ik_1\cdot x_1}\cdots\mathrm{e}^{-ik_n\cdot x_n}.\end{aligned} \tag{563}$$

In fact, for convenience, we introduce the coefficients $\overline{\Gamma}^{(n)}(k_1,\cdots,k_n)$ that are easier to manipulate. They are defined by the following expression

$$\Gamma^{(n)}(k_1,\cdots,k_n) \equiv (2\pi)^4\delta(k_1+\cdots k_n)\,\overline{\Gamma}^{(n)}(k_1,\cdots k_n), \tag{564}$$

Then, we insert the Fourier expansion (563) into the expression (562) of the effective action. The integrals over space of the exponentials lead to Dirac functions in momentum. As a consequence, one arrives at the following expression

$$\Gamma[\Phi_c] = V\sum_{n=0}^{\infty}\frac{1}{n!}\overline{\Gamma}^{(n)}(\boldsymbol{k}_i=0)\Phi_c^n, \tag{565}$$

where, as already mentioned, $V = (2\pi)^4\delta(0)$ is the space-time volume. One finally reaches the result that

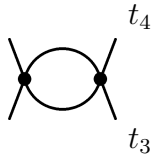
$$V_{\rm eff}(\phi_{\rm c}) = -\sum_{n=0}^{\infty}\frac{1}{n!}\overline{\Gamma}^{(n)}(\boldsymbol{k}_i = 0)\Phi_{\rm c}^n. \tag{566}$$

Therefore, in order to calculate the effective potential, one just has to evaluate the proper vertices with vanishing momenta. At one loop, this gives rise to the following sum

$V_{\rm eff} =$ $+\cdots$. (567)

It turns out that this series can be summed up as we are now going to show.

In order to see how it works, let us calculate the second diagram in the above expansion. For this purpose, we start with the following diagram

$\mathcal{D} =$ 

(568)

According to the Feynman rules (402) and (403), this diagram is equivalent to the expression

$$\begin{aligned}\mathcal{D} &= (-i\lambda)^2\int {\rm d}u{\rm d}v D(t_1-u)D(t_2-u)D^2(u-v)\\ &\quad\times D(v-t_3)D(v-t_4),\end{aligned} \tag{569}$$

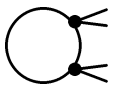
where u and v denotes the two internal vertices. Upon using Eq. (451), this expression takes the form [here, we use the quantum-mechanical theory introduced in Sec. 24.5.1 where the "mass" is denoted ω]

$$\begin{aligned}\mathcal{D} &= (-i\lambda)^2\int {\rm d}u{\rm d}v\int\frac{{\rm d}E_1}{2\pi}\frac{{\rm d}E_2}{2\pi}\frac{{\rm d}E_3}{2\pi}\frac{{\rm d}E_4}{2\pi}\frac{{\rm d}E}{2\pi}\frac{{\rm d}\tilde{E}}{2\pi}\frac{i{\rm e}^{-iE_1(t_1-u)}}{E_1^2-\omega^2}\frac{i{\rm e}^{-iE_2(t_2-u)}}{E_2^2-\omega^2}\frac{i{\rm e}^{-iE_3(v-t_3)}}{E_3^2-\omega^2}\\ &\quad\times\frac{i{\rm e}^{-iE_4(v-t_4)}}{E_4^2-\omega^2}\frac{i{\rm e}^{-iE(u-v)}}{E^2-\omega^2}\frac{i{\rm e}^{-i\tilde{E}(u-v)}}{\tilde{E}^2-\omega^2}.\end{aligned} \tag{570}$$

The integrals over u and v are easily performed and leads to two Dirac functions. Then, a further integration over $\tilde{E}$ can be performed and we are left with the expression

$$\begin{aligned}\mathcal{D} &= (-i\lambda)^2(2\pi)^2\delta(E_1+E_2-E_3-E_4)\int\frac{{\rm d}E_1}{2\pi}\frac{{\rm d}E_2}{2\pi}\frac{{\rm d}E_3}{2\pi}\frac{{\rm d}E_4}{2\pi}\frac{{\rm d}E}{2\pi}\frac{i{\rm e}^{-iE_1t_1}}{E_1^2-\omega^2}\frac{i{\rm e}^{-iE_2t_2}}{E_2^2-\omega^2}\frac{i{\rm e}^{-iE_3t_3}}{E_3^2-\omega^2}\frac{i{\rm e}^{-iE_4t_4}}{E_4^2-\omega^2}\\ &\quad\times\frac{i}{E^2-\omega^2}\frac{i}{(E_3+E_4-E)^2-\omega^2}.\end{aligned} \tag{571}$$

In order to compute the expression (567) (or rather the contribution of the second diagram to this expression), we have seen that we must amputate the diagram by appending the external leg propagators. This amounts to ignore the factors $i/(E_i^2-\omega^2)$ in the above expression. Moreover, since the graph is given by the Fourier transform of $\mathcal{D}$, this means that only the expression inside the above quadruple integrals over E_i should be considered. Finally, the "momentum" (in the present context the energy) of the external legs should be put to zero (*i.e.* we require all the E_i's left in the expression of $\mathcal{D}$ to vanish) and we must factorize out a Dirac function expressing energy conservation [times $(2\pi)^2$ since we are considering a $0+1$ theory]. This means that Eq. (571) takes the following form

 $= (-i\lambda)^2\int\frac{{\rm d}E}{2\pi}\frac{i}{E^2-\omega^2}\frac{i}{E^2-\omega^2}$. (572)

The fact that the propagator appears squared is of course not fortuitous. It comes from the fact that we have two "double legs" attached to the loop. For n "double legs", it is clear that we would have the propagator to the power n and an overall factor $(-i\lambda)^n$. As a consequence, at one loop, Eq. (566) can be expressed as

$$V_{\rm eff}\left(\Phi_{\rm c}\right) = -\sum_{p=1}^{\infty}\frac{\Phi_{\rm c}^{2p}}{2p}\left(\frac{\lambda 4!}{2}\right)^p\int\frac{{\rm d}E}{2\pi}\frac{1}{(E^2-\omega^2)^p}, \tag{573}$$

where we have taken into account a symmetry factor and where we have noted $n = 2p$, p being the number of external points. The sum can now be performed and we obtain

$$V_{\text{eff}}(\Phi_c) = -\frac{1}{2}\int_{-\infty}^{\infty}\frac{dE}{2\pi}\ln\left(1+\frac{12\lambda\Phi_c^2}{E^2-\omega^2}\right) \tag{574}$$

$$= \frac{1}{2}\left(\omega-\sqrt{12\lambda\Phi_c^2-\omega^2}\right) \tag{575}$$

$$= \frac{\hbar}{2}\left(-m+\sqrt{m^2+12\lambda\Phi_c^2}\right), \tag{576}$$

where we have re-established $\hbar$ and where we have taken into account the sign difference for ω (or m) due to the fact that we work with a $(-+++)$ signature, see the remark after Eq. (518). This expression should be compared to Eq. (538). In fact the first term just comes from the fact that the method of the effective potential assumes that $V_{\text{eff}}(0) = 0$. Therefore, in order to compare with the Gaussian effective potential, one must ignore this rescaling. As a consequence, the total potential (including the leading order in $\hbar^0$) reads

$$V_{\text{eff}}(\Phi_c) = \frac{1}{2}m^2\Phi_c^2 + \frac{\hbar}{2}\sqrt{m^2+12\lambda\Phi_c^2}. \tag{577}$$

One sees that the vacuum energy is predicted to be $V_{\text{eff}}(0) = \hbar m/2$ a result which is not realistic in the non perturbative regime.

In fact the comparison can be made in an even more explicit manner if one remembers that the one loop Gaussian effective potential consists in retaining the dominant term is a systematic expansion in $\hbar$. As a consequence, one should ignore the $\hbar^2$ term in Eq. (536) and the $\hbar$ term in Eq. (537). In particular, this last approximation leads to the following explicit solution for the gap equation

$$\overline{\Omega}^2 = m^2 + 12\lambda\Phi_0^2 = \frac{d^2V}{d\Phi_0^2}. \tag{578}$$

If we then use Eq. (538) and neglects the $\hbar^2$ term in that equation, one obtains

$$V_{\text{G}}(\Phi_0) = V(\Phi_0) + \frac{\hbar}{2}\sqrt{\frac{d^2V}{d\Phi_0^2}} + \mathcal{O}(\hbar^2) \tag{579}$$

$$= \frac{1}{2}m^2\Phi_0^2 + \frac{\hbar}{2}\sqrt{m^2+12\lambda\Phi_0^2} + \cdots, \tag{580}$$

that is to say exactly Eq. (577). We conclude that the effective potential is in fact contained in the Gaussian effective potential approach. It just consists in neglecting higher power of $\hbar$ since this is a one loop approach. As a consequence, the effective potential cannot be reliable in the non perturbative regime where the quantum effects are strong.

Having demonstrated that the Gaussian effective potential approach is an efficient tool, we now need to study how it works in quantum field theory. Then, we will able to use it in order to calculate the energy density of the vacuum state.

24.6.3 The Gaussian Effective Potential in Field Theory and the Vacuum Energy

We now turn to the main question of this section, namely the calculation of the Gaussian effective potential in quantum field theory [59, 60, 63, 65]. For this purpose, we now consider a model similar to the one considered in the previous subsections, namely a free model plus a quartic self interaction

$$\mathcal{H} = \frac{1}{2}\dot{\Phi}^2 + \frac{1}{2}\delta^{ij}\partial_i\Phi\partial_j\Phi + \frac{1}{2}m_{\text{B}}^2\Phi^2 + \lambda_{\text{B}}\Phi^4, \tag{581}$$

where $m_{\rm B}$ and $\lambda_{\rm B}$ are the bare mass and coupling constant, respectively. Then, we write the field operator as (we work in d space-time dimensions)

$$\begin{aligned}\Phi\left(t,\boldsymbol{x}\right) &= \Phi_0+\frac{1}{\left(2\pi\right)^{(d-1)/2}}\int\frac{\mathrm{d}^{d-1}\boldsymbol{k}}{\sqrt{2\omega_\Omega(k)}}\Big(c_{\boldsymbol{k}}\mathrm{e}^{-i\omega t+i\boldsymbol{k}\cdot\boldsymbol{x}}\\ &\quad +c^\dagger_{\boldsymbol{k}}\mathrm{e}^{i\omega t-i\boldsymbol{k}\cdot\mathbf{x}}\Big),\end{aligned}\tag{582}$$

with

$$\omega_\Omega(k)\equiv\sqrt{k^2+\Omega^2}.\tag{583}$$

The quantity Φ_0 represents a classical and constant field. The parameter Ω denotes the mass of the excitations around this classical value. The idea at the basis of the calculation is the same as before, namely minimizing the energy with respect to the parameter Ω. Using the properties of the creation and annihilation operators, it is straightforward to calculate the mean value of the Hamiltonian in the vacuum state. We find

$$\begin{aligned}\langle\mathcal{H}\rangle\left(\Omega,\Phi_0\right) &= I_1(\Omega)+\frac{1}{2}\left(m_{\rm B}^2-\Omega^2\right)I_0(\Omega)+\frac{1}{2}m_{\rm B}^2\Phi_0^2\\ &+ \lambda_{\rm B}\Phi_0^4+6\lambda_{\rm B}\Phi_0^2I_0(\Omega)+3\lambda_{\rm B}I_0^2(\Omega),\end{aligned}\tag{584}$$

where we have defined the integrals $I_p(\Omega)$ by

$$I_p(\Omega)=\frac{1}{\left(2\pi\right)^{d-1}}\int\frac{\mathrm{d}^{d-1}\boldsymbol{k}}{2\omega_\Omega(k)}\omega_\Omega^{2p}(k).\tag{585}$$

Of course, the integrals $I_p(\Omega)$ can be divergent and this signals, as usual in field theory, the need for renormalization. In fact, these integrals can be performed exactly and one is led to

$$I_p(\Omega)=\frac{1}{2(4\pi)^{(d-1)/2}}\frac{\Gamma[(2-d-2p)/2]}{\Gamma[(1-2p)/2]}\Omega^{d+2p-2}.\tag{586}$$

It is clear that, for some values of d and p, the above expression does not exist and is only formal.

As already mentioned, in order to calculate the Gaussian effective potential, one must vary $\langle\mathcal{H}\rangle$ with respect to Ω. Upon using Eq. (584) and $\mathrm{d}I_p/\mathrm{d}\Omega=(2p-1)\Omega I_{p-1}$, one arrives at the following "gap" equation, the field theory counterpart of Eq. (537)

$$\Omega^2=m_{\rm B}^2+12\lambda_{\rm B}\left[\Phi_0^2+I_0\left(\Omega\right)\right].\tag{587}$$

This is an algebraic equation for Ω. Its solution can be written as $\overline{\Omega}=\overline{\Omega}\left(\Phi_0;m_{\rm B},\lambda_{\rm B}\right)$, that is to say it depends on Φ_0 and on the bare parameters (in the following we no longer write the dependence on the bare parameters). Then, the Gaussian effective potential is defined by inserting this solution into Eq. (584). This leads to

$$V_{\rm G}(\Phi_0)\equiv\langle\mathcal{H}\rangle\left[\overline{\Omega}(\Phi_0),\Phi_0\right].\tag{588}$$

Upon using the gap equation, one can also find a more compact expression, namely

$$\begin{aligned}V_{\rm G}(\Phi_0) &= I_1\left[\overline{\Omega}\left(\Phi_0\right)\right]-3\lambda_{\rm B}I_0^2\left[\overline{\Omega}\left(\Phi_0\right)\right]+\frac{1}{2}m_{\rm B}^2\Phi_0^2\\ &+\lambda_{\rm B}\Phi_0^4.\end{aligned}\tag{589}$$

At this level, this expression is in fact still formal because of the divergences of the integrals I_p. Therefore, we must first tame those divergences. This is done by means of a renormalization of the parameters appearing in the Lagrangian of the model.

Renormalization consists in parametrizing the physical quantities in terms of new, observable parameters, different from the bare parameters that appear in the Lagrangian. We now define the renormalized mass by

$$m_{\rm R}^2 = \left.\frac{{\rm d}^2 V_{\rm G}}{{\rm d}\Phi_0^2}\right|_{\Phi_0=0}. \tag{590}$$

Upon using Eq. (589) and the following relation [which comes from differentiating Eq. (587)]

$$\frac{{\rm d}\overline{\Omega}}{{\rm d}\Phi_0} = \frac{\Phi_0}{\overline{\Omega}}\frac{12\lambda_{\rm B}}{1+6\lambda_{\rm B} I_{-1}(\overline{\Omega})}, \tag{591}$$

one finds that

$$\begin{aligned}\frac{{\rm d}^2 V_{\rm G}}{{\rm d}\Phi_0^2} &= m_{\rm B}^2 + 12\lambda_{\rm B}\left[\Phi_0^2 + I_0(\overline{\Omega})\right] \\ &\quad -\frac{(12\lambda_{\rm B}\Phi_0)^2 I_{-1}(\overline{\Omega})}{1+6\lambda_{\rm B} I_{-1}(\overline{\Omega})}.\end{aligned} \tag{592}$$

As a consequence, putting $\Phi_0 = 0$ in the above equation, the renormalized mass can be expressed as

$$m_{\rm R}^2 = m_{\rm B}^2 + 12\lambda_{\rm B} I_0(\overline{\Omega}_0), \tag{593}$$

where $\overline{\Omega}_0$ denotes the solution of the gap equation (587) when $\Phi_0 = 0$, *i.e.* $\Omega^2 = m_{\rm B}^2 + 12\lambda_{\rm B} I_0(\Omega)$. It is obvious that $\overline{\Omega}_0$ is not a function of Φ_0 but depends only on $m_{\rm B}$ and $\lambda_{\rm B}$. In fact, this can also been seen from Eq. (587) since it tells us that $m_{\rm R}^2 = \overline{\Omega}_0^2$. As a consequence, the above equation can be re-written as

$$m_{\rm B}^2 = m_{\rm R}^2 - 12\lambda_{\rm B} I_0(m_{\rm R}). \tag{594}$$

We are now in a position where we can calculate the Gaussian effective potential in terms of the renormalized mass. For this purpose, we restart from Eq. (588) and substitute the expression (594). One obtains

$$\begin{aligned}V_{\rm G}(\Phi_0) &= I_1(\overline{\Omega}) + \frac{1}{2}\left(m_{\rm R}^2 - \overline{\Omega}^2\right) I_0(\overline{\Omega}) + \frac{1}{2} m_{\rm R}^2 \Phi_0^2 \\ &\quad +\lambda_{\rm B}\Phi_0^4 + 6\lambda_{\rm B}\left[I_0(\overline{\Omega}) - I_0(m_{\rm R})\right]\Phi_0^2 \\ &\quad +3\lambda_{\rm B} I_0^2(\overline{\Omega}) - 6\lambda_{\rm B} I_0(\overline{\Omega})\, I_0(m_{\rm R}).\end{aligned} \tag{595}$$

However, at this stage, the expression of the Gaussian effective potential is not yet a function of the renormalized mass only as one still needs to express $I_0(\overline{\Omega})$ and $I_1(\overline{\Omega})$ in terms of $I_0(m_{\rm R})$ and $I_1(m_{\rm R})$. There are several ways to do it, see Ref. [59]. Here we choose to use a method where a cut-off is temporarily introduced. We work in $d = 4$ dimensions to simplify the calculations. Let us exemplify the method with I_0. We have

$$I_0(\Omega) = \frac{1}{2(2\pi)^3}\int {\rm d}^3\boldsymbol{k}\,\omega_\Omega^{-1} \tag{596}$$

$$= \frac{1}{4\pi^2}\int_0^M {\rm d}k k^2\left(k^2+\Omega^2\right)^{-1/2}, \tag{597}$$

where we have performed the angular integrals and where we have introduced a cut-off M to regulate the otherwise divergent integral. This gives

$$\begin{aligned}I_0(\Omega) &= \frac{1}{4\pi^2}\left[\frac{M}{2}\sqrt{\Omega^2+M^2}\right. \\ &\quad \left. -\frac{\Omega^2}{2}\ln\left(M+\sqrt{\Omega^2+M^2}\right) + \frac{\Omega^2}{2}\ln\Omega\right].\end{aligned} \tag{598}$$

As a consequence, one can write

$$I_0(\Omega) - I_0(m) = \frac{1}{2}\left(m^2-\Omega^2\right) I_{-1}(m) + \frac{1}{4\pi^2}\left[\frac{M}{2}\sqrt{\Omega^2+M^2} - \frac{M}{2}\sqrt{m^2+M^2} + \frac{\Omega^2}{2}\ln\left(\frac{M+\sqrt{m^2+M^2}}{M+\sqrt{\Omega^2+M^2}}\right) + \frac{\Omega^2}{2}\ln\Omega - \frac{m^2}{2}\ln m + \frac{1}{2}\left(m^2-\Omega^2\right)\left(\frac{M}{\sqrt{m^2+M^2}} + \ln m\right)\right], \tag{599}$$

where we have used

$$I_{-1}(\Omega) = \frac{1}{4\pi^2}\left[\frac{-M}{\sqrt{\Omega^2+M^2}} + \ln\left(M+\sqrt{\Omega^2+M^2}\right) - \ln\Omega\right], \tag{600}$$

an expression which can be obtained with the same method as described above. Then we take the limit $M \to +\infty$ and Eq. (599) becomes

$$I_0(\Omega) - I_0(m) = -\frac{1}{2}\left(\Omega^2-m^2\right) I_{-1}(m) + \frac{m^2}{16\pi^2}\left[x\ln x - (x-1)\right], \tag{601}$$

where $x \equiv \Omega^2/m^2$. Notice that, at this stage of the calculation, there is no need to write x in terms of $\overline{\Omega}$ and/or $m_{\rm R}$. The above expression is valid for any Ω and m. Following Ref. [59], we call the function between the square brackets in the above equation $L_2(x)$. Exactly in the same manner, one can also show that

$$I_1(\Omega) - I_1(m) = \frac{1}{2}\left(\Omega^2-m^2\right) I_0(m) - \frac{1}{8}\left(\Omega^2-m^2\right)^2 I_{-1}(m) + \frac{m^4}{32\pi^2}L_3(x), \tag{602}$$

with $L_3(x) = [2x^2\ln x - 2(x-1) - 3(x-1)^2]/4$. Then, upon using Eqs. (601) and (602) in Eq. (595), one obtains

$$V_{\rm G}(\Phi_0) = I_1\left(m_{\rm R}\right) - 3\lambda_{\rm B} I_0^2\left(m_{\rm R}\right) + \frac{1}{2}m_{\rm R}^2\Phi_0^2 + \lambda_{\rm B}\Phi_0^4 + \frac{m_{\rm R}^4}{32\pi^2}L_3(\overline{x}) + \frac{3}{4}\lambda_{\rm B} m_{\rm R}^4\left(\overline{x}-1\right)^2 I_{-1}^2\left(m_{\rm R}\right) + \frac{1}{8}I_{-1}\left(m_{\rm R}\right) m_{\rm R}^4(\overline{x}-1)\left\{(\overline{x}-1) - \frac{3\lambda_{\rm B}}{2\pi^2}\left[L_2(\overline{x}) + \frac{16\pi^2}{m_{\rm R}^2}\Phi_0^2\right]\right\} - \frac{m_{\rm R}^4}{32\pi^2}L_2(\overline{x})\left\{(\overline{x}-1) - \frac{3\lambda_{\rm B}}{8\pi^2}\left[L_2(\overline{x}) + 32\pi^2\frac{\Phi_0^2}{m_{\rm R}^2}\right]\right\}, \tag{603}$$

with $\overline{x} \equiv \overline{\Omega}^2/m_{\rm R}^2$. Finally, the gap equation (587) can also be re-written in terms of the renormalized mass. The corresponding expression reads

$$(\overline{x}-1)\left[1 + 6\lambda_{\rm B} I_{-1}\left(m_{\rm R}\right)\right] = \frac{3\lambda_{\rm B}}{4\pi^2}\left[L_2(\overline{x}) + 16\pi^2\frac{\Phi_0^2}{m_{\rm R}^2}\right]. \tag{604}$$

We see that the dependence on Φ_0 of the Gaussian effective potential is quite complicated. Indeed, solving the gap equation provides us with $\overline{x} = \overline{x}(\Phi_0)$ that we must then insert in Eq. (603).

We shall now discuss the renormalization of the coupling constant $\lambda_{\rm B}$. It is defined by

$$\lambda_{\rm R} = \frac{1}{4!}\frac{{\rm d}^4 V_{\rm G}}{{\rm d}\Phi_0^4}\bigg|_{\Phi_0=0}. \tag{605}$$

Proceeding in the same manner as before, one obtains the following equation for the quantity $\lambda_{\rm R}$

$$\lambda_{\rm R} = \lambda_{\rm B}\frac{1-12\lambda_{\rm B}I_{-1}(m_{\rm R})}{1+6\lambda_{\rm B}I_{-1}(m_{\rm R})}. \tag{606}$$

The renormalization of the coupling constant has been discussed at length in Refs. [59, 63] and we will not repeat all the details since, here, our main concern is the vacuum energy. In fact we will focus on the case named "precarious" in Ref. [59]. Let us return to Eq. (606); one can invert this expression and express the bare coupling constant in terms of the renormalized one. One obtains

$$\begin{aligned}\lambda_{\rm B} &= \frac{1-6\lambda_{\rm R}I_{-1}(m_{\rm R})}{24I_{-1}(m_{\rm R})} \\ &\times\left\{1\pm\sqrt{1-\frac{48\lambda_{\rm R}I_{-1}(m_{\rm R})}{[1-6\lambda_{\rm R}I_{-1}(m_{\rm R})]^2}}\right\}.\end{aligned} \tag{607}$$

Then, from Eq. (600), we see that $I_{-1}(m_{\rm R})$ is in fact a logarithmically divergent quantity when the cut-off M is sent to infinity. Therefore, we must treat $I_{-1}(m_{\rm R})$ as an arbitrarily large quantity. In this limit, if one chooses the plus sign in Eq. (607), one obtains

$$\lambda_{\rm B} = -\frac{1}{2}\lambda_{\rm R} + \mathcal{O}\left[\frac{1}{I_{-1}(m_{\rm R})}\right]. \tag{608}$$

As shown in Ref. [59], this case is not viable. Intuitively, it is clear that a large negative coupling constant leads to a potential which is not bounded from below. Therefore, we are left with the minus sign for which one finds that

$$\begin{aligned}\lambda_{\rm B} &= -\frac{1}{6I_{-1}(m_{\rm R})}\left\{1+\frac{1}{2\lambda_{\rm R}I_{-1}(m_{\rm R})}\right. \\ &\left.+\mathcal{O}\left[\frac{1}{I_{-1}^2(m_{\rm R})}\right]\right\}.\end{aligned} \tag{609}$$

In passing, it is worth signaling that this equation can be re-expressed in a different manner as follows. Let us first introduce the new characteristic scale $M_{\rm C}$ (not to be confused with the cut-off M discussed before) defined by

$$\frac{1}{\lambda_{\rm R}} \equiv -\frac{1}{4\pi^2}\ln\left(\frac{m_{\rm R}^2}{M_{\rm C}^2}\right). \tag{610}$$

Then, if we use the fact that

$$I_{-1}(M_{\rm C}) - I_{-1}(m_{\rm R}) = -\frac{1}{8\pi^2}L_1(x), \tag{611}$$

where $L_1(x) = \ln x$ [an expression which can be derived with the same method that has led to Eqs. (599) and (602)], the bare coupling constant can be re-written as

$$\lambda_{\rm B} = -\frac{1}{6I_{-1}(M_{\rm C})}. \tag{612}$$

Therefore, it has a negative but infinitesimal value which implies that the stability properties of the corresponding scenario are much better compared to the case (608). As argued in Ref. [59], besides simplifying the expression of the renormalized coupling constant, the scale $M_{_C}$ plays an important role in a complete study of the precarious case (even if, in the present context, this is not apparent since we focus on the vacuum energy only). Inserting the expression (609) [or Eq. (612)] into Eq. (603), one can re-write the effective potential as

$$\begin{aligned} V_{_G}(\Phi_0) &= I_1\left(m_{_R}\right) - 3\lambda_{_B} I_0^2\left(m_{_R}\right) + \frac{1}{2} m_{_R}^2 \overline{x} \Phi_0^2 \\ &+ \frac{m_{_R}^4}{32\pi^2} L_3(\overline{x}) - \frac{m_{_R}^4}{16\lambda_{_R}} \left(\overline{x}-1\right)^2 \\ &+ \mathcal{O}\left[\frac{1}{I_{-1}(m_{_R})}\right], \end{aligned} \tag{613}$$

while the gap equation (604) becomes

$$\overline{x} - 1 = \frac{\lambda_{_R}}{4\pi^2}\left[L_2\left(\overline{x}\right) + 16\pi^2 \frac{\Phi_0^2}{m_{_R}^2}\right]. \tag{614}$$

Let us notice that, in Eq. (613), we have left the quantity $\lambda_{_B}$ into the field independent term $I_1\left(m_{_R}\right) - 3\lambda_{_B} I_0^2\left(m_{_R}\right)$ since it represents the zero point energy that will be treated in detail in the next sub-section 24.6.4. Finally, if one uses the explicit form of the functions L_2 and L_3, then the Gaussian effective potential takes the form

$$\begin{aligned} V_{_G}(\Phi_0) &= I_1\left(m_{_R}\right) - 3\lambda_{_B} I_0^2\left(m_{_R}\right) + \frac{1}{2} m_{_R}^2 \overline{x} \Phi_0^2 \\ &+ \frac{m_{_R}^4}{128\pi^2}\Bigg[2\overline{x}^2 \ln \overline{x} - 2(\overline{x}-1) - 3(\overline{x}-1)^2 \\ &- 2\frac{4\pi^2}{\lambda_{_R}}(\overline{x}-1)^2\Bigg], \end{aligned} \tag{615}$$

and the gap equation can be re-expressed as

$$(\overline{x}-1)\left(1 + \frac{4\pi^2}{\lambda_{_R}}\right) - 16\pi^2 \frac{\Phi_0^2}{m_{_R}^2} = \overline{x} \ln \overline{x}. \tag{616}$$

The two equations (615) and (616) give a complete expression of the Gaussian effective potential for the renormalized precarious "$\lambda\Phi^4$" theory [59, 63].

As we have done for the quantum mechanics case, it is interesting to compare the above result to the standard effective potential approach. We know that this potential can be obtained from the Gaussian effective potential method by neglecting the terms proportional to $\hbar^2$. Therefore, we restart from Eq. (584) and neglect the term $3\lambda_{_B} I_0^2$ since this corresponds to the standard one loop method (*i.e.* it is of order $\hbar^2$). One obtains

$$\begin{aligned} \langle \mathcal{H} \rangle (\Omega, \Phi_0) &= I_1(\Omega) + \frac{1}{2}\left(m_{_B}^2 - \Omega^2\right) I_0(\Omega) + \frac{1}{2} m_{_B}^2 \Phi_0^2 \\ &+ \lambda_{_B} \Phi_0^4 + 6\lambda_{_B} \Phi_0^2 I_0(\Omega) + \mathcal{O}(\hbar^2). \end{aligned} \tag{617}$$

As a consequence, this leads to the new following gap equation

$$\Omega^2 = m_{_B}^2 + 12\lambda_{_B} \Phi_0^2. \tag{618}$$

Clearly, this relation is the exact counter-part of Eq. (578). Then, one has to calculate the renormalized mass and coupling constant. one arrives at

$$m_{_R}^2 = m_{_B}^2 + 12\lambda_{_B} I_0(m_{_B}), \tag{619}$$

$$\lambda_{_R} = \lambda_{_B}\left[1 - 18\lambda_{_B} I_{-1}(m_{_B})\right]. \tag{620}$$

These equations should be compared to Eqs. (594) and (606). We notice that it is now $m_{\rm B}$ that appears in the argument of the integrals rather than $m_{\rm R}$ as in Eqs. (594) and (606). This is because Eq. (618) implies that $\overline{\Omega}(\Phi_0 = 0) = m_{\rm B}$ and not $\overline{\Omega}(\Phi_0 = 0) = m_{\rm R}$ as before. The final step consists in using the two previous equation in Eq. (617). One obtains

$$V_{\rm G}(\Phi_0) = I_1(m_{\rm B}) + \frac{1}{2}m_{\rm R}^2\Phi_0^2 + \lambda_{\rm R}\Phi_0^4 + \frac{m_{\rm B}^4}{64\pi^2}\left[\left(1+12\lambda_{\rm B}\frac{\Phi_0^2}{m_{\rm B}^2}\right)^2 \ln\left(1+12\lambda_{\rm B}\frac{\Phi_0^2}{m_{\rm B}^2}\right) - 12\lambda_{\rm B}\frac{\Phi_0^2}{m_{\rm B}^2} - 216\lambda_{\rm B}^2\frac{\Phi_0^4}{m_{\rm B}^4}\right] + \cdots, \tag{621}$$

the dots corresponding to the higher order terms in $\hbar$. If one neglects those terms, then $V_{\rm G}$ reduces to $V_{\rm eff}$. The corresponding expression is nothing but the Coleman-Weinberg potential. The previous analysis confirms that the effective potential is indeed a particular case of the Gaussian effective potential approach. An even more detailed comparisons of the two methods can be found in Refs. [59, 63]. Here, we do not pursue this issue and now come to the question of the vacuum energy problem.

24.6.4 Vacuum Energy and Gaussian Effective Potential

We are now in a position where the main problem of this section can be addressed. From Eqs. (603) and (613), we see that the vacuum energy density predicted by the Gaussian effective potential method (*i.e.* the term left after having taken $\Phi_0 = 0$ in those expressions) can be expressed as

$$\rho_{\rm vac} = I_1(m_{\rm R}) - 3\lambda_{\rm B} I_0(m_{\rm R}). \tag{622}$$

As we have already stressed many times, it is important to remember that this result is a priori non perturbative. Using the expression of the precarious coupling constant (609), one obtains

$$\rho_{\rm vac} = I_1(m_{\rm R}) + \frac{1}{2}\frac{I_0(m_{\rm R})}{I_{-1}(m_{\rm R})}\left[1+\frac{1}{2\lambda I_{-1}(m_{\rm R})}\right] I_0(m_{\rm R}). \tag{623}$$

The expressions of the integrals I_p have been given in Eqs. (585) and (586). The three integrals that appear in the above expression are all divergent. Nevertheless, the ratio I_0/I_{-1} is finite and reads

$$\frac{I_0(m_{\rm R})}{I_{-1}(m_{\rm R})} = \frac{m_{\rm R}^2}{2-d}. \tag{624}$$

Nevertheless, even after having inserted the above ratio into Eq. (623), this expression still contains singular terms. Explicitly, one has

$$\rho_{\rm vac} = \frac{m_{\rm R}^4}{4(2-d)}\left[-m_{\rm R}^{d-4}\frac{4-d}{2\sqrt{\pi}(4\pi)^{(d-1)/2}}\Gamma\left(-\frac{d}{2}\right) + \frac{1}{\lambda_{\rm R}(2-d)}\right]. \tag{625}$$

However, writing this expression for $d = 4+\epsilon$ and taking the limit $\epsilon \to 0$, we have the remarkable result that a second cancellation occurs and that the final expression is in fact finite. It can be expressed as

$$\rho_{\rm vac} = \frac{m_{\rm R}^4}{128\pi^2}\left[1 - 2\ln\left(\frac{m_{\rm R}^2}{M_{\rm C}^2}\right)\right], \tag{626}$$

where we have used Eq. (612). Even more remarkable is that the above equation is very similar to Eq. (392). In particular, the vacuum energy is proportional to fourth power, not of the cut-off, but of the mass of the corresponding particle and the scale $M_{\rm C}$ is present as a logarithmic

correction. Therefore, we conclude that our analysis based on the Gaussian effective potential has totally confirmed the results presented in the previous sections. We end this section by noticing that the vacuum energy predicted by the Coleman Weinberg potential is, see Eq. (621), $\rho_{\text{vac}} = I_1(m_{\text{B}})$ and is therefore different from the one obtained before, see Eq. (626). Moreover, it seems that this expression remains divergent.

24.7 Contribution from Other Fields

So far, for simplicity, we have only treated the case of a scalar field. However, in order to be more realistic, it is clear that one must also evaluate the vacuum energy for other type of fields. This question is the purpose of the present section. As we will see, the corresponding calculations are in fact very similar. In fact, the important question is the dependence of the overall coefficient which multiplies ρ_{vac} with the spin of the particle. As is well-known, this is at the origin of a remarkable cancellation which leads to the concept of super-symmetry. Therefore, in the following, we will pay special attention to this issue.

24.7.1 Fermion Fields

We start with the case of a Dirac spinor field describing a spin 1/2 particle [16, 17, 18, 19, 20, 21, 22]. For convenience, we quickly remind basics fact about the quantum field theory of a spin 1/2 field. Then, we turn to the calculation of ρ_{vac}.

The Lagrangian of the free model is [we treated the mass-less case in Eq. (331)]

$$\mathcal{L}_{\text{Dirac}} = -\overline{\Psi}\left(i\gamma^{\mu}\partial_{\mu} - m\right)\Psi, \tag{627}$$

where γ^{μ} are matrices satisfying

$$\{\gamma^{\mu}, \gamma^{\nu}\} = -2g^{\mu\nu}. \tag{628}$$

The above equation is in fact valid for any metric tensor. Here, we just consider the flat space-time case and, therefore, we will always consider that $g_{\mu\nu} = \eta_{\mu\nu}$. The quantity $\Psi(t, \boldsymbol{x})$ represents a four components Dirac spinor and $\overline{\Psi} \equiv \Psi^{\dagger}\gamma^0$ [this definition was already mentioned just after Eq. (329)]. Here, we use the Dirac representation which means that the γ^{μ} matrices can be taken to be equal to [recall that, in Eq. (327), we used the chiral representation]

$$\begin{aligned} \gamma^0 &= \begin{pmatrix} \mathbb{I}_2 & 0 \\ 0 & -\mathbb{I}_2 \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix}, \\ \gamma^5 &= \begin{pmatrix} 0 & \mathbb{I}_2 \\ \mathbb{I}_2 & 0 \end{pmatrix}, \end{aligned} \tag{629}$$

where σ^i are the standard two-dimensional Pauli matrices. The equation of motion for the spinor Ψ is easily obtained from Eq. (627). Varying the Dirac Lagrangian, it is straightforward to show that

$$\left(i\gamma^{\mu}\partial_{\mu} - m\right)\Psi = 0, \tag{630}$$

which is of course nothing but the Dirac equation. Then, as we have done for the scalar field case, we now expand the Ψ operator in Fourier modes according to

$$\begin{aligned} \Psi(t, \boldsymbol{x}) = \frac{1}{(2\pi)^{3/2}} \int \frac{\mathrm{d}\boldsymbol{k}}{\sqrt{2\omega(k)}} \sum_{r=1}^{2} \Big[& b_{\boldsymbol{k}}^{r} u(\boldsymbol{k}, r) \mathrm{e}^{-i\omega t + i\boldsymbol{k}\cdot\boldsymbol{x}} \\ & + \left(d_{\boldsymbol{k}}^{r}\right)^{\dagger} v(\boldsymbol{k}, r) \mathrm{e}^{i\omega t - i\boldsymbol{k}\cdot\boldsymbol{x}} \Big]. \end{aligned} \tag{631}$$

As appropriate for a spinor field, the operators $b_{\boldsymbol{k}}^{r}$ and $d_{\boldsymbol{k}}^{r}$ satisfy anti-commutation rules. The index r represents the two polarization states. Inserting the expansion (631) into the Dirac equation (630) leads to two equations, namely

$$\left(\gamma^0\omega - \gamma^i k_i - m\right) u(\boldsymbol{k}, r) = 0, \tag{632}$$

$$\left(\gamma^0\omega - \gamma^i k_i + m\right) v(\boldsymbol{k}, r) = 0. \tag{633}$$

Upon using the explicit form (629), the first equation can be re-written as

$$\begin{pmatrix} (m-\omega)\,\mathbb{I}_2 & \sigma^i k_i \\ -\sigma^i k_i & (m+\omega)\,\mathbb{I}_2 \end{pmatrix} u(\boldsymbol{k}, r) = 0. \tag{634}$$

In order to have a non-trivial solution, the determinant of this matrix should vanish and this leads to

$$w(k) = \sqrt{k^2 + m^2}, \tag{635}$$

where we have used the following properties of the Pauli matrices, $\sigma^i k_i \sigma^j k_j = k^2$. On the other hand, Eq. (634) also implies that

$$u(\boldsymbol{k}, r) = \begin{pmatrix} \psi \\ \dfrac{\sigma^i k_i}{m+\omega}\psi \end{pmatrix}, \tag{636}$$

where ψ is an arbitrary two components spinor. Choosing the solution to be an eigenstate of the spin operator along the z-direction (in the rest frame of the particle), the two components of $u(\boldsymbol{k}, r)$ can be expressed as

$$u(\boldsymbol{k}, 1) = \sqrt{m+\omega}\begin{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \dfrac{\sigma^i k_i}{m+\omega}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{pmatrix}, \tag{637}$$

and

$$u(\boldsymbol{k}, 2) = \sqrt{m+\omega}\begin{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ \dfrac{\sigma^i k_i}{m+\omega}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{pmatrix}. \tag{638}$$

It is easy to check that this spinor is normalized according to

$$\overline{u}(\boldsymbol{k}, r) u(\boldsymbol{k}, s) = 2m\delta_{sr}. \tag{639}$$

In the same manner, the solution $v(\boldsymbol{k}, r)$, corresponding to the other branch of the Fourier expansion, can be written as

$$v(\boldsymbol{k}, 1) = \sqrt{m+\omega}\begin{pmatrix} \dfrac{\sigma^i k_i}{m+\omega}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{pmatrix}, \tag{640}$$

and

$$v(\boldsymbol{k}, 2) = \sqrt{m+\omega}\begin{pmatrix} \dfrac{\sigma^i k_i}{m+\omega}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{pmatrix}. \tag{641}$$

The normalization of this spinor satisfies a relation almost identical to that of $u(\boldsymbol{k}, r)$ except for the presence of a minus sign

$$\overline{v}(\boldsymbol{k}, r) v(\boldsymbol{k}, s) = -2m\delta_{sr}. \tag{642}$$

Finally, we also have

$$\overline{u}(\boldsymbol{k}, r) v(\boldsymbol{k}, s) = 0, \tag{643}$$

which can easily be obtained from the above expressions.

All the previous considerations are standard textbooks calculations. We now turn to our main goal, namely the calculation of the vacuum energy. Therefore, the next step consists in calculating the energy momentum tensor. Either using the Noether procedure or a variation of the action with respect to vierbeins lead to the equation

$$T_{\mu\nu} = -\frac{i}{4}\left(\overline{\Psi}\gamma_\mu\partial_\nu\Psi + \overline{\Psi}\gamma_\nu\partial_\mu\Psi - \partial_\mu\overline{\Psi}\gamma_\nu\Psi - \partial_\nu\overline{\Psi}\gamma_\mu\Psi\right). \tag{644}$$

Of course, this expression is symmetric in μ and ν as it should. Endowed with this formula, one can now evaluate the various components of the stress energy tensor in the vacuum state. Let us first calculate the energy density. One has

$$\langle\rho\rangle = -\frac{i}{4}\langle 0\,\big|\left(-2\Psi^\dagger\partial_0\Psi + 2\partial_0\Psi^\dagger\Psi\right)\big|\,0\rangle, \tag{645}$$

since $\gamma^0 = -\gamma_0$ and $(\gamma^0)^2 = \mathbb{I}_4$. Inserting the Fourier expansion (631) into the above expression, one arrives at

$$\langle\rho\rangle = -\frac{1}{(2\pi)^3}\frac{1}{2}\int \mathrm{d}\boldsymbol{k}\sum_{s=1}^{2} v^\dagger(\boldsymbol{k},s)v(\boldsymbol{k},s). \tag{646}$$

This equation can be further simplified. Upon using the explicit form of the spinors given above, one can show that

$$\sum_{s=1}^{2} v^\dagger(\boldsymbol{k},s)v(\boldsymbol{k},s) = 4\omega(k), \tag{647}$$

and, therefore, we obtain the following expression for the vacuum energy

$$\langle\rho\rangle = -\frac{1}{(2\pi)^3}\frac{4}{2}\int \mathrm{d}\boldsymbol{k}\,\omega(k). \tag{648}$$

Several comments are in order at this point. From the above expression, we see that we find minus four times the corresponding result for a scalar field, see Eq. (364). We have a factor four because we have two particles (the particle and its anti-particle) with two polarization state each. Of course, the most striking aspect of the above equation is that $\langle\rho\rangle$ is negative. It is therefore interesting to understand the origin of the minus sign in more details. For this purpose, let us calculate the Hamiltonian operator associated to the Dirac field

$$H = \int \mathrm{d}^3\boldsymbol{x}\, T_{00}. \tag{649}$$

Inserting the Fourier expansion (631) of the spinor Ψ into the above equation, it is straightforward to arrive at [127]

$$H = \int \mathrm{d}\boldsymbol{k}\sum_{s=1}^{2}\omega(k)\left[(b^s_{\boldsymbol{k}})^\dagger\, b^s_{\boldsymbol{k}} - d^s_{\boldsymbol{k}}\,(d^s_{\boldsymbol{k}})^\dagger\right]. \tag{653}$$

However, as we have already recalled, a spinor field is a quantity which anti-commutes and, therefore, one has $\left\{d^s_{\boldsymbol{k}}, \left(d^r_{\boldsymbol{p}}\right)^\dagger\right\} = \delta\left(\boldsymbol{k}-\boldsymbol{p}\right)\delta_{rs}$. As a consequence, the Hamiltonian (653) can be re-written as

$$H = \int \mathrm{d}\boldsymbol{k}\sum_{s=1}^{2}\omega(k)\left[N_b + N_d - \delta(0)\right], \tag{654}$$

[127] For convenience, let us notice that the following relationships are useful for this calculation

$$u^\dagger(\boldsymbol{k},s)u(\boldsymbol{k},r) = 2\omega\delta_{rs}, \tag{650}$$
$$v^\dagger(\boldsymbol{k},s)v(\boldsymbol{k},r) = 2\omega\delta_{rs}, \tag{651}$$
$$v^\dagger(\boldsymbol{k},s)v(-\boldsymbol{k},r) = 0. \tag{652}$$

These three equations are responsible for the fact that only the quantity $\omega(k)$ appears into the expression of the Hamiltonian (653).

where $N_b \equiv (b^s_{\boldsymbol{k}})^\dagger\, b^s_{\boldsymbol{k}}$ and $N_d \equiv (d^s_{\boldsymbol{k}})^\dagger\, d^s_{\boldsymbol{k}}$ are the particle and anti-particle number operators respectively. Of course, their mean value vanishes in the vacuum state which contains no particle. The zero point energy is thus given by the formally infinite term $\delta(0)$. Since this term appears with a minus sign in the Hamiltonian, the corresponding energy density is indeed negative. The origin of this minus sign is the anti-commuting properties of the creation and annihilation operators. We conclude that the vacuum energy density is negative because we deal with fermions which are anti-commuting objects.

Let us now turn to the calculation of the pressure. Upon using Eq. (644), one has

$$\begin{aligned}
\langle p\rangle &= \frac{1}{3}\perp^{\mu\nu} T_{\mu\nu} && (655)\\
&= -\frac{i}{6}g^{00}\big(\overline{\Psi}\gamma_0\partial_0\Psi - \partial_0\overline{\Psi}\gamma_0\Psi\big) - \frac{i}{6}g^{ij}\big(\overline{\Psi}\gamma_i\partial_j\Psi \\
&- 2\partial_i\overline{\Psi}\gamma_j\Psi\big) - \frac{i}{6}u^0u^0\big(\overline{\Psi}\gamma_0\partial_0\Psi - \partial_0\overline{\Psi}\gamma_0\Psi\big) && (656)\\
&= -\frac{i}{6}\delta^{ij}\left(\overline{\Psi}\gamma_i\partial_j\Psi - 2\partial_i\overline{\Psi}\gamma_j\Psi\right). && (657)
\end{aligned}$$

From the above expression, we see that we have two terms to calculate. Using the definition of $\overline{\Psi}$, the first one reads

$$\begin{aligned}
\langle 0\left|\Psi^\dagger\gamma^0\gamma^i\partial_i\Psi\right|0\rangle &= -\frac{i}{(2\pi)^3}\int\frac{\mathrm{d}\boldsymbol{k}}{\sqrt{2\omega(k)}}k_i \\
&\times \sum_{s=1}^{2} v^\dagger(\boldsymbol{k},s)\gamma^0\gamma^i v(\boldsymbol{k},s), && (658)
\end{aligned}$$

while the second one takes the form

$$\langle 0\left|\Psi^\dagger\gamma^0\gamma^i\partial_i\Psi\right|0\rangle = -\langle 0\left|\partial_i\Psi^\dagger\gamma^0\gamma^i\Psi\right|0\rangle. \quad (659)$$

As a consequence, one finds that the vacuum pressure is given by the following expression

$$\begin{aligned}
\langle p\rangle &= -\frac{1}{3}\frac{1}{(2\pi)^3}\int\frac{\mathrm{d}\boldsymbol{k}}{\sqrt{2\omega(k)}} \\
&\times \sum_{s=1}^{2} v^\dagger(\boldsymbol{k},s)\gamma^0\gamma^i k_i v(\boldsymbol{k},s). && (660)
\end{aligned}$$

One must now evaluate the sum present in the above integral. We have seen that the spinors v obeys the Dirac equation $(\gamma^0\omega - \gamma^i k_i + m)\, v(\boldsymbol{k},s) = 0$. Using this last equation, one deduces that

$$\begin{aligned}
v^\dagger(\boldsymbol{k},s)\gamma^0\gamma^i k_i v(\boldsymbol{k},s) &= \omega v^\dagger(\boldsymbol{k},s)v(\boldsymbol{k},s) \\
&+ v^\dagger(\boldsymbol{k},s)\gamma^0 m v(\boldsymbol{k},s). && (661)
\end{aligned}$$

Then, one can sum over the polarization and use the normalization of the spinors established before, see for instance Eqs. (639), (642) and (643), to obtain that

$$\begin{aligned}
&\sum_{s=1}^{2} v^\dagger(\boldsymbol{k},s)\gamma^0\gamma^i k_i v(\boldsymbol{k},s) = 4\omega^2 \\
&+\sum_{s=1}^{2} v^\dagger(\boldsymbol{k},s)\gamma^0 m v(\boldsymbol{k},s). && (662)
\end{aligned}$$

On the other hand, with the explicit expressions (640) and (641), it is easy to directly check that $\sum_{s=1}^{2} v^{\dagger}(\mathbf{k},s)\gamma^0 m v(\mathbf{k},s) = -4m^2$ from which one can establish that

$$\sum_{s=1}^{2} v^{\dagger}(\boldsymbol{k},s)\gamma^0\gamma^i k_i v(\boldsymbol{k},s) = 4\omega^2 - 4m^2 = 4k^2, \tag{663}$$

which is exactly the sum needed to evaluate the pressure. Therefore, our final expression for $\langle p\rangle$ can be written as

$$\langle p\rangle \;=\; -\frac{1}{(2\pi)^3}\frac{4}{6}\int \mathrm{d}\boldsymbol{k}\frac{k^2}{\omega(k)}. \tag{664}$$

We see that, again, we recover exactly the same result as for the scalar field but with the additional multiplicative factor -4, see Eq. (366). This means that all the previous considerations discussed in the scalar field case also apply to the fermion case. In particular, the regularization of the above divergent integrals would proceed in the same way. Introducing a cut-off would break Lorentz invariance and would lead to an incorrect equation of state. On the other hand, using, say, dimensional regularization produces the correct vacuum equation of state and leads to a regularized vacuum energy which is minus four times the regularized scalar field vacuum energy. We do not need to repeat all these calculations here and we now move to the case of a vector field.

24.7.2 Proca Fields

Having treated the case of a spin 1/2 field in the previous sub-section, we now turn to the case of a Proca field [16, 17, 18, 19, 20, 21, 22]. A Proca field $A_\mu(t,\boldsymbol{x})$ is a massive vector, spin 1, field. The corresponding action can be written as

$$S[A_\mu] \equiv -\int \mathrm{d}^4x\sqrt{-g}\left(\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m^2 A_\mu A^\mu\right), \tag{665}$$

where $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$ and m is the mass of the particle. Again, we will restrict our considerations to flat space-time and, therefore, we will take for the metric determinant $g = \eta = 1$.

Before turning to the calculation of the vacuum stress energy momentum, we quickly remind some basics fact about the quantization of this type of fields. Varying with respect to A_μ, one obtains the equation of motion

$$\partial_\mu F^{\mu\nu} - m^2 A^\nu = 0. \tag{666}$$

Taking the derivative of this expressions and noting that $\partial_\nu\partial_\mu F^{\mu\nu} = 0$ (since this is the contraction of a symmetric expression with an anti-symmetric one), one arrives at

$$\partial_\nu A^\nu = 0, \tag{667}$$

provided that $m \neq 0$, which we assume in this section. As a consequence, the equation of motion of the vector field is given by

$$\partial_\nu\partial^\nu A_\mu - m^2 A_\mu = 0, \tag{668}$$

which is nothing but a Klein-Gordon equation for each component of A_μ. As usual, we Fourier expand the field in terms of creation and annihilation operators. This leads to the following expression

$$\begin{aligned} A_\mu(t,\boldsymbol{x}) \;=\; & \frac{1}{(2\pi)^{3/2}}\int\frac{\mathrm{d}\boldsymbol{k}}{\sqrt{2\omega(k)}}\sum_{\alpha=0}^{4}\epsilon_\mu^\alpha(\boldsymbol{k})\Big[a_{\boldsymbol{k}}^\alpha \mathrm{e}^{-i\omega t + i\boldsymbol{k}\cdot\boldsymbol{x}} \\ & + (a_{\boldsymbol{k}}^\alpha)^\dagger\,\mathrm{e}^{i\omega t - i\boldsymbol{k}\cdot\boldsymbol{x}}\Big], \end{aligned} \tag{669}$$

where, as usual, one has $\omega(k) = \sqrt{k^2 + m^2}$. Of course, the operators $a^\alpha_{\boldsymbol{k}}$ satisfy commutation rules since a vector field is a bosonic field. The sum over α is a sum over four vectors that span space-time and form a basis. At this point, A_μ satisfies the equation of motion but not yet the constraint $\partial_\nu A^\nu = 0$. Working out this constraint, it is easy to show that this amounts to $\epsilon^\alpha_\mu k^\mu a^\alpha_{\boldsymbol{k}} = 0$. If we now choose a frame where the wave-vector is aligned along the z-direction, then the basis can be taken to be

$$\epsilon^0_\mu = \begin{pmatrix} -\omega/m \\ 0 \\ 0 \\ k/m \end{pmatrix}, \quad \epsilon^1_\mu = \begin{pmatrix} 0 \\ \hat{\epsilon}^1_x \\ \hat{\epsilon}^1_y \\ 0 \end{pmatrix}, \tag{670}$$

$$\epsilon^2_\mu = \begin{pmatrix} 0 \\ \hat{\epsilon}^2_x \\ \hat{\epsilon}^2_y \\ 0 \end{pmatrix}, \quad \epsilon^3_\mu = \begin{pmatrix} -k/m \\ 0 \\ 0 \\ \omega/m \end{pmatrix}, \tag{671}$$

where $\hat{\epsilon}^{1,2}$ are vectors perpendicular to the wave-vector (the time components of ϵ^0 and ϵ^3 carry a minus sign because we show here the covariant vectors). But we have $\epsilon^0_\mu k^\mu \neq 0$ despite the fact that this is the only choice such that $\epsilon^0_\mu \epsilon^{\mu 3} = 0$. Therefore, the only possibility left in order to satisfy the constraint is to take $a^0_{\boldsymbol{k}} = 0$. As a consequence, the Fourier expansion of the vector potential is now given by

$$A_\mu(t, \boldsymbol{x}) = \frac{1}{(2\pi)^{3/2}} \int \frac{\mathrm{d}\boldsymbol{k}}{\sqrt{2\omega(k)}} \sum_{\alpha=1}^{3} \epsilon^\alpha_\mu(\boldsymbol{k}) \Big[a^\alpha_{\boldsymbol{k}} \mathrm{e}^{-i\omega t + i\boldsymbol{k}\cdot\boldsymbol{x}} + (a^\alpha_{\boldsymbol{k}})^\dagger \mathrm{e}^{i\omega t - i\boldsymbol{k}\cdot\boldsymbol{x}} \Big], \tag{672}$$

where the three vectors appearing in this decomposition are still given in Eqs. (670) and (671).

Let us now come to our main subject and calculate the stress-energy tensor. There are several ways to carry out this calculation. Here, we choose to return to Eq. (665) and to vary this action with respect to the space-time metric tensor. Then, in order to work with the stress energy tensor in flat space-time, we will take $g_{\mu\nu} = \eta_{\mu\nu}$ but, of course, only after having performed the variation. Straightforward manipulations lead to the following expression

$$T_{\mu\nu} \equiv -\frac{2}{\sqrt{-g}} \frac{\delta S[A_\mu]}{\delta g^{\mu\nu}} \tag{673}$$

$$= -g_{\mu\nu} \left(\frac{1}{4} g^{\lambda\alpha} g^{\kappa\beta} F_{\lambda\kappa} F_{\alpha\beta} + \frac{1}{2} m^2 g^{\alpha\beta} A_\alpha A_\beta \right) + g^{\alpha\beta} F_{\mu\alpha} F_{\nu\beta} + m^2 A_\mu A_\nu. \tag{674}$$

We are now in a position where we can compute the various components of $T_{\mu\nu}$. Let us start with the energy density. From the above expression, one deduces that

$$T_{00} = \frac{1}{2}\delta^{ij} F_{0i} F_{0j} + \frac{1}{4}\delta^{ij}\delta^{kl} F_{ik} F_{jl} + \frac{1}{2} m^2 A_0^2 + \frac{1}{2} m^2 \delta^{ij} A_i A_j. \tag{675}$$

Then, in order to evaluate the above equation, we have to separately calculate the four terms that appear in this expression. Let us first give the relevant results needed to calculate the first

term. Upon using Eq. (672), it is easy to establish that

$$\begin{aligned}\delta^{ij}\langle 0|\partial_0 A_i \partial_0 A_j|0\rangle &= \frac{1}{(2\pi)^3}\int \frac{\mathrm{d}\boldsymbol{k}}{2\omega(k)}\delta^{ij} \\ &\times \sum_{\alpha=1}^{3}\epsilon_i^\alpha \epsilon_j^\alpha \omega^2, \end{aligned} \tag{676}$$

$$\begin{aligned}\delta^{ij}\langle 0|\partial_0 A_i \partial_j A_0|0\rangle &= -\frac{1}{(2\pi)^3}\int \frac{\mathrm{d}\boldsymbol{k}}{2\omega(k)}\delta^{ij} \\ &\times \sum_{\alpha=1}^{3}\epsilon_i^\alpha \epsilon_0^\alpha \omega k_j, \end{aligned} \tag{677}$$

and,

$$\begin{aligned}\delta^{ij}\langle 0|\partial_i A_0 \partial_j A_0|0\rangle &= \frac{1}{(2\pi)^3}\int \frac{\mathrm{d}\boldsymbol{k}}{2\omega(k)}\delta^{ij} \\ &\times \sum_{\alpha=1}^{3}\epsilon_0^\alpha \epsilon_0^\alpha k_i k_j. \end{aligned} \tag{678}$$

In the same manner, the second term, containing the space space component of the field strength, can be expressed as

$$\begin{aligned}\delta^{ij}\delta^{kl}\langle 0|F_{ik}F_{jl}|0\rangle &= \frac{2}{(2\pi)^3}\int \frac{\mathrm{d}\boldsymbol{k}}{2\omega(k)}\delta^{ij}\delta^{kl} \\ &\times \sum_{\alpha=1}^{3}\left(\epsilon_k^\alpha \epsilon_l^\alpha k_i k_j - \epsilon_k^\alpha \epsilon_j^\alpha k_i k_l\right). \end{aligned} \tag{679}$$

The third term can be written as

$$\langle 0|A_0^2|0\rangle = \frac{1}{(2\pi)^3}\int \frac{\mathrm{d}\boldsymbol{k}}{2\omega(k)}\sum_{\alpha=1}^{3}\epsilon_0^\alpha \epsilon_0^\alpha, \tag{680}$$

and, finally, the fourth and last term takes the form

$$\delta^{ij}\langle 0|A_i A_j|0\rangle = \frac{1}{(2\pi)^3}\int \frac{\mathrm{d}\boldsymbol{k}}{2\omega(k)}\sum_{\alpha=1}^{3}\delta^{ij}\epsilon_i^\alpha \epsilon_j^\alpha. \tag{681}$$

Having calculated the four terms appearing in the expression of T_{00}, one must now combine them. Using the relationships satisfied by the vectors basis[128], one obtains the following ex-

[128]For convenience, let us give the two equations needed to complete the calculation. It involves simple combination of the polarization vectors. Using the explicit representation given by Eqs. (670) and (671), one arrives at

$$\delta^{ij}\sum_{\alpha=1}^{3}\epsilon_i^\alpha \epsilon_j^\alpha = 2 + \frac{\omega^2}{m^2}, \tag{682}$$

$$\sum_{\alpha=1}^{3}\epsilon_0^\alpha \epsilon_0^\alpha = \frac{k^2}{m^2}. \tag{683}$$

Then, these two expressions can be combined to lead to the following relation which is what is needed in order to simplify the expression of the vacuum energy density

$$\begin{aligned}\sum_{\alpha=1}^{3}\left(\epsilon_k^\alpha \epsilon_l^\alpha k_i k_j - \epsilon_k^\alpha \epsilon_j^\alpha k_i k_l\right) &= k^2\left(2 + \frac{\omega^2}{m^2}\right) - \frac{k^2\omega^2}{m^2} \\ &= 2. \end{aligned} \tag{684}$$

pression for the vacuum energy density

$$\langle \rho \rangle = \frac{1}{(2\pi)^3} \frac{3}{2} \int \mathrm{d}\boldsymbol{k}\, \omega(k). \tag{685}$$

Based on our experience with the calculation of the vacuum energy of a spin 1/2 field, this expression is totally expected. As before, it is equal to the scalar field energy density times a multiplicative factor which takes into account the spin of the particle. Here, for a massive vector field, there are three polarization states and, as a consequence, the multiplicative factor is three.

In order to be exhaustive, let us now evaluate the pressure. The calculation proceeds exactly as before and it is easy to show that

$$\perp^{\mu\nu} T_{\mu\nu} = T_{00} + m^2 A_0^2 - m^2 \delta^{ij} A_i A_j, \tag{686}$$

from which, using the relations already established before, we immediately deduce that

$$\langle p \rangle = \frac{1}{(2\pi)^3} \frac{3}{6} \int \mathrm{d}\boldsymbol{k}\, \frac{k^2}{\omega(k)}. \tag{687}$$

Of course, this is again the expected formula, namely the scalar field pressure times a multiplicative factor equals to three. Clearly, there is no need to repeat the discussion about the regularization of the above divergent integrals. It proceeds exactly as explained before.

24.7.3 Mass-less Vector Field

Finally, we consider the case where the vector field is mass-less, $m = 0$ [16, 17, 18, 19, 20, 21, 22]. This is of course necessary if we want to realistically evaluate the vacuum energy density since we know that this type of field exist in Nature. If the mass vanishes, this means that the dispersion relation is now given by

$$\omega(k) = k, \tag{688}$$

as appropriate for photons. In that case, contrary to the massive case, we can no longer establish that $\partial_\nu A^\nu = 0$. Therefore, the equation of motion takes the form

$$\partial_\mu \partial^\mu A^\nu - \partial^\nu \left(\partial_\mu A^\mu \right) = 0. \tag{689}$$

At this stage, we need to fix a gauge. As is well-known, it is convenient to work in the Coulomb gauge defined by the following conditions

$$A_0 = 0, \quad \partial_i A^i = 0. \tag{690}$$

In particular, the last condition implies that $A_i(t, \boldsymbol{x})$ is a transverse field. As a consequence, its Fourier expansion can be written as

$$\begin{aligned} A_i\left(t, \boldsymbol{x}\right) &= \frac{1}{(2\pi)^{3/2}} \int \frac{\mathrm{d}\boldsymbol{k}}{\sqrt{2\omega(k)}} \sum_{\alpha=1}^{2} \epsilon_i^\alpha(\boldsymbol{k}) \Big[a_{\boldsymbol{k}}^\alpha \mathrm{e}^{-i\omega t + i\boldsymbol{k}\cdot\boldsymbol{x}} \\ &\quad + \left(a_{\boldsymbol{k}}^\alpha\right)^\dagger \mathrm{e}^{i\omega t - i\boldsymbol{k}\cdot\boldsymbol{x}} \Big]. \end{aligned} \tag{691}$$

The two three-vectors ϵ^1 and ϵ^2 are perpendicular to the wave-vectors $\boldsymbol{k}$. In the mass-less case, only two degrees of freedom remain.

The calculation of the stress-energy tensor is now straightforward. Using the calculation of the previous section, in particular the expression of the time-time component of the stress-energy tensor (675), one finds that

$$T_{00} = \frac{1}{2} \delta^{ij} \partial_0 A_i \partial_0 A_j + \frac{1}{4} \delta^{ij} \delta^{kl} F_{ik} F_{jl}, \tag{692}$$

and, since we now have,

$$\delta^{ij}\sum_{\alpha=1}^{2}\epsilon_i^\alpha\epsilon_j^\alpha = 2, \tag{693}$$

$$\sum_{\alpha=1}^{2}\left(\epsilon_k^\alpha\epsilon_l^\alpha k_i k_j - \epsilon_k^\alpha\epsilon_j^\alpha k_i k_l\right) = 2k^2, \tag{694}$$

one finally arrives at the following expression for the vacuum energy density

$$\langle\rho\rangle = \frac{1}{(2\pi)^3}\frac{2}{2}\int \mathrm{d}\boldsymbol{k}\,\omega(k), \tag{695}$$

that is to say the scalar field energy density times two, as expected since there are two polarization states. For the pressure, using the same approach, one immediately obtains

$$\langle p\rangle = \frac{1}{(2\pi)^3}\frac{2}{6}\int \mathrm{d}\boldsymbol{k}\,\frac{k^2}{\omega(k)}, \tag{696}$$

where, as usual, the factor two in front of the whole expression originates from the the two polarization states of a mass-less vector particle.

We conclude by stating the main result established in this section. For any type of fields, the vacuum energy density can be written as

$$\langle\rho\rangle = \frac{1}{(2\pi)^3}\frac{s}{2}\int \mathrm{d}\boldsymbol{k}\,\omega(k), \tag{697}$$

where s represents the number of polarization states ($s = 1$ for a scalar field, $s = 4$ for a spinor field, $s = 3$ for a massive vector field and $s = 2$ for a mass-less vector field). The pressure is given by a similar expression. When these divergent integrals are regularized, one can show that, again for any type of field, this leads to the Lorentz invariant form of the vacuum equation of state, namely $\langle p\rangle = -\langle\rho\rangle$. In principle, the results of this section makes the accurate evaluation of the vacuum energy density in the standard model of particle physics possible. However, before discussing this point, we need to say something about the fact that ρ_{vac} is positive for bosons and negative for fermions.

24.7.4 Application: Super-Symmetry

Motivation

The fact that the sign of the zero-point energy density is different for fermions and bosons, see Eqs. (364) and (648), suggests a very simple way of solving the cosmological constant problem. It is clear that one can design a theory where the two contributions are equal in absolute value such that the final result is exactly zero. Let us see how it works in practice. For this purpose, let us consider again the Lagrangian of a massive real scalar field, see also Eq. (302)

$$\mathcal{L}_\phi = -\frac{\alpha}{2}\eta^{\mu\nu}\partial_\mu\Phi\partial_\nu\Phi - \frac{1}{2}m^2\Phi^2, \tag{698}$$

where we have introduced a constant α in front of the kinetic term. Before we took $\alpha = 1$ and it is interesting to discuss why we made this choice (after all, how do we know that $\alpha = 1$?). As usual, see Eq. (357), one Fourier expands the scalar field as

$$\Phi(t,\boldsymbol{x}) = \frac{1}{(2\pi)^{3/2}}\int\frac{\mathrm{d}\boldsymbol{k}}{\sqrt{2\omega(k,\alpha)}}\Big(c_{\boldsymbol{k}}\mathrm{e}^{-i\omega t+i\boldsymbol{k}\cdot\boldsymbol{x}} + c_{\boldsymbol{k}}^\dagger\mathrm{e}^{i\omega t-i\boldsymbol{k}\cdot\boldsymbol{x}}\Big), \tag{699}$$

with a new expression for the quantity ω, compare to Eq. (358), namely

$$\omega(k,\alpha) = \sqrt{k^2 + \frac{m^2}{\alpha^2}}, \tag{700}$$

since this is the condition for Eq. (699) to be a solution of the "new" equation of motion [see also Eq. (356)]

$$-\alpha\ddot{\phi} + \alpha\delta^{ij}\partial_i\phi\partial_j\phi - m^2\phi = 0. \tag{701}$$

The new dispersion relation can easily be understood. We see that the quantity α simply renormalizes the value of the mass m. This makes sense since, if $m = 0$, we clearly see that the relativistic relation $\omega = k$ is not affected by the presence of the constant α.

The next step is to derive the commutation relations for the creation and annihilation operators. Let us assume for the moment that we have the standard relation (359)

$$\left[c_{\boldsymbol{k}}, c^\dagger_{\boldsymbol{p}}\right] = \delta\left(\boldsymbol{k} - \boldsymbol{p}\right). \tag{702}$$

Then, from the expression of the Lagrangian, one can calculate the conjugate momentum. It reads $\Pi(t,\boldsymbol{x}) = -\alpha\dot{\Phi}(t,\boldsymbol{x})$. This implies that the commutation relation between the field and its conjugate momentum now reads

$$[\phi(t,\boldsymbol{x}), \pi(t,\boldsymbol{y})] = i\alpha\delta(\boldsymbol{x} - \boldsymbol{y}), \tag{703}$$

an expression which reduces to the standard commutation relation only if $\alpha = 1$. Therefore, if one wants to have $[\Phi, \Pi] = i\delta(\boldsymbol{x}-\boldsymbol{y})$ and, at the same time, $[c_{\boldsymbol{k}}, c^\dagger_{\boldsymbol{p}}] = \delta\left(\boldsymbol{k} - \boldsymbol{p}\right)$ there is no other choice that taking $\alpha = 1$. It is worth noticing that working with the two previous commutation relations (702), (703) and $\alpha \neq 1$ would lead to a different prediction for the energy density of the vacuum. Indeed, the stress-energy tensor is now given by

$$T_{\mu\nu} = \alpha\partial_\mu\Phi\partial_\nu\Phi - g_{\mu\nu}\left(\frac{\alpha}{2}g^{\alpha\beta}\partial_\alpha\Phi\partial_\beta\Phi + \frac{1}{2}m^2\Phi^2\right), \tag{704}$$

which implies that

$$T_{00} = \frac{\alpha}{2}\dot{\Phi}^2 + \frac{\alpha}{2}\delta^{ij}\partial_i\Phi\partial_j\Phi + \frac{1}{2}m^2\Phi^2. \tag{705}$$

From this expression, it is easy to show that

$$\langle\rho\rangle = \frac{\alpha}{(2\pi)^3}\frac{1}{2}\int \mathrm{d}\boldsymbol{k}\,\omega(k,\alpha), \tag{706}$$

an expression that should be compared to Eq. (364).

Of course, one can also demonstrate that the consistency can be re-establish if the commutation relations for the creation and annihilation operators are taken to be $[c_{\boldsymbol{k}}, c^\dagger_{\boldsymbol{p}}] = \delta\left(\boldsymbol{k} - \boldsymbol{p}\right)/\alpha$. In this case the commutation relation of the field and its conjugate momentum have the correct form. Moreover, one has

$$\langle 0|\dot{\Phi}^2|0\rangle = \frac{1}{(2\pi)^3}\int\frac{\mathrm{d}\boldsymbol{k}}{2\omega(k,\alpha)}\frac{\omega^2(k,\alpha)}{\alpha}, \tag{707}$$

$$\langle 0|\delta^{ij}\partial_i\Phi\partial_j\Phi|0\rangle = \frac{1}{(2\pi)^3}\int\frac{\mathrm{d}\boldsymbol{k}}{2\omega(k,\alpha)}\frac{k^2}{\alpha}, \tag{708}$$

$$\langle 0|\Phi^2|0\rangle = \frac{1}{(2\pi)^3}\int\frac{\mathrm{d}\boldsymbol{k}}{2\omega(k,\alpha)}\frac{1}{\alpha}, \tag{709}$$

[these formulas should be compared to Eqs. (360), (361) and (362), which implies that

$$\langle\rho\rangle = \frac{1}{(2\pi)^3}\frac{1}{2}\int \mathrm{d}\boldsymbol{k}\,\omega(k,\alpha), \tag{710}$$

that is to say the "standard" result.

Therefore, in conclusion, the "1/2" coefficient in the expression of the vacuum energy density is only obtained if the commutation relations and the normalization of the field are properly and consistently chosen. In particular, if one chooses to work with a coefficient "1/2" in front of the kinetic term (canonical normalization), then there is no other choice than working with $[c_{\boldsymbol{k}}, c^{\dagger}_{\boldsymbol{p}}] = \delta(\boldsymbol{k} - \boldsymbol{p})$. Or, in other words, the previous commutation relation implies that a "1/2" in front of the kinetic term must be present.

We now apply the previous discussion to the cases of a complex scalar field and to a Majorana spinor. The Lagrangian for a complex scalar field should be written as

$$\mathcal{L} = -g^{\mu\nu}\partial_\mu\Phi^*\partial_\nu\Phi - m^2\Phi^*\Phi, \tag{711}$$

with, this time, no "1/2" in front of the whole expression. Indeed, if one writes that

$$\phi = \frac{1}{\sqrt{2}}(A_1 + iA_2), \tag{712}$$

then the Lagrangian takes the form of the sum of two Lagrangian for the two real canonically normalized fields A_1 and A_2.

The previous discussion has also important implications for the Lagrangian of a Majorana spinor. A Majorana spinor is a spinor satisfying

$$\Psi^{\rm c}_{\rm M} = \Psi_{\rm M}, \tag{713}$$

where the charge conjugate spinor $\Psi^{\rm c}$ is defined by

$$\Psi^{\rm c} \equiv C\overline{\Psi}^{\rm T}. \tag{714}$$

In this expression the symbol "T" denotes the transpose matrix. The matrix C is the charge conjugation operator that can be represented by $i\gamma^2\gamma^0$ and satisfies $C\gamma^{\rm T}_\mu C^{-1} = -\gamma_\mu$. We have of course $(\Psi^{\rm c})^{\rm c} = \Psi$. From a Dirac spinor Ψ one can always write two Majorana spinors

$$\Psi_{\rm M1} = \frac{1}{\sqrt{2}}(\Psi + \Psi^{\rm c}), \tag{715}$$

$$\Psi_{\rm M2} = \frac{1}{i\sqrt{2}}(\Psi - \Psi^{\rm c}). \tag{716}$$

Of course, the above expressions can always be inverted and re-written as

$$\Psi = \frac{1}{\sqrt{2}}(\Psi_{\rm M1} + i\Psi_{\rm M2}), \tag{717}$$

$$\Psi^{\rm c} = \frac{1}{\sqrt{2}}(\Psi_{\rm M1} - i\Psi_{\rm M2}). \tag{718}$$

The two previous expression can be inserted in the Dirac Lagrangian (627) (taking $m = 0$). This leads to the following expression

$$\begin{aligned}\mathcal{L}_{\rm Dirac} &= -\frac{1}{\sqrt{2}}\left(\overline{\Psi}_{\rm M1} - i\overline{\Psi}_{\rm M2}\right) i\gamma^\mu\partial_\mu\frac{1}{\sqrt{2}}\left(\overline{\Psi}_{\rm M1} + i\overline{\Psi}_{\rm M2}\right)\\ &= -\frac{i}{2}\overline{\Psi}_{\rm M1}\gamma^\mu\partial_\mu\Psi_{\rm M1} - \frac{i}{2}\overline{\Psi}_{\rm M2}\gamma^\mu\partial_\mu\Psi_{\rm M2}\\ &\quad -\frac{i}{2}\left(-i\overline{\Psi}_{\rm M2}\gamma^\mu\partial_\mu\Psi_{\rm M1} + i\overline{\Psi}_{\rm M1}\gamma^\mu\partial_\mu\Psi_{\rm M2}\right).\end{aligned} \tag{719}$$

But, using partial integrations, one easily shows that

$$\overline{\Psi}_{\rm M2}\gamma^\mu\partial_\mu\Psi_{\rm M1} = -\left(\partial_\mu\overline{\Psi}_{\rm M2}\right)\gamma^\mu\Psi_{\rm M1} = +\overline{\Psi}_{\rm M1}\gamma^\mu\partial_\mu\Psi_{\rm M2}, \tag{720}$$

so that the last term vanishes. Therefore, one conclude that the Lagrangian for a Majorana spinor is given by

$$\mathcal{L}_{\text{Majorana}} = -\frac{1}{2}\overline{\Psi}_{\text{M}}\left(i\gamma^{\mu}\partial_{\mu} - m\right)\Psi_{\text{M}}, \tag{721}$$

and that the expansion of a Majorana spinor operator in Fourier can be expressed as, see also Eq. (631)

$$\begin{aligned}\Psi_{\text{M}}(t,\boldsymbol{x}) &= \frac{1}{(2\pi)^{3/2}}\int\frac{\mathrm{d}\boldsymbol{k}}{\sqrt{2\omega(k)}}\sum_{r=1}^{2}\Big[b^{r}_{\boldsymbol{k}}u(\boldsymbol{k},r)\mathrm{e}^{-i\omega t+i\boldsymbol{k}\cdot\boldsymbol{x}}\\ &\quad + (b^{r}_{\boldsymbol{k}})^{\dagger}v(\boldsymbol{k},r)\mathrm{e}^{i\omega t-i\boldsymbol{k}\cdot\boldsymbol{x}}\Big],\end{aligned} \tag{722}$$

where we notice that only the operators $b^{r}_{\boldsymbol{k}}$ are present which reflects the fact that, for a Majorana spinor, the particle and the anti-particle are the same. But the most important aspect is the coefficient "1/2" in Eq. (721). Clearly, it means that if we perform the calculation of $\langle\rho\rangle$ again, we are going to find

$$\langle\rho\rangle = -\frac{1}{(2\pi)^{3}}\frac{2}{2}\int\mathrm{d}\boldsymbol{k}\,\omega(k), \tag{723}$$

that is to say an expression with an overall factor 2 instead of 4 for a Dirac fermion, see Eq. (648).

We are now in a position to address the main point of this section. Let us consider the following model described by a complex scalar field and a Marojana spinor

$$\mathcal{L} = -g^{\mu\nu}\partial_{\mu}\Phi^{*}\partial_{\nu}\Phi - m^{2}_{\text{B}}\phi^{*}\phi - \frac{1}{2}\overline{\Psi}_{\text{M}}\left(i\gamma^{\mu}\partial_{\mu} - m_{\text{F}}\right)\Psi_{\text{M}}, \tag{724}$$

where m_{B} is the mass of the bosonic field and m_{F} is the mass of the fermionic field. From the above considerations, it is obvious to calculate the vacuum energy density. It reads

$$\begin{aligned}\langle\rho\rangle &= \frac{1}{(2\pi)^{3}}\frac{2}{2}\int\mathrm{d}\boldsymbol{k}\,\sqrt{k^{2}+m^{2}_{\text{B}}}\\ &\quad -\frac{1}{(2\pi)^{3}}\frac{2}{2}\int\mathrm{d}\boldsymbol{k}\,\sqrt{k^{2}+m^{2}_{\text{F}}}.\end{aligned} \tag{725}$$

We have a factor two for the complex scalar field since we have seen that a complex scalar field is equivalent to two real scalar fields and a factor -2 for the Marojana spinor as shown before. In general, the quantity $\langle\rho\rangle$ is not zero because $m_{\text{B}} \neq m_{\text{F}}$. But if we now assume that

$$m_{\text{B}} = m_{\text{F}}, \tag{726}$$

then the zero-point energy density exactly cancels out and we have solved the cosmological constant problem. The above model is in fact nothing but the simplest super-symmetric model, the so-called Wess-Zumino model [70, 71, 72, 73, 19, 74].

The Vacuum Energy in Super-Symmetry

Having motivated the idea of super-symmetry, we now study in more detail its implications. In particular, we would like to check that the type of cancellation studied before is generic in super-symmetry. In order to simplify the discussion, we start with considering the Lagrangian (724) with $m_{\text{B}} = m_{\text{F}} = 0$. We will now use the standard super-symmetric notation with the dotted two-components Weyl spinors [19, 74]. In this formalism, the Majorana spinor can be written as

$$\Psi_{\text{M}} = \begin{pmatrix}\Psi_{\alpha}\\ \overline{\Psi}^{\dot{\alpha}}\end{pmatrix}. \tag{727}$$

Then, using that the Dirac matrices can be expressed as

$$\gamma^\mu = \begin{pmatrix} 0 & (\sigma^\mu)_{\alpha\dot{\alpha}} \\ (\overline{\sigma}^\mu)^{\dot{\alpha}\alpha} & 0 \end{pmatrix}, \tag{728}$$

we find that the Lagrangian (721) can be re-written as

$$\mathcal{L}_{\text{Majorana}} = -\frac{i}{2}\overline{\Psi}_{\dot{\alpha}} (\overline{\sigma}^\mu)^{\dot{\alpha}\alpha} \partial_\mu \Psi_\alpha - \frac{i}{2}\Psi^\alpha (\sigma^\mu)_{\alpha\dot{\alpha}} \partial_\mu \overline{\Psi}^{\dot{\alpha}}, \tag{729}$$

where we used that (by definition) $\overline{\Psi}_{\dot{\alpha}} = (\Psi_\alpha)^*$ and $\Psi^\alpha = (\overline{\Psi}^{\dot{\alpha}})^*$. The next step consists in using the relation $\Psi^\alpha (\sigma^\mu)_{\alpha\dot{\alpha}} \partial_\mu \overline{\Psi}^{\dot{\alpha}} = -\partial_\mu \overline{\Psi}_{\dot{\alpha}} (\overline{\sigma}^\mu)^{\dot{\alpha}\alpha} \Psi_\alpha$ (the spinor components are Grassmann variables and, therefore, anti-commute, see below) and, then, to transform the resulting expression by integration by part. In this way, one shows that the Lagrangian (729) is in fact twice the first term. As a consequence, the Wess-Zumino model is described by

$$\mathcal{L}_{\text{W-Z}} = -\eta^{\mu\nu}\partial_\mu \Phi^\dagger \partial_\nu \Phi - i\overline{\Psi}_{\dot{\alpha}} (\overline{\sigma}^\mu)^{\dot{\alpha}\alpha} \partial_\mu \Psi_\alpha. \tag{730}$$

Let us now consider a super-symmetric transformation. In order to understand what it really means, let us use the analogy with a gauge transformation [19, 74]. For definiteness, we consider a SU(2) transformation. Let v be a SU(2) doublet of spinor fields. After an infinitesimal transformation, this doublet v becomes a new doublet v' such that

$$v' = \mathrm{e}^{i\epsilon^j \sigma_j/2} v \simeq \left(\mathbb{I}_2 + i\epsilon^j \frac{\sigma_j}{2}\right) v, \tag{731}$$

or

$$\delta v = i\epsilon^j \frac{\sigma_j}{2} v, \tag{732}$$

where σ_j are the Pauli matrices, the generators of the group. The quantity ϵ^j is a vector (or a collection of three parameters) specifying the transformation. In super-symmetry, the transformation is characterized by a Weyl spinor ξ^α which, therefore, plays the role of the quantity ϵ^j. A super-symmetric transformation transforms a boson into a fermion and vice-versa. As a consequence, one can write

$$\delta\Phi = \sqrt{2}\xi^\alpha \Psi_\alpha, \tag{733}$$

$$\delta\Psi_\alpha = -i\sqrt{2} (\sigma^\mu)_{\alpha\dot{\beta}} \overline{\xi}^{\dot{\beta}} \partial_\mu \Phi, \tag{734}$$

where the factors i or $\sqrt{2}$ are introduced for future convenience. Before showing that the Wess-Zumino Lagrangian is indeed invariant under a super-symmetric transformation, one needs to introduce a last result. We would like the Lagrangian to be invariant without using the equation of motion (in other words, we would like the Lagrangian to be invariant "off-shell"). In this case, one needs to add a term such that the new Lagrangian reads

$$\mathcal{L}_{\text{W-Z}} = -\eta^{\mu\nu}\partial_\mu \Phi^\dagger \partial_\nu \Phi - i\overline{\Psi}_{\dot{\alpha}} (\overline{\sigma}^\mu)^{\dot{\alpha}\alpha} \partial_\mu \Psi_\alpha - F^\dagger F, \tag{735}$$

where F is a complex scalar field. It is easy to see that this term is harmless since the corresponding equation of motion is just $F = 0$. But this term is necessary if one wants to match the bosonic and fermionic degrees of freedom. A Weyl spinor has two-complex components and we now have two complex scalar fields. Obviously, we also need to modify our super-symmetric transformation and they now take the form

$$\delta\Phi = \sqrt{2}\xi^\alpha \Psi_\alpha, \tag{736}$$

$$\delta\Psi_\alpha = \sqrt{2}\xi_\alpha F - i\sqrt{2} (\sigma^\mu)_{\alpha\dot{\beta}} \overline{\xi}^{\dot{\beta}} \partial_\mu \Phi, \tag{737}$$

$$\delta F = i\sqrt{2} (\partial_\mu \Psi^\alpha) (\sigma^\mu)_{\alpha\dot{\beta}} \overline{\xi}^{\dot{\beta}}. \tag{738}$$

Our goal is now to show that the Lagrangian (735) is indeed invariant under the three transformations (736), (737) and (738).

In fact, the key to show that the Lagrangian is invariant is to be able to evaluate the above transformations for the complex conjugate quantities. We have

$$\begin{aligned}\delta\Phi^\dagger &= \sqrt{2}\,(\xi^\alpha\Psi_\alpha)^\dagger = \sqrt{2}\,(\Psi_\alpha)^\dagger\,(\xi^\alpha)^\dagger \\ &= \sqrt{2}\,(\Psi_\alpha)^*\,(\xi^\alpha)^* = \sqrt{2}\,\overline{\Psi}_{\dot\alpha}\overline{\xi}^{\dot\alpha} \\ &= \sqrt{2}\,\overline{\xi}_{\dot\alpha}\overline{\Psi}^{\dot\alpha} = -\sqrt{2}\,\overline{\xi}^{\dot\alpha}\overline{\Psi}_{\dot\alpha}. \end{aligned} \tag{739}$$

We cannot review here the complete dotted spinors formalism in detail, see Ref. [74], but in order to make the calculations reasonably self-consistent, it is nevertheless interesting to recall a few things. The first point to discuss is the definition of the dagger symbol for Grassmann variables. In fact, a very clear discussion can be found in p. 35 of Ref. [74] and there is no need here to entirely repeat it. In brief, one needs to remember that we deal with quantum fields. Usually, the dagger symbols means complex conjugation and transposition. For a scalar field, it simply amounts to take the complex conjugate of the mode function and put (or remove) a dagger to the creation and annihilation operators. For the spinor component Ψ_α, this is the same (here it is important to remember that one discusses one component only and not the spinor viewed as a column and, therefore, we "cannot" take the transpose of the matrix -there is no matrix- !) and one defines $\Psi^\dagger_{\dot\alpha}$ as the complex conjugate of the corresponding spinor component times the creation and annihilation operators with or without the dagger symbol. However, the question arises as how to evaluate the Hermitian conjugate of a product of Grassmann variables. Here, Ref. [74] shows that the most convenient way is to define $(\xi^\alpha\Psi_\alpha)^\dagger = \Psi^\dagger_\alpha\xi^{\alpha\dagger}$, that is to say something similar to what we are used to in the case of two matrices even if, again, we do not deal with matrices here but with components of Weyl spinors. This definition justifies the first step in the previous equation.

Then, it is also interesting to discuss how the two last equations are obtained. Let us write $\overline{\Psi}_{\dot 1} = a$, $\overline{\Psi}_{\dot 2} = b$ and $\overline{\xi}^{\dot 1} = c$, $\overline{\xi}^{\dot 2} = d$, where a, b, c and d are (complex) Grassman variables. As a consequence, one has $\overline{\Psi}_{\dot\alpha}\overline{\xi}^{\dot\alpha} = ac + bd$. On the other hand, $\overline{\xi}_{\dot\alpha} \equiv \epsilon_{\dot\alpha\dot\beta}\overline{\xi}^{\dot\beta}$, where

$$\epsilon_{\dot\alpha\dot\beta} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{740}$$

from which we deduce that $\overline{\xi}_{\dot 1} = d$ and $\overline{\xi}_{\dot 2} = -c$. In the same manner, one has $\overline{\Psi}^{\dot\alpha} = \epsilon^{\dot\alpha\dot\beta}\overline{\Psi}_{\dot\beta}$ with

$$\epsilon^{\dot\alpha\dot\beta} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \tag{741}$$

from which one obtains $\overline{\psi}^{\dot 1} = -b$ and $\overline{\psi}^{\dot 2} = a$. Using these results, one has $\overline{\xi}_{\dot\alpha}\overline{\Psi}^{\dot\alpha} = -db - ca = ac + bd = \overline{\Psi}_{\dot\alpha}\overline{\xi}^{\dot\alpha}$, which is the wanted result. But, using again the expressions of the components obtained above, one also notices that $\overline{\xi}_{\dot\alpha}\overline{\Psi}^{\dot\alpha} = -\overline{\xi}^{\dot\alpha}\overline{\Psi}_{\dot\alpha}$ which is the last result used to establish the expression (739) of $\delta\Phi^\dagger$. In a more compact notation, one has

$$\begin{aligned}\overline{\xi}_{\dot\alpha}\overline{\Psi}^{\dot\alpha} &= \epsilon_{\dot\alpha\dot\beta}\overline{\xi}^{\dot\beta}\overline{\Psi}^{\dot\alpha} = \overline{\xi}^{\dot\beta}\epsilon_{\dot\alpha\dot\beta}\overline{\Psi}^{\dot\alpha} = -\overline{\xi}^{\dot\beta}\epsilon_{\dot\beta\dot\alpha}\overline{\Psi}^{\dot\alpha} \\ &= -\overline{\xi}^{\dot\beta}\overline{\Psi}_{\dot\beta}, \end{aligned} \tag{742}$$

the key point being the fact that the matrix $\epsilon_{\dot\alpha\dot\beta}$ is antisymmetric.

We now need the transformation of $(\delta\Psi_\alpha)^* = \delta\overline{\Psi}_{\dot\alpha}$. one has

$$\delta\overline{\Psi}_{\dot\alpha} = \sqrt{2}\,\overline{\xi}_{\dot\alpha}F^\dagger + i\sqrt{2}\left[(\sigma^\mu)_{\alpha\dot\beta}\right]^*\xi^\beta\partial_\mu\phi^\dagger, \tag{743}$$

$$= \sqrt{2}\,\overline{\xi}_{\dot\alpha}F^\dagger + i\sqrt{2}\,(\sigma^\mu)_{\beta\dot\alpha}\,\xi^\beta\partial_\mu\Phi^\dagger. \tag{744}$$

Finally, following the same lines, the transformation of $\delta F^\dagger$ can be expressed as

$$\begin{aligned}\delta F^\dagger &= -i\sqrt{2}(\overline{\xi}^{\dot\beta})^\dagger\left(\partial_\mu\Psi^\alpha\right)^\dagger\left[\left(\sigma^\mu\right)_{\alpha\dot\beta}\right]^* &(745)\\ &= -i\sqrt{2}\xi^\beta\left(\partial_\mu\overline{\Psi}^{\dot\alpha}\right)\left(\sigma^\mu\right)_{\beta\dot\alpha}. &(746)\end{aligned}$$

We are now in a position where one can compute the variation of the super-symmetric Lagrangian. Collecting the above results, one obtains

$$\begin{aligned}\delta\mathcal{L}_{\rm w-z} &= -\eta^{\mu\nu}\partial_\mu\delta\Phi^\dagger\partial_\nu\Phi-\eta^{\mu\nu}\partial_\mu\Phi^\dagger\partial_\nu\delta\Phi-i\delta\overline{\Psi}_{\dot\alpha}\left(\overline{\sigma}^\mu\right)^{\dot\alpha\alpha}\partial_\mu\Psi_\alpha-i\overline{\Psi}_{\dot\alpha}\left(\overline{\sigma}^\mu\right)^{\dot\alpha\alpha}\partial_\mu\delta\Psi_\alpha-\delta F^\dagger F-F^\dagger\delta F &(747)\\ &= -i\sqrt{2}F^\dagger\overline{\xi}_{\dot\alpha}\left(\overline{\sigma}^\mu\right)^{\dot\alpha\alpha}\partial_\mu\Psi_\alpha-i\sqrt{2}F^\dagger\left(\partial_\mu\Psi^\alpha\right)\left(\sigma^\mu\right)_{\dot\alpha\beta}\overline{\xi}^{\dot\beta}-i\sqrt{2}\,\overline{\Psi}_{\dot\alpha}\left(\overline{\sigma}^\mu\right)^{\dot\alpha\alpha}\xi_\alpha\partial_\mu F\\ &\quad+i\sqrt{2}\xi^\beta\left(\sigma^\mu\right)_{\beta\dot\alpha}\left(\partial_\mu\overline{\Psi}^{\dot\alpha}\right)F+\sqrt{2}\eta^{\mu\nu}\overline{\xi}^{\dot\alpha}\partial_\mu\overline{\Psi}_{\dot\alpha}\partial_\nu\Phi-\sqrt{2}\left(\overline{\sigma}^\mu\right)^{\dot\alpha\alpha}\left(\sigma^\nu\right)_{\alpha\dot\beta}\overline{\Psi}_{\dot\alpha}\overline{\xi}^{\dot\beta}\partial_\mu\partial_\nu\Phi-\sqrt{2}\eta^{\mu\nu}\xi^\alpha\partial_\nu\Psi_\alpha\partial_\mu\Phi^\dagger\\ &\quad+\sqrt{2}\left(\sigma^\nu\right)_{\beta\dot\alpha}\left(\overline{\sigma}^\mu\right)^{\dot\alpha\alpha}\xi^\beta\partial_\mu\Psi_\alpha\partial_\nu\Phi^\dagger. &(748)\end{aligned}$$

Each term in this expression needs to be studied in detail. The first term is also equal to $i\sqrt{2}F^\dagger\partial_\mu\Psi^\alpha\left(\sigma^\mu\right)_{\alpha\dot\alpha}\overline{\xi}^{\dot\alpha}$ and, therefore, exactly cancels the second one. The third one is, after integration by parts, $i\sqrt{2}\left(\partial_\mu\overline{\Psi}_{\dot\alpha}\right)\left(\overline{\sigma}\right)^{\dot\alpha\alpha}\xi_\alpha F$, which is also equal to $-i\sqrt{2}\xi^\alpha\left(\sigma^\mu\right)_{\alpha\dot\alpha}\left(\partial_\mu\overline{\Psi}^{\dot\alpha}\right)F$. This cancels exactly the fourth term. As a consequence, all the term involving the field F give a total derivative contribution only. It remains the four terms involving the scalar field Φ and $\Phi^\dagger$. We first focus on the terms involving Φ only. For this purpose, we use the following relation [19, 74]

$$\left(\overline{\sigma}^\mu\right)^{\dot\alpha\alpha}\left(\sigma^\nu\right)_{\alpha\dot\beta}=\eta^{\mu\nu}\delta^{\dot\alpha}{}_{\dot\beta}+2\left(\overline{\sigma}^{\mu\nu}\right)^{\dot\alpha}{}_{\dot\beta}, \tag{749}$$

where

$$\overline{\sigma}^{\mu\nu}\equiv\frac{1}{4}\left(\overline{\sigma}^\mu\sigma^\nu-\overline{\sigma}^\nu\sigma^\mu\right)=-\overline{\sigma}^{\nu\mu}. \tag{750}$$

Therefore, the second term involving Φ [*i.e.* the sixth term in the above formula (748)] can be re-written as

$$\begin{aligned}&-\sqrt{2}\left(\overline{\sigma}^\mu\right)^{\dot\alpha\alpha}\left(\sigma^\nu\right)_{\alpha\dot\beta}\overline{\Psi}_{\dot\alpha}\overline{\xi}^{\dot\beta}\partial_\mu\partial_\nu\Phi\\ =\;&-\sqrt{2}\left[\eta^{\mu\nu}\delta^{\dot\alpha}{}_{\dot\beta}+2\left(\overline{\sigma}^{\mu\nu}\right)^{\dot\alpha}{}_{\dot\beta}\right]\overline{\Psi}_{\dot\alpha}\overline{\xi}^{\dot\beta}\partial_\mu\partial_\nu\Phi\\ =\;&-\sqrt{2}\eta^{\mu\nu}\overline{\Psi}_{\dot\alpha}\overline{\xi}^{\dot\alpha}\partial_\mu\partial_\nu\Phi, &(751)\end{aligned}$$

the second term in this equation leading to a vanishing contribution since $\overline{\sigma}^{\mu\nu}$ is anti-symmetric. The next manipulation is to integrate by part. This gives

$$\begin{aligned}&-\sqrt{2}\left(\overline{\sigma}^\mu\right)^{\dot\alpha\alpha}\left(\sigma^\nu\right)_{\alpha\dot\beta}\overline{\Psi}_{\dot\alpha}\overline{\xi}^{\dot\beta}\partial_\mu\partial_\nu\Phi\\ =\;&-\sqrt{2}\eta^{\mu\nu}\overline{\Psi}_{\dot\alpha}\overline{\xi}^{\dot\alpha}\partial_\mu\partial_\nu\Phi=\sqrt{2}\eta^{\mu\nu}\partial_\mu\overline{\Psi}_{\dot\alpha}\overline{\xi}^{\dot\alpha}\partial_\nu\Phi\\ =\;&\sqrt{2}\eta^{\mu\nu}\overline{\xi}_{\dot\alpha}\partial_\mu\overline{\Psi}^{\dot\alpha}\partial_\nu\Phi=-\sqrt{2}\eta^{\mu\nu}\overline{\xi}^{\dot\alpha}\partial_\mu\overline{\Psi}_{\dot\alpha}\partial_\nu\Phi.\end{aligned} \tag{752}$$

Therefore, this term exactly cancels the first term involving Φ [or the fifth term in the formula (748)]. Clearly, this also works for the two terms involving $\Phi^\dagger$. We have thus shown, without using the equation of motion, that the variation of the Wess-Zumino Lagrangian (735) is a total derivative. As a consequence, the Wess-Zumino model is indeed super-symmetric, that is to say invariant under the super-symmetric transformations (736), (737) and (738).

We have seen before that the constant spinor ξ^α was the equivalent of the vector ϵ^j in the case of the SU(2) transformation considered at the beginning of this section. But what is the equivalent of the generator $\sigma_j/2$ for a super-symmetric transformation? Moreover, we know that the generators obey the formula $[\sigma_i/2,\sigma_j/2]=i\epsilon_{ijk}\sigma_k/2$. What is the corresponding equation satisfied by the super-symmetric generators? We now turn to these questions.

Using again the analogy with the SU(2) transformation, we expect that an infinitesimal super-symmetric transformation to have the form

$$\delta_\xi \Phi \simeq i \left(\xi^\alpha Q_\alpha + \overline{\xi}_{\dot{\alpha}} \overline{Q}^{\dot{\alpha}} \right) \Phi. \tag{753}$$

Here Q_α is the super-symmetric generator, the equivalent of $\sigma_j/2$. Clearly, since $\xi^\alpha Q_\alpha$ must be a scalar, Q_α is a Weyl spinor as the notation indicates.

In order to find the super-symmetric algebra, we first compute the quantity $(\delta_\eta \delta_\xi - \delta_\xi \delta_\eta)\,\Phi$ using Eqs. (736), (737) and (738). It is straightforward to show that

$$\begin{aligned}
&(\delta_\eta \delta_\xi - \delta_\xi \delta_\eta)\,\Phi = 2\xi^\alpha \eta_\alpha F - 2i\xi^\alpha \left(\sigma^\mu\right)_{\alpha\dot{\beta}} \overline{\eta}^{\dot{\beta}} \partial_\mu \Phi \\
&-2\eta^\alpha \xi_\alpha F + 2i\eta^\alpha \left(\sigma^\mu\right)_{\alpha\dot{\beta}} \overline{\xi}^{\dot{\beta}} \partial_\mu \Phi \qquad (754) \\
&= -2i\xi^\alpha \left(\sigma^\mu\right)_{\alpha\dot{\beta}} \overline{\eta}^{\dot{\beta}} \partial_\mu \Phi + 2i\eta^\alpha \left(\sigma^\mu\right)_{\alpha\dot{\beta}} \overline{\xi}^{\dot{\beta}} \partial_\mu \Phi, \qquad (755)
\end{aligned}$$

since $\xi^\alpha \eta_\alpha = \eta^\alpha \xi_\alpha$. But, of course, we can evaluate the same commutator using the expression (753) for the infinitesimal super-symmetric transformation. This leads to

$$\begin{aligned}
&(\delta_\eta \delta_\xi - \delta_\xi \delta_\eta)\,\Phi = -\left(\eta^\alpha Q_\alpha + \overline{\eta}_{\dot{\alpha}} \overline{Q}^{\dot{\alpha}} \right) \\
&\times \left(\xi^\beta Q_\beta + \overline{\xi}_{\dot{\beta}} \overline{Q}^{\dot{\beta}} \right) \Phi + \left(\xi^\alpha Q_\alpha + \overline{\xi}_{\dot{\alpha}} \overline{Q}^{\dot{\alpha}} \right) \\
&\times \left(\eta^\beta Q_\beta + \overline{\eta}_{\dot{\beta}} \overline{Q}^{\dot{\beta}} \right) \Phi. \qquad (756)
\end{aligned}$$

Then, expanding this expression (carefully taking into account the fact that the spinor components are Grassmann variables), we arrive at

$$\begin{aligned}
(\delta_\eta \delta_\xi - \delta_\xi \delta_\eta)\,\Phi \;&=\; \eta^\alpha \left(Q_\alpha Q_\beta + Q_\beta Q_\alpha \right) \xi^\beta \Phi \\
&+\; \overline{\eta}_{\dot{\alpha}} \left(\overline{Q}^{\dot{\alpha}} \overline{Q}^{\dot{\beta}} + \overline{Q}^{\dot{\beta}} \overline{Q}^{\dot{\alpha}} \right) \overline{\xi}_{\dot{\beta}} \Phi \\
&+\; \eta^\alpha \left(Q_\alpha \overline{Q}^{\dot{\beta}} + \overline{Q}^{\dot{\beta}} Q_\alpha \right) \overline{\xi}_{\dot{\beta}} \Phi \\
&-\; \xi^\alpha \left(Q_\alpha \overline{Q}^{\dot{\beta}} + \overline{Q}^{\dot{\beta}} Q_\alpha \right) \overline{\eta}_{\dot{\beta}} \Phi. \qquad (757)
\end{aligned}$$

One must now compare Eq. (755) with Eq. (757). First of all, we notice that, in Eq. (755), there is no term with two undotted or two dotted spinors ξ and η. This implies that

$$\left\{ Q_\alpha, Q_\beta \right\} = \left\{ Q^{\dot{\alpha}}, Q^{\dot{\beta}} \right\} = 0. \tag{758}$$

Then, a comparison of the first "cross-terms" leads to

$$\left\{ Q_\alpha, \overline{Q}_{\dot{\beta}} \right\} \Phi = -2i \left(\sigma^\mu\right)_{\alpha\dot{\beta}} \partial_\mu \Phi, \tag{759}$$

and a similar relation for the second cross-term. Remembering that $P_\mu = -i\partial_\mu$ (the minus sign is in fact just a convention; the problem with super-symmetry is that the literature contains many different conventions (A nice treatment of this question can be found in the first problem of Ref. [75], Appendix C), one arrives at

$$\left\{ Q_\alpha, \overline{Q}_{\dot{\beta}} \right\} = 2 \left(\sigma^\mu\right)_{\alpha\dot{\beta}} P_\mu. \tag{760}$$

This equation shows that two successive super-symmetric transformations are in fact equivalent to a space-time translation. As we will see, this has far-reaching consequences. However, for

the subject discussed here, the above formula has also very important implications. Indeed, working it out explicitly, one has

$$Q_1\overline{Q}_{\dot{1}} + \overline{Q}_{\dot{1}}Q_1 = 2P_3 + 2P_0, \tag{761}$$
$$Q_2\overline{Q}_{\dot{2}} + \overline{Q}_{\dot{2}}Q_2 = -2P_3 + 2P_0, \tag{762}$$

from which we deduce

$$4P_0 = Q_1\overline{Q}_{\dot{1}} + \overline{Q}_{\dot{1}}Q_1 + Q_2\overline{Q}_{\dot{2}} + \overline{Q}_{\dot{2}}Q_2. \tag{763}$$

Therefore, since the Hamiltonian is nothing but the time translation generator P_0, this implies that

$$\langle 0\,|H|\,0\rangle = \frac{1}{4}\left\langle 0\left|Q_1\overline{Q}_{\dot{1}} + \overline{Q}_{\dot{1}}Q_1 + Q_2\overline{Q}_{\dot{2}} + \overline{Q}_{\dot{2}}Q_2\right|0\right\rangle \geq 0. \tag{764}$$

We have learned two things: firstly, if the vacuum state is not super-symmetric, *i.e.* if super-symmetry is spontaneously broken (the underlying theory is super-symmetric but the solution or the state in which the system is placed is not), then the vacuum energy is necessarily positive. Secondly, another remarkable consequence of the above results is that, when the vacuum state is super-symmetric, that is to say when

$$Q_\alpha|0\rangle = 0, \tag{765}$$

one has $\langle 0\,|H|\,0\rangle = 0$, *i.e.* the vacuum energy is automatically zero. We have thus proven that the cancellation discussed at the beginning of this section is in fact not a coincidence. Moreover, we have identified under which conditions it is valid: it is always true for a super-symmetric system. This is the reason why super-symmetry has the potential to solve the cosmological constant problem. It gives a concrete example of a symmetry which forces the vacuum energy to vanish.

Unfortunately, we know that super-symmetry must be broken in the real world. This comes from the fact that each boson (fermion) of the standard model is not observed to have a super-symmetric partner of the same mass. This means that vacuum energy is now given by Eq. (725). Let us evaluate it with our "incorrect" method which consists in regularizing the divergent integrals with a cut-off M, see Eq. (371). We find

$$\begin{aligned}\langle\rho\rangle_{\text{SUSY}} &= \frac{M^4}{8\pi^2}\left(1 + \frac{m_{\text{B}}^2}{M^2} + \cdots\right) \\ &\quad -\frac{M^4}{8\pi^2}\left(1 + \frac{m_{\text{F}}^2}{M^2} + \cdots\right) \qquad (766)\\ &= \frac{M^2}{8\pi^2}\left(m_{\text{B}}^2 - m_{\text{F}}^2\right). \qquad (767)\end{aligned}$$

We see that, even if super-symmetry is broken, the most divergent part $\propto M^4$ of the energy density cancel out. If we now properly estimate the energy density by means of dimensional regularization (392), one obtains that

$$\langle\rho\rangle_{\text{SUSY}} = \frac{m_{\text{B}}^4}{32\pi^2}\ln\left(\frac{m_{\text{B}}^4}{\mu^2}\right) - \frac{m_{\text{F}}^4}{32\pi^2}\ln\left(\frac{m_{\text{F}}^4}{\mu^2}\right). \tag{768}$$

Therefore, roughly speaking, one can estimate that vacuum energy is now given by $\simeq M_{\text{SUSY}}^4$ where M_{SUSY} is the typical difference between the mass of the super-partners, which is also the super-symmetric breaking mass.

A last remark is in order. In this section we have considered super-symmetry in flat space-time while the cosmological constant problem is formulated in curved space-time. In order to be consistent, we should therefore treat super-symmetry in this last context. As is well-known, this leads to super-gravity [19, 75]. Regarding the vacuum energy question, super-gravity presents important differences compared to super-symmetry. In particular, one loses the result that vacuum energy is always positive. But, this also shows that discussing the zero-point energy density issue in curved space-time can bring new aspects to the problem. This is the reason why, in the next section, we address this question.

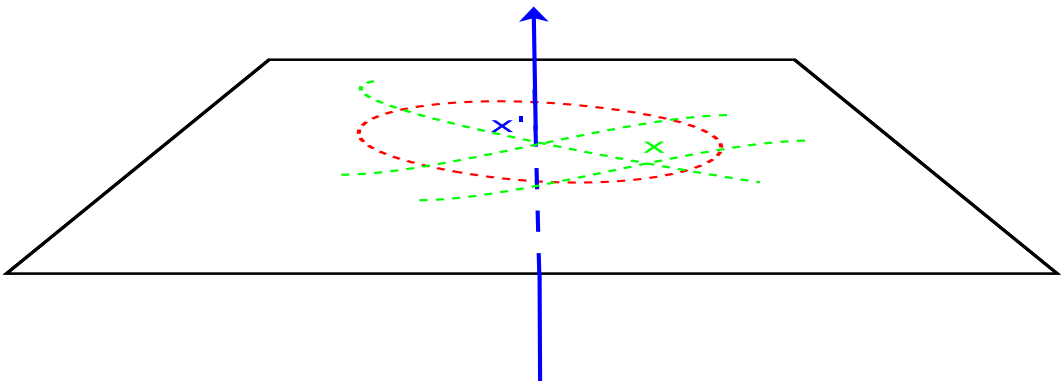


Figure 156: Sketch of the Riemann coordinates used in order to derive an approximate expression for the Green function, see Eq. (799). The plot aims at illustrating the fact that the Riemann coordinates are a local concept. Since it is sufficient to establish the Green function locally (and not in the entire curved manifold), one can always endow the neighborhood of a point x' (denoted here by the region surrounded by the closed dotted line) with special coordinates x^μ such that the calculation is simplified. It is important to notice that this does not restrict the generality of the obtained results.

24.8 The Vacuum Energy Density in Curved Space-time

In the previous sections, we have computed the vacuum energy density in flat space-time. This may seem inappropriate since we have argued that the problem occurs only in curved space-time. Here we show how to justify the previous approach. Intuitively, this is justified because we deal with ultra-violet divergences for which the large scale structure of the curved manifold should not play a too important role. In other words, in the ultra-violet regime, one only probes the local properties of space-time and, locally, one cannot distinguish a curved manifold from the Minkowski space-time. However, we need to put this intuitive reasoning on solid grounds. For this purpose, let us consider our toy model again where matter is simply represented by a scalar field. The only difference with the previous sections is that, now, this quantum field lives in a curved space-time. Therefore, in order to have meaningful Einstein equations, one must assume that the quantum average of the stress energy tensor sources the Einstein equations. In other words, we start from [53]

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda_{\rm B} g_{\mu\nu} = \kappa \langle T_{\mu\nu} \rangle \,, \tag{769}$$

Then, by analogy with Eq. (299), we define the effective action W to be

$$\langle T_{\mu\nu} \rangle = -\frac{2}{\sqrt{-g}} \frac{\delta W}{\delta g^{\mu\nu}}. \tag{770}$$

The effective action can be written in term of the sourceless generating functional $Z[0]$, namely

$$W = -i \ln Z[0], \tag{771}$$

where we recall that the explicit expression of $Z[0]$ is given by, see also Eq. (393),

$$Z[0] = \int \mathcal{D}\phi \, {\rm e}^{iS_{\rm matter}[\phi]}, \tag{772}$$

the quantity $S_{\rm matter}$ being the action of the scalar field. The above result can be proven in the following way. Varying $Z[0]$, one obtains

$$\delta Z[0] = \int \mathcal{D}\phi \, i\delta S_{\rm matter} \, {\rm e}^{iS_{\rm matter}[\phi]} \tag{773}$$

$$= \langle 0 \, | i\delta S_{\rm matter} | \, 0 \rangle \,. \tag{774}$$

As a consequence,

$$-\frac{2}{\sqrt{-g}}\frac{\delta W}{\delta g^{\mu\nu}} = -\frac{2}{\sqrt{-g}}\frac{-i}{Z[0]}\frac{\delta Z[0]}{\delta g^{\mu\nu}} \tag{775}$$

$$= \frac{1}{Z[0]}\left\langle 0\left|-\frac{2}{\sqrt{-g}}\frac{\delta S_{\text{matter}}}{\delta g^{\mu\nu}}\right|0\right\rangle \tag{776}$$

$$= \frac{1}{Z[0]}\langle 0\,|T_{\mu\nu}|\,0\rangle = \langle T_{\mu\nu}\rangle\,, \tag{777}$$

since the mean value is defined by $\langle\cdots\rangle \equiv \langle 0|\cdots|0\rangle/\langle 0|0\rangle$.

The next step is to calculate W explicitly in our case, *i.e.* for a scalar field. We now show that the effective action can be expressed in terms of the Green function. The Green function is defined by the following expression [53]

$$\left(-g^{\mu\nu}\nabla_\mu\nabla_\nu + m^2\right) G_{\text{F}}(x,y) = \frac{1}{\sqrt{-g}}\delta(x-y). \tag{778}$$

It is easy to show that $G_{\text{F}}(x,y) = iD_{\text{F}}(x-y)$, where the propagator has been defined in Eq. (397). Integrating by part the action of the scalar field $\Phi(x)$, the generating function can be written as

$$Z[J] = \int \mathcal{D}\Phi \exp\left[-\frac{i}{2}\int \mathrm{d}^d x\sqrt{-g}\Phi(x)\left(-g^{\mu\nu}\nabla_\mu\nabla_\nu + m^2\right)\Phi(x) + i\int \mathrm{d}^d x\sqrt{-g}J(x)\Phi(x)\right], \tag{779}$$

where we have considered the d-dimensional case for future convenience. This path integral is a Gaussian integral and, therefore, can be evaluated using the standard techniques. For $J = 0$ one has

$$Z[0] = (2\pi)^{\infty/2}\left(\det K\right)^{-1/2} \tag{780}$$

$$= (2\pi)^{\infty/2}\mathrm{e}^{\frac{1}{2}\mathrm{tr}\left(\ln K^{-1}\right)}, \tag{781}$$

where K is the following operator, appearing in Eq. (779),

$$K(x,y) \equiv i\sqrt{-g}\left(-g^{\mu\nu}\nabla_\mu\nabla_\nu + m^2\right)\delta^d(x-y). \tag{782}$$

But the inverse of the operator K is in fact the propagator $D_{\text{F}}(x,y)$. A pedagogical, but simplistic, demonstration of this fact is the following one. The generating functional can be written as

$$Z[0] = \int \Pi_x \mathrm{d}\Phi_i \mathrm{e}^{-\frac{1}{2}\sum\limits_{x,y}\Phi_x K_{xy}\Phi_y}, \tag{783}$$

where, in order to reproduce Eq. (782) (or, rather, to mimic this formula), the matrix K_{xy} is taken to be $K_{xy} = K\delta_{xy}$ and the "number K is defined" by $K \equiv i\sqrt{-g}\left(-g^{\mu\nu}\nabla_\mu\nabla_\nu + m^2\right)$. Then, the inverse of the matrix K must obey

$$\sum_z K_{xz}K^{-1}_{zy} = \delta_{xy}, \tag{784}$$

that is to say, using the definition of K_{xy},

$$KK^{-1}_{xy} = \delta_{xy}. \tag{785}$$

But this equation is nothing but Eq. (778). This shows that, as announced, $K^{-1}_{xy} = \left(D_{\text{F}}\right)_{xy}$. As a consequence, one can write

$$Z[0] = (2\pi)^{\infty/2}\mathrm{e}^{\frac{1}{2}\mathrm{tr}\left(\ln D_{\text{F}}\right)}, \tag{786}$$

or,

$$W = -\frac{i}{2}\mathrm{tr}\left(\ln D_{\mathrm{F}}\right) - i\ln\left[(2\pi)^{\infty/2}\right]. \tag{787}$$

We have reduced the evaluation of the effective action to the evaluation of the Green function (or the propagator) in curved space-time. We now turn to this question.

For this purpose, we introduce the Riemann normal coordinates x^μ with origin at the point x', see Fig. 156 and Refs. [76, 77, 78]. As it will become gradually clear, the motivation for such a choice is as follows. Since we want to see how the Einstein equations are modified, one only needs to evaluate the Green function locally. In this case, one can introduce coordinates that drastically simplify the problem, *i.e.* the calculation. This is exactly the role played by the Riemann normal coordinates. In these coordinates, the metric can be written as [76, 77]

$$g_{\mu\nu} = \eta_{\mu\nu} - \frac{1}{3}R_{\mu\alpha\nu\beta}x^\alpha x^\beta - \frac{1}{6}R_{\mu\alpha\nu\beta;\gamma}x^\alpha x^\beta x^\gamma + \left(-\frac{1}{20}R_{\mu\alpha\nu\beta;\gamma\delta} + \frac{2}{45}R_{\alpha\mu\beta\lambda}R^\lambda{}_{\gamma\nu\delta}\right)x^\alpha x^\beta x^\gamma x^\delta + \cdots \tag{788}$$

Then, we insert this metric into the Green function equation (778) and retain terms with coefficients involving four derivatives of the metric or fewer. A lengthy calculation leads to

$$\begin{aligned}
&\eta^{\mu\nu}\partial_\mu\partial_\nu\overline{G} - \left(m^2 - \frac{1}{6}\right)\overline{G} - \frac{1}{3}R_\alpha{}^\nu\partial_\nu\overline{G} + \frac{1}{3}R^\mu{}_\alpha{}^\nu{}_\beta x^\alpha x^\beta\partial_\mu\partial_\nu\overline{G} + \frac{1}{6}R_{;\alpha}x^\alpha\overline{G} + \left(-\frac{1}{3}R_\alpha{}^\nu{}_{;\beta} + \frac{1}{6}R_{\alpha\beta}{}^{;\nu}\right)x^\alpha x^\beta\partial_\nu\overline{G} \\
+ &\frac{1}{6}R^\mu{}_\alpha{}^\nu{}_{\beta;\gamma}x^\alpha x^\beta x^\gamma\partial_\mu\partial_\nu\overline{G} + \frac{1}{12}R_{;\alpha\beta}x^\alpha x^\beta\overline{G} + \left(-\frac{1}{30}R_\alpha{}^\lambda R_{\lambda\beta} + \frac{1}{60}R^\kappa{}_\alpha{}^\lambda{}_\beta R_{\kappa\lambda} + \frac{1}{60}R^{\lambda\mu\kappa}{}_\alpha R_{\lambda\mu\kappa\beta} - \frac{1}{120}R_{;\alpha\beta}\right. \\
+ &\left.\frac{1}{40}\eta^{\mu\nu}\partial_\mu\partial_\nu R_{\alpha\beta}\right)x^\alpha x^\beta\overline{G} + \left(-\frac{3}{20}R^\nu{}_{\alpha;\beta\gamma} + \frac{1}{10}R_{\alpha\beta}{}^{;\nu}{}_\gamma - \frac{1}{60}R^\kappa{}_\alpha{}^\nu{}_\beta R_{\kappa\lambda} + \frac{1}{15}R^\kappa{}_{\alpha\lambda\beta}R_\kappa{}^\nu{}_\gamma{}^\lambda\right)x^\alpha x^\beta x^\gamma\partial_\nu\overline{G} \\
+ &\left(\frac{1}{20}R^\mu{}_\alpha{}^\nu{}_{\beta;\gamma\delta} + \frac{1}{15}R^\mu{}_{\alpha\lambda\beta}R^\lambda{}_\gamma{}^\nu{}_\delta\right)x^\alpha x^\beta x^\gamma x^\delta\partial_\mu\partial_\nu\overline{G} = -\delta^d(x),
\end{aligned} \tag{789}$$

where all the indices are raised or lowered by the Minkowski metric and where we have defined a rescaled Green function according to $G_{\mathrm{F}}(x,y) \equiv g^{-1/4}\overline{G}(x,y)$. The next step is to Fourier transform the above equation. In this way the local expansion which, in real space, looks like an expansion in the coordinates x^μ is transformed, in momentum space, as an expansion in terms of the inverse of k^μ, as appropriate for a local expansion. More precisely, defining

$$\overline{G}(x,x') = \frac{1}{(2\pi)^d}\int \mathrm{d}^d k\, \mathrm{e}^{i\eta^{\alpha\beta}k_\alpha x_\beta}\overline{G}(k), \tag{790}$$

we can expand the Green function as $\overline{G}(k) = \overline{G}_0(k) + \overline{G}_1(k) + \cdots$ where $\overline{G}_i(k)$ is the quantity appearing in front of a coefficient involving the i-th derivatives of the metric tensor [76, 77]. Let us notice that the Green function is a function of x' in the sense that x^μ represents the coordinates of a point x defined such x' is at the origin. Another remark is that we define the Fourier transform of the Green function with an overall factor $(2\pi)^{-d}$ such that the coefficient of the Dirac function in the right hand side of Eq. (789) is minus one and does not contain factors of π, thanks to the formula $\delta^d(x) = 1/(2\pi)^d\int \mathrm{d}^d k\, \mathrm{e}^{ikx}$.

Introducing this expansion into Eq. (789) and consistently identifying the terms of a given order leads to the desired result. This iterative procedure gives

$$\begin{aligned}
\overline{G}(k) &= \frac{1}{k^2+m^2} + \frac{R}{6}\frac{1}{(k^2+m^2)^2} + \frac{i}{6}R_{;\alpha}\frac{1}{k^2+m^2}\frac{\partial}{\partial k_\alpha}\left[\left(k^2+m^2\right)^{-1}\right] + \frac{R^2}{36}\frac{1}{(k^2+m^2)^3} \\
&\quad + a_{\alpha\beta}\frac{1}{k^2+m^2}\frac{\partial^2}{\partial k_\alpha\partial k_\beta}\left[\left(k^2+m^2\right)^{-1}\right] + \cdots, \qquad (791) \\
&= \frac{1}{k^2+m^2} + \frac{R}{6}\frac{1}{(k^2+m^2)^2} + \frac{i}{12}R_{;\alpha}\frac{\partial}{\partial k_\alpha}\left[\left(k^2+m^2\right)^{-2}\right] + \frac{1}{3}a_{\alpha\beta}\partial^\alpha\partial^\beta\left[\left(k^2+m^2\right)^{-2}\right] \\
&\quad + \left(\frac{R^2}{36} - \frac{2}{3}a^\lambda{}_\lambda\right)\frac{1}{(k^2+m^2)^3} + \cdots, \qquad (792)
\end{aligned}$$

where the coefficient $a_{\alpha\beta}$ can be expressed as

$$a_{\alpha\beta} = -\frac{3}{40}R_{;\alpha\beta} - \frac{1}{40}\eta^{\mu\nu}\partial_\mu\partial_\nu R_{\alpha\beta} + \frac{1}{30}R_\alpha{}^\lambda R_{\lambda\beta} - \frac{1}{60}R^\kappa{}_\alpha{}^\lambda{}_\beta R_{\kappa\lambda} + \frac{1}{60}R^{\lambda\mu\kappa}{}_\alpha R_{\lambda\mu\kappa\beta}. \tag{793}$$

Inserting this last expression into Eq. (790) leads to

$$\overline{G}(x,x') = \frac{1}{(2\pi)^d}\int \mathrm{d}^n k\, \mathrm{e}^{i\eta^{\alpha\beta}k_\alpha x_\beta}\left[1 + f_1(x,x')\left(-\frac{\partial}{\partial m^2}\right) + f_2(x,x')\left(-\frac{\partial}{\partial m^2}\right)^2 + \cdots\right]\frac{1}{k^2+m^2}, \tag{794}$$

with

$$f_1(x,x') = \frac{R}{6} + \frac{1}{12}R_{;\alpha}x^\alpha - \frac{1}{3}a_{\alpha\beta}x^{\alpha\beta}, \tag{795}$$

$$f_2(x,x') = \frac{R^2}{72} - \frac{1}{3}a^\lambda{}_\lambda \tag{796}$$

The final step of the calculation consists in expressing the Green function in the so-called proper time formalism [76, 77]. Firstly, we replace the quantity $(k^2+m^2)^{-1}$ with

$$\frac{1}{k^2+m^2} = \int_0^\infty i\mathrm{d}s \mathrm{e}^{-is\left(k^2+m^2\right)}, \tag{797}$$

and, secondly, we make use of the result

$$\begin{aligned}&\frac{1}{(2\pi)^d}\int \mathrm{d}^n k\, \mathrm{e}^{i\eta^{\alpha\beta}k_\alpha x_\beta - is\left(k^2+m^2\right)} \\ &= \frac{i}{(4\pi)^{d/2}}\,(is)^{-d/2}\,\mathrm{e}^{-ism^2+i\sigma/(2s)},\end{aligned} \tag{798}$$

where $\sigma = \eta^{\alpha\beta}x_\alpha x_\beta/2$. This leads to our final expression for the Green function, namely

$$\begin{aligned}\overline{G}(x,x') &= \frac{i}{(4\pi)^{d/2}}\int_0^\infty \frac{i\mathrm{d}s}{(is)^{d/2}}\mathrm{e}^{-im^2s-\sigma/(2is)} \\ &\quad\times\left[1+isf_1+(is)^2 f_2+\cdots\right].\end{aligned} \tag{799}$$

Of course, one could always push the expansion to higher orders in inverse of the momenta in order to obtain higher orders in s. But, in the present context, this will not be necessary.

We now have all the tools to evaluate the effective action (787). We now seek an explicit expression for the quantity $\ln(D_{\mathrm{F}})$. As already mentioned, D_{F} can be viewed as the inverse of the "continuous matrix" K which means that, formally, one can write

$$D_{\mathrm{F}} = \int_0^\infty \mathrm{e}^{-iKs} i\mathrm{d}s, \tag{800}$$

where we have used again Eq. (797). This expression allows us to express the matrix element of e^{-iKs}. Indeed, given that $D_{\mathrm{F}} = -ig^{-1/4}\overline{G}$, a direct comparison of Eq. (799) with the previous relation leads to

$$\begin{aligned}\left(\mathrm{e}^{-iKs}\right)_{xx'} &= \frac{g^{-1/4}}{(4\pi)^{d/2}}\frac{\mathrm{e}^{-im^2s-\sigma/(2is)}}{(is)^{d/2}} \\ &\times \left[1+isf_1+(is)^2 f_2+\cdots\right].\end{aligned} \tag{801}$$

Now, let us consider the following integral

$$\begin{aligned}\int_\lambda^\infty \mathrm{e}^{-iKs}\frac{i\mathrm{d}s}{is} &= -\int_{i\lambda K}^\infty \frac{\mathrm{e}^{-t}}{t}\mathrm{d}t \\ &= -\gamma - \ln(i\lambda K) - \sum_{k=1}^\infty \frac{(-1)^{k+1}(i\lambda K)^k}{kk!}.\end{aligned} \tag{802}$$

In the limit $\lambda \to 0$, up to an infinite constant that can be ignored, only the logarithm survives. Therefore, we can define the logarithm of the operator K as

$$-\ln K = \ln D_{\rm F} = \int_0^\infty {\rm e}^{-iKs} \frac{i{\rm d}s}{is}. \tag{803}$$

Then, using Eq. (801), one obtains

$$\begin{aligned} \left(\ln D_{\rm F}\right)_{xx'} &= \frac{g^{-1/4}(x)}{(4\pi)^{d/2}} \int_0^\infty \frac{{\rm e}^{-im^2 s - \sigma(x,x')/(2is)}}{(is)^{d/2}} \\ &\times \left[1 + is f_1(x,x') \right. \\ &\left. + (is)^2 f_2(x,x') + \cdots \right] \frac{i{\rm d}s}{is}. \end{aligned} \tag{804}$$

As a consequence, we have now an explicit expression for the effective action, namely

$$W = -\frac{i}{2} \int {\rm d}^d x \sqrt{-g} \lim_{x' \to x} \left(\ln D_{\rm F}\right)_{xx'}, \tag{805}$$

or

$$\begin{aligned} W &= \int {\rm d}^d x \sqrt{-g} \left\{ -\frac{i}{2} \frac{1}{(4\pi)^{d/2}} \int_0^\infty \frac{{\rm e}^{-im^2 s}}{(is)^{d/2}} \right. \\ &\left. \times \left[1 + is f_1 + (is)^2 f_2 + \cdots \right] \frac{i{\rm d}s}{is} \right\}. \end{aligned} \tag{806}$$

In the coincident limit $x' \to x$, the determinant of the metric tensor has been taken to one since it is calculated in x', that is to say for $x = 0$. For the same reason, we also have considered that $\sigma = 0$. The above expression can be evaluated explicitly in terms of Euler function and one obtains

$$\begin{aligned} W &= \int {\rm d}^d x \sqrt{-g} \left\{ -\frac{i}{2} \frac{1}{(4\pi)^{d/2}} \left(\frac{m}{\mu}\right)^{d-4} \right. \\ &\times \left[m^4 \Gamma\left(-\frac{d}{2}\right) + m^2 f_1 \Gamma\left(1 - \frac{d}{2}\right) \right. \\ &\left. \left. + f_2 \Gamma\left(2 - \frac{d}{2}\right) + \cdots \right] \right\}, \end{aligned} \tag{807}$$

where we have introduced the scale μ in order to maintain the correct dimension of the effective action. The above expression can be analyzed in a dimensional regularization scheme. In this case, one is led to the following expression

$$\begin{aligned} iW &= \int {\rm d}^d x \sqrt{-g} \left\{ \left[\frac{2}{\epsilon} - \gamma - \ln\left(\frac{m^2}{4\pi\mu^2}\right) \right] \right. \\ &\left. \times \left(\frac{m^4}{64\pi^2} - \frac{m^2 f_1}{32\pi^2} + \frac{f_2}{32\pi^2} \right) + \cdots \right\}, \end{aligned} \tag{808}$$

which is our final result. The i factor can be absorbed into a constant shift so we do not need to worry about it. If we compare the effective action with the original Lagrangian (297), we see that the first divergent term in the above expression can be absorbed into a redefinition of the cosmological constant (the second term can be viewed as a redefinition of the Newton constant and the third term leads to the appearance of new terms in the action, see Ref. [53]). Explicitly, one has

$$\Lambda_{\text{eff}} = \Lambda_{\text{B}} + \kappa \frac{m^4}{64\pi^2} \left[\frac{2}{\epsilon} - \gamma - \ln\left(\frac{m^2}{4\pi\mu^2} \right) \right], \tag{809}$$

that is to say exactly the same expression as before (except an unimportant factor 3/2), see for instance Eq. (391). As a consequence in a $\overline{\text{MS}}$ scheme [16], we can remove the divergence and obtain exactly the same expression derived before, see Eq. (392), namely

$$\Lambda_{\text{eff}} = \Lambda_{\text{B}} + \kappa \frac{m^4}{64\pi^2} \ln\left(\frac{m^2}{\mu^2} \right). \tag{810}$$

The calculations of this section justify our previous approach in flat space-time. Indeed, we have just proven that all the results derived before in flat space-time are in fact valid in the more rigorous approach where the curvature of space-time is properly taken into account. This confirms that the vacuum energy does not scale as the cut-off to the power four, see Eqs. (371) and (374) but is in fact proportional to the mass of the particle to the power four times a logarithmic factor depending on the renormalization scale. In some sense, the calculation in flat space-time can be considered as a computational trick: it is sufficient to calculate the vacuum energy in this simple framework since, from the previous considerations, we know that this is also the result that consistently emerges from an approach where the curvature of space-time is properly included. Of course, physically, the reason for this success is not so surprising and was already mentioned before. Since we deal with the ultraviolet behavior of the theory, a local analysis should necessarily lead to a correct and consistent result.

It is also worth mentioning that Eqs. (392) and (810) recently attracted lot of attention, see Refs. [47, 51]. In particular, as we discussed at length in Sec. 24.4.1 and as was emphasized in these articles, these equations are the consequence of the fact that the regularization scheme used to tame the divergence of the vacuum energy must respect Lorentz invariance. In fact, these equations have been known for a long time, see in particular Eq. (6.50) of the standard textbook [53]. Therefore, the fact that the cosmological constant is proportional to the mass of the particle to the power four, and not to the cut-off to the power four as usually claimed, see again Eqs. (371) and (374), is in fact not a new result. Of course, it modifies a lot our estimate of vacuum energy. For instance, the correct result gives zero for the photons while, for the same situation, the result based on the wrong regularization scheme gives infinity. Clearly, this means that the correct regularization scheme leads to a cosmological constant much smaller than the one obtained from the wrong approach. As a consequence, one could hope that the cosmological constant problem is in fact just an artifact due to the use of an incorrect regularization method. We turn to this question in the next section but, unfortunately, we will see that this is not the case.

24.9 The Value of the Cosmological Constant

We are now in a position to conclude the first part of this review. The considerations of the previous sections lead to the following expression for the vacuum energy [51]

$$\rho_{\text{vac}} = \sum_i n_i \frac{m_i^4}{64\pi^2} \ln\left(\frac{m_i^2}{\mu^2} \right) + \rho_{\text{B}} + \rho_{\text{vac}}^{\text{EW}} + \rho_{\text{vac}}^{\text{QCD}} + \cdots, \tag{811}$$

where ρ_{B} indicates the contribution coming from the bared cosmological constant and where the dots means that other phase transition could contribute. Let us now evaluate the first term more precisely for the standard model of particles physics [51]. In this case, one has one scalar field, the Higgs boson with $n_{\text{H}} = 1$, $m_{\text{H}} \simeq 125$ GeV, the six quarks (fermions) for which $n_{\text{quarks}} = -4$ and $m_{\text{t}} \simeq 171.2\text{GeV}$, $m_{\text{b}} \simeq 4.2\text{GeV}$, $m_{\text{c}} \simeq 1.27\text{GeV}$, $m_{\text{s}} \simeq 0.104\text{GeV}$, $m_{\text{u}} \simeq 0.24\text{GeV}$ and $m_{\text{d}} \simeq 0.48\text{GeV}$, the leptons (fermions) with $n_{\text{leptons}} = 4$ and $m_{\text{e}} \simeq 0.511\text{MeV}$,

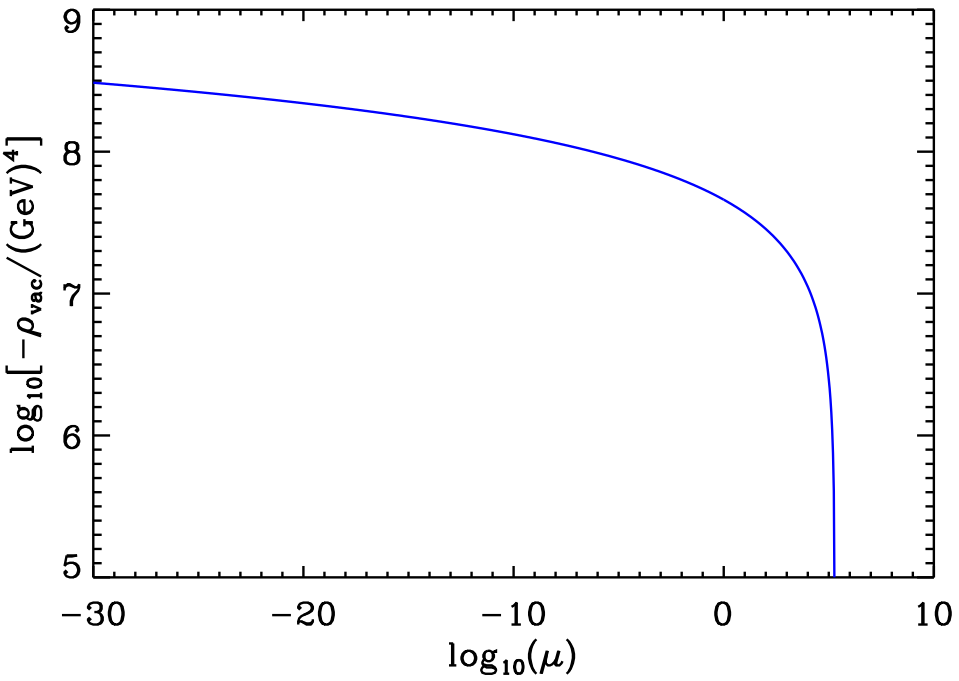


Figure 157: Evolution of the vacuum energy density [more precisely the first term in Eq. (811) versus the renormalization scale μ. In the range considered here, the vacuum energy density is negative. The "divergence" observed around $\log_{10}\mu \simeq 5$ does not correspond to a new physical effect but just signal that $\rho_{\rm vac}$ becomes positive.

$m_\mu \simeq 105$MeV, $m_\tau \simeq 1.77$MeV, the neutrinos (fermions) the mass of which is so small that we can ignore them and, finally, the gauge bosons (massive vector fields), $n_{\rm Z} = 3$, $m_{\rm Z} \simeq 91$GeV and $n_{\rm W^\pm} = 3$, $m_{\rm W^\pm} \simeq 80$GeV. Of course, we also have the photon but since $m_\gamma = 0$, as already mentioned before, it does not contribute to the vacuum energy. The only piece missing in order to calculate $\rho_{\rm vac}$ in Eq. (811) is the renormalization scale μ. It was argued in Ref. [51] that, because we use the photons coming from the supernovae to determine the cosmological constant and because these photons couple to the metric the expansion rate of which is characterized by the Hubble constant, one should take $\mu \sim \sqrt{E_\gamma E_{\rm grav}}$ with $E_{\rm grav} \simeq H_0 \simeq 3.7 \times 10^{-41}\,{\rm GeV}$ and with the energy of the photons corresponding to the wavelength $\lambda \simeq 500\,{\rm nm}$. This leads to $\mu \simeq 3 \times 10^{-25}\,{\rm GeV}$ and implies that Eq. (811) reads

$$\rho_{\rm vac} \simeq -2 \times 10^8\,{\rm GeV}^4 + \rho_{\rm B} + \rho_{\rm vac}^{\rm EW} + \rho_{\rm vac}^{\rm QCD} + \cdots . \tag{812}$$

If one does not want to estimate μ, one can simply plot $\rho_{\rm vac}$ as a function of μ as done in Fig. 157. Clearly, regardless of the precise value of μ, we are very far from $\rho_{\rm vac} \simeq 10^{72}\,{\rm GeV}^4$ always mentioned in the literature.

At this stage, Eq. (812) can be seen as the "prediction" of the standard model for the vacuum energy. At first sight, the calculation is straightforward and, in order to obtain the number in Eq. (812), we have used techniques of regularization that are known to work in similar contexts (*i.e.* when one computes a cross-section in particle physics) and to lead to a very good agreement with experiments. Therefore, we now turn to the question of what is known experimentally about the vacuum energy.

24.10 Measuring the Cosmological Constant in Cosmology

Having discussed how the vacuum energy can be calculated, we now address the question of how it is determined and/or constrained experimentally/observationally. Here, we should

distinguish the case of cosmology, treated in this section, and the other experiments, treated in the following sections. The difference is that, in cosmology, one claims a detection of Λ while, as we will see, the other experiments can only put an upper limit on its value. Moreover, in cosmology, there are in fact many different and independent observables that can be used to measure the cosmological constant [1445, 1524, 79, 80, 1677, 82]. All these techniques are reviewed in other articles [40, 41] and, therefore, it is not our purpose here to discuss this question in details. However, for completeness and also because this has some impact on the theoretical status of the cosmological constant problem, we now discuss how Λ can be determined. We will do so from a theoretical perspective leaving, as mentioned before, the technical aspects to other articles [40, 41].

24.10.1 The Accelerating Universe

As is well-known, the claim that vacuum energy has been measured (and not only constrained) was first established from the discovery that the expansion of the universe is accelerated (although one can find many claims about the value of the cosmological constant in the history of cosmology) [1445, 1524]. Since the above mentioned claim has clearly very important consequences for physics, one needs to be very precise at this stage and this is why, in what follows, we will review the exact origin of this result.

In order to construct a cosmological model we apply general relativity to the Universe as a whole [83]. The cosmological principle implies that the Universe is, on large scales, homogeneous and isotropic. As a consequence, the metric is given by the following expression

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\gamma^{(3)}_{ij}\mathrm{d}x^i\mathrm{d}x^j\,, \tag{813}$$

where $\gamma^{(3)}_{ij}$ is the metric of the three-dimensional space-like sections and where $a(t)$ is the scale factor. The expansion is characterized by the Hubble parameter, $H = \dot{a}/a$ where a dot means a derivative with respect to cosmic time. In the standard model of cosmology, matter is assumed to be a collection of N perfect fluids (at least in the simplest version) and, as a consequence, its stress-energy tensor is given by the following expression

$$T_{\mu\nu} = \sum_{i=1}^{N} T^{(i)}_{\mu\nu} = (\rho_{\scriptscriptstyle\mathrm{T}} + p_{\scriptscriptstyle\mathrm{T}})u_\mu u_\nu + p_{\scriptscriptstyle\mathrm{T}} g_{\mu\nu}\,, \tag{814}$$

where $\rho_{\scriptscriptstyle\mathrm{T}}$ is the (total) energy density and $p_{\scriptscriptstyle\mathrm{T}}$ the (total) pressure. These two quantities are related by the equation of state

$$p_{\scriptscriptstyle\mathrm{T}} = w(\rho_{\scriptscriptstyle\mathrm{T}}), \tag{815}$$

[in general, there is an equation of state per fluid considered *i.e.* $p_i = w_i(\rho_i)$]. The vector u_μ is the four velocity and satisfies the relation $u_\mu u^\mu = -1$. In terms of cosmic time this means that $u^\mu = (1, \mathbf{0})$. The fact that the stress-energy tensor is conserved, $\nabla^\alpha T_{\alpha\mu} = 0$, amounts to

$$\dot{\rho}_{\scriptscriptstyle\mathrm{T}} + 3\frac{\dot{a}}{a}(\rho_{\scriptscriptstyle\mathrm{T}} + p_{\scriptscriptstyle\mathrm{T}}) = 0,. \tag{816}$$

This expression is obtained when one takes $\mu = 0$. The case $\mu = i$ does not lead to an interesting equation for the background. Moreover, if we assume that the fluids are separately conserved, then the above relation is in fact true for each fluid, *i.e.* $\dot{\rho}_i + 3H(\rho_i + p_i) = 0$.

We are now in a position to write down the Einstein equations (306) for the Universe. We arrive at the following non-linear, second order, differential equations

$$\frac{\dot{a}^2}{a^2} + \frac{k}{a^2} = \frac{\kappa}{3}\sum_{i=1}^{N}\rho_i + \frac{\Lambda_{\scriptscriptstyle\mathrm{eff}}}{3}\,, \tag{817}$$

$$-\left(2\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{k}{a^2}\right) = \kappa\sum_{i=1}^{N} p_i - \Lambda_{\scriptscriptstyle\mathrm{eff}}\,, \tag{818}$$

From the previous formulas, as already signaled in Eq. (307), we see that we can also define ρ_{Λ_B} by $\rho_{\Lambda_B} = \Lambda_B/\kappa$ and p_{Λ_B} by $p_{\Lambda_B} = -\rho_{\Lambda_B} = -\Lambda_B/\kappa$. In this case, the cosmological constant is described as fluid with a constant energy density and pressure and with an equation of state $p_{\Lambda_B} + \rho_{\Lambda_B} = 0$, i.e.

$$w_{\rm vac} \equiv \frac{p_{\Lambda_B}}{\rho_{\Lambda_B}} = -1\,. \tag{819}$$

This is of course consistent with Eq. (301). It is worth noticing that the vacuum energy is described by a fluid with a negative effective pressure.

Combining the two Einstein equations, one gets an equation which permits to calculate the acceleration of the scale factor

$$\frac{\ddot{a}}{a} = -\frac{\kappa}{6}\sum_{i=1}^{N}(\rho_i + 3p_i) + \frac{2}{3}\Lambda_{\rm eff}\,. \tag{820}$$

Let us notice that, in order to obtain the above formula, we have not assumed anything about the curvature k. This equation is especially interesting because it gives the condition leading to an accelerated expansion, namely

$$\rho_T + 3p_T < 0\,. \tag{821}$$

Since the energy density of matter must be positive, we see that the above condition requires a negative pressure, *i.e.* some exotic form of matter. But, this is of course exactly what happens when the vacuum energy is the dominant fluid in the universe. Then, Eq. (820) can be written as $\ddot{a}/a \simeq 2\Lambda_{\rm eff}/3$ which is positive for a positive effective cosmological constant. Therefore, in the framework of the standard model described before, the discovery that $\ddot{a} > 0$ means that a fluid with a negative pressure is driving the expansion.

24.10.2 The Hubble Diagram

In this section, we explain how the measurement of the Hubble diagram can be used to infer the value of the cosmological constant.

Let us first consider a source in flat space-time whose absolute luminosity is given by L (energy emitted per unit of time). This quantity can be written as

$$L = \frac{n_E \hbar\nu}{\Delta t}\,, \tag{822}$$

where n_E is the number of photons emitted during the time interval Δt. Suppose that we observe the source from a distance d_L using a telescope, the surface of the mirror being ΔS. The number of photons received is given by

$$n_R = n_E\left(\frac{\Delta S}{4\pi d_L^2}\right). \tag{823}$$

On the other hand, the flux of photons (or apparent luminosity *i.e.* energy per unit of time and surface) through the telescope is, by definition,

$$\Phi_R = \frac{n_R \hbar\nu}{\Delta S \Delta t}\,, \tag{824}$$

from which we deduce that

$$\Phi_R = \frac{L}{4\pi d_L^2}\,. \tag{825}$$

We have just recovered the well-known fact that the apparent luminosity of a source decreases as the inverse squared of the distance. On the other hand, we can use this relation as an operational definition of the distance, namely $d_L = \sqrt{L/(4\pi\Phi_R)}$.

The previous considerations can be used as a definition of distance in cosmology. For this purpose, let us now consider the same situation but in a curved space-time with the following metric [a form more explicit than the one given in Eq. (813)]

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)(\mathrm{d}r^2 + r^2\mathrm{d}\Omega_2^2)\,. \tag{826}$$

We assume that the source is located at the co-moving coordinate r_1 while the observer is located at the origin of the coordinates. In order to reproduce the above flat space-time calculation in this new situation, we first need to calculate the surface of the sphere $t = t_{\mathrm{R}}$ =const., $r = r_1$ =const. surrounding the source. The surface is just given by the integral of the covariant volume element, namely

$$S = \int \sqrt{g}\,\mathrm{d}^2x = \int_0^{\pi}\int_0^{2\pi} \sqrt{a_{\mathrm{R}}^4 r_1^4}\,\sin\theta \mathrm{d}\theta\mathrm{d}\varphi = 4\pi a_{\mathrm{R}}^2 r_1^2\,. \tag{827}$$

This means that the number of photons received in the telescope can now be expressed as

$$n_{\mathrm{R}} = n_{\mathrm{E}}\left(\frac{\Delta S}{4\pi a_{\mathrm{R}}^2 r_1^2}\right). \tag{828}$$

But we have also to take into account the two following phenomena. Firstly, the energy of the photons has changed during the propagation from the source to the observer. If λ is the wavelength, we have $\lambda = (2\pi/k)a(t) \propto \nu^{-1}$, from which we deduce that $E_{\mathrm{R}}/E_{\mathrm{E}} = \nu_{\mathrm{R}}/\nu_{\mathrm{E}} = a_{\mathrm{E}}/a_{\mathrm{R}}$, that is to say we have the redshift due to the expansion. Secondly, if the photons are emitted as a burst of duration Δt_{E}, they will be received as burst of duration $\Delta t_{\mathrm{R}} = \Delta t_{\mathrm{E}} a_{\mathrm{R}}/a_{\mathrm{E}}$. We are now in a position where we can calculate the apparent luminosity. We obtain

$$\Phi_{\mathrm{R}} = \frac{L}{4\pi a_{\mathrm{R}}^2 r_1^2}\left(\frac{a_{\mathrm{E}}}{a_{\mathrm{R}}}\right)^2 = \frac{L}{4\pi a_{\mathrm{R}}^2 r_1^2}\frac{1}{(1+z)^2}\,, \tag{829}$$

where z is the redshift of the source. Now if we use the operational definition of the luminosity distance (825), we find

$$d_{\mathrm{L}}(z) = a_{\mathrm{R}} r_1 (1+z) = a_{\mathrm{R}}(1+z)\int_{t_{\mathrm{E}}}^{t_{\mathrm{R}}} \frac{c\mathrm{d}\tau}{a(\tau)}\,. \tag{830}$$

We see from the last expression that the distance of a source depends on the behavior of the scale factor between the time of emission and the time of reception. On the contrary, if we measure the luminosity distance versus the redshift, we can learn about this behavior. Measuring the apparent luminosity is quite easy and, clearly, the difficulty lies in estimating the absolute luminosity of the source. It is also convenient to work in terms of the distance modulus, $m - M$, since the data are often presented in this way. The quantity m (or sometimes m_{bol}) is the apparent bolometric magnitude. It is related to the apparent bolometric luminosity ℓ by

$$\ell = 10^{-2m/5} \times 2.52 \times 10^{-5}\,\mathrm{erg/cm^2} \times \mathrm{sec}\,. \tag{831}$$

For instance the sun has an apparent bolometic magnitude of -26.85. On the other hand, the absolute bolometric magnitude M is the apparent bolometric magnitude that the source would have at a distance of 10 pc and is given by

$$L = 10^{-2M/5} \times 3.02 \times 10^{35}\,\mathrm{erg/sec}\,. \tag{832}$$

For instance, the absolute bolometric magnitude of the sun is 4.72. As a consequence, using these two definitions, the distance modulus is defined by

$$d_{\mathrm{L}}(z) \equiv \frac{c}{H_0}\bar{d}_{\mathrm{L}}(z) \equiv 10^{1+(m-M)/5} \times 10^{-6}\mathrm{Mpc}\,, \tag{833}$$

such that $\bar{d}_{\rm L}$ is dimensionless and $c/H_0 = 3000\,h^{-1}$Mpc is of course nothing but the Hubble length (we have re-established the speed of light for convenience). From the above expression, one immediately deduces that

$$\mu(z) \equiv m - M = 5\log_{10}\bar{d}_{\rm L}(z) + 25 + 5\log_{10}\left(\frac{c}{H_0}\right)\,, \tag{834}$$

where c is expressed in km $\times$ s^{-1}, H_0 in km$\times$s^{-1} $\times$ Mpc^{-1} and $d_{\rm L}$ in Mpc.

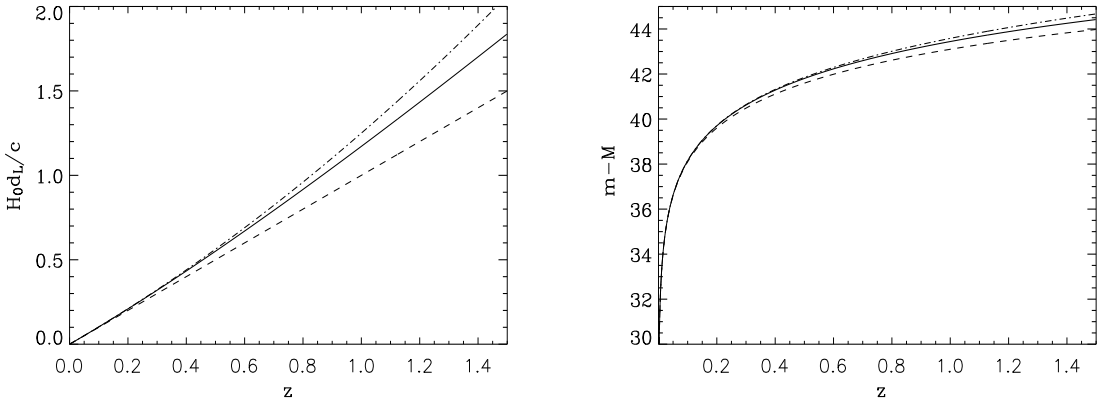


Figure 158: Luminosity distance $d_{\rm L}(z)$ (left panel) and distance modulus $\mu(z)$ (right panel) in a matter dominated Universe. The solid line corresponds to the exact result. The dashed line is the linear law $d_{\rm L} \simeq cz/H_0$ and the dotted dashed line represents the approximated result at second order, see Eq. (840). We have used $h = 0.72$.

In practice we only have access to $d_{\rm L}$ over a limited range of redshifts. Therefore, one can derive a general expression by Taylor expanding the scale factor. For small redshifts, we have

$$a(t) = a(t_{\rm R}) + \dot{a}(t_{\rm R})(t-t_{\rm R}) + \frac{1}{2}\ddot{a}(t_{\rm R})(t-t_{\rm R})^2 + \cdots = a(t_{\rm R})\left[1 + H_{\rm R}(t-t_{\rm R}) - \frac{1}{2}H_{\rm R}^2 q_{\rm R}(t-t_{\rm R})^2 + \cdots\right], \tag{835}$$

where $q \equiv -a\ddot{a}/(\dot{a})^2$ is the acceleration parameter. A positive acceleration parameter corresponds to a decelerating universe whether a negative one corresponds to an accelerating universe. Using the previous expression, the luminosity distance becomes

$$\begin{aligned} d_{\rm L}(z) &= c(1+z)\int_{t_{\rm E}}^{t_{\rm R}} {\rm d}\tau\left[1 - H_{\rm R}(\tau - t_{\rm R}) + \frac{1}{2}H_{\rm R}^2 q_{\rm R}(\tau - t_{\rm R})^2 + H_{\rm R}^2 q_{\rm R}(\tau - t_{\rm R})^2 + \cdots\right] &(836)\\ &= c(1+z)\left[-(t_{\rm E} - t_{\rm R}) + \frac{1}{2}H_{\rm R}(t_{\rm E} - t_{\rm R})^2 - \frac{1}{6}H_{\rm R}^2 q_{\rm R}(t_{\rm E} - t_{\rm R})^3 - \frac{1}{3}H_{\rm R}^2 q_{\rm R}(t_{\rm E} - t_{\rm R})^3 + \cdots\right]. &(837)\end{aligned}$$

As expected the result only depends on the time of flight $t_{\rm E} - t_{\rm R}$. In order to obtain our final result, we must express this time of flight in terms of the redshift. The redshift of the source is defined by $z = a(t_{\rm R})/a(t_{\rm E}) - 1$. Using the Taylor expansion of the scale factor, we end up with

$$z = -H_{\rm R}(t_{\rm E} - t_{\rm R}) + H_{\rm R}^2\left(1 + \frac{1}{2}q_{\rm R}\right)(t_{\rm E} - t_{\rm R})^2 + \cdots\,, \tag{838}$$

which can be easily inverted to give the expression of the time of flight, namely

$$t_{\rm E} - t_{\rm R} = -\frac{z}{H_{\rm R}} + \frac{1}{H_{\rm R}}\left(1 + \frac{1}{2}q_{\rm R}\right)z^2 + \cdots\,. \tag{839}$$

The last step is to replace the above expression into the Eq. (837). This gives

$$d_{\rm L}(z) = \frac{c}{H_0}\left[z + \frac{1}{2}(1-q_0)z^2 + \cdots\right], \tag{840}$$

where we have taken into account that the time of reception is the present time (denoted with the subscript "0"). We see that the measurement of the luminosity distance versus the redshift allows us to determine the Hubble parameter today. If one goes sufficiently far in redshifts, one can determine whether the universe is accelerating or not via the accelerating parameter q_0. It is worth noticing that this last conclusion is independent of the underlying gravity theory (provided this is a metric theory). The typical behavior of the luminosity distance and of the modulus is illustrated in Fig. 158 for a matter dominated universe. Let us also notice that the method that we have just presented for pedagogical reasons is of course not used in practice because we have now measured up to redshifts of $z \simeq 1.4$ for which the Taylor expansion is not valid (and, of course, in the real world the analysis is always more complicated that the simplified version discussed above). In this case, one must use an exact expression, that is to say one must compute the integral in Eq. (830) exactly.

If we now use general relativity (*i.e.* a specific theory for gravity), one can go further and relates the value of the acceleration parameter to the matter content of the Universe. The critical energy density $\rho_{\rm cri}$, defined by $\rho_{\rm cri} \equiv 3H^2/\kappa$, is used to evaluate the contribution of each species through the parameter $\Omega_i \equiv \rho_i/\rho_{\rm cri}$ where ρ_i is the energy density of the species i at present time. Then, the Friedman equation can be re-written as

$$\sum_{i=1}^{N} \Omega_i = 1 + \frac{k}{a^2 H^2} \,, \tag{841}$$

and the acceleration parameter can be re-expressed according to

$$q_0 = \frac{1}{2} \sum_{i=1}^{N} \Omega_i \left(1 + 3w_i\right) \,. \tag{842}$$

The universe contains pressure-less gas (cold dark matter and ordinary baryonic matter) for which $w = 0$ but also radiation ($w = 1/3$) and, finally, vacuum energy ($w = -1$). In this case, one obtains

$$q_0 = \frac{1}{2}\Omega_{\rm m} + \Omega_\gamma - \Omega_{\rm vac} \,. \tag{843}$$

Today, the contribution of radiation is negligible since we know from the measurement of the CMB that $\Omega_\gamma h^2 \simeq 2.48 \times 10^{-5}$ [1677]. Therefore, the acceleration parameter is in fact given by $q_0 \simeq \Omega_{\rm m}/2 - \Omega_{\rm vac}$. On the other, we also know that the universe is spatially flat which means that $\Omega_{\rm m} + \Omega_{\rm vac} \simeq 1$ [1677]. As a consequence, from the measurement of $q_0 \simeq -0.67 \pm 0.25$ [1445, 1524], one can deduce the two quantities $\Omega_{\rm m}$ and $\Omega_{\rm vac}$. On finds that $\Omega_{\rm m} \simeq 0.3$ and $\Omega_{\rm vac} \simeq 0.7$. This is how one reaches the conclusion that our universe is accelerating and that vacuum energy dominates the present energy content. Remarkably, as announced at the beginning of this section, it is a detection and a measurement of $\rho_{\rm vac}$. One finds

$$\rho_{\rm vac} \simeq \Omega_{\rm vac} \rho_{\rm cri} \simeq 10^{-47}\,{\rm GeV}^4, \tag{844}$$

a number that should be compared with Eq. (812).

At this stage, several remarks are in order. Firstly, we notice the large mismatch between the theoretical expectation (812) and the above, observationally determined, number (844), something like 54 orders of magnitude (but much less than the 122 orders of magnitude often quoted in the literature). This is of course nothing but the cosmological constant problem although we will be more accurate with regards to its definition in the subsequent sections. Something is clearly wrong but, as will be discussed, it is difficult to identify where the mistake is. Secondly, it is clear that the above conclusion rests on many assumptions. For instance, we have assumed that the universe is homogeneous and isotropic [42] and that gravity is well described by general relativity. This means that, a priori, the cosmological constant problem is present in this context only. Thirdly, we have also implicitly assumed that the reason for the accelerated expansion is vacuum energy. This is clearly a natural idea since there is a term in our theory that must be present since it satisfies all the required properties (covariance, energy conservation) and that precisely leads to the observed phenomenon. A priori, nothing more can be asked to a new theoretical framework. But, given the mismatch discussed before, one can also doubt the identification of the source of the accelerated expansion with vacuum energy. As a

matter of fact, it is possible to construct models where the universe accelerates because of some new source of matter (often named "dark energy"). This is for instance the case of quintessence and/or galileons [43] models where a scalar field is responsible for the acceleration. In this case, one can no longer claim that we have detected the vacuum energy and that its value is given by Eq. (844). However, this does not solve the cosmological constant problem because, even if the source of the acceleration is another fluid, we still have the constraint $\rho_{\rm vac} < \rho_{\rm cri}$. Clearly, the number (812) does not satisfy this inequality which means that the problem is still present. In this sense, the constraint deduced from cosmology is very important: even if dark energy is not the cosmological constant, the observation that the energy density today is the critical energy density severely limits the value of $\Lambda_{\rm eff}$. Fourthly, in principle, it is possible to check experimentally whether the reason for the accelerated expansion is vacuum energy or something else like quintessence [1497, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 448, 97, 98, 99] (more recently, in Ref. [100], the current acceleration was also explained by a scalar field experiencing quantum fluctuations during inflation). Measuring the equation of state parameter w is the clue since, if this is a constant equals to -1, then one can be sure that the source of the acceleration is the cosmological constant. Unfortunately, this measurement is difficult [101].

One more general grounds, it is clear that it would be better to have constraints, or even a measurement, of vacuum energy in an experimental context which is is not cosmology where, as reminded before, measurements are difficult and always subject to many biases and systematic errors. In the next sections, we explore whether this is possible.

24.11 Measuring the Cosmological Constant Elsewhere Than In Cosmology?

In the previous section, we have written the Einstein equation in an homogeneous and isotropic universe. Here, we investigate the influence of the cosmological term in a spherically symmetric situation. This will allow us to study how the presence of $\Lambda_{\rm eff}$ affects the motion of the planets in our solar system and how it modifies the energy levels of the Hydrogen atom.

24.11.1 The Static and Spherically Symmetric Gravitational Field in Presence of a Cosmological Constant

Let us analyze the influence of the cosmological constant in a static and spherically symmetric situation [102]. This means that, in spherical coordinates, one can write the metric tensor as [102]

$$\mathrm{d}s^2 = -B(r)\mathrm{d}t^2 + A(r)\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\varphi^2\right), \tag{845}$$

where $A(r)$ and $B(r)$ are two free functions. We seek solutions of the Einstein equations (306) with a vanishing stress energy tensor. It is easy to show that these equations can be expressed as

$$R_{\mu\nu} - \Lambda_{\rm eff} g_{\mu\nu} = 0. \tag{846}$$

The next step is to calculate the components of the Ricci tensor for the metric tensor given by Eq. (845). Straightforward calculations lead to the following expressions

$$R_{tt} = \frac{B}{A}\left(\frac{1}{2}\frac{B''}{B} - \frac{1}{4}\frac{B'^2}{B^2} - \frac{1}{4}\frac{A'B'}{AB} + \frac{1}{r}\frac{B'}{B}\right), \tag{847}$$

$$R_{rr} = -\frac{1}{2}\frac{B''}{B} + \frac{1}{4}\frac{B'^2}{B^2} + \frac{1}{4}\frac{A'B'}{AB} + \frac{1}{r}\frac{B'}{B}, \tag{848}$$

$$R_{\theta\theta} = \frac{1}{A}\left(\frac{r}{2}\frac{A'}{A} - \frac{r}{2}\frac{B'}{B} - 1\right) + 1, \tag{849}$$

$$R_{\varphi\varphi} = R_{\theta\theta}\sin^2\theta, \tag{850}$$

the other components being zero. In the above equations, a prime denotes a derivative with respect to the radial coordinate r.

We are now in a position where one can determine the solution of the field equations. If we take the combination $R_{tt}/B + R_{rr}/A = -\Lambda_{\text{eff}}B/B + \Lambda_{\text{eff}}A/A = 0$ [from the Einstein equations (846)], one obtains

$$\frac{1}{rA}\left(\frac{B'}{B} + \frac{A'}{A}\right) = 0, \tag{851}$$

which implies that $B(r) = C/A(r)$, where C is a constant. Using this expression, one can re-express $R_{\theta\theta}$ as

$$R_{\theta\theta} = \frac{B(r)}{C}\left(-r\frac{B'}{B} - 1\right) + 1, \tag{852}$$

and, as a consequence, the corresponding Einstein equation now reads

$$rB'(r) + B(r) = C - C\Lambda_{\text{eff}}r^2. \tag{853}$$

This equation can easily be integrated and one obtains

$$B(r) = D + \frac{C}{r} - D\Lambda_{\text{eff}}\frac{r^2}{3}, \tag{854}$$

where D is another integration constant. The constants C and D are fixed in such a way that, in absence of a cosmological constant, one recovers the standard Schwarschild solution. This amounts to take $D = 1$ and $C = 2GM$, where M is the mass of the central body and G the Newton constant. This completes our calculation since the two functions $A(r)$ and $B(r)$ are now completely specified, namely

$$A(r) = \frac{1}{B(r)}, \quad B(r) = 1 - \frac{2GM}{r} - \frac{\Lambda_{\text{eff}}r^2}{3}. \tag{855}$$

We can now compare this solution to the standard Schwarschild solution. The fact that the time-time component of the metric tensor is inversely proportional to the r-r component is still true but the radial dependence of the function $B(r)$ is modified by the presence of a cosmological constant. In the Newtonian limit, we have $g_{tt} \simeq 1 + 2V(r)$, where $V(r)$ is the Newtonian potential. In the present context, this implies that

$$V(r) \simeq -\frac{GM}{r} - \frac{\Lambda_{\text{eff}}r^2}{6}. \tag{856}$$

Therefore, we obtain the usual gravitational potential corrected by a term proportional to the cosmological constant. This term looks like an inverted harmonic oscillator. This corresponds to a force felt by a test body of mass m given by

$$\boldsymbol{F} = \frac{GMm}{r^2}\boldsymbol{u} - \frac{m\Lambda_{\text{eff}}r}{3}\boldsymbol{u}, \tag{857}$$

where $\boldsymbol{u}$ is a vector directed towards the central body of mass M. The cosmological constant force is directed in the other direction and can therefore be viewed as a kind of "anti-gravity force". We also notice that it is proportional to r which means that its effect will be more important on large scales. This last property is the clue to understand why measuring the cosmological constant is more efficient on large scales (the best example being of course cosmology). This means that the modification of the planet orbits and/or the energy levels of the atoms will certainly be a small effect. On the other hand, we know that measurements in, say, atomic physics are extremely accurate and, a priori, there is the hope that this accuracy could compensate the smallness of the effect. Unfortunately, as we are now going to study, this will not be the case.

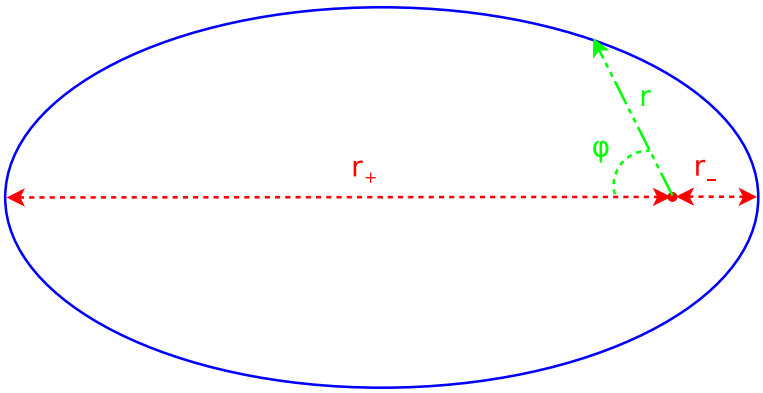


Figure 159: Sketch of the elliptic planet orbit with the aphelion and perihelion. The orbit is planar and the position of the planet can be characterized by the coordinates (r, φ).

24.11.2 Planets Orbits

Our goal in this section is to study how the planet orbits are modified by the presence of vacuum energy [103, 104, 105]. We will follow the treatment of Ref. [103]. We assume that the gravitational field created by the sun and the cosmological constant is given by Eqs. (845) and (855) with $M = M_\odot$. The motion of a planet is represented by the motion of a test body in this background gravitational field. The corresponding geodesic equation can be written as

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} + \Gamma^\mu_{\nu\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\frac{\mathrm{d}x^\lambda}{\mathrm{d}\tau} = 0, \tag{858}$$

where τ is an affine parameter (not necessarily the proper time). Then, after having evaluated the Christoffel symbols for the metric (845), one can express these equations explicitly. One obtains

$$\frac{\mathrm{d}^2 t}{\mathrm{d}\tau^2} + \frac{B'}{B}\frac{\mathrm{d}t}{\mathrm{d}\tau}\frac{\mathrm{d}r}{\mathrm{d}\tau} = 0, \tag{859}$$

$$\frac{\mathrm{d}^2 r}{\mathrm{d}\tau^2} + \frac{B'}{2A}\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 + \frac{A'}{2A}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 - \frac{r}{A}\left(\frac{\mathrm{d}\theta}{\mathrm{d}\tau}\right)^2 - \frac{r\sin^2\theta}{A}\left(\frac{\mathrm{d}\varphi}{\mathrm{d}\tau}\right)^2 = 0, \tag{860}$$

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}\tau^2} + \frac{2}{r}\frac{\mathrm{d}r}{\mathrm{d}\tau}\frac{\mathrm{d}\theta}{\mathrm{d}\tau} - \cos\theta\sin\theta\left(\frac{\mathrm{d}\varphi}{\mathrm{d}\tau}\right)^2 = 0, \tag{861}$$

$$\frac{\mathrm{d}^2\varphi}{\mathrm{d}\tau^2} + \frac{2}{r}\frac{\mathrm{d}r}{\mathrm{d}\tau}\frac{\mathrm{d}\varphi}{\mathrm{d}\tau} + 2\frac{\cos\theta}{\sin\theta}\frac{\mathrm{d}\theta}{\mathrm{d}\tau}\frac{\mathrm{d}\varphi}{\mathrm{d}\tau} = 0, \tag{862}$$

Let us now solve these equations. Firstly, we can assume that the orbit stays in the $\theta = \pi/2$ plane. This automatically solves Eq. (861). Then, if we divide Eq. (859) by $\mathrm{d}t/\mathrm{d}\tau$ and Eq. (862) by $\mathrm{d}\varphi/\mathrm{d}\tau$, we obtain two new equations, namely

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left[\ln\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right) + \ln B\right] = 0, \tag{863}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left[\ln\left(\frac{\mathrm{d}\varphi}{\mathrm{d}\tau}\right) + \ln r^2\right] = 0. \tag{864}$$

The first equation can be used to define the affine parameter. We choose it such that it satisfies $\mathrm{d}t/\mathrm{d}\tau = 1/B$. Then, the second equation tells us that

$$r^2\frac{\mathrm{d}\varphi}{\mathrm{d}\tau} = J, \tag{865}$$

where J is a constant. Finally, Eq. (860) can be re-written as

$$\frac{\mathrm{d}^2 r}{\mathrm{d}\tau^2} + \frac{A'}{2A}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 - \frac{J^2}{Ar^3} + \frac{B'}{2AB^2} = 0. \tag{866}$$

If we multiply this last equation by $2A\mathrm{d}r/\mathrm{d}\tau$, one arrives at

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left[A\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2+\frac{J^2}{r^2}-\frac{1}{B}\right]=0. \tag{867}$$

This means that the quantity inside the square brackets is a constant. Let us call this constant $-E$. Then, this gives an equation expressing how the radial coordinate r varies with the angle φ,

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\varphi}\right)^2=\frac{r^4}{A(r)}\left[\frac{1}{J^2B(r)}-\frac{1}{r^2}-\frac{E}{J^2}\right]. \tag{868}$$

It is easy to express the constants E and J in terms of the aphelion and perihelion r_+ and r_-, which are the two points on the orbit such that $\mathrm{d}r/\mathrm{d}\varphi=0$, see Fig. 159 (where the elements of the orbits are represented). This gives

$$\frac{1}{J^2} = \frac{r_+^{-2}-r_-^{-2}}{B_+^{-1}-B_-^{-1}}, \tag{869}$$

$$E = \frac{r_+^2/B_+-r_-^2/B_-}{r_+^2-r_-^2}, \tag{870}$$

with $B_\pm$ defined by $B_\pm\equiv B(r_\pm)$.

So far, we have been very general and we have never specified the function $B(r)$. From now on, we are going to use the explicit form of this function, see Eq. (855). Let us notice in particular that, in absence of a cosmological constant and in the limit where $GM/r\ll 1$, the two above expressions reduce to

$$\frac{E-1}{2}\simeq\frac{GM}{r_++r_-},\quad J^2\simeq 2GMD, \tag{871}$$

where $D\equiv r_+r_-/(r_++r_-)$ (of course, not to be confused with the integration constant introduced in the previous sub-section). As is shown in the following footnote, these are exactly the relations obtained in the Newtonian case. More generally, Eq. (868) takes the form

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\varphi}\right)^2 = \frac{1-E}{J^2}r^4+\frac{2GME}{J^2}r^3-r^2+2GMr \\ +\frac{\Lambda_{\rm eff}r^4}{3}+\frac{E\Lambda_{\rm eff}r^6}{3J^2}. \tag{872}$$

Performing the standard change of variable, $u\equiv 1/r$ and differentiating with respect to φ, we obtain [129]

$$\frac{\mathrm{d}^2u}{\mathrm{d}\varphi^2}+u=\frac{E}{J^2}GM+3GMu^2-\frac{E}{J^2}\frac{\Lambda_{\rm eff}}{3u^3}. \tag{881}$$

[129] It is interesting to compare this result with the Newtonian analysis. In this case, the Lagrangian is given by

$$L=\frac{m}{2}\left(\dot r^2+r^2\dot\varphi^2\right)+\frac{GMm}{r}, \tag{873}$$

where a dot denotes a derivative with respect to time. Again we have chosen the orbit to be in the plane $\theta=\pi/2$. This leads to the following equations of motion

$$m\ddot r-mr\dot\varphi^2+\frac{GMm}{r^2} = 0, \tag{874}$$

$$mr^2\ddot\varphi+2mr\dot r\dot\varphi = 0. \tag{875}$$

The second equation gives $\dot\varphi=J/r^2$ while the first one can be re-written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\dot r^2}{2}+\frac{J^2}{2r^2}-\frac{GM}{r}\right)=0 \tag{876}$$

This is the differential equation that needs to be solved in order to find the trajectory of the planet. It should be compared with its Newtonian counterpart, see Eq. (879). One can check that, in this regime (without a cosmological constant), the two equations are indeed identical since according to Eq. (871), $E \to 1$, $\Lambda_{\text{eff}} = 0$ and the term $3GMu^2$ can be neglected. The relativistic term $3GMu^2$ is responsible for the precession of the perihelion and the last term represents the correction due to the cosmological constant.

We now turn to the solution of Eq. (881). Let us first review how one calculates the precession of the perihelion without a cosmological constant [102]. The idea is to restart from Eq. (872). For $\Lambda_{\text{eff}} = 0$, the polynomial on the right hand side can be written as

$$\frac{1-E}{J^2} r(r-r_+)(r-r_-)(r-\epsilon) \tag{882}$$

since we know it has to vanish at $r = r_\pm$. The only quantity that remains to be determined in this expression is ϵ. Comparing the above relation with Eq. (872), we deduce that

$$\frac{1-E}{J^2} r_+ r_- \epsilon = 2GM. \tag{883}$$

Using the expressions derived before, it is easy to show that $\epsilon = 2GM/(1-2GM/D)$. As a consequence, one can write

$$\begin{aligned} \varphi_+ - \varphi_- &= \frac{J}{\sqrt{1-E}} \int_{r_-}^{r_+} \frac{1}{\sqrt{(r-r_+)(r-r_-)}} \\ &\quad \times \frac{1}{\sqrt{1-\epsilon/r}} \frac{\mathrm{d}r}{r}. \end{aligned} \tag{884}$$

There are several methods to evaluate this integral. As a matter of fact, it can be integrated exactly in terms of Elliptic functions [48, 49]. But, being given that ϵ is a small quantity, it is better to expand the result in this parameter. This leads to easier calculations that are, moreover, more explicit. In addition, the method can be generalized to the case of a non-vanishing cosmological constant, which is of course our main goal. In fact, if we restart from Eq. (868) and use that $B^{-1} \simeq 1 + 2GM/r + \cdots$, we see that the term in the squared bracket in Eq. (868) is a quadratic function of $1/r$ which vanishes at $r = r_\pm$. This means that one necessarily has

$$\frac{1}{J^2 B(r)} - \frac{E}{J^2} - \frac{1}{r^2} \simeq C \left(\frac{1}{r} - \frac{1}{r_-}\right)\left(\frac{1}{r} - \frac{1}{r_+}\right). \tag{885}$$

The quantity between the parenthesis is a constant and we call it $-E$. Then, this equation can be re-expressed as

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\varphi}\right)^2 = -\frac{2E}{J^2} r^4 + \frac{2GM}{J^2} r^3 - r^2. \tag{877}$$

From this equation one can deduce an expression for E and J in terms of the aphelion and the perihelion. One gets

$$E = \frac{GM}{r_+ + r_-}, \quad J^2 = 2GMD. \tag{878}$$

Introducing the variable $u = 1/r$ and differentiating once more with respect to φ, one arrives at

$$\frac{\mathrm{d}^2 u}{\mathrm{d}\varphi^2} + u = \frac{GM}{J^2}. \tag{879}$$

This equation can easily be integrated. The solution reads

$$u(\varphi) = F\cos(\varphi - \varphi_0) + \frac{GM}{J^2} = \frac{GM}{J^2}(1 - e\cos\varphi), \tag{880}$$

where F and φ_0 are two arbitrary constants. Clearly, one can always take $\varphi_0 = 0$. The above trajectory represents, as is well-known, an ellipse with eccentricity e.

Then, writing Eq. (868) in terms of u rather than r, one arrives at

$$\left(\frac{\mathrm{d}u}{\mathrm{d}\varphi}\right)^2 \simeq \frac{C}{A(r)}\left(u-u_-\right)\left(u-u_+\right). \tag{886}$$

Now, when $\Lambda_{\text{eff}} \neq 0$, as is clear from Eq. (872), the term in the square bracket of Eq. (868) is no longer a quadratic function of $1/r$ even if the expansion of the function $B^{-1}(r)$ is used. However, this term still vanishes at $r = r_\pm$ and the integral will be dominated by the contributions coming from the vicinity of these points, see also Eq. (884). The idea is then to continue to work with Eq. (885) and to take into account the cosmological constant in the definition of C [103]. This should provide a reasonable approximation of $\varphi_+ - \varphi_-$. Using Eqs. (869) and (870), one can write

$$\begin{aligned}
&\frac{1}{J^2 B(r)} - \frac{E}{J^2} - \frac{1}{r^2} \\
= \;& \frac{r_-^2\left[B^{-1}(r) - B_-^{-1}\right] - r_+^2\left[B^{-1}(r) - B_+^{-1}\right]}{r_+^2 r_-^2\left(B_+^{-1} - B_-^{-1}\right)} - \frac{1}{r^2} \\
\simeq \;& C\left(\frac{1}{r} - \frac{1}{r_-}\right)\left(\frac{1}{r} - \frac{1}{r_+}\right).
\end{aligned} \tag{887}$$

Then, if we differentiate both sides twice with respect to u, this gives

$$C = \frac{\left(u_+ - u_-\right)\left(u_+ - u_-\right)A''(u)}{2\left(u_+ - u_-\right)A'(u)} - 1, \tag{888}$$

$$\simeq -1 + \left.\frac{uA''}{A'}\right|_{u=L^{-1}}, \tag{889}$$

where $L \equiv 2D$ is the semilatus rectum. In the above expression, we have used $A_+ - A_- \simeq \left(u_+ - u_-\right)A'\left(L^{-1}\right)$. So far, the method is completely general provided that $A^{-1} = B$. We now use the explicit form of A^{-1}. Expanding everything in GMu and in Λ_{eff}, one obtains [103]

$$C \simeq -1 + \frac{4GM}{L} + \frac{\Lambda_{\text{eff}} L^3}{GM} + \frac{4\Lambda_{\text{eff}} L^3}{3} + \cdots. \tag{890}$$

The final step consists in using the above expression in Eq. (886). One finds [103]

$$\begin{aligned}
\varphi_+ - \varphi_- &\simeq \int_{u_-}^{u_+} \frac{A^{1/2}(u)\mathrm{d}u}{\left[C\left(u-u_+\right)\left(u-u_-\right)\right]^{1/2}} \\
&= \pi\left(1 + \frac{3}{2}\frac{2GM}{L} + \frac{\Lambda_{\text{eff}} L^3}{2GM} + \cdots\right),
\end{aligned} \tag{891}$$

where we have expanded $A^{1/2}C^{-1/2}$ and where we have used the fact that $\int_{u_-}^{u_+} \mathrm{d}u/[(u-u_+)(u-u_-)]^{1/2} = \ln(u_+ - u_-) - \ln(u_+ - u_-) + 2i\pi/2$, the complex i being exactly cancel by the i coming from the fact that C, which appears in a square root, is negative [its expansion starts with -1, see Eq. (890)]. As a consequence, we obtain

$$\Delta\varphi \equiv 2\left(\varphi_+ - \varphi_-\right) - 2\pi \simeq 6\pi\frac{GM}{L} + \Delta\varphi_\Lambda, \tag{892}$$

where the first term is the standard relativistic perihelion precession while [103]

$$\Delta\varphi_\Lambda \simeq \frac{\pi\Lambda_{\text{eff}} L^3}{GM}, \tag{893}$$

is the contribution due to the cosmological constant. The above equation is the main result of this sub-section. It gives the perihelion precession due to the presence of vacuum energy. Let us notice that the scaling $\Delta\varphi_{\Lambda} \propto L^3$ indicates that the precession is larger for planets that are far from the Sun. It is of course totally consistent with the fact that the effect of the cosmological constant is more important on large scales.

We now use the above result to constrain the amount of vacuum energy present in our universe. For this purpose, let us now calculate this perihelion shift for, say, Mercury (as mentioned before, the effect is larger for other planets but this does not modify the conclusions obtained below; estimates for Venus, Earth and Mars can be found in Ref.[105]). In this case, we have $r_+ \simeq 68.8 \times 10^6$km and $r_- \simeq 46 \times 10^6$km. This implies that $D \simeq 27.7 \times 10^6$km. We also have $GM \simeq 2.95$km. Finally, it is interesting to calculate in terms of the vacuum energy density observed in cosmology, see Eq. (844). For the cosmological constant, this leads to the following number $\Lambda_{\text{obs}} \simeq 8.81\, h^2 \times 10^{-48}\text{km}^{-2}$. This leads to the following expression

$$\Delta\varphi_{\Lambda} \simeq 1.6h^2 \times 10^{-24} \left(\frac{\Lambda}{\Lambda_{\text{obs}}}\right) \quad \text{rad/revolution}, \tag{894}$$

or, since Mercury completes 415 revolutions each century and there are $360 \times 60 \times 60/(2\pi)$ arc-seconds per radian,

$$\Delta\varphi_{\Lambda} \simeq 2.7h^2 \times 10^{-16} \left(\frac{\Lambda}{\Lambda_{\text{obs}}}\right) \quad \text{arc-second/century}, \tag{895}$$

which is, of course, completely unobservable. The reason for such a result has already been mentioned. The solar system has a characteristic scale (say a few astronomical units) which is too small for the cosmological constant to leave a sizable imprint. However, even if one cannot measure the vacuum energy density with this method, it can nevertheless be used to constrain its value. Indeed, the uncertainty in the precession of the perihelion of Mercury is about $0.1''$ par century, which means that

$$\Delta\varphi_{\Lambda} \simeq 2.7h^2 \times 10^{-16} \left(\frac{\rho_{\Lambda}}{\rho_{\text{cri}}}\right) < 0.1, \tag{896}$$

implying that

$$\rho_{\text{vac}} \lesssim 10^{15}\rho_{\text{cri}} \simeq 3 \times 10^{-32}\,\text{GeV}^4. \tag{897}$$

Of course, this limit is not competitive with what one has obtained in cosmology. However, when compared to Eq. (812), it leads to an interesting piece of information which allows us to define what the cosmological constant problem is more accurately. Indeed, we see that there is a contradiction between the theoretical expectation (812) and the observations of the planet trajectories in our solar system. In other words, the predictions of Eq. (812), if true, would lead to orbits drastically different from what we observe. Therefore, even without the cosmological observations and only from the motion of the planets in the solar system, one sees that our calculation of vacuum energy must be incorrect. Put it differently, it would be incorrect to say that the cosmological constant problem originates only from the recent astrophysical observations. In some sense, the problem is much worse. Only by observing the planets in the solar system, a class of observations which seems to be much more straightforward than the cosmological observations, we know that our regularization method of the vacuum energy must be flawed.

24.11.3 Modifications of the Atomic Levels

We have just seen that observing the motion of planets in our solar system can tell us something about the value of the vacuum energy. A priori, another possibility to observe the influence of the cosmological constant is through the modification of the atomic energy levels that its presence would cause. In this section, we investigate this question in the case of the Hydrogen atom since this is technically simple [106, 107, 108, 109, 110, 111]. We will follow the treatment of Ref. [106]. In particular, since the problem is still spherically symmetric, one can use the metric tensor (845) obtained before.

The motion of an electron around the nucleus, living in a static and spherically symmetric manifold, is described by the following Dirac equation

$$\left(\Gamma^{\mu} D_{\mu} + m\right) \Psi = 0, \tag{898}$$

where Ψ is a Dirac spinor, see Sec. 24.7.1 and m the mass of the electron. In the above equation, Γ_{μ} are the Dirac matrices satisfying the relation

$$\{\Gamma^{\mu}, \Gamma^{\nu}\} = 2g^{\mu\nu}. \tag{899}$$

Let us notice that we have already presented this formula, see Eq. (628). Here, we slightly change the notation and now denote the Dirac matrices in curved space-time by Γ^{μ} (notice that we have also included a factor i in the definition of Γ^{μ}). The matrices Γ^{μ} are related to the Dirac matrices in flat space-time, γ^{μ}, by $\Gamma_{\mu} = e^{\alpha}{}_{\mu}\gamma_{\alpha}$, where $e^{\alpha}{}_{\mu}$ are the vierbein fields such that $g_{\mu\nu} = e^{\alpha}{}_{\mu} e^{\beta}{}_{\nu} \eta_{\alpha\beta}$. The symbol D_{μ} denotes the covariant derivative acting on the spinor, namely

$$D_{\mu}\Psi = \left(\partial_{\mu} - \omega_{\mu}\right)\Psi, \tag{900}$$

where the spinorial connection can be expressed as

$$\omega_{\mu} \equiv -\frac{1}{4}\gamma_{\alpha}\gamma_{\beta} e^{\alpha}{}_{\nu} g^{\nu\lambda} \left(\nabla_{\mu} e^{\beta}{}_{\lambda}\right) + iqA_{\mu}, \tag{901}$$

with $\nabla_{\mu} e^{\beta}{}_{\lambda} \equiv \partial_{\mu} e^{\beta}{}_{\lambda} - \Gamma^{\sigma}_{\mu\lambda} e^{\beta}{}_{\sigma}$ and A_{μ} is the vector field describing the electromagnetic field acting on the electron (q is the charge). The Dirac equation can also be re-written as

$$i\partial_0 \Psi = H\Psi, \tag{902}$$

where the effective Hamiltonian is given by

$$\begin{aligned} H &= -i\left(g^{00}\right)^{-1} \Gamma^0 \Gamma^i \left(\partial_i \Psi - \omega_i \Psi\right) + i\Gamma_0 \Psi \\ &\quad -i\left(g^{00}\right)^{-1} \Gamma^0 m \Psi. \end{aligned} \tag{903}$$

In this case, the Dirac equation looks like an effective Schrödinger equation.

It is clear that the typical dimension of an atom is very small in comparison to the typical curvature scale of the manifold. Therefore, as was done for the calculation of the Green function in Sec. 24.8, one can perform a local analysis. For this purpose, we now use the Fermi normal coordinates [78]. In these coordinates, the metric tensor takes the form

$$g_{00} = -1 - R_{0\ell 0m} x^{\ell} x^{m} + \cdots, \tag{904}$$

$$g_{0i} = -\frac{2}{3} R_{0\ell im} x^{\ell} x^{m} + \cdots, \tag{905}$$

$$g_{ij} = \delta_{ij} - \frac{1}{3} R_{i\ell jm} x^{\ell} x^{m} + \cdots, \tag{906}$$

where the dots indicate higher order terms. This implies that the veirbein fields can be expressed as

$$e^{\alpha}{}_{0} = \delta^{\alpha}{}_{0} - \frac{1}{2} R^{\alpha}{}_{\ell 0m} x^{\ell} x^{m}, \tag{907}$$

$$e^{\alpha}{}_{i} = \delta^{\alpha}{}_{i} - \frac{1}{2} R^{\alpha}{}_{\ell im} x^{\ell} x^{m}, \tag{908}$$

and the Christoffel symbols can be written as

$$\Gamma^{0}_{ij} = \frac{1}{3}\left(R_{0ijm} + R_{0jim}\right) x^{m}, \tag{909}$$

$$\Gamma^{0}_{0i} = R_{0iom} x^{m}, \tag{910}$$

$$\Gamma^{i}_{jk} = \frac{1}{3}\left(R_{jikm} + R_{kijm}\right) x^{m}, \quad \Gamma^{0}_{00} = 0, \tag{911}$$

$$\Gamma^{i}_{0j} = R_{0mji} x^{m}, \quad \Gamma^{i}_{00} = R_{0i0m} x^{m}. \tag{912}$$

Using the above results, this finally leads to the following expression for the spinorial connection [106]

$$\begin{aligned}\omega_0 &= \frac{1}{2}\gamma_0\gamma_j R^j{}_{00m}x^m + \frac{1}{4}\gamma_k\gamma_j R^{kj}{}_{0m}x^m \\ &\quad +iqA_0, \qquad (913)\\ \omega_i &= \frac{1}{4}\gamma_0\gamma_j R^{0j}{}_{im}x^m + \frac{1}{8}\gamma_k\gamma_j R^{kj}{}_{im}x^m \\ &\quad +iqA_i. \qquad (914)\end{aligned}$$

The above formulas will allow us to calculate the effective Hamiltonian (903) explicitly. But before reaching this stage, we must also determine vector field A_μ. The Lagrangian of the corresponding gauge field is given by, see also Eq. (665)

$$S = -\int \mathrm{d}^4x\sqrt{-g}\left(\frac{1}{4}g^{\alpha\beta}g^{\mu\nu}F_{\mu\alpha}F_{\nu\beta} - j^\mu A_\mu\right), \qquad (915)$$

from which we deduce that the equations of motion read

$$g^{\mu\nu}\nabla_\mu\nabla_\nu A_\gamma - R^\sigma{}_\gamma A_\sigma = -j_\gamma. \qquad (916)$$

At first order in the curvature, these equations can be expressed as [106]

$$\delta^{ij}\partial_i\partial_j A_0 + \frac{1}{3}R_{i\ell jm}x^\ell x^m\partial^i\partial^j A_0 + \frac{5}{3}R_{i00m}x^m\partial^i A_0 + 2R^k{}_{i0m}x^m\partial^i A_k - \frac{2}{3}R_{im}x^m\partial^i A_0 = -j_0 \qquad (917)$$

$$\begin{aligned}&\delta^{ij}\partial_i\partial_j A_k + \frac{1}{3}R_{i\ell jm}x^\ell x^m\partial^i\partial^j A_k + \frac{2}{3}\delta^{ij}\left(R^\alpha{}_{kim} + R^\alpha{}_{ikm}\right)x^m\partial_j A_\alpha - \frac{2}{3}R^\sigma{}_k A_\sigma - \frac{1}{3}R^p{}_{00k}A_p \\ &-\frac{1}{3}R_{i00m}x^m\partial^i A_k - \frac{2}{3}R_{pm}x^m\partial^p A_k = j_k. \qquad (918)\end{aligned}$$

For a one electron atom with a nucleus of charge Ze, one can take

$$j^0 = -Ze\,\delta(\boldsymbol{r}), \quad j^k = 0. \qquad (919)$$

In what follows, we determine the vector perturbatively in the curvature. At zeroth order, in flat space-time, the equation for the time component reduces to

$$\delta^{ij}\partial_i\partial_j A_0 = Ze\,\delta(\boldsymbol{r}), \qquad (920)$$

whose solution is of course

$$A_0^{(0)} = -\frac{Ze}{4\pi}\frac{1}{r}. \qquad (921)$$

Then, we write $A_0 = A_0^{(0)} + \delta A_0$, where δA_0 is first order in the curvature. By inserting this expression into the full equation (917) and using the zeroth order solution (921), one arrives at [106]

$$\delta^{ij}\partial_i\partial_j\delta A_0 + \frac{1}{3}\frac{Ze}{4\pi r^3}\left(3R^0{}_{\ell 0m} - R_{\ell m}\right)x^\ell x^m = 0. \qquad (922)$$

The solution of the above equation reads [106]

$$\begin{aligned}\delta A_0 &= \frac{1}{12}\frac{Ze}{4\pi}\left(R + 4R_{00}\right)r \\ &\quad +\frac{1}{12}\frac{Ze}{4\pi}\left(3R^0{}_{\ell 0m} - R_{\ell m}\right)\frac{x^\ell x^m}{r}. \qquad (923)\end{aligned}$$

In the same manner, one can determine the space component. Obviously, at zeroth order, one has $A_k^{(0)} = 0$. The first order perturbation δA_k obeys the equation [106]

$$\delta^{ij}\partial_i\partial_j\delta A_k + \frac{2}{3}\frac{Ze}{4\pi}R^0{}_k\frac{1}{r} + \frac{2}{3}\frac{Ze}{4\pi}R^0{}_{ikm}\frac{x^i x^m}{r^3} = 0, \qquad (924)$$

whose solution can be written as [106]

$$\delta A_k = \frac{1}{2}\frac{Ze}{4\pi}R_{0k}r + \frac{1}{6}\frac{Ze}{4\pi}R^0{}_{lkm}\frac{x^\ell x^m}{r}. \tag{925}$$

This completes our calculation of the electromagnetic field.

Having determined the vierbeins fields (*i.e.* the quantities describing the gravitational force) and the gauge field (*i.e.* the quantity describing the electromagnetic force), we are now in a position where we can calculate the Hamiltonian (903). In order to make contact with the usual treatment of the Hydrogen atom, it is more convenient to introduce the standard Dirac matrices α_i and β related to the flat space-time Dirac matrices by

$$\gamma_0 = i\beta, \quad \gamma_i = -i\beta\alpha_i. \tag{926}$$

As is well-known, they satisfy

$$\alpha_i\alpha_j + \alpha_j\alpha_i = 2\delta_{ij}, \tag{927}$$

$$\alpha_i\beta + \beta\alpha_i = 0, \quad \beta^2 = 1. \tag{928}$$

Then, we use the expressions of the spinorial connections (913), of the gauge field (921), (925) and of the Dirac matrices in the equation (903) of the Hamiltonian. This leads to the following expression

$$\begin{aligned} H &= -i\alpha^i\partial_i + m\beta - \frac{\zeta}{r} - \frac{i}{2}R_{0\ell 0m}x^\ell x^m\alpha_i\partial^i - \frac{i}{6}R_{i\ell jm}x^\ell x^m\alpha^j\partial^i - \frac{i}{6}R_{0\ell jm}x^\ell x^m\alpha^j\alpha^i\partial_i - \frac{i}{2}R_{0\ell im}x^\ell x^m\partial^i \\ &+\frac{i}{8}\alpha^i\alpha^j\alpha^k R_{jkim}x^m + \frac{i}{4}\alpha^i\alpha^j R_{0jmi}x^m + \frac{i}{4}\alpha^i\alpha^j R_{0mij}x^m - \frac{i}{2}\alpha^j R_{0j0m}x^m + \frac{1}{6}\zeta R_{0\ell mi}x^\ell x^m\frac{\alpha^i}{r} + \frac{1}{2}\zeta R_{0i}\alpha^i r \\ &+\frac{1}{12}\zeta\left(R + 4R_{00}\right)r - \frac{1}{12}\zeta\left(R_{\ell m} + 3R_{0\ell 0m}\right)\frac{x^\ell x^m}{r} - \frac{m}{6}R_{i\ell 0m}x^\ell x^m\beta\alpha^i + \frac{m}{2}R_{0\ell 0m}x^\ell x^m\beta. \end{aligned} \tag{929}$$

In this expression we have defined $\zeta \equiv Ze^2/(4\pi)$ and we have taken $q = -e$. The three first terms are the usual Dirac Hamiltonian. The other terms represent the correction due to space-time curvature. Since the above expression is quite complicated, it is interesting to estimate the order of magnitude of each term. The curvature is given by D^{-2} where D is a typical distance. Each x^i is typically of the order of the Bohr radius, *i.e.* $x^i \simeq a_0 \simeq m^{-1}e^{-2} \simeq m^{-1}\zeta^{-1}$. As a consequence, ∂_i is of order $m\zeta$. This also means that the typical velocity is given by $m\zeta/m \simeq \zeta$. Of course the matrices α^i and β are dimensionless but, in the non relativistic limit $\alpha^i\partial_i$ must reduce to $\simeq mv^2$ which implies that $\alpha^i \simeq v \simeq \zeta$. In the following we consider that $\zeta = Z\alpha$ where α is the fine structure constant is a small number. Then, the dominant term in the above expression is the last one, which means that the Hamiltonian can be approximated as

$$\begin{aligned} H &= H_0 + H_{\rm int} \simeq \boldsymbol{\alpha}\cdot\boldsymbol{p} + m\beta - \frac{\zeta}{r} \\ &+\frac{1}{2}mR_{0\ell 0m}x^\ell x^m\beta, \end{aligned} \tag{930}$$

where we have used $p_i = -i\partial_i$. If we use the following representation for the Dirac matrices

$$\alpha^i = \begin{pmatrix} 0 & \sigma^i \\ \sigma^i & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} \mathbb{I}_2 & 0 \\ 0 & -\mathbb{I}_2, \end{pmatrix} \tag{931}$$

and write the Dirac spinor as

$$\Psi = \mathrm{e}^{iEx^0}\begin{pmatrix}\phi \\ \chi\end{pmatrix}, \tag{932}$$

then the Dirac equation reduces to two equations that can be expressed as

$$\boldsymbol{\alpha}\cdot\boldsymbol{p}\,\chi + \left(m - \frac{\zeta}{r} + \frac{1}{2}mR_{0\ell 0m}x^\ell x^m\right)\phi = E\phi, \tag{933}$$

$$\boldsymbol{\alpha}\cdot\boldsymbol{p}\,\phi - \left(m + \frac{\zeta}{r} + \frac{1}{2}mR_{0\ell 0m}x^\ell x^m\right)\chi = E\chi. \tag{934}$$

From the second relation we can express χ in terms of $\boldsymbol{\alpha}\cdot\boldsymbol{p}\,\phi$ and then we can insert this expression into the first equation. This leads to

$$\left(E+m+\frac{\zeta}{r}+\frac{1}{2}mR_{0\ell 0m}x^\ell x^m\right)^{-1}(\boldsymbol{\alpha}\cdot\boldsymbol{p})^2\,\phi+\left(m-\frac{\zeta}{r}+\frac{1}{2}mR_{0\ell 0m}x^\ell x^m\right)\phi \;=\; E\phi \tag{935}$$

or, using $\boldsymbol{\alpha}\cdot\boldsymbol{p}\,\phi=p^2\,\phi$ and expanding the denominator of the first term in the above equation

$$\left[1-\frac{1}{E+m}\frac{\zeta}{r}-\frac{m}{2(E+m)}R_{0\ell 0m}x^\ell x^m\right]\frac{p^2}{E+m}\phi-\left(\frac{\zeta}{r}-\frac{1}{2}mR_{0\ell 0m}x^\ell x^m\right)\phi \;=\; (E-m)\phi. \tag{936}$$

But, if we now perform an additional expansion assuming non-relativistic velocities for the electron (which, in practice, is a good approximation), one arrives at the following expression for the term $p^2/(E+m)$

$$\frac{p^2}{E+m}=\frac{p}{m+m+p^2/(2m)+\cdots}\simeq\frac{p^2}{2m}+\cdots \tag{937}$$

As a consequence, we conclude that the non relativistic limit is given by

$$H_{\rm NR}\phi=(E-m)\phi, \tag{938}$$

where the non-relativistic Hamiltonian can be written as [106]

$$\begin{aligned}H_{\rm NR} &= H_0+H_{\rm int}=\frac{p^2}{2m}-\frac{\zeta}{r}\\ &\quad+\frac{1}{2}mR_{0\ell 0m}x^\ell x^m\beta.\end{aligned} \tag{939}$$

This is our final result. Endowed with this equation, one can calculate, by means of perturbations theory, the modification of the ground state due to the presence of a cosmological constant. This leads to (the arguments of E indicates the value of the quantum numbers n and ℓ)

$$\Delta E(1,0) \;=\; \frac{1}{2}mR_{0\ell 0m}\left\langle 1s\left|x^\ell x^m\right|1s\right\rangle \tag{940}$$

$$\;=\; \frac{1}{2}mR_{0\ell 0m}\frac{1}{3}\delta^{\ell m}\left\langle 1s\left|r^2\right|1s\right\rangle \tag{941}$$

$$\;=\; \frac{1}{6}mR_{00}\left\langle 1s\left|r^2\right|1s\right\rangle \tag{942}$$

The ground state wave function is given by (at this order of perturbations theory, there is no need to take into account the modification on the wave-function; one can use the standard one)

$$|1s\rangle=\frac{1}{\sqrt{\pi a_0^3}}\mathrm{e}^{-r/a_0}, \tag{943}$$

where we recall that a_0 is the Bohr radius. This implies that

$$\left\langle 1s\left|r^2\right|1s\right\rangle=\frac{4\pi}{\pi a_0^3}\int_0^\infty r^4\mathrm{e}^{-r/a_0}\mathrm{d}r=\frac{4\pi}{\pi a_0^3}\frac{3a_0^5}{4}=3a_0^2. \tag{944}$$

The vacuum Einstein equations are, see Eq. (846), $R_{\mu\nu}=\Lambda_{\rm eff}g_{\mu\nu}$ and, therefore, at first order in the cosmological constant one $R_{00}=-\Lambda_{\rm eff}$ (here, of course, we only compute the corrections due to the cosmological constant and not those originating from the Earth gravitational field). This means that

$$\Delta E(1,0) \;=\; -\frac{1}{2}mc^2\Lambda_{\rm eff}a_0^2, \tag{945}$$

where we have re-established the speed of light for convenience. In fact, this result could have been anticipated from the Newtonian approach. Indeed, we have seen that the Newtonian potential, in presence of a cosmological constant, is given by $-\Lambda_{\text{eff}}mc^2r^2/6$, see Eqs. (856) and (857). Using the perturbation theory with this potential exactly leads Eq. (945). Let us now evaluate the correction. Using that the reduced mass of the atom is essentially the electron mass, $m \simeq 9.1 \times 10^{-31}\,\text{kg}$ and that the Bohr radius is $a_0 = 0.52 \times 10^{-10}\,\text{m}$, one obtains

$$\Delta E(1,0) = -6 \times 10^{-69} h^2 \left(\frac{\Lambda}{\Lambda_{\text{obs}}}\right) \text{eV}, \tag{946}$$

a quantity completely unobservable, probably for ever. For comparison the Lamb shift is of the order $10^{-6}\,\text{eV}$. At this stage this does not come as a surprise. As discussed in the case of planet orbits, the size of the system is so small that the influence of the cosmological constant is negligible. However, requiring that the shift due to the cosmological constant be smaller than the Lamb shift leads to

$$\rho_{\text{vac}} \lesssim 1.6 \times 10^{62} \rho_{\text{cri}} \simeq 1.3 \times 10^{16} \text{GeV}^4. \tag{947}$$

This result is less good than the one obtained from planet motion by about 50 orders of magnitude. Moreover, the theoretical expectation (812) satisfies this inequality. Let us notice in passing that the result based on the wrong regularization scheme, $\rho_{\text{vac}} \simeq M_{\text{Pl}}^4$ does not satisfy the previous bound.

We now have a clearer view of what the cosmological constant problem is. On one hand, one can calculate the vacuum energy density by means of, apparently, well known techniques of regularization. These techniques usually lead very robust results, fully consistent with the observations. For instance, the radiative corrections to the calculation of cross-sections have proven in very good agreement with various high energy physics experiments. One the other hand, various observational results (and not only in cosmology) indicate that the previous calculation is not correct. The theoretical expectation value for the vacuum energy density is so large that it should have already be seen by many experiences. Moreover, in cosmology, under a certain number of hypothesis, one even measures ρ_{vac} which is found to be many orders of magnitude smaller than the number derived from the theory. Since, as mentioned before, the theoretical framework seems to be robust (let us recall again that it involves calculations that can be found in the first pages of any textbook on quantum field theory) and since the observations seem to be convincing, we are facing a genuine mystery. However, before accepting this conclusion, it is worth checking that no loophole is present in the previous reasoning. This will be the goal of the next sections. A first question that can be asked is whether the zero-point fluctuations that are at the origin of the problem are a real physical phenomenon or just an artifact of the formalism of quantum field theory. We turn to this issue in the next section.

24.12 Do the Vacuum Fluctuations Really Exist?

We have seen in the previous sections that the zero-point fluctuations manifest themselves as the vacuum energy. Since this leads to conclusions that seem to be in contradiction with observations, it is legitimate to ask whether these fluctuations really exist in Nature. In particular, we want to investigate whether the zero-point fluctuations could manifest themselves in other physical phenomena. The answer to this question is usually positive and one experiment which is considered as a proof that the vacuum fluctuations are real is the measurement of the Lamb shift [112, 18]. In the next sub-section, we discuss this phenomenon.

24.12.1 The Lamb Shift

Let us consider the "motion" of an electron in an atom. Its "location" is described by the vector $\boldsymbol{r}$. The vector $\boldsymbol{r}$ is going to fluctuate because of the interaction between the electron and the zero-point fluctuations of the electromagnetic field. This interaction will slightly modify the position of the atomic levels leading to the Lamb shift. Let us study in details how this effect can be derived. Let $V(\boldsymbol{r})$ be the Coulomb potential which determines the properties of the atom. At the point $\boldsymbol{r} + \delta\boldsymbol{r}$, the potential can be Taylor expanded according to

$$V(\boldsymbol{r} + \delta\boldsymbol{r}) = V(\boldsymbol{r}) + \frac{\partial V}{\partial x_i}\delta x_i + \frac{1}{2}\frac{\partial^2 V}{\partial x_i \partial x_j}\delta x_i \delta x_j + \cdots. \tag{948}$$

Then, if we time average the previous expression, one obtains

$$\langle V(\boldsymbol{r}+\delta\boldsymbol{r})\rangle \simeq V(\boldsymbol{r}) + \frac{\partial V}{\partial x_i}\langle \delta x_i\rangle + \frac{1}{2}\frac{\partial^2 V}{\partial x_i \partial x_j}\langle \delta x_i \delta x_j\rangle + \cdots, \tag{949}$$

and, because the problem is spherically symmetric, we have $\langle \delta x_i\rangle = 0$ and $\langle \delta x_i \delta x_j\rangle = \langle \delta r^2\rangle \delta_{ij}/3$. The first term in Eq. (949) is the unperturbed Coulomb potential and the last one (since the second one vanishes) can be viewed as a small perturbation of the Hamiltonian, namely

$$\Delta H = \frac{1}{6}\frac{\partial^2 V}{\partial x_i \partial x_j}\left\langle \delta r^2\right\rangle \delta_{ij}. \tag{950}$$

In an atom with atomic number Z, the Coulomb potential is given by

$$V = -\frac{Z\alpha}{r}, \tag{951}$$

where $\alpha \simeq 1/137$ is the fine structure constant. This means that $\delta_{ij}\partial^2 V/(\partial x_i \partial x_j) = 4\pi Z\alpha\delta(\boldsymbol{r})$.

Let now us assume that the atom is placed in the quantum state $|\ell, m\rangle$. As we have already discussed in Sec. 24.11.3, at first order, the time-independent perturbations theory tells us that the correction to the energy of the level (ℓ, m) is given by [57]

$$\Delta E(n,\ell) = \langle \ell, m|\Delta H|\ell, m\rangle. \tag{952}$$

Using Eq. (950), one obtains the following expression for the energy displacement

$$\Delta E(n,\ell) = \frac{2\pi}{3} Z\alpha \left|\Psi_{n\ell}(0)\right|^2 \left\langle \delta r^2\right\rangle, \tag{953}$$

where $\Psi_{n\ell}$ is the wave-function in the corresponding quantum state. Since the wave-function at the origin can be expressed as [57]

$$|\Psi_{n\ell}(0)| = \frac{1}{\pi^{1/2}}\left(\frac{mZ\alpha}{n}\right)^{3/2}\delta_{\ell 0}, \tag{954}$$

where m is the electron mass ($m \simeq 0.511\,\mathrm{MeV}$) one finally arrives at [18]

$$\Delta E(n,\ell) = \frac{(2mZ\alpha)^3}{12}\frac{Z\alpha}{n^3}\left\langle \delta r^2\right\rangle \delta_{\ell 0}. \tag{955}$$

Therefore, we have reduced the problem to the calculation of the quantity $\langle \delta r^2\rangle$. In order to evaluate this quantity, we must now evaluate the displacement of the electron. The electron "moves" under the influence of the electric field within the atom and, as a consequence, δr_i obeys the following equation

$$m\frac{\mathrm{d}^2}{\mathrm{d}t^2}\delta r_i = eE_i, \tag{956}$$

where $E_\mu \equiv u^\nu F_{\mu\nu} = -\dot{A}_i$ is the electric field (or, rather, its vacuum fluctuating component). Using the usual expansion of the vector potential in terms of creation and annihilation operators, see Eq. (691), one can write

$$\begin{aligned} E_i(t,\boldsymbol{x}) &= \frac{-1}{(2\pi)^{3/2}}\int \frac{\mathrm{d}\boldsymbol{k}}{\sqrt{2\omega(k)}}\sum_{\alpha=1}^{2}\epsilon_i^\alpha(\boldsymbol{k}) \\ &\times \left[-i\omega a_{\boldsymbol{k}}^\alpha \mathrm{e}^{-i\omega t + i\boldsymbol{k}\cdot\boldsymbol{x}} + i\omega\left(a_{\boldsymbol{k}}^\alpha\right)^\dagger \mathrm{e}^{i\omega t - i\boldsymbol{k}\cdot\boldsymbol{x}}\right], \end{aligned} \tag{957}$$

In the same manner, one has to Fourier expand δr_i which is considered as an operator in this context. One obtains

$$\begin{aligned}\delta r_i\,(t,\boldsymbol{x}) &= \frac{1}{(2\pi)^{3/2}}\int \mathrm{d}\boldsymbol{k}\sum_{\alpha=1}^{2}\epsilon_i^{\alpha}(\boldsymbol{k})\Big[\delta r(\boldsymbol{k},t)a_{\boldsymbol{k}}^{\alpha}\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}}\\ &\quad +\delta r^{*}(\boldsymbol{k},t)\,(a_{\boldsymbol{k}}^{\alpha})^{\dagger}\,\mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\Big].\end{aligned} \tag{958}$$

The Fourier amplitude of the displacement operator can be calculated from the equation of motion (956). This leads to the following expression

$$\delta r(\boldsymbol{k},t)=\frac{e}{m}\frac{1}{\sqrt{2\omega^3}}\mathrm{e}^{-i\omega t}. \tag{959}$$

From this expression, one can now evaluate the quantity $\langle\delta r^2\rangle$. One finds

$$\begin{aligned}\left\langle\delta r^2\right\rangle &= \left\langle 0\left|\delta^{ij}\delta r_i\delta r_j\right|0\right\rangle\\ &= \frac{1}{(2\pi)^3}\int \mathrm{d}\boldsymbol{k}\sum_{\alpha=1}^{2}\delta^{ij}\epsilon_i^{\alpha}\epsilon_j^{\alpha}\delta r(\boldsymbol{k},t)\delta r^{*}(\boldsymbol{k},t)\end{aligned} \tag{960}$$

$$= \frac{2\alpha}{m^2\pi}\int_0^{\infty}\frac{\mathrm{d}\omega}{\omega}, \tag{961}$$

where we have used $\alpha=e^2/(4\pi)$. Therefore, inserting the above formula into Eq. (955), one obtains the following equation for the energy shift [18]

$$\Delta E(n,\ell)=\frac{4mZ^4\alpha^5}{3\pi n^3}\int_0^{\infty}\frac{\mathrm{d}\omega}{\omega}\delta_{\ell 0}. \tag{962}$$

This equation is the main result of this section. It gives the energy levels displacement due to the presence of the electromagnetic zero-point fluctuations.

As usual this expression is divergent and needs to be regularized. In a simplified treatment, one simply assumes, see Ref. [18], that the wavelengths in the above sum must be larger than the Compton wavelength of the electron, which implies $\omega<\omega_{\max}\simeq m$ and that $\omega>\omega_{\min}\simeq 1/a_0\simeq 1/[1/(\alpha m)]\simeq m\alpha$, where a_0 is the Bohr radius. In this case, the previous results reduces to

$$\Delta E(n,\ell)\simeq\frac{4mZ^4\alpha^5}{3\pi n^3}\ln\left(\frac{1}{\alpha}\right)\delta_{\ell 0}. \tag{963}$$

Let us now evaluate this quantity for the Hydrogen atom. Clearly, there is no effect for $n=1$. But for $n=2$ and $\ell=0$, one has

$$\Delta E(2,0)\simeq\frac{4m\alpha^5}{3\pi 8}\ln\left(\frac{1}{\alpha}\right)\delta_{\ell 0}\simeq 668\,\mathrm{MHz}, \tag{964}$$

where we have used $m\simeq 0.511\mathrm{MeV}=h\nu$ with $h=4.135\times 10^{-15}\mathrm{eV}\times\mathrm{s}$.

This shift has been experimentally observed and is usually taken as a proof that the zero point fluctuations are real since they lead to an observed physical phenomenon. Based on this result, it now seems difficult to argue that the zero-point fluctuations are just an artifact of the formalism of quantum field theory. Moreover, there exists another experiment, the Casimir effect, where the effects of the vacuum fluctuations can be observed. In the next section, we investigate this case.

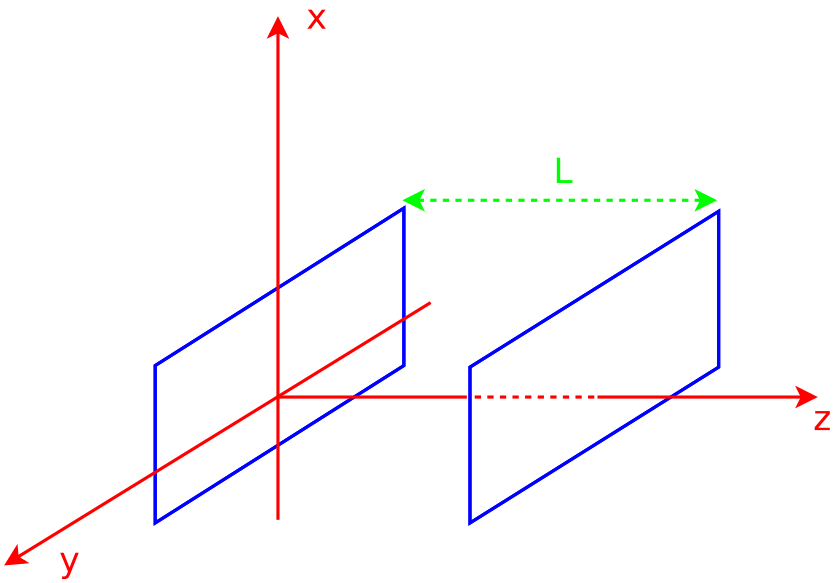


Figure 160: Sketch of a Casimir cavity of width L.

24.12.2 The Casimir Effect

In this section, we study the Casimir effect [113, 114, 115, 116, 117, 118, 119]. Since its discovery, this effect has been studied in great details and it is clear that, here, one cannot review all the literature. We will give a description of this phenomenon in a simplified context only and will refer to Refs. [113, 114, 115] for readers interested in learning more about this interesting subject.

The Casimir Force

Let us consider an experimental situation where we have two conducting plates in the (x, y) plane separated by a distance L long the z direction, see Fig. 160. For simplicity, instead of considering the electromagnetic field in the cavity (which is the case in the real world, especially when measurements are performed), we will just treat the case of a real massive scalar field $\Phi(t, \boldsymbol{x})$. This field obeys the Klein-Gordon equation already studied and solved before, see Eq. (356), except that now the boundary conditions are modified by the presence of the two plates. Let us write $\Phi(t, \boldsymbol{x}) = X(x)Y(y)Z(z)T(t)$. Then, inserting this anzatz into the Klein-Gordon equation, one obtains

$$-\frac{\ddot{T}}{T} + \frac{1}{X}\frac{\mathrm{d}^2X}{\mathrm{d}x^2} + \frac{1}{Y}\frac{\mathrm{d}^2Y}{\mathrm{d}y^2} + \frac{1}{Z}\frac{\mathrm{d}^2Z}{\mathrm{d}z^2} - m^2 = 0, \tag{965}$$

where m is the mass of the scalar particle. Clearly, one can separate the variables and, as a consequence, the above equation can be split into four differential equations, namely

$$-\frac{\ddot{T}}{T} - m^2 = C^2 \tag{966}$$

$$\frac{1}{X}\frac{\mathrm{d}^2X}{\mathrm{d}x^2} = k_x^2 \tag{967}$$

$$\frac{1}{Y}\frac{\mathrm{d}^2Y}{\mathrm{d}y^2} = k_y^2 \tag{968}$$

$$\frac{1}{Z}\frac{\mathrm{d}^2Z}{\mathrm{d}z^2} = k_z^2, \tag{969}$$

where the constants C, k_x, k_y and k_z are related by

$$C - k_x^2 - k_y^2 - k_z^2 = 0. \tag{970}$$

The solution in the x and y direction can be expressed in terms of standard plane waves, $X(x) = A_x \mathrm{e}^{ik_x x} + B_x \mathrm{e}^{-ik_x x}$ and $Y(y) = A_y \mathrm{e}^{ik_y y} + B_y \mathrm{e}^{-ik_y y}$. The quantities $A_{x,y}$ and $B_{x,y}$ are just integration constants. Of particular interest is of course the solution along the z-axis. It can also be written as

$$Z(z) = A_z \mathrm{e}^{ik_z z} + B_z \mathrm{e}^{-ik_z z}, \tag{971}$$

where A_z and B_z are two integration constants. But the difference with the case of a free field is that the boundary conditions "feel" the presence of the two plates. These boundary conditions are given by $Z(0) = 0$ and $Z(L) = 0$. The first one implies that $A_z = -B_z$ while the second one gives

$$2i \sin\left(k_z L\right) = 0, \tag{972}$$

or, equivalently,

$$k_z = n\frac{\pi}{L}, \tag{973}$$

where n is an integer. Finally, the equation for the function $T(t)$ can also be easily solved and the solution reads

$$T(t) = A_t \mathrm{e}^{i\omega t} + B_t \mathrm{e}^{-i\omega t}, \tag{974}$$

where $\omega^2 = C^2 + m^2$. Therefore, upon using Eq. (970), the frequency ω takes the form

$$\omega = \sqrt{k_x^2 + k_y^2 + n^2\frac{\pi^2}{L^2} + m^2}. \tag{975}$$

In this expression, k_x and k_y take continuous values.

Based on the previous considerations, one can now write the field operator as an expansion in terms of creation and annihilation operators. But this expansion encodes the fact that, along the z-axis, the wave-number is discrete. Concretely, we have

$$\Phi(t,\boldsymbol{x}) = \int \frac{\mathrm{d}k_x}{(2\pi)^{1/2}} \int \frac{\mathrm{d}k_y}{(2\pi)^{1/2}} \sum_{n=1}^{\infty} \sqrt{\frac{2}{L}} \frac{1}{\sqrt{2\omega}} \sin\left(n\frac{\pi}{L}z\right) \left(c_{k_x,k_y,n} \mathrm{e}^{-i\omega t + ik_x x + ik_y y} + c^{\dagger}_{k_x,k_y,n} \mathrm{e}^{i\omega t - ik_x x - ik_y y}\right). \tag{976}$$

The operators $c_{k_x,k_y,n}$ and $c^{\dagger}_{k_x,k_y,n}$ obey the usual commutation relations for bosons. It is interesting to check that the above expansion is indeed consistent with the commutation relations between the field and its conjugate momentum. For this purpose, let us compute the following quantity

$$\begin{aligned}\int \mathrm{d}x \int \mathrm{d}y \int_0^L \mathrm{d}z\, \Phi\, \mathrm{e}^{-ip_x x - ip_y y} \sin\left(m\frac{\pi}{L}z\right) &= \int \frac{\mathrm{d}k_x}{(2\pi)^{1/2}} \int \frac{\mathrm{d}k_y}{(2\pi)^{1/2}} \sum_{n=1}^{\infty} \int \mathrm{d}x \int \mathrm{d}y \sqrt{\frac{2}{L}} \frac{1}{\sqrt{2\omega}} \Big(c_{k_x,k_y,n} \mathrm{e}^{-i\omega t + ik_x x + ik_y y} \\ &\quad + c^{\dagger}_{k_x,k_y,n} \mathrm{e}^{i\omega t - ik_x x - ik_y y}\Big) \mathrm{e}^{-ip_x x - p_y y} \int_0^L \mathrm{d}z \sin\left(m\frac{\pi}{L}z\right) \sin\left(n\frac{\pi}{L}z\right). \end{aligned} \tag{977}$$

Using the fact that $\int_0^L \mathrm{d}z \sin(n\pi z/L)\sin(m\pi z/L) = (L/2)\delta_{mn}$ (as can be checked directly) and $\int \mathrm{d}x \mathrm{e}^{i(k_x - p_x)x} = (2\pi)\delta(k_x - p_x)$ one arrives at

$$\int \mathrm{d}x \int \mathrm{d}y \int_0^L \mathrm{d}z\, \Phi(t,\boldsymbol{x})\, \mathrm{e}^{-ip_x x - ip_y y} \sin\left(m\frac{\pi}{L}z\right) = \frac{2\pi}{2}\sqrt{\frac{L}{\omega}} \Big(c_{p_x,p_y,m} \mathrm{e}^{-i\omega t} + c^{\dagger}_{-p_x,-p_y,m} \mathrm{e}^{i\omega t}\Big). \tag{978}$$

In the same manner, one can also evaluate the following integral involving the conjugate momentum $\Pi(t,\boldsymbol{x}) = \dot{\Phi}(t,\boldsymbol{x})$ (a dot means a derivative with respect to time)

$$\int \mathrm{d}x \int \mathrm{d}y \int_0^L \mathrm{d}z\, \Pi(t,\boldsymbol{x})\, \mathrm{e}^{-ip_x x - ip_y y} \sin\left(m\frac{\pi}{L}z\right) = \frac{2\pi}{2} i\omega \sqrt{\frac{L}{\omega}} \Big(-c_{p_x,p_y,m} \mathrm{e}^{-i\omega t} + c^{\dagger}_{-p_x,-p_y,m} \mathrm{e}^{i\omega t}\Big) \tag{979}$$

From the two above expression, one can deduce a formula expressing the creation and the annihilation operators in terms of the field operator and its conjugate momentum. One obtains

$$c_{p_x,p_y,m} = \sqrt{\frac{\omega}{L}}\frac{\mathrm{e}^{i\omega t}}{2\pi}\left[\int \mathrm{d}x\mathrm{d}y \int_0^L \mathrm{d}z\, \Phi\, \mathrm{e}^{-ip_x x - ip_y y}\sin\left(m\frac{\pi}{L}z\right) + \frac{i}{\omega}\int \mathrm{d}x\mathrm{d}y \int_0^L \mathrm{d}z\, \Pi\, \mathrm{e}^{-ip_x x - ip_y y}\sin\left(m\frac{\pi}{L}z\right)\right],$$

$$c^\dagger_{p_x,p_y,m} = \sqrt{\frac{\omega}{L}}\frac{\mathrm{e}^{-i\omega t}}{2\pi}\left[\int \mathrm{d}x\mathrm{d}y \int_0^L \mathrm{d}z\, \Phi\, \mathrm{e}^{ip_x x + ip_y y}\sin\left(m\frac{\pi}{L}z\right) - \frac{i}{\omega}\int \mathrm{d}x\mathrm{d}y \int_0^L \mathrm{d}z\, \Pi\, \mathrm{e}^{ip_x x + ip_y y}\sin\left(m\frac{\pi}{L}z\right)\right].$$

We are now in a position where we can compute the commutator of a creation and an annihilation operators. It can be written as

$$\begin{aligned}\left[c_{p_x,p_y,n}, c^\dagger_{q_x,q_y,m}\right] = \frac{\mathrm{e}^{i(\omega_p-\omega_q)t}}{(2\pi)^2}\frac{\sqrt{\omega_p\omega_q}}{L}\int \mathrm{d}x\mathrm{d}y \int_0^L \mathrm{d}z \int \mathrm{d}\overline{x}\mathrm{d}\overline{y} \int_0^L \mathrm{d}\overline{z}\Bigg[&\Phi(t,\boldsymbol{x})\,\mathrm{e}^{-ip_x x - ip_y y}\sin\left(n\frac{\pi}{L}z\right)\\ &+\frac{i}{\omega_p}\Pi(t,\boldsymbol{x})\,\mathrm{e}^{-ip_x x - ip_y y}\sin\left(n\frac{\pi}{L}z\right), \Phi(t,\overline{\boldsymbol{x}})\,\mathrm{e}^{iq_x\overline{x}+iq_y\overline{y}}\sin\left(m\frac{\pi}{L}\overline{z}\right)\\ &-\frac{i}{\omega_q}\Pi(t,\overline{\boldsymbol{x}})\,\mathrm{e}^{iq_x\overline{x}+iq_y\overline{y}}\sin\left(m\frac{\pi}{L}\overline{z}\right)\Bigg].\end{aligned} \tag{980}$$

Then, we make use of the canonical relation $[\Phi(t,\boldsymbol{x}),\Pi(t,\boldsymbol{y})] = i\delta^3(\boldsymbol{x}-\boldsymbol{y})$ in the previous formula. This leads to the following complicated expression

$$\begin{aligned}\left[c_{p_x,p_y,n}, c^\dagger_{q_x,q_y,m}\right] &= \frac{\mathrm{e}^{i(\omega_p-\omega_q)t}}{(2\pi)^2}\frac{\sqrt{\omega_p\omega_q}}{L}\int \mathrm{d}x\mathrm{d}y \int_0^L \mathrm{d}z \int \mathrm{d}\overline{x}\mathrm{d}\overline{y} \int_0^L \mathrm{d}\overline{z}\Bigg\{-\frac{i}{\omega_q}\mathrm{e}^{-ip_x x - ip_y y + iq_x\overline{x}+iq_y\overline{y}}\sin\left(n\frac{\pi}{L}z\right)\\ &\times\sin\left(m\frac{\pi}{L}\overline{z}\right)[\Phi(t,\boldsymbol{x}),\Pi(t,\bar{\boldsymbol{x}})] + \frac{i}{\omega_p}\mathrm{e}^{-ip_x x - ip_y y + iq_x\overline{x}+iq_y\overline{y}}\sin\left(n\frac{\pi}{L}z\right)\sin\left(m\frac{\pi}{L}\overline{z}\right)\\ &\times[\Pi(t,\boldsymbol{x}),\Phi(t,\overline{\boldsymbol{x}})]\Bigg\}\end{aligned} \tag{981}$$

$$\begin{aligned}&= \frac{\mathrm{e}^{i(\omega_p-\omega_q)t}}{(2\pi)^2}\frac{\sqrt{\omega_p\omega_q}}{L}\left(\frac{1}{\omega_q}+\frac{1}{\omega_p}\right)\int \mathrm{d}x\mathrm{d}y \int_0^L \mathrm{d}z\mathrm{e}^{-i(p_x-q_x)x - i(p_y-q_y)y}\sin\left(n\frac{\pi}{L}z\right)\sin\left(m\frac{\pi}{L}z\right)\\ &= \frac{\omega}{L}\frac{2}{\omega}\delta(p_x-q_x)\delta(p_y-q_y)\frac{L}{2}\delta_{mn} = \delta(p_x-q_x)\delta(p_y-q_y)\delta_{mn}.\end{aligned} \tag{982}$$

We have thus reached our goal, namely we have shown that the canonical commutation relation between the field and its conjugate momentum implies the usual commutation relation between the creation and the annihilation operator.

We are now in a position where we can evaluate the energy density of the field in the Casimir cavity. It is of course given by the same expression as in the free case, see Eq. (363). This means that one should first calculate the three terms that participate in the expression of T_{00}, see Eqs. (360), (361) and (362). Let us start with $\langle 0|\dot{\Phi}^2|0\rangle$. Using the Fourier expansion of the field and the commutation relations established above, one obtains

$$\begin{aligned}\langle 0|\dot{\Phi}^2|0\rangle &= \int\frac{\mathrm{d}k_x}{2\pi}\int\frac{\mathrm{d}k_y}{2\pi}\sum_{n=1}^{\infty}\frac{2}{L}\frac{\omega^2}{2\omega}\\ &\times\sin^2\left(n\frac{\pi}{L}z\right),\end{aligned} \tag{983}$$

and, if we integrate over the cavity volume

$$\int_{\text{Cavity}} \mathrm{d}\boldsymbol{x}\,\langle 0|\dot{\Phi}^2|0\rangle = \frac{D^2}{2}\int\frac{\mathrm{d}k_x}{2\pi}\int\frac{\mathrm{d}k_y}{2\pi}\sum_{n=1}^{\infty}\omega, \tag{984}$$

where D^2 is the area of the plates (of course, the two parallel plates have the same surface). In the same manner, one finds

$$\int\limits_{\text{Cavity}} \mathrm{d}\boldsymbol{x}\, \langle 0|\delta^{ij}\partial_i\Phi\partial_j\Phi|0\rangle = \frac{D^2}{2}\int\frac{\mathrm{d}k_x}{2\pi}\int\frac{\mathrm{d}k_y}{2\pi} \times \sum_{n=1}^{\infty}\frac{1}{2\omega}\left(k_x^2+k_y^2+n^2\frac{\pi^2}{L^2}\right). \tag{985}$$

Finally, the last relevant term reads

$$\int\limits_{\text{Cavity}} \mathrm{d}\boldsymbol{x}\, \langle 0|\Phi^2|0\rangle = \frac{D^2}{2}\int\frac{\mathrm{d}k_x}{2\pi}\int\frac{\mathrm{d}k_y}{2\pi}\sum_{n=1}^{\infty}\frac{1}{2\omega}. \tag{986}$$

Putting everything together and using the definition (363), the expression of the energy of the field within the empty cavity takes the form

$$\int\limits_{\text{Cavity}} \mathrm{d}\boldsymbol{x}\, \langle 0|T_{00}|0\rangle \tag{987}$$

$$= \left(\frac{D}{2\pi}\right)^2\int\mathrm{d}k_x\int\mathrm{d}k_y\sum_{n=1}^{\infty}\frac{\omega}{2} \tag{988}$$

$$= \left(\frac{D}{2\pi}\right)^2\int\mathrm{d}k_x\int\mathrm{d}k_y\sum_{n=1}^{\infty}\frac{1}{2}\sqrt{k_x^2+k_y^2+n^2\frac{\pi^2}{L^2}+m^2}. \tag{989}$$

Of course, this expression has exactly the expected form. Roughly speaking it can be written as "$\sum\frac{1}{2}\omega$", see Eq. (975), the only difference being that, along the z-axis, the sum takes into account the fact that the momentum is discrete. Another common point with our previous considerations is that the above number is actually divergent and must be regularized. In the following, we assume $m=0$ for simplicity and use dimensional regularization. Therefore, the expression of the energy can now be written as

$$E = \left(\frac{D}{2\pi}\right)^d\int\mathrm{d}^d\boldsymbol{k}_\perp\sum_{n=1}^{\infty}\frac{1}{2}\sqrt{k_\perp^2+n^2\frac{\pi^2}{L^2}} \tag{990}$$

$$= \left(\frac{D}{2\pi}\right)^d\int\mathrm{d}k_\perp k_\perp^{d-1}\mathrm{d}^{d-1}\Omega\sum_{n=1}^{\infty}\frac{1}{2}\sqrt{k_\perp^2+n^2\frac{\pi^2}{L^2}} = \left(\frac{D}{2\pi}\right)^d\frac{2\pi^{d/2}}{\Gamma(d/2)}\int\mathrm{d}k_\perp k_\perp^{d-1}\sum_{n=1}^{\infty}\frac{1}{2}\sqrt{k_\perp^2+n^2\frac{\pi^2}{L^2}} \tag{991}$$

where d is the "dimension" of the plates (*i.e.* $d=2$ in the real world) and $k_\perp$ denotes the wave-vector living in the sub-manifold defined by the plates. Then using the change of variables $y=Lk_\perp/(n\pi)$ and the following definition [48, 49] of the Euler's integral of first kind

$$B(x,y) = 2\int_0^{\infty}\frac{t^{2x-1}}{(1+t^2)^{x+y}}\mathrm{d}t = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}, \tag{992}$$

the energy can be re-expressed as

$$E = \left(\frac{D}{2\pi}\right)^d\frac{\pi^{d/2}}{2}\left(\frac{\pi}{L}\right)^{d+1}\frac{\Gamma(-d/2-1/2)}{\Gamma(-1/2)}\sum_{n=1}^{\infty}n^{d+1}. \tag{993}$$

The above formula can be expressed in term of the Riemann zeta function defined by the following expression

$$\zeta(s) \equiv \sum_{n=1}^{\infty} n^{-s}. \tag{994}$$

Moreover, upon using the two following equations

$$\zeta(-1-d) = 2^{-1-d}\pi^{-2-d}\sin\left[\frac{\pi}{2}(-1-d)\right] \times\Gamma(2+d)\zeta(2+d), \tag{995}$$

$$\Gamma\left(-\frac{d}{2}-\frac{1}{2}\right)\sin\left[\frac{\pi}{2}(-1-d)\right] = \frac{\pi}{\Gamma(d/2+3/2)}, \tag{996}$$

one can re-write the energy as

$$\frac{E}{D^d} = -\frac{1}{L^{d+1}}\frac{\Gamma(1+d/2)}{2^{d+2}\pi^{d/2+1}}\zeta(2+d). \tag{997}$$

The Casimir force arises because of a shift in the vacuum energy when the plates are present compared to the situation where they are absent. So the relevant quantity is in fact $\Delta E/D^d \equiv (E-E_0)/D^d$ where E_0/D^d is the above quantity in the limit $L \to +\infty$. As a consequence, for $d=2$, one arrives at

$$\frac{\Delta E}{D^2} = -\frac{\pi^2}{1440L^3}, \tag{998}$$

where we have used $\zeta(4) = \pi^4/90$. Therefore, the corresponding force is given by [113, 114, 115, 116]

$$F = -\frac{\partial}{\partial L}\left(\frac{\Delta E}{D^2}\right) = -\frac{\pi^2}{480L^4}, \tag{999}$$

which is one half the result obtained in the case of the electromagnetic field since we have only one state of polarization instead of two. We notice that the force is attractive (this could change in the case of more complicated geometries).

It is also interesting to check that the final result is independent from the regularization scheme used to obtain it. In order to test this idea in the simplest situation (*i.e.* where the calculations are easy), let us consider again the energy of a Casimir cavity but, this time, with $d=0$. In this case, our zeta-function regularized result reads

$$\frac{E}{D^0} = -\frac{1}{L}\frac{\Gamma(1)\zeta(2)}{4\pi} = -\frac{\pi}{24L}, \tag{1000}$$

since $\Gamma(1)=1$ and $\zeta(2)=\pi^2/6$. One the other hand, we can return to the expression of E in Eq. (993). In the case $d=0$, the energy density $\rho(L) = E/L$ reads

$$\rho(L) = \frac{\pi}{2L^2}\sum_{n=1}^{\infty} n. \tag{1001}$$

In order to regularize this divergent quantity, we now proceed differently and introduce an exponential cut-off such that (previously, we argued that introducing a cut-off was not a good method because this breaks Lorentz invariance. Here, of course, this objection is no longer valid as the presence of the two plates along the z-axis obviously breaks translation invariance; as a consequence working with a cut-off seems legitimate in the present context)

$$\rho(L,\alpha) = \frac{\pi}{2L^2}\sum_{n=1}^{\infty} n\,\mathrm{e}^{-n\alpha/L}. \tag{1002}$$

The sum is easy to perform and one obtains

$$\rho(L,\alpha) = \frac{\pi}{8L^2}\sinh^{-2}\left(\frac{\alpha}{2L}\right) \quad (1003)$$

$$\simeq \frac{\pi}{2\alpha^2} - \frac{\pi}{24L^2} + \cdots \quad (1004)$$

Now, the regularized Casimir energy is the shift energy in the vacuum after having sent the cut-off to zero; in other words

$$E = L\lim_{\alpha\to 0}\left[\rho(L,\alpha) - \lim_{L\to\infty}\rho(L,\alpha)\right] \quad (1005)$$

$$= -\frac{\pi}{24L}, \quad (1006)$$

that is to say exactly the same expression obtained with another regularization scheme. It is reassuring to notice that the final result is independent of the regularization method used.

The Casimir Stress-Energy Tensor

In this section, we aim at calculating the stress energy tensor of a scalar field in a Casimir cavity [119, 113, 114, 115, 116]. Our goal is of course to compare this tensor with the vacuum stress-energy tensor $T_{\mu\nu} = -\rho_{\text{vac}} g_{\mu\nu}$. Here, we use a method based on the calculations of the Green function. This is yet another method of calculation of the Casimir effect, different from the two ones exposed before and, therefore, it will be interesting to compare this approach with the treatments used previously.

Let us start with the calculation of the Green function. It obeys the following equation

$$-\eta^{\mu\nu}\partial_\mu\partial_\nu G(x,\overline{x}) = \delta^{(4)}(x^\mu - \overline{x}^\mu). \quad (1007)$$

It is interesting to compare the above formula with Eq. (778). It is clear that this is the same equation but with $g_{\mu\nu} = \eta_{\mu\nu}$ (flat space-time) and $m = 0$ (mass-less field). In order to solve this equation, it is convenient to Fourier transform $G(x,\overline{x})$ in frequency and transverse momentum. Explicitly, one writes

$$G(x,\overline{x}) = \int \frac{\mathrm{d}\omega}{2\pi}\frac{\mathrm{d}\boldsymbol{k}_\perp}{(2\pi)^2}\mathrm{e}^{-i\omega\left(t-\overline{t}\right)+ik_x(x-\overline{x})+ik_y(y-\overline{y})} \times g(z,\overline{z};\omega,k_x,k_y). \quad (1008)$$

Inserting this expression into the Green equation (1007), one obtains

$$-\left(\frac{\mathrm{d}^2}{\mathrm{d}z^2} + \lambda^2\right) g(z,\overline{z}) = \delta(z-\overline{z}), \quad (1009)$$

where we have defined

$$\lambda^2 \equiv \omega^2 - k_\perp^2 = \omega^2 - k_x^2 - k_y^2. \quad (1010)$$

Our goal is now to determine explicitly the Green function. Let us start with the solution within the plates, *i.e.* $0 < z, \overline{z} < L$. When $z \neq \overline{z}$, the solution can be written

$$g(z,\overline{z}) = A(\overline{z})\cos(\lambda z) + B(\overline{z})\sin(\lambda z), \quad (1011)$$

where $A(\overline{z})$ and $B(\overline{z})$ are two unknown function of $\overline{z}$. We must now take into account the boundary conditions. If $z < \overline{z}$, the condition $g(z=0,\overline{z}) = 0$ implies that $A = 0$ and therefore $g = B\sin(\lambda z)$ (for $z < \overline{z}$). On the other hand, if $z > \overline{z}$, the condition $g(z=L,\overline{z}) = 0$ means that

$$g = A\cos(\lambda z) - A\frac{\cos(\lambda L)}{\sin(\lambda L)}\sin(\lambda z), \quad z > \overline{z}. \quad (1012)$$

We have two unknown quantities and, therefore, we need two equations. One is provided by writing the continuity of the Green function when $z = \overline{z}$. The other is obtained by integrating Eq. (1009) from $\overline{z} - \epsilon$ to $\overline{z} + \epsilon$. This gives

$$-\int_{\overline{z}-\epsilon}^{\overline{z}+\epsilon} \frac{\mathrm{d}^2 g}{\mathrm{d}z^2}\mathrm{d}z - \lambda^2 \int_{\overline{z}-\epsilon}^{\overline{z}+\epsilon} g\mathrm{d}z = \int_{\overline{z}-\epsilon}^{\overline{z}+\epsilon} \delta(z-\overline{z})\mathrm{d}z, \tag{1013}$$

which leads to

$$-\left[\frac{\mathrm{d}g}{\mathrm{d}z}(\overline{z}+\epsilon) - \frac{\mathrm{d}g}{\mathrm{d}z}(\overline{z}-\epsilon)\right] = 1. \tag{1014}$$

Therefore, taking the limit $\epsilon \to 0$, the solutions of these two equations (*i.e.* the one expressing the continuity of the Green function and the one we have just derived) read

$$A(\overline{z}) = \frac{1}{\lambda}\sin(\lambda\overline{z}), \tag{1015}$$

$$B(\overline{z}) = \frac{1}{\lambda}\cos(\lambda\overline{z}) - \frac{\cos(\lambda L)}{\sin(\lambda L)}A(\overline{z}). \tag{1016}$$

As a consequence, the Green function between the plates can be expressed as [116]

$$g(z,\overline{z}) = \begin{cases} \dfrac{\sin(\lambda z)\sin(\lambda L - \lambda\overline{z})}{\lambda\sin(\lambda L)}, & z < \overline{z}, \\ \dfrac{\sin(\lambda\overline{z})\sin(\lambda L - \lambda z)}{\lambda\sin(\lambda L)}, & z > \overline{z}. \end{cases} \tag{1017}$$

The same method can be used to determine the Green function outside the cavity. Let us apply it for z and $\overline{z}$ larger than L, *i.e.* on the right hand side of the cavity. In this case, it is more convenient to re-write the general solution (1011) as

$$g(z,\overline{z}) = C(\overline{z})\mathrm{e}^{i\lambda z} + D(\overline{z})\mathrm{e}^{-i\lambda z}. \tag{1018}$$

When $z > \overline{z}$, we require only one branch to be present which amounts to take $D = 0$. For $z < \overline{z}$, the boundary condition is still $g(L,\overline{z}) = 0$. Then, straightforward manipulations lead to [116]

$$g(z,\overline{z}) = \begin{cases} \dfrac{1}{\lambda}\mathrm{e}^{i\lambda(\overline{z}-L)}\sin(\lambda z - \lambda L), & z < \overline{z}, \\ \dfrac{1}{\lambda}\mathrm{e}^{i\lambda(z-L)}\sin(\lambda\overline{z} - \lambda L), & z > \overline{z}, \end{cases} \tag{1019}$$

which completes our determination of the Green function. The calculation of the Green function outside the cavity, but on the left hand side, proceeds exactly in the same way. The corresponding expression is obtained from the one above with $L = 0$, reflecting the fact that the boundary condition is now $g(0,\overline{z}) = 0$.

Endowed with the Green function of the problem, one can now turn to our main goal, namely the calculation of the stress-energy tensor. We have already seen that, see also Eq. (397),

$$G(\boldsymbol{r},\overline{\boldsymbol{r}}) = i\left\langle 0\left|T\left[\Phi(\boldsymbol{r})\Phi(\overline{\boldsymbol{r}})\right]\right|0\right\rangle. \tag{1020}$$

But we know that the T_{00} component of the stress energy tensor can be expressed as, see Eq. (303),

$$T_{00} = \frac{1}{2}\left[(\partial_0\Phi)^2 + (\partial_x\Phi)^2 + (\partial_y\Phi)^2 + (\partial_z\Phi)^2\right]. \tag{1021}$$

Therefore, if we apply the following operator to the Green function in Eq. (1020)

$$\frac{1}{2}\left[\partial_0\partial_{\overline{0}} + \partial_x\partial_{\overline{x}} + \partial_y\partial_{\overline{y}} + \partial_z\partial_{\overline{z}}\right], \tag{1022}$$

where $\partial_{\overline{\mu}}$ means a derivative with respect to $\overline{x}^{\mu}$, to $G(\boldsymbol{r},\overline{\boldsymbol{r}})$ and, then, take the space-time points to be the same $\boldsymbol{r}=\overline{\boldsymbol{r}}$, one should obtain an expression for $\langle T_{00}\rangle$. Using Eqs. (1008) and (1010), the result reads

$$\begin{aligned}\langle T_{00}\rangle &= \frac{1}{2i}\int\frac{\mathrm{d}\omega}{2\pi}\frac{\mathrm{d}k_x\mathrm{d}k_y}{(2\pi)^2}\left(\omega^2+k_x^2+k_y^2+\partial_z\partial_{\overline{z}}\right)\\ &\quad\times g(z,\overline{z})\Big|_{z=\overline{z}}.\end{aligned}\tag{1023}$$

Then, we use the expression of the Green function inside the plates (1017) and one arrives at

$$\begin{aligned}\langle T_{00}\rangle &= \frac{1}{2i}\int\frac{\mathrm{d}\omega}{2\pi}\frac{\mathrm{d}k_x\mathrm{d}k_y}{(2\pi)^2}\frac{1}{\lambda\sin(\lambda L)}\Big[(\omega^2+k_x^2+k_y^2)\sin(\lambda z)\sin(\lambda L-\lambda L)-\lambda^2\cos(\lambda z)\cos(\lambda L-\lambda L)\Big] &(1024)\\ &= -\frac{1}{2i}\int\frac{\mathrm{d}\omega}{2\pi}\frac{\mathrm{d}k_x\mathrm{d}k_y}{(2\pi)^2}\frac{1}{\lambda\sin(\lambda L)}\left[\omega^2\cos(\lambda L)-k^2\cos(2\lambda z-\lambda L)\right]. &(1025)\end{aligned}$$

In order to evaluate this integral, we perform the two Wick rotations $\omega\to i\zeta$ and $\lambda\to i\kappa$. This leads to

$$\langle T_{00}\rangle=-\frac{1}{2}\int_0^{\infty}\frac{\mathrm{d}\zeta}{2\pi}\int\frac{\mathrm{d}k_x\mathrm{d}k_y}{(2\pi)^2}\frac{1}{\kappa\sinh(\kappa L)}\left[\zeta^2\cosh(\kappa L)+k^2\cosh(2\kappa z-\kappa L)\right].\tag{1026}$$

Then, we introduce polar coordinates in the plane (ζ,k), *i.e.* $\zeta=\kappa\cos\theta$ and $k=\kappa\sin\theta$. This gives

$$\begin{aligned}\langle T_{00}\rangle &= -\frac{1}{4\pi^2}\int_0^{\infty}\mathrm{d}\kappa\kappa\int_0^{\pi/2}\mathrm{d}\theta\kappa^2\frac{\sin\theta}{\sinh(\kappa L)}\left[\cos^2\theta\cosh(\kappa L)+\sin^2\theta\cosh(2\kappa z-\kappa L)\right] &(1027)\\ &= -\frac{1}{12\pi^2}\int_0^{\infty}\mathrm{d}\kappa\kappa^3\left[\frac{\cosh(\kappa L)}{\sinh(\kappa L)}+2\frac{\cosh(2\kappa z-\kappa L)}{\sinh(\kappa L)}\right]. &(1028)\end{aligned}$$

Finally, writing that

$$\frac{\cosh(\kappa L)}{\sinh(\kappa L)}=1+\frac{2}{\mathrm{e}^{2\kappa L}-1},\tag{1029}$$

the expression of $\langle T_{00}\rangle$ can be re-expressed as

$$\begin{aligned}\langle T_{00}\rangle &= -\frac{1}{6\pi^2}\int_0^{\infty}\mathrm{d}\kappa\kappa^3\left[\frac{1}{\mathrm{e}^{2\kappa L}-1}+\frac{1}{2}\right.\\ &\quad\left.+\frac{\mathrm{e}^{2\kappa z}+\mathrm{e}^{2\kappa(L-z)}}{\mathrm{e}^{2\kappa L}-1}\right].\end{aligned}\tag{1030}$$

The last step of the calculation consists in evaluating the last term [we call it $g(z)$ in what follows; not to be confused with the Green function] of the above equation. One has

$$\begin{aligned}g(z) &= -\frac{1}{6\pi^2}\frac{1}{16L^4}\int_0^{\infty}\mathrm{d}y\,y^3\frac{\mathrm{e}^{y/L}+\mathrm{e}^{y(1-z/L)}}{\mathrm{e}^{y}-1} &(1031)\\ &= -\frac{1}{6\pi^2}\frac{1}{16L^4}\sum_{n=1}^{\infty}\int_0^{\infty}\mathrm{d}y\,y^3\left[\mathrm{e}^{y(z/L-n)}\right.\\ &\quad\left.+\mathrm{e}^{y(1-n-z/L)}\right], &(1032)\end{aligned}$$

where we have expanded the denominator in a geometric series. The two above integrals can be expressed in terms of the Hurwitz zeta function defined by [48, 49]

$$\zeta(s,a) \equiv \sum_{m=0}^{\infty} \frac{1}{(m+a)^s}, \tag{1033}$$

and this leads to

$$g(z) = -\frac{1}{16\pi^2 L^4}\left[\zeta\left(4,\frac{z}{L}\right)+\zeta\left(4,1-\frac{z}{L}\right)\right]. \tag{1034}$$

If we use the fact that $\int_0^{+\infty} \kappa^3/(\mathrm{e}^{2\kappa L}-1)\mathrm{d}\kappa = \pi^4/(240L^4)$, one arrives at our final expression for the time-time component of the stress energy tensor, namely

$$\langle T_{00}\rangle = -\frac{\pi^2}{1440L^4} - \frac{1}{6\pi^2}\int_0^{\infty}\mathrm{d}\kappa\frac{\kappa^3}{2} + g(z). \tag{1035}$$

Of course we notice that the first term in the above equation is similar to Eq. (998).

Let us now evaluate $\langle T_{00}\rangle$ outside the cavity. The calculation proceeds along the same lines but we now need to use the Green function given by Eq. (1019). Straightforward manipulations lead to

$$\langle T_{00}\rangle = -\frac{1}{6\pi^2}\int_0^{\infty}\mathrm{d}\kappa\frac{\kappa^3}{2} - \frac{1}{16\pi^2}\frac{1}{(z-L)^4}. \tag{1036}$$

We obtain the same structure as inside the cavity, namely a divergent term and a surface divergence term, *i.e.* a term which is divergent only on the plate. As explained in Ref. [116], this type of terms can be removed by restoring conformal invariance.

Let us now calculate the quantity $\langle T_{zz}\rangle$ inside the cavity. Clearly, following the previous considerations, it can be written as

$$\begin{aligned}\langle T_{zz}\rangle &= \frac{1}{2i}\int\frac{\mathrm{d}\omega}{2\pi}\frac{\mathrm{d}k_x\mathrm{d}k_y}{(2\pi)^2}\left(\partial_t\partial_{\overline{t}} - \partial_x\partial_{\overline{x}} - \partial_y\partial_{\overline{y}} + \partial_z\partial_{\overline{z}}\right)\\ &\quad\times g(z,\overline{z})\Big|_{z=\overline{z}}.\end{aligned} \tag{1037}$$

Following exactly the same procedure as before, we arrive at

$$\langle T_{zz}\rangle = -3\frac{\pi^2}{1440L^4} - \frac{3}{6\pi^2}\int_0^{\infty}\mathrm{d}\kappa\frac{\kappa^3}{2}. \tag{1038}$$

Let us now calculate $\langle T_{zz}\rangle$ outside the plates. As we did for the time time component, we must now use the Green function (1019). This gives

$$\begin{aligned}\langle T_{zz}\rangle &= \frac{1}{2i}\int\frac{\mathrm{d}\omega}{2\pi}\frac{k_x\mathrm{d}k_y}{(2\pi)^2}\frac{\mathrm{e}^{i\lambda(z-L)}}{\lambda}\lambda^2\Big[\sin\left(\lambda z-\lambda L\right)\\ &\quad + i\cos\left(\lambda z-\lambda L\right)\Big]\end{aligned} \tag{1039}$$

$$= -\frac{3}{6\pi^2}\int_0^{\infty}\frac{\kappa^3}{2}\mathrm{d}\kappa. \tag{1040}$$

Of course, this quantity is divergent.

Finally, in order to complete the determination of the stress-energy tensor, one must calculate $\langle T_{xx}\rangle = \langle T_{yy}\rangle$. It can be expressed as

$$\begin{aligned}\langle T_{xx}\rangle &= \frac{1}{2i}\int\frac{\mathrm{d}\omega}{2\pi}\frac{\mathrm{d}k_x\mathrm{d}k_y}{(2\pi)^2}\left(\partial_t\partial_{\overline{t}}+\partial_x\partial_{\overline{x}}-\partial_y\partial_{\overline{y}}-\partial_z\partial_{\overline{z}}\right)\\ &\quad\times g(z,\overline{z})\Big|_{z=\overline{z}} \qquad (1041)\\ &= \frac{1}{2i}\int\frac{\mathrm{d}\omega}{2\pi}\frac{\mathrm{d}k_x\mathrm{d}k_y}{(2\pi)^2}\frac{1}{\lambda\sin(\lambda L)}\Big[\left(\omega^2+k_x^2-k_y^2\right)\\ &\quad\times\sin(\lambda z)\sin(\lambda L-\lambda z)\\ &\quad+\lambda^2\cos(\lambda z)\cos(\lambda L-\lambda z)\Big]. \qquad (1042)\end{aligned}$$

A first step consists in dealing with the terms proportional to $k_x^2-k_y^2$. In fact these terms are of the form $\int \mathrm{d}\boldsymbol{k}_\perp(k_x^2-k_y^2)\mathcal{F}(k_\perp)$, where $\mathcal{F}(k_\perp)$ represents the other terms in the integral which are function of $k_x^2+k_y^2$ only. Therefore, one can evaluate them by going to polar coordinates in the plane (k_x,k_y). This leads to an expression of the form $\int \mathrm{d}k_\perp\mathrm{d}\psi k_\perp^2(\cos^2\psi-\sin^2\psi)\mathcal{F}(k_\perp)=0$, which shows that the corresponding contribution vanishes. Then, following the same steps as before, one reduces the above expression to

$$\begin{aligned}\langle T_{xx}\rangle &= \frac{1}{12\pi^2}\int_0^\infty\mathrm{d}\kappa\frac{\kappa^3}{\sinh(\kappa L)}\Big[\cosh(2\kappa z-\kappa L)+\\ &\quad+2\cosh(\kappa z)\cosh(\kappa L-\kappa z)\Big]=-\langle T_{00}\rangle, \qquad (1043)\end{aligned}$$

where we have used that $2\cosh(\kappa z)\cosh(\kappa L-\kappa z)=\cosh(\kappa L)+\cosh(2\kappa z-\kappa L)$. Of course, as already mentioned, this also shows that $\langle T_{yy}\rangle=-\langle T_{00}\rangle$. All the other components being zero, this completes our calculation of the stress energy in the cavity. Only the calculations of $\langle T_{xx}\rangle=\langle T_{yy}\rangle$ outside the plates remains to be done and we now turn to this question. Considering again Eq. (1041), we have

$$\begin{aligned}\langle T_{xx}\rangle &= \frac{1}{2i}\int\frac{\mathrm{d}\omega}{2\pi}\frac{\mathrm{d}k_x\mathrm{d}k_y}{(2\pi)^2}\frac{\mathrm{e}^{i\lambda(z-L)}}{\lambda}\Big[\omega^2\sin(\lambda z-\lambda L)\\ &\quad-i\lambda^2\cos(\lambda z-\lambda L)\Big]=-\langle T_{00}\rangle. \qquad (1044)\end{aligned}$$

We are now in a position where we can write the full stress energy tensor. Inside the cavity, it takes the form

$$\begin{aligned}\langle T_{\mu\nu}\rangle_{\text{inside}} &= (\rho_{\text{vac}}+\rho_{\text{Casimir}})\begin{pmatrix}1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&3\end{pmatrix}\\ &\quad+g(z)\begin{pmatrix}1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&0\end{pmatrix}, \qquad (1045)\end{aligned}$$

where we have defined the two quantities ρ_{vac} and ρ_{Casimir} by

$$\rho_{\text{vac}}=-\frac{1}{6\pi^2}\int_0^\infty\mathrm{d}\kappa\frac{\kappa^3}{2},\quad \rho_{\text{Casimir}}=-\frac{\pi^2}{1440L^4}. \qquad (1046)$$

Outside the cavity (more precisely on the right hand side), the results derived above imply that

$$\langle T_{\mu\nu}\rangle_{\text{outside}} = \rho_{\text{vac}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} - \frac{1}{16\pi^2(z-L)^4} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{1047}$$

As discussed in Ref. [116], the terms proportional to $g(z)$ and $(z-L)^{-4}$ are surface divergent terms. They can be removed in the case where conformal invariance is re-established and, for that reason, are not present in electromagnetism, *i.e.* when one considers an electric field inside the cavity rather than a scalar field. In this case, we see that the difference between the stress-energy tensors inside and outside the cavity reads

$$\Delta\langle T_{\mu\nu}\rangle = \rho_{\text{Casimir}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}, \tag{1048}$$

which is a finite result. The infinite part has been removed by subtraction. Finally, the stress energy tensor can be written in a covariant form, namely [119]

$$\Delta\langle T_{\mu\nu}\rangle = \rho_{\text{Casimir}} \left(4\hat{z}^\mu \hat{z}^\nu - g^{\mu\nu}\right), \tag{1049}$$

where $\hat{z}^\mu$ in the unit vector in the z direction.

At this point, several comments are in order. Firstly, we notice in Eqs. (1048) and (1049) that the energy density (*i.e.* the time-time component) is similar to the expressions already derived in Sec. 24.12.2 by other methods. This confirms that this calculation is independent of the regularization scheme used to obtain the result. Secondly, it is clear that the stress-energy tensor (1049) is not similar to the cosmological constant stress-energy tensor $T_{\mu\nu} = -\rho_{\text{vac}} g_{\mu\nu}$. This illustrates the difference between the Casimir case and the cosmological constant case. In the latter case, one measures the absolute value of the vacuum energy while, in the former case, one is only sensitive to the difference of vacuum energy from one side of the plates to the other.

Thirdly, the Casimir Force has been observed in the laboratory and this is usually taken as another evidence in favor of the existence of the zero-point fluctuations. After all, since we observe the Casimir force and since this force is due to a change in the structure of the vacuum that is perturbed by the presence of the plates, the fact is that the vacuum must exist as a real physical phenomenon and is not an artifact of the quantum field theory formalism. In addition, this also validates the methods of regularization used to calculate ρ_{vac}. In the context of the Casimir effect, these methods seem to work very well and to lead to predictions that are in agreement with the experiments. Therefore, why would the very same techniques, used in a similar context, fail to regularize the cosmological constant?

Fourthly, the claim that the calculation and the observation of the Casimir force are evidences in favor of the reality of vacuum fluctuations has been challenged in Refs. [120, 121]. It is indeed possible to derive the expression of the Casimir force without referring at all to zero point fluctuations, by means of the "source theory" approach to quantum field theory. However, to our knowledge, there is no attempt to do the same for the Lamb shift (but this would certainly be an interesting exercise). Therefore, it seems that zero-points fluctuations are nevertheless necessary to quantum field theory even if the above remark should lead us to tone down the claim that they have been seen in the laboratory. Let us also notice that, very recently, another situation where the vacuum fluctuations can lead to an observable effect has been studied in Ref. [122]. This article investigates whether the quantum zero point fluctuations could cause the rotation of a small nano device. This "rotational vacuum effect" has in fact a similar origin than the Casimir effect.

Finally, it should be clear that, if the Lamb shift and the Casimir effect seem to indicate that the vacuum fluctuations are real, they do not say anything with regards to their gravitational properties. In other words, they do not say how the vacuum fluctuations weigh (it

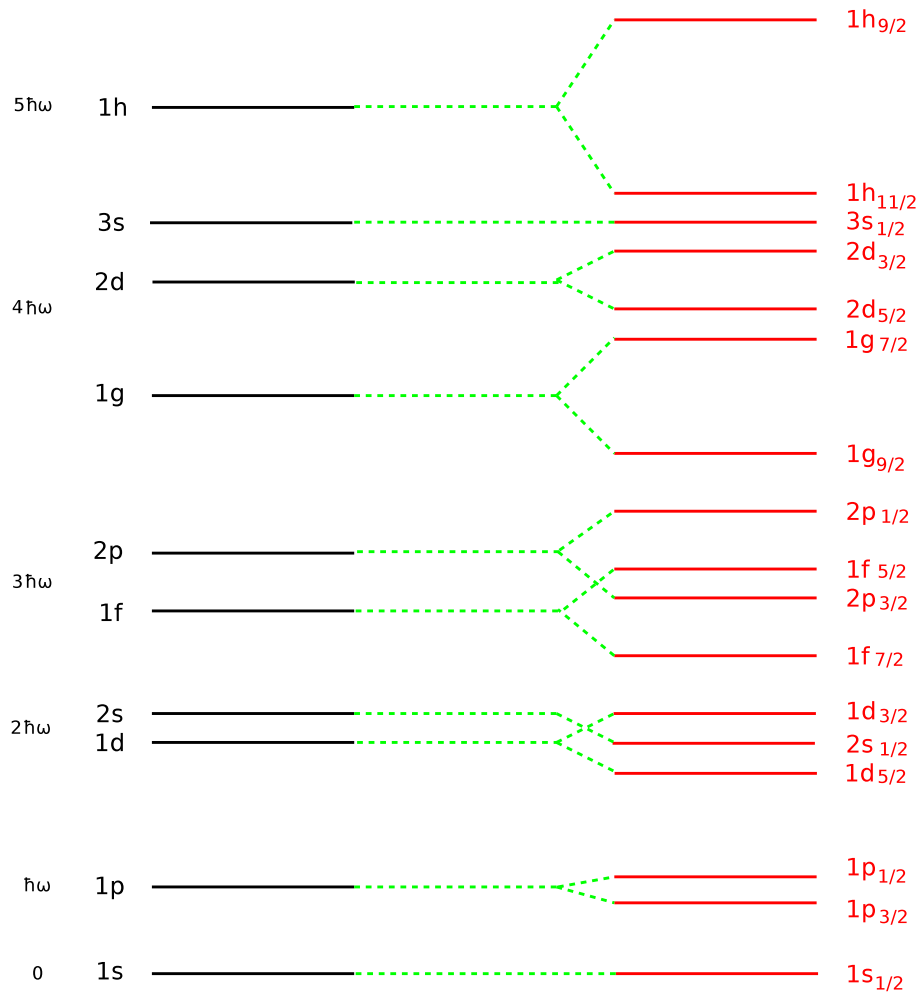


Figure 161: Energy levels of the nucleus. On the left-hand side are represented the levels in the approximation where the Woods-Saxon potential is approximated by a parabola. In this case, each level (which includes a series of sub-levels) is separated by the energy $\hbar\omega$. On the right hand side are represented the levels where the spin orbit interaction (1051) is taken into account. This breaks the degeneracy of the sub-levels and reproduces the magic numbers.

is worth noticing here that the gravitational properties of a Casimir cavity have been studied in Refs. [123, 124, 125, 126, 127, 128, 129]). It is clear that there exists the possibility that the zero-point fluctuations do exist but have non-standard gravitational properties, *i.e.* that they abnormally weigh. It is therefore interesting to investigate if there is an experiment which measures whether this is true or not. In the next section, we turn to this question.

24.13 Do the Quantum Fluctuations Gravitate?

In this section, we study the gravitational property of the Lamb shift [12, 130]. We will follow the treatment of Ref. [130]. In Sec. 24.12.1, we showed that the interaction of the electron with the "vacuum electric field" inside the atom leads to a shift of the atomic levels. In some sense, this means that the zero-point fluctuations affect the mass (energy) of the atom. Therefore, by studying how the atom falls down, one can put constraints on the gravitational properties of the vacuum fluctuations. This is the idea that we pursue in this section. However, since the corresponding effect is small for an atom and is larger for a nucleus (of course, the Lamb shift also exists in this case), it is better to study the latter case. For this reason, we start with recalling some basics facts about nuclear models.

The main idea is that the interaction between a given nucleon and the other $A-1$ nucleons

can be mimicked by a Woods-Saxon potential given by

$$V_{\rm ws}(r) = -V_0 \left[1 - \exp\left(\frac{r-R}{a}\right)\right], \tag{1050}$$

where a describes how abrupt the potential is while R is the radius of the nucleus. As is well-known one can take $R = r_0 A^{1/3}$ where $r_0 \simeq 1.3 \times 10^{-15}\,{\rm m} = 1.3\,{\rm fm}$. The quantity V_0 represents the depth of the potential and can be taken to be $V_0 \simeq 45\,{\rm MeV}$. In fact, in order to have a reasonable model, one must add to the Woods-Saxon potential a spin orbit interaction term which can be expressed as

$$V_{\rm L,s} = -f(r)\boldsymbol{L}\cdot\boldsymbol{s}, \tag{1051}$$

where $\boldsymbol{L}$ is the angular momentum of the nucleon and $\boldsymbol{s}$ its spin. With a good approximation, one can write $f(r) \simeq 24 A^{-2/3}\,{\rm MeV}$. Then, the Hamiltonian of the nucleus can be written as

$$H = \sum_{j=1}^{A}\left[\frac{p_j^2}{2m_j} + V_{\rm ws}(r_j) - f(r_j)\boldsymbol{L}_j\cdot\boldsymbol{s}_j\right]. \tag{1052}$$

The total wave-function can be taken as the product of the individual wave functions (in fact, more precisely, one should consider a Slater determinant in order to take into account the fact that the total wave function must be anti-symmetrized), that is to say

$$\Psi\left(\boldsymbol{r}_1, \cdots, \boldsymbol{r}_A\right) = \prod_{j=1}^{A}\Psi_j\left(\boldsymbol{r}_j\right). \tag{1053}$$

Each individual wave-function can be obtained from a Schrödinger equation given by (for the moment, one ignores the spin orbit term since it can be treated as a perturbation)

$$\left[-\frac{\hbar^2}{2\mu}\nabla^2 + V_{\rm ws}(r_j)\right]\Psi_j(r_j) = E_j\Psi_j(r_j), \tag{1054}$$

where μ is the reduced mass,

$$\frac{1}{\mu} \equiv \frac{1}{m_{\rm p}} + \frac{1}{(A-1)m_{\rm p}}, \tag{1055}$$

or $\mu \simeq m_{\rm p}(A-1)/A \simeq m_{\rm p} \simeq 1.6 \times 10^{-27}\,{\rm kg}$, $m_{\rm p}$ being the proton mass. Unfortunately, the Schrödinger equation cannot be solved analytically for the Woods-Saxon potential and, therefore, we are forced to make approximations. The simplest one is just to describe the Woods-Saxon potential with a parabola, *i.e.*

$$V_{\rm ws}(r) \simeq -V_0\left(1 - \frac{r^2}{R^2}\right) = -V_0 + \frac{1}{2}\mu\omega_0^2 r^2, \tag{1056}$$

with $\omega_0 \equiv \sqrt{2V_0/(\mu R^2)}$. In this case, the question becomes solvable. We can use the spherical symmetry of the problem and write the wave function as

$$\Psi_j(\boldsymbol{r_j}) = \frac{u_{n,\ell}(r_j)}{r_j} Y_{\ell m}\left(\theta_j, \varphi_j\right), \tag{1057}$$

where $Y_{\ell m}$ is a spherical harmonics [48, 49]. The radial function $u_{n\ell}$ is controlled by the equation

$$\left[-\frac{\hbar^2}{2\mu}\frac{{\rm d}^2}{{\rm d}r_j^2} + \frac{\ell(\ell+1)}{2\mu r_j^2} + \frac{1}{2}\mu\omega_0^2 r^2\right]u_{n,\ell} = (E_j + V_0)u_{n,\ell}, \tag{1058}$$

which can be solved in terms of Laguerre polynomials L_n^α [48, 49]. Explicitly, one has

$$\Psi(r_j) = C r_j^\ell \mathrm{e}^{-\mu\omega_0 r_j^2/(2\hbar)} L_{n-1}^{\ell+1/2}\left(\frac{\mu\omega_0}{\hbar} r^2\right) Y_{\ell m}\left(\theta_j, \varphi_j\right), \tag{1059}$$

where C is a normalization constant to be determined. In the above expression, one has $n \geq 1$ and $\ell \geq 0$. The energy levels are given by the expression

$$E(n,\ell) = -V_0 + \hbar\omega_0\left[2(n-1) + \ell + \frac{3}{2}\right]. \tag{1060}$$

The factor $3/2 = 3 \times 1/2$ originates from the fact that we have a three-dimensional harmonic oscillator. The quantum numbers have been chosen so that the ground state is $n = 1$, $\ell = 0$ which allows the spectroscopic notation "$1s$", see Fig. 161.

The only thing which remains to be done is to determine the constant C. Obviously, this is done by normalizing the wave-function, *i.e.*

$$|C|^2 \int_0^\infty \mathrm{d}r_j r_j^2 \frac{|u_{n,\ell}|^2}{r_j^2} \int \mathrm{d}\Omega_j Y_{\ell m} = 1, \tag{1061}$$

Using the solution obtained above and performing the change of variable $\rho = (\mu\omega_0/\hbar)r^2$, the previous condition reduces to

$$\frac{|C|^2}{2}\left(\frac{\mu\omega_0}{\hbar}\right)^{-\ell-3/2} \times \int_0^\infty \rho^{\ell+1/2}\mathrm{e}^{-\rho} L_{n-1}^{\ell+1/2}(\rho) L_{n-1}^{\ell+1/2}(\rho) = 1. \tag{1062}$$

Then, using Eq. (7.414.3) of Ref. [49], one obtains

$$|C|^2 = 2(n-1)!\left(\frac{\mu\omega_0}{\hbar}\right)^{\ell+3/2}\Gamma^{-1}\left(n+\ell+\frac{1}{2}\right). \tag{1063}$$

We are now in a position where we can compute the value of the wave-function at $r_j = 0$ for $\ell = 0$. This gives

$$|\Psi(0)|^2 = |C|^2\left[L_{n-1}^{1/2}(0)\right]^2 \frac{1}{4\pi}. \tag{1064}$$

Using the fact that $L_{n-1}^{1/2}(0) = \binom{n-1/2}{n-1} = \Gamma(n+1/2)/[\sqrt{\pi}(n-1)!/2]$, one obtains

$$|\Psi_{n,\ell=0}(0)|^2 = \frac{2}{\pi^2}\frac{\Gamma(n+1/2)}{(n-1)!}\left(\frac{\mu\omega_0}{\hbar}\right)^{3/2}. \tag{1065}$$

This completes the first part of the calculation.

We have established on very general grounds that the shift in energy due to the interaction of a charged particle (in Sec. 24.12.1, the electron in the atom, here the proton in the nucleus) is given by Eq. (953). The means value of the square of the displacement has been calculated in Eq. (961). Applying these formulas to the case of the protons in the nucleus, one obtains [130]

$$\Delta E(n,0) \simeq \frac{4\alpha^2}{3m_{\mathrm{p}}^2}\sum_i Z^{(i)}\left|\Psi^{(i)}_{n,\ell=0}(0)\right|^2 \int_0^\infty \frac{\mathrm{d}\omega}{\omega}. \tag{1066}$$

As usual, this infinite integral must be regularized and usually leads to logarithmic corrections, see for instance Eq. (963). We will ignore this factor in what follows since this does not affect

too much the final result. As explained in Ref. [130], the sum over "(i)" runs only over s-wave since this is the only way to get $\Psi_{n,\ell}(0) \neq 0$. Moreover, still following Ref. [130], it is reasonable to approximate the field experienced by a proton by a field produced by the other protons in inner shells. This is why $Z^{(i)}$ represents the total number of protons present in shells inner than the proton "(i)" under consideration.

Our goal is now to test how the vacuum fluctuations weigh. For this purpose, we write the gravitational mass of a given nucleus as [130]

$$m_{\rm g} = m_{\rm i} - \eta \Delta E, \tag{1067}$$

where $m_{\rm i}$ is the inertial mass. The parameter η is a priori a free parameter which controls to which extent the weak equivalence principle would be violated. If $\eta = 0$, there is no violation of the universality of free fall. A convenient measure of potential violations of the weak equivalence principle is given by the Eötvös ratio

$$\eta(1,2) \equiv \frac{m_{\rm g}(1)}{m_{\rm i}(1)} - \frac{m_{\rm g}(2)}{m_{\rm i}(2)}, \tag{1068}$$

for two bodies 1 and 2. Using the previous calculations, this ratio can be expressed as

$$\begin{aligned} \eta(1,2) &= \eta \frac{4\alpha^2}{3m_{\rm p}^3} \left[\frac{1}{A_1} \sum_i Z_1^{(i)} \left| \Psi_{n,\ell=0}^{(1,i)}(0) \right|^2 \right. \\ &\left. - \frac{1}{A_2} \sum_i Z_2^{(i)} \left| \Psi_{n,\ell=0}^{(2,i)}(0) \right|^2 \right], \end{aligned} \tag{1069}$$

where we have used $m_{\rm i} \sim A m_{\rm p}$, a relation valid since we work at first order in the parameter η. Then, we use the fact that

$$(\mu\omega_0)^{3/2} = \left(\frac{2m_{\rm p}V_0}{r_0^2} \right)^{3/4} \frac{1}{A^{1/2}} \tag{1070}$$

to express the Eötvös ratio as [130]

$$\begin{aligned} \eta(1,2) &= \eta \frac{4\alpha^2}{3} \left(\frac{2V_0}{m_{\rm p}^3 r_0^2} \right)^{3/4} \frac{2}{\pi^2} \\ &\times \left[\frac{1}{A_1^{3/2}} \sum_i Z_1^{(i)} \frac{\Gamma(n+1/2)}{(n-1)!} \right. \\ &\left. - \frac{1}{A_2^{3/2}} \sum_i Z_2^{(i)} \frac{\Gamma(n+1/2)}{(n-1)!} \right]. \end{aligned} \tag{1071}$$

Let us evaluate this number for aluminum ($Z = 13$, $A = 27$) and Platinum ($Z = 78$, $A = 195$). In the case of aluminum, since we only have 13 protons, the only s-shell protons are on the $1\,s_{1/2}$ level (in fact, we have two protons in $1\,s_{1/2}$, four protons in $1\,p_{3/2}$, two protons in $1\,p_{1/2}$ and five protons in $1\,d_{5/2}$, see Fig. 161). As a consequence, there is no inner protons for those protons (since, obviously, there is no inner level than the ground state $1\,s_{1/2}$) and the corresponding contribution in the Eötvös ratio vanishes. In the case of Platinum, all the levels are filled up until the level $1\,h_{11/2}$ which contains height protons, see Fig. 161. Therefore, we have two protons on $1\,s_{1/2}$, two protons on $2\,s_{1/2}$ and two on $3\,s_{1/2}$, see again Fig. 161. This means that the $3\,s_{1/2}$ protons see four inner protons and the $2\,s_{1/2}$ protons see two inner protons. As a consequence, one can write

$$\begin{aligned} \eta({\rm Al},{\rm Pt}) &\simeq -\eta \frac{4\alpha^2}{3} \left(\frac{2V_0}{m_{\rm p}^3 r_0^2} \right)^{3/4} \frac{2}{\pi^2} \frac{1}{A_{\rm Pt}^{3/2}} \\ &\times \left[\frac{2}{(2-1)!} \Gamma\left(2+\frac{1}{2}\right) + \frac{4}{(3-1)!} \Gamma\left(3+\frac{1}{2}\right) \right] \end{aligned} \tag{1072}$$

Using the values of V_0 and r_0 chosen previously this leads to

$$\eta(\mathrm{Al}, \mathrm{Pt}) \simeq -5.6 \times 10^{-10}\, \eta \tag{1073}$$

Given the fact that $|\eta(\mathrm{Al}, \mathrm{Pt})| \lesssim 10^{-12}$ [131, 132], we reach the conclusion that

$$\eta \lesssim 1.7 \times 10^{-3}. \tag{1074}$$

Therefore, the vacuum fluctuations seem to have standard gravitational properties to a very good accuracy.

The conclusion of this section is that it seems difficult to argue that the zero-point fluctuations do not gravitate. Clearly, this makes the cosmological constant problem more acute. However, the results obtained here are by no means a final proof. For instance, it seems reasonable to assume that the stress energy tensor of the Lamb shift is not of the cosmological constant type. Therefore, the fact that a nucleus falls down normally does not necessarily imply that the cosmological constant weighs according to the Einstein equations. Nevertheless, no sign of inconsistency appears in the above situation and one is obviously tempted to assume that zero-point fluctuations couple to gravity as any other type of matter. Again, this makes the cosmological constant problem more mysterious.

24.14 The Weak Equivalence Principle in Quantum Mechanics

In the previous sections, we have argued that the zero-point fluctuations seem to be real and seem to gravitate normally. Therefore, these two reasons cannot be invoked to avoid the cosmological constant problem. However, there is another (but related) issue that it is interesting to investigate. When we consider the vacuum stress energy tensor in the Einstein equations, we implicitly assume that the zero-point fluctuations obey the weak equivalence principle, *i.e.* that they "fall" in a gravitational field as any other type of matter. Clearly the zero point fluctuations are of quantum-mechanical origin. The weak equivalence principle is well established (and tested) in classical physics and general relativity is based on this principle. But what is the status of this principle in quantum mechanics? If it is not valid in quantum mechanics, maybe we do not have the right to couple the quantum zero point fluctuations in the standard way in the Einstein equations? Could it be a way to avoid the problem? In order to discuss this issue, we will consider two situations, the Collela, Overhausser and Werner experiment (COW) and the quantum Galileo experiment.

24.14.1 The Collela, Overhausser and Werner (COW) Experiment

The COW experiment [133, 134, 135] consists in sending neutrons in an interferometer placed in a weak gravitational field, see Fig. 162. The presence of the gravitational field affects the wave-function of the neutrons and changes the relative phases between the two beams. When the interferometer is rotated, the gravitational phase shift is modified because the height difference between the two arms is changed. This effect can then be detected (and was detected) in the interference pattern. The COW experiment is conceptually very important. This was the first experiment to measure the effect of the gravitational field on the wave-function. In fact, it shows that the gravitation is just an ordinary force as far as the Schrödinger equation is concerned.

Let us now describe the experiment in more detail. The neutron interferometer is represented in Fig. 162. Let us assume that, along one of the two paths, a phase shift $\Delta\varphi$ is introduced. Our first goal is to evaluate the intensity seen by the detectors O_2 and O_3. After diffraction, the wave function of the neutron takes the form

$$\Psi(\boldsymbol{r}) = \Phi T(\theta) \mathrm{e}^{i\boldsymbol{k}_0\cdot\boldsymbol{r}} + \Phi D(\theta) \mathrm{e}^{-i\boldsymbol{k}_\mathrm{G}\cdot\boldsymbol{r}}, \tag{1075}$$

where $\boldsymbol{k}_0$ is the incident wave-number and $\boldsymbol{k}_\mathrm{G}$ the wave number after Bragg diffraction (which, therefore, must satisfy the Bragg condition). The quantity Φ is the amplitude of the incident wave and $T = T(\theta)$ (for "transmitted") and $D = D(\theta)$ (for "diffracted") are two coefficients describing the amplitude of the two branches of the wave-function after the neutron has emerged

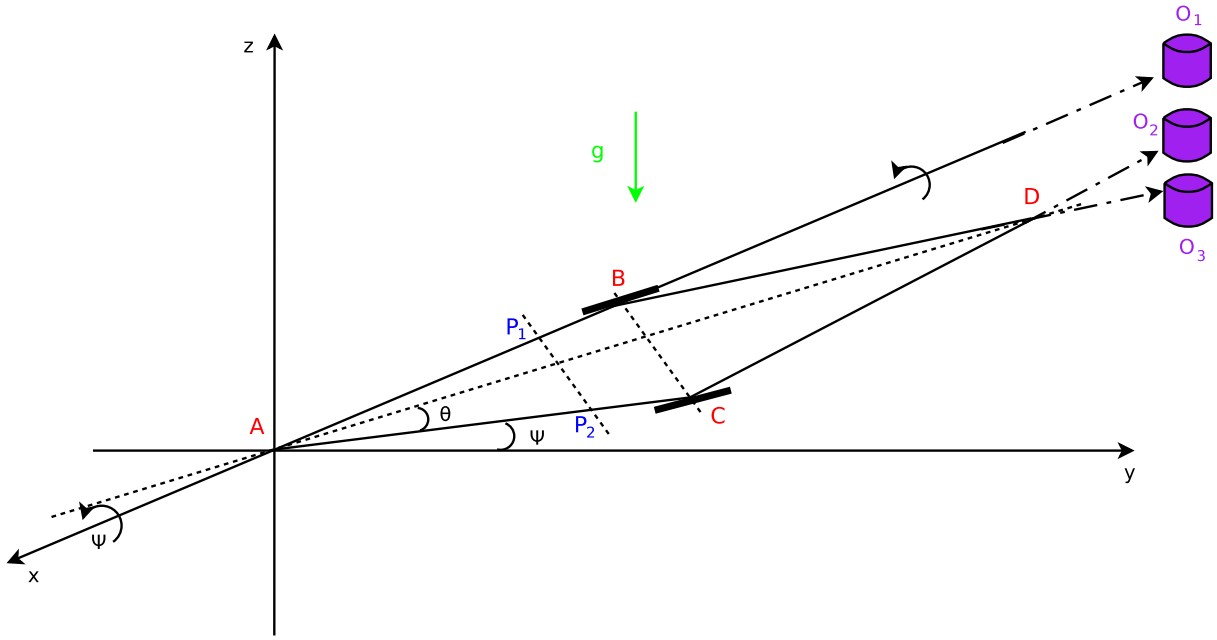


Figure 162: Representation of the Collela, Overhausser and Werner (COW) interferometer. The interferometer is placed in a vertical gravitational field (represented by the green arrow) and is rotated around the axis AB.

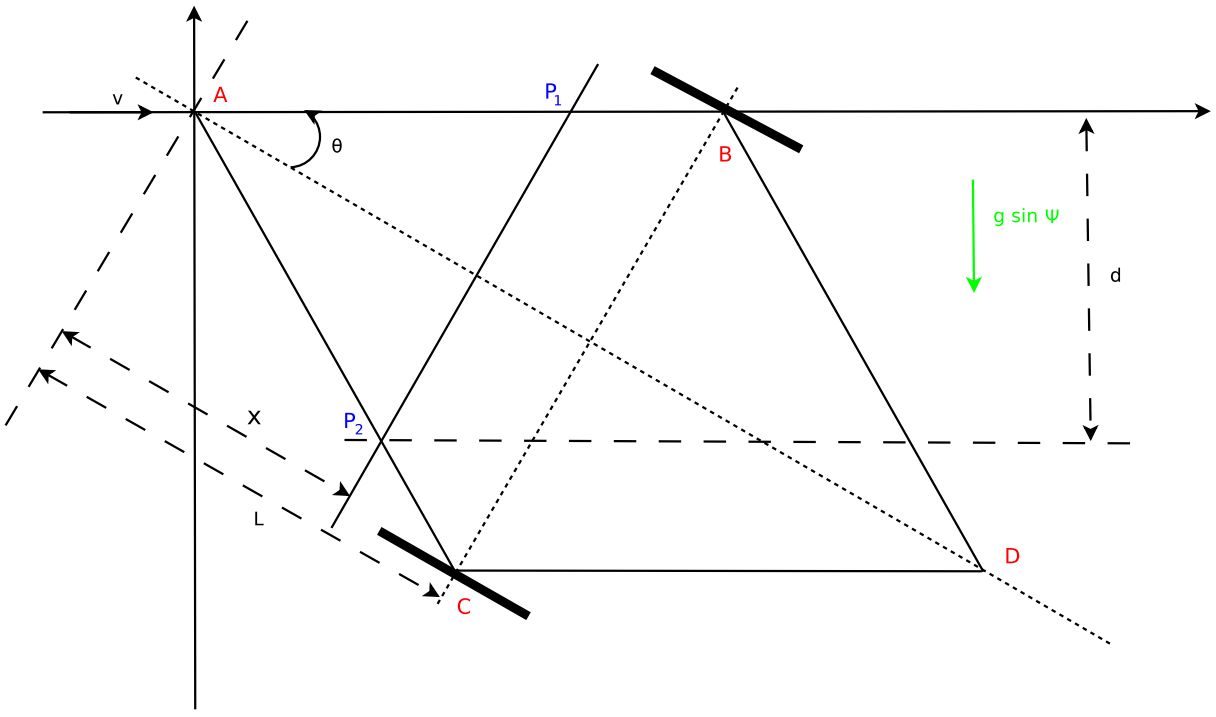


Figure 163: Representation of the COW interferometer in the plane of the two beams.

from the crystal. Let us now consider the the path $ABDO_3$. At point B, the wave function is given by (the origin of the coordinates is chosen to be at point A)

$$\Psi(AB) = \Phi T(\theta) \mathrm{e}^{i\boldsymbol{k}_0\cdot\boldsymbol{r}_\mathrm{B}}, \tag{1076}$$

since the wave was just transmitted. Then, at point B, we have a diffraction and, therefore, the wave-function at point D reads

$$\Psi(ABD) = \Phi T(\theta) \mathrm{e}^{i\boldsymbol{k}_0\cdot\boldsymbol{r}_\mathrm{B}} D(\theta) \mathrm{e}^{i\boldsymbol{k}_\mathrm{G}\cdot(\boldsymbol{r}_\mathrm{D}-\boldsymbol{r}_\mathrm{B})}. \tag{1077}$$

Finally, at point D, we have another diffraction but this time with opposite angle. The leads to

$$\begin{aligned}\Psi(ABDO_3) &= \Phi T(\theta) \mathrm{e}^{i\boldsymbol{k}_0\cdot\boldsymbol{r}_\mathrm{B}} D(\theta) \mathrm{e}^{i\boldsymbol{k}_0\cdot(\boldsymbol{r}_\mathrm{D}-\boldsymbol{r}_\mathrm{B})} \\ &\quad \times D(-\theta) \mathrm{e}^{i\boldsymbol{k}_\mathrm{G}\cdot(\boldsymbol{r}-\boldsymbol{r}_\mathrm{D})}.\end{aligned} \tag{1078}$$

We also assume that, along that path, there is a phase shift $\Delta\varphi$ that we do not specify for the moment. As a consequence, the total wave function can be written as

$$\begin{aligned}\Psi(ABDO_3) &= \Phi T(\theta) \mathrm{e}^{i\boldsymbol{k}_0\cdot\boldsymbol{r}_\mathrm{B}} D(\theta) \mathrm{e}^{i\boldsymbol{k}_0\cdot(\boldsymbol{r}_\mathrm{D}-\boldsymbol{r}_\mathrm{B})} \\ &\quad \times D(-\theta) \mathrm{e}^{i\Delta\varphi} \mathrm{e}^{i\boldsymbol{k}_0\cdot(\boldsymbol{r}-\boldsymbol{r}_\mathrm{D})},\end{aligned} \tag{1079}$$

where we have introduced the factor $\mathrm{e}^{i\Delta\varphi}$.

It is straightforward to follow the same procedure and to establish the expression for $\Psi(ACDO_3)$. This gives the expression of the wave function at detector O_3, namely $\Psi(O_3) = \Psi(ABDO_3) + \Psi(ACDO_3)$,

$$\Psi(O_3) = \Phi \left[T(\theta) \mathrm{e}^{i\boldsymbol{k}_0\cdot\boldsymbol{r}_\mathrm{B}} D(\theta) \mathrm{e}^{i\boldsymbol{k}_0\cdot(\boldsymbol{r}_\mathrm{D}-\boldsymbol{r}_\mathrm{B})} D(-\theta) \mathrm{e}^{i\Delta\varphi} + D(\theta) \mathrm{e}^{i\boldsymbol{k}_0\cdot\boldsymbol{r}_\mathrm{C}} D(-\theta) \mathrm{e}^{i\boldsymbol{k}_0\cdot(\boldsymbol{r}_\mathrm{D}-\boldsymbol{r}_\mathrm{C})} T(\theta) \right] \mathrm{e}^{i\boldsymbol{k}_0\cdot(\boldsymbol{r}-\boldsymbol{r}_\mathrm{D})}. \tag{1080}$$

Using this expression, one can calculate the expected intensity. Straightforward manipulations leads to the following expression

$$I(O_3) = |\Psi(O_3)|^2 = \alpha\left(1 + \cos\Delta\varphi\right), \tag{1081}$$

where α is a constant that can be determined from the above expression and which depends on the incident flux and the crystal structure. Then, one can repeat the same analysis and determine the wave function at detector O_2. It reads

$$\Psi(O_2) = \Phi \left[T(\theta) \mathrm{e}^{i\boldsymbol{k}_0\cdot\boldsymbol{r}_\mathrm{B}} D(\theta) \mathrm{e}^{i\boldsymbol{k}_\mathrm{G}\cdot(\boldsymbol{r}_\mathrm{D}-\boldsymbol{r}_\mathrm{B})} T(-\theta) \mathrm{e}^{i\Delta\varphi} + D(\theta) \mathrm{e}^{i\boldsymbol{k}_\mathrm{G}\cdot\boldsymbol{r}_\mathrm{C}} D(-\theta) \mathrm{e}^{i\boldsymbol{k}_0\cdot(\boldsymbol{r}_\mathrm{D}-\boldsymbol{r}_\mathrm{C})} D(\theta) \right] \mathrm{e}^{i\boldsymbol{k}_\mathrm{G}\cdot(\boldsymbol{r}-\boldsymbol{r}_\mathrm{D})}. \tag{1082}$$

The intensity at detector O_2 can be computed along the same lines and one is led to

$$I(O_2) = \beta - \alpha\cos\Delta\varphi, \tag{1083}$$

where α and β are just constants that do not play an important role in what follows [of course, the constant α is the same as in Eq. (1081)].

Having estimated the form of the signal expected for any phase shift $\Delta\varphi$, we now turn to its calculation in the case of the COW experiment. For this purpose, we return to the Schrödinger equation which can be written as

$$\nabla^2\Psi + \omega^2\Psi = 0, \tag{1084}$$

with

$$\omega^2 = \frac{2m}{\hbar^2}\left(E - V\right). \tag{1085}$$

Now, let us assume that the potential is given by

$$V = V_0 + \Delta V(\boldsymbol{r}), \tag{1086}$$

where V_0 is a constant and ΔV a small space-dependent small perturbation, *i.e.* $\Delta V \ll V_0$. If the perturbation does not change abruptly, the Wentzel-Kramers-Brillouin (WKB) wave-function

$$\Psi \propto \exp\left[i \int \omega(\boldsymbol{r}) \mathrm{d}\boldsymbol{r} \right], \tag{1087}$$

is a very good approximation to the actual solution. Introducing the notation $k_0 = 2m(E - V_0)/\hbar^2$ and using that $p = mv = \hbar k_0$, an expansion of the phase in $\Delta V/V_0$ leads to the following expression

$$\Psi \propto \exp\left(i\boldsymbol{k}_0 \cdot \boldsymbol{r} - \frac{1}{\hbar}\int\limits_{\mathcal{P}} \Delta V \mathrm{d}t \right), \tag{1088}$$

where the subscript "$\mathcal{P}$" indicates that one integrates along the non perturbed path. From the above equation, we conclude that

$$\Delta\varphi = -\frac{1}{\hbar}\int\limits_{\mathcal{P}} \Delta V \mathrm{d}t. \tag{1089}$$

The same result can also be expressed differently. Using energy conservation, on can write

$$\frac{p^2}{2m} + V_0 + \Delta V = E, \tag{1090}$$

and, since $\boldsymbol{p} = \boldsymbol{p}_0 + \delta\boldsymbol{p}$, one obtains

$$\frac{p_0^2}{2m} + \frac{\boldsymbol{p}_0}{m} \cdot \delta\boldsymbol{p} + V_0 + \Delta V = E, \tag{1091}$$

or,

$$\boldsymbol{v}_0 \cdot \delta\boldsymbol{p} = -\Delta V. \tag{1092}$$

As a consequence, the phase shift can be re-written as the following expression

$$\Delta\varphi = \frac{1}{\hbar}\int\limits_{\mathcal{P}} \delta\boldsymbol{p} \cdot \mathrm{d}\boldsymbol{r}, \tag{1093}$$

where the integral should again be calculated along the path of the neutron.

Let us now determine the phase shift for the COW experiment. Let us consider two point P_1 and P_2 that correspond to one beam transmitted and one beam diffracted in the interferometer, see Fig. 163. The potential difference between those two points is given by

$$V(P_2) - V(P_1) = mgd\sin\psi, \tag{1094}$$

where d can be expressed as $d = AP_1\cos(\pi/2 - 2\theta)$. Given that the quantity x in Fig. 163 can be written as $x = AP_1\cos\theta$, one deduces that $d = 2x\sin\theta$. As a consequence,

$$V(P_2) - V(P_1) = 2m_{\mathrm{grav}}gx\sin\theta\sin\psi, \tag{1095}$$

where we have now carefully written that the mass which appeared in the above expression is the gravitational mass (since this is the "coupling constant" to the gravitational field). This implies that the phase shift can be written as

$$\Delta\varphi = -\frac{2}{\hbar}\int\limits_{0}^{L} 2m_{\mathrm{g}}g_0 x\sin\theta\sin\psi\frac{\mathrm{d}x}{\cos\theta}, \tag{1096}$$

where we have taken into account the fact that the line P_1P_2 is "moving" with a speed $v\cos\theta$. The factor two in front of the whole expression comes from the fact that the phase shift accumulates and is, along the total path, twice its value in the fist part of the interferometer. The integration over the variable x is easily done and we obtain

$$\Delta\varphi = -2\frac{m_{\mathrm{g}}g_0}{\hbar v}L^2\tan\theta\sin\psi. \tag{1097}$$

Finally, noticing that the total area of the interferometer is nothing but $\mathcal{A} \equiv 2L^2 \tan\theta$ and that the de Broglie relation is $v = h/(m_{\rm in}\lambda)$, where $m_{\rm in}$ is the inertial mass and λ the wavelength of the incident beam, we arrive at [133, 134, 135]

$$\Delta\varphi = -\frac{2\pi}{h^2} m_{\rm grav} m_{\rm in} g_0 \mathcal{A}\lambda \sin\psi. \tag{1098}$$

This expression is quite remarkable as it involves the gravitational and the inertial mass together with the Planck constant. Moreover, we deal with an observable quantity which does not depend on the ratio of the gravitational mass to the inertial mass, as is usually the case, but on the product of these quantities. We also see that the intensity of the signal at detectors O_2 and/or O_3 depend on the orientation of the interferometer ψ.

However, in practice, one must deal with two other effects [133, 134, 135]. The first one is the Sagnac effect due to the rotation of Earth. Classically, the Hamiltonian of a neutron is given by

$$H = \frac{p^2}{2m_{\rm in}} + m_{\rm grav}\boldsymbol{g}\cdot\boldsymbol{r} - \boldsymbol{\omega}\cdot\boldsymbol{L}, \tag{1099}$$

where $\boldsymbol{L} = \boldsymbol{r}\times\boldsymbol{p}$ is the angular momentum of the neutron and $\boldsymbol{\omega}$ the angular velocity of Earth. Upon using the Hamilton equations, this implies that

$$\boldsymbol{p} = m_{\rm in}\,\dot{\boldsymbol{r}} + m_{\rm in}\,\boldsymbol{\omega}\times\boldsymbol{r}. \tag{1100}$$

Therefore, one obtains a Sagnac phase shift which can be expressed as

$$\Delta\varphi_{\rm Sagnac} = \frac{m_{\rm in}}{\hbar}\oint(\boldsymbol{\omega}\times\boldsymbol{r})\cdot{\rm d}\boldsymbol{r}, \tag{1101}$$

where we have used Eq. (1093). Using Stokes theorem, this can also be written as

$$\Delta\varphi_{\rm Sagnac} = \frac{4\pi m_{\rm in}}{h}\boldsymbol{\omega}\cdot\boldsymbol{\mathcal{A}}, \tag{1102}$$

where $\boldsymbol{\mathcal{A}}$ is the vector associated to the interferometer area. If Γ is the incident neutron beam west of due south and δ the co-latitude, the Sagnac phase shift takes the form

$$\Delta\varphi_{\rm Sagnac} = \frac{4\pi m_{\rm in}}{h}\omega\mathcal{A}\left(\cos\psi\cos\delta + \sin\psi\sin\Gamma\sin\delta\right). \tag{1103}$$

The COW experiment is such that the incident beam is directed due south which implies that $\Gamma = 0$. In the case, one obtains

$$\Delta\varphi_{\rm Sagnac} = \frac{4\pi m_{\rm i}}{h}\omega\mathcal{A}\cos\delta\cos\psi, \tag{1104}$$

a dependence in ψ which is different from the gravitational phase shift, see Eq. (1098).

Finally, there is a shift due to the fact that the interferometer bends and/or warps under its own weight. The COW experiment claim that this effect can be described by the following equation

$$\Delta\varphi_{\rm bend} = -q_{\rm bend}\sin\psi. \tag{1105}$$

This expression can be justified by noticing that the bending effect depends on the rotation angle. Clearly, this effect is difficult to estimate from first principles.

Therefore, the total phase shift expected in the COW experiment is the sum of the three contributions discussed before, namely the gravity, Sagnac and bending shifts. This leads to the following expression

$$\Delta\varphi = -q_{\rm grav}\sin\psi + q_{\rm Sagnac}\cos\psi - q_{\rm bend}\sin\psi, \tag{1106}$$

where, in order to take into account Eqs. (1098) and (1104), we have defined $q_{\rm grav} \equiv -2\pi m_{\rm grav} m_{\rm in} g_0 \mathcal{A}\lambda/h^2$ and $q_{\rm Sagnac} = 4\pi m_{\rm in}\omega\mathcal{A}\cos\delta/h$. The above expression can be re-written as

$$\Delta\varphi = q\sin\left(\psi - \psi_0\right), \tag{1107}$$

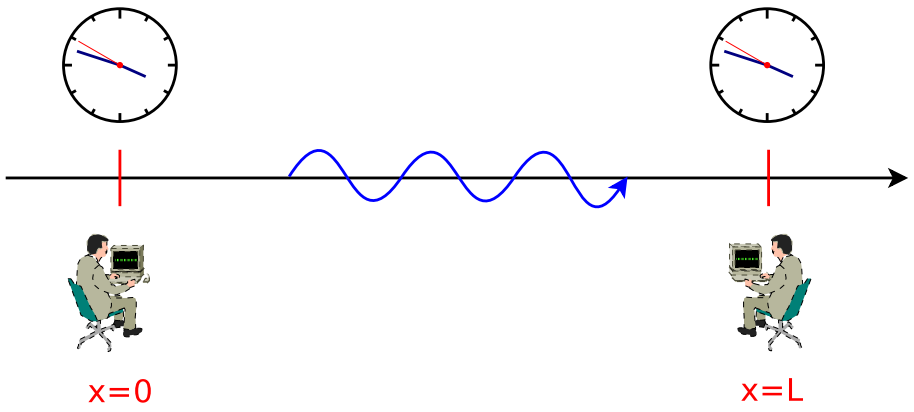


Figure 164: Sketch of a quantum "Peres" clock. The clock is coupled to the particle traveling from the point $x = 0$ to the point $x = L$. When the particle crosses $x = 0$, the clock changes its state and starts running. When the particle reaches $x = L$, the clock stops and a comparison of the two clock states permits a measurement of the time of flight of the particle.

where $q^2 = \left(q_{\rm grav} + q_{\rm bend}\right)^2 + q_{\rm Sagnac}^2$ and $\tan\psi_0 = q_{\rm Sagnac}/\left(q_{\rm grav} + q_{\rm bend}\right)$. In practice, the Sagnac effect is only 2.5% of the gravitational effect but we see it leads to a global shift of the oscillatory pattern.

Let us now describe the result of the experiment. According to the previous considerations, the intensity observed at detector O_3 is

$$I(O_3) = \alpha\left\{1 + \cos\left[q\sin\left(\psi - \psi_0\right)\right]\right\}. \tag{1108}$$

By Fourier transforming this signal, one can extract the frequency $q_{\rm grav}$ of this oscillations. If one repeat this procedure for different wavelengths, one can fit the dependence of $q_{\rm grav}$ with λ which allows us to determine the product $m_{\rm grav}m_{\rm in}$. Let us recall that the neutron mass is usually obtained by mass spectroscopy on the deuteron which leads to the following value

$$m_{\rm n} = m_{\rm D} - m_{\rm p} + \frac{E_\gamma}{c^2} = 1.6747\times10^{-24}{\rm g}, \tag{1109}$$

where E_γ is deuteron binding energy. On the other hand, the COW experiment has found [133]

$$\left(m_{\rm grav}m_{\rm in}\right)^{1/2} = (1.675\pm0.003)\times10^{-24}\,{\rm g}. \tag{1110}$$

This shows that the weak equivalence principle seems to be satisfied even in a purely quantum mechanical situation. This also shows that gravity can be "coupled" to the Schrödinger equation in a standard way. The COW experiment is therefore very important since this was the first time that this was demonstrated. Moreover, if we consider that $m_{\rm n} = m_{\rm in}$ then the two previous equations shows that the COW experiment proves that $m_{\rm in} = m_{\rm grav}$ even in the quantum domain.

24.14.2 The Quantum Galileo Experiment

It is well-known that, classically, the motion of a body in a constant gravitational field only depends on the ratio $m_{\rm grav}/m_{\rm in}$ (which is measured to be unity). However, if one writes the Schrödinger equation for the same situation, namely

$$i\hbar\frac{\partial\Psi(t,z)}{\partial t} = -\frac{\hbar^2}{2m_{\rm in}}\frac{\partial^2\Psi(t,z)}{\partial z^2} + m_{\rm grav}gz\Psi(t,z), \tag{1111}$$

it is apparent that $m_{\rm in}$ and $m_{\rm grav}$ no longer cancel out. Therefore, it is interesting to study how the universality of the free fall is recovered in quantum mechanics. We investigate this point in the following sections.

The Salecker-Wigner-Peres Clock

In this section, we present a simple model of a quantum clock, the so-called "Salecker-Wigner-Peres clock" [136, 137, 138, 139, 140]. We introduce this system because we want to study how a quantum particle falls down in a constant gravitational field. One way to to do that is to measure the time of flight of the particle and to compare it with its classical analogue. Hence the question of how to define the time of flight becomes important. As is well-known, the measurement of time (or, to be more accurate, of time of arrival) is problematic in quantum mechanics [141, 142, 143, 144, 145]. In particular, there is no time operator because time is not a dynamical variable. One way out is to introduce a quantum system, the "clock", that is coupled to the particle and changes its state when the particle crosses the starting and the ending points along its "trajectory", see Fig. 164. Reading these two states then gives information about the time of flight. We now study how this general idea works in practice.

We assume that the clock has an odd number $N = 2j + 1$ of states. The Hamiltonian of the clock is given by

$$H_{\text{clock}} = \omega J = -i\hbar\omega \frac{\partial}{\partial\theta}, \tag{1112}$$

where the quantity θ can be viewed as describing the position of the clock's hand. The corresponding states of the clock are represented by the following wave-function

$$u_n(\theta) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{in\theta}, \tag{1113}$$

that are eigenfunction of H_{clock} since $H_{\text{clock}} u_n(\theta) = n\hbar\omega u_n(\theta)$. It is also easy to show that

$$\mathrm{e}^{-iH_{\text{clock}}t/\hbar} u_n(\theta) = u_n(\theta - \omega t), \tag{1114}$$

which gives the time evolution of the system.

It is also convenient to introduce another basis $v_s(\theta)$ defined by $(s = 0, \cdots, N-1)$

$$v_s(\theta) = \frac{1}{\sqrt{N}} \sum_{n=-j}^{n=j} \mathrm{e}^{-\frac{2i\pi sn}{N}} u_n(\theta) \tag{1115}$$

$$= \frac{1}{\sqrt{2\pi N}} \frac{\sin\left[N\left(\theta - 2\pi s/N\right)/2\right]}{\sin\left[\left(\theta - 2\pi s/N\right)/2\right]}. \tag{1116}$$

If N is large enough, these functions peak at $\theta = 2\pi s/N$ and this corresponds to the hand of the clock pointing to the s-th hour with an angle uncertainty of $\pm\pi/N$ (hence, as expected, if N is large, the accuracy is good). Moreover, using Eq. (1114), we have

$$\mathrm{e}^{-iH_{\text{clock}}t/\hbar} v_s(\theta) = v_s(\theta - \omega t). \tag{1117}$$

This means that if we evaluate the above equation for a time τ such that $\omega\tau = 2\pi/N$, then one obtains

$$\mathrm{e}^{-iH_{\text{clock}}\tau/\hbar} v_s(\theta) = v_{s+1\,(\mathrm{mod}\,N)}(\theta), \tag{1118}$$

which confirms the interpretation given before.

Let us now consider a free particle traveling along the x-axis. We want to measure the time of flight between $x = 0$ and $x = L$, see Fig. 164. In order for the clock to record this time, we must of course couple it to the particle. Following Ref. [137], we assume that the Hamiltonian of the total system "particle+clock" can be written as

$$H = \frac{p^2}{2m} + \mathcal{P}(x) H_{\text{clock}}, \tag{1119}$$

where $\mathcal{P}(x)$ is the projector operator which is one if $0 < x < L$ and zero otherwise and m is the mass of the particle. We write the eigen wave-functions of the system as

$$\Psi(t, x, \theta) = \psi(x, \theta)\, \mathrm{e}^{-iEt/\hbar}$$

$$= \frac{1}{\sqrt{N}} \sum_{n=-j}^{n=j} \psi_n(x) u_n(\theta) \mathrm{e}^{-iEt/\hbar}, \tag{1120}$$

and inserting this form into the Schrödinger equation, one obtains

$$\sum_{n=-j}^{n=j}\left[-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\psi_n}{\mathrm{d}x^2}+n\hbar\omega\mathcal{P}(x)\psi_n(x)\right.$$
$$\left.-E\psi_n(x)\right]u_n(\theta)=0. \tag{1121}$$

Therefore, we just have to solve the time-independent Schrödinger equation in presence of a rectangular barrier. Of course, this exercise is solved in any textbooks on quantum mechanics, see for instance Ref. [57]. It is straightforward to show that

$$\psi_n(x)=\begin{cases}A\mathrm{e}^{ikx}+B\mathrm{e}^{-ikx}, & x<0,\\ C\mathrm{e}^{ipx}+D\mathrm{e}^{-ipx}, & 0<x<L,\\ E\mathrm{e}^{ikx}, & x>L,\end{cases} \tag{1122}$$

where we have defined

$$k = \frac{\sqrt{2mE}}{\hbar}, \tag{1123}$$
$$p = \frac{\sqrt{2m(E-n\hbar\omega)}}{\hbar}=k\sqrt{1-\epsilon}, \tag{1124}$$

with $\epsilon\equiv n\hbar\omega/E$, a quantity which measures how the clock perturbs the particle. The quantities A, B, C, D and E are integration constants that are fixed by requiring the wave-function and its derivative to be continuous at the barrier. Notice that, after the barrier there is only one branch; this is of course because we assume that the particle is coming from the left, see Fig. 164. Matching ψ_n and $\mathrm{d}\psi_n/\mathrm{d}x$ at $x=0$ and $x=L$, one obtains

$$A = \frac{1}{4p}\left[(k+p)\mathrm{e}^{i(k-p)L}-(k-p)\mathrm{e}^{i(k+p)L}\right]E+\frac{1}{4k}\left[(k+p)\mathrm{e}^{i(k-p)L}+(k-p)\mathrm{e}^{i(k+p)L}\right]E, \tag{1125}$$
$$B = \frac{1}{4p}\left[(k+p)\mathrm{e}^{i(k-p)L}-(k-p)\mathrm{e}^{i(k+p)L}\right]E-\frac{1}{4k}\left[(k+p)\mathrm{e}^{i(k-p)L}+(k-p)\mathrm{e}^{i(k+p)L}\right]E. \tag{1126}$$

Then, we assume that the clock does not perturb the particle too much and, as a consequence, that $\epsilon\ll 1$. In this case, $B\simeq 0$ and $A\simeq \mathrm{e}^{i(k-p)L}E$. Indeed, we have $p=k\sqrt{1-\epsilon}$ which, at leading order, simply gives $k\simeq p$ (but, of course, one should use the next to leading order when one estimates the argument of the exponentials, see below). Then, initially, this means that the wave-function can be written as, see Eq. (1120)

$$\psi(x,\theta)=\frac{1}{\sqrt{N}}\sum_{n=-j}^{n=j}A\mathrm{e}^{ikx}u_n(\theta)=A\mathrm{e}^{ikx}v_0(\theta),\quad x\ll 0, \tag{1127}$$

because k does not depend on n, see Eq. (1123). We see that, initially, the clock points to zeroth hour, as expected. After the barrier, using the previous considerations, the expression of the wave-function can be written as, see Eq. (1120)

$$\psi(x,\theta)\simeq\frac{A}{\sqrt{N}}\mathrm{e}^{ikx}\sum_{n=-j}^{n=j}\mathrm{e}^{-i(k-p)L}u_n(\theta),\quad x\gg L, \tag{1128}$$

where we have used Eq. (1115). But we have

$$p\simeq k\left(1-\frac{\epsilon}{2}\right)\simeq k-\frac{n\omega}{(2E/m)^{1/2}}, \tag{1129}$$

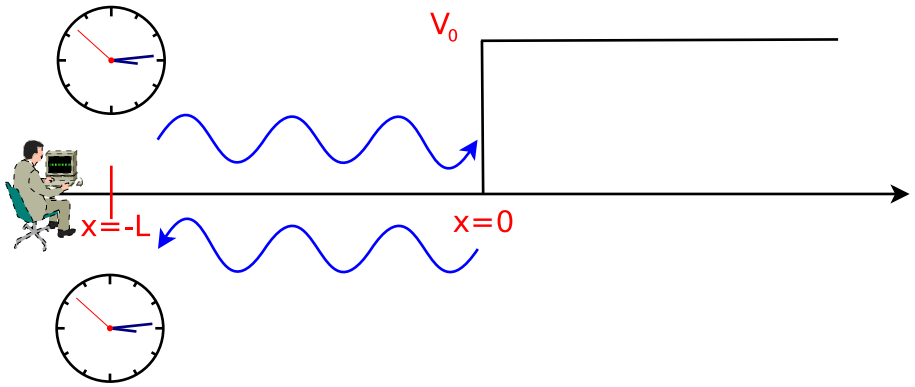


Figure 165: Measurement of an out-and-back time of flight. An observer, located at $x = -L$ sends a particle towards a square potential step located at $x = 0$. The physicist turns on the clock when the particle leaves and turn it off when the reflected wave (*i.e.* the reflected particle) comes back and reaches $x = -L$.

which implies that (for $x \gg L$)

$$\psi(x,\theta) = \frac{A}{\sqrt{2\pi N}} \mathrm{e}^{ikx} \sum_{n=-j}^{n=j} \mathrm{e}^{in\theta - in\omega L/(2E/m)^{1/2}} \tag{1130}$$

$$= A\mathrm{e}^{ikx} v_0 \left(\theta - \frac{\omega L}{\sqrt{2E/m}}\right). \tag{1131}$$

This means, see Eq. (1117), that the clock now indicates the time

$$t = \frac{L}{(2E/m)^{1/2}} = \frac{L}{v_{\rm clas}}, \tag{1132}$$

where $v_{\rm clas} \equiv (2E/m)^{1/2}$ is the classical velocity. Therefore, the clock has measured a time of flight which is nothing but the classical time of flight $L/v_{\rm clas}$ [137].

We now consider the situation represented in Fig. 165, see Ref. [140]. An observer, located at $x = -L$, sends a particle towards a reflecting barrier of height V_0 located at $x = 0$. When the particle leaves the Peres clock is turned on. Then the particle bounces back and when it comes back at the observer position, the Peres clock is turned off. With this experimental set up, one can measure the out-and-back time of flight. Following the same approach as before, the wave-function can be written as

$$\psi_n(x) = \begin{cases} A\mathrm{e}^{ikx} + B\mathrm{e}^{-ikx}, & x < 0, \\ C\mathrm{e}^{-px}, & x > 0, \end{cases} \tag{1133}$$

where we now have

$$k = \frac{\sqrt{2m(E - n\hbar\omega)}}{\hbar} = \frac{\sqrt{2mE}}{\hbar}\sqrt{1-\epsilon} \tag{1134}$$

$$p = \frac{\sqrt{2m(V_0 + n\hbar\omega - E)}}{\hbar}$$

$$= \frac{\sqrt{2m(V_0 - E)}}{\hbar}\left(1 + \frac{E}{V_0 - E}\epsilon\right)^{1/2}. \tag{1135}$$

Then, we have to match the wave-function and its derivative at $x = 0$. This gives

$$B = \frac{k - ip}{k + ip} A \equiv RA, \tag{1136}$$

$$C = \frac{2k}{k + ip} A. \tag{1137}$$

As before, we assume that, initially, the clock was is a given state, here $v_0(\theta)$. Therefore, the wave-function in the region $x < 0$ can be expressed as, see Eq. (1120)

$$\psi(x, \theta) = \frac{A}{\sqrt{N}} \sum_{n=-j}^{n=j} \left(\mathrm{e}^{ikx} + R\mathrm{e}^{-ikx} \right) u_n(\theta). \tag{1138}$$

But $|R| = 1$ and its phase is given by $-2 \arctan(p/k)$. As a consequence, one has

$$\begin{aligned} \psi(x = -L, \theta) &= \frac{A}{\sqrt{N}} \sum_{n=-j}^{n=j} \Big[\mathrm{e}^{-ikL} \\ &+ \mathrm{e}^{-2i \arctan(p/k)} \mathrm{e}^{ikL} \Big] u_n(\theta). \end{aligned} \tag{1139}$$

The next step is similar to the case of a free particle, namely we expand the phases of the two wave-function branches in terms of the parameter ϵ. This gives

$$\begin{aligned} \psi(x = -L, \theta) &= \frac{A}{\sqrt{2\pi N}} \sum_{n=-j}^{n=j} \exp\left[-\frac{iL}{\hbar}\sqrt{2mE} + in\left(\theta + \frac{\omega L}{v_{\mathrm{clas}}}\right)\right] + \exp\left[-2i \arctan\left(\sqrt{\frac{V_0 - E}{E}}\right) + \frac{iL}{\hbar}\sqrt{2mE} \right. \\ &+ \left. in\left(\theta - \frac{\omega L}{v_{\mathrm{clas}}} - \frac{\hbar\omega}{\sqrt{E(V_0 - E)}}\right)\right] \end{aligned} \tag{1140}$$

$$\begin{aligned} &= \frac{A}{\sqrt{N}} \exp\left(-\frac{iL}{\hbar}\sqrt{2mE}\right) v_0\left(\theta + \frac{\omega L}{v_{\mathrm{clas}}}\right) + \frac{A}{\sqrt{N}} \exp\left[-2i \arctan\left(\sqrt{\frac{V_0 - E}{E}}\right) + \frac{iL}{\hbar}\sqrt{2mE}\right] \\ &\times v_0\left(\theta - \frac{\omega L}{v_{\mathrm{clas}}} - \frac{\hbar\omega}{\sqrt{E(V_0 - E)}}\right), \end{aligned} \tag{1141}$$

and this allows us to directly read the time indicated by the clock. For the first branch, it is $-L/v_{\mathrm{clas}}$ while for the second one, it is given by $L/v_{\mathrm{clas}} + \hbar/\sqrt{E(V_0 - E)}$. As a consequence, the time of flight of the particle can be expressed as

$$\Delta t = \frac{2L}{v_{\mathrm{clas}}} + \frac{\hbar}{\sqrt{E(V_0 - E)}}. \tag{1142}$$

Therefore, the time of flight is the classical one, $2L/v_{\mathrm{clas}}$, but, this time, there is an additional contribution. This one can be understood as follows. The quantity Δt can be re-written as [140]

$$\Delta t = \frac{2(L + d)}{v_{\mathrm{clas}}}, \tag{1143}$$

where $d = 1/p$, see Eq. (1135) (evaluated at $\epsilon = 0$) is the penetration depth into the potential step. This means that there is an additional delay due to the tunnel effect and the fact that the particle has a non vanishing probability to be below the barrier [140]. In this regime, the "velocity" is nevertheless given by its classical value v_{clas} [140].

This concludes this section on the Peres clock. We have seen that this is a useful device to measure times of flight. We now use it to study the motion of a quantum particle in an uniform gravitational field.

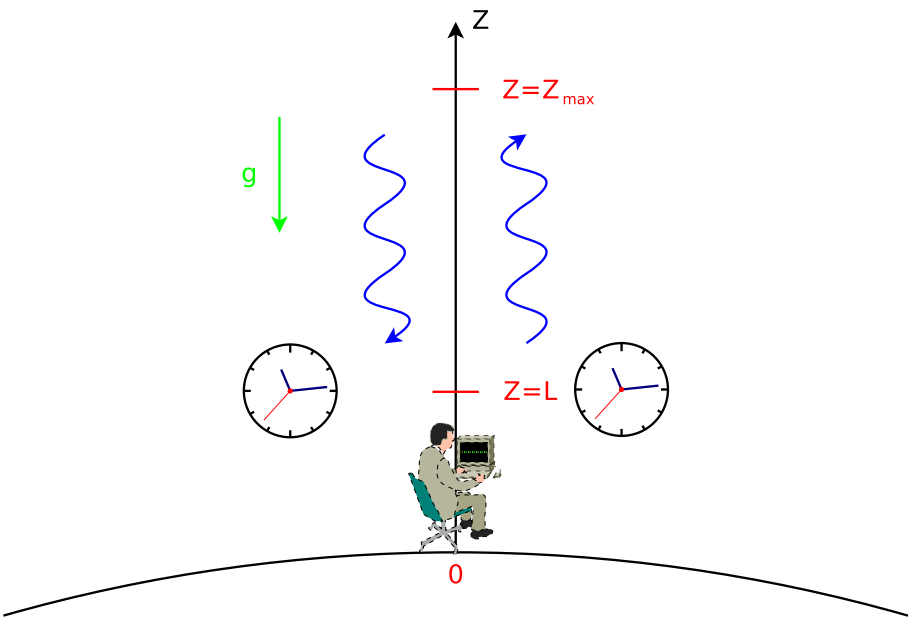


Figure 166: Sketch of a quantum particle in an uniform gravitational field. The particle is sent upward, the initial time being recorded by a quantum Peres clock. The particle reaches a maximum altitude and then falls back to the detector at $z = L$ where the Peres clock measures this return time.

The Gravitational Case

In this section, we study the status of the weak equivalence principle and of the universality of the free fall in quantum mechanics [146, 140, 139, 147, 148, 149, 150].

Following Ref. [140], let us consider the situation where a particle is sent upwards in a uniform gravitational field, see Fig. 166. If z denotes the vertical coordinate, the corresponding classical situation is described by the following Lagrangian

$$L(\dot{z}, z) = \frac{m_{\rm in}}{2}\dot{z}^2 - m_{\rm grav} g z, \tag{1144}$$

where g is the gravitational field and where we have made the difference between the inertial mass $m_{\rm in}$ and the gravitational mass (the "gravitational charge") $m_{\rm g}$. The solution to the classical equation of motion reads

$$z(t) = -\frac{1}{2}\frac{m_{\rm grav}}{m_{\rm in}} g t^2 + v_{\rm ini} t + L, \tag{1145}$$

where $z = L$ is the initial altitude (at $t = 0$) and $v_{\rm ini}$ is the initial velocity, related to the energy E of the particle by the following expression

$$v_{\rm ini} = \sqrt{2g\frac{m_{\rm grav}}{m_{\rm in}}}\left(\frac{E}{m_{\rm grav} g} - L\right)^{1/2}. \tag{1146}$$

Of course, the motion of the particle only depends on the ratio $m_{\rm grav}/m_{\rm in}$. The maximum altitude, $z_{\rm max}$, reached by the particle, see Fig. 166, is given by

$$z_{\rm max} = \frac{E}{m_{\rm grav} g}, \tag{1147}$$

which means that, from its initial position to $z_{\rm max}$, the particle can rise a distance of $E/(m_{\rm grav} g) - L$. After having reached $z_{\rm max}$, the particle falls back and the up-to-down classical time of flight reads

$$\Delta t = 2\sqrt{\frac{2}{g}}\left(\frac{m_{\rm in}}{m_{\rm g}}\right)^{1/2}\left(\frac{E}{m_{\rm grav} g} - L\right)^{1/2}. \tag{1148}$$

If $m_{\rm grav} = m_{\rm in}$, this equation can also be written as $\Delta t = 2v_{\rm ini}/g$.

Then, we would like to consider the same situation but from a quantum-mechanical point of view [140]. In particular, we would like to measure the quantum time of flight. Given the discussion of the previous section, it is clear that we must couple the quantum particle to a Peres clock located at $z = L$. Then, the next step is, of course, to solve the Schrödinger equation. Following the previous considerations, we just have to consider the equivalent of Eq. (1121), but for the case of an uniform gravitational field. It reads

$$\frac{{\rm d}^2\psi_n}{{\rm d}z^2} + \left[\frac{2m_{\rm in}\left(E - n\hbar\omega\right)}{\hbar^2} - \frac{2m_{\rm in}m_{\rm grav}g}{\hbar^2}z\right]\psi_n = 0. \tag{1149}$$

It is straightforward to obtain that the solution can be expressed in terms of Airy functions [48, 49] ${\rm Ai}(z)$ and ${\rm Bi}(z)$. If we retain only the branch that leads to a bounded wave-function, the result is given by

$$\psi(z,\theta) = \frac{A}{\sqrt{N}}\sum_{n=-j}^{n=j}{\rm Ai}\left(\frac{z-b}{a}\right)u_n(\theta), \tag{1150}$$

where A is constant fixed by the normalization of the wave-function and a and b are defined by

$$a = \left(\frac{\hbar^2}{2m_{\rm in}m_{\rm grav}g}\right)^{1/3}, \tag{1151}$$

$$b = \frac{E - n\hbar\omega}{m_{\rm grav}g}. \tag{1152}$$

The next step consists in evaluating the wave-function at $x = L$, with the assumption that the return point is located far from the maximum altitude. This means that $L - b$ is a negative quantity and that $|L-b|$ is large [more precisely, one has $L-b = L-E/(m_{\rm grav}g)+E\epsilon/(m_{\rm grav}g)$ and the quantity $L - E/(m_{\rm grav}g)$ is minus the distance the particle can rise, see the remark after Eq. (1147). Therefore, $L - b < 0$ and, if this distance is large, $|L - b|$ is indeed a large quantity]. In this limit, one has

$$\psi(L,\theta) \simeq \frac{A}{2i\sqrt{\pi N}}\sum_{n=-j}^{n=j}\left(\frac{z-b}{a}\right)^{-1/4} \times\left({\rm e}^{i\xi+i\pi/4} - {\rm e}^{-i\xi-i\pi/4}\right)u_n(\theta), \tag{1153}$$

where $\xi \equiv (2/3)[(b - L)/a]^{3/2}$. Explicitly, this quantity can be expressed as

$$\xi = \frac{2}{3a^{3/2}}\left[\frac{E}{m_{\rm grav}g}(1-\epsilon) - L\right]^{3/2} \tag{1154}$$

$$\simeq \frac{2}{3a^{3/2}}\left(\frac{E}{m_{\rm grav}g} - L\right)^{3/2} - \frac{1}{a^{3/2}}\left(\frac{E}{m_{\rm grav}g} - L\right)^{1/2}\frac{n\hbar\omega}{m_{\rm grav}g}. \tag{1155}$$

In order to calculate the difference between the two times indicated by the clock and corresponding to the two branches of the wave-function, it is in fact sufficient to estimate the part of the phase shift that depends on n. Therefore, it is clear that the phase shift $\pm\pi/4$ in the expression (1153) will not contribute because it is n-independent. Moreover, the term $[(z - b)/a]^{-1/4}$ is n-dependent but is common to the two branches of the wave-functions. As a consequence, it will cancel out when the phase shift is determined (but would participate to the "absolute" time

indicated by the clock when the particle leaves $z = L$ and returns to $z = L$). Summarizing, the up-and down time of flight is given by

$$\Delta t = \frac{2}{a^{3/2}}\left(\frac{E}{m_{\rm grav}g} - L\right)^{1/2}\frac{\hbar}{m_{\rm grav}g} \tag{1156}$$

$$= 2\sqrt{\frac{2}{g}}\left(\frac{m_{\rm in}}{m_{\rm grav}}\right)^{1/2}\left(\frac{E}{m_{\rm grav}g} - L\right)^{1/2}, \tag{1157}$$

that is to say exactly Eq. (1148), see Ref. [140]. Several remarks are in order here. Firstly, although it is of quantum-mechanical origin, the above equation does not contain $\hbar$ that cancels out. Secondly, this result seems to indicate that the weak equivalence principle holds at the quantum level. Thirdly, it may appear surprising that the quantum Δt is not corrected by a term corresponding to the penetration depth, see Sec. 24.14.2, in particular Eq. (1143), as it is the case for a square potential. The interpretation given in Ref. [140] is that there is a finite probability that the particle tunnels below the barrier but there is also a finite probability that the particle is scattered back before the classical turning point. If these two probabilities cancel then one obtains the classical prediction. Fourthly, the above result is valid only if the return time is measured far from the maximum altitude. What happens if, on the contrary, the particle starts from a position which is close to the turning point? To investigate this case, we re-start from Eq. (1150) and write the solution as

$$\psi(z,\theta) = \frac{A}{\sqrt{N}}\sum_{n=-j}^{n=j}\frac{\sqrt{y}}{3}\left\{\left[{\rm e}^{i\pi/3}J_{1/3}\left(\zeta\right) + {\rm e}^{-i\pi/3}J_{-1/3}\left(\zeta\right)\right] + \left[\left(1-{\rm e}^{i\pi/3}\right)J_{1/3}\left(\zeta\right) + \left(1-{\rm e}^{-i\pi/3}\right)J_{-1/3}\left(\zeta\right)\right]\right\}u_n(\theta), \tag{1158}$$

where $y \equiv (b-z)/a > 0$ and $\zeta \equiv 2y^{3/2}/3$. In this way, we have identified the incident (first term in the squared bracket) and the reflected wave (second term in the square bracket). Indeed, using the asymptotic behavior of the Bessel functions at infinity, one can check that each branches give precisely ${\rm e}^{\pm(i\xi+i\pi/4)}$ if $z = L$, see Eq. (1153). If we now consider the limit $y \ll 1$, then one arrives at

$$\psi(z,\theta) = \frac{A}{\sqrt{N}}\sum_{n=-j}^{n=j}\left\{\left[\frac{3^{-2/3}}{2\Gamma(2/3)} + \frac{3^{-4/3}}{2\Gamma(4/3)}y\right] + i\sqrt{3}\left[-\frac{3^{-2/3}}{2\Gamma(2/3)} + \frac{3^{-4/3}}{2\Gamma(4/3)}y\right] + \left[\frac{3^{-2/3}}{2\Gamma(2/3)} + \frac{3^{-4/3}}{2\Gamma(4/3)}y\right]\right.$$
$$\left. -i\sqrt{3}\left[-\frac{3^{-2/3}}{2\Gamma(2/3)} + \frac{3^{-4/3}}{2\Gamma(4/3)}y\right] + \cdots\right\}u_n(\theta), \tag{1159}$$

the two first terms corresponding to the incident wave and the two last ones to the reflected wave. The two branches are of course complex conjugate to each other. Therefore, if we denotes by $\aleph$ the complex number corresponding to the two first terms in the above equation, then the wave-function takes the form

$$\psi(z,\theta) \simeq \frac{A}{\sqrt{N}}\sum_{n=-j}^{n=j}|\aleph|\left({\rm e}^{i\arg\aleph} + {\rm e}^{-i\arg\aleph}\right)u_n(\theta), \tag{1160}$$

where

$$\arg\aleph = \arctan\left[\sqrt{3}\frac{3^{2/3}\Gamma(2/3)y - 3^{4/3}\Gamma(4/3)}{3^{2/3}\Gamma(2/3)y + 3^{4/3}\Gamma(4/3)}\right]. \tag{1161}$$

The rest of the calculation proceeds as before. We first expand the phase in ϵ and, then, take the limit $E/(m_{\rm g}g) - L \to 0$. This leads to

$$\arg\aleph \simeq -\frac{\pi}{3} - \frac{3^{5/6}}{6}\frac{\Gamma(2/3)}{\Gamma(4/3)}\frac{\hbar}{m_{\rm grav}ga}n\omega. \tag{1162}$$

As a consequence, one finds that the time of flight is given by

$$\Delta t \simeq 2\frac{3^{5/6}}{6}\frac{\Gamma(2/3)}{\Gamma(4/3)}2^{1/3}\left(\frac{m_{\rm in}}{m_{\rm grav}}\right)^{1/3}\left(\frac{\hbar}{m_{\rm grav}g^2}\right)^{1/3} \tag{1163}$$

$$\simeq 1.59\times\left(\frac{m_{\rm in}}{m_{\rm grav}}\right)^{1/3}\left(\frac{\hbar}{m_{\rm grav}g^2}\right)^{1/3}, \tag{1164}$$

in rough agreement with Ref. [140]. This time, we notice that $\hbar$ has not canceled in the final expression and that the time of flight is different from its classical counterpart (1148). This also means, as already mentioned, that the cancellation which leads to a quantum time of flight equals to the classical one is in fact valid only if the particle starts far from the turning point.

We conclude that the weak equivalence principle (or the universality of the free fall) can be extended to the quantum regime. The linear potential seems to possess precisely the shape which leads to identical times of flight. It is important to realize that this property depends on the shape of the potential and would not be obtained with another shape, see Ref. [140]. Moreover, this seems to be true only far from the turning point. If one identifies the wave-vector to $k = (2m_{\rm in}m_{\rm grav}g/\hbar^2)^{1/3}$, see Ref. [151], the de Broglie relation $v \simeq \hbar k/m$ allows us to estimate a velocity. Then, from the quantum time of flight, see Eq. (1164), we can construct a length $\simeq [\hbar^2/(m^2 g)]^{1/3}$ which typically controls the validity of the approximation just mentioned.

Atom Trampoline

In this section, we want to briefly mention the case of the "atom trampoline" [150, 151]. Let us consider again the experimental set up of the previous section but let us now assume that a reflecting wall has been installed on the ground, at $z = 0$. Then, the boundary condition $\psi_n(0) = 0$ leads to $\mathrm{Ai}(-b/a) = 0$. If z_n denotes the zero of the Airy function, this means that we now have a discrete spectrum of energy levels given by

$$E_n = -z_n \left(\frac{\hbar^2 m_{\rm grav}^2 g^2}{2m_{\rm in}}\right)^{1/3}. \tag{1165}$$

It is interesting to notice that this result depends on the ratio $m_{\rm grav}^2/m_{\rm in}$ and not on $m_{\rm grav}/m_{\rm in}$. Moreover, this spectrum has been observed (with ultra-cold neutrons which leads to E_n of the order of $\sim 10^{-12}\,$eV) in the gravitational field of the Earth [152, 153]. Therefore, this result confirms the discussion about the COW experiment described in Sec. 24.14.1. It experimentally establishes that the gravitational force "couples" to the Schrödinger equation in a standard way.

Schrödinger Equation in an Accelerated Frame

It is also frequent to refer to the weak equivalence principle as the property stating that, locally, the effect of a constant gravitational field can be mimicked by an accelerating frame. Therefore, it is interesting to study whether this claim holds in quantum mechanics. This question has been studied in Refs. [146, 150, 151]. Here, we follow the treatment of Ref. [146].

Let us consider the free Schröndinger equation (in one dimension to simplify the problem). It reads

$$i\hbar\frac{\partial\Psi(t,z)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi(t,z)}{\partial z^2}. \tag{1166}$$

Then, let us consider the following coordinate transformation

$$z' = z - vt - \frac{1}{2}gt^2 \tag{1167}$$

$$t' = t. \tag{1168}$$

It is straightforward to show that the function $\varphi(t',z')$ defined according to

$$\begin{aligned}\Psi(t,z) &= \varphi\left[t'(t), z'(t,z)\right] \\ &\quad \times \exp\left[\frac{imv}{\hbar}\left(z' + \frac{vt'}{2}\right)\right. \\ &\quad \left. + \frac{imgt'}{\hbar}\left(z' + \frac{vt'}{2} + \frac{gt'^2}{6}\right)\right],\end{aligned} \tag{1169}$$

satisfies the equation [146]

$$i\hbar\frac{\partial\varphi(t',z')}{\partial t'} = -\frac{\hbar^2}{2m}\frac{\partial^2\varphi(t',z')}{\partial z'^2} + mgz'\varphi(t',z'). \tag{1170}$$

The potential in the above equation is precisely of the form describing a constant gravitational field. In this sense, we find that the equivalence between a constant gravitational field and an accelerated frame propagates to quantum mechanics.

Falling Composite Quantum Objects

In the previous sections, we have studied how a quantum particle falls in a uniform gravitational field. For instance, we have investigated under which circumstances one can say that the weak equivalence principle holds in quantum mechanics. Here, we would like to study the same question but in the case of a composite quantum object. In particular, it is interesting to ask whether the motion of the composite system (the atom) can depend on its internal state.

Let us start by recalling the Ehrenfest theorem [57] for a quantum particle. It states that the time derivatives of the mean position and momentum are given by (for simplicity, we consider the case of a one-dimensional system since we have in mind a particle falling down along the z-axis)

$$\frac{\mathrm{d}\langle z\rangle}{\mathrm{d}t} = \frac{\langle p_z\rangle}{m_{\mathrm{in}}}, \tag{1171}$$

$$\frac{\mathrm{d}\langle p_z\rangle}{\mathrm{d}t} = -\left\langle\frac{\partial V}{\partial z}\right\rangle. \tag{1172}$$

In the case of a linear potential, $V(z) = m_{\mathrm{grav}}gz$, the two previous equations expressing the Ehrenfest theorem can be combined in such a way that we arrive at

$$\frac{\mathrm{d}^2\langle z\rangle}{\mathrm{d}t^2} = -\frac{m_{\mathrm{grav}}}{m_{\mathrm{in}}}g. \tag{1173}$$

This means that the average position $\langle z\rangle(t)$ evolves exactly as the classical z given in Eq. (1145). However, let us also notice that the variance of the position is not a function of the ratio $m_{\mathrm{grav}}/m_{\mathrm{in}}$ only, see in particular Refs. [146, 147, 148].

After this short reminder, let us consider a bound system of two particles interacting through a potential V which only depends on the distance between them. We also assume that this system is embedded in a region where there is an exterior field described by the potential $U(\boldsymbol{r})$. The corresponding Lagrangian reads

$$\begin{aligned} L\left(\dot{\boldsymbol{r}}_1,\dot{\boldsymbol{r}}_2,\boldsymbol{r}_1,\boldsymbol{r}_2\right) &= \frac{m_{\mathrm{in},1}}{2}\dot{\boldsymbol{r}}_1^2+\frac{m_{\mathrm{in},2}}{2}\dot{\boldsymbol{r}}_2^2-V\left(\boldsymbol{r}_1-\boldsymbol{r}_2\right)\\ &\quad -U\left(\boldsymbol{r}_1\right)-U\left(\boldsymbol{r}_2\right), \end{aligned} \tag{1174}$$

where $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are the positions of the two particles and $m_{\mathrm{in},1}$, $m_{\mathrm{in},2}$ their mass. If we now introduce the center of mass position vector,

$$\boldsymbol{r}_{\mathrm{G}} = \frac{m_{\mathrm{in},1}}{m_{\mathrm{in},1}+m_{\mathrm{in},2}}\boldsymbol{r}_1+\frac{m_{\mathrm{in},2}}{m_{\mathrm{in},1}+m_{\mathrm{in},2}}\boldsymbol{r}_2, \tag{1175}$$

and the relative position vector,

$$\boldsymbol{r} = \boldsymbol{r}_1-\boldsymbol{r}_2, \tag{1176}$$

then Eq. (1174) takes the form

$$\begin{aligned} L\left(\dot{\boldsymbol{r}}_{\mathrm{G}},\dot{\boldsymbol{r}},\boldsymbol{r}_{\mathrm{G}},\boldsymbol{r}\right) &= \frac{M}{2}\dot{\boldsymbol{r}}_{\mathrm{G}}^2+\frac{\mu}{2}\boldsymbol{r}^2-V(\boldsymbol{r})\\ &\quad -U(\boldsymbol{r}_1)-U(\boldsymbol{r}_2), \end{aligned} \tag{1177}$$

where $M\equiv m_{\mathrm{in},1}+m_{\mathrm{in},2}$ is the total mass and $\mu\equiv m_{\mathrm{in},1}m_{\mathrm{in},2}/(m_{\mathrm{in},1}+m_{\mathrm{in},2})$ is the reduced one. Of course, in the above expression $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ must be viewed as function of $\boldsymbol{r}_{\mathrm{G}}$ and $\boldsymbol{r}$[130].

[130]The relation between these quantities is given by

$$\boldsymbol{r}_1 = \boldsymbol{r}_{\mathrm{G}}+\frac{m_{\mathrm{in},2}}{M}\boldsymbol{r}, \tag{1178}$$

$$\boldsymbol{r}_2 = \boldsymbol{r}_{\mathrm{G}}-\frac{m_{\mathrm{in},1}}{M}\boldsymbol{r}. \tag{1179}$$

From the above considerations, it is straightforward to obtain the Hamiltonian of the system. We arrive at the following expression

$$H = \frac{\boldsymbol{p}_{\mathrm{G}}^2}{2M} + \frac{\boldsymbol{p}^2}{2\mu} + V(\boldsymbol{r}) + U(\boldsymbol{r}_1) + U(\boldsymbol{r}_2), \tag{1180}$$

where $\boldsymbol{p}_{\mathrm{G}} \equiv \boldsymbol{p}_1 + \boldsymbol{p}_2$ and $\boldsymbol{p} = (\mu/m_{\mathrm{in},1})\boldsymbol{p}_1 - (\mu/m_{\mathrm{in},2})\boldsymbol{p}_2$. If we now assume that the external field is a uniform gravitational field (directed along the z-axis), then it is easy to show that

$$\begin{aligned} U(\boldsymbol{r}_1) + U(\boldsymbol{r}_2) &= (m_{\mathrm{grav},1} + m_{\mathrm{grav},2})\, g z_{\mathrm{G}} \\ &\quad + \frac{1}{M}\left(m_{\mathrm{grav},1} m_{\mathrm{in},2} - m_{\mathrm{grav},2} m_{\mathrm{in},1}\right) g z \qquad (1181) \\ &\equiv \overline{U}(z_{\mathrm{G}}) + \tilde{U}(z), \qquad (1182) \end{aligned}$$

where, according to Eq. (1176), $z \equiv z_1 - z_2$. As a consequence, the Hamiltonian (1180) takes the form

$$H = H_{\mathrm{G}} + H_{\mathrm{rel}}, \tag{1183}$$

where the center of mass and relative Hamiltonians can be expressed as

$$H_{\mathrm{G}} = \frac{\boldsymbol{p}_{\mathrm{G}}^2}{2M} + \overline{U}(z_{\mathrm{G}}) \tag{1184}$$

$$H_{\mathrm{rel}} = \frac{\boldsymbol{p}^2}{2\mu} + V(\boldsymbol{r}) + \tilde{U}(z). \tag{1185}$$

The physical interpretation of these equations is clear. The center of mass behaves as a free particle embedded in a uniform gravitational field. On the other hand, the internal states of the composite system are determined by the potential $V(\boldsymbol{r})$ but are also affected by the gravitational field through the term $\tilde{U}(z)$. In practice, one expects that $\tilde{U}(z) \ll V(\boldsymbol{r})$ and, therefore, the perturbations of the energy levels will be either very small or even totally negligible. Moreover, because $[z, p_{\mathrm{G},z}] = [z_{\mathrm{G}}, p_z] = 0$, the two above Hamiltonians commute

$$[H_{\mathrm{G}}, H_{\mathrm{rel}}] = 0, \tag{1186}$$

which implies that the total state space is in fact the tensorial product of the center of mass state space and of the relative state space.

Let us now study the behavior of $\langle z_{\mathrm{G}} \rangle$. One has

$$\begin{aligned} \frac{\mathrm{d}\langle z_{\mathrm{G}} \rangle}{\mathrm{d}t} &= \frac{1}{i\hbar}\langle [z_{\mathrm{G}}, H_{\mathrm{G}} + H_{\mathrm{rel}}] \rangle = \frac{1}{i\hbar}\langle [z_{\mathrm{G}}, H_{\mathrm{G}}] \rangle \\ &= \frac{\langle p_{\mathrm{G},z} \rangle}{M}. \qquad (1187) \end{aligned}$$

In the same manner, one can now calculate the evolution of $\langle p_{\mathrm{G},z} \rangle$. One obtains

$$\begin{aligned} \frac{\mathrm{d}\langle p_{\mathrm{G},z} \rangle}{\mathrm{d}t} &= \frac{1}{i\hbar}\langle [p_{\mathrm{G},z}, H_{\mathrm{G}} + H_{\mathrm{rel}}] \rangle = \frac{1}{i\hbar}\langle [p_{\mathrm{G},z}, H_{\mathrm{G}}] \rangle \\ &= \frac{1}{i\hbar}\left\langle \left[p_{\mathrm{G},z}, \overline{U}(z_{\mathrm{G}})\right] \right\rangle \\ &= -\left(m_{\mathrm{grav},1} + m_{\mathrm{grav},2}\right) g. \qquad (1188) \end{aligned}$$

Therefore, combining Eqs. (1187) and (1188), we arrive at the following expression

$$\frac{\mathrm{d}^2\langle z_{\mathrm{G}} \rangle}{\mathrm{d}t^2} = -\frac{m_{\mathrm{grav},1} + m_{\mathrm{grav},2}}{m_{\mathrm{in},1} + m_{\mathrm{in},2}} g. \tag{1189}$$

Of course, this equation is similar to Eq. (1173) and this implies that the center of mass mean value of the quantum composite system will follow the classical trajectory. In particular, this

means that $\langle z_{\rm G}\rangle$ does not depend in which internal energy state the system is placed. Therefore, if, for instance, we consider a falling Hydrogen atom, the behavior of the position mean value of the atom will not depend of whether the atom is in the state, say, $1s$ or $2p$. We also notice that the coefficient in front of g in Eq. (1189) is the ratio of the total gravitational mass to the total inertial mass. In this sense, the weak equivalence principle is also satisfied by the composite quantum systems.

Finally, it is also interesting to repeat the gedanken experiment in Fig. 166 but with a composite system. It seems reasonable to couple the Peres clock to the center of mass of the system which means that the Hamiltonian of the system is still given by

$$H = H_{\rm G} + H_{\rm rel}, \tag{1190}$$

where, now, the center of mass and relative Hamiltonians can be expressed as

$$H_{\rm G} = \frac{\boldsymbol{p}_{\rm G}^2}{2M} + \overline{U}(z_{\rm G}) + \mathcal{P}(z_{\rm G})H_{\rm clock} \tag{1191}$$

$$H_{\rm rel} = \frac{\boldsymbol{p}^2}{2\mu} + V(\boldsymbol{r}) + \tilde{U}(z). \tag{1192}$$

Then, following Eq. (1120), we write the wave-function of the system as

$$\Psi\left(t, \boldsymbol{r}_{\rm G}, \boldsymbol{r}, \theta\right) = \frac{1}{\sqrt{N}} \sum_{n=-j}^{n=j} \psi_n(\boldsymbol{r}_{\rm G})\chi_n(\boldsymbol{r})u_n(\theta)\mathrm{e}^{-iEt/\hbar}. \tag{1193}$$

The next step consists in inserting this wave-function in the Schrödinger equation in order to obtain the equivalent of Eq. (1121). The result can be written as

$$\sum_{n=-j}^{n=j}\Bigg\{\chi_n(\boldsymbol{r})\left[-\frac{\hbar^2}{2M}\Delta_{\rm G}\psi_n(\boldsymbol{r}_{\rm G}) + \overline{U}(z_{\rm G})\psi_n(\boldsymbol{r}_{\rm G}) + n\hbar\omega\mathcal{P}(z_{\rm G})\psi_n(\boldsymbol{r}_{\rm G}) - E_{\rm G}\psi_n(\boldsymbol{r}_{\rm G})\right] + \psi_n(\boldsymbol{r}_{\rm G})\left[-\frac{\hbar^2}{2m}\Delta\chi_n(\boldsymbol{r}) + \tilde{U}(z)\chi_n(\boldsymbol{r}) - E_{\rm rel}\chi_n(\boldsymbol{r})\right]\Bigg\}u_n(\theta) = 0, \tag{1194}$$

where $E = E_{\rm G} + E_{\rm rel}$. This implies that only the center of mass part, $\psi_n(\boldsymbol{r}_{\rm G})$, of the total wave-function will depend on n. As a consequence, the internal part should in fact be written as $\chi(\boldsymbol{r})$ instead of $\chi_n(\boldsymbol{r})$. This means that $\chi(\boldsymbol{r})$ cannot cause a rotation of the Peres clock. Therefore, the time indicated by the clock will not depend on the internal state of the composite system. We conclude that the time of flight of an atom, defined and calculated in terms of the Peres clock, does not depend in which quantum state this atom is placed.

This concludes the section on the weak equivalence principle in quantum mechanics. The result is mixed. We have shown that, very often, it is satisfied in the framework of quantum mechanics. But, on the other hand, it seems sometimes to be modified by quantum effects, see in particular Eq. (1164). Maybe this indicates that its application to the more complicated case of quantum vacuum fluctuations of a field is rather suspicious?

24.15 Conclusions

We are now in a position where we can really formulate the cosmological constant problem. Vacuum fluctuations seem to exist in Nature and to have normal gravitational properties. Therefore, it seems natural to postulate that they participate in the value of the cosmological constant. However, when one tries to use well controlled techniques of quantum field theory to compute their energy density, one obtains a number which seems to be in contradiction with the measurement of Λ that one obtains in cosmology and with the constraints that one deduces from (for instance) the motion of the planets in our solar system. The difficulty of the problem stems from the fact that, in order to remove the above contradiction, one necessarily has to abandon something that is considered as robust, *i.e.* renormalization in quantum field theory, weak equivalence principle, high accuracy measurements of the expansion of the universe, etc ....

Even if this is not the purpose of this article, let us conclude this review with a few words about the solutions that have been proposed to solve the cosmological constant problem, for a complete overview of the subject, see for instance Ref. [154]. Of course, the most obvious solution is super-symmetry [19, 74, 75]. However, as we have discussed in Sec. 24.7.4, super-symmetry has to be broken and this destroys the "miraculous" cancellation of the various terms participating the vacuum energy. In fact, the only common point between these various contributions is gravity. Therefore, this suggests that the mechanism that cancels the cosmological constant makes use of gravity. This idea has been explored in the so-called adjustment mechanisms [155, 728, 157]. These models work reasonably well but, unfortunately, they suffer from a severe disease. Indeed, in these scenarios, the Newton constant is time-dependent and it turns out that the variation of the gravitational coupling is too strong to be compatible with the known constraints on $|\dot{G}/G|$. Therefore, although the idea is very attractive at first sight, it seems that it does not yet exist a realistic realization of it. It is also worth signaling that back-reaction mechanisms have also been studied in Refs. [158, 159]. Here, the idea is that the long wavelength (super-Hubble) cosmological perturbations are described by a stress-energy tensor that exactly cancels the cosmological constant.

Another class of solution is based on quantum gravity and quantum cosmology [160, 161, 162]. It is argued that the no-boundary Hartle-Hawking wave-function is such that it peaks at $\Lambda = 0$. However, this approach suffers from various limitations among which is the fact that the path integral is not properly defined and that probabilities are not positive definite (this comes from the fact that the Wheeler-De Witt equation is in fact similar to a Klein-Gordon equation. The positivity of probabilities in quantum cosmology can be only be restored in the WKB approximation).

Recently, following the pioneered work of Ref. [163], there have been many attempts to solve the cosmological constant problem in the framework of extra dimensions, see chapter seven of Ref. [154] or Ref. [164]. Another popular solution, also based on high-energy physics considerations, is the landscape approach [12] in string theory [165]. This approach assumes that all the pocket universes of the landscape are populated during eternal inflation. Combining this fact with the anthropic principle [1818], it is argued that the most probable value of Λ is approximately the value observed today. This solution suffers from the difficulty of defining a measure on the landscape [167] and the use of the anthropic principle in this context has been criticized in Refs. [168, 169].

Finally, another possibility that has recently been investigated is to modify the Einstein equations such that they become blind to a stress energy tensor of the form $T_{\mu\nu} \propto g_{\mu\nu}$, see Refs. [170, 171, 172].

In order to be complete, let us mention that models of analogue gravity have also been developed to study the vacuum energy problem [173, 174, 175]. In these models, one can calculate the equivalent of the cosmological constant and study precisely its origin since the physics controlling the microscopic constituents is explicitly known. These models are based on many different types of analogue physical systems such as condensed matter systems [176], super-fluid Helium [177], Fermi liquid [178], Bose-Einstein condensates [179, 180] etc ... Then, one can show that a simple calculation of the ground state may lead to an incorrect result, the only possibility in order to reach the correct answer being to fully take into account the microscopic theory. This would indicate that a correct calculation of the cosmological constant must necessarily be based on the fine structure of space-time, *i.e.* probably on a consistent theory of quantum gravity.

As the variety of the subjects studied is this review shows, the cosmological constant problem is a very rich one. Given its difficulty, it is probable that it will have to wait for a new theory beyond the current standard model to find its resolution. This theory will have to describe the gravitational properties of the vacuum fluctuations and to regulate the ultra-violet infinities that appear in the calculation of vacuum energy density. This list of requirements almost constitutes an identity card for a theory of quantum gravity. But the most interesting aspect maybe goes the other way round. Given the fact that cosmology has enabled us to grasp observational signatures related to the vacuum energy, maybe these experimental results will help us to deduce and to establish a convincing theory of quantum gravity? This is the reason why the cosmological constant problem appears to be so important and so interesting.

Part VII
Experimental Tests

25 Tests of Well-Known Properties of the Universe

25.1 Test that Universe is Expanding

(???? I'll find some articles)

25.2 Test that Universe is Approximately Flat

(???? I'll find some articles)

26 Some Mathematical Methods for Analysis of Experiments

26.1 Analysis

Increasingly, theorists and even busy experimentalists are turning their attention to the fundamental question of how best to analyze a set of data. The main reason for this focus is that the quality and quantity of data have improved dramatically over the past decade. There is every reason to believe that this trend will continue. Anisotropies in the temperature of the CMB have been measured by dozens of experiments already. The satellites MAP and Planck will take these measurements to the next stage, but there is no reason to think this will be the last stage. There are still polarization and very small scale anisotropies to be measured. The power spectrum of matter is probed in a variety of ways; activity here, too, shows no sign of letting up. After the completion of the Sloan Digital Sky Survey and 2DF, the two largest redshift surveys to date, surveyors have begun planning large weak lensing missions and even deeper galaxy surveys. These larger data sets create new challenges in analysis.

A wonderful/disturbing example of these challenges was given recently by Julian Borrill. He used scaling arguments to show that a brute-force algorithm for making a map from the raw data of the Boomerang CMB anisotropy experiment would take 12 years to run on current computers! Already, data sets are far too large for brute-force calculations. And things are rapidly getting more dire. Since typically the number of arithmetic operations scales as the number of pixels cubed, and since MAP and Planck will have of order 10 to 100 times more data than Boomerang, the situation cries out for creative solutions.

Another reason for the recent focus on analysis is one I hope to convey in this chapter: analysis is exciting. The techniques that have been proposed to deal with the complexity of forthcoming data sets are beautiful. The elegance of these techniques is of course enhanced by their importance. But the elegance is there; for this reason alone, it is well worth working through some of these recent advances.

26.1.1 the Likelihood Function

The basic building block of contemporary analysis is the likelihood function. This is defined as the probability that a given experiment would get the data it did given a theory. This seemingly simple definition is exceedingly powerful. Once we have the likelihood function, with a caveat or two, we can determine the parameters of the theory (best estimate is the place in parameter space where the likelihood function is largest) along with errors (determined by the width of the likelihood function). We start with a simple example and move on to the likelihood function for the CMB and then a galaxy survey.

Simple Example

Suppose you want to weigh somebody. Since you are a scientist, you know that, in addition to the measurement, you should also report an uncertainty. So you set up 100 different scales and record the person's weight on each of these different scales. Given these 100 numbers, what value should you report for the weight and the uncertainty in the weight? We all know the answer to this question, so let's introduce the formalism of the likelihood function in this simple context.

The likelihood function is the probability of getting the hundred numbers given a theory. Our theory will be that each measurement is the sum of a constant signal (the person's weight) w and noise, with the noise drawn from a Gaussian distribution with mean zero and variance σ_w^2. Thus our "theory" has two free parameters, w and σ_w. If only one data point d was taken, the probability of getting d given the theory would be

$$P\left[d|w,\sigma_w\right] := \mathcal{L}\left(d;w,\sigma_w\right) = \frac{1}{\sqrt{2\pi\sigma_w^2}}\exp\left\{-\frac{(d-w)^2}{2\sigma_w^2}\right\}$$

Here and throughout, $P[x|y]$ denotes the probability of x given y. Equation (11.1) simply restates the assumptions that $d-w$ is equal to noise and that the noise is drawn from a Gaussian distribution with standard deviation σ_w. In the limit that σ_w becomes very small, this function becomes sharply peaked at $d = w$. Since we are making $N_{\rm m} = 100$ independent measurements, the likelihood function is the product of all the individual likelihood functions. That is,

$$\mathcal{L} = \frac{1}{\left(2\pi\sigma_w^2\right)^{N_m/2}}\exp\left\{-\frac{\sum_{i=1}^{N_{\rm wi}}\left(d_i-w\right)^2}{2\sigma_w^2}\right\}$$

Notice that, although the data are distributed as a Gaussian, the likelihood function is not Gaussian in all the theoretical parameters (it is in w but is not in σ_w).

We are interested in the value of the theoretical parameters w and σ_w. Thus, we don't want $P\left[\{d_i\}\,|w,\sigma_u\right]$. which is the likelihood function we have computed. Rather, we want $P\left[u:\sigma_{w'}\mid\{d_i\}\right]$. To obtain the latter from the former we can use a simple relation from elementary probability theory,

$$P[B\cap A] = P[B|A]P[A]$$

$$= P[A|B]P[B].$$

In this context $A = \{d_i\}$ and $B = \{w,\sigma_w\}$, so the equality between the two lines of Eq. (11.3) means that

$$P\left[w,\sigma_w\mid\{d_i\}\right] = \frac{P\left[\{d_i\}\,|w,\sigma_w\right]P\left[w,\sigma_w\right]}{P\left[\{d_i\}\right]}.$$

The denominator can be rewritten by realizing that when we integrate the probability $P\left[w,\sigma_w|d\right]$ over all values of the parameters w,σ_w, we must get 1. So the denominator is equal to the integral of the numerator over w,σ_w. As a result, the denominator does not depend on the parameters w and σ_w, so it does not affect the place in parameter space where the likelihood function peaks or the width of the likelihood function. For the most part, then, we are free to ignore it.

To get the probability we want we need the likelihood function - the first term in the numerator-and also the prior probability $P\left[w,\sigma_w\right]$. If we possess prior information about these quantities, we might use this information here. If we want to be conservative, and assume nothing, we put in a uniform prior for the parameters. Then,

$$P\left[w,\sigma_w\mid\{d_i\}\right]\propto\mathcal{L},$$

the proportionality constant being independent of the parameters and therefore of little interest. Many people find this idea of using prior information disturbing. Indeed, even the conservative choice of a uniform prior is not as innocent as it sounds. If we had taken the

parameter to be σ_w^2 instead of σ_w and we had assumed a prior uniform in σ_w^2 (i.e., that equal intervals of σ_w^2 are equally likely), we would get a different answer for the final probability (try it!). Nonetheless, the dependence on the prior is a problem only in cases where the data are not very discriminatory. If the data do have discriminatory power, then the likelihood function $P\left[\{d_i\}\,|w,\sigma_w\right]$ will be sharply peaked and any reasonable prior will not affect the final results.

We can now find best-fit values for our parameters w and σ_w. Simply find the place in parameter space where the likelihood function is largest. In this simple example, we can proceed analytically by differentiating $\mathcal{L}$ with respect to each of the parameters. First consider the derivative with respect to w.

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{\sum_{j=1}^{N_\mathrm{m}} (d_j - w)}{\sigma_w^2 \left(2\pi\sigma_w^2\right)^{N_\mathrm{m}/2}} \exp\left\{ -\frac{\sum_{i=1}^{N_\mathrm{m}} (d_i - w)^2}{2\sigma_w^2} \right\}.$$

For this derivative to be zero, we set the prefactor

$$\sum_{i=1}^{N_\mathrm{m}} (d_i - w) = 0$$

or equivalently, the likelihood is a maximum when

$$w = \bar{w} = \frac{1}{N_\mathrm{m}} \sum_{i=1}^{N_\mathrm{m}} d_i$$

the expected answer. Similarly, we can find what the most probable value of σ_w^2 is by computing

$$\frac{\partial \mathcal{L}}{\partial \sigma_w^2} = \mathcal{L}\left[-\frac{N_\mathrm{m}}{2\sigma_w^2} + \frac{\sum_{i=1}^{N_\mathrm{m}} (d_i - w)^2}{2\sigma_w^4} \right]$$

and setting it equal to zero. Solving for the variance σ_w^2, we find a most probable value of

$$\sigma_w^2 = \frac{\sum_{i=1}^{N_\mathrm{m}} (d_i - \bar{w})^2}{N_\mathrm{m}}$$

again the expected result.

We have found the best-fit values of our theoretical parameters. What is the error in these best-fit values? The error is just proportional to the width of the likelihood function. A simple way to approximate the width is to assume that $\mathcal{L}$ is Gaussian in the parameters, or equivalently that $\ln \mathcal{L}$ is quadratic in the parameters. We know in general that the variance of a Gaussian distribution is twice the inverse of the coefficient of the quadratic term, so we can simply identify the variance (the square of the error) by computing this coefficient. Let's work this out explicitly for w:

$$\begin{aligned} \ln \mathcal{L}(w) &= \ln \mathcal{L}(\bar{w}) + \frac{1}{2} \frac{\partial^2 \ln \mathcal{L}}{\partial w^2}\bigg|_{w=\bar{w}} (w - \bar{w})^2 \\ &= \ln \mathcal{L}(\bar{w}) - \frac{N_\mathrm{m}}{2\sigma_w^2}(w - \bar{w})^2. \end{aligned}$$

Thus the width of the likelihood function at its maximum is $\sigma_w / N_\mathrm{m}^{1/2}$. This is the one-sigma error in our determination of w. This too is familiar: as more measurements are taken, the noise gets beaten down by a factor of 1 over the square root of the number of independent measurements. It is important to reiterate that the uncertainty on our estimate of the weight is not equal to σ_w.

Two numbers then sum up all $N_{\rm m}$ measurements: our best guess for the person's weight - in this case $\bar{w}$ given by Eq. (11.8) - and the error on this estimate, here equal to $\sigma_w/N_{\rm m}^{1/2}$. Therefore, we can compress all 100 measurements into just two by rewriting the likelihood function as

$$\mathcal{L} = \frac{1}{\sqrt{2\pi C_N}} \exp\left\{\frac{-(w-\bar{w})^2}{2C_N}\right\}$$

where the variance due to noise is now

$$C_N = \frac{\sigma_w^2}{N_{\rm m}}$$

This form of the likelihood has precisely the same maximum and width as does the form with all $N_{\rm m}$ data points. Thus it has compressed all the information in the likelihood function into two numbers, $\bar{w}$ and C_N. Moving away from the weight metaphor, we can apply the above to the CMB. Instead of a true weight, the signal s is the true CMB temperature at a given point on the sky. The many measurements of the signal correspond to many measurements of the temperature at that point. The signal at that spot is a constant, and the data are the sum of the constant signal plus noise (atmospheric, instrumental, or both). The compression of all the different measurements into one data point with an associated error as in Eq. (11.12) is called map-making. We will take this up in more detail in Section 11.5. Now, though, we must move beyond the likelihood of Eq. (11.12). For we know that no theory predicts a value for the temperature at a particular position on the sky; i.e., no theory predicts s. Rather all theories predict a distribution of temperatures, from which s at a given pixel is drawn. We must now incorporate this distribution into the likelihood function.

CMB Likelihood

Let's convert the notation of the previous subsection to the CMB. The true temperature anisotropy in a given spot on the sky s replaces w, while the data point $\bar{w}$ (really the average of many measurements) becomes the estimated value of this temperature anisotropy, call it Δ.[1] The variance of this estimator C_N, which represents the spread of the measurements, is also given. How can this set of data (Δ, C_N) be compared with theory? The simplest theories, such as inflation, predict that the signal in a given spot on the sky is drawn from a Gaussian distribution. So the probability that the sky temperature falls in a range between s and $s+ds$ is

$$P(s)ds = \frac{1}{\sqrt{2\pi C_S}} \exp\left\{\frac{-s^2}{2C_S}\right\} ds$$

Here C_S is the variance expected due to the signal alone, independent of any noise. This variance is directly related to the C_l 's in a manner we will explore in Section 11.2.

In order to get the likelihood function, we have to convolve the probability distribution of Eq. (11.14) with the likelihood function of Eq. (11.12). Schematically

$$P\left[\Delta|C_S\right] = \sum_s P[\Delta|s]P\left[s|C_S\right]$$

More concretely, the likelihood function is an integral over all possible values of the true anisotropy:

$$\mathcal{L} = \int_{-\infty}^{\infty} \frac{ds}{\sqrt{2\pi C_S}} \exp\left\{\frac{-s^2}{2C_S}\right\} \frac{1}{\sqrt{2\pi C_N}} \exp\left\{\frac{-(\Delta-s)^2}{2C_N}\right\}$$

The argument of the exponential here is quadratic in s so it is straightforward to carry out the integration over s. Let us rewrite the argument of the exponential as

[1] In Section 11.5 we explore ways to go from the raw data, the timestream, to the pixelized map represented by Δ. For now, we assume that this step has already been taken.

$$-\frac{s^2 C}{2C_S C_N} + \Delta s/C_N - \frac{\Delta^2}{2C_N} = -\frac{C}{2C_S C_N}\left[s - \frac{C_S \Delta}{C}\right]^2 + \frac{C_S \Delta^2}{2C_N} - \frac{\Delta^2}{2C_N}$$

where the full covariance matrix is defined as

$$C := C_S + C_N$$

Changing variables in the integral over s to $x := s - C_S\Delta/C$ leads to

$$\begin{aligned}\mathcal{L} &= \frac{1}{\sqrt{2\pi C_N}} \exp\left\{\frac{C_S \Delta^2}{2CC_N} - \frac{\Delta^2}{2C_N}\right\} \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi C_S}} \exp\left\{-\frac{Cx^2}{2C_S C_N}\right\} \\ &= \sqrt{\frac{1}{2\pi C}} \exp\left\{\frac{-\Delta^2}{2C}\right\}.\end{aligned}$$

This is our final expression for the likelihood function for a one pixel experiment. This form is exactly what one expects: the measured temperature should be distributed like a Gaussian with a variance given by the sum of the variances due to noise and signal.

We can easily generalize Eq. (11.19) to the more realistic case with a measurement of N_p pixels. Then the likelihood function is

$$\mathcal{L} = \frac{1}{(2\pi)^{N_p/2}(\det C)^{1/2}} \exp\left\{-\frac{1}{2}\Delta C^{-1}\Delta\right\}$$

where now Δ is the data vector consisting of all N_p measurements and C is the full covariance matrix. In general, the noise covariance matrix can often be close to diagonal, but the theoretical covariance matrix is not diagonal. Thus, the hard computational part of evaluating the likelihood function is taking the determinant and the inverse of the $N_p \times N_p$ matrix C. The passage from one theoretical parameter (C_S) to a full matrix of parameters creates complications besides the computational. If there was only parameter, observers could quote results in the form of one number. Now that all correlations need to be included, one needs to allow for many different theoretical parameters, in principle all $N_p\left(N_p+1\right)/2$ elements of the (symmetric) covariance matrix C_S. In practice of course this is not done. First of all, the covariance in all theories depends on the angular distance between two points, so elements of the matrix corresponding to two sets of points separated by the same distance are identical. Equivalently, a given theory is associated with a full set of C_l 's; as we will see shortly, these can be used to construct all the elements of C_S. The second simplification is that most experiments have not been sensitive to individual C_l 's but rather to the average power over a range of l, i.e., in a given band. So, analysts typically fit for bandpowers, a fitting which requires even fewer parameters to be determined.

As mentioned the matrix $C = C_S + C_N$ is typically not diagonal. However, we can get some nice insight into the likelihood function by considering the special case when C is diagonal and proportional to the identity matrix (all diagonal elements the same). In that case,

$$\mathcal{L} \propto \frac{1}{\left(C_S + C_N\right)^{N_p/2}} \exp\left\{-\frac{1}{2}\frac{\sum_{i=1}^{N_p}\Delta_i^2}{C_S + C_N}\right\}.$$

We can easily find the value of C_S which maximizes the likelihood function in this case. Differentiating with respect to C_S leads to

$$\frac{\partial \mathcal{L}}{\partial C_S} = \mathcal{L} \times \left[\frac{-\left(N_p/2\right)}{\left(C_S + C_N\right)} + \frac{1}{2}\frac{\sum_{i=1}^{N_p}\Delta_i^2}{\left(C_S + C_N\right)^2}\right].$$

If we set this to zero, we find that the likelihood function is maximized at

$$C_S = \frac{1}{N_p}\sum_{i=1}^{N_p}\Delta_i^2 - C_N$$

Thus a useful rule of thumb for estimating the signal in a CMB experiment is to calculate the variance of the data points (the first term on the right in Eq. (11.23)) and compare it with the average noise per pixel (the second term on the right). If the data has larger variance than the noise, the theoretical signal is simply the difference between the two.

We can also calculate the error on this determination of C_S. As we saw in Section 11.1.1, the error is related to the second derivative of the log of the likelihood function:

$$\sigma_{C_S} = \left(\frac{-\partial^2 \ln \mathcal{L}}{\partial C_S^2}\right)^{-1/2}$$

In this case, it is easy to calculate the derivative. Differentiating Eq. (11.22) once more leads to

$$\frac{\partial^2 \ln \mathcal{L}}{\partial {C_S}^2} = \frac{(N_p/2)}{(C_S + C_N)^2} - \frac{\sum_{i=1}^{N_p}\Delta_i^2}{(C_S + C_N)^3}$$

At the peak of the likelihood we can replace the $\sum_{i=1}^{N_p}\Delta_i^2$ by $N_p\,[C_S + C_N]$, so

$$\sigma_{C_s} = \sqrt{\frac{2}{N_p}}\,(C_S + C_N)$$

Equation (11.26) is a simplified version of a very handy, useful formula which can be used to assess how accurately a given experiment will determine parameters. This simplified version gives the errors on our one theoretical parameter, C_S. The more general formula gives the corresponding errors when the free parameters are the C_l 's themselves. In that case,

$$\sigma_{C_l} = \sqrt{\frac{2}{(2l+1)f_s}}\,(C_l + C_{N,l})\,.$$

The only change moving from Eq. (11.26) to Eq. (11.27) is that the number of pixels - or equivalently the number of independent measurements - has been replaced by $(2l+1)f_s$, where f_s is the fraction of the sky covered. This makes perfect sense, for in the full sky limit, one can measure at best $2l + 1a_{lm}$'s; that is, one can sample the distribution characterized by C_l only $2l + 1$ times. In fact, this is a fundamental limit on the accuracy with which we can measure the C_l 's. Even if there is no noise ($C_N = 0$), there remains a fundamental uncertainty in the theoretical parameters (either C_S or C_l) due to the fact that we only have one sky on which to take measurements. This limit, which we have already encountered in Chapter 8, is called sample variance, or in the limit of an all-sky survey cosmic variance.

Galaxy Surveys

At first, one might think that analysis of galaxy surveys would be completely different from CMB analysis. There are a number of differences. The galaxy distribution is fundamentally 3D, while the CMB anisotropies are a function of angular position only. Also, CMB experiments measure a continuous field, the temperature field, a function of position. Galaxy surveys count discrete objects (galaxies). A survey is simply a list of positions of these objects. Another difference is that the CMB temperatures are drawn from a Gaussian distribution, whereas the galaxies supposedly trace the underlying mass distribution, which - at least on small scales - has already "gone nonlinear." Nonlinearities inevitably produce non-Gaussianity, even if the primordial distribution is Gaussian.

Despite these, and other, differences, analysts have in recent years come to realize that many of the same techniques can be applied to data from both the CMB and galaxy surveys. To solidify the CMB-galaxy survey connection, we need to formalize the concept of a pixel. In

the case of the CMB, the notion of a pixel is so natural that I didn't even bother to define it above. For galaxy surveys, following the treatment of Tegmark et al. (1998), we can define the data in pixel i as

$$\Delta_i := \int d^3x \psi_i(\vec{x}) \left[\frac{n(\vec{x}) - \bar{n}(\vec{x})}{\bar{n}(\vec{x})} \right]$$

Here $n(\vec{x})$ is the galaxy density at $\vec{x}$ and $\bar{n}$ is the expected number of galaxies at $\vec{x}$, i.e., the number there would be if the distribution was uniform. The weighting function ψ_i, which determines the pixelization, will be discussed shortly, but first let's understand operationally how to determine n and $\bar{n}$ from a survey. A simple way is to divide the volume into small sub-volumes, each of which is much smaller than the total survey, but large enough to contain many (e.g., greater than 10) galaxies. The density of a given sub-volume is then the number of galaxies in it divided by its volume. For a uniform survey, the average density $\bar{n}$ would just be the total number of galaxies divided by the total volume. [2]

There are two popular choices for ψ_i, choices which determine the pixelization. First is "counts-in-cells," wherein

[2] A caveat: if the pixels as defined by ψ_i are overlapping, more care must be taken in computing $\bar{n}$ for then the same galaxy could appear in more than one pixel.

$$\psi_i^{\rm CIC}(\vec{x}) = \begin{cases} \bar{n}(\vec{x}) & \text{if } \vec{x} \text{ is in the } i \text{ th sub-volume} \\ 0 & \text{otherwise} \end{cases} .$$

In this case, the sub-volumes themselves are the pixels, and Δ_i is the over(under)density in the i th sub-volume. Another useful pixelization scheme is a set of Fourier pixels, which emerge from choosing

$$\psi_i^{\rm Fourier}(\vec{x}) = \frac{e^{i\vec{k}_i \cdot \vec{x}}}{V} \begin{cases} 1 & \vec{x} \text{ inside the survey volume} \\ 0 & \vec{x} \text{ outside survey volume} \end{cases} .$$

Here V is the volume of the survey. In this case, the pixels are not spatial, but rather live in the Fourier domain. Still, even in this case, Δ_i is the fractional overdensity in the pixel.

No matter which pixelization is chosen, one cannot hope to write down a simple expression for the likelihood function, the probability of getting a set of $\{\Delta_i\}$ given a theory. The theory of galaxy formation is simply too complicated. Indeed, even if one assumes that the galaxy density perfectly traces the mass overdensity, the complications from gravity alone make the likelihood function non-Gaussian. Nonetheless, progress can still be made by noting that the expectation value of Δ_i is zero by construction, with a covariance matrix [3]

$$\left\langle \Delta_i \Delta_j^* \right\rangle = (C_S)_{ij} + (C_N)_{ij}$$

exactly like the CMB case. We will discuss the signal covariance matrix in detail in Section 11.2. The noise covariance matrix is actually easier than the corresponding CMB matrix, which depends on the atmosphere, pointing, instrumental noise, scan strategy, and other experimental details. In a galaxy survey, even if there was no signal, the expected value of the square of the density, $\langle n^2(\vec{x}) \rangle$, would still differ from $\bar{n}^2$ simply because there are only a finite number of galaxies in a given subvolume. Thus, even in the absence of any intrinsic clustering of the galaxies $\langle \Delta_i \Delta_j \rangle$ would be nonzero because of Poisson noise. You can show in Exercise 3 that the covariance matrix due to Poisson noise is

$$(C_N)_{ij} = \int d^3x \frac{\psi_i(\vec{x}) \psi_j^*(\vec{x})}{\bar{n}(\vec{x})}$$

Armed with this noise covariance matrix and the signal covariance matrix we explore next, galaxy survey analysts can use many of the same techniques as the CMB analysts.

26.1.2 Signal Covariance Matrix

Theory

Until now, we have sidestepped the question of how the expected variance in a given experiment is related to the underlying power spectrum. That is, we have learned that the predictions

of a given theory are a set of C_l 's and $P(k)$. If we want to relate theory to experiment - and we do! — we need to know how to turn this set of predictions into a covariance matrix C_S.

[3] The angular brackets here denote an average over the distribution from which Δ_i is drawn.

CMB Window Functions

For simplicity, let us first consider the diagonal element of the covariance matrix:

$$C_{S,ii} := \langle s_i s_i \rangle \quad (\text{ no sum over } i)$$

where the average $\langle \ldots \rangle$ is over many realizations of the theoretical distribution and the subscript i labels the pixel. The temperature difference reported in each pixel can be expressed as

$$s_i = \int d\hat{n} \Theta(\hat{n}) B_i(\hat{n})$$

where B_i is the beam pattern at the i th pixel and Θ is the underlying temperature. As an example, the beam pattern from the MSAM experiment is shown in Figure 11.1. It is typical of the patterns produced by many CMB experiments: the difference of the temperature in two (or more) regions of the sky, and in each region the temperature is sampled by a beam which is roughly Gaussian.

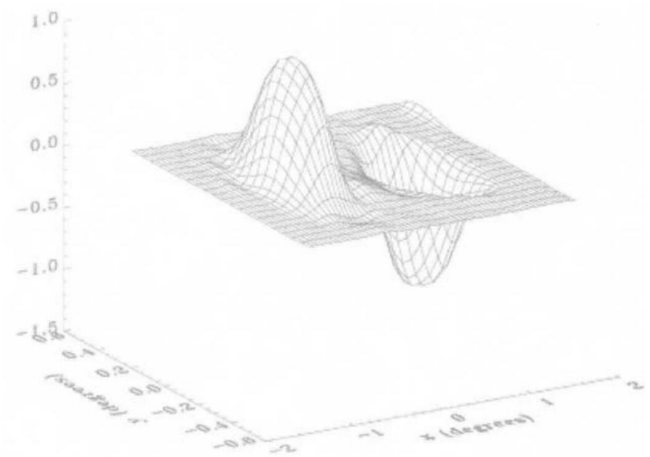


Figure 11.1. The beam pattern for the MSAM experiment (see Wilson et al., 2000, for a summary). The anisotropy reported at a given pixel is roughly the difference between the temperature at \pm 0.5 degrees from the center of the pixel. The beamwidth is also on the order of half a degree.

To find C_S we square Eq. (11.34) after expressing the temperature field as an expansion over spherical harmonics as in Eq. (8.60):

$$\frac{C_{S,ii}}{T^2} = \int d\hat{n} \int d\hat{n}' B_i(\hat{n}) B_i(\hat{n}') \sum_{lm} Y_{lm}(\hat{n}) \sum_{l'm'} Y^*_{l'm'}(\hat{n}') \langle a_{lm} a^*_{l'm'} \rangle$$

Using Eq. (8.63), we find that the sums over $l'm'$ collapse to give

$$\frac{C_{S,ii}}{T^2} = \int d\hat{n} \int d\hat{n}' B_i(\hat{n}) B_i(\hat{n}') \sum_{l} C_l \sum_{m} Y_{lm}(\hat{n}) Y^*_{lm}(\hat{n}') .$$

But $\sum_m Y_{lm}(\hat{n}) Y^*_{lm}(\hat{n}') = (2l+1) P_l(\hat{n} \cdot \hat{n}') / 4\pi$, so

$$\frac{C_{S,ii}}{T^2} = \sum_l \frac{2l+1}{4\pi} C_l W_{l.ii}$$

where the window function is defined as

$$W_{l,ii} := \int d\hat{n} \int d\hat{n}' B_i(\hat{n}) B_i(\hat{n}') P_l(\hat{n} \cdot \hat{n}')$$

Until now, we have been thinking of $\hat{n}, \hat{n}'$ as three-dimensional unit vectors. If $\hat{n}'$ and $\hat{n}$ are sufficiently close to each other, though, we can use the flat space approximation. The three-dimensional unit vectors can be safely approximated as two-dimensional vectors $\vec{x}, \vec{x}'$ in the transverse directions. The distance between $\vec{x}$ and $\vec{x}'$ (measured in radians) is then equal to the angle between $\hat{n}$ and $\hat{n}'$. In this limit, the argument of the Legendre polynomial in Eq. (11.38) becomes

$$\vec{n} \cdot \vec{n}' = \cos(|\vec{x} - \vec{x}'|).$$

The diagonal window function is therefore

$$W_{l,ii} = \int d^2x \int d^2x' B_i(\vec{x}) B_i(\vec{x}') P_l(\cos(|\vec{x} - \vec{x}'|))$$

A useful property of Legendre polynomials is that they become equal to the zeroorder Bessel function in the limit of large l (the small-angle limit we are working under here). So,

$$\begin{aligned} P_l(\cos(|\vec{x} - \vec{x}'|)) &\to J_0(l|\vec{x} - \vec{x}'|) \\ &= \frac{1}{2\pi} \int_0^{2\pi} d\phi e^{-il|\vec{x}-\vec{x}'|\cos\phi}, \end{aligned}$$

where the last line is an integral representation of the Bessel function (Eq. (C.21)). We can simplify further by promoting l to a 2D vector with direction chosen so that the angle between $\vec{l}$ and $\vec{x} - \vec{x}'$ is equal to ϕ. Then, the argument of the exponential simplifies to $-i\vec{l} \cdot (\vec{x} - \vec{x}')$. This form is so useful because the $\vec{x}$ integral for example is now

$$\int d^2x B_i(\vec{x}) e^{-i\vec{l}\cdot\vec{x}} := \tilde{B}_i(\vec{l})$$

where $\tilde{B}_i$ is the Fourier transform of the beam pattern. The $\vec{x}'$ integral is the complex conjugate of this, so the window function simplifies to

$$W_{l,ii} = \frac{1}{2\pi} \int_0^{2\pi} d\phi \left|\tilde{B}_i(\vec{l})\right|^2$$

Thus calculating the window function reduces to a two-step process: - Calculate the 2D Fourier transform of the beam pattern.

- Find the angular average of the square of this transform.

The window function is a function of the experiment only and indeed contains information about the beam size and chopping angle of the experiment. However, it is not the whole story. A complete evaluation of the likelihood function entails calculating all of the elements of the covariance matrix C_S. A given off-diagonal element of the matrix is given by Eqs. (11.37) and (11.38), with one of the indices i changed to j. The matrix is symmetric, so it is characterized by $N_p(N_p + 1)/2$ elements where N_p is the number of pixels. Thus, for an N_p-pixel experiment there are really $N_p(N_p + 1)/2$ window functions!

Examples of CMB Window Functions

Gaussian Beam. Let us take a break from formalism and calculate a simple (diagonal) window function. Consider a Gaussian beam; this is a good approximation to many CMB experiments. The beam pattern for the i th pixel is

$$B_i(\vec{x}) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(\vec{x}-\vec{x}_i)^2}{2\sigma^2}\right)$$

We may choose $\vec{x}_i$ to be zero for the window function computation. The Fourier transform of the beam is also a Gaussian,

$$\begin{aligned}\tilde{B}_i(\vec{l}) &= \frac{1}{2\pi\sigma^2}\int d^2x e^{-i\vec{l}\cdot\vec{x}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \\ &= e^{-l^2\sigma^2/2}.\end{aligned}$$

In this simple case, $\tilde{B}$ does not depend on the direction of $\vec{l}$, so there is no need to take the angular average. The window function is then simply the square of the Fourier transform,

$$W_{l,ii} = e^{-l^2\sigma^2}$$

The window function falls off sharply at large l. Large l corresponds to small angular scales. Structure on scales smaller than the beam size is inevitably washed away and undetectable. Figure 11.2 illustrates the series of steps, from beam function to Fourier transform to window function.

There are two subtleties associated with the Gaussian window function. First, one must avoid the temptation to set σ equal to the number which is often quoted in papers, the full width half maximum (FWHMI). The latter is twice the value of x: for which $B(\vec{x})$ drops to half of its maximum. So $\sigma = FWHM/(\sqrt{8\ln(2)}) = 0.4245$ FWHM. The second subtlety has to do with normalization. It is crucial to determine how observers have normalized their output. The prefactor in Eq. (11.44) ensures that if the temperature field was uniform, the reported temperture would be equal to the underlying one. In this simple case, the choice is obvious; in general factors of 2 can easily be lost.

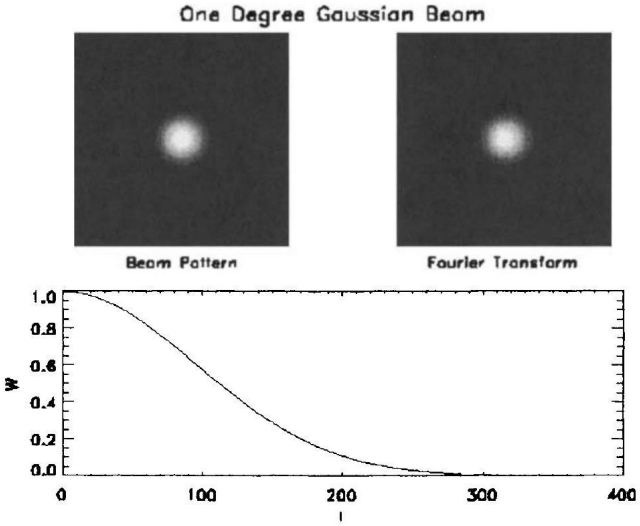


Figure 11.2. A one-degree Gaussian beam is shown in the upper left panel and its Fourier transform, also a Gaussian, in the upper right. The resulting window function is shown in the bottom panel.

Differencing a Gaussian Beam. As another straightforward example, let us consider an experiment which takes the difference between the temperatures at two adjacent points on the sky. For simplicity, let us first assume that the Gaussian beam is infinitely small, so it can be approximated as a Dirac delta function. Then,

$$B(x,y) = \delta(y)\left[\delta\left(x-x_0\right) - \delta\left(x+x_0\right)\right]$$

where the chopping angle, or the distance between the plus and minus position, is $2x_0$. The Fourier transform of this is straightforward:

$$\tilde{B}(\vec{l}) = 2i\sin\left(l_x x_0\right)$$

The window function is the angular average over all $\vec{l}$ directions. Choosing the angle between $\vec{l}$ and the x - axis to be ϕ, we have

$$\begin{aligned} W_l &= \frac{4}{2\pi}\int_0^{2\pi} d\phi \sin^2\left[l x_0 \cos\phi\right] \\ &= \frac{1}{\pi}\int_0^{2\pi} d\phi \left(1 - \cos\left[2 l x_0 \cos\phi\right]\right) \\ &= 2\left(1 - P_l\left[\cos\left(2x_0\right)\right]\right) \end{aligned}$$

where the last line follows from Eq. (11.41).

Until now, we have neglected the finite width of the beam. However, this turns out to be very simple to rectify. A realistic beam will be the convolution of the chop described by Eq. (11.47) with a finite beam size:

$$\begin{aligned} B(x,y) = \frac{1}{2\pi\sigma^2}\int dx' dy' \exp\left\{-\frac{(x-x')^2 + (y-y')^2}{2\sigma^2}\right\} \\ \times \delta\left(y'\right)\left[\delta\left(x' - x_0\right) - \delta\left(x' + x_0\right)\right] \end{aligned}$$

Recall, though, that the Fourier transform of the convolution of two functions is simply equal to the product of the two Fourier transforms. The angular averaging over this product is unaffected since the Gaussian has no angular dependence. Therefore, the final window function is

$$W_l = e^{-l^2\sigma^2}\left(1 - P_l\left[\cos\left(2x_0\right)\right]\right)$$

This window function is shown in Figs. 11.3 and 11.4 along with the beam pattern

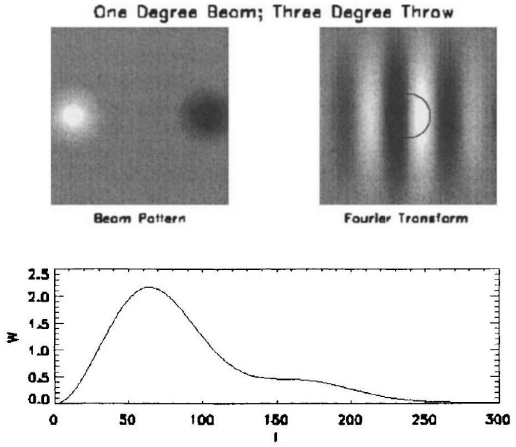
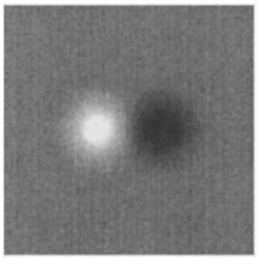
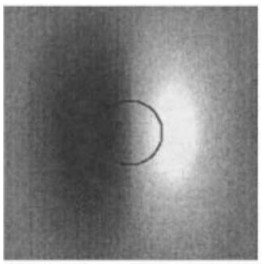


Figure 11.3. Differencing a Gaussian beam. Upper left panel shows the beam pattern and the upper right its Fourier transform. The circle in the upper right corresponds to $l = 50$. The bottom panel shows the window function. One Degrae Beam; One Degrae Throw



Beom Pottern



Fourler Tranefarm

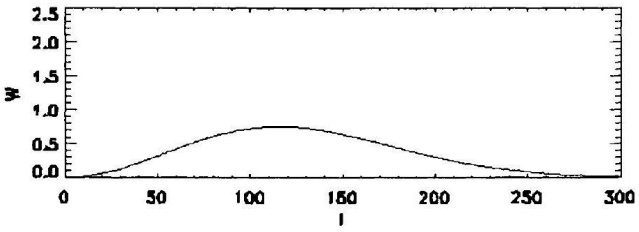


Figure 11.4. Same as Figure 11.3 except the chopping angle is now much smaller. Note the lack of support at large scales due to the reduced chopping angle. This can also be seen in the Fourier transform which vanishes near the center of the circle.

for two different chopping angles. Note that for l much smaller than $1/x_0$, the Fourier transform vanishes. So, unlike the undifferenced beam, there is no support for small l. That is, chopping removes information about structure on large scales. As the chopping angle gets larger, more and more information is obtained about large scales.

Window Functions for Galaxy Surveys

We now consider the signal covariance matrix for galaxy surveys. By setting the term in square brackets in Eq. (11.28) to $\delta(\vec{x})$, we see that the signal covariance matrix is equal to

$$(C_S)_{ij} = \langle \Delta_i \Delta_j \rangle|_{\text{no noise}} = \int d^3x d^3x' \psi_i(\vec{x}) \psi_j(\vec{x}') \, \xi(\vec{x} - \vec{x}')$$

The correlation function ξ appears here because it is equal to the expectation value of the product of two overdensities, Eq. (9.7). Since ξ is the Fourier transform of the power spectrum, we see that the signal covariance matrix for a galaxy survey is

$$\begin{aligned}(C_S)_{ij} &= \int d^3x d^3x' \int \frac{d^3k}{(2\pi)^3} \frac{d^3k'}{(2\pi)^3} \frac{d^3k''}{(2\pi)^3} \tilde{\psi}_i(\vec{k}) \tilde{\psi}_j^*\left(\vec{k}'\right) P(k'') \, e^{i\left[\vec{k}+\vec{k}''\right]\cdot\vec{x} - i\left[\vec{k}'+\vec{k}''\right]\cdot\vec{x}'} \\ &= \int \frac{d^3k}{(2\pi)^3} P(k) \tilde{\psi}_i(\vec{k}) \tilde{\psi}_j^*(\vec{k})\end{aligned}$$

The second equality follows simply after integrating over $\vec{x}$ and $\vec{x}'$ to get 3D Dirac delta functions and using these to perform the integrals over $\vec{k}'$ and $\vec{k}''$. It is convenient to define the window function as the angular part of this integral

$$W_{ij}(k) := \int \frac{d\Omega_k}{4\pi} \tilde{\psi}_\imath(\vec{k}) \tilde{\psi}_j^*(\vec{k})$$

so that

$$(C_S)_{ij} = \int_0^\infty \frac{dk}{k} \left[\frac{k^3 P(k)}{2\pi^2}\right] W_{ij}(k)$$

Notice that the window function for galaxy surveys has the identical form as that for CMB experiments. In both cases, it is the angular average of the square of the Fourier transform of the weighting function (either B or ψ). Also, you should recognize the quantity in square

brackets in Eq. (11.55) as $\Delta^2(k)$, the contribution to the variance per $\ln(k)$. Let's turn to some examples of window functions of galaxy surveys.

Volume-Limited Survey. Consider a survey which observes all galaxies within a radius R from us. If we use Fourier pixelization (Eq. (11.30)), then the Fourier transform of the weighting function is

$$\tilde{\psi}_i(\vec{k}) = \int\limits_{|\vec{x}|<R} \frac{d^3x}{V} e^{-i\vec{k}\cdot\vec{x}} e^{i\vec{k}_2\cdot\vec{x}}$$

We will shortly carry out this integral, square, and then average over all angles to get the diagonal window function of Eq. (11.54). First, though, let's ask what we expect qualitatively. Equation (11.56) is the Fourier transform of the survey volume as a function of $\vec{k} - \vec{k}_i$. The survey volume is a sphere of radius R. In general, when a function is confined to a region $x < R$, the Fourier transform is confined to $k < 1/R$. Here then, $\tilde{\psi}$ will be nonzero only when $\left|\vec{k} - \vec{k}_i\right|$ is less than $1/R$. The window function therefore should peak at $k = k_i$ and have a width of order $1/R$.

More quantitatively, the integral in Eq. (11.56) is

$$\begin{aligned}\tilde{\psi}_i(\vec{k}) &= \frac{4\pi}{V\left|\vec{k} - \vec{k}_i\right|} \int\limits_0^R dx x \sin\left(\left|\vec{k} - \vec{k}_i\right| x\right) \\ &= \frac{4\pi}{V\left(\left|\vec{k} - \vec{k}_i\right|\right)^3} \left[-\left|\vec{k} - \vec{k}_i\right| R\cos\left(\left|\vec{k} - \vec{k}_i\right| R\right) + \sin\left(\left|\vec{k} - \vec{k}_i\right| R\right)\right] (1\end{aligned}$$

The diagonal window function is the angular average of the square of this. Defining $y := \left|\vec{k} - \vec{k}_i\right| R$, this angular average is

$$\begin{aligned}W_{ii}(k) &= \frac{(4\pi R^3)^2}{V^2} \int\limits_{-1}^{1} \frac{d\mu}{2} \int\limits_0^{2\pi} \frac{d\phi}{2\pi} \frac{(\sin y - y\cos y)^2}{y^6} \\ &= \frac{8\pi^2 R^6}{V^2} \int\limits_{-1}^{1} \frac{d\mu}{y^6} (\sin y - y\cos y)^2\end{aligned}$$

where μ is the cosine of the angle between $\vec{k}$ and $\vec{k}_i$. Integrating over y instead of μ leads to

$$W_{ii}(k) = \frac{9}{2kk_iR^2} \int\limits_{|k-k,|R}^{(k+k,)R} \frac{dy}{y} j_1^2(y)$$

since the volume of the survey $V = 4\pi R^3/3$. This window function is shown in Figure 11.5 for several different values of k_iR. Notice that modes with wavelength much smaller than the size of the survey $k_iR \gg 1$ do indeed have window functions sharply peaked at $k = k_i$, with a width of order $1/R$. The largest wavelength modes, however, pick up contributions from all scales (e.g., the $k_iR = 3$ curve in Figure 11.5). Not surprisingly, surveys do not do a good job measuring the power on wavelengths comparable to their sizes.

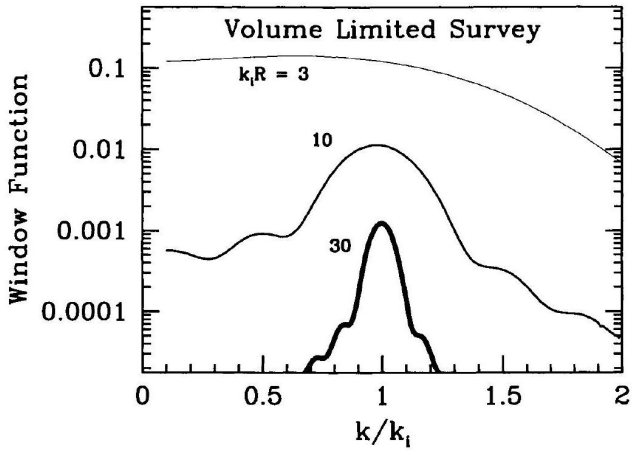


Figure 11.5. The window function for a volume-limited survey. Narrow window functions enable one to determine the power spectrum at the wavenumber of interest more accurately. Modes with wavelengths comparable to the size of the survey (e.g., here $k_iR = 3$) have broad window functions.

The height of the window function is also important, for it determines the amplitude of the signal covariance matrix. When considering modes with wavelengths k_i^{-1} much smaller than R, you will show in Exercise 7 that

$$(C_S)_{ii} \simeq \frac{P(k_i)}{V}.$$

It is instructive to compare this to the corresponding element of the noise matrix, as given in Eq. (11.32). For the diagonal elements. $|\psi_i|^2$ in the integrand is $1/V^2$ as long as $\vec{x}$ is in the survey volume. Thus,

$$(C_N)_{ii} = \frac{1}{\bar{n}V}$$

equal to the inverse of the total number of galaxies in the survey. The ratio of the diagonal elements of the signal and noise covariance matrices is therefore

$$\frac{(C_S)_{ii}}{(C_N)_{ii}} \simeq P(k_i)\,\bar{n}$$

Harking back to Eq. (11.27), we identify this ratio as the ratio of cosmic variance to Poisson noise. A rough estimate is that $\bar{n} \sim 1h^3\mathrm{Mpc}^{-3}$, so cosmic variance dominates as long as the power spectrum is larger than $1h^{-3}\mathrm{Mpc}^3$. Looking back to Figure 7.11, we see that on large scales, this is always satisfied. On small scales, eventually the power spectrum does drop beneath $1h^{-3}\mathrm{Mpc}^3$, the linear power spectrum of standard CDM at $k \sim 10h\mathrm{Mpc}^{-1}$. On very small scales, therefore, Poisson noise becomes important.

Pencil-Beam Survey. Now consider a survey which is very deep, but also very narrow, with the general shape of a pencil (Figure 11.6). The Fourier transform of the

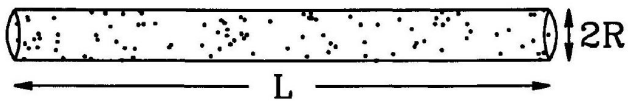


Figure 11.6. A pencil-beam survey with depth L much larger than the typical size of the narrow dimensions, R. The z-axis in the text is taken to be aligned with the long dimension.

weighting function in this case is

$$\tilde{\psi}_i(\vec{k}) = \int \frac{d^3x}{\pi R^2 L} e^{i(\vec{k}_\mathrm{k}-\vec{k})\cdot\vec{x}} \Theta(z + L/2)\Theta(L/2 - z)\Theta\left(R^2 - x^2 - y^2\right)$$

where Θ is a step function, equal to 1 if its argument is positive and zero otherwise. It is easiest to carry out this integral in cylindrical coordinates, wherein $x^2 + y^2 = r^2$. If we define $\vec{q} := \vec{k}_i - \vec{k}$, then

$$\tilde{\psi}_i(\vec{k}) = \frac{1}{\pi R^2 L} \int_{-L/2}^{L/2} dz e^{iq_z z} \int_0^R dr r \int_0^{2\pi} d\theta e^{iq_r r \cos\theta}$$

The azimuthal integral can be done using Eq. (C.17), the integral over z using Eq. (C.15), so that

$$\tilde{\psi}_i(\vec{k}) = \frac{2}{R^2} j_0\left(q_z L/2\right) \int_0^R dr r J_0\left(q_r r\right)$$

Finally the integral over r is $RJ_1\left(q_r R\right)/q_r$, which you can see by differentiating the integral with respect to $q_r R$ and then using Eq. (C.22). Therefore, the Fourier transform of the weighting function is

$$\tilde{\psi}_i(\vec{k}) = \frac{2}{(q_r R)} j_0\left(q_z L/2\right) J_1\left(q_r R\right)$$

Equation (11.66) indicates that the Fourier transform of the weighting function is anisotropic. Indeed, even before deriving the various flavors of Bessel functions, we should have expected this Fourier transform to be compact along the q_z direction and broad in the transverse plane, i.e., shaped like a disk. This flows from our intuition that the Fourier transform of a function localized within a radius R should be localized within a region $1/R$. Indeed, in the z direction, $j_0\left(q_z L/2\right)$ falls off once q_z gets larger than $2/L$. The same holds for $J_1\left(q_r R\right)$: it becomes small for $q_r > 1/R$. The ringing associated with these Bessel functions is a manifestation of the fact that the Fourier transform of a top-hat function oscillates for large wavenumbers (e.g., Exercise 6).

To get the window function for a pencil-beam survey, we need to average Eq. (11.66) over all directions $\hat{k}$. This average will differ for different $\vec{k}_i$. Let's choose $\vec{k}_i$ to point in the $\hat{z}$-direction as one concrete example. The averaging will pick up contributions only when $\vec{q} := \vec{k}_i - \vec{k}$ has z-component smaller than L^{-1} and transverse component smaller than R^{-1}. Since the transverse component of $\vec{k}_i$ is zero in this example, many $\vec{k}$ will contribute, as long as their transverse component is smaller than R^{-1}. Therefore, we expect the window function to get contributions from many wavenumbers, not to be sharply peaked around $\left|\vec{k}_i\right|$. A similar argument holds for other directions $\hat{k}_i$. Figure 11.7 shows the window function for a pencilbeam survey. As expected, it is broader than that for a symmetric, volume-limited survey: a given scale k_i picks up contributions from smaller scales $k > k_i$.

Summary

We have determined the signal covariance matrix for CMB experiments which measure a filtered version of the temperature in a given set of pixels and for galaxy surveys which measure the overdensity in a given set of pixels. Not surprisingly, the fundamental relation between the covariance matrix and and the underlying power spectrum is very similar in both cases: the connection is provided by a window function determined by the experimental/observational specifications. It is interesting to point out that we have encountered natural window functions in Chapters 9 and 10. The angular correlation function of Section 9.1 is simply the signal covariance matrix of measurements of the 2D galaxy distribution. Recall from Eq. (9.13) that this too is an integral of the 3D power spectrum convolved with a window function (we called it a kernel back then). The same is true for the peculiar velocity (Eq. (9.29)) and the shear field that can be measured with weak lensing (Eq. (10.32)). In all of those cases, the window function is determined partly by the

Pencil Beam Survey

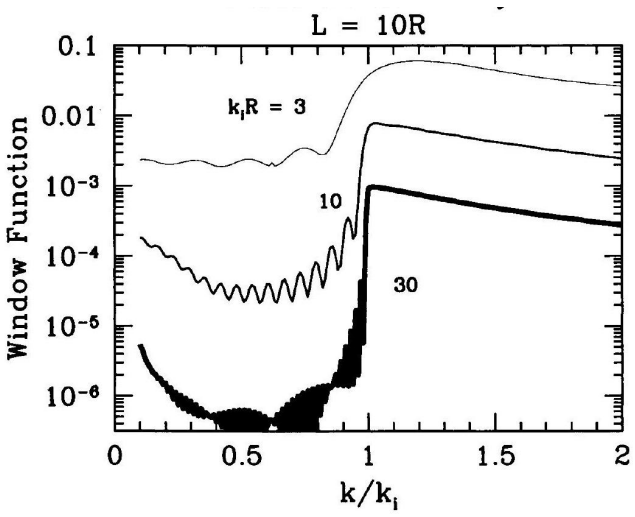


Figure 11.7. The window function for a pencil-beam survey with length 10 times larger than width. These window functions are for $\vec{k}_2$ pointing in the $\hat{z}$-direction, along the long dimension of the survey.

observational strategy (e.g., how deep one goes), but also by the intrinsic nature of the measurement itself. For example, the 2D galaxy distribution intrinsically is a projection of the 3D distribution. So it is seductive to think of CMB anisotropy experiments and 3D galaxy surveys as more pristine measurements of the power spectrum. Analysts are hard to seduce, though. They recognize that, mathematically, the different sets of measurements can all be analyzed in the same fashion. So the likelihood function can be used on virtually all cosmological observations. It is a very powerful tool.

It is also simple. Were it not for the size of modern cosmological data sets, we would be done. At least in the Gaussian case (CMB or even galaxy surveys on large scales), the likelihood function is given by Eq. (11.20). The data points are simply the pixelized temperatures or overdensities, while the covariance matrix is the sum of the noise and signal covariances. The noise covariance matrix is usually estimated from the data itself. while the signal covariance matrix is computed by convolving the theory (i.e., the C_i 's or the power spectrum) with the window function. In principle. then, one could compute this likelihood function at many points in parameter space, find its maximum (this constitutes the set of best-fit parameters), and the contour delineating the region in which. say; 95% of the volume lies. This contour would then be the 95% confidence region of the parameters. Many experiments of the previous decade, especially CMB experiments, have been analyzed in this brute-force fashion. Times are changing, though, and the bruteforce approach has already become impractical.

26.1.3 Estimating the Likelihood Function

Theory

To illustrate the need for new techniques of likelihood computation, let us consider a concrete example: the data set from the Boomerang anisotropy experiment (Netterfield et al., 2001). There are 57,000 pixels on the sky covered in this set. The theory and noise covariance matrices are both nondiagonal and both $57,000 \times 57,000$ dimensional matrices. Inverting these beasts with present computers is possible, although slow. [4] If we needed to invert only once, this might be acceptable. But, we need to evaluate the likelihood function at many points in parameter space to find its maximum and the region at which it falls to, say, 5% of its maximum. This would be barely manageable if the parameter space was one-dimensional. A onedimensional fit, though, would lose most of the information contained in the map. The data are actually sensitive to the power on many different scales. Therefore, the parameter space - the amplitude of the power on these many different scales is multidimensional, "multi-" here of order 20. The likelihood function should in principle be computed about 10 times in each dimension, for a total of 10^{20}com putations. Since each inversion takes several hours, this is not feasible. All of these estimates are for the Boomerang experiment. The MAP satellite will have 10 times as many pixels and be sensitive to a wider range of scales. Planck will be more sensitive still.

Thus, we need new techniques, shortcuts, for evaluating the likelihood function and finding its maximum and its width.

Karhunen-Love Techniques

The first technique was discovered many years ago and reinvented by a number of people over the past few years to deal with both CMB data (Bunn, 1995; Bond, 1995) and with data from galaxy surveys (Vogeley and Szalay, 1996). It is a method for speeding up the computation of the likelihood function. The fundamental idea is simple: any experiment, no matter how good, will have many modes which are useless, fundamentally contaminated by noise. [5] If it was obvious which modes were most noisy, then we could greatly simplify the likelihood calculation by not using those modes. If only 10% of the modes carried useful information - and this figure is roughly what is found in many present-day experiments - then the data set would be reduced by a factor of 10. The covariance matrices would now be $(N/10) \times (N/10)$ and the inversion (which scales as N^3) would speed up by a factor of 1000. It is a nice, simple idea. The only problem is finding the modes which are useful.

[4] Note that the problems outlined here assume that we are handed the map of 57.000 pixels. Mapmaking is actually the most difficult computational part of the analysis!

"A mode here is defined as a linear combination of the dat a points. If the signal and noise covariance matrices were diagonal, then it would be simple to ascertain which modes had high signal-to-noise. The pixels with diagonal elements $C_S > C_N$ would have signal-to-noise greater than 1 ; the others would be low-signal modes. The problem is to identify the low modes in the more realistic case where the covariance matrices are not diagonal. This is precisely what KarhunenLove does. To illustrate the technique, let us first write it down formally and then work out a simple example explicitly.

We assume there are N_p data points, Δ_i. Each data point is presumed to be the sum of both signal s_i and noise n_i. Each of these are assumed to be uncorrelated (the noise knows nothing about the signal and vice versa). Thus the full covariance matrix is

$$\langle \Delta_i \Delta_j \rangle := C_{ij} = C_{S.ij} + C_{N,ij}$$

The Karhunen-Love method takes advantage of the fact that instead of computing the likelihood function using Δ_i and its covariance matrix C, we could instead use rotated data

$$\Delta'_i := R_{ij} \Delta_j$$

where R is a real matrix. The covariance matrix associated with Δ' will be

$$\begin{aligned} C'_{ij} &= \langle (R\Delta)_i (R\Delta)_j \rangle \\ &= R_{ii'} R_{jj'} C_{i'j'} \end{aligned}$$

In matrix notation this is simply

$$C' = RCR^T$$

where R^T is the transpose of R.

The Karhunen-Love method consists of three such rotations.

1. R_1: Diagonalize C_N
2. R_2 : Set $C'_N = I$
3. R_3: Diagonalize C'_S

The first step is always possible since C_N is a real, symmetric matrix. Once C_N has been diagonalized, it is trivial to perform step 2: simply choose R_2 to be diagonal with elements equal to 1 over the square root of the diagonal elements that emerge from step 1. Finally, step 3 is straightforward since again C_S is a real symmetric matrix. Let's evaluate the new theory and noise covariance matrices. The theory matrix is

$$C'_S = R_3 R_2 R_1 C_S R_1^T R_2 R_3^T$$

Note that, since R_2 is diagonal it is equal to its transpose. The matrix C'_S is a diagonal matrix. Now consider C'_N. After step 2, it was simply the identity matrix. So we need consider only the effect of step 3. In fact, since R_3 is unitary, it has no effect $\left(R_3 I R_3^T = I\right)$. Thus, C'_N is still equal to the identity matrix. This has profound implications. It means that the elements of (the diagonal matrix) C'_S are a measure of the signal-to-noise squared of the modes! The data points

$$\Delta'_i = (R_3 R_2 R_1)_{ij}\, \Delta_j$$

then have diagonal covariance matrix

$$\left\langle \Delta'_i \Delta'_j \right\rangle = \begin{cases} 1 + C'_{S,ii} & i = j \\ 0 & i \neq j \end{cases} .$$

These modes can be ordered according to their signal-to-noise values. Modes with large C'_S can be kept; those with C'_S significantly smaller than 1 can be eliminated from the analysis.

Let us work through a simple example to see how Karhunen-Love picks out the highest signal-to-noise modes. The example is a simple two-pixel experiment with diagonal noise:

$$C_N = \begin{pmatrix} \sigma_n^2 & 0 \\ 0 & \sigma_n^2 \end{pmatrix}$$

The signal covariance matrix does have correlations between the two pixels so

$$C_S = \sigma_s^2 \begin{pmatrix} 1 & \epsilon \\ \epsilon & 1 \end{pmatrix}$$

where σ_s is the expected rms in the pixel and $-1 < \epsilon < 1$ measures how correlated the signal is between the two pixels. Steps 1 and 2 of the Karhunen-Love method are particularly simple since C_N is diagonal. Thus,

$$R_1 = I \quad ; \quad R_2 = \frac{1}{\sigma_n} I.$$

To complete step 3, we need to diagonalize

$$R_2 R_1 C_S R_1^T R_2 = \frac{\sigma_s^2}{\sigma_n^2} \begin{pmatrix} 1 & \epsilon \\ \epsilon & 1 \end{pmatrix}$$

To diagonalize the matrix in Eq. (11.77), we must solve

$$\frac{\sigma_s^2}{\sigma_n^2} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & \epsilon \\ \epsilon & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} C'_{S,11} & 0 \\ 0 & C'_{S,22} \end{pmatrix}$$

for the rotation angle θ. Carrying out the multiplication on the left side leads to

$$\frac{\sigma_s^2}{\sigma_n^2} \begin{pmatrix} 1 + \epsilon \sin(2\theta) & \epsilon \cos(2\theta) \\ \epsilon \cos(2\theta) & 1 - \epsilon \sin(2\theta) \end{pmatrix} = \begin{pmatrix} C'_{S.11} & 0 \\ 0 & C'_{S.22} \end{pmatrix}.$$

Equality in the off-diagonal elements holds if $\theta = \pi/4$, so the new theory covariance matrix is

$$C'_S = \frac{\sigma_s^2}{\sigma_n^2} \begin{pmatrix} 1 + \epsilon & 0 \\ 0 & 1 - \epsilon \end{pmatrix}$$

The rotation matrix which diagonalized the theory matrix is the first one on the left in Eq. (11.78) with $\theta = \pi/4$, so

$$R_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

The new modes are $\Delta' = R_3 R_2 \Delta$; explicitly, they are

$$\Delta_1' = \frac{(\Delta_1 + \Delta_2)}{\sqrt{2}\sigma_n}$$
$$\Delta_2' = \frac{(-\Delta_1 + \Delta_2)}{\sqrt{2}\sigma_n}.$$

The new modes (Eq. (11.82)) and their covariance matrix (Eq. (11.80)) are easy to understand if we consider the special cases of $\epsilon = 0$ and $\epsilon = +1$. If $\epsilon = 0$ the two modes have the same signal-to-noise, σ_s/σ_n. If the expected signal is large, these modes both carry information; if not, the noise swamps the signal. In either case, each mode - the sum and the difference - is equally (un)important. If $\epsilon = +1$ then the theory predicts the same signal in each pixel. In that case, the difference mode (Δ_2') is worthless, since only noise contributes to it. We see this from the fact that $(C_S')_{22}$ goes to zero as $\epsilon \to 1$. Its signal-to-noise is zero. The other mode - the sum of the two pixels - has signal-to-noise of $\sqrt{2}\sigma_s/\sigma_n$ since the two measurements beat down the noise by a factor of $\sqrt{2}$. This would of course have emerged from the full 2×2 likelihood analysis. But, using both modes in the analysis is a waste of time, a waste which is detected and obliterated by the Karhunen-Love method.

Bunn (1995) and Bond (1995) independently analyzed the COBE data by looking at Karhunen-Love modes. Figure 11.8 shows several such modes: clearly the ones with the highest signal-to-noise are the large-scale modes, indicating that COBE was sensitive to large-angle anisotropy. The smallest signal-to-noise modes are the small-scale modes which COBE did not have the resolution to measure.

Vogeley and Szalay (1996) first applied this technique to galaxy surveys. In this context, the Karhunen-Love method has another useful feature. Recall from Section 11.2.3 that on large scales, the signal covariance matrix is larger than Poisson noise, while on small scales the reverse is true. When we order the modes, then, large-scale modes will have the largest ratio of signal-to-noise. Thus the KarhunenLove basis will preferentially pick out large-scale modes. This is extremely useful because we are often most interested in eliminating small-scale modes - which are afflicted by nonlinearities and bias - from an analysis. The Karhunen-Love method does this automatically! An example is shown in Figure 11.9.

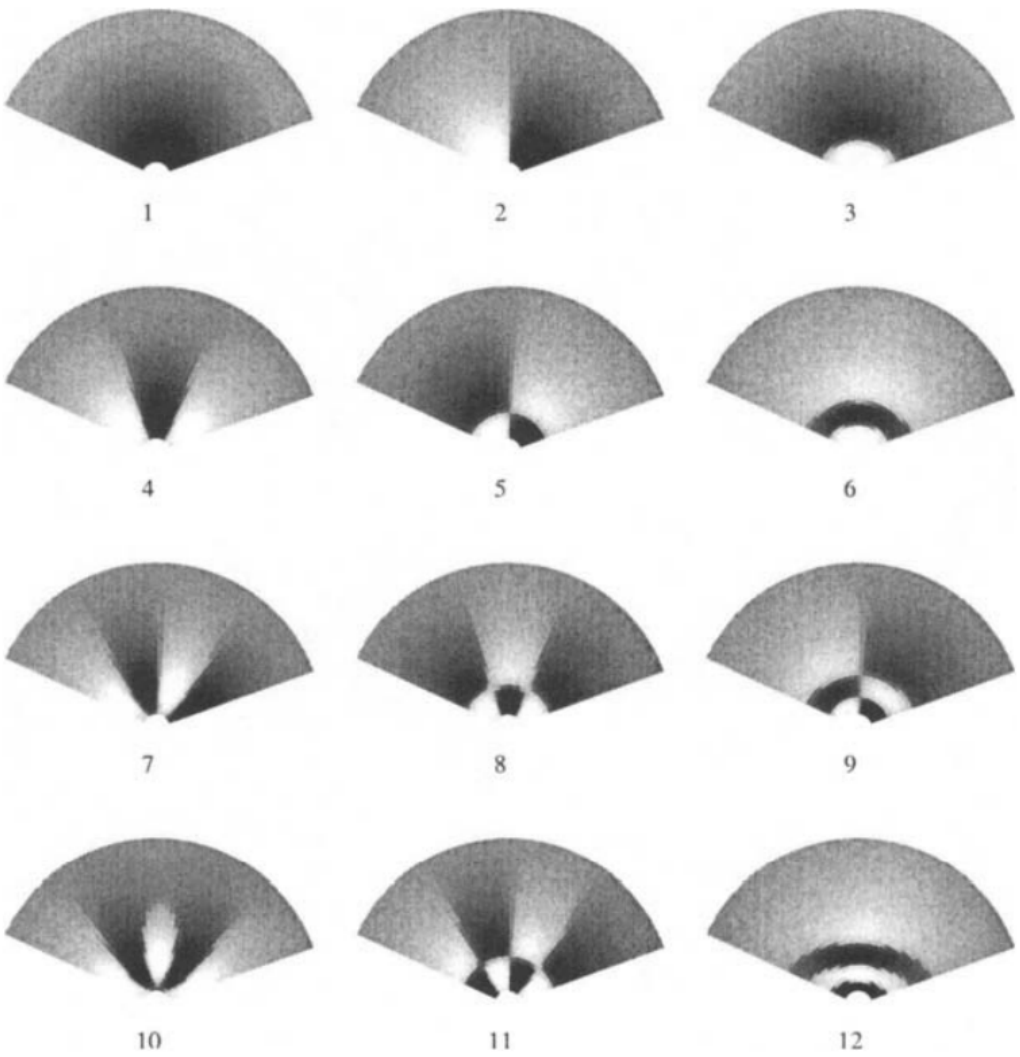


Figure 11.9. The 12 modes with highest signal-to-noise for the CFA2 survey (from Vogeley and Szalay, 1996). Modes pick out the large-scale structure of the galaxy distribution. In each case, we are at the bottom of the slice, and the top region is farthest from us.

For example, the first eigenmode roughly weights pixels only on the basis of their distance from us. It essentially takes the difference between the number of galaxies close to us and the number at moderate distances. The second mode takes a different component of the dipole, the difference between the number of galaxies on the left and the number on the right. Modes with lower weight take successively higher moments of the galaxy distribution.

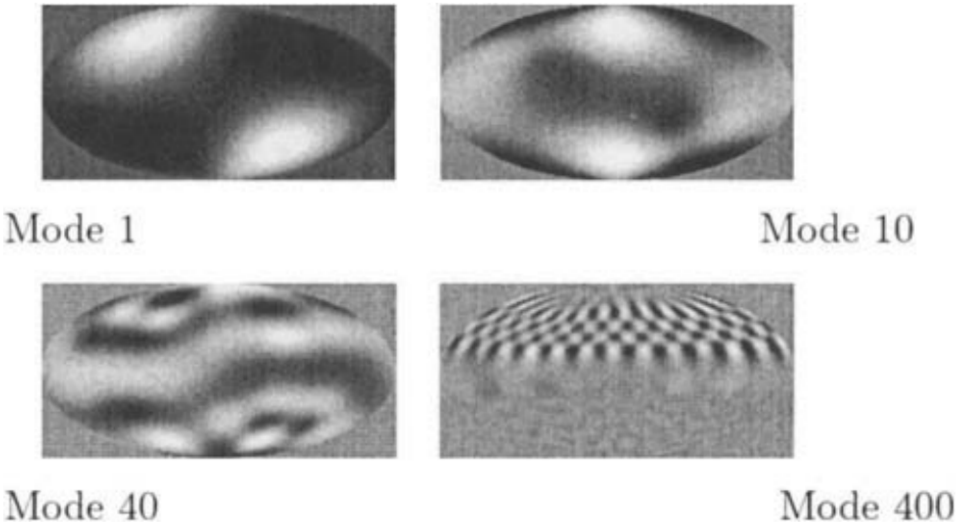


Figure 11.8. Modes of the COBE anisotropy experiment (from Bunn, 1995) ordered according to their signal-to-noise. The mode with the largest signal-to-noise (upper left) is sensitive predominantly to the quadrupole, while the mode with the smallest signal (bottom right) is sensitive to much smaller scale structure.

There are several drawbacks to the Karhunen-Love method. First, we need to assume a C_S to begin with, in order to identify the modes that are worthless. Although this might seem like a big problem - the modes that are thrown out for one choice of C_S could conceivably be important for another - people who have studied the issue assure us that it is not. They claim that the choice of the important modes is relatively insensitive to the input, assumed C_S. Another drawback is computational. Once the important modes are chosen, C'_S needs to be recalculated at many points in parameter space. In most of parameter space it will not be diagonal at all (it is only diagonal at the special point, the C_S that was chosen as the input spectrum initially). Thus at every point in parameter space, we still need to invert nondiagonal covariance matrices. This drawback is of course partially offset by the fact that - by virtue of the much smaller size of the matrices - the computation is now much faster. Nonetheless in many instances this is not enough to make the full computation managable. We must find still other ways of reducing the computational burden.

Before turning to one such way, let me mention one more use of the KarhunenLove method. It is an extremely useful consistency check. This is perhaps best illustrated with an example. The Python experiment (Coble et al., 1999) measured anisotropy over a large (for that time) area, with very low signal-to-noise. In principle, this is a good idea because the large area beats down cosmic variance. In practice, though, it presents challenges because it is difficult to check the consistency of the data. One typically wants to break up the data into several subsets and make sure that each subset sees the same signal. With Python, this was very difficult because noise dominated each subset. One way to check for consistency then is to work with Karhunen-Love modes. In this basis, each data point d'_i should be drawn from a Gaussian distribution with variance equal to $1 + C'_{S.ii}$. The histogram of $d'_i / \left(1 + C'_{S.ii}\right)^{1/2}$ should then look Gaussian. Figure 11.10 shows this histogram using a preliminary noise matrix. There were 650 measurements, but about 70 of these were eliminated for reasons that needn't concern us, so ignore the central spike at $d = 0$. If the remaining 580 data points were distributed as a Gaussian we would expect about two of them (0.3%) to have absolute value greater than 3, and none of them to have absolute value greater than 4. In fact, Figure 11.10 shows that nine modes are more than 4-sigma away from zero. The distribution is definitely not Gaussian with this noise model.

This analysis led the Python team to question the model they had constructed for the noise covariance. (Adjacent points were more correlated than they had allowed for.) Redoing the analysis with a new noise matrix led to the results in Figure 11.11. This technique for testing data quality has been used in a number of other venues, often identifying signs of trouble.

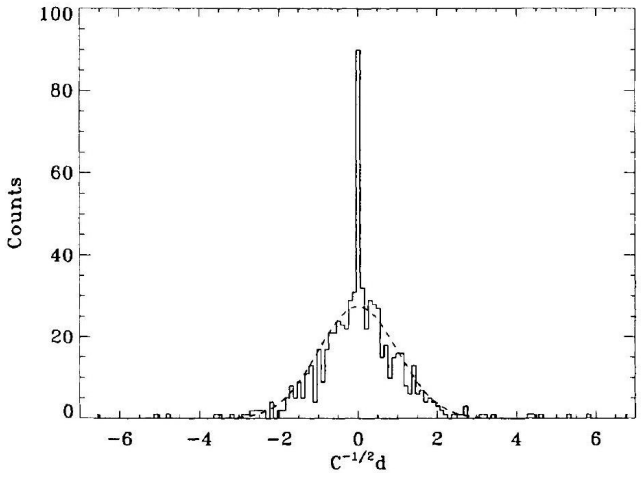


Figure 11.10. Histogram of data from one modulation of the Python CMB anisotropy measurement with a preliminary noise matrix. The data are in Karhunen-Love basis in which the covariance matrix is diagonal, so $C^{-1/2}d$ should be distributed as a Gaussian with variance equal to unity. The central spike should be ignored as 70 of the modes have been set to zero. The best fit Gaussian is the solid line. The counts are lower than the best-fit Gaussian in the central region, but above it in the tails.

Optimal Quadratic Estimator

One simple way of speeding up the likelihood calculation is to employ one of many successful root-finding algorithms. We are searching for the place where the likelihood function is a maximum, so we want to find

$$\left.\frac{\partial \mathcal{L}}{\partial \lambda}\right|_{\lambda=\bar{\lambda}} = 0$$

where for simplicity I've assumed that the likelihood function depends on only one parameter λ (we'll generalize this shortly) and $\bar{\lambda}$ is its value at the maximum of the likelihood.

An efficient way to find the root is to consider the derivative of the likelihood function evaluated at some trial point $\lambda = \lambda^{(0)}$. Expand this derivative around $\lambda^{(0)}$ in a Taylor expansion:

$$\mathcal{L}_{,\lambda}(\bar{\lambda}) = \mathcal{L}_{,\lambda}\left(\lambda^{(0)}\right) + \mathcal{L}_{,\lambda\lambda}\left(\lambda^{(0)}\right)\left(\bar{\lambda} - \lambda^{(0)}\right) + \ldots$$

where I have introduced the notation of writing partial derivatives as subscripted commas:

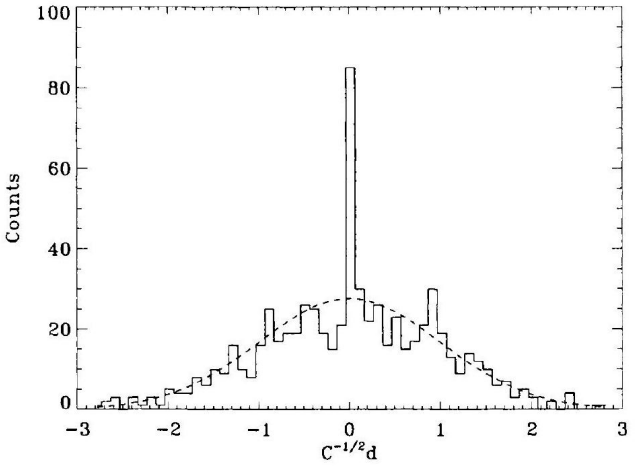


Figure 11.11. Same as Figure 11.10, but this time with an improved noise matrix. Note that there are no data points with absolute value greater than $3C^{1/2}$. The counts are consistent with a Gaussian distribution.

$$\mathcal{L}_{,\lambda} := \frac{\partial \mathcal{L}}{\partial \lambda} \quad ; \quad \mathcal{L}_{,\lambda\lambda} := \frac{\partial^2 \mathcal{L}}{\partial \lambda \partial \lambda}.$$

Since $\mathcal{L}$ is maximized at $\bar{\lambda}$, its derivative vanishes there; hence the right-hand side of Eq. (11.84) must also vanish. Setting it to zero leads to a simple expression for $\bar{\lambda}$:

$$\bar{\lambda} \simeq \lambda^{(0)} - \frac{\mathcal{L}_{,\lambda}\left(\lambda^{(0)}\right)}{\mathcal{L}_{,\lambda\lambda}\left(\lambda^{(0)}\right)}$$

where the $\simeq$ sign acknowledges that we have neglected higher order terms in the Taylor expansion in Eq. (11.84).

The solution in Eq. (11.86) assumes that the likelihood function is a quadratic function of the parameter λ. In fact, it is nothing of the sort: even in the simplest cases the likelihood function is not a quadratic function. For example, far from its maximum, $\mathcal{L}$ typically becomes exponentially small. A much better approximation therefore is that $\mathcal{L}$ is a Gaussian function, so that $\ln(\mathcal{L})$ is quadratic in λ. We can repeat the derivation above since the place where $\mathcal{L}$ is maximized is also the place where $\ln \mathcal{L}$ is a maximum. The estimator for λ is now

$$\hat{\lambda} = \lambda^{(0)} - \frac{(\ln \mathcal{L})_{,\lambda}\left(\lambda^{(0)}\right)}{(\ln \mathcal{L})_{,\lambda\lambda}\left(\lambda^{(0)}\right)}$$

Figure 11.12 illustrates the first iteration of this algorithm, called the NewtonRaphson method.

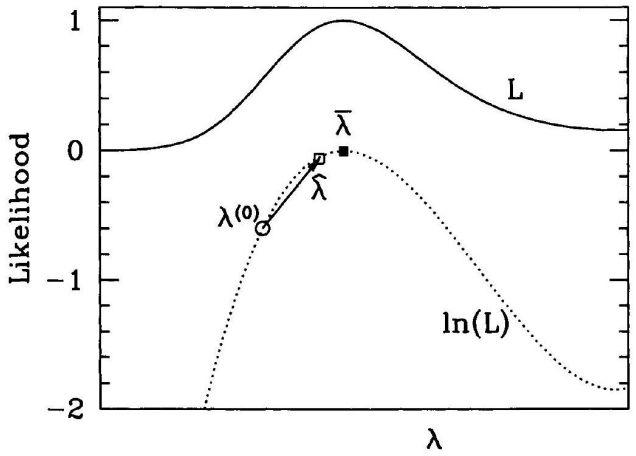


Figure 11.12. A sample likelihood function and its derivatives. The root finding alogrithm starts from the point $\lambda^{(0)}$ (open circle) and moves to $\hat{\lambda}$ (open square) which is quite close to the true maximum of the likelihood, $\bar{\lambda}$ (filled square). The Newton-Raphson technique does this by evaluating the first and second derivatives of $\ln(L)$ at the trial point $\lambda^{(0)}$. The method would not work as well if the derivatives of L were used, because L is not even approximately quadratic away from $\bar{\lambda}$.

To find an estimate for the best-fit value of λ, we need to calculate the derivatives in Eq. (11.87). In the case of the CMB, $\mathcal{L}$ is given explicitly by Eq. (11.20). Thus we need to differentiate the log of Eq. (11.20):

$$(\ln \mathcal{L})_{,\lambda} = \frac{\partial}{\partial \lambda}\left[-\frac{1}{2}\ln(\det C) - \frac{1}{2}\Delta C^{-1}\Delta\right].$$

The covariance matrix here C depends on the theoretical parameter, λ. We can use the identity $\ln\det(C) = \mathrm{Tr}\ln(C)$ and the fact that $C^{-1}_{,\lambda} = -C^{-1}C_{,\lambda}C^{-1}$ to get

$$(\ln \mathcal{L})_{,\lambda} = \frac{1}{2}\Delta C^{-1}C_{,\lambda}C^{-1}\Delta - \frac{1}{2}\,\mathrm{Tr}\left[\mathrm{C}^{-1}\mathrm{C}_{,\lambda}\right]$$

Here, the trace of $C^{-1}C_{,\lambda}$ is the sum of $(C^{-1})_{ij}\,\partial C_{ji}/\partial\lambda$ over all pixels i and j. Getting the second derivative requires more of the same types of steps. We find

$$(\ln \mathcal{L})_{,\lambda\lambda} = -\Delta C^{-1} C_{,\lambda} C^{-1} C_{,\lambda} C^{-1} \Delta + \frac{1}{2} \mathrm{Tr}\left[\mathrm{C}^{-1}\mathrm{C}_{,\lambda}\mathrm{C}^{-1}\mathrm{C}_{,\lambda}\right] + \frac{1}{2}\left(\Delta C^{-1} C_{,\lambda\lambda} C^{-1} \Delta - \mathrm{Tr}\left[\mathrm{C}^{-1}\mathrm{C}_{,\lambda\lambda}\right]\right).$$

Equation (11.90) gives the second derivative of $\ln \mathcal{L}$ with respect to the parameter λ. By definition, this is minus the curvature of the likelihood function:

$$\mathcal{F} = -\frac{\partial^2 \ln \mathcal{L}}{\partial \lambda^2}$$

The curvature is particularly important when evaluated at the maximum of the likelihood function, for there it measures how rapidly the likelihood falls away from the maximum (since the first derivative vanishes). If the curvature is small, then the likelihood changes slowly and the data are not very constraining: the resulting uncertainities on the parameter will be large. Conversely, large curvature translates into small uncertainties. [6]

We could take the ratio of Eqs. (11.89) and (11.90) to get an estimate for the maximum of the likelihood. This is not what is usually done. Rather, one typically sets $\Delta\Delta \to \langle \Delta\Delta \rangle = C$ in the second derivative. Upon doing this, the last line of Eq. (11.90) vanishes, and we are left with

$$\hat{\lambda} = \lambda^{(0)} + F^{-1}_{\lambda\lambda} \frac{\Delta C^{-1} C_{,\lambda} C^{-1} \Delta - \mathrm{Tr}\left[\mathrm{C}^{-1}\mathrm{C}_{,\lambda}\right]}{2}$$

where F is defined as

$$F_{\lambda\lambda} := \langle \mathcal{F} \rangle = \frac{1}{2} \mathrm{Tr}\left[\mathrm{C}_{,\lambda}\mathrm{C}^{-1}\mathrm{C}_{,\lambda}\mathrm{C}^{-1}\right]$$

That is, F is the average of the curvature over many realizations of signal and noise, both of which in this case are assumed to be drawn from Gaussian distributions. Here C and its derivatives are evaluated at the input point, $C\left(\lambda^{(0)}\right)$.

There are several important features about Eq. (11.92), our estimator for the value of λ which maximizes the likelihood function. As the title of this subsection promised, it is a quadratic estimator: it is of the general form $A\Delta^2 + B$. The only hard part was determining the coefficients which lead to the best algorithm for finding the root. In the spirit of the Newton-Raphson method, Eq. (11.92) is best used iteratively. One assumes an input spectrum, uses it to determine a new input parameter ($\hat{\lambda}$ in Eq. (11.92)), then uses the new input parameter to find a new best-fit value, and so on until the process converges. In practice, analysts have found that very few iterations are needed until convergence. Nonetheless, we must be somewhat wary of quadratic estimators. It is possible that they will lead us to local maxima in parameter space. Finally, the foregoing discussion assumes there

[6] The correspondence between the curvature matrix and parameter uncertainties is often more quantitative than this. See Numerical Recipes (Press et al., 1992), Chapter 15, for a detailed discussion. The bottom line, though, is that, in many cases, the inverse of the curvature matrix is a very good approximation to the error matrix on the parameters. E.g., in the one parameter case we are presently considering, the $1-\sigma, 68\%$ confidence level on λ is approximately equal to the inverse square root of the curvature matrix. is only one free parameter, but Eqs. (11.92) and (11.93) are easily generalizable to the more relevant case when many parameters are allowed to vary. If we have many parameters $\lambda_\alpha = \lambda_1, \lambda_2, \ldots$, then the quadratic estimator for each is

$$\hat{\lambda}_\alpha = \lambda^{(0)}_\alpha + F^{-1}_{\alpha\beta} \frac{\Delta C^{-1} C_{,\beta} C^{-1} \Delta - \mathrm{Tr}\left[C^{-1} C_{,\beta}\right]}{2}$$

where the Fisher matrix is defined as

$$F_{\alpha\beta} := \left\langle -\frac{\partial^2 (\ln \mathcal{L})}{\partial \lambda_\alpha \partial \lambda_\beta} \right\rangle = \frac{1}{2} \mathrm{Tr}\left[\mathrm{C}_{,\alpha}\mathrm{C}^{-1}\mathrm{C}_{,\beta}\mathrm{C}^{-1}\right].$$

Putting aside the details which led to Eq. (11.94), we can appreciate that the result is remarkable. We can now hope to find the values of the parameters which maximize the likelihood function without blindly covering the whole parameter space. A very small number of matrix manipulations suffice to determine these bestfit values. This is a huge advantage, too good to pass up considering the alternative.

We still need to find a way to evaluate the errors on the parameters. Had we evaluated $\mathcal{L}$ everywhere, we could easily identify the region in parameter space ruled out at, say, the 95% level. How do we identify such regions using the quadratic estimator? To answer this question we need to remove ourselves from the derivation above and simply notice that Eq. (11.94) is an estimator for the true best-fit values of the parameters. If we think of it in this way - rather than as the result of a rootfinding algorithm - we can study its distribution. Since the distributions for both signal and noise are known (they are assumed Gaussian with covariance matrices C_S and C_N, respectively), we can calculate the estimator's expectation value and variance.

First, let's consider its expectation value:

$$\left\langle \hat{\lambda}_\alpha \right\rangle = \lambda_\alpha^{(0)} + F_{\alpha\beta}^{-1} \frac{\langle \Delta C^{-1} C_{,\beta} C^{-1} \Delta \rangle - \mathrm{Tr}\left[\mathrm{C}^{-1}\mathrm{C}_{,\beta}\right]}{2}$$

Here the covariance matrix, its derivatives, and the Fisher matrix have all been evaluated at the trial point $\lambda_\alpha = \lambda_\alpha^{(0)}$. The expectation value $\langle \Delta\Delta \rangle$ on the other hand is equal to the true covariance matrix, $C\left(\bar{\lambda}_\alpha\right)$. We can expand $C\left(\bar{\lambda}_\alpha\right)$ about $\lambda_\alpha^{(0)}$:

$$C\left(\bar{\lambda}_\alpha\right) \simeq C + C_{,\alpha'}\left(\bar{\lambda}_{\alpha'} - \lambda_{\alpha'}^{(0)}\right)$$

Therefore the expectation value of the quadratic estimator is

$$\left\langle \hat{\lambda}_\alpha \right\rangle = \lambda_\alpha^{(0)} + \frac{1}{2} F_{\alpha\beta}^{-1} \left\{ \mathrm{Tr}\left[\mathrm{C}^{-1}\mathrm{C}_{,\beta}\mathrm{C}^{-1}\mathrm{C}\right] \right.$$

$$\left. + \mathrm{Tr}\left[\mathrm{C}^{-1}\mathrm{C}_{,\beta}\mathrm{C}^{-1}\mathrm{C}_{,\alpha'}\right]\left(\bar{\lambda}_{\alpha'} - \lambda_{\alpha'}^{(0)}\right) - \mathrm{Tr}\left[\mathrm{C}^{-1}\mathrm{C}_{,\beta}\right] \right\}$$

The first term in brackets cancels the third. The remaining trace is twice the Fisher matrix, so upon multiplying by F^{-1}, we are left with

$$\left\langle \hat{\lambda}_\alpha \right\rangle = \bar{\lambda}_\alpha$$

So the quadratic estimator we have been considering is unbiased: the expectation values of the set of $\hat{\lambda}_\alpha$ are equal to the true parameters $\bar{\lambda}_\alpha$, no matter what set of parameters are assumed at the outset.

We are also interested in the variance of the estimator:

$$\left\langle \left(\hat{\lambda}_\alpha - \bar{\lambda}_\alpha\right)\left(\hat{\lambda}_\beta - \bar{\lambda}_\beta\right) \right\rangle = \left(F^{-1}\right)_{\alpha\beta}$$

This equality, which I will leave as an exercise, holds if we are truly at the maximum of the likelihood function and if the data points really are distributed as a Gaussian. If these conditions hold, then the expected errors on the parameters are equal to the square root of the diagonal elements of F^{-1}. This is a magic limit, for there is a theorem, the Cramer-Rao inequality, that states that no method can measure the parameters with errors smaller than this (e.g., Kendall and Stuart, 1969). This makes sense since the errors from a full likelihood computation could not possibly be smaller than the width of the likelihood. This width in turn is determined by the curvature, and the Fisher matrix is simply the ensemble average of the curvature. Equation (11.100) tells us that on average, the quadratic estimator of Eq. (11.94) will reach this optimal limit.

Given any point in parameter space, we can calculate the associated Fisher matrix. Thus a simple way to assign error bars to the parameters determined via the quadratic estimator is to use the Fisher matrix evaluated at that point in parameter space. Bond, Jaffe, and Knox (1998), among others, have shown that this prescription works well: i.e., it agrees with a more complete tracing out of the likelihood contours.

Equation (11.100) is useful for other reasons as well. As is apparent from Eq. (11.95), the Fisher matrix - and hence the expected errors on any set of parameters - can be evaluated

without any data. It will serve us well in Section 11.4.3 when we set out to determine how well upcoming experiments will be able to determine parameters.

We have derived the quadratic estimator in a way which might lead you to believe that it is restricted to the CMB. Namely, our derivation assumed that the likelihood function is Gaussian, true for the CMB but not for galaxy surveys. Even without the assumption of a Gaussian likelihood function, though, the quadratic estimator of Eq. (11.94) can be applied to galaxy surveys. Like any quadratic estimator, it has a mean and a variance. We have just seen that, for Gaussian distributions, it has the lowest variance possible. On large scales, where the galaxy distribution is Gaussian, therefore, it is extremely relevant. Even on small scales, where nonlinearities add to the variance, it is often competitive with other, more traditional estimators.

26.1.4 the Fisher Matrix: Limits and Applications

The Fisher matrix plays a key role in describing the ability of a given experiment to constrain parameters. It is difficult to gain much insight, though, from the definition in Eq. (11.95). Fortunately, in the case of full sky coverage, the Fisher matrix can be computed analytically. This analytic computation can then be extended-via a plausibility argument - to the more realistic case of partial sky coverage. This calculation is presented next for both the CMB and galaxy surveys. The most popular use of the Fisher matrix is as a tool for forecasting. How well do we expect a given experiment (even a hypothetical one) to determine cosmological parameters? The Fisher matrix is ideally suited for this task, and we will see some startling expectations from upcoming experiments.

CMB

The trace in Eq. (11.95) is a sum of the diagonal elements of the matrix $[C_{,\alpha}C^{-1}C_{,\beta}C^{-1}]_{ij}$ where i, j index the pixels used in the map. There are two decisions that need to be made before the Fisher matrix can be computed. First, what pixelization scheme should we use, and second what parameters λ_α are we interested in? For a full-sky CMB experiment, we choose as our parameters the C_l 's themselves. That is, we take each individual C_l as a free parameter and ask how well an experiment can determine it. To avoid confusion (both the covariance matrix and the C_l 's are C 's), let's call each parameter λ_l instead of C_l, at least while working through the algebra. This answers the second question. The best way to deal with the first question - how to pixelize - in the case of the CMB is to use the a_{lm} 's. That is, instead of using the pixelized temperatures $\Theta(\hat{n})$, use

$$a_{lm} = \int d\Omega Y^*_{lm}(\hat{n})\Theta(\hat{n})$$

as the data values. Each pixel then is labeled by l and m, so a given row (or column) in the covariance matrix corresponds to a fixed valued of l and m. Explicitly, since we start with the quadrupole,

$$C = \begin{pmatrix} C_{l=2,m=-2;l'=2,m'=-2} & C_{2,-2;2,-1} & \cdots & C_{2,-2;2,2} & C_{2,-2;3,-3} & \cdots \\ C_{2,-1;2,-2} & C_{2,-1;2,-1} & \cdots & C_{2,-1;2,2} & C_{2,-1;3,-3} & \cdots \\ & & & \vdots & & \\ C_{3,-3;2,-2} & C_{3,-3;2,-1} & \cdots & C_{3,-3;2,2} & C_{3,-3;3,-3} & \cdots \\ & & & \vdots & & \end{pmatrix}.$$

As usual the covariance matrix is the sum of the signal and noise covariance matrices. From Eq. (8.63), the signal covariance matrix would be $\delta_{ll'}\delta_{mm'}\lambda_l$ (remember that we're using λ_l instead of C_l) if the window function were unity. Let's assume that the experiment measures the anisotropy with a beam size σ. Then the signal covariance matrix must be multiplied by $e^{-l^2\sigma^2}$. The noise covariance matrix is a little trickier. You will show in Exercise 11 that, in the case of uncorrelated, uniform noise, it is $\delta_{ll'}\delta_{mm'}w^{-1}$. Here w is the weight defined as

$$w = \left[(\Delta\Omega)\sigma_n^2\right]^{-1}$$

where $\Delta\Omega$ is the size in radians of the real space pixels and σ_n is the noise per pixel. Putting these two together, we have

$$C_{lm;l'm'} = \delta_{ll'}\delta_{mm'}\left[\lambda_l e^{-l^2\sigma^2} + w^{-1}\right]$$

With these simple assumptions, we can take the inverse of the covariance matrix C and also find its derivative with respect to the parameters, the λ_l 's. The inverse of the covariance matrix is

$$\left(C^{-1}\right)_{lm:l'm'} = \delta_{ll'}\delta_{mm'}\left[\lambda_l e^{-l^2\sigma^2} + w^{-1}\right]^{-1}$$

while the derivative of the covariance matrix with respect to the parameter λ_α is

$$C_{lm;l'm',\alpha} = \delta_{ll'}\delta_{mm'}\delta_{l\alpha}e^{-l^2\sigma^2}$$

We can now construct the Fisher matrix; the only difficult task will be keeping track of indices. Very explicitly,

$$\begin{aligned}
F_{\alpha\alpha'} &= \frac{1}{2}C_{lm;l'm',\alpha}C^{-1}_{l'm':l''m''}C_{l''m'':l'''m''',\alpha'}C^{-1}_{l'''m''':lm} \\
&= \frac{1}{2}\left(\delta_{ll'}\delta_{mm'}\delta_{l\alpha}e^{-l^2\sigma^2}\right)\left(\frac{\delta_{l'l''}\delta_{m'm''}}{\lambda_{l'}e^{-l'^2\sigma^2}+w^{-1}}\right)\left(\delta_{l''l'''}\delta_{m''m'''}\delta_{l''\alpha'}e^{-l''^2\sigma^2}\right) \\
&\times\left(\frac{\delta_{l'''l}\delta_{m'''m}}{\lambda_l e^{-l^2\sigma^2}+w^{-1}}\right)
\end{aligned}$$

with the implicit sum over $ll'l''l'''mm'm''m'''$. Consider first the Kronecker deltas with subscripts m, m', m'', and m'''. There are four of these; summing over all the subscripts besides m contracts these four to

$$\sum_{m'm''m'''}\delta_{mm'}\delta_{m'm''}\delta_{m''m'''}\delta_{m'''m} = \delta_{mm}.$$

Then, summing over all m leads to a factor of $2l+1$. The remaining sums over l and its cousins leads to a simple factor of $\delta_{\alpha\alpha'}$. Therefore in the all-sky limit, the Fisher matrix for a CMB experiment is

$$F_{ll'} = \frac{2l+1}{2}\delta_{ll'}e^{-2l^2\sigma^2}\left[C_l e^{-l^2\sigma^2} + w^{-1}\right]^{-2}$$

where I have gone back to C_l here since the covariance matrix does not appear. In an all-sky survey, therefore, the Fisher matrix for the C_l 's is diagonal. There are no correlations between adjacent C_l 's. The errors on a given C_l expected from an all-sky experiment can be read off from Eq. (11.109). The errors are equal to $\sqrt{F^{-1}}$, so

$$\delta C_l = \sqrt{\frac{2}{2l+1}}\left[C_l + w^{-1}e^{l^2\sigma^2}\right].$$

As anticipated in Eq. (11.26), there are thus two sources of error: (i) cosmic variance, proportional to the signal itself C_l and (ii) noise-atmospheric or instrumental as encoded in the weight w and the smoothing determined by the beam width σ. The factor of $2l+1$ in the denominator also traces back to Eq. (11.26); it is the number of independent samples used to estimate a given C_l.

No experiment will ever cover the entire sky, since the CMB cannot be observed in the plane of our galaxy. Even MAP and Planck, two satellites designed to map CMB anisotropy from space, will therefore cover a fraction of the sky $f_{\rm sky} < 1$. Recalling that the factor of $2l+1$ in the denominator of Eq. (11.26) counts the number of samples, we could guess that this factor must be multipled by $f_{\rm sky}$. This leads to

$$\delta C_l = \sqrt{\frac{2}{(2l+1)f_{\rm sky}}}\left[C_l + w^{-1}e^{l^2\sigma^2}\right].$$

This formula enables one to project the errors obtainable for any given experiment. The three characteristics of the experiment which determine the error on C_l are the sky coverage; the weight; and the beam width.

Galaxy Surveys

The analogue of the all-sky CMB experiment is a volume-limited galaxy survey as the volume gets arbitrarily large. This limit applies to all modes with wavelengths k^{-1} much smaller than the typical size of the survey. We have already computed the signal and noise covariance matrices in this limit, Eqs. (11.60) and (11.61), so the covariance matrix for Fourier pixels is

$$C_{\vec{k}_i,\vec{k}_j} = \frac{\delta_{ij}}{V}\left(P\left(k_i\right) + \frac{1}{\bar{n}}\right).$$

To compute the Fisher matrix, we will need the inverse of this and its derivative with respect to whatever parameters we choose. The inverse is simple,

$$C^{-1}_{\vec{k}_i,\vec{k}_j} = \frac{\delta_{ij}V}{P\left(k_i\right) + \frac{1}{n}}$$

For our parameters, we will choose the amplitude of the power spectrum in a set of narrow k-bins, each with width Δk. The power in the bin with $k_\alpha < k < k_\alpha + \Delta k$ will be denoted P_α. The derivative of the covariance matrix with respect to P_α is therefore

$$C_{ij,\alpha} := \frac{\partial C_{\vec{k}_i,\vec{k}_j}}{\partial P_\alpha} = \frac{\delta_{ij}}{V} d_{i\alpha}$$

with

$$d_{i\alpha} := \begin{cases} 1 & k_\alpha < \left|\vec{k}_i\right| < k_\alpha + \Delta k \\ 0 & \text{otherwise} \end{cases}.$$

The Fisher matrix is now

$$\begin{aligned} F_{\alpha\alpha'} &= \frac{1}{2} C_{ij,\alpha} C^{-1}_{jj'} C_{j'i',\alpha'} C^{-1}_{i'i} \\ &= \frac{1}{2}\delta_{ij} d_{i\alpha} \frac{\delta_{jj'}}{P\left(k_j\right) + \frac{1}{n}} \delta_{j'i'} d_{i'\alpha'} \frac{\delta_{i'i}}{P\left(k_i\right) + \frac{1}{n}}. \end{aligned}$$

The sums over j, j, i' are straightforward, leaving

$$F_{\alpha\alpha'} = \frac{1}{2} \frac{\sum_i d_{i\alpha} d_{i\alpha'}}{\left(P_\alpha + \frac{1}{n}\right)\left(P_{\alpha'} + \frac{1}{\bar{n}}\right)}.$$

As long as the k-bins do not overlap, k_i cannot be in two different bins so the product $d_{i\alpha} d_{i\alpha'}$ requires $\alpha = \alpha'$. The sum then is over all vectors $\vec{k}_i$ in a spherical shell with radius k_α and width Δk. This sum is $4\pi k_\alpha^2 \Delta k V$. Therefore, the Fisher matrix is diagonal, with elements

$$F_{\alpha\alpha'} = \delta_{\alpha\alpha'} \frac{4\pi k_\alpha^2 \Delta k V}{2\left(P_\alpha + \frac{1}{\bar{n}}\right)^2}.$$

The error on the power spectrum in this limit is the inverse square root,

$$\delta P_\alpha = \sqrt{\frac{2}{4\pi k_\alpha^2 \Delta k V}}\left(P_\alpha + \frac{1}{\bar{n}}\right).$$

This is identical in form to the errors on the C_l 's. The denominator in the prefactor counts the number of modes in a given estimate; the first term is cosmic variance; and the last is the noise, in this case Poisson noise.

Forecasting

One of the great promises of upcoming cosmological experiments is that they will determine many of the presently unknown cosmological parameters. How does one predict the expected uncertainties in cosmological parameters from future experiments? The answer is surprisingly simple. Let's consider a CMB experiment as an example. Start with the following:

- A set of C_l 's that are assumed to describe the true universe - The uncertainty on C_l from a given experiment, δC_l, assumed to be given by Eq. (11.111)
- The set of cosmological parameters, $\{\lambda_\alpha\}$, for which we want to forecast errors

The observed C_l 's in this universe will be close to the true C_l 's; indeed, if we form

$$\chi^2\left(\{\lambda_\alpha\}\right) = \sum_l \frac{\left(C_l\left(\{\lambda_\alpha\}\right) - C_l^{\rm obs}\right)^2}{\left(\delta C_l\right)^2},$$

then we expect this χ^2 to reach a minimum at the point in parameter space where $\lambda_\alpha = \bar{\lambda}_\alpha$, the actual values of the parameters. Of course, we do not now know what those values are, but even without that information, we can ask how quickly $\chi^2\left(\{\lambda_\alpha\}\right)$ changes as λ_α moves away from $\bar{\lambda}_\alpha$. If it increases rapidly, then the errors on the parameters will be very small; if the χ^2 changes little, then there will be large errors on the parameters.

To quantify this, we can expand χ^2 about its minimum at $\bar{\lambda}_\alpha$. Let's first do this in the case of one parameter; the generalization to many parameters will be straightforward. In the one-parameter case,

$$\chi^2(\lambda) = \chi^2(\bar{\lambda}) + \mathcal{F}(\lambda - \bar{\lambda})^2$$

The linear term in Eq. (11.121) vanishes since χ^2 is a minimum at $\bar{\lambda}$. The coefficient of the quadratic term is

$$\mathcal{F} = \frac{1}{2}\frac{\partial^2\chi^2}{\partial\lambda^2}\bigg|_{\lambda=\bar{\lambda}}$$

The curvature here, $\mathcal{F}$, measures how rapidly the χ^2 changes away from its minimum. As such, the error on λ is simply $1/\sqrt{\mathcal{F}}$. So all we need to do in order to determine the expected errors on a parameter is calculate $\mathcal{F}$. Note that $\mathcal{F}$ is the curvature of the likelihood function only if the likelihood function is equal to $e^{-\chi^2/2}$, that is, if the errors on the C_l 's are Gaussian distributed. In fact, they are not, so $\mathcal{F}$ as given by Eq. (11.122) is not really the curvature, $-\partial^2 \ln\mathcal{L}/\partial\lambda^2$. Nonetheless, the distribution is close enough to Gaussian that the error estimates that arise are expected to be quite accurate.

The second derivative of χ^2 contains two terms:

$$\mathcal{F} = \sum_l \frac{1}{\left(\delta C_l\right)^2}\left[\left(\frac{\partial C_l}{\partial\lambda}\right)^2 + \left(C_l - C_l^{\rm obs}\right)\frac{\partial^2 C_l}{\partial\lambda^2}\right]$$

The second term in the sum over l is traditionally neglected. The idea (as elucidated in Numerical Recipes) is that the difference $C_l - C_l^{\rm obs}$ will sometimes be negative, sometimes positive. On average, there will be much cancellation, so the first term will dominate. Thus, the general practice is to take

$$\mathcal{F} \to \sum_l \frac{1}{\left(\delta C_l\right)^2}\frac{\partial C_l}{\partial\lambda}\frac{\partial C_l}{\partial\lambda}$$

Equivalently, you can think of the dropping of the second term as taking an average over the whole distribution, thereby replacing the curvature matrix with the Fisher matrix (although again keeping in mind that this is not the true curvature or Fisher matrix since the C_l 's are not distributed as Gaussians). The generalization of this to many parameters is simply

$$F_{\alpha\beta} = \sum_l \frac{1}{(\delta C_l)^2} \frac{\partial C_l}{\partial \lambda_\alpha} \frac{\partial C_l}{\partial \lambda_\beta}$$

In order to predict how accurately parameters will be known, then, we simply need to know the experiment's specifications (to determine δC_l) and the derivatives of the C_l 's around their (assumed) true values.

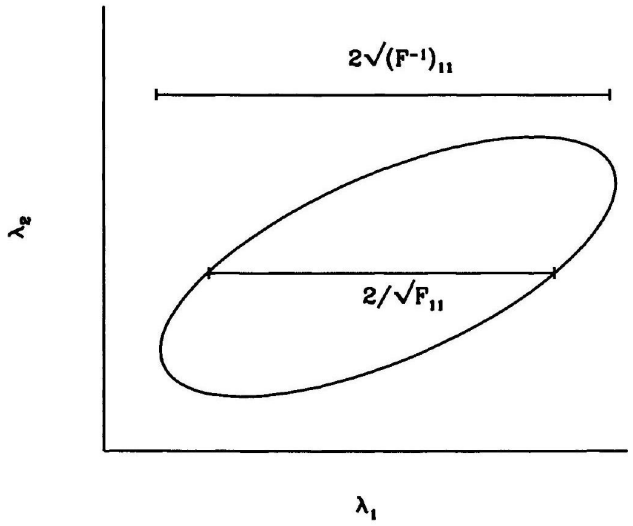


Figure 11.13. Error ellipse in a 2D parameter space. If no prior information is known about λ_2, then the error on λ_1 is $\sqrt{(F^{-1})_{11}}$. If, however, λ_2 is fixed, the error on $\lambda_1, (F_{11})^{-1/2}$ is smaller.

Assuming a Gaussian distribution, the one-sigma uncertainty on a one parameter fit is $1/\sqrt{F}$. How about if more than one parameter is allowed to vary? Figure 11.13 illustrates the situation in a two-dimensional setting. If the parameter λ_2 is assumed known, then the error on λ_1 is still $1/\sqrt{F_{11}}$. However, if λ_2 is allowed to vary, the error on λ_1 is now $\sqrt{(F^{-1})_{11}}$. It is instructive to prove this explicitly by noting that we are assuming that the joint probability for the two parameters is

$$P(\lambda_1, \lambda_2) \propto \exp\left\{-\frac{1}{2}\lambda_i F_{ij} \lambda_j\right\}$$

where I have assumed that the distribution peaks at $\lambda_i = 0$ for simplicity. Allowing λ_2 to vary is equivalent to integrating this probability distribution over all possible values of λ_2. This is referred to as marginalizing over λ_2. Then,

$$\begin{aligned} P(\lambda_1) &= \int d\lambda_2 P(\lambda_1, \lambda_2) \\ &\propto \exp\left\{-\frac{\lambda_1^2}{2}\left(\frac{F_{11}F_{22} - F_{12}F_{21}}{F_{22}}\right)\right\} \end{aligned}$$

where the second line comes from carrying out the λ_2 integration explicitly. The term in parentheses in the exponential- $-\left[F_{11}F_{22} - F_{12}F_{21}\right]/F_{22}$ - is simply equal to $1/\left(F^{-1}\right)_{11}$. So the one-sigma error is indeed given by $\sqrt{(F^{-1})_{11}}$.

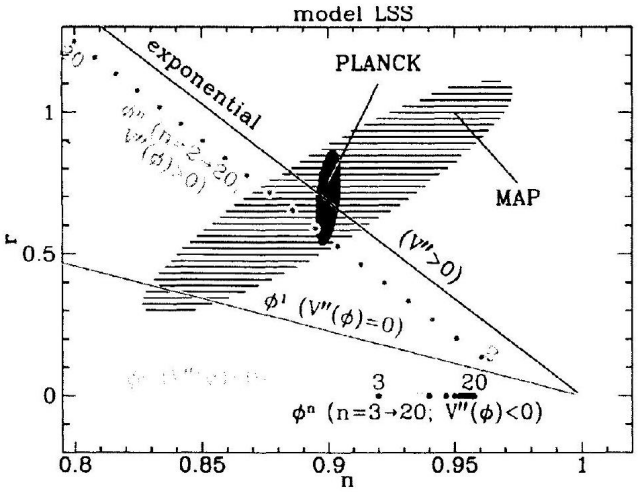


Figure 11.14. Expected 95% uncertainty on the inflationary parameters n and r from MAP and Planck (from Dodelson, Kinney, and Kolb, 1997). (See color Plate 11.14.) Three other parameters (normalization, Ω_B, and h) have been marginalized over. Every inflationary model gives a unique prediction somewhere in this plane; many such predictions are plotted.

Figure 11.14 shows the expected uncertainties from MAP and Planck for two inflationary parameters, those which determine the primordial spectrum. Note that I have fixed the "true" model to be one in which the spectral index $n = 0.9$ and the tensor-to-scalar ratio $r = 0.7$. Different fiducial models often lead to quite different error bars. The ellipse in Figure 11.14 has been drawn after marginalizing over Table 11.1. Marginalized Errors for Λ CDM for various experiments.

	Map	Planck	Map+SDSS	Planck+SDSS
h	0.22	0.13	0.029	0.022
Ω_m	0.24	0.14	0.036	0.027
Ω_Λ	0.19	0.11	0.042	0.024
$\ln(\Omega_b h^2)$	0.060	0.010	0.050	0.010
m_ν(eV)	0.58	0.26	0.33	0.21
Y_P	0.020	0.013	0.020	0.013
n	0.048	0.008	0.040	0.008
r	0.18	0.012	0.16	0.012
τ	0.022	0.004	0.021	0.004

three other cosmological parameters: normalization, baryon density, and Hubble constant. To do this, one starts with the five-dimensional Fisher matrix, inverts, and then considers only the 2×2 part of the inverse. This 2×2 part defines the ellipses drawn in the figure.

Also plotted in Figure 11.14 are the predictions from a wide variety of inflationary models. The figure argues persuasively that we will indeed be able to distinguish among different inflationary models with upcoming CMB experiments. This is a remarkable finding: we will learn about physics at 10^{15}GeV or higher using CMB data. If more parameters are allowed to vary, then the errors on any one parameter naturally get larger. However, this can be offset by using other observations, most notably those from large-scale structure. Table 11.1 presents the marginalized errors on a number of parameters expected from the MAP, Planck, and Sloan Digital Sky Survey experiments (Eisenstein, Hu, and Tegmark, 1999). Here the errors include the measurement of polarization in MAP and more importantly Planck.

26.1.5 Mapmaking and Inversion

Until now, we have discussed ways of analyzing a map, a collection of pixels with data Δ and noise covariance matrix C_N. How does one make such a map? How do we go from the timestream of data to a pixelized set of spatial Δ 's? Most work on this issue has focused on the temperature maps in the CMB, so I'll use this as an example. We will see that mapmaking is essentially an inversion problem. So the mapmaking techniques discussed here are applicable to a broad range of problems in physics and astronomy.

Let's first state the problem. An experiment amounts to a timestream of data, d_t. Each number in the timestream corresponds to data taken at a given point on the sky. The data are assumed to be composed of signal plus noise:

$$d_t = P_{ti} s_i + \eta_t$$

Here subscript t denotes a given element in the timestream; i denotes spatial pixel; P is the pointing matrix; s is the underlying temporally constant, but spatially varying signal; and η is the temporal noise. The pointing matrix encodes information about where the receiver is pointing. That is, it associates with each temporal measurement t a given pixel i. It is an $N_t \times N_p$ matrix where N_t is the number of temporal measurements and N_p is the number of spatial pixels. The pointing matrix has a special form: every row has only one nonzero entry equal to 1, corresponding to the pixel on the sky being observed at the time denoted by the row. Each column, however, typically has many nonzero entries corresponding to all the times a given spot has been observed. The noise η is assumed to be Gaussian with a covariance matrix N. There are techniques to determine N directly from the data, but to simplify the discussion, we will assume that N is known.

What is the best way to make a map from this timestream? One forms a χ^2,

$$\chi^2 := \sum_{tt'ij} \left(d_t - P_{ti} s_i\right) N^{-1}_{tt'} \left(d_{t'} - P_{t'j} s_j\right)$$

and minimizes with respect to the signal s. Indeed, if the noise is Gaussian, then the likelihood function is proportional to $e^{-\chi^2/2}$ and minimizing the χ^2 is equivalent to maximizing $\mathcal{C}$. Taking the derivative with respect to s_i leads to

$$\frac{\partial \chi^2}{\partial s_i} = -2 \sum_{tt'j} P_{ti} N^{-1}_{tt'} \left(d_{t'} - P_{t'j} s_j\right)$$

Set the derivative equal to zero:

$$\sum_{tt'j} P_{ti} N^{-1}_{tt'} P_{t'j} s_j = \sum_{tt'j} P_{ti} N^{-1}_{tt'} d_{t'}$$

The terms multiplying s_j on the left are an $N_p \times N_p$ matrix,

$$\left(C_N^{-1}\right)_{ij} := \sum_{tt'} P_{ti} N^{-1}_{tt'} P_{t'j}$$

Multiply both sides by the inverse of this (C_N itself) to find that the χ^2 is minimized when s is equal to

$$\Delta_i = C_{N,ij} P_{ti} N^{-1}_{tt'} d_{t'}$$

In matrix notation,

$$\Delta = C_N P^T N^{-1} d$$

where ^{T} denotes transpose. The noise matrix of this map is equal to

$$C_N = \left(P^T N^{-1} P\right)^{-1}$$

a fact which you can verify by taking $\langle \Delta\Delta \rangle$. A simple limit of Eq. (11.134) emerges when the timestream noise is diagonal and uniform (this is unrealistic). In that case, the elements of C_N become

$$C_{N.ij} \to N \left(\sum_t P^T_{it} P_{tj} \right)^{-1}$$

where now N is simply a number, the diagonal element of the timestream noise. Recall that for a given t, P_{ti} is nonzero for only one pixel i. So the product $P^T_{it} P_{tj}$ vanishes unless $i = j$ and the receiver at time t was pointing at pixel i. Thus the sum over t counts the number

of times the receiver sampled pixel i; call this number π_i. In this artificial case of uniform, uncorrelated noise, therefore, the noise covariance matrix for the map is diagonal with elements N/π_i. This makes sense: as a given pixel is sampled more times, the standard deviation goes down as $\pi_i^{-1/2}$. The map now becomes

$$\Delta_i \to \frac{1}{\pi_i} \sum_t P_{it}^T d_t$$

That is, one simply averages all the data points corresponding to the given pixel (exactly like Eq. (11.8))

Figure 11.15 shows a more realistic implementation of Eq. (11.134), from the long-duration balloon flight of Boomerang (Netterfield et al., 2001), launched on December 29,1998. The map covers a region a region of about 700 square degrees with 7 -arcminute pixels. Therefore, the map required $N_p \simeq 50,000$, while the timestream contained of order 2×10^8 data points. Some tricks were needed to avoid direct inversion and multiplication of the large matrices P and N. Nonetheless, the basis for the Boomerang map, indeed for all CMB maps, is Eq. (11.134).

The raw data from which the map is made need not be the timestream. Instead, the raw data could consist of a series of modulations, e.g., the difference between the temperature at two points. Reconstructing a map from a set of modulations sounds like a much different problem than doing the same from the timestream. In fact, it is mathematically identical: the data d is the sum of a signal and noise. The signal can be thought of as a matrix acting on the underlying temperature field. This matrix does not have the exact form as the pointing matrix (i.e., only one nonzero element in each row), but it is a matrix nonetheless, and all the operations in Eqs. (11.134) and (11.135) can be carried out. There is a big advantage in using a map constructed in this fashion as opposed to the modulated data themselves. Ultimately, the main use of the data will be to estimate parameters in a likelihood analysis. As we have seen, one must construct the signal covariance matrix in order to carry out such an analysis. The signal covariance matrix for a map is extremely simple - the window function is simply $P_l(\cos\theta_{ij})$ - whereas that for modulated data is quite difficult to obtain (recall Section 11.2).

Equation (11.134) and the corresponding noise matrix in Eq. (11.135) are even more general than this. They apply to any problem in which the data are a sum of some processed signal and noise, i.e., to an extraordinarily wide range of problems in physics and astronomy. Consider just two examples. First, the angular correlation

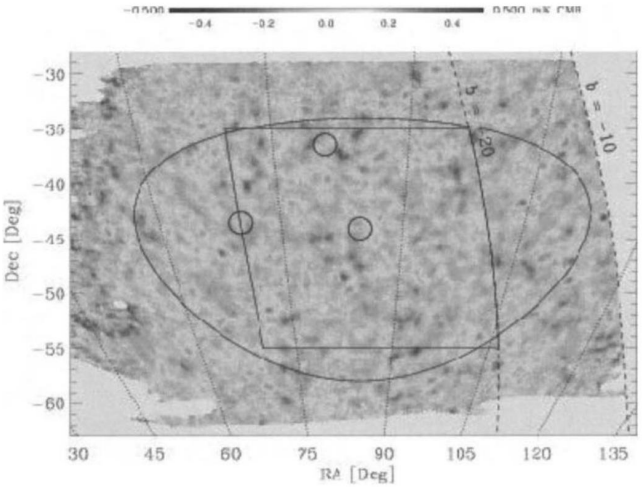


Figure 11.15. A map of the CMB temperature from observations by Boomerang (Netterfield et al., 2001), a long-duration balloon flight at the South Pole. (See also color Plate 11.15.) Hot and cold spots have amplitudes as large as 500μK. Circles shows quasars identified in these radio observations. The large elliptical region delineates data analyzed to obtain bandpowers. The rectangular region is an earlier data set.

function is a sum of the 3D power spectrum processed by a kernel plus noise. One can apply Eq. (11.134) directly to obtain the 3D power spectrum, simply replacing the pointing matrix with the kernel. In the next section, we will see another example, the extraction of the CMB signal from data contaminated by foregrounds.

26.1.6 Systematics

A systematic error is one which remains even after averaging over many data samples. Systematic errors are the most worrisome aspect of most cosmological observations. Knowing this, observers typically take many precautions against them, submitting raw data to a wide variety of consistency checks. Many of these tests are the result of common sense and intuition. Nothing much formal can be said about them. Here I want to focus on several... systematic ways of dealing with such effects.

Foregrounds

One of the biggest obstacles to observing the anisotropies in the CMB are foregrounds, other sources of radiation which also emit at microwave frequencies. The list of foregrounds is long and includes anything in space that might come between us and the radiation left over from the Big Bang. There is dust, synchrotron radiation, and free-free or bremsstrahlung emission, all emanating from our galaxy (but extending to regions of the sky far from the galactic plane). There are also extragalactic sources of radiation, point sources and clusters of galaxies. All of these have the potential to contaminate an experiment searching only for a cosmic signal. The magnitude of this problem is hinted at in the nomenclature. Usually in science, a possible source of systematic error is called a background. In CMB physics, we cannot call these things backgrounds: the "B" in CMB precludes that possibility. We must acknowledge that the cosmic signal is coming from farther away than any possible contaminant and we must deal with the real possibility that the cosmic signal will be smaller than some of these contaminants.

The problem of foregrounds has in the past few years been demonstrated to be manageable. There are a number of reasons for this good fortune (it is good fortune: if we were living deeper in the galaxy foreground amplitudes would be considerably larger). First observers have been very successful at finding the coolest portions of the sky and using only these regions. Second, foreground amplitudes have proven to be smaller than the cosmic signal in a fairly wide part of frequency space. Figure 11.16 shows the intensity of several galactic foregrounds and the CMB. The amplitudes of each of these components varies across the sky, but the relative amplitudes shown in Figure 11.16 are fairly typical. At very high frequencies, dust dominates, and at very low frequencies synchrotron and bremsstrahlung become important. But, in the range from 30 to 200GHz, the CMB anisotropies often have the largest intensities.

The final reason foregrounds can be managed - and the one I want to focus on here - can also be gleaned from Figure 11.16. The spectral shapes of the foregrounds are all different, different from one another and from the blackbody shape of the CMB anisotropy. This raises the possibility that detections at different frequencies can be used to extract the CMB signal from the foregrounds. As analysts, we must find the optimal way to perform this extraction. Given measurements at several different frequencies, what is the best algorithm for finding the cosmic signal? How effective do we expect this extraction to be?

First we need to set up some notation. Instead of intensity (or brightness) B, which has units of ergs $\mathrm{cm}^{-2}\mathrm{sec}^{-1}$ Hz^{-1}, it is convenient to introduce the brightness temperature or antenna temperature, defined as

$$\begin{aligned} T_{\mathrm{ant}} &:= \frac{B}{2\nu^2} \\ &= 2\pi\nu f, \end{aligned}$$

which has dimensions of kelvins (recall that we are working in units with $k_B = \hbar = c = 1$). The frequency ν is related to the momentum we used in earlier chapters via $p = 2\pi\nu$ and f is the familiar phase space density.

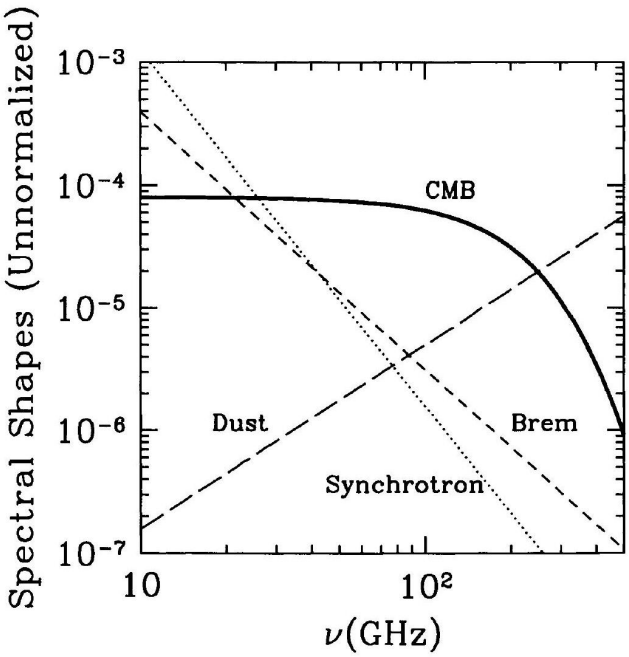


Figure 11.16. Spectral shapes of the dominant galactic foregrounds and the CMB blackbody temperature anisotropy.

For the CMB we know that f is given by Eq. (4.35), so the antenna temperature is

$$\frac{T_{\rm ant}^{\rm cmb}}{T} = \frac{x}{e^x - 1} + \Theta \frac{x^2 e^x}{(e^x - 1)^2}$$

where

$$x := \frac{p}{T} = \frac{2\pi\nu}{T}$$

The first term in Eq. (11.139) is the monopole, which does not interest us. The second contains information about the shape of the spectrum of the CMB anisotropies. It is useful then to neglect the first term and write

$$\frac{T_{\rm ant}^{\rm cmb}(\nu)}{T} = \Theta W^{\rm cmb}(2\pi\nu/T)$$

where W^α will be the shape vector for different components. That for the CMB is

$$W^{\rm cmb} = \frac{x^2 e^x}{(e^x - 1)^2}$$

Every component can be written in this fashion: the product of an amplitude (in this case Θ), which is frequency independent, and the shape vector, which carries no information about the amplitude. For a blackbody distribution like the CMB, the amplitude has a name: Θ is called the thermodynamic temperature. Note that the CMB shape vector goes to 1 in the limit of small frequencies, the Rayleigh-Jeans limit. At high frequencies, it falls off exponentially.

I have called W a shape vector implying that it has a number of components. These components are the different frequency channels at which a given experiment takes data. We will label these with a subscript c while allowing for many possible foreground components:

$$\vec{W}^\alpha = W_c^\alpha \quad c = 1, \ldots, N_{ch} \quad ; \quad \alpha = 0, \ldots, N_{\rm foregrounds}$$

Here the CMB component is associated with index $\alpha = 0$. Thus, the data from a given experiment in a given pixel (we focus on only one spatial pixel) on the sky is in the form of a set of antenna temperatures, d_c, at all the different frequency channels. Our model is that this data set is the sum of the contributions from the CMB, foregrounds, and Gaussian noise:

$$d_c = \sum_{\alpha=0}^{N_{\rm foregrounds}} W_c^\alpha \Theta^\alpha + n_c$$

We assume that we know the covariance matrix of the noise N and all the spectral shapes [7]W_c^α. The question is, how do we best determine the CMB anisotropy?

This problem has the exact same form as does the general inversion problem of Section 11.5. We want to obtain estimates of the amplitude Θ^α. We can immediately write down the minimum variance estimator for Θ^α:

$$\Delta^\alpha = (C_N)_{\alpha\beta} W_c^\beta N_{cd}^{-1} d_d$$

with covariance matrix

$$\left(C_N^{-1}\right)_{\alpha\beta} = W_c^\alpha \left(N^{-1}\right)_{cd} W_d^\beta$$

Let's work out a simple example to bring these formulae, which are so ubiquitous, to life.

7This assumption is true for the CMB, and very nearly true for synchrotron and bremsstrahlung which have shapes which are fairly constant over the sky. The assumption is worst for dust. A number of groups have explored the consequences of incorrect assumptions about the shapes or allowed for some freedom in the shapes. Consider an experiment taking measurements at two frequencies, both in the Rayleigh-Jeans regime. The shape vector for the CMIB then has two components, both equal to 1:

$$\vec{W}^0 = (1, 1)$$

Let's also assume that the noise is uncorrelated from one frequency channel to the next and is uniform with diagonal element, σ_n^2. First let's consider the case of zero foregrounds. In that case, the covariance matrix is just a single number, the inverse of

$$\begin{aligned} WN^{-1}W &\to \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1/\sigma_n^2 & 0 \\ 0 & 1/\sigma_n^2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \frac{2}{\sigma_n^2}. \end{aligned}$$

The inverse square root of this is the noise, $\sigma_n/\sqrt{2}$. The two channels reduce the noise by a factor of the square root of 2 (if there were three channels in our example, the factor would be $\sqrt{3}$, etc.). The mimimum variance estimator is given by Eq. (11.145),

$$\begin{aligned} \Delta &\to \frac{\sigma_n^2}{2}(1,1) \begin{pmatrix} 1/\sigma_n^2 & 0 \\ 0 & 1/\sigma_n^2 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \\ &= \frac{d_1 + d_2}{2}. \end{aligned}$$

We simply average the two data points.

Now suppose there is one foreground to worry about, say synchrotron emission, with shape vector

$$\vec{W}^1 = (1, 1/2)$$

Typically, the intensity of synchrotron emission scales as ν^{-1} (see, e.g., Rybicki and Lightman, 1979), so its antenna temperature falls off as ν^{-3}. Thus a shape vector $(1, 1/2)$ follows from observing at, e.g., $\nu_1 = 20$GHz and $\nu_2 = 25$GHz.

Now the covariance matrix C_N is a two by two matrix. Then Eq. (11.146) becomes

$$\begin{aligned} C_N^{-1} = W_c^\alpha \left(N^{-1}\right)_{cd} W_d^\beta &= \begin{pmatrix} 1 & 1 \\ 1 & 1/2 \end{pmatrix} \begin{pmatrix} 1/\sigma_n^2 & 0 \\ 0 & 1/\sigma_n^2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1/2 \end{pmatrix} \\ &= \frac{1}{\sigma_n^2} \begin{pmatrix} 2 & 3/2 \\ 3/2 & 5/4 \end{pmatrix}. \end{aligned}$$

Already at this stage, we can glean some important information about the experiment. Recall from the discussion in Section 11.4 that the inverses of the diagonal elements of C_N^{-1} are the unmarginalized variances, the errors if all other parameters are known. In this case, there is one parameter besides the amplitude of the CMB, the amplitude of the foreground. If we assume that is known, then the error on the CMB temperature will be the inverse square root of the oo component of the matrix in Eq. (11.151). This is $\sigma_n/\sqrt{2}$, in agreement with the calculation above. To find the errors if we know nothing about the foreground amplitudes, take the inverse of this to find the covariance matrix,

$$C_N = 4\sigma_n^2 \begin{pmatrix} 5/4 & -3/2 \\ -3/2 & 2 \end{pmatrix}$$

The oo component of this gives the marginalized variance, $5\sigma_n^2$. The ratio of the marginalized error to the unmarginalized error is a measure of how severely the unknown foregrounds degrade our ability to determine the CMB temperature. It is called the foreground degradation factor, or simply the FDF. In this case, the FDF is equal to $\sqrt{10}$.

In this example, we can now determine the minimum variance estimator for the CMB temperature. Following Eq. (11.145), we write

$$\begin{aligned} \Delta^0 &= 4\sigma_n^2(5/4, -3/2) \begin{pmatrix} 1 & 1 \\ 1 & 1/2 \end{pmatrix} \begin{pmatrix} 1/\sigma_n^2 & 0 \\ 0 & 1/\sigma_n^2 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \\ &= -d_1 + 2d_2. \end{aligned}$$

This should be no surprise. The best estimator for the CMB temperature is completely insensitive to the amplitude of the foreground component. For, if the foreground really does have shape vector $(1, 1/2)$, the linear combination $-d_1 + 2d_2$ from the foreground is equal to zero.

In real life, one must find the minimum variance estimators at many different spatial pixels. The formula of Eq. (11.145) remains identical in this more general case: one simply adds an index for spatial pixel. It is often most convenient to work with the a_{lm} 's instead of the temperature as a function of angular coordinates. Then, the minimum variance estimator for a_{lm} often weights the different frequency channels differently depending upon how the noise in each channels scales with l. An example is shown in Figure 11.17, based on the five frequencies of the MAP experiment. The figure shows how the best estimator for a_{lm} weighs the five different frequencies in the absence of foregrounds. At low l, all the channels have similar noise, so the best estimator is just the average of the five. The beam size is frequency dependent, however, largest at the lowest frequencies. Therefore, at high l, the lowest channels have no signal. Only the highest frequency channel can be used. Indeed, one sees that the minimum variance estimator gradually drops a channel at a time as l gets larger.

Often prior information on the foregrounds exists, in the form of an estimate of the power spectrum of each foreground component, i.e., its C_l. This prior information can be incorporated into the minimum variance estimate; see, e.g.,

D-15.

Figure 11.18 shows the results of accounting for foregrounds in the MAP experiment. The difference between the weighting scheme in this figure and that in Figure

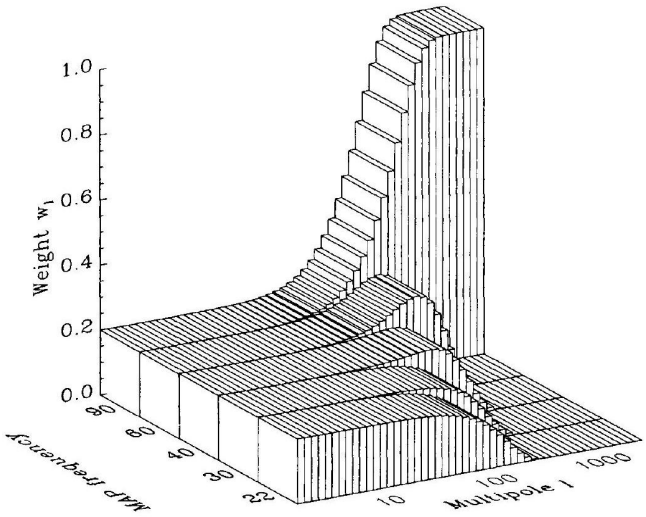


Figure 11.17. The minimum variance linear combination of the frequency channels of MAP in the absence of foregrounds (from Tegmark et al., 2000). The noise at low l is identical in all channels, so the minimum variance estimator weights them equally. At high l, the highest frequency channel has the lowest noise, so the best estimator uses only that channel.

11.17 is striking, especially at low l. No longer does one weight all the different channels identically. Rather, a complicated set of weights must be used to project out the foreground contamination.

Note that foregrounds do indeed fit the definition of a systematic effect. If one neglected synchrotron emission in the previous example, the naive estimate of the CMB temperature $(d_1 + d_2)$ would be wrong no matter how small the noise. You might argue that, over the whole sky, the average "wrongness" would cancel out, since there are an equal number of hot and cold spots of foregrounds. The power spectrum, though, the key quantity of interest, would be contaminated: it would be the sum of $C_l^{\rm CMB}$ and $C_l^{\rm synch}$, again even if there was no noise.

Mode Subtraction

A common problem in cosmological observations is that a particular mode is contaminated by some external source. An example might be a region of space very close to our galaxy, where foregrounds are particularly important. In a galaxy survey, it might be a dusty region. Dust tends to absorb high-frequency light, and so redden an object's image. This leads to less flux in blue bands, often pushing a galaxy outside the flux limits of a survey. This push can be crucial, because typ-

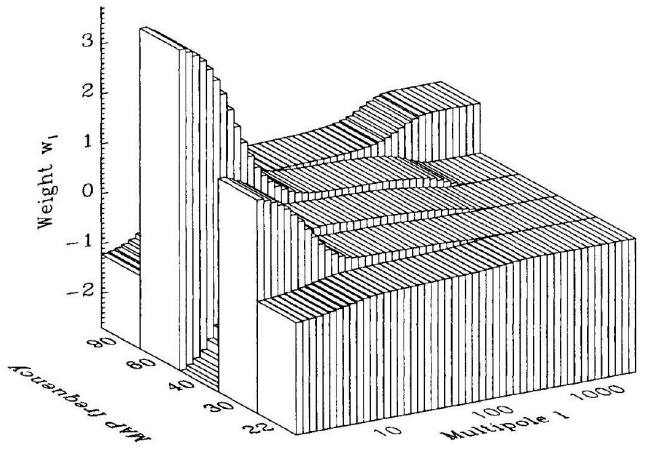


Figure 11.18. The minimum variance linear combination of the frequency channels of MAP in the presence of foregrounds (from Tegmark et al., 2000). Compare with Figure 11.17 to see that, especially at low l, foreground contamination dictates a more complicated linear combination of the channels.

ically most of a survey's galaxies lie close to the flux limits. Yet another example is a CMB experiment which has no sensitivity to the average temperature over a given set of pixels because of atmospheric contamination. Another is a galaxy survey in which a given stripe is contaminated because a CCD went bad at the time of observation. And there are many more examples.

One way of dealing with such contamination is to subtract it off. This is commonly done in galaxy surveys by applying the reddening correction due to dust from an external dust map (the best one at present is by Schlegel, Finkbeiner, and Davis, 1998).

There is another way of dealing with mode contamination, one which is rapidly growing in importance, as precision cosmology becomes a reality. This technique is based on the twin observations that (i) often there is quite a bit of uncertainty in the amplitude of the contamination and (ii) a given experiment often measures many, many modes. Since the second fact is true, we can often do without the offending mode entirely without losing too much information. Since the first fact is true, we often should do without the offending mode, for it may lead us to an incorrect place in parameter space.

How can we eliminate a contaminated mode from an experiment? Let's start with the case where the contaminated mode is a single spatial pixel. In that case, one simple way to ensure that the pixel carries no weight in the likelihood analysis is to add to the covariance matrix a huge number in the diagonal element corre- sponding to the offending pixel. Then, no matter the value observed in this pixel, the likelihood function will not be affected. This simple idea-adding large noise to a contaminated mode - can be extended to more complicated modes, those not localized to one spatial pixel.

As an example, consider a two-pixel CMB experiment, in which the atmosphere contributes identically to both pixels. Thus, the average temperature of the CMB cannot be determined. The contaminant is assumed to be 100% correlated in the two pixels, so we add to the noise covariance matrix

$$C_{\rm con} = \kappa \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

where κ is a very large number. Suppose the noise - in the absence of this constraint - is uniform and uncorrelated. Then the new noise covariance matrix is

$$C_N = \sigma_n^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \kappa \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

In the likelihood function, we add the noise covariance matrix to the signal covariance matrix. Again, for simplicity assume that the pixels are far enough away from each other so that the signal is uncorrelated. Then, the likelihood function depends on the full covariance matrix

$$C = \begin{pmatrix} \sigma_n^2 + \sigma_s^2 + \kappa & \kappa \\ \kappa & \sigma_n^2 + \sigma_s^2 + \kappa \end{pmatrix}.$$

In particular, the likelihood function depends on the determinant and inverse of this matrix (recall Eq. (11.20)). In the limit that κ is very large, the determinant becomes $2\kappa\left(\sigma_n^2 + \sigma_s^2\right)$ and the inverse is

$$C^{-1} \to \frac{1}{2\left(\sigma_n^2 + \sigma_s^2\right)} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Therefore, in this limit, the likelihood function becomes,

$$\mathcal{L} \to \frac{1}{2\pi\sqrt{2\kappa\left(\sigma_n^2 + \sigma_s^2\right)}} \exp\left\{-\frac{\left(\Delta_1 - \Delta_2\right)^2}{2\left(\sigma_n^2 + \sigma_s^2\right)}\right\}$$

That is, apart from an irrelevant normalization constant, the likelihood function is a Gaussian in $\Delta_1 - \Delta_2$ (the difference between the observed temperatures in the two pixels) with variance equal to $\sigma_n^2 + \sigma_s^2$. Thus adding the constraint matrix of Eq. (11.154) is our way of telling the likelihood function to ignore the average temperature.

In this simple example, we could have written down Eq. (11.158) from the outset, but with more complicated modes, the constraint formalism is extremely powerful. What is the generalization of Eq. (11.154) for an arbitrary contamination? Suppose the contaminated mode is of the form m_i where index i labels the pixels. Thus, in the average example, m_i would be equal to $(1, 1)$. One adds to the noise covariance matrix the outer product of this vector times a large number κ:

$$(C_{\rm con})_{ij} = \kappa m_i m_j.$$

This is precisely what we did above for the average, but this expression is now completely general and allows for elimination of any contaminated mode. These matrices, often called constraint matrices, have been used extensively in recent CMB analyses, most notably in the interferometric experiment, DASI.

SUGGESTED READING

I am not a professional statistician and this chapter no doubt glosses over some important concepts in statistics. Nonetheless, I believe this chapter does do justice to the recent flurry of activity in cosmological analysis. Readers interested in more general statistics treatments might consult The Advanced Theory of Statistics (Kendall and Stuart). An essential reference is Numerical Recipes (Press et al.) for all numerical work, and Chapter 15 especially for some of the analysis issues discussed here. A couple of nice early papers on the CMB are Readhead et al. (1989) and Bond et al. (1991). The former takes the likelihood function further than we did here. For example, it deals with frequentist tests and defines such terms as significance and power, which are extremely important in statistics. The Bond et al. paper is a concise introduction to Bayesian analysis of a CMB experiment. Among the ideas explained clearly there are CMB window functions; the likelihood function; dealing with an unknown average; and, to top it off, the idea that CMB experiments will probe the baryon density.

The discussion in Section 11.1.3 is based on a similar treatment in Tegmark et al. (1998), which cemented the connection between CMB analyses and galaxy surveys. It also deals with pixelization schemes other than just the two in Section 11.1.3. Window functions for the CMB are discussed in many places. A nice recent treatment is given by Souradeep and Ratra (2001).

Karhunen and Love decompositions were introduced to the cosmology community by Bunn (1995) and Bond (1995), who used them to analyze COBE, and by Vogeley and Szalay (1996) in the context of galaxy surveys. The optimal quadratic estimator was introduced by Tegmark (1997) and Bond, Jaffe, and Knox (1998), the latter using the Newton-Raphson motivation I stressed in the text. The former focused on the minimum variance aspect, which you can prove in

D-10.

Earlier, Feldman, Kaiser, and Peacock (1994) had computed an optimal estimator for galaxy surveys which turns out to be the small-scale limit of the optimal quadratic estimator. The Fisher matrix was introduced by Fisher (1935). Knox (1995) computed the Fisher matrix in the all-sky CMB case, Tegmark et al. (1998) for galaxy surveys. Jungman et al. (1996) used the Fisher matrix (although they didn't call it that) to give the first forecast of parameter determination. There have been many follow-up papers improving and tweaking various parts of the forecast. Some of the improvements are discussed by Eisenstein, Hu, and Tegmark (1999). The curvature matrix and the covariance matrix of errors on parameters are covered in detail in Numerical Recipes.

Mapmaking and indeed many of the issues in CMB analysis are reviewed by Bond et al. (1999). Foregrounds have been discussed by many authors. The text borrows most heavily from Dodelson (1997) and Tegmark et al. (2000). Other works of note include Tegmark and Efstathiou (1996) and Bouchet and Gispert (1999). Mode subtraction along the lines discussed in Section 11.6.2 was introduced by Bond, Jaffe, and Knox (1998). A nice example of its use is in Halverson et al. (2005).

Exercises

D-1.

In the simple example of Section 11.1.1, show that a prior uniform in σ_n^2 gives a final probability distribution in σ_n different from the one in Eq. (11.5).

D-2.

In the simple example of Section 11.1.1, we found the error on the signal s. What is the error on the other theoretical parameter of the model σ_n?

D-3.

Derive the expression for the covariance matrix due to Poisson sampling, Eq. (11.32).
(a) Divide the survey region into small sub-volumes. Assume that the number of galaxies in a given sub-volume is drawn from a Poisson distribution with mean $\bar{n}$ (assume $\bar{n}$ is constant in all sub-volumes for simplicity),

$$P(n) = \frac{(\bar{n})^n e^{-\bar{n}}}{n!}$$

Determine the expectation values $\langle n \rangle$ and $\langle n^2 \rangle$ for this distribution.
(b) Rewrite Eq. (11.28) as

$$\Delta_i = v \sum_\alpha \psi_i \left(\vec{x}_\alpha\right) \left[\frac{n\left(\vec{x}_\alpha\right) - \bar{n}}{\bar{n}}\right]$$

where α indexes each sub-volume of size v. Using the results of (a), and assuming that there is no intrinsic clustering, determine $\langle \Delta_i \Delta_j \rangle$. Show that it is given by Eq. (11.32). You'll have to change the sums back into integrals.

D-4.

Determine the noise covariance matrix in a galaxy survey using countsin-cells. Assume the cells are spherical with radius R, and find $\left(C_N\right)_{ij}$ as a function of the separation between two cell centers; call it r_{ij}. Assume that the survey is volume limited, that is, that $\bar{n}$ is constant everywhere within a volume V.

D-5.

Do a full likelihood analysis of the University of California at Santa Barbara's 1990-1 CMB experiment carried out at the South Pole (Gaier et al., 1992)
(a) Determine the window function. The chopping angle was 2.1° and the beam width (FWHM) 1.35°. The anisotropy was measured at nine positions, each separated by 2.1° on the sky. Neglect off-diagonal elements.
(b) Fit a flat band power (i.e., take $\mathcal{C} := l(l+1)C_l/2\pi$ constant) to the following data:

Position	$\delta T(\mu\mathrm{K})$	$\sigma_n(\mu\mathrm{K})$
1	-30.5	25.9
2	-3.2	26.5
3	29.2	26.1
4	-10.8	26.3
5	-8.7	28.8
6	23.1	26.4
7	4.7	26.5
8	-24.7	26.6
9	20.3	25.8

Assume that the noise and the signal are both uncorrelated from pixel to pixel. (c) The likelihood function you obtain should peak at band power equal to zero. Find the 95% CL upper limit on the band power, defined by the point for which

$$\int_0^{\mathcal{C}_U} d\mathcal{C}\mathcal{L}(\mathcal{C}) = 0.95 \int_0^{\infty} d\mathcal{C}\mathcal{L}(\mathcal{C})$$

This upper limit was reported around the same time as the COBE detection of anisotropies. Compare the two results.

D-6.

Find the Fourier transform of the top-hat function $f(x) = \Theta(x+ R)\Theta(R - x)$ where Θ is the step function, equal to 1 when its argument is positive and zero otherwise.

D-7.

Find the diagonal elements of the covariance matrix C_S for a volume limited survey for modes $k_i \gg R^{-1}$. Show that they are given by Eq. (11.60).

D-8.

Compute the off-diagonal window function for the two types of galaxy surveys mentioned in Section 11.2.3, a volume-limited survey and a pencil-beam survey. Take both $\vec{k}_i$ and $\vec{k}_j$ parallel to the z-axis (which in the pencil beam survey is aligned with the long distance L). Plot the window function at $k = k_i$ as a function of k_j. At the point $k_i = k_j$, you should recapture the corresponding points in Figures 11.5 and 11.7.

D-9.

Prove Eq. (11.100). Assume there is only one parameter λ and that the likelihood function is Gaussian in the overdensities Δ. Further assume that you have iterated enough times so that the input parameter $\lambda^{(0)}$ is equal to the true value $\bar{\lambda}$.

D-10.

Derive the optimal quadratic estimator of Eq. (11.92) by minimizing the variance of a general quadratic estimator subject to the constraint that its expectation value is unbiased. Hint: Use a Lagrange multiplier to enforce the constraint.

D-11.

Consider an all-sky CMB experiment with spatial pixels of area $\Delta\Omega$. Assume that the experiment measures the temperature in each pixel with Gaussian noise σ_n. The noise is thus assumed to be uniform (the same everywhere on the sky) and uncorrelated (from one pixel to the next). Determine the noise covariance matrix for a_{lm}. If the pixel size is cut in half (for the same experiment), each pixel will get less observing time by a factor of 2. The noise will then go up for each pixel by a factor of $\sqrt{2}$. Show that these two changes (smaller pixels; more noise per pixel) leave the noise covariance matrix for a_{lm} unchanged.

D-12.

Estimate the expected errors on C_l for the following experiments: (i) COBE, (ii) Boomerang, and (iii) MAP, and (iv) Planck.

D-13.

The full-sky limits for the Fisher matrix derived in Section 11.4 can be used to find the optimal quadratic estimator. The results should not surprise you.

(a) Given a set of a_{lm} from a full-sky CMB experiment with uniform weight w, find the optimal quadratic estimator for C_l.

(b) Given a set of pixelized Fourier overdensities from a $3D$ galaxy survey, find the optimal quadratic estimator for $P(k)$.

(c) From your answers to (a) and (b), discuss qualitatively the effectiveness of the optimal quadratic estimator. When do you expect it to perform differently from the naive quadratic estimator, $\sum_m |a_{lm}|^2/(2l+1)$ for C_l and $\sum_{\vec{k}} \left|\delta_{\vec{k}}\right|^2/(4\pi k^2 \Delta k)$ for $P(k)$ (where Δk is the width of the k-bin)?

D-14.

Show that the noise covariance matrix of a map using the estimator in Eq. (11.134) is C_N, as given in Eq. (11.135)

D-15.

Suppose one had prior information about a foreground, in the form of an assumed power spectrum, C_l^α, where α labels the foreground component.

(a) Find the miminum variance estimator for the temperature and the associated covariance matrix.

(b) Consider the example of Section 11.6.1 with two frequencies in the RayleighJeans regime and one foreground with shape vector $\vec{W}^1 = (1, 1/2)$. What is the best estimator of the CMB temperature if the foreground has assumed mean equal to zero and variance equal to that of the noise (σ_n^2)? What is the new error on the determination of the CMB temperature? Compare both of these to the case treated in the text when no information about the foreground amplitude was assumed.

27 Planck 2018 results. VI. Cosmological parameters

We present cosmological parameter results from the final full-mission *Planck* measurements of the cosmic microwave background (CMB) anisotropies, combining information from the temperature and polarization maps and the lensing reconstruction. Compared to the 2015 results, improved measurements of large-scale polarization allow the reionization optical depth to be measured with higher precision, leading to significant gains in the precision of other correlated parameters. Improved modelling of the small-scale polarization leads to more robust constraints on many parameters, with residual modelling uncertainties estimated to affect them only at the $0.5\,\sigma$ level. We find good consistency with the standard spatially-flat 6-parameter ΛCDM cosmology having a power-law spectrum of adiabatic scalar perturbations (denoted "base ΛCDM" in this paper), from polarization, temperature, and lensing, separately and in combination. A combined analysis gives dark matter density $\Omega_{\rm c}h^2 = 0.120 \pm 0.001$, baryon density $\Omega_{\rm b}h^2 = 0.0224 \pm 0.0001$, scalar spectral index $n_{\rm s} = 0.965 \pm 0.004$, and optical depth $\tau = 0.054 \pm 0.007$ (in this abstract we quote 68 % confidence regions on measured parameters and 95 % on upper limits). The angular acoustic scale is measured to 0.03 % precision, with $100\theta_* = 1.0411 \pm 0.0003$. These results are only weakly dependent on the cosmological model and remain stable, with somewhat increased errors, in many commonly considered extensions. Assuming the base-ΛCDM cosmology, the inferred (model-dependent) late-Universe parameters are: Hubble constant $H_0 = (67.4 \pm 0.5)$ km s^{-1}Mpc^{-1}; matter density parameter $\Omega_{\rm m} = 0.315 \pm 0.007$; and matter fluctuation amplitude $\sigma_8 = 0.811 \pm 0.006$. We find no compelling evidence for extensions to the base-ΛCDM model. Combining with baryon acoustic oscillation (BAO) measurements (and considering single-parameter extensions) we constrain the effective extra relativistic degrees of freedom to be $N_{\rm eff} = 2.99 \pm 0.17$, in agreement with the Standard Model prediction $N_{\rm eff} = 3.046$, and find that the neutrino mass is tightly constrained to $\sum m_\nu <$

0.12 eV. The CMB spectra continue to prefer higher lensing amplitudes than predicted in base ΛCDM at over $2\,\sigma$, which pulls some parameters that affect the lensing amplitude away from the ΛCDM model; however, this is not supported by the lensing reconstruction or (in models that also change the background geometry) BAO data. The joint constraint with BAO measurements on spatial curvature is consistent with a flat universe, $\Omega_K = 0.001 \pm 0.002$. Also combining with Type Ia supernovae (SNe), the dark-energy equation of state parameter is measured to be $w_0 = -1.03 \pm 0.03$, consistent with a cosmological constant. We find no evidence for deviations from a purely power-law primordial spectrum, and combining with data from BAO, BICEP2, and Keck Array data, we place a limit on the tensor-to-scalar ratio $r_{0.002} < 0.06$. Standard big-bang nucleosynthesis predictions for the helium and deuterium abundances for the base-ΛCDM cosmology are in excellent agreement with observations. The *Planck* base-ΛCDM results are in good agreement with BAO, SNe, and some galaxy lensing observations, but in slight tension with the Dark Energy Survey's combined-probe results including galaxy clustering (which prefers lower fluctuation amplitudes or matter density parameters), and in significant, $3.6\,\sigma$, tension with local measurements of the Hubble constant (which prefer a higher value). Simple model extensions that can partially resolve these tensions are not favoured by the *Planck* data.

27.1 Introduction

Since their discovery [346], temperature anisotropies in the cosmic microwave background (CMB) have become one of the most powerful ways of studying cosmology and the physics of the early Universe. This paper reports the final results on cosmological parameters from the Planck Collaboration.[131] Our first results were presented in Planck Collaboration XVI [276, hereafter PCP13]. These were based on temperature (TT) power spectra and CMB lensing measurements from the first 15.5 months of *Planck* data combined with the Wilkinson Microwave Anisotropy Probe (WMAP) polarization likelihood at multipoles $\ell \leq 23$ [42] to constrain the reionization optical depth τ. Planck Collaboration XIII [284, hereafter PCP15] reported results from the full *Planck* mission (29 months of observations with the High Frequency Instrument, HFI), with substantial improvements in the characterization of the *Planck* beams and absolute calibration (resolving a difference between the absolute calibrations of WMAP and *Planck*). The focus of PCP15, as in PCP13, was on temperature observations, though we reported preliminary results on the high-multipole TE and EE polarization spectra. In addition, we used polarization measurements at low multipoles from the Low Frequency Instrument (LFI) to constrain the value of τ.

Following the completion of PCP15, a concerted effort by the *Planck* team was made to reduce systematics in the HFI polarization data at low multipoles. First results were presented in Planck Collaboration Int. XLVI [297], which showed evidence for a lower value of the reionization optical depth than in the 2015 results. Further improvements to the HFI polarization maps prepared for the 2018 data release are described in Planck Collaboration III [291]. In this paper, we constrain τ using a new low-multipole likelihood constructed from these maps. The improvements in HFI data processing since PCP15 have very little effect on the TT, TE, and EE spectra at high multipoles. However, this paper includes characterizations of the temperature-to-polarization leakage and relative calibrations of the polarization spectra enabling us to produce a combined TT,TE,EE likelihood that is of sufficient fidelity to be used to test cosmological models (although with some limitations, which will be described in detail in the main body of this paper). The focus of this paper, therefore, is to present updated cosmological results from *Planck* power spectra and CMB lensing measurements using temperature and polarization.

PCP13 showed that the *Planck* data were remarkably consistent with a spatially-flat ΛCDM cosmology with purely adiabatic, Gaussian initial fluctuations, as predicted in simple inflationary models. We refer to this model, which can be specified by six parameters, as "base" ΛCDM in this paper. Note that in the base ΛCDM cosmology we assume a single minimal-mass neutrino eigenstate. We investigated a grid of one- and two-parameter extensions to the base-ΛCDM cosmology (varying, for example, the sum of neutrino masses, effective number of relativistic

[131] *Planck* (https://www.esa.int/Planck) is a project of the European Space Agency (ESA) with instruments provided by two scientific consortia funded by ESA member states and led by Principal Investigators from France and Italy, telescope reflectors provided through a collaboration between ESA and a scientific consortium led and funded by Denmark, and additional contributions from NASA (USA).

degrees of freedom $N_{\rm eff}$, spatial curvature Ω_K, or dark-energy equation of state w_0), finding no statistically significant preference for any departure from the base model. These conclusions were reinforced using the full *Planck* mission data in PCP15.

The analyses reported in PCP13 and PCP15 revealed some discrepancies (often referred to as "tensions") with non-*Planck* data in the context of ΛCDM models (e.g., distance-ladder measurements of the Hubble constant and determinations of the present-day amplitude of the fluctuation spectrum), including other CMB experiments [350]. As a result, it is important to test the fidelity of the *Planck* data as thoroughly as possible. First, we would like to emphasize that where it has been possible to compare data between different experiments at the map level (therefore eliminating cosmic variance), they have been found to be consistent within the levels set by instrument noise, apart from overall differences in absolute calibration; comparisons between WMAP and *Planck* are described by Huang et al. [186], between the Atacama Cosmology Telescope (ACT) and *Planck* by Louis et al. [225], and between the South Pole Telescope (SPT) and *Planck* by Hou et al. [181]. There have also been claims of internal inconsistencies in the *Planck* TT power spectrum between frequencies [347] and between the ΛCDM parameters obtained from low and high multipoles [6]. In addition, the *Planck* TT spectrum preferred more lensing than expected in the base-ΛCDM model (quantified by the phenomenological $A_{\rm L}$ parameter defined in Sect. 27.2.3) at moderate statistical significance, raising the question of whether there are unaccounted for systematic effects lurking within the *Planck* data. These issues were largely addressed in Planck Collaboration XI [283], PCP15, and in an associated paper, Planck Collaboration Int. LI [299]. We revisit these issues in this paper at the cosmological parameter level, using consistency with the *Planck* polarization spectra as an additional check. Since 2013, we have improved the absolute calibration (fixing the amplitudes of the power spectra), added *Planck* polarization, full-mission *Planck* lensing , and produced a new low-multipole polarization likelihood from the *Planck* HFI. Nevertheless, the key parameters of the base-ΛCDM model reported in this paper, agree to better than $1\,\sigma_{2013}$[132] with those determined from the nominal mission temperature data in PCP13, with the exception of τ (which is lower in the 2018 analysis by $1.1\,\sigma_{2013}$). The cosmological parameters from *Planck* have remained remarkably stable since the first data release in 2013.

The results from *Planck* are in very good agreement with simple single-field models of inflation [279, 288]. We have found no evidence for primordial non-Gaussianity [280, 287], setting stringent upper limits. Nor have we found any evidence for isocurvature perturbations or cosmic defects [see PCP15 and 288]. *Planck*, together with Bicep/Keck [51] polarization measurements, set tight limits on the amplitude of gravitational waves generated during inflation. These results are updated in this paper and in the companion papers, describing more comprehensive tests of inflationary models [296] and primordial non-Gaussianity [295]. The *Planck* results require adiabatic, Gaussian initial scalar fluctuations, with a red-tilted spectrum. The upper limits on gravitational waves then require flat inflationary potentials, which has stimulated new developments in inflationary model building [see e.g., 129, 196, 141, 14, and references therein]. Some authors [189, 188] have come to a very different conclusion, namely that the *Planck*/Bicep/Keck results require special initial conditions and therefore disfavour inflation. This controversy lies firmly in the theoretical domain [see e.g., 157, 223], since observations of the CMB constrain only a limited number of e-folds during inflation, not the initial conditions. Post *Planck*, inflation remains a viable and attractive mechanism for accounting for the structure that we see in the Universe.

The layout of this paper is as follows. Section 27.2 describes changes to our theoretical modelling since PCP15 and summarizes the likelihoods used in this paper. More comprehensive descriptions of the power-spectrum likelihoods are given in Planck Collaboration V [293], while the 2018 *Planck* CMB lensing likelihood is described in detail in Planck Collaboration VIII [294]. Section 27.3 discusses the parameters of the base-ΛCDM model, comparing parameters derived from the *Planck* TT, TE, and EE power spectra. Our best estimates of the base-ΛCDM cosmological parameters are derived from the full *Planck* TT,TE,EE likelihood combined with *Planck* CMB lensing and an HFI-based low-multipole polarization likelihood to constrain τ. We compare the *Planck* TE and EE spectra with power spectra measured from recent ground-based experiments in Sect. 27.4.

The *Planck* base-ΛCDM cosmology is compared with external data sets in Sect. 27.5. CMB power spectrum measurements suffer from a "geometric degeneracy" [see 116] which limits their ability to constrain certain extensions to the base cosmology (for example, allowing Ω_K or w_0 to vary). *Planck* lensing measurements partially break the geometric degeneracy, but it is broken

[132]Here σ_{2013} is the standard deviation quoted on parameters in PCP13.

very effectively with the addition of baryon acoustic oscillation (BAO) measurements from galaxy surveys. As in PCP13 and PCP15 we use BAO measurements as the primary external data set to combine with *Planck*. We adopt this approach for two reasons. Firstly, BAO-scale determinations are relatively simple geometric measurements, with little scope for bias from systematic errors. Secondly, the primary purpose of this paper is to present and emphasize the *Planck* results. We therefore make minimal use of external data sets in reporting our main results, rather than combining with many different data sets. Exploration of multiple data sets can be done by others using the Monte Carlo Markov chains and *Planck* likelihoods released through the Planck Legacy Archive (PLA).[133] Nevertheless, Sect. 27.5 presents a comprehensive survey of the consistency of the *Planck* base-ΛCDM cosmology with different types of astrophysical data, including Type 1a supernovae, redshift-space distortions, galaxy shear surveys, and galaxy cluster counts. These data sets are consistent with the *Planck* base-ΛCDM cosmology with, at worst, moderate tensions at about the $2.5\,\sigma$ level. Distance-ladder measurements of the Hubble constant, H_0, are an exception, however. The latest measurement from Riess et al. [306] is discrepant with the *Planck* base-ΛCDM value for H_0 at about the $4.4\,\sigma$ level. This large discrepancy, and its possible implications for cosmology, is discussed in Sect. 27.5.4.

Section 27.6 investigates the internal consistency of the *Planck* base-ΛCDM parameters, presenting additional tests using the TE and EE spectra, as well as a discussion of systematic uncertainties. Results from our main grid of parameter constraints on one- or two-parameter extensions to the base-ΛCDM cosmology are presented in Sect. **??**. That section also includes discussions of more complex models of dark energy and modified gravity [updating the results presented in 285], primordial nucleosynthesis, reionization, recombination, and dark matter annihilation. Section 28.11 summarizes our main conclusions.

27.2 Methodology and likelihoods

27.2.1 Theoretical model

The definitions, methodology, and notation used in this paper largely follow those adopted in the earlier Planck Collaboration papers dealing with cosmological parameters (PCP13, PCP15). Our baseline assumption is the ΛCDM model with purely adiabatic scalar primordial perturbations with a power-law spectrum. We assume three neutrinos species, approximated as two massless states and a single massive neutrino of mass $m_\nu = 0.06\,\mathrm{eV}$. We put flat priors on the baryon density $\omega_\mathrm{b} \equiv \Omega_\mathrm{b}h^2$, cold dark matter density $\omega_\mathrm{c} \equiv \Omega_\mathrm{c}h^2$, an approximation to the observed angular size of the sound horizon at recombination θ_MC, the reionization optical depth τ, the initial super-horizon amplitude of curvature perturbations A_s at $k = 0.05\,\mathrm{Mpc}^{-1}$, and the primordial spectral index n_s. Other parameter definitions, prior limits, and notation are described explicitly in table 1 of PCP13; the only change is that we now take the amplitude prior to be flat in $\log A_\mathrm{s}$ over the range $1.61 < \log(10^{10}A_\mathrm{s}) < 3.91$ (which makes no difference to *Planck* results, but is consistent with the range used for some external data analyses).

Changes in our physical modelling compared with PCP15 are as follows.

- For modelling the small-scale nonlinear matter power spectrum, and calculating the effects of CMB lensing, we use the `halofit` technique [344] as before, but now replace the Takahashi et al. [354] approach with `HMcode`, the fitting method of Mead et al. [237, 236], as implemented in `camb` [221].

- For each model in which the fraction of baryonic mass in helium Y_P is *not* varied independently of other parameters, the value is now set using an updated big-bang nucleosynthesis (BBN) prediction by interpolation on a grid of values calculated using version 1.1 of the `PArthENoPE` BBN code [271, version 2.0 gives identical results]. We now use a fixed fiducial neutron decay-constant value of $\tau_\mathrm{n} = 880.2\,\mathrm{s}$, neglecting uncertainties. Predictions from `PArthENoPE` for the helium mass fraction ($Y_\mathrm{P} \approx 0.2454$, nucleon fraction $Y_\mathrm{P}^\mathrm{BBN} \approx 0.2467$ from *Planck* in ΛCDM) are lower than those from the code of Pitrou et al. [272] for the same value of τ_n by $\Delta Y_\mathrm{P} \approx 0.0005$; however, other parameter results would be consistent to well within 0.1σ. See Sect. **??** for further discussion of BBN parameter uncertainties and code variations.

[133] https://pla.esac.esa.int

Building upon many years of theoretical effort, the computation of CMB power spectra and the related likelihood functions has now become highly efficient and robust. Our main results are based upon the lensed CMB power spectra computed with the August 2017 version of the `camb`[134] Boltzmann code [221] and parameter constraints are based on the July 2018 version of `CosmoMC`[135] [219, 218]. We have checked that there is very good consistency between these results and equivalent results computed using the `class` Boltzmann code [56] and `MontePython` sampler [30, 64]. Marginalized densities, limits, and contour plots are generated using updated adaptive kernel density estimates (with corrections for boundary and smoothing biases) as calculated using the `getdist` package[136] (also part of `CosmoMC`), which improves average accuracy for a given number of posterior samples compared to the version used in our previous analyses.

A few new derived parameters have been added to the output of the `CosmoMC` chains to allow comparisons and combinations with external data sets. A full description of all parameters is provided in the tables presented in the Explanatory Supplement [274], and parameter chains are available on the PLA.

27.2.2 Power spectra and likelihoods

Since the 2015 *Planck* data release, most of the effort on the low-level data processing has been directed to improving the fidelity of the polarization data at low multipoles. The first results from this effort were reported in Planck Collaboration Int. XLVI [297] and led to a new determination of the reionization optical depth, τ. The main results presented in this paper are based on the 2018 HFI maps produced with the `SRoll` mapmaking algorithm described in detail in [291], supplemented with LFI data described in Planck Collaboration II [290].

Because *Planck*-HFI measures polarization by differencing the signals measured by polarization-sensitive bolometers (PSBs), a number of instrumental effects need to be controlled to achieve high precision in the absolute calibrations of each detector. These include: effective gain variations arising from nonlinearities in the analogue-to-digital electronics and thermal fluctuations; far-field beam characterization, including long bolometer time constants; and differences in detector bandpasses. The `SRoll` mapmaking solution for the 100–353 GHz channels minimizes map residuals between all HFI detectors at a given frequency, using absolute calibrations based on the orbital dipole, together with a bandpass-mismatch model constructed from spatial templates of the foregrounds and a parametric model characterizing the remaining systematics. We refer the reader to Planck Collaboration III [291] for details of the implementation of `SRoll`. The fidelity of the `SRoll` maps can be assessed using various null tests (e.g., splitting the data by half-mission, odd-even surveys, and different detector combinations) and by the consistency of the recovered Solar dipole solution. These tests are described in [291] and demonstrate that the Solar dipole calibration is accurate to about one part in 10^4 for the three lowest-frequency HFI channels. Large-scale intensity-to-polarization leakage, caused by calibration mismatch in the `SRoll` maps, is then reduced to levels $\lesssim 10^{-6}\mu\mathrm{K}^2$ at $\ell > 3$.

The low-multipole polarization likelihood used in this paper is based on the `SRoll` polarization maps and series of end-to-end simulations that are used to characterize the noise properties and remaining biases in the `SRoll` maps. This low-multipole likelihood is summarized in Sect. 27.2.2 and is described in more detail in Planck Collaboration V [293].

As in previous *Planck* papers, the baseline likelihood is a hybrid, patching together a low-multipole likelihood at $\ell < 30$ with a Gaussian likelihood constructed from pseudo-cross-spectrum estimates at higher multipoles. Correlations between the low and high multipoles are neglected. In this paper, we have used two independent high-multipole TT,TE,EE likelihoods.[137] The `Plik` likelihood, which is adopted as the baseline in this paper, is described in Sect. 27.2.2, while the `CamSpec` likelihood is described in Sect. 27.2.2 and Appendix **??**. These two likelihoods are in very good agreement in TT, but show small differences in TE and EE, as described below and in the main body of this paper. Section 27.2.3 summarizes the *Planck* CMB lensing likelihood, which is described in greater detail in Planck Collaboration VIII [294].

Before summarizing the high-multipole likelihoods, we make a few remarks concerning the 2018 `SRoll` maps. The main aim of the `SRoll` processing is to reduce the impact of systematics

[134] https://camb.info

[135] https://cosmologist.info/cosmomc/

[136] https://getdist.readthedocs.io/

[137] We use roman letters, such as TT,TE,EE, to refer to particular likelihood combinations, but use italics, such as TT, when discussing power spectra more generally.

at low multipoles and hence the main differences between the 2015 and 2018 HFI maps are at low multipoles. Compared to the 2015 HFI maps, the `SRoll` maps eliminate the last 1000 HFI scanning rings (about 22 days of observations) because these were less thermally stable than the rest of the mission. `SRoll` uses higher resolution maps to determine the destriping offsets compared to the 2015 maps, leading to a reduction of about 12 % in the noise levels at 143 GHz [see figure 10 of 291]. A tighter requirement on the reconstruction of Q and U values at each pixel leads to more missing pixels in the 2018 maps compared to 2015. These and other changes to the 2018 *Planck* maps have very little impact on the temperature and polarization spectra at high multipoles (as will be demonstrated explicitly in Fig. 175 below).

There are, however, data-processing effects that need to be accounted for to create an unbiased temperature+polarization likelihood at high multipoles from the `SRoll` maps. In simplified form, the power absorbed by a detector at time t on the sky is

$$P(t) = G\,\{I + \rho\,[Q\cos 2(\psi(t) + \psi_0) + U\sin 2(\psi(t) + \psi_0)]\} + n(t), \tag{1195}$$

where I, Q, and U are the beam-convolved Stokes parameters seen by the detector at time t, G is the effective gain (setting the absolute calibration), ρ is the detector polarization efficiency, $\psi(t)$ is the roll angle of the satellite, ψ_0 is the detector polarization angle, and $n(t)$ is the noise. For a perfect polarization-sensitive detector, $\rho = 1$, while for a perfect unpolarized detector, $\rho = 0$. The polarization efficiencies and polarization angles for the HFI bolometers were measured on the ground and are reported in [315]. For polarization-sensitive detectors the ground-based measurements of polarization angles were measured to an accuracy of approximately 1° and the polarization efficiencies to a quoted accuracy of 0.1–0.3 %. The `SRoll` mapmaking algorithm assumes the ground-based measurements of polarization angles and efficiencies, which cannot be separated because they are degenerate with each other. Errors in the polarization angles induce leakage from E to B modes, while errors in the polarization efficiencies lead to gain mismatch between I, Q and U. Analysis of the *Planck* TB and EB spectra (which should be zero in the absence of parity-violating physics) reported in Planck Collaboration III [291], suggest errors in the polarization angles of $\lesssim 0.5°$, within the error estimates reported in [315]. However, systematic errors in the polarization efficiencies are found to be several times larger than the [315] determinations (which were limited to characterizations of the feed and detector sub-assemblies and did not characterize the system in combination with the telescope) leading to effective calibration offsets in the polarization spectra. These polarization efficiency differences, which are detector- and hence frequency-dependent, need to be calibrated to construct a high-multipole likelihood. To give some representative numbers, the Rosset et al. [315] ground-based measurements estimated polarization efficiencies for the PSBs, with typical values of 92–96 % at 100 GHz, 83–93 % at 143 GHz, and 94–95 % at 217 GHz (the three frequencies used to construct the high-multipole polarization likelihoods). From the `SRoll` maps, we find evidence of systematic errors in the polarization efficiencies of order 0.5–1 % at 100 and 217 GHz and up to 1.5 % at 143 GHz. Differences between the main beams of the PSBs introduce temperature-to-polarization leakage at high multipoles. We use the `QuickPol` estimates of the temperature-polarization beam transfer function matrices, as described in Hivon et al. [177], to correct for temperature-to-polarization leakage. Inaccuracies in the corrections for effective polarization efficiencies and temperature-to-polarization leakage are the main contributors to systematic errors in the *Planck* polarization spectra at high multipoles.

In principle, B-mode polarization spectra contain information about lensing and primordial tensor modes. However, for *Planck*, B-mode polarization spectra are strongly noise dominated on all angular scales. Given the very limited information contained in the *Planck* B-mode spectra (and the increased complexity involved) we do not include B-mode power spectra in the likelihoods; however, for an estimate of the lensing B-mode power spectrum see Planck Collaboration VIII 294, hereafter PL2018.

The baseline `Plik` likelihood

The `Plik` high-multipole likelihood [described in detail in 293, hereafter PPL18] is a Gaussian approximation to the probability distributions of the TT, EE, and TE angular power spectra, with semi-analytic covariance matrices calculated assuming a fiducial cosmology. It includes multipoles in the range $30 \leq \ell \leq 2508$ for TT and $30 \leq \ell \leq 1996$ for TE and EE, and is constructed from half-mission cross-spectra measured from the 100-, 143-, and 217-GHz HFI frequency maps.

The TT likelihood uses four half-mission cross-spectra, with different multipole cuts to avoid multipole regions where noise dominates due to the limited resolution of the beams and to ensure foreground contamination is correctly handled by our foreground model: 100×100 ($\ell = 30$–1197); 143×143 ($\ell = 30$–1996); 143×217 ($\ell = 30$–2508); and 217×217 ($\ell = 30$–2508). The TE and EE likelihoods also include the 100×143 and 100×217 cross-spectra to improve the signal-to-noise ratio, and have different multipole cuts: 100×100 ($\ell = 30$–999); 100×143 ($\ell = 30$–999); 100×217 ($\ell = 505$–999); 143×143 ($\ell = 30$–1996); 143×217 ($\ell = 505$–1996); and 217×217 ($\ell = 505$–1996). The 100-, 143-, and 217-GHz intensity maps are masked to reduce Galactic dust, CO, extended sources, and point-source contamination (a different point-source mask is used at each frequency), as well as badly-conditioned/missing pixels, effectively retaining 66, 57, and 47 % of the sky after apodization, respectively (see equation 10 in PCP15 for a definition of the effective sky fraction). The apodization is applied to reduce the mask-induced correlations between modes, and reduces the effective sky fraction by about 10 % compared to the unapodized masks. The 100-, 143-, and 217-GHz maps in polarization are masked only for Galactic contamination and badly-conditioned or missing pixels, effectively retaining 70, 50, and 41 % of the sky after apodization, respectively.

The baseline likelihood uses the different frequency power spectra without coadding them, modelling the foreground and instrumental effects with nuisance parameters that are marginalized over at the parameter estimation level, both in temperature and in polarization. To reduce the size of the covariance matrix and data vector, the baseline `Plik` likelihood uses binned band powers, which give an excellent approximation to the unbinned likelihood for smooth theoretical power spectra. Unbinned versions of the likelihoods are also available and provide almost identical results to the binned spectra for all of the theoretical models considered in our main parameter grid (Sect. **??**).

The major changes with respect to the 2015 `Plik` likelihood are the following.

- *Beams.* In 2015, the effective beam window functions were calculated assuming the same average sky fraction at all frequencies. In this new release, we apply beam window functions calculated for the specific sky fraction retained at each frequency. The impact on the spectra is small, at the level of approximately 0.1 % at $\ell = 2000$.
- *Dust modelling in TT.* The use of intensity-thresholded point-source masks modifies the power spectrum of the Galactic dust emission, since such masks include point-like bright Galactic dust regions. Because these point-source masks are frequency dependent, a different dust template is constructed from the 545-GHz maps for each power spectrum used in the likelihood. This differs from the approach adopted in 2015, which used a Galactic dust template with the same shape at all frequencies. As in 2015, the Galactic dust amplitudes are then left free to vary, with priors determined from cross-correlating the frequency maps used in the likelihood with the 545-GHz maps. These changes produce small correlated shifts in the dust, cosmic infrared background (CIB), and point-source amplitudes, but have negligible impact on cosmological parameters.
- *Dust modelling in TE and EE.* Dust amplitudes in TE are varied with Gaussian priors as in 2015, while in EE we fix the dust amplitudes to the values obtained using the cross-correlations with 353-GHz maps, for the reasons detailed in PPL18. The choice of fixing the dust amplitudes in EE has a small impact (of the order of $0.2\,\sigma$) on the base-ΛCDM results when combining into the full "TT,TE,EE," `Plik` likelihood because EE has lower statistical power compared to TT or TE; however, dust modelling in EE has a greater effect when parameters are estimated from EE alone (e.g., fixing the dust amplitude in EE lowers $n_{\rm s}$ by $0.8\,\sigma$, compared to allowing the dust amplitude to vary.)
- *Correction of systematic effects in the polarization spectra.* In the 2015 *Planck* analysis, small differences in the inter-frequency comparisons of TE and EE foreground-corrected polarization power spectra were identified and attributed to systematics such as temperature-to-polarization leakage and polarization efficiencies, which had not been characterized adequately at the time. For the 2018 analysis we have applied the following corrections to the `Plik` spectra.

  - *Beam-leakage correction.* The TE and EE pseudo-spectra are corrected for temperature-to-polarization leakage caused by beam mismatch, using polarized beam matrices computed with the `QuickPol` code described in Hivon et al. [177]. The beam-leakage correction template is calculated using fiducial theoretical spectra computed from the best-fit ΛCDM cosmology fitted to the TT data, together with `QuickPol` estimates of the HFI polarized beam transfer-function matrices. This template is then included in our data model. The correction for beam leakage has a larger impact on TE than on EE. For base-ΛCDM cosmology, correcting for the leakage induces shifts of $\lesssim 1\,\sigma$ when constraining parameters

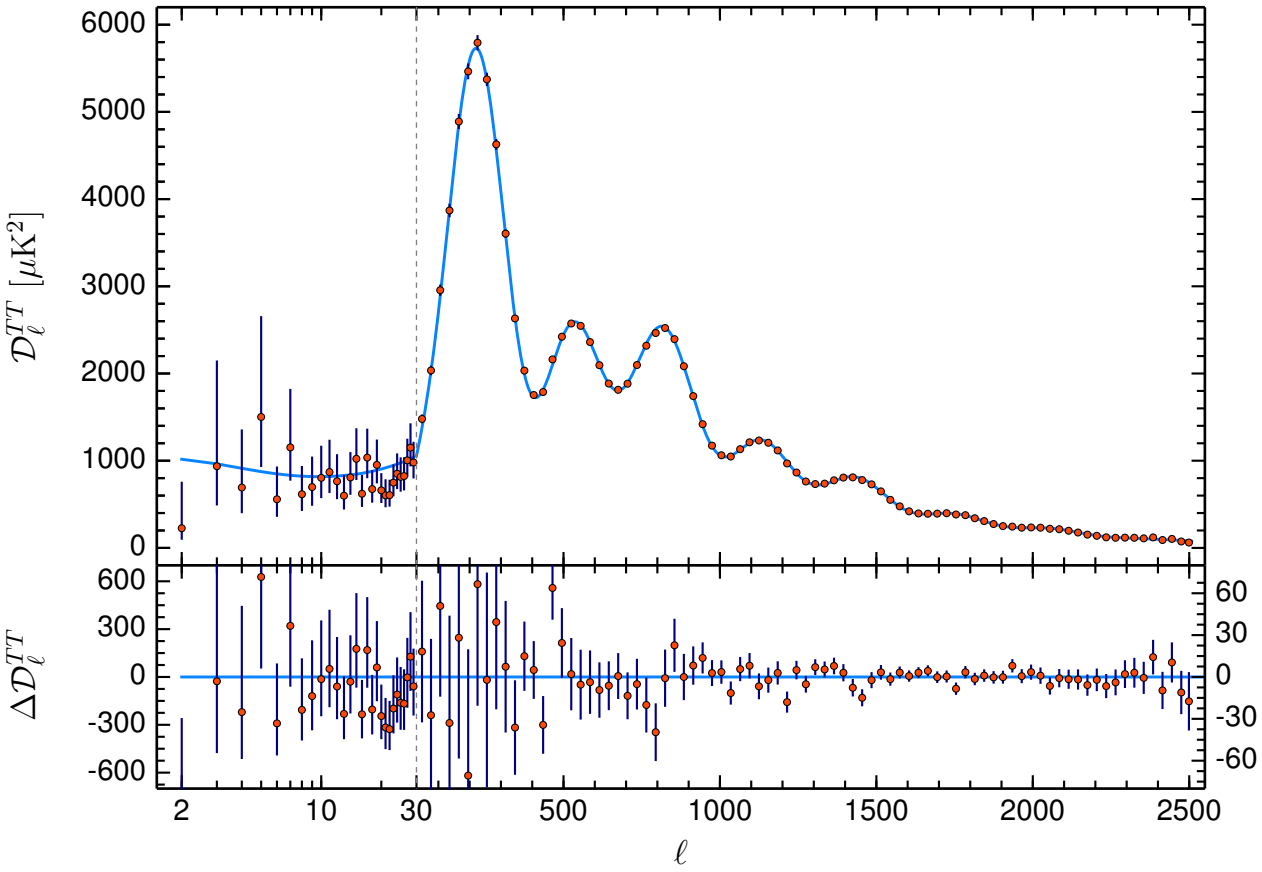


Figure 167: *Planck* 2018 temperature power spectrum. At multipoles $\ell \geq 30$ we show the frequency-coadded temperature spectrum computed from the `Plik` cross-half-mission likelihood, with foreground and other nuisance parameters fixed to a best fit assuming the base-ΛCDM cosmology. In the multipole range $2 \leq \ell \leq 29$, we plot the power spectrum estimates from the `Commander` component-separation algorithm, computed over 86 % of the sky. The base-ΛCDM theoretical spectrum best fit to the *Planck* TT,TE,EE+lowE+lensing likelihoods is plotted in light blue in the upper panel. Residuals with respect to this model are shown in the lower panel. The error bars show $\pm 1\,\sigma$ diagonal uncertainties, including cosmic variance (approximated as Gaussian) and not including uncertainties in the foreground model at $\ell \geq 30$. Note that the vertical scale changes at $\ell = 30$, where the horizontal axis switches from logarithmic to linear.

with TT,TE,EE, namely $+1.1\,\sigma$ for ω_b, $-0.7\,\sigma$ for ω_c, $+0.7\,\sigma$ for θ_MC, and $+0.5\,\sigma$ for n_s, with smaller changes for other parameters.

- *Effective polarization efficiencies.* We estimate the effective polarization efficiencies of the `SRoll` maps by comparing the frequency polarization power spectra to fiducial spectra computed from the best-fit base-ΛCDM model determined from the temperature data. The details and limitations of this procedure are described in PPL18 and briefly summarized further below. Applying these polarization efficiency estimates, we find relatively small shifts to the base-ΛCDM parameters determined from the TT,TE,EE likelihood, with the largest shifts in ω_b ($+0.4\,\sigma$), ω_c ($+0.2\,\sigma$), and n_s ($+0.2\,\sigma$). The parameter shifts are small because the polarization efficiencies at different frequencies partially average out in the coadded TE spectra (see also Fig. 175, discussed in Sect. 27.3).
- *Correlated noise in auto-frequency cross-spectra and subpixel effects.* The likelihood is built using half-mission cross-spectra to avoid noise biases from auto-spectra. However,

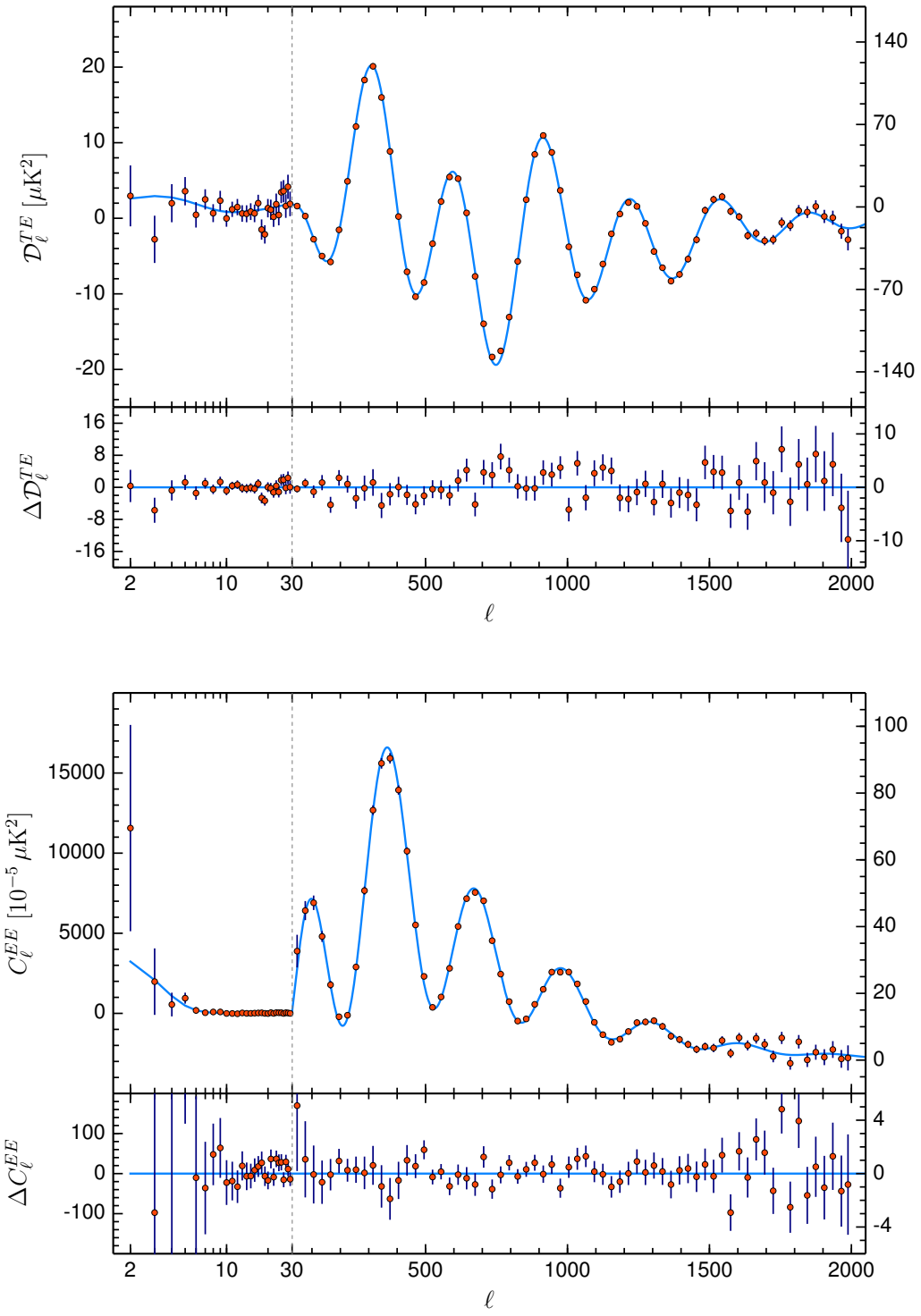


Figure 168: *Planck* 2018 TE (top) and EE (bottom) power spectra. At multipoles $\ell \geq 30$ we show the coadded frequency spectra computed from the `Plik` cross-half-mission likelihood with foreground and other nuisance parameters fixed to a best fit assuming the base-ΛCDM cosmology. In the multipole range $2 \leq \ell \leq 29$, we plot the power spectra estimates from the `SimAll` likelihood (though only the EE spectrum is used in the baseline parameter analysis at $\ell \leq 29$). The best-fit base-ΛCDM theoretical spectrum fit to the *Planck* TT,TE,EE+lowE +lensing likelihood is plotted in light blue in the upper panels. Residuals with respect to this model are shown in the lower panels. The error bars show Gaussian $\pm 1\sigma$ diagonal uncertainties including cosmic variance. Note that the vertical scale changes at $\ell = 30$, where the horizontal axis switches from logarithmic to linear.

small residual correlated noise contributions may still be present. The pixelization of the maps introduces an additional noise term because the centroid of the "hits" distribution of the detector samples in each pixel does not necessarily lie at the pixel centre. The impact of correlated noise is evaluated using the end-to-end simulations described in Planck Collaboration III [291], while the impact of subpixel effects is estimated with analytic calculations. Both effects are included in the `Plik` data model, but have negligible impact on cosmological parameters.

Of the systematic effects listed above, correction for the polarization efficiencies has the largest uncertainty. We model these factors as effective polarization calibration parameters c_ν^{EE}, defined at the power spectrum level for a frequency spectrum $\nu \times \nu$.[138] To correct for errors in polarization efficiencies and large-scale beam-transfer function errors, we recalibrate the TE and EE spectra against a fiducial theoretical model to minimize

$$\chi^2 = (C^{\rm D} - {\rm G}C^{\rm Th}){\rm M}^{-1}(C^{\rm D} - {\rm G}C^{\rm Th}), \tag{1196a}$$

with respect to the c_ν^{EE} parameters contained in the diagonal calibration matrix G with elements

$${\rm G}_{i,i} = \left(\frac{1}{\sqrt{c_\nu^{XX}c_{\nu'}^{YY}}} + \frac{1}{\sqrt{c_{\nu'}^{XX}c_\nu^{YY}}}\right)_{i,i}, \tag{1196b}$$

where the index $i = 1, N$ runs over the multipoles ℓ and frequencies $\nu \times \nu'$ of the spectra contained in the $C^{\rm D}$ data vector of dimension N; $C^{\rm D}$ contains the C_ℓ frequency spectra either for $XY = TE$ or $XY = EE$, fit separately. In Eq. (1196a), M is the covariance matrix for the appropriate spectra included in the fit, while the c_ν^{TT} temperature calibration parameters are fixed. We perform the fit only using multipoles $\ell = 200$–1000 to minimize the impact of inaccuracies in the foreground modelling or noise, and we test the stability of the results by fitting either one frequency spectrum or all the frequency spectra at the same time. The recalibration is computed with respect to a fiducial model vector $C^{\rm Th}$ because the *Planck* polarization spectra are noisy and it is not possible to inter-calibrate the spectra to a precision of better than 1 % without invoking a reference model. The fiducial theoretical spectra $C_\ell^{\rm Th}$ contained in $C^{\rm Th}$ are derived from the best-fit temperature data alone, assuming the base-ΛCDM model, adding the beam-leakage model and fixing the Galactic dust amplitudes to the central values of the priors obtained from using the 353-GHz maps. This is clearly a model-dependent procedure, but given that we fit over a restricted range of multipoles, where the TT spectra are measured to cosmic variance, the resulting polarization calibrations are insensitive to small changes in the underlying cosmological model.

In principle, the polarization efficiencies found by fitting the TE spectra should be consistent with those obtained from EE. However, the polarization efficiency at 143×143, c_{143}^{EE}, derived from the EE spectrum is about $2\,\sigma$ lower than that derived from TE (where the σ is the uncertainty of the TE estimate, of the order of 0.02). This difference may be a statistical fluctuation or it could be a sign of residual systematics that project onto calibration parameters differently in EE and TE. We have investigated ways of correcting for effective polarization efficiencies: adopting the estimates from EE (which are about a factor of 2 more precise than TE) for both the TE and EE spectra (we call this the "map-based" approach); or applying independent estimates from TE and EE (the "spectrum-based" approach). In the baseline `Plik` likelihood we use the map-based approach, with the polarization efficiencies fixed to the efficiencies obtained from the fits on EE: $\left(c_{100}^{EE}\right)_{\rm EE\,fit} = 1.021$; $\left(c_{143}^{EE}\right)_{\rm EE\,fit} = 0.966$; and $\left(c_{217}^{EE}\right)_{\rm EE\,fit} = 1.040$. The `CamSpec` likelihood, described in the next section, uses spectrum-based effective polarization efficiency corrections, leaving an overall temperature-to-polarization calibration free to vary within a specified prior.

The use of spectrum-based polarization efficiency estimates (which essentially differs by applying to EE the efficiencies given above, and to TE the efficiencies obtained fitting the TE spectra, $\left(c_{100}^{EE}\right)_{\rm TE\,fit} = 1.04$, $\left(c_{143}^{EE}\right)_{\rm TE\,fit} = 1.0$, and $\left(c_{217}^{EE}\right)_{\rm TE\,fit} = 1.02$), also has a small, but non-negligible impact on cosmological parameters. For example, for the ΛCDM model, fitting the `Plik` TT,TE,EE+lowE likelihood, using spectrum-based polarization efficiencies,

[138] Thus, the polarization efficiency for a cross-frequency spectrum $\nu \times \nu'$ in, e.g., EE is $\sqrt{c_\nu^{EE} \times c_{\nu'}^{EE}}$.

we find small shifts in the base-ΛCDM parameters compared with ignoring spectrum-based polarization efficiency corrections entirely; the largest of these shifts are $+0.5\,\sigma$ in ω_b, $+0.1\,\sigma$ in ω_c, and $+0.3\,\sigma$ in n_s (to be compared to $+0.4\,\sigma$ in ω_b, $+0.2\,\sigma$ in ω_c, and $+0.2\,\sigma$ in n_s for the map-based case). Furthermore, if we introduce the phenomenological A_L parameter (discussed in much greater detail in Sect. 27.6.2), using the baseline TT,TE,EE+lowE likelihood gives $A_L = 1.180 \pm 0.065$, differing from unity by 2.7σ (the value of A_L is unchanged with respect to the case where we ignore polar efficiencies entirely, 1.180 ± 0.065). Switching to spectrum-based polarization efficiency corrections changes this estimate to $A_L = 1.142 \pm 0.066$ differing from unity by 2.1σ. Readers of this paper should therefore not over-interpret the *Planck* polarization results and should be aware of the sensitivity of these results to small changes in the specific choices and assumptions made in constructing the polarization likelihoods, which are not accounted for in the likelihood error model. To emphasize this point, we also give results from the `CamSpec` likelihood (see, e.g., Table 24), described in the next section, which has been constructed independently of `Plik`. We also note that if we apply the `CamSpec` polarization masks and spectrum-based polarization efficiencies in the `Plik` likelihood, then the cosmological parameters from the two likelihoods are in close agreement.

The coadded 2018 `Plik` temperature and polarization power spectra and residuals with respect to the base-ΛCDM model are shown in Figs. 167 and 168.

The `CamSpec` likelihood

The `CamSpec` temperature likelihood was used as the baseline for the first analysis of cosmological parameters from *Planck*, reported in PCP13, and was described in PPL13. A detailed description of `CamSpec` and its generalization to polarization is given in [117]. For PCP15, the `CamSpec` temperature likelihood was unaltered from that adopted in PPL13, except that we used half-mission cross-spectra instead of detector-set cross-spectra and made minor modifications to the foreground model. For this set of papers, the `CamSpec` temperature analysis uses identical input maps and masks as `Plik` and is unaltered from PCP15, except for the following details.

- In previous versions we used half-ring difference maps (constructed from the first and second halves of the scanning rings within each pointing period) to estimate noise. In this release we have used differences between maps constructed from odd and even rings. The use of odd-even differences makes almost no difference to the temperature analysis, since the temperature spectra that enter the likelihood are signal dominated over most of the multipole range. However, the odd-even noise estimates give higher noise levels than half-ring difference estimates at multipoles $\lesssim 500$ (in qualitative agreement with end-to-end simulations), and this improves the χ^2 of the polarization spectra. This differs from the `Plik` likelihood, which uses the half-ring difference maps to estimate the noise levels, together with a correction to compensate for correlated noise, as described in PPL18.
- In PCP15, we used power-spectrum templates for the CIB from the halo models described in [281]. The overall amplitude of the CIB power spectrum at 217 GHz was allowed to vary as one of the "nuisance" parameters in the likelihood, but the relative amplitudes at 143×217 and 143×143 were fixed to the values given by the model. In the 2018 analysis, we retain the template shapes from [281], but allow free amplitudes at 217×217, 143×217, and 143×143. The CIB is ignored at 100 GHz. We made these changes to the 2018 `CamSpec` likelihood to reduce any source of systematic bias associated with the specific model of [281], since this model is uncertain at low frequencies and fails to match *Herschel*-SPIRE measurements [361] of the CIB anisotropies at 350 and 500 μm for $\ell \gtrsim 3000$ [228]. This change was implemented to see whether it had any impact on the value of the lensing parameter A_L (see Sect. 27.6.2); however, it has a negligible effect on A_L or on other cosmological parameters. The `Plik` likelihood retains the 2015 model for the CIB.
- In PCP15 we used a single functional form for the Galactic dust power spectrum template, constructed by computing differences of 545×545 power spectra determined using different masks. The dust template was then rescaled to match the dust amplitudes at lower frequencies for the masks used to form the likelihood. In the 2018 `CamSpec` likelihood we use dust templates computed from the 545×545 spectra, using masks with exactly the same point-source holes as those used to compute the 100×100, 143×143, 143×217, and 217×217 power spectra that are used in the likelihood. The `Plik` likelihood adopts a similar approach and the `CamSpec` and `Plik` dust templates are in very good agreement.

In forming the temperature likelihood, we apply multipole cuts to the temperature spectra

as follows: $\ell_{\rm min} = 30$, $\ell_{\rm max} = 1200$ for the 100×100 spectrum; $\ell_{\rm min} = 30$, $\ell_{\rm max} = 2000$ for the 143×143 spectrum; and $\ell_{\rm min} = 500$, $\ell_{\rm max} = 2500$ for 143×217 and 217×217. As discussed in previous papers, the $\ell_{\rm min}$ cuts applied to the 143×217 and 217×217 spectra are imposed to reduce any potential systematic biases arising from Galactic dust at these frequencies. A foreground model is included in computing the covariance matrices, assuming that foregrounds are isotropic and Gaussian. This model underestimates the contribution of Galactic dust to the covariances, since this component is anisotropic on the sky. However, dust always makes a very small contribution to the covariance matrices in the `CamSpec` likelihood. [228] describe a simple model to account for the Galactic dust contributions to covariance matrices.

It is important to emphasize that these changes to the 2018 `CamSpec` TT likelihood are largely cosmetic and have very little impact on cosmological parameters. This can be assessed by comparing the `CamSpec` TT results reported in this paper with those in PCP15. The main changes in cosmological parameters from the TT likelihood come from the tighter constraints on the optical depth, τ, adopted in this paper.

In polarization, `CamSpec` uses a different methodology to `Plik`. In temperature, there are a number of frequency-dependent foregrounds at high multipoles that are described by a physically motivated parametric model containing "nuisance" parameters. These nuisance parameters are sampled, along with cosmological parameters, during Markov chain Monte Carlo (MCMC) exploration of the likelihood. The TT likelihood is therefore a power-spectrum-based component-separation tool and it is essential to retain cross-power spectra for each distinct frequency combination. For the *Planck* TE and EE spectra, however, Galactic dust is by far the dominant foreground contribution. At the multipoles and sensitivities accessible to *Planck*, polarized point sources make a negligible contribution to the foreground [as verified by ACTPol and SPTpol; 226, 169], so the only foreground that needs to be subtracted is polarized Galactic dust emission. As described in PCP15, we subtract polarized dust emission from each TE/ET and EE spectrum using the 353-GHz half-mission maps. This is done in an analogous way to the construction of 545-GHz-cleaned temperature maps described in PCP15 and Appendix **??**. Since the 353-GHz maps are noisy at high multipoles we use the cleaned spectra at multipoles ≤ 300 and extrapolate the dust model to higher multipoles by fitting power laws to the dust estimates at lower multipoles.

The polarization spectra are then corrected for temperature-to-polarization leakage and effective polarization efficiencies as described below, assuming a fiducial theoretical power spectrum. The corrected TE/ET spectra and EE spectra for all half-mission cross-spectra constructed from 100-, 143-, and 217-GHz maps are then coadded to form a single TE spectrum and a single EE spectrum for the `CamSpec` likelihood. The polarization part of the `CamSpec` likelihood therefore contains no nuisance parameters other than overall calibration factors c_{TE} and c_{EE} for the TE and EE spectra. Since the `CamSpec` likelihood uses coadded TE and EE spectra, we do not need to bin the spectra to form a TT,TE,EE likelihood. The polarization masks used in `CamSpec` are based on 353–143 GHz polarization maps that are degraded in resolution and thresholded on $P = (Q^2 + U^2)^{1/2}$. The default `CamSpec` polarization mask used for the 2018 analysis preserves a fraction $f_{\rm sky} = 57.7\,\%$ and is apodized to give an effective sky fraction (see equation 10 of PCP15) of $f^{\rm W}_{\rm sky} = 47.7\,\%$. We use the same polarization mask for all frequencies. The `CamSpec` polarization masks differ from those used in the `Plik` likelihood, which uses intensity-thresholded masks in polarization (and therefore a larger effective sky area in polarization, as described in the previous section).

To construct covariance matrices, temperature-to-polarization leakage corrections, and effective polarization efficiencies, we need to adopt a fiducial model. For the 2018 analysis, we adopted the best-fit `CamSpec` base-ΛCDM model from PCP15 to construct a likelihood from the 2018 temperature maps. We then ran a minimizer on the TT likelihood, imposing a prior of $\tau = 0.05 \pm 0.02$, and the best-fit base-ΛCDM cosmology was adopted as our fiducial model. To deal with temperature-to-polarization leakage, we used the `QuickPol` polarized beam matrices to compute corrections to the TE and EE spectra assuming the fiducial model. The temperature-to-polarization leakage corrections are relatively small for TE spectra (although they have some impact on cosmological parameters, consistent with the behaviour of the `Plik` likelihood described in the previous section), but are negligible for EE spectra.

To correct for effective polarization efficiencies (including large-scale transfer functions arising from errors in the polarized beams) we recalibrated each TE, ET, and EE spectrum against

the fiducial model spectra by minimizing

$$\chi^2 = \sum_{\ell_1\ell_2} (C^{\rm D}_{\ell_1} - \alpha_{\rm P} C^{\rm Th}_{\ell_1}) {\rm M}^{-1}_{\ell_1\ell_2} (C^{\rm D}_{\ell_2} - \alpha_{\rm P} C^{\rm Th}_{\ell_2}), \tag{1197}$$

with respect to $\alpha_{\rm P}$, where $C^{\rm D}_\ell$ is the beam-corrected data spectrum (TE, ET, or EE) corrected for temperature-to-polarization leakage, M is the covariance matrix for the appropriate spectrum, and the sums extend over $200 \leq \ell \leq 1000$. We calibrate each TE and EE spectrum individually, rather than computing map-based polarization calibrations. Although there is a good correspondence between spectrum-based calibrations and map-based calibrations, we find evidence for some differences, particularly for the 143×143 EE spectrum in agreement with the `Plik` analysis. Unlike `Plik`, we adopt spectrum-based calibrations of polarization efficiencies in preference to map-based calibrations.

As in temperature, we apply multipole cuts to the polarization spectra prior to coaddition in order to reduce sensitivity to dust subtraction, beam estimation, and noise modelling. For TE/ET spectra we use: $\ell_{\rm min} = 30$ and $\ell_{\rm max} = 1200$ for the 100×100, 100×143 and 100×217 spectra; $\ell_{\rm min} = 30$ and $\ell_{\rm max} = 2000$ for 143×143 and 143×217; and $\ell_{\rm min} = 500$ and $\ell_{\rm max} = 2500$ for the 217×217 cross-spectrum. For EE, we use: $\ell_{\rm min} = 30$ and $\ell_{\rm max} = 1000$ for 100×100; $\ell_{\rm min} = 30$ and $\ell_{\rm max} = 1200$ for 100×143; $\ell_{\rm min} = 200$ and $\ell_{\rm max} = 1200$ for 100×217; $\ell_{\rm min} = 30$ and $\ell_{\rm max} = 1500$ for 143×143; $\ell_{\rm min} = 300$ and $\ell_{\rm max} = 2000$ for 143×217; and $\ell_{\rm min} = 500$ and $\ell_{\rm max} = 2000$ for 217×217. Since dust is subtracted from the polarization spectra, we do not include a dust model in the polarization covariance matrices. Note that at low multipoles, $\ell \lesssim 300$, Galactic dust dominates over the CMB signal in EE at all frequencies. We experimented with different polarization masks and different multipole cuts and found stable results from the `CamSpec` polarization likelihood.

To summarize, for the TT data `Plik` and `CamSpec` use very similar methodologies and a similar foreground model, and the power spectra used in the likelihoods only differ in the handling of missing pixels. As a result, there is close agreement between the two temperature likelihoods. In polarization, different polarization masks are applied and different methods are used for correcting Galactic dust, effective polarization calibrations, and temperature-to-polarization leakage. In addition, the polarization covariance matrices differ at low multipoles. As described in Appendix **??**, the two codes give similar results in polarization for base ΛCDM and most of the extensions of ΛCDM considered in this paper, and there would be no material change to most of the science conclusions in this paper were one to use the `CamSpec` likelihood in place of `Plik`. However, in cases where there are differences that could have an impact on the scientific interpretation (e.g., for $A_{\rm L}$, $\sum m_\nu$, and Ω_K) we show results from both codes. This should give the reader an impression of the sensitivity of the science results to different methodologies and choices made in constructing the polarization blocks of the high-multipole likelihoods.

The low-ℓ likelihood

The HFI low-ℓ polarization likelihood is based on the full-mission HFI 100-GHz and 143-GHz Stokes Q and U low-resolution maps, cleaned through a template-fitting procedure using LFI 30-GHz [290] and HFI 353-GHz maps,[139] which are used as tracers of polarized synchrotron and thermal dust, respectively (for details about the cleaning procedure see PPL18). Power spectra are calculated based on a quadratic maximum-likelihood estimation of the cross-spectrum between the 100- and 143-GHz data, and the multipole range used spans $\ell = 2$ to $\ell = 29$.

We only use the EE likelihood ("lowE") for the main parameter results in this paper. The likelihood code, called `SimAll`, is based on the power spectra. It is constructed using an extension of the `SimBaL` algorithm presented in Planck Collaboration Int. XLVI [297], using 300 end-to-end simulations characterizing the HFI noise and residual systematics [see 291, for details] to build an empirical probability distribution of the EE spectra (ignoring the off-diagonal correlations). The TE spectrum at low multipoles does not provide tight constraints compared to EE because of cosmic variance. However, PPL18 discusses the TE spectra at low multipoles constructed by cross-correlating the `Commander` component-separated map with the

[139] The polarized synchrotron component is fitted only at 100 GHz, being negligible at 143 GHz. For the polarized dust component, following the prescription contained in Planck Collaboration III [291], the low-ℓ HFI polarization likelihood uses the 353-GHz polarization-sensitive-bolometer-only map.

100- and 143-GHz maps. The TE spectra show excess variance compared to simulations at low multipoles, most notably at $\ell = 5$ and at $\ell = 18$ and 19, for reasons that are not understood. No attempt has been made to fold in `Commander` component-separation errors in the statistical analysis. We have therefore excluded the TE spectrum at low multipoles (with the added benefit of simplifying the construction of the `SimAll` likelihood). Little information is lost by discarding the TE spectrum. Evidently, further work is required to understand the behaviour of TE at low multipoles; however, as discussed in PPL18, the τ constraint derived from TE to $\ell_{\rm max} = 10$ ($\tau = 0.051 \pm 0.015$) is consistent with results derived from the `SimAll` EE likelihood summarized below.

Using the `SimAll` likelihood combined with the low-ℓ temperature `Commander` likelihood [see 292], varying $\ln(10^{10} A_{\rm s})$ and τ, but fixing other cosmological parameters to those of a fiducial base-ΛCDM model (with parameters very close to those of the baseline ΛCDM cosmology in this paper), PPL18 reports the optical depth measurement[140] Planck2018CosmolParam/

$$Planck2018CosmolParam/\tau = 0.0506 \pm 0.0086 \quad (68\,\%,\ \text{lowE}). \tag{1198}$$

This is significantly tighter than the LFI-based constraint used in the 2015 release ($\tau = 0.067 \pm 0.022$), and differs by about half a sigma from the result of Planck Collaboration Int. XLVI [297] ($\tau = 0.055 \pm 0.009$). The latter change is driven mainly by the removal of the last 1000 scanning rings in the 2018 `SRoll` maps, higher variance in the end-to-end simulations, and differences in the 30-GHz map used as a synchrotron tracer [see appendix A of 290]. The impact of the tighter optical depth measurement on cosmological parameters compared to the 2015 release is discussed in Sect. 27.3.6. The error model in the final likelihood does not fully include all modelling uncertainties and differences between likelihood codes, but the different approaches lead to estimates of τ that are consistent within their respective $1\,\sigma$ errors.

In addition to the default `SimAll` lowE likelihood used in this paper, the LFI polarization likelihood has also been updated for the 2018 release, as described in detail in PPL18. It gives consistent results to `SimAll`, but with larger errors ($\tau = 0.063 \pm 0.020$); we give a more detailed comparison of the various τ constraints in Sect. **??**.

The low-ℓ temperature likelihood is based on maps from the `Commander` component-separation algorithm, as discussed in detail in Planck Collaboration IV [292], with a Gibbs-sample-based Blackwell-Rao likelihood that accurately accounts for the non-Gaussian shape of the posterior at low multipoles, as in 2015. The CMB maps that are used differ in several ways from the 2015 analysis. Firstly, since the 2018 analysis does not produce individual bolometer maps (since it is optimized to reduce large-scale polarization systematics) the number of foreground components that can be constrained is reduced compared to 2015. The 2018 `Commander` analysis only fits the CMB, a single general low-frequency power-law component, thermal dust, and a single CO component with spatially constant line ratios between 100, 217, and 353 GHz. Secondly, the 2018 analysis is based only on *Planck* data and so does not including the WMAP and Haslam 408-MHz maps. Finally, in order to be conservative with respect to CO emission, the sky fraction has been reduced to 86 % coverage, compared to 93 % in 2015. The net effect is a small increase in errors, and the best-fit data points are correspondingly slightly more scattered compared to 2015. The (arbitrary) normalization of the `Commander` likelihood was also changed, so that a theory power spectrum equal to the best-fit power spectrum points will, by definition, give $\chi^2_{\rm eff} = 0$.

Likelihood notation

Throughout this paper, we adopt the following labels for likelihoods: (i) *Planck* TT+lowE denotes the combination of the high-ℓ TT likelihood at multipoles $\ell \geq 30$, the low-ℓ temperature-only `Commander` likelihood, and the low-ℓ EE likelihood from `SimAll`; (ii) labels such as *Planck* TE+lowE denote the TE likelihood at $\ell \geq 30$ plus the low-ℓ EE `SimAll` likelihood; and (iii)

[140] The corresponding marginalized amplitude parameter is $\ln(10^{10} A_{\rm s}) = 2.924 \pm 0.052$, which gives $A_{\rm s}$ about 10 % lower than the value obtained from the joint fits in Sect. 27.3. The τ constraints quoted here are lower than the joint results, since the small-scale power has a preference for higher $A_{\rm s}$ (and hence higher τ for the well-measured $A_{\rm s} e^{-2\tau}$ combination) at high multipoles, related to the preference for more lensing discussed in Sect. 27.6.

Planck TT,TE,EE+lowE denotes the combination of the combined likelihood using TT, TE, and EE spectra at $\ell \geq 30$, the low-ℓ temperature `Commander` likelihood, and the low-ℓ `SimAll` EE likelihood. For brevity we sometimes drop the "*Planck*" qualifier where it should be clear, and unless otherwise stated high-ℓ results are based on the `Plik` likelihood. TE correlations at $\ell \leq 29$ are not included in any of the results presented in this paper.

Uncertainties on cosmological parameters

To maximize the accuracy of the results, various choices can be made in the construction of the high-multipole likelihoods. Examples of these are the sky area, noise models, multipole ranges, frequencies, foreground parameterization, and priors, as detailed for this release of *Planck* data in PPL18. The cosmological parameters and their uncertainties depend on these options. It is therefore necessary to test the sensitivity of the results with respect to such choices. In particular, when removing or adding independent information (e.g., by lifting or adding priors, or by measuring parameters from different multipole ranges), we *do expect* cosmological parameters to shift. The crucial question, however, is whether these are in agreement with statistical expectations. If they are consistent with being statistical excursions, then the noise model, along with foreground and instrumental nuisance parameters (e.g., polarization efficiencies), may be a consistent representation of the data. In this case, the uncertainties quoted in this paper should accurately describe the combined noise and sample variance due to finite data. Different choices of sky area, multipole range, etc., will produce changes in the parameters, but they will be adequately described by the quoted uncertainties. On the other hand, if the shifts *do not* agree with statistical expectations, they might be an indication of unmodelled systematic effects.

In PPL18 we discuss a series of tests indicating the overall robustness of our results. Internal to the `Plik` likelihood code, we consider the CMB spectra, errors, and resulting parameters as we vary the input data, ℓ range, sky area, etc. We also consider the effect of known sources of systematic uncertainty, such as high-frequency oscillations in the raw time-ordered data and temperature-to-polarization leakage. We further test the baseline likelihood using extensive simulations; these tests demonstrate the solidity of our results. As a specific example, when lifting all priors on nuisance parameters (such as calibration and foregrounds), the posterior mean on the number of relativistic species $N_{\rm eff}$ shifts upwards by about $1\,\sigma$. We quantify in PPL18 that this is statistically not anomalous, since lifting priors reduces information and, as a consequence, error bars also increase.

Only in a small number of areas, do such tests show mild internal disagreements at the level of spectra and parameters. One example is the higher than expected χ^2 of the `Plik` TE frequency-likelihood, which can be traced back to a small mismatch between the different cross-frequency spectra. When we co-add the foreground-cleaned frequency TE spectra into one CMB spectrum (which is less sensitive to such a mismatch), the related χ^2 is in better agreement with expectations. A second example is the choice of polarization-efficiency corrections, which has a small impact on the final results and is further discussed below.

We have also compared the results from the `Plik` likelihood with those obtained with `CamSpec` in Sects. 27.2.2 and 27.2.2 and Appendix **??**, as well as in PPL18 [see also 117]. Some of the likelihood choices (e.g., sky area and multipole range) will give different detailed results within the expected sample variance. Others, such as the models for noise (bias-corrected half-ring difference for `Plik` versus odd-even rings for `CamSpec`) and polarization efficiency, may give a hint of residual systematic uncertainties. If we restrict ourselves to temperature, the `Plik` and `CamSpec` likelihoods are in excellent accord, with most parameters agreeing to better than $0.5\,\sigma$ ($0.2\,\sigma$ on the ΛCDM model). On the other hand, we find indications (discussed in more detail in PPL18) that the polarization efficiencies of the frequency-channel maps differ when measured in the TE or EE spectra, and the `Plik` and `CamSpec` likelihoods have explored different choices of polarization efficiency corrections. This and polarization-noise modelling may be responsible for differences in the details of the resulting polarization spectra and parameters.

For the base-ΛCDM model, the results from `Plik` and `CamSpec` for the TT,TE,EE likelihoods are in good agreement (see Table 24), again with most parameters agreeing to better than $0.5\,\sigma$. We also find differences between the `Plik` and `CamSpec` TTTEEE likelihoods for some extended models, especially for the single-parameter extensions with $A_{\rm L}$ (at $0.7\,\sigma$) and $\Omega_{\rm K}$ (at $0.5\,\sigma$); these differences are discussed in Sects. 27.6.2 and **??**, respectively, where we show results for both likelihoods. For both $A_{\rm L}$ and $\Omega_{\rm K}$, the `Plik` TT,TE,EE likelihood pulls away from the base-ΛCDM model with a slightly higher significance than the `CamSpec` TT,TE,EE likelihood. The

is due, at least in part, to the choice of how to model polarization efficiencies, as discussed in PPL18. For the Ω_K case, for example, the $\Delta\chi^2$ between the ΛCDM and ΛCDM+Ω_K models for TT,TE,EE+lowE is $\Delta\chi^2 = 11$, of which 8.3 $\Delta\chi$ points are due to the improvement of the `Plik` TT,TE,EE likelihood. Using spectrum-based polarization efficiencies, instead of map-based ones[141] reduces that total difference to $\Delta\chi^2 = 5.2$, of which $\Delta\chi^2 = 4.6$ is due to the `Plik` likelihood. This is in agreement with the $\Delta\chi^2$ value obtained for these models by `CamSpec`, which uses spectrum-based polarization efficiencies, with $\Delta\chi^2 = 4.3$.

Other details of choices in the likelihood functions impact the difference in parameters; however, these comprise both expected statistical fluctuations (due to differing raw data cuts and sky coverage) and possible residual systematic errors. For both extended models the *Planck* TTTEEE likelihoods are usually combined with other data to break parameter degeneracies. For these parameters, the addition of either *Planck* lensing or BAO data overwhelms any differences between the `Plik` and `CamSpec` likelihoods and so we find almost identical results.

In this paper we therefore do not explicitly model an increase in error bars due to these residual systematic errors — any such characterization would inevitably be incomplete, and it would also be impossible to give the necessary probabilistic characterization required for meaningful quantitative error bars. Instead our best-fit values, posterior means, errors and limits should (as always) be considered as conditional on the cosmological model and our best knowledge of the *Planck* instruments and astrophysical foregrounds, as captured by the baseline likelihoods.

27.2.3 The CMB lensing likelihood

The CMB photons that arrive here today traverse almost the entire observable Universe. Along the way their paths are deflected by gradients in the gravitational potentials associated with inhomogeneities in the Universe [55]. The dominant effects [e.g., 220, 161] are a smoothing of the acoustic peaks, conversion of E-mode polarization to B-mode polarization, and generation of a connected 4-point function, each of which can be measured in high angular resolution, low-noise observations, such as those from *Planck*.

Planck was the first experiment to measure the lensing signal to sufficient precision for it to become important for the determination of cosmological parameters, providing sensitivity to parameters that affect the late-time expansion, geometry, and clustering [277, hereafter PL2013]. In Planck Collaboration XV [286, hereafter PL2015] the *Planck* lensing reconstruction was improved by including polarization information. The *Planck* lensing measurement is still the most significant detection of CMB lensing to date. In this final data release we report a measurement of the power spectrum of the lensing potential, $C_L^{\phi\phi}$, from the 4-point function, with a precision of around 2.6 % on the amplitude, as discussed in detail in PL2018. We demonstrate the robustness of the reconstruction to a variety of tests over lensing multipoles $8 \leq L \leq 400$, and conservatively restrict the likelihood to this range to reduce the impact of possible systematics. Compared to 2015, the multipole range is extended from $L_{\rm min} = 40$ down to $L_{\rm min} = 8$, with other analysis changes mostly introducing random fluctuations in the band powers, due to improvements in the noise modelling and the somewhat different mixture of frequencies being used in the foreground-cleaned `SMICA` maps [see 292]. The signal-to-noise per multipole is almost the same as in 2015, which, combined with the wider multipole range, makes the likelihood just slightly more powerful than in 2015. CMB lensing can provide complementary information to the *Planck* CMB power spectra, since it it probes much lower redshifts, including $z \lesssim 2$, when dark energy becomes important. The lensing effect depends on the propagation of photons on null geodesics, and hence depends on the background geometry and Weyl potential (the combination of scalar metric perturbations that determines the Weyl spacetime curvature tensor; see e.g. [220]).

We approximate the lensing likelihood as Gaussian in the estimated band powers, making perturbative corrections for the small dependence of band powers on the cosmology, as described in PL2015. We neglect correlations between the 2- and 4-point functions, which are negligible at *Planck* sensitivity [327, 261]. As in PL2015, band powers at multipoles $L > 400$ are less robust than over $8 \leq L \leq 400$, with some evidence for a curl-test failure, and possibly also systematic differences between individual frequencies that we were unable to resolve. Multipoles at $L < 8$

[141] As explained in Sections 27.2.2 and 27.2.2, the "map-based" approach applies the same polarization efficiency corrections estimated from EE to both the TE and EE spectra, while the "spectrum-based" approach applies independent estimates obtained from TE and EE to the TE and EE spectra, respectively.

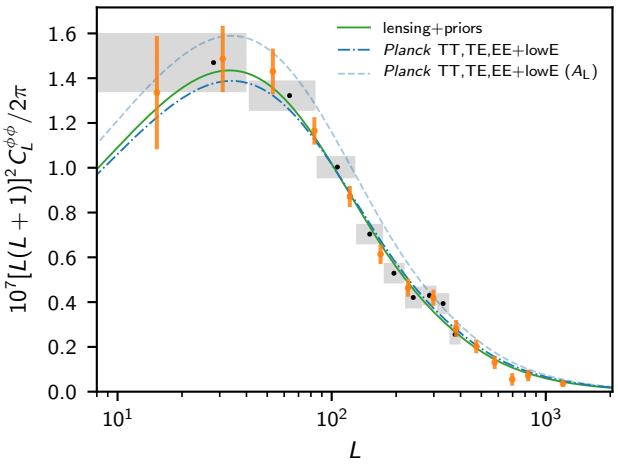


Figure 169: CMB lensing-potential power spectrum, as measured by *Planck* (see PL2018 for a detailed description of this measurement). Orange points show the full range of scales reconstructed with a logarithmic binning, while grey bands show the error and multipole range of the conservative band powers used for the likelihood, with black points showing the average multipole of the band weight. The solid line shows the best ΛCDM fit to the conservative points alone, and the dot-dashed line shows the prediction from the best fit to the *Planck* CMB power spectra alone. The dashed line shows the prediction from the best fit to the CMB power spectra when the lensing amplitude $A_{\rm L}$ is also varied ($A_{\rm L} = 1.19$ for the best-fit model; see Sect. 27.6.2 for a detailed discussion of $A_{\rm L}$).

are very sensitive to the large mean-field correction on these scales, and hence are sensitive to the fidelity of the simulations used to estimate the mean field. As described above, our baseline cosmological results therefore conservatively use only the multipole range $8 \leq L \leq 400$.

The *Planck* measurements of $C_L^{\phi\phi}$ are plotted in Fig. 169, where they are compared to the predicted spectrum from the best-fitting base-ΛCDM model of Sect. 27.3, and Fig. 170 shows the corresponding broad redshift ranges that contribute to the lensing band powers in the ΛCDM model. Fig. 169 shows that the lensing data are in excellent agreement with the predictions inferred from the CMB power spectra in the base-ΛCDM model ($\chi^2_{\rm eff} = 8.9$ for 9 binned conservative band-power measurements, $\chi^2_{\rm eff} = 14.0$ for 14 bins over the full multipole range; we discuss agreement in extensions to the ΛCDM model in more detail below). The lensing data prefer lensing power spectra that are slightly tilted towards less power on small scales compared to the best fit to the CMB power spectra. This small tilt pulls joint constraints a small fraction of an error bar towards parameters that give a lower lensing amplitude on small scales. Parameter results from the full multipole range would be a little tighter and largely consistent with the conservative band powers, although preferring slightly lower fluctuation amplitudes (see PL2018).

As described in detail in PL2018, the lensing likelihood (in combination with some weak priors) can alone provide ΛCDM parameter constraints that are competitive with current galaxy lensing and clustering, measuring Planck2018CosmolParam/

$$Planck2018CosmolParam/\sigma_8\Omega_{\rm m}^{0.25} = 0.589 \pm 0.020 \quad \begin{matrix}(68\,\%,\ Planck \\ \text{lensing}).\end{matrix} \tag{1199}$$

Combined with BAO (see Sect. 27.5.1 below) and a baryon density prior to break the main degeneracy between H_0, $\Omega_{\rm m}$, and σ_8 (described in PL2015), individual parameters H_0, $\Omega_{\rm m}$, and σ_8 can also separately be constrained to a precision of a few percent. We use $\Omega_{\rm b}h^2 = 0.0222 \pm 0.0005$ [motivated by the primordial deuterium abundance measurements of 82, see

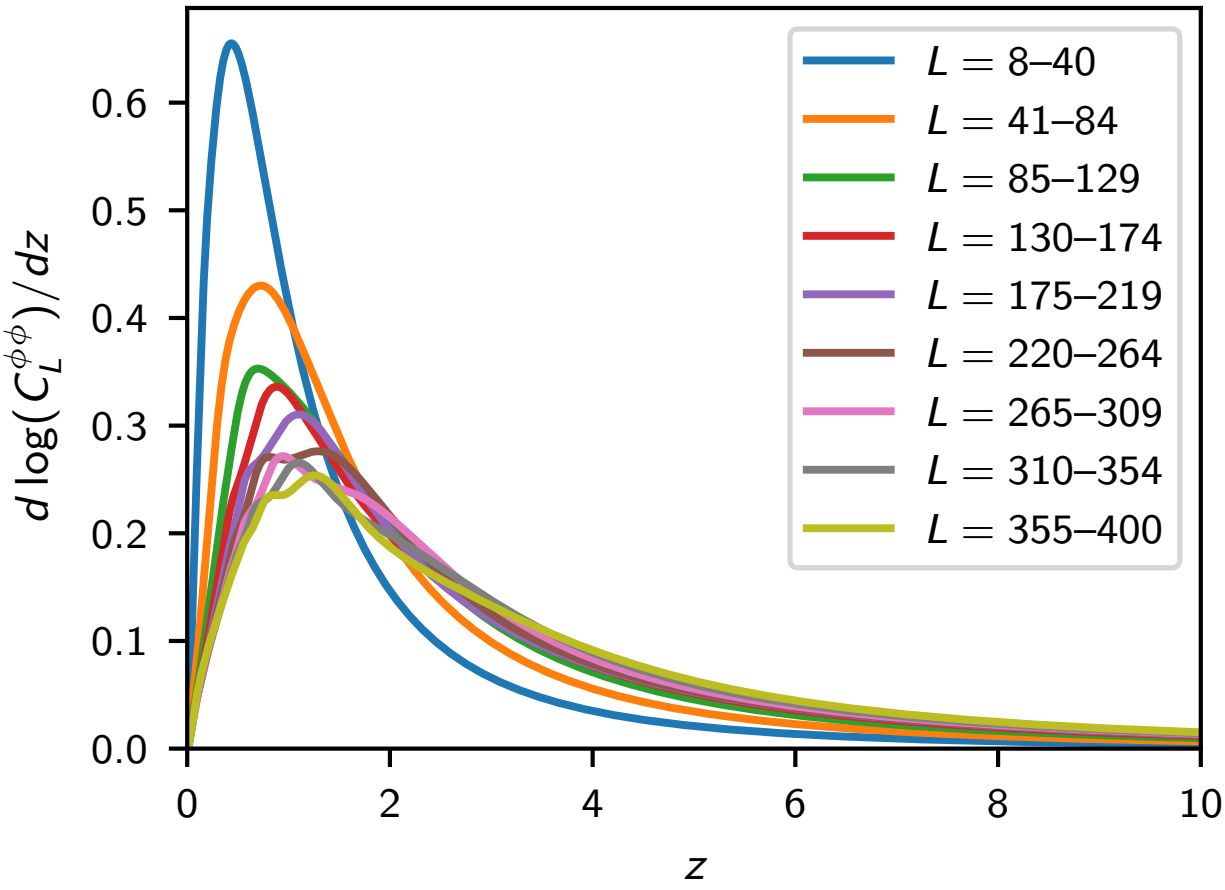


Figure 170: Contributions to the conservative CMB lensing band powers (see text and Fig. 169) as a function of redshift in the base-ΛCDM model (evaluated here, and only here, using the Limber approximation [227] on all scales). Multipole ranges of the corresponding band powers are shown in the legend.

also Sect. **??**], which gives

$$\left.\begin{aligned} Planck2018CosmolParam/H_0 &= 67.9^{+1.2}_{-1.3}\ \mathrm{km\ s^{-1}Mpc^{-1}},\\ \sigma_8 &= 0.811 \pm 0.019,\\ \Omega_\mathrm{m} &= 0.303^{+0.016}_{-0.018}, \end{aligned}\right\}\ 68\,\%, \text{lensing+BAO}. \qquad (1200)$$

The constraints of Eq. (1199) and (1200) in are in very good agreement with the estimates derived from the *Planck* power spectra and are independent of how the *Planck* power spectra depend on the cosmological model at high multipoles. This is a strong test of the internal consistency of the *Planck* data. The *Planck* lensing constraints in Eqs. (1199) and (1200), and the consistency of these results with the *Planck* power spectrum likelihoods, should be borne in mind when comparing *Planck* results with other astrophysical data (e.g., direct measurements of H_0 and galaxy shear surveys, see Sect. 27.5).

In this paper, we focus on joint constraints with the main *Planck* power spectrum results, where the lensing power spectrum tightens measurements of the fluctuation amplitude and improves constraints on extended models, especially when allowing for spatial curvature.

A peculiar feature of the *Planck* TT likelihood, reported in PCP13 and PCP15, is the favouring of high values for the lensing consistency parameter A_L (at about $2.5\,\sigma$). This result is discussed in detail in Sect. 27.6.2. It is clear from Fig. 169, however, that the *Planck* lensing likelihood prefers values of A_L close to unity and cosmological parameters that are close

to those of the best-fit base-ΛCDM parameters derived from the *Planck* TT,TE,EE+lowE +lensing likelihood (i.e., without allowing $A_{\rm L}$ to vary).

27.3 Constraints on base ΛCDM

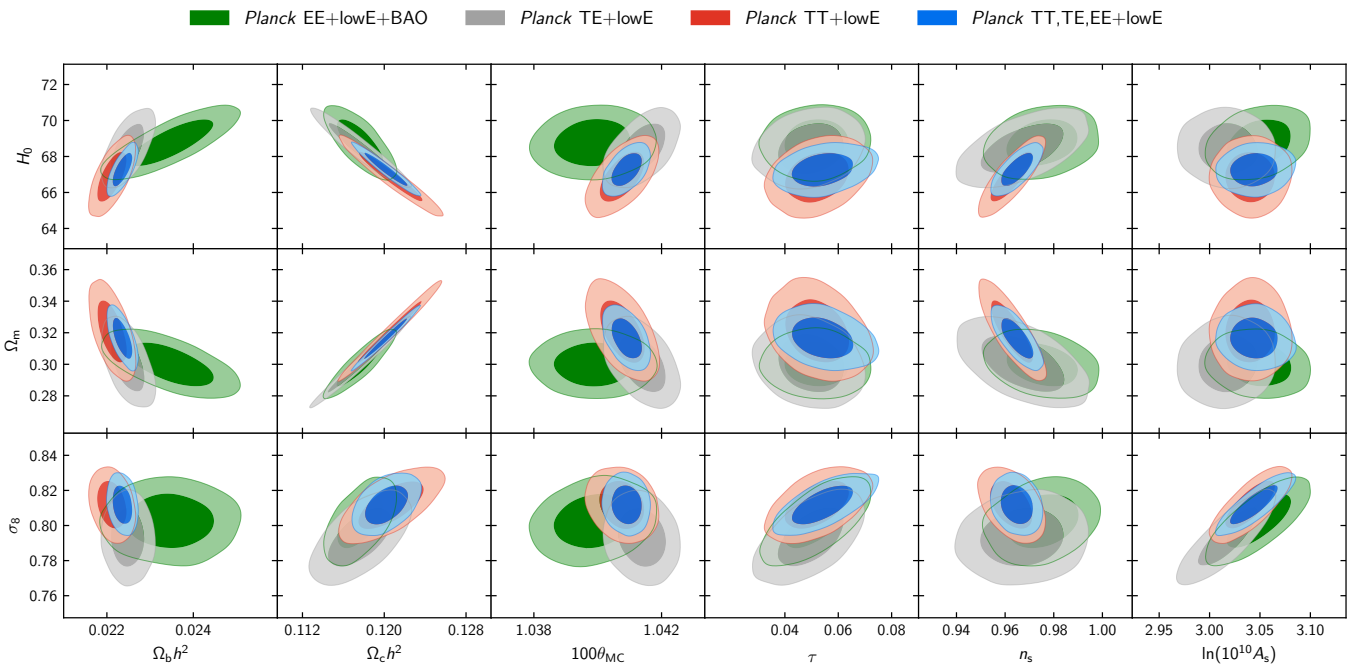


Figure 171: Constraints on parameters of the base-ΛCDM model from the separate *Planck* EE, TE, and TT high-ℓ spectra combined with low-ℓ polarization (lowE), and, in the case of EE also with BAO (described in Sect. 27.5.1), compared to the joint result using *Planck* TT,TE,EE+lowE. Parameters on the bottom axis are our sampled MCMC parameters with flat priors, and parameters on the left axis are derived parameters (with H_0 in km s^{-1}Mpc^{-1}). Contours contain 68 % and 95 % of the probability.

The *Planck* measurement of seven acoustic peaks in the CMB temperature power spectrum allows cosmological parameters to be constrained extremely accurately. In previous papers, we have focussed on parameters derived from the TT power spectrum. The TE and EE polarization spectra provide a powerful consistency check on the underlying model and also help to break some partial parameter degeneracies. The goal of this section is to explore the consistency of cosmological parameters of the base-ΛCDM cosmology determined from TT, TE, and EE spectra and to present results from the combinations of these spectra, which are significantly more precise that those determined using TT alone.

Figure 171 shows 2-dimensional marginalized constraints on the six MCMC sampling parameters of the base-ΛCDM model used to explore the parameter posteriors, plotted against the following derived parameters: the Hubble constant H_0, late-time clustering amplitude σ_8 and matter density parameter $\Omega_{\rm m}$ (defined including a 0.06-eV mass neutrino). Table 24 gives individual parameter constraints using our baseline parameter combination *Planck* TT,TE,EE +lowE+lensing. These represent the legacy results on the cosmological ΛCDM parameters from the *Planck* satellite, and are currently the most precise measurements coming from a single CMB experiment. We give the best-fit values, as well as the marginalized posterior mean values, along with the corresponding 68 % probability intervals. Table 24 also quantifies the small changes in parameters that are found when using the `Plik` and `CamSpec` high-ℓ polarization analyses described in Sect. 27.2.2 and Appendix **??**. Table 25 gives marginalized parameter constraints from the various CMB spectra, individually and without CMB lensing, including a wider variety of derived parameters of physical interest.

We now discuss in more detail the parameters that are most directly measured by the data and how these relate to constraints on individual parameters of more general interest.

Table 24: Base-ΛCDM cosmological parameters from *Planck* TT,TE,EE+lowE+lensing. Results for the parameter best fits, marginalized means and 68 % errors from our default analysis using the `Plik` likelihood are given in the first two numerical columns. The `CamSpec` likelihood results give some idea of the remaining modelling uncertainty in the high-ℓ polarization, though parts of the small shifts are due to slightly different sky areas in polarization. The "Combined" column give the average of the `Plik` and `CamSpec` results, assuming equal weight. The combined errors are from the equal-weighted probabilities, hence including some uncertainty from the systematic difference between them; however, the differences between the high-ℓ likelihoods are so small that they have little effect on the $1\,\sigma$ errors. The errors do not include modelling uncertainties in the lensing and low-ℓ likelihoods or other modelling errors (such as temperature foregrounds) common to both high-ℓ likelihoods. A total systematic uncertainty of around $0.5\,\sigma$ may be more realistic, and values should not be overinterpreted beyond this level. The best-fit values give a representative model that is an excellent fit to the baseline likelihood, though models nearby in the parameter space may have very similar likelihoods. The first six parameters here are the ones on which we impose flat priors and use as sampling parameters; the remaining parameters are derived from the first six. Note that $\Omega_{\rm m}$ includes the contribution from one neutrino with a mass of 0.06 eV. The quantity $\theta_{\rm MC}$ is an approximation to the acoustic scale angle, while θ_* is the full numerical result.

Parameter	`Plik` best fit	`Plik` [1]	`CamSpec` [2]	$([2]-[1])/\sigma_1$	Combined
$\Omega_{\rm b}h^2$	0.022383	0.02237 ± 0.00015	0.02229 ± 0.00015	-0.5	0.02233 ± 0.00015
$\Omega_{\rm c}h^2$	0.12011	0.1200 ± 0.0012	0.1197 ± 0.0012	-0.3	0.1198 ± 0.0012
$100\theta_{\rm MC}$	1.040909	1.04092 ± 0.00031	1.04087 ± 0.00031	-0.2	1.04089 ± 0.00031
τ	0.0543	0.0544 ± 0.0073	$0.0536^{+0.0069}_{-0.0077}$	-0.1	0.0540 ± 0.0074
$\ln(10^{10}A_{\rm s})$	3.0448	3.044 ± 0.014	3.041 ± 0.015	-0.3	3.043 ± 0.014
$n_{\rm s}$	0.96605	0.9649 ± 0.0042	0.9656 ± 0.0042	$+0.2$	0.9652 ± 0.0042
$\Omega_{\rm m}h^2$	0.14314	0.1430 ± 0.0011	0.1426 ± 0.0011	-0.3	0.1428 ± 0.0011
H_0 [km s^{-1}Mpc^{-1}]	67.32	67.36 ± 0.54	67.39 ± 0.54	$+0.1$	67.37 ± 0.54
$\Omega_{\rm m}$	0.3158	0.3153 ± 0.0073	0.3142 ± 0.0074	-0.2	0.3147 ± 0.0074
Age [Gyr]	13.7971	13.797 ± 0.023	13.805 ± 0.023	$+0.4$	13.801 ± 0.024
σ_8	0.8120	0.8111 ± 0.0060	0.8091 ± 0.0060	-0.3	0.8101 ± 0.0061
$S_8 \equiv \sigma_8(\Omega_{\rm m}/0.3)^{0.5}$	0.8331	0.832 ± 0.013	0.828 ± 0.013	-0.3	0.830 ± 0.013
$z_{\rm re}$	7.68	7.67 ± 0.73	7.61 ± 0.75	-0.1	7.64 ± 0.74
$100\theta_*$	1.041085	1.04110 ± 0.00031	1.04106 ± 0.00031	-0.1	1.04108 ± 0.00031
$r_{\rm drag}$ [Mpc]	147.049	147.09 ± 0.26	147.26 ± 0.28	$+0.6$	147.18 ± 0.29

27.3.1 Acoustic scale

The acoustic oscillations in ℓ seen in the CMB power spectra correspond to a sharply-defined acoustic angular scale on the sky, given by $\theta_* \equiv r_*/D_{\rm M}$ where r_* is the comoving sound horizon at recombination quantifying the distance the photon-baryon perturbations can influence, and $D_{\rm M}$ is the comoving angular diameter distance[142] that maps this distance into an angle on the sky. *Planck* measures Planck2018CosmolParam/

$$Planck2018CosmolParam/100\theta_* = 1.04097 \pm 0.00046 \quad \begin{matrix}(68\,\%,\ Planck\ \mathrm{TT} \\ +\mathrm{lowE}),\end{matrix} \tag{1201}$$

[142]The quantity $D_{\rm M}$ is $(1+z)D_{\rm A}$, where $D_{\rm A}$ is the usual angular diameter distance.

Table 25: Parameter 68 % intervals for the base-ΛCDM model from *Planck* CMB power spectra, in combination with CMB lensing reconstruction and BAO. The top group of six rows are the base parameters, which are sampled in the MCMC analysis with flat priors. The middle group lists derived parameters. The bottom three rows show the temperature foreground amplitudes $f^{TT}_{\ell=2000}$ for the corresponding frequency spectra (expressed as the contribution to $D^{TT}_{\ell=2000}$ in units of $(\mu\mathrm{K})^2$). In all cases the helium mass fraction used is predicted by BBN (posterior mean $Y_\mathrm{P} \approx 0.2454$, with theoretical uncertainties in the BBN predictions dominating over the *Planck* error on $\Omega_\mathrm{b}h^2$). The reionization redshift mid-point z_re and optical depth τ here assumes a simple tanh model (as discussed in the text) for the reionization of hydrogen and simultaneous first reionization of helium. Our baseline results are based on *Planck* TT,TE,EE +lowE+lensing (as also given in Table 24).

arameter	TT+lowE 68% limits	TE+lowE 68% limits	EE+lowE 68% limits	TT,TE,EE+lowE 68% limits	TT,TE,EE+lowE+lensing 68% limits	TT,TE,EE+lowE+lensing+ 68% limits
. . . . .	0.02212 ± 0.00022	0.02249 ± 0.00025	0.0240 ± 0.0012	0.02236 ± 0.00015	0.02237 ± 0.00015	0.02242 ± 0.00014
. . . . .	0.1206 ± 0.0021	0.1177 ± 0.0020	0.1158 ± 0.0046	0.1202 ± 0.0014	0.1200 ± 0.0012	0.11933 ± 0.00091
MC	1.04077 ± 0.00047	1.04139 ± 0.00049	1.03999 ± 0.00089	1.04090 ± 0.00031	1.04092 ± 0.00031	1.04101 ± 0.00029
. . . . .	0.0522 ± 0.0080	0.0496 ± 0.0085	0.0527 ± 0.0090	$0.0544^{+0.0070}_{-0.0081}$	0.0544 ± 0.0073	0.0561 ± 0.0071
$^{10}A_\mathrm{s}$)	3.040 ± 0.016	$3.018^{+0.020}_{-0.018}$	3.052 ± 0.022	3.045 ± 0.016	3.044 ± 0.014	3.047 ± 0.014
. . . . .	0.9626 ± 0.0057	0.967 ± 0.011	0.980 ± 0.015	0.9649 ± 0.0044	0.9649 ± 0.0042	0.9665 ± 0.0038
$\mathrm{m\,s^{-1}\,Mpc^{-1}}$]	66.88 ± 0.92	68.44 ± 0.91	69.9 ± 2.7	67.27 ± 0.60	67.36 ± 0.54	67.66 ± 0.42
. . . . .	0.679 ± 0.013	0.699 ± 0.012	$0.711^{+0.033}_{-0.026}$	0.6834 ± 0.0084	0.6847 ± 0.0073	0.6889 ± 0.0056
. . . . .	0.321 ± 0.013	0.301 ± 0.012	$0.289^{+0.026}_{-0.033}$	0.3166 ± 0.0084	0.3153 ± 0.0073	0.3111 ± 0.0056
^{2}	0.1434 ± 0.0020	0.1408 ± 0.0019	$0.1404^{+0.0034}_{-0.0039}$	0.1432 ± 0.0013	0.1430 ± 0.0011	0.14240 ± 0.00087
^{3}	0.09589 ± 0.00046	0.09635 ± 0.00051	$0.0981^{+0.0016}_{-0.0018}$	0.09633 ± 0.00029	0.09633 ± 0.00030	0.09635 ± 0.00030
. . . . .	0.8118 ± 0.0089	0.793 ± 0.011	0.796 ± 0.018	0.8120 ± 0.0073	0.8111 ± 0.0060	0.8102 ± 0.0060
$\sigma_8(\Omega_\mathrm{m}/0.3)^{0.5}$	0.840 ± 0.024	0.794 ± 0.024	$0.781^{+0.052}_{-0.060}$	0.834 ± 0.016	0.832 ± 0.013	0.825 ± 0.011
$_\mathrm{n}^{25}$	0.611 ± 0.012	0.587 ± 0.012	0.583 ± 0.027	0.6090 ± 0.0081	0.6078 ± 0.0064	0.6051 ± 0.0058
. . . . .	7.50 ± 0.82	$7.11^{+0.91}_{-0.75}$	$7.10^{+0.87}_{-0.73}$	7.68 ± 0.79	7.67 ± 0.73	7.82 ± 0.71
s	2.092 ± 0.034	2.045 ± 0.041	2.116 ± 0.047	$2.101^{+0.031}_{-0.034}$	2.100 ± 0.030	2.105 ± 0.030
$_\mathrm{s}e^{-2\tau}$	1.884 ± 0.014	1.851 ± 0.018	1.904 ± 0.024	1.884 ± 0.012	1.883 ± 0.011	1.881 ± 0.010
Gyr]	13.830 ± 0.037	13.761 ± 0.038	$13.64^{+0.16}_{-0.14}$	13.800 ± 0.024	13.797 ± 0.023	13.787 ± 0.020
. . . . .	1090.30 ± 0.41	1089.57 ± 0.42	$1087.8^{+1.6}_{-1.7}$	1089.95 ± 0.27	1089.92 ± 0.25	1089.80 ± 0.21
pc]	144.46 ± 0.48	144.95 ± 0.48	144.29 ± 0.64	144.39 ± 0.30	144.43 ± 0.26	144.57 ± 0.22
. . . . .	1.04097 ± 0.00046	1.04156 ± 0.00049	1.04001 ± 0.00086	1.04109 ± 0.00030	1.04110 ± 0.00031	1.04119 ± 0.00029
. . . . .	1059.39 ± 0.46	1060.03 ± 0.54	1063.2 ± 2.4	1059.93 ± 0.30	1059.94 ± 0.30	1060.01 ± 0.29
[Mpc]	147.21 ± 0.48	147.59 ± 0.49	146.46 ± 0.70	147.05 ± 0.30	147.09 ± 0.26	147.21 ± 0.23
$\mathrm{Mpc^{-1}}$]	0.14054 ± 0.00052	0.14043 ± 0.00057	0.1426 ± 0.0012	0.14090 ± 0.00032	0.14087 ± 0.00030	0.14078 ± 0.00028
. . . . .	3411 ± 48	3349 ± 46	3340^{+81}_{-92}	3407 ± 31	3402 ± 26	3387 ± 21
$\mathrm{Mpc^{-1}}$]	0.01041 ± 0.00014	0.01022 ± 0.00014	$0.01019^{+0.00025}_{-0.00028}$	0.010398 ± 0.000094	0.010384 ± 0.000081	0.010339 ± 0.000063
,eq	0.4483 ± 0.0046	0.4547 ± 0.0045	0.4562 ± 0.0092	0.4490 ± 0.0030	0.4494 ± 0.0026	0.4509 ± 0.0020
. . . . .	31.2 ± 3.0			29.5 ± 2.7	29.6 ± 2.8	29.4 ± 2.7
$\times 217$	33.6 ± 2.0			32.2 ± 1.9	32.3 ± 1.9	32.1 ± 1.9
. . . . .	108.2 ± 1.9			107.0 ± 1.8	107.1 ± 1.8	106.9 ± 1.8

corresponding to a precise 0.05 % measurement of the angular scale $\theta_* = (0.^\circ59643 \pm 0.^\circ00026)$. The angular scales of the peaks in the polarization spectrum and cross-spectrum are different, since the polarization at recombination is sourced by quadrupolar flows in the photon fluid, which are out of phase with the density perturbations. The polarization spectra can, however, be used to measure the same acoustic scale parameter, giving a stringent test on the assumption of purely adiabatic perturbation driving the oscillations. From the polarization spectra we find

$$100\theta_* = 1.04156 \pm 0.00049 \quad (68\,\%,\ \textit{Planck}\ \text{TE+lowE}), \tag{1202a}$$
$$100\theta_* = 1.04001 \pm 0.00086 \quad (68\,\%,\ \textit{Planck}\ \text{EE+lowE}), \tag{1202b}$$

in excellent agreement with the temperature measurement. The constraint from TE is of similar precision to that from TT: although the polarization data are much noisier, the TE and EE spectra have more distinct acoustic peaks, which helps improve the signal-to-noise ratio of the acoustic scale measurement. Using the combined likelihood we find: Planck2018CosmolParam/

$$\textit{Planck2018CosmolParam}/100\theta_* = 1.04109 \pm 0.00030 \quad (68\,\%,\ \text{TT,TE,EE+lowE}), \tag{1203}$$

a measurement with 0.03 % precision.[143]

Because of its simple geometrical interpretation, θ_* is measured very robustly and almost independently of the cosmological model (see Table **??**). It is the CMB analogue of the transverse baryon acoustic oscillation scale $r_{\rm drag}/D_{\rm M}$ measured from galaxy surveys, where $r_{\rm drag}$ is the comoving sound horizon at the end of the baryonic-drag epoch (see Sect. 27.5.1). In ΛCDM, the CMB constraint can be expressed as a tight 0.04 %-precision relation between $r_{\rm drag}\,h$ and $\Omega_{\rm m}$ as Planck2018CosmolParam/

$$\textit{Planck2018CosmolParam}/\left(\frac{r_{\rm drag}h}{\rm Mpc}\right)\left(\frac{\Omega_{\rm m}}{0.3}\right)^{0.4} = 101.056 \pm 0.036 \quad (68\,\%,\ \text{TT,TE,EE+lowE}). \tag{1204}$$

The sound horizon $r_{\rm drag}$ depends primarily on the matter, baryon, and radiation densities, which for fixed observed CMB temperature today,[144] gives a 0.05 % constraint on the combination Planck2018CosmolParam/

$$\textit{Planck2018CosmolParam}/\Omega_{\rm m}^{0.3}h(\Omega_{\rm b}h^2)^{-0.16} = 0.87498 \pm 0.00052 \quad (68\,\%,\ \text{TT,TE,EE+lowE}). \tag{1205}$$

Marginalizing out the dependence on the baryon density, the remaining degeneracy between the matter density and Hubble parameters is well approximated by a constraint on the parameter combination $\Omega_{\rm m}h^3$ [265]. We find a 0.3 % constraint from *Planck*: Planck2018CosmolParam/

$$\textit{Planck2018CosmolParam}/\Omega_{\rm m}h^3 = 0.09633 \pm 0.00029 \quad (68\,\%,\ \text{TT,TE,EE+lowE}), \tag{1206}$$

corresponding to an anti-correlation between the matter density $\Omega_{\rm m}h^2$ and the Hubble parameter. This correlation can also be seen in Fig. 171 as an anti-correlation between the dark-matter density $\Omega_{\rm c}h^2$ and H_0, and a corresponding positive correlation between $\Omega_{\rm c}h^2$ and $\Omega_{\rm m}$.

[143] Doppler aberration due to the Earth's motion means that θ_* is expected to vary over the sky at the 10^{-3} level; however, averaged over the likelihood masks, the expected bias for *Planck* is below $0.1\,\sigma$.

[144] We take $T_0 = 2.7255$K [134], with the ±0.0006K error having negligible impact on results.

27.3.2 Hubble constant and dark-energy density

The degeneracy between $\Omega_{\rm m}$ and H_0 is not exact, but the constraint on these parameters individually is substantially less precise than Eq. (1206), giving

$$\left.\begin{array}{r} Planck2018CosmolParam/H_0 = (67.27 \pm 0.60)\ \text{km s}^{-1}\text{Mpc}^{-1}, \\ \Omega_{\rm m} = 0.3166 \pm 0.0084, \end{array}\right\} \begin{array}{l} 68\,\%,\ \text{TT,TE,EE} \\ +\text{lowE}. \end{array} \quad (1207)$$

It is important to emphasize that the values given in Eq. (1207) assume the base-ΛCDM cosmology with minimal neutrino mass. These estimates are highly model dependent and this needs to be borne in mind when comparing with other measurements, for example the direct measurements of H_0 discussed in Sect. 27.5.4. The values in Eq. (1207) are in very good agreement with the independent constraints of Eq. (1200) from *Planck* CMB lensing+BAO. Including CMB lensing sharpens the determination of H_0 to a 0.8 % constraint: Planck2018CosmolParam/

$$Planck2018CosmolParam/H_0 = (67.36 \pm 0.54)\ \text{km s}^{-1}\text{Mpc}^{-1} \quad \begin{array}{l} (68\,\%,\ \text{TT,TE,EE} \\ +\text{lowE} \\ +\text{lensing}). \end{array} \quad (1208)$$

This value is our "best estimate" of H_0 from *Planck*, assuming the base-ΛCDM cosmology.

Since we are considering a flat universe in this section, a constraint on $\Omega_{\rm m}$ translates directly into a constraint on the dark-energy density parameter, giving Planck2018CosmolParam/

$$\Omega_\Lambda = 0.6847 \pm 0.0073 \quad \begin{array}{l} (68\,\%,\ \text{TT,TE,EE+lowE} \\ +\text{lensing}). \end{array} \quad (1209)$$

In terms of a physical density, this corresponds to $\Omega_\Lambda h^2 = 0.3107 \pm 0.0082$, or cosmological constant $\Lambda = (4.24 \pm 0.11) \times 10^{-66}\,\text{eV}^2 = (2.846 \pm 0.076) \times 10^{-122}\, m_{\rm Pl}^2$ in natural units (where $m_{\rm Pl}$ is the Planck mass).

27.3.3 Optical depth and the fluctuation amplitude

Since the CMB fluctuations are linear up to lensing corrections, and the lensing corrections are largely oscillatory, the average observed CMB power spectrum amplitude scales nearly proportionally with the primordial comoving curvature power spectrum amplitude $A_{\rm s}$ (which we define at the pivot scale $k_0 = 0.05\,\text{Mpc}^{-1}$). The sub-horizon CMB anisotropies are however scattered by free electrons that are present after reionization, so the observed amplitude actually scales with $A_{\rm s}e^{-2\tau}$, where τ is the reionization optical depth (see Sect. **??** for further discussion of reionization constraints). This parameter combination is therefore well measured, with the 0.6 % constraint Planck2018CosmolParam/

$$Planck2018CosmolParam/A_{\rm s}e^{-2\tau} = (1.884 \pm 0.012) \times 10^{-9} \quad \begin{array}{l} (68\,\%,\ \text{TT,TE,EE} \\ +\text{lowE}). \end{array} \quad (1210)$$

In this final *Planck* release the optical depth is well constrained by the large-scale polarization measurements from the *Planck* HFI, with the joint constraint Planck2018CosmolParam/

$$\tau = 0.0544^{+0.0070}_{-0.0081} \quad (68\,\%,\ \text{TT,TE,EE+lowE}). \quad (1211)$$

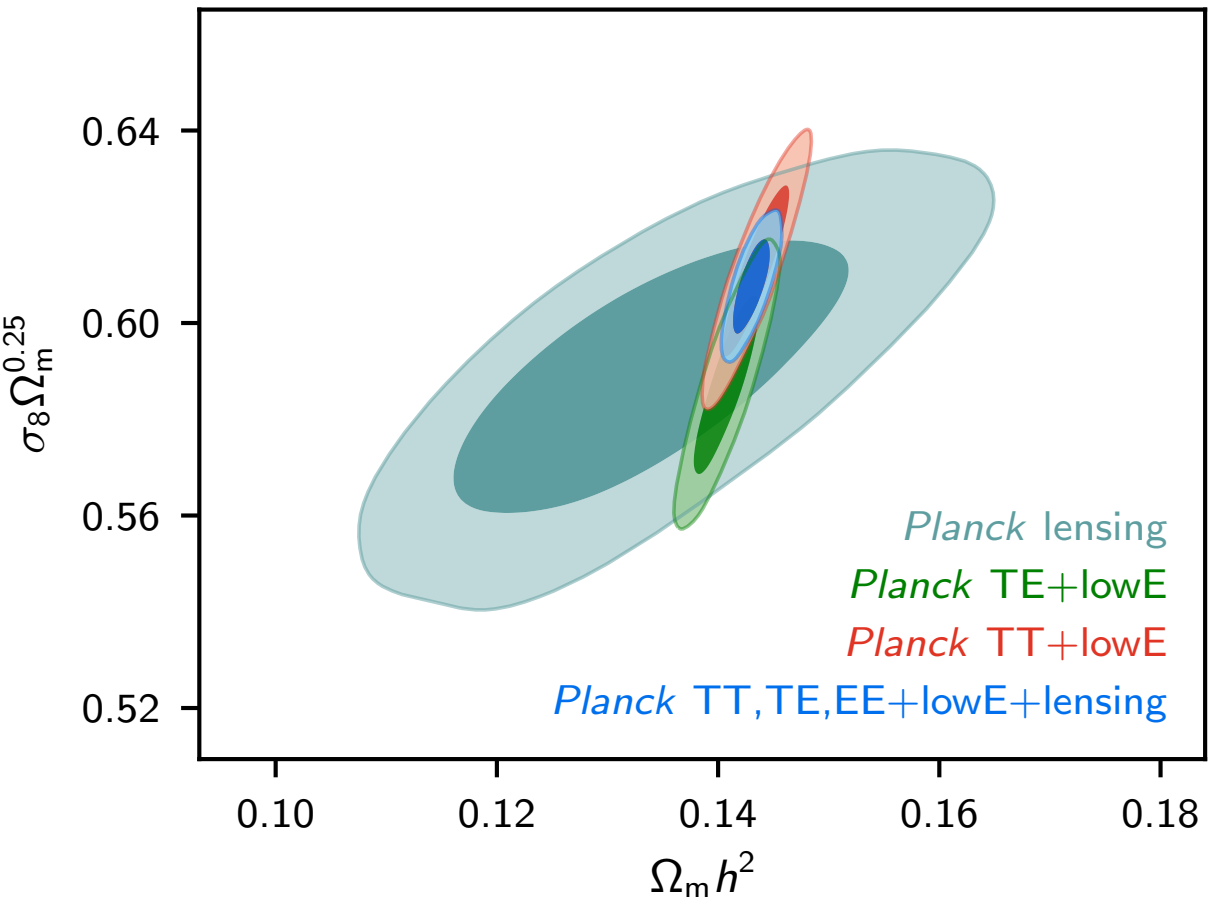


Figure 172: Base-ΛCDM 68 % and 95 % marginalized constraint contours for the matter density and $\sigma_8\Omega_m^{0.25}$, a fluctuation amplitude parameter that is well constrained by the CMB-lensing likelihood. The *Planck* TE, TT, and lensing likelihoods all overlap in a consistent region of parameter space, with the combined likelihood substantially reducing the allowed parameter space.

Assuming simple tanh parameterization of the ionization fraction,[145] this implies a mid-point redshift of reionization Planck2018CosmolParam/

$$z_{\rm re} = 7.68 \pm 0.79 \quad (68\,\%,\ \text{TT,TE,EE+lowE}), \tag{1212}$$

and a one-tail upper limit of $z_{\rm re} < 9.0$ (95 %). This is consistent with observations of high-redshift quasars that suggest the Universe was fully reionized by $z \approx 6$ [61]. We do not include the astrophysical constraint that $z_{\rm re} \gtrsim 6.5$ in our default parameter results, but if required results including this prior are part of the published tables on the Planck Legacy Archive

[145] For reference, the ionization fraction $x_{\rm e} = n_{\rm e}/n_{\rm H}$ in the tanh model is assumed to have the redshift dependence [217]:

$$x_e = \frac{1 + n_{\rm He}/n_{\rm H}}{2}\left[1 + \tanh\left(\frac{y(z_{\rm re}) - y(z)}{\Delta y}\right)\right],$$

where $y(z) = (1+z)^{3/2}$, $\Delta y = \frac{3}{2}(1+z_{\rm re})^{1/2}\Delta z$, with $\Delta z = 0.5$. Helium is assumed to be singly ionized with hydrogen at $z \gg 3$, but at lower redshifts we add the very small contribution from the second reionization of helium with a similar tanh transition at $z = 3.5$.

(PLA). A more detailed discussion of reionization histories consistent with *Planck* and results from other *Planck* likelihoods is deferred to Sect. **??**.

The measurement of the optical depth breaks the $A_s e^{-2\tau}$ degeneracy, giving a 1.5 % measurement of the primordial amplitude: Planck2018CosmolParam/

$$Planck2018CosmolParam/A_s = (2.101^{+0.031}_{-0.034}) \times 10^{-9} \quad \begin{matrix}(68\,\%,\ \mathrm{TT,TE,EE} \\ +\mathrm{lowE}).\end{matrix} \tag{1213}$$

Since the optical depth is reasonably well constrained, degeneracies with other cosmological parameters contribute to the error in Eq. (1213). From the temperature spectrum alone there is a significant degeneracy between $A_s e^{-2\tau}$ and $\Omega_m h^2$, since for fixed θ_*, larger values of these parameters will increase and decrease the small-scale power, respectively. This behaviour is mitigated in our joint constraint with polarization because the polarization spectra have a different dependence on $\Omega_m h^2$; polarization is generated by causal sub-horizon quadrupole scattering at recombination, but the temperature spectrum has multiple sources and is also sensitive to non-local redshifting effects as the photons leave the last-scattering surface [see, e.g., 142, for further discussion].

Assuming the ΛCDM model, the *Planck* CMB parameter amplitude constraint can be converted into a fluctuation amplitude at the present day, conventionally quantified by the σ_8 parameter. The CMB lensing reconstruction power spectrum also constrains the late-time fluctuation amplitude more directly, in combination with the matter density. Figure 172 shows constraints on the matter density and amplitude parameter combination $\sigma_8 \Omega_m^{0.25}$ that is well measured by the CMB lensing spectrum (see PL2015 for details). There is good consistency between the temperature, polarization, and lensing constraints here, and using their combination significantly reduces the allowed parameter space. In terms of the late-time fluctuation amplitude parameter σ_8 we find the combined result Planck2018CosmolParam/

$$\sigma_8 = 0.8111 \pm 0.0060 \quad \begin{matrix}(68\,\%,\ \textit{Planck}\ \mathrm{TT,TE,EE} \\ +\mathrm{lowE}+\mathrm{lensing}).\end{matrix} \tag{1214}$$

Measurements of galaxy clustering, galaxy lensing, and clusters can also measure σ_8, and we discuss consistency of these constraints within the ΛCDM model in more detail in Sect. 27.5.

27.3.4 Scalar spectral index

The scale-dependence of the CMB power spectrum constrains the slope of the primordial scalar power spectrum, conventionally parameterized by the power-law index n_s, where $n_s = 1$ corresponds to a scale-invariant spectrum. The matter and baryon densities also affect the scale-dependence of the CMB spectra, but in a way that differs from a variation in n_s, leading to relatively mild degeneracies between these parameters. Assuming that the primordial power spectrum is an exact power law we find Planck2018CosmolParam/

$$n_s = 0.9649 \pm 0.0042 \quad \begin{matrix}(68\,\%,\ \textit{Planck}\ \mathrm{TT,TE,EE} \\ +\mathrm{lowE}+\mathrm{lensing}),\end{matrix} \tag{1215}$$

which is $8\,\sigma$ away from scale-invariance ($n_s = 1$), confirming the red tilt of the spectrum at high significance in ΛCDM. Section **??** and [296] discuss the implications of this result for models of inflation and include constraints on models with primordial tensor modes and a scale-dependent scalar spectral index.

27.3.5 Matter densities

The matter density can be measured from the CMB spectra using the scale-dependence of the amplitude, since for fixed θ_* a larger matter density reduces the small-scale CMB power.

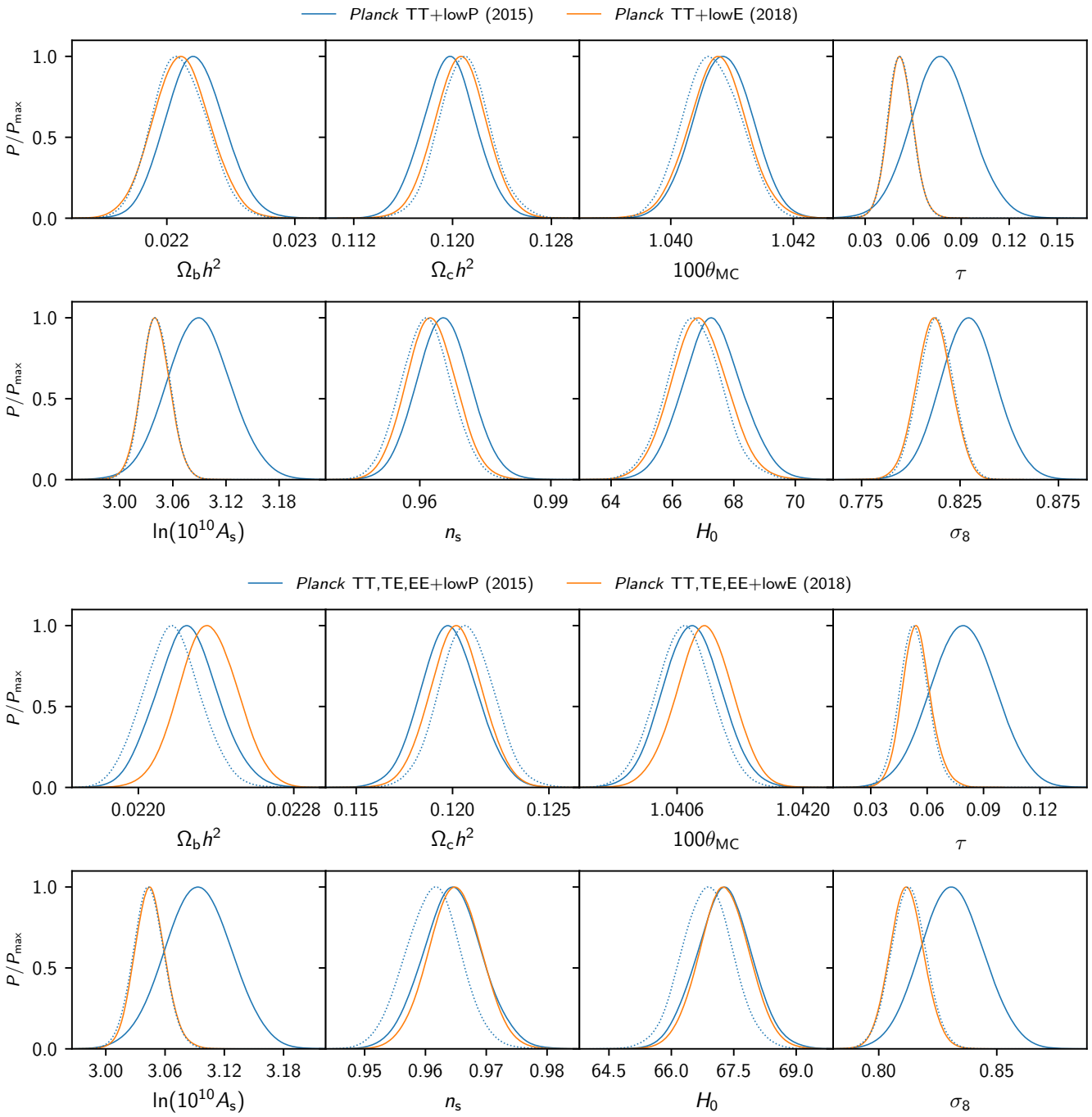


Figure 173: Comparison between the 2015 and 2018 marginalized ΛCDM parameters. Dotted lines show the 2015 results, replacing the 2015 "lowP" low-ℓ polarization likelihood with the new 2018 "lowE" `SimAll` likelihood, isolating the impact of the change in the low-ℓ polarization likelihood (and hence the constraints on τ).

The matter density also affects the amount of lensing in the CMB spectra and the amplitude of the CMB-lensing reconstruction spectrum. The matter density is well constrained to be

Planck2018CosmolParam/

$$Planck2018CosmolParam/\Omega_{\rm m}h^2 = 0.1430 \pm 0.0011 \quad \begin{array}{l}(68\,\%,\ Planck\\ \text{TT,TE,EE+lowE}\\ \text{+lensing}).\end{array} \tag{1216}$$

The matter mostly consists of cold dark matter, with density constrained at the percent level: Planck2018CosmolParam/

$$Planck2018CosmolParam/\Omega_{\rm c}h^2 = 0.1200 \pm 0.0012 \quad \begin{array}{l}(68\,\%,\ Planck\\ \text{TT,TE,EE+lowE}\\ \text{+lensing}).\end{array} \tag{1217}$$

Changes in the baryon density affect the spectrum in characteristic ways, modifying the relative heights of the even and odd acoustic peaks, due to the effect of baryons on the depth of first and subsequent acoustic (de)compressions. Despite comprising less than a sixth of the total matter content, the baryon effects on the power spectra are sufficiently distinctive that the baryon-density parameter is measured at sub-percent level accuracy with *Planck*: Planck2018CosmolParam/

$$Planck2018CosmolParam/\Omega_{\rm b}h^2 = 0.02237 \pm 0.00015 \quad \begin{array}{l}(68\,\%,\ Planck\\ \text{TT,TE,EE+lowE}\\ \text{+lensing}).\end{array} \tag{1218}$$

There is a partial degeneracy with $n_{\rm s}$, which can also affect the relative heights of the first few peaks. This is most evident in TE, but is reduced in TT because of the larger range of scales that are measured by *Planck* with low noise.

27.3.6 Changes in the base-ΛCDM parameters between the 2015 and 2018 data releases

Figure 173 compares the parameters of the base-ΛCDM model measured from the final data release with those reported in PCP15. To differentiate between changes caused by the new lowE polarization likelihood, and therefore generated by the change in the measured optical depth to reionization, we also show the result of using the 2015 likelihoods in combination with the 2018 lowE polarization likelihood at low multipoles. Figure 173 includes the results for both *Planck* TT+lowE and *Planck* TT,TE,EE+lowE.[146]

The main differences in ΛCDM parameters between the 2015 and the 2018 releases are caused by the following effects.

• New polarization low-ℓ likelihood. The use of the new HFI low-ℓ polarization likelihood in place of the 2015 LFI likelihood is the largest cause of shifts between the 2015 and 2018 parameters. The lowering and tightening of the constraint on τ is responsible for a $1\,\sigma$ decrease of $\ln(10^{10}A_{\rm s})$ through the $A_{\rm s}e^{-2\tau}$ degeneracy. This in turn decreases the smoothing due to gravitational lensing at high multipoles, which is compensated by an increase of about $1\,\sigma$ in $\omega_{\rm c}$. This decreases the amplitude of the first acoustic peak, so $n_{\rm s}$ shifts to a lower value by about $0.5\,\sigma$ to restore power. Further adjustments are then achieved by the changes of θ_* and $\omega_{\rm b}$ by about $0.5\,\sigma$.

[146] The published 2015 parameter constraints and chains had a small error in the priors for the polarization Galactic foregrounds, which was subsequently corrected in the published likelihoods. The impact on cosmological parameters was very small. Here we compare with the uncorrected 2015 chains, not the published 2015 likelihood.

• Polarization corrections in the high-ℓ likelihood. As described in detail in Sect. 27.2.2, the largest changes from 2015 are caused by corrections applied to the polarization spectra. To isolate the causes of shifts introduced by changes in the high-ℓ likelihood, Fig. 174 compares 2018 results neglecting corrections to the polarization spectra with results from the 2015 high-ℓ likelihood combined with the 2018 lowE likelihood (so that both sets of results are based on similar constraints on τ). The shift towards larger values in $\omega_{\rm b}$ by around $1\,\sigma$ is mainly caused by the beam-leakage correction in the TE high-ℓ likelihood, which is also responsible for an increase of approximately $0.5\,\sigma$ in $n_{\rm s}$, compensating for the shift in $n_{\rm s}$ as a result of the change in τ since 2015. The beam-leakage correction also changes $\omega_{\rm c}$ (by $-0.7\,\sigma$) and $\theta_{\rm MC}$ ($+0.7\,\sigma$). The other corrections implemented in 2018 have a smaller impact on the ΛCDM parameters, as described in detail in Planck Collaboration V [293].

Figure 175 presents the differences between the coadded spectra from 2018 and 2015. This plot shows the stability of the TT spectra, while also demonstrating that the main differences in polarization between the 2015 and 2018 releases are caused by the 2018 corrections for polarization efficiencies and beam leakage.

27.4 Comparison with high-resolution experiments

As discussed in PCP13 and PCP15, *Planck* TT spectra are statistically much more powerful than temperature data from current high-resolution experiments such as the Atacama Cosmology Telescope [ACT, e.g., 92] and the South Pole Telescope [SPT, e.g., 350, 146]. As a result, the *Planck* temperature data dominate if they are combined with ACT and SPT data. In PCP15, the high-resolution temperature data were used only to constrain low-amplitude components of the foreground model, which are otherwise weakly constrained by *Planck* data alone (with very little impact on cosmological parameters). We adopt the same approach in this paper.

Since the publication of PCP15, [181] have performed a direct map-based comparison of the SPT temperature data at 150 GHz with the *Planck* 143-GHz maps over the same area of sky (covering 2540 deg^2), finding no evidence for any systematic error in either data set after accounting for an overall difference in calibration. Temperature power spectrum comparisons between *Planck* and SPT are reported in a companion paper by Aylor et al. [32]. They find cosmological parameters for base ΛCDM derived from *Planck* and SPT over the same patch of sky and multipole range to be in excellent agreement. In particular, by comparing parameters determined over the multipole range 650–2000 from both experiments, the reduction in sample variance allows a test that is sensitive to systematic errors that could cause shifts in parameter posteriors comparable to the widths of the PCP15 posteriors. The parameters determined over the SPT sky area differ slightly, but not significantly, from the best-fit ΛCDM parameters reported in PCP15 based on a much larger area of sky. [32] also find a tendency for the base-ΛCDM parameters derived from SPT to shift as the multipole range is increased, but at low statistical significance.

Polarization measurements have become a major focus for ground-based CMB experiments. High resolution TE and EE spectra have been measured by the ACT Polarimeter (ACTPol) and the polarization-sensitive receiver of SPT (SPTpol). Following two seasons of observations, ACTPol has covered 548 deg^2 along the celestial equator at 149 GHz with data and analysis presented in Naess et al. [244] and Louis et al. [226]. The ACTPol spectra span the multipole range $350 < \ell < 9000$. SPTpol polarization spectra from 100 deg^2 in the southern hemisphere at 150 GHz were first reported in Crites et al. [87] and recently extended to 500 deg^2 [169]. The SPTpol spectra span the multipole range $50 < \ell < 8000$. In contrast, the *Planck* TE and EE power spectra lose statistical power at multipoles $\gtrsim 1500$. The ACTPol and SPTpol spectra are compared with the *Planck* TE and EE spectra in Fig. 176. The polarization spectra measured from these three very different experiments are in excellent agreement.

For the base-ΛCDM cosmology, the cosmological parameters should have converged close to their true values by multipoles ~ 2000. Since ACTPol and SPTpol cover a much smaller sky area than *Planck* the errors on their TE and EE spectra are larger than those of *Planck* at low multipoles (see Fig. 176). As a consequence, the current ACTPol and SPTpol polarization constraints on the parameters of the base-ΛCDM cosmology are much weaker than those derived from *Planck*. The ACTPol results [226] are consistent with the *Planck* base-ΛCDM parameters and showed a small improvement in constraints on extensions to the base cosmology that affect the damping tail. Similar results were found by SPTpol, though Henning et al. [169] noted a

Table 26: Minimum χ^2 values fitting the SPTpol spectra to the best-fit *Planck* and SPTpol ΛCDM cosmologies (as described in the text). N_b gives the number of band powers in each spectrum. The deviation of χ^2_min from the expectation $\langle\chi^2_\mathrm{min}\rangle = N_\mathrm{dof}$ is given by the columns labelled N_σ, where $N_\sigma = (\chi^2_\mathrm{min} - N_\mathrm{dof})/\sqrt{2N_\mathrm{dof}}$, and $N_\mathrm{dof} = N_\mathrm{b} - 8$. The last two columns give χ^2_p for parameter differences (Eq. 1219) and the associated PTEs.

		Planck cosmology		SPT cosmology			
SPTpol spectrum	N_b	χ^2_min	N_σ	χ^2_min	N_σ	χ^2_p	PTE
$TE + EE$	112	146.1	2.91	137.4	2.31	9.85	0.08
TE	56	71.4	2.38	70.3	2.27	3.38	0.64
EE	56	67.3	1.96	61.4	1.37	8.21	0.15

$\gtrsim 2\,\sigma$ tension with the base-ΛCDM model and found a trend for the parameters of the base-ΛCDM model to drift away from the *Planck* solution as the SPTpol likelihood is extended to higher multipoles. To assess these results we have performed some tests of the consistency of the latest *Planck* results and the SPTpol spectra.

As a reference model for SPTpol we adopt the base-ΛCDM parameters for the combined TE+EE fit to the SPTpol data from table 5 of [169]. It is worth noting that the best-fit SPTpol cosmology is strongly excluded by the *Planck* TT spectra and by the *Planck* $TE+EE$ spectra. We use the *Planck* TT,TE,EE+lowE+lensing base-ΛCDM best-fit cosmology (as plotted in Fig. 176) as a reference model for *Planck*. For each model, we ran the public version of the SPTpol likelihood code,[147] sampling the nuisance parameters using the same priors as in [169]. The best-fit values of χ^2 are listed in Table 26. As in [169], in assigning significance levels to these values, we take the number of degrees of freedom to be equal to the number of band powers minus eight, corresponding to five cosmological parameters (ω_b, ω_c, θ_MC, n_s, $A_\mathrm{s}e^{-2\tau}$) and three nuisance parameters with flat priors.

As found by [169], the SPTpol TE spectrum gives nearly identical values of χ^2 for both the SPTpol and *Planck* cosmologies and so does not differentiate between them; however, the χ^2 values are high, at the $2.3\,\sigma$ level. The SPTpol EE spectrum provides weaker constraints on cosmological parameters than the TE spectrum and is clearly better fit by the SPTpol cosmology. If the SPTpol covariance matrix is accurate, the combined TE+EE SPTpol data disfavour the *Planck* ΛCDM cosmology quite strongly and disfavour any 6-parameter ΛCDM cosmology. For ΛCDM models, outliers distributed over a wide range of multipoles contribute to the high χ^2 values, notably at $\ell = 124$, 324, 1874, 2449, and 3249 in TE, and $\ell = 1974$ and 6499 in EE.

We can assess consistency of the parameter differences, $\mathbf{\Delta p}$, between the two experiments by computing,

$$\chi^2_p = \mathbf{\Delta p}^\mathsf{T} \mathrm{C}_p^{-1} \mathbf{\Delta p}, \tag{1219}$$

where C_p is the covariance matrix for SPTpol parameters (we neglect the errors in the *Planck* parameters, which are much smaller). Values for χ^2_p are given in Table 26 together with probabilities to exceed (PTEs) computed from a χ^2 distribution with five degrees of freedom. We find no evidence for any statistically significant inconsistency between the two sets of parameters, even for the combined $TE + EE$ SPTpol likelihood. We also note that the parameter $A_\mathrm{s}e^{-2\tau}$ makes quite a large contribution to χ^2_p for the $TE + EE$ and EE spectra, but is sensitive to possible systematic errors in the SPTpol polarization efficiency calibration [169, which, as discussed, is not well understood]. Varying the maximum multipole used in the SPTpol likelihood (ℓ_max), we find that the parameters of the SPTpol TE+EE cosmology converge by $\ell_\mathrm{max} = 2500$; higher multipoles do not contribute significantly to the SPTpol base-ΛCDM solution.

[169] reported a trend for the parameters of the base-ΛCDM cosmology to change as the SPTpol likelihood is extended to higher multipoles, which they suggested may be an indication

[147] Downloaded from http://pole.uchicago.edu/public/data/henning17/ . Note that we discovered errors in the way that the covariances matrices were loaded for separate TE and EE analyses, which have been corrected in the analysis presented here.

of new physics. However, this effect is not of high statistical significance and cannot be tested by the *Planck* spectra, which become less sensitive than the SPTpol spectra at multipoles $\gtrsim 1500$. The consistency of the base-ΛCDM cosmology at high multipoles in polarization should become clearer in the near future as more polarization data are accumulated by ACTPol and SPTpol.

27.5 Comparison with other astrophysical data sets

27.5.1 Baryon acoustic oscillations

As in PCP13 and PCP15 baryon acoustic oscillation (BAO) measurements from galaxy redshift surveys are used as the primary non-CMB astrophysical data set in this paper. The acoustic scale measured by BAOs, at around 147 Mpc, is much larger than the scale of virialized structures. This separation of scales makes BAO measurements insensitive to nonlinear physics, providing a robust geometrical test of cosmology. It is for this reason that BAO measurements are given high weight compared to other non-CMB data in this and in previous *Planck* papers. BAO features in the galaxy power spectrum were first detected by [80] and [119]. Since their discovery, BAO measurements have improved in accuracy via a number of ambitious galaxy surveys. As demonstrated in PCP13 and PCP15 BAO results from galaxy surveys have been consistently in excellent agreement with the best-fit base-ΛCDM cosmology inferred from *Planck*. More recently, the redshift reach of BAO measurements has been increased using quasar redshift surveys and Lyman-α absorption lines detected in quasar spectra.

Figure 177 summarizes the latest BAO results, updating figure 14 of PCP15. This plot shows the acoustic-scale distance ratio $D_{\rm V}(z)/r_{\rm drag}$ measured from surveys with effective redshift z, divided by the mean acoustic-scale ratio in the base-ΛCDM cosmology using *Planck* TT,TE,EE +lowE+lensing. Here $r_{\rm drag}$ is the comoving sound horizon at the end of the baryon drag epoch and $D_{\rm V}$ is a combination of the comoving angular diameter distance $D_{\rm M}(z)$ and Hubble parameter $H(z)$:

$$D_{\rm V}(z) = \left[D_{\rm M}^2(z)\frac{cz}{H(z)}\right]^{1/3}. \tag{1220}$$

The grey bands in the figure show the $\pm1\,\sigma$ and $\pm2\,\sigma$ ranges allowed by *Planck* in the base-ΛCDM cosmology.

Compared to figure 14 of PCP15, we have replaced the Baryon Oscillation Spectroscopic Survey (BOSS) LOWZ and CMASS results of Anderson et al. [23] with the latest BOSS data release 12 (DR12) results summarized by [15]. That paper reports "consensus" results on BAOs (weighting together different BAO analyses of BOSS DR12) reported by [314], [359], and [50] in three redshift slices with effective redshifts $z_{\rm eff} = 0.38$, 0.51, and 0.61. These new measurements, shown by the red triangles in Fig. 177, are in good agreement with the *Planck* base-ΛCDM cosmology.

By using quasars, it has become possible to extend BAO measurements to redshifts greater than unity. [28] have measured the BAO scale D_V at an effective redshift of $z_{\rm eff} = 1.52$ using a sample of quasars from the extended Baryon Oscillation Survey (eBOSS). This measurement is shown by the red circle in Fig. 177 and is also in very good agreement with *Planck*. The results of the Ata et al. [28] analysis also agree well with other analyses of the eBOSS quasar sample [e.g., 147].

At even higher redshifts BAOs have been measured in the Lyman α spectra of quasars [102, 136, 37, 114, 96, 57]. In the first preprint version of this paper, we compared the *Planck* results with those from BAO features measured from the flux-transmission correlations of Sloan Digital Sky Survey (SDSS) DR12 quasars [37] and with the cross-correlation of the Ly α forest with SDSS quasars [114]. The combined result on $D_{\rm M}/r_{\rm drag}$ from these analyses was about $2.3\,\sigma$ lower than expected from the best-fit *Planck* base-ΛCDM cosmology. The [37] and du Mas des Bourboux et al. [114] analyses have been superseded by equivalent studies of a larger sample of SDSS DR14 quasars reported in [96] and [57]. The combined result for $D_{\rm M}/r_{\rm drag}$ from these analyses [as quoted by 57] is plotted as the orange hexagon on Fig. 177 and lies within $1.7\,\sigma$ of the *Planck* best-fit model. The errors on these high-redshift BAO measurements are still quite large in comparison with the galaxy measurements and so we do not include them in our default BAO compilation.[148]

[148]The first preprint version of this paper showed that the inclusion of the [37] and du Mas des Bourboux

The more recent BAO analyses solve for the positions of the BAO feature in both the line-of-sight and transverse directions (the distortion in the transverse direction caused by the background cosmology is sometimes called the Alcock-Paczynski effect, Alcock & Paczynski 16), leading to joint constraints on the angular diameter distance $D_{\rm M}(z_{\rm eff})$ and the Hubble parameter $H(z_{\rm eff})$. These constraints for the BOSS DR12 analysis are plotted in Fig. 178. Samples from the *Planck* TT+lowE and *Planck* TT,TE,EE+lowE+lensing likelihood are shown in green and red, respectively, demonstrating that BAO and *Planck* polarization data with lensing consistently pull parameters in the same direction (towards slightly lower $\Omega_{\rm c}h^2$). We find the same behaviour for *Planck* when adding polarization and lensing to the *TT* likelihood separately. This demonstrates the remarkable consistency of the *Planck* data, including polarization and CMB lensing with the galaxy BAO measurements. Evidently, the *Planck* base-ΛCDM parameters are in good agreement with both the isotropized $D_{\rm V}$ BAO measurements plotted in Fig. 177, and with the anisotropic constraints plotted in Fig. 178.

In this paper, we use the 6dFGS and SDSS-MGS measurements of $D_{\rm V}/r_{\rm drag}$ [47, 313] and the final DR12 anisotropic BAO measurements of Alam et al. [15]. Since the WiggleZ volume partially overlaps that of the BOSS-CMASS sample, and the correlations have not been quantified, we do not use the WiggleZ results in this paper. It is clear from Fig. 177 that the combined BAO likelihood for the lower redshift points is dominated by the BOSS measurements.

In the base-ΛCDM model, the *Planck* data constrain the Hubble constant H_0 and matter density $\Omega_{\rm m}$ to high precision:

$$\left.\begin{aligned} Planck2018CosmolParam/H_0 &= (67.36 \pm 0.54)\ \text{km s}^{-1}\text{Mpc}^{-1}, \\ \Omega_{\rm m} &= 0.3158 \pm 0.0073, \end{aligned}\right\}\ \begin{array}{l}68\,\%,\ \text{TT,TE,EE} \\ \text{+lowE+lensing.}\end{array} \tag{1221}$$

With the addition of the BAO measurements, these constraints are strengthened to

$$\left.\begin{aligned} Planck2018CosmolParam/H_0 &= (67.66 \pm 0.42)\ \text{km s}^{-1}\text{Mpc}^{-1}, \\ \Omega_{\rm m} &= 0.3111 \pm 0.0056, \end{aligned}\right\}\ \begin{array}{l}68\,\%,\ \text{TT,TE,EE} \\ \text{+lowE+lensing} \\ \text{+BAO.}\end{array} \tag{1222}$$

These numbers are in very good agreement with the constraints given in Eq. (1200), which exclude the high-multipole *Planck* likelihood. Section 27.5.4 discusses the consistency of direct measurements of H_0 with these estimates and Hubble parameter measurements from the line-of-sight component of BAOs at higher redshift.

As discussed above, we have excluded Ly α BAOs from our default BAO compilation. The full likelihood for the combined Ly α and Ly α-quasar cross-correlations reported in du Mas des Bourboux et al. [114] is not yet available; nevertheless, we can get an indication of the impact of including these measurements by assuming uncorrelated Gaussian errors on $D_{\rm M}/r_{\rm drag}$ and $r_{\rm drag}H$. Adding these measurements to *Planck* TT,TE,EE+lowE and our default BAO compilation shifts H_0 higher, and $\Omega_{\rm m}h^2$ and σ_8 lower, by approximately $0.3\,\sigma$. The joint *Planck*+BAO result then gives $D_{\rm M}/r_{\rm drag}$ and $r_{\rm drag}H$ at $z = 2.4$ lower by 0.25 and 0.3 of *Planck*'s σ, leaving the overall $2.3\,\sigma$ tension with these results almost unchanged. As shown by Aubourg et al. [29], it is difficult to construct well-motivated extensions to the base-ΛCDM model that can resolve the tension with the Ly α BAOs. Further work is needed to assess whether the discrepancy between *Planck* and the Ly α BAO results is a statistical fluctuation, is caused by small systematic errors, or is a signature of new physics.

27.5.2 Type Ia supernovae

The use of type Ia supernovae (SNe) as standard candles has been of critical importance to cosmology, leading to the discovery of cosmic acceleration [307, 266]. For ΛCDM models,

et al. [114] Ly α BAO results had a minor impact on the parameters of the base-ΛCDM cosmology. The impact of the more recent Ly α results of [96] and [57] will be even lower, since they are in closer agreement with the *Planck* best-fit cosmology.

however, SNe data have little statistical power compared to *Planck* and BAO and in this paper they are used mainly to test models involving evolving dark energy and modified gravity. For these extensions of the base cosmology, SNe data are useful in fixing the background cosmology at low redshifts, where there is not enough volume to allow high precision constraints from BAO.

In PCP15 we used the "Joint Light-curve Analysis" (JLA) sample constructed from the SNLS and SDSS SNe plus several samples of low redshift SNe described in Betoule et al. [46, 45] and Mosher et al. [241]. In this paper, we use the new "Pantheon" sample of Scolnic et al. [329], which adds 276 supernovae from the Pan-STARRS1 Medium Deep Survey at $0.03 < z < 0.65$ and various low-redshift and HST samples to give a total of 1048 supernovae spanning the redshift range $0.01 < z < 2.3$. The Pantheon compilation applies cross-calibrations of the photometric systems of all of the sub-samples used to construct the final catalogue [328], reducing the impact of calibration systematics on cosmology[149]. The Pantheon data are compared to the predictions of the *Planck* TT,TE,EE+lowE+lensing base-ΛCDM model best fit in Fig. 179. The agreement is excellent. The JLA and Pantheon samples are consistent with each other (with Pantheon providing tighter constraints on cosmological parameters) and there would be no significant change to our science conclusions had we chosen to use the JLA sample in this paper. To illustrate this point we give results for a selection of models using both samples in the parameter tables available in the PLA; Fig. 183, illustrating inverse-distance-ladder constraints on H_0 (see Sect. 27.5.4), shows a specific example.

27.5.3 Redshift-space distortions

The clustering of galaxies observed in a redshift survey exhibits anisotropies induced by peculiar motions (known as redshift-space distortions, RSDs). Measurement of RSDs can provide constraints on the growth rate of structure and the amplitude of the matter power spectrum [e.g., 263]. Since it uses non-relativistic tracers, RSDs are sensitive to the time-time component of the metric perturbation or the Newtonian potential. A comparison of the amplitude inferred from RSDs with that inferred from lensing (sensitive to the Weyl potential, see Sect. **??**). provides a test of General Relativity.

Measurements of RSDs are usually quoted as constraints on $f\,\sigma_8$, where for models with scale-independent growth $f = d\ln D/d\ln a$. For ΛCDM, $d\ln D/d\ln a \approx \Omega_{\rm m}^{0.55}(z)$. We follow PCP15, defining

$$f\,\sigma_8 \equiv \frac{\left[\sigma_8^{(\rm vd)}(z)\right]^2}{\sigma_8^{(\rm dd)}(z)}, \tag{1223}$$

where $\sigma_8^{(\rm vd)}$ is the density-velocity correlation in spheres of radius $8\,h^{-1}$Mpc in linear theory.

Measuring $f\,\sigma_8$ requires modelling nonlinearities and scale-dependent bias and is considerably more complicated than estimating the BAO scale from galaxy surveys. One key problem is deciding on the precise range of scales that can be used in an RSD analysis, since there is a need to balance potential systematic errors associated with modelling nonlinearities against reducing statistical errors by extending to smaller scales. In addition, there is a partial degeneracy between distortions caused by peculiar motions and the Alcock-Paczynski effect. Nevertheless, there have been substantial improvements in modelling RSDs in the last few years, including extensive tests of systematic errors using numerical simulations. Different techniques for measuring $f\,\sigma_8$ are now consistent to within a few percent [15].

Figure 180, showing $f\,\sigma_8$ as a function of redshift, is an update of figure 16 from PCP15. The most significant changes from PCP15 are the new high precision measurements from BOSS DR12, shown as the red points. These points are the "consensus" BOSS D12 results from Alam et al. [15], which averages the results from four different ways of analysing the DR12 data [49, 153, 323, 324]. These results are in excellent agreement with the *Planck* base ΛCDM cosmology (see also Fig. 181) and provide the tightest constraints to date on the growth rate of fluctuations. We have updated the VIPERS constraints to those of the second public data release [269] and added a data point from the Galaxy and Mass Assembly (GAMA) redshift survey [53]. Two new surveys have extended the reach of RSD measurements (albeit with large

[149]We use the November 2018 data file available from https://github.com/dscolnic/Pantheon/, which includes heliocentric redshifts and no bulk-flow corrections for $z > 0.08$.

errors) to redshifts greater than unity: the deep FASTSOUND emission line redshift survey [250]; and the BOSS DR14 quasar survey [373]. We have also added a new low redshift estimate of $f\,\sigma_8$ from [187] at an effective redshift of $z_{\rm eff} = 0.023$, which is based on correlating deviations from the mean magnitude-redshift relation of SNe in the Pantheon sample with estimates of the nearby peculiar velocity field determined from the 6dF Galaxy Survey [348]. As can be seen from Fig. 180, these growth rate measurements are consistent with the *Planck* base-ΛCDM cosmology over the entire redshift range $0.023 < z_{\rm eff} < 1.52$.

Since the BOSS-DR12 estimates provide the strongest constraints on RSDs, it is worth comparing these results with *Planck* in greater detail. Here we use the "full-shape consensus" results[150] on D_V, $f\,\sigma_8$, and $F_{\rm AP}$ for each of the three redshift bins from [15] and the associated 9×9 covariance matrix, where $F_{\rm AP}$ is the Alcock-Paczinski parameter,

$$F_{\rm AP}(z) = D_{\rm M}(z)\frac{H(z)}{c}. \tag{1224}$$

Figure 181 shows the constraints from BOSS-DR12 on $f\,\sigma_8$ and $F_{\rm AP}$ marginalized over D_V. *Planck* base-ΛCDM constraints are shown by the red and green contours. For each redshift bin, the *Planck* best-fit values of $f\,\sigma_8$ and $F_{\rm AP}$ lie within the 68 % contours from BOSS-DR12. Figure 181 highlights the impressive consistency of the base-ΛCDM cosmology from the high redshifts probed by the CMB to the low redshifts sampled by BOSS.

27.5.4 The Hubble constant

Perhaps the most controversial tension between the *Planck* ΛCDM model and astrophysical data is the discrepancy with traditional distance-ladder measurements of the Hubble constant H_0. PCP13 reported a value of $H_0 = (67.3 \pm 1.2)$ km s^{-1}Mpc^{-1} for the base-ΛCDM cosmology, substantially lower that the distance-ladder estimate of $H_0 = (73.8 \pm 2.4)$ km s^{-1}Mpc^{-1} from the SH0ES[151] project [308] and other H_0 studies [e.g., 137, 140]. Since then, additional data acquired as part of the SH0ES project [309, 310, hereafter R18] has exacerbated the tension. R18 conclude that $H_0 = (73.48 \pm 1.66)$ km s^{-1}Mpc^{-1}, compared to our *Planck* TT,TE,EE +lowE+lensing estimate from Table 1 of $H_0 = (67.27 \pm 0.60)$ km s^{-1}Mpc^{-1}. Using Gaia parallaxes Riess et al. [311] slightly tightened their measurement to $H_0 = (73.52 \pm 1.62)$ km s^{-1}Mpc^{-1}. Recently Riess et al. [306] then used improved measurements of LMC Cepheids to further tighten[152] the constraint to $H_0 = (74.03 \pm 1.42)$ km s^{-1}Mpc^{-1}. Interestingly, the central values of the SH0ES and *Planck* estimates have hardly changed since the appearance of PCP13, but the errors on both estimates have shrunk so that the discrepancy has grown from around $2.5\,\sigma$ in 2013 to $3.5\,\sigma$ today [$4.4\,\sigma$ using 306]. This discrepancy has stimulated a number of investigations of possible systematic errors in the either the *Planck* or SH0ES data, which have failed to identify any obvious problem with either analysis [e.g., 347, 6, 299, 115, 69, 375, 135]. It has also been argued that the Gaussian likelihood assumption used in the SH0ES analysis leads to an overestimate of the statistical significance of the discrepancy [127].

Recently, [138] have reported a determination of H_0 using the tip of the red giant branch as a distance estimator. This analysis gives $H_0 = (69.8 \pm 1.9)$ km s^{-1}Mpc^{-1}, i.e., intermediate between the SH0ES measurement and the *Planck* base-ΛCDM value. There has been some controversy (see Yuan et al. [369]) concerning the calibration of the tip of the red giant branch adopted in Freedman et al. [138], though a recent reanalysis by Freedman et al. [139] yields a value of H_0 that is almost identical to that reported in Freedman et al. [138].

Measurements of the Hubble constant using strong gravitational-lensing time delays are also higher than the *Planck* base-ΛCDM value. The most recent results, based on six strongly lensed quasars, give $H_0 = 73.3^{+1.7}_{-1.8}$ km s^{-1}Mpc^{-1} [366], which is about $3.2\,\sigma$ higher than the *Planck* value. A number of other techniques have been used to infer H_0, including stellar

[150] When using RSDs to constraint dark energy in Sect. **??**, we use the alternative $D_{\rm M}$, H, and $f\,\sigma_8$ parameterization from [15] for consistency with the DR12 BAO-only likelihood that we use elsewhere.

[151] SN, H0, for the Equation of State of dark energy.

[152] By default in this paper (and in the PLA) we use the Riess et al. [310] number (available at the time we ran our parameter chains) unless otherwise stated; using the updated number would make no significant difference to our conclusions.

ages [e.g., 193, 151], distant megamasers [305, 208, 144] and gravitational-wave standard sirens [3]. These measurements span a range of values. Nevertheless, there is a tendency for local determinations to sit high compared to the *Planck* base-ΛCDM value, with the SH0ES Cepheid-based measurement giving the most statistically significant discrepancy.

In this paper, we take the R18 estimate at face value and include it as a prior in combination with *Planck* in some of the parameter tables available on the PLA. The interested reader can then assess the impact of the R18 measurement on a wide range of extensions to the base-ΛCDM cosmology.

We already mentioned in Sect. 27.5.1 that BAO measurements along the line of sight constrain $H(z)r_{\rm drag}$. *Planck* constrains $r_{\rm drag}$ to a precision of 0.2 % for the base-ΛCDM model and so the BAO measurements can be accurately converted into absolute measurements of $H(z)$. This is illustrated by Fig. 182, which shows clearly how well the *Planck* base-ΛCDM cosmology fits the BAO measurements of $H(z)$ over the redshift range 0.3–2.5, yet fails to match the R18 measurement of H_0 at $z = 0$. The model is also consistent with the most recent Ly α BAO measurements at $z \approx 2.3$.

PCP13 and PCP15 emphasized that this mismatch between BAO measurements and forward distance-ladder measurements of H_0 is not sensitive to the *Planck* data at high multipoles. For example, combining WMAP with BAO measurements leads to $H_0 = (68.14 \pm 0.73)$ km s^{-1}Mpc^{-1} for the base-ΛCDM cosmology, which is discrepant with the R18 value at the 2.9 σ level.

Heavens et al. [164], Cuesta et al. [88], and Aubourg et al. [29] showed that the combination of CMB, BAO, and SNe data provides a powerful "inverse-distance-ladder" approach to constructing a physically calibrated distance-redshift relation down to very low redshift. For the base-ΛCDM model, this inverse-distance-ladder approach can be used to constrain H_0 without using any CMB measurements at all, or by only using constraints on the CMB parameter $\theta_{\rm MC}$ [see also 44, 7, 105, 213]. This is illustrated in Fig. 183, which shows how the constraints on H_0 and $\Omega_{\rm m}$ converge to the *Planck* values as more data are included. The green contours show the constraints from BAO and the Pantheon SNe data, together with a BBN constraint on the baryon density ($\Omega_{\rm b}h^2 = 0.0222 \pm 0.0005$) based on the primordial deuterium abundance measurements of Cooke et al. [82, see Sect. **??**]. The dashed contours in this figure show how the green contours shift if the Pantheon SNe data are replaced by the JLA SNe sample. Adding *Planck* CMB lensing (grey contours) constrains $\Omega_{\rm m}h^2$ and shifts H_0 further away from the R18 measurement. Using a "conservative" *Planck* prior of $100\theta_{\rm MC} = 1.0409 \pm 0.0006$ (which is consistent with all of the variants of ΛCDM considered in this paper to within 1 σ, see Table **??**) gives the red contours, with $H_0 = (67.9 \pm 0.8)$ km s^{-1}Mpc^{-1} and $\Omega_{\rm m} = 0.305 \pm 0.001$, very close to the result using the full *Planck* likelihood (blue contours). Evidently, there is a significant problem in matching the base-ΛCDM model to the R18 results and this tension is not confined exclusively to the *Planck* results.

The question then arises of whether there is a plausible extension to the base-ΛCDM model that can resolve the discrepancy. Table **??** summarizes the *Planck* constraints on H_0 for variants of ΛCDM considered in this paper. H_0 remains discrepant with R18 in all of these cases, with the exception of models in which we allow the dark energy equation of state to vary. For models with either a fixed dark energy equation-of-state parameter, w_0, or time-varying equation of state parameterized by w_0 and w_a (see Sect. **??** for definitions and further details), *Planck* data alone lead to poor constraints on H_0. However, for most physical dark energy models where $p_{\rm de} \geq -\rho_{\rm de}$ (so $w_0 > -1$), and the density is only important after recombination, H_0 can only decrease with respect to ΛCDM if the measured CMB acoustic scale is maintained, making the discrepancy with R18 worse. If we allow for $w_0 < -1$, then adding BAO and SNe data is critical to obtain a useful constraint [as pointed out by 29], and we find

$$H_0 = (68.34 \pm 0.81) \text{ km s}^{-1}\text{Mpc}^{-1}, \quad (w_0 \text{ varying}), \tag{1225a}$$

$$H_0 = (68.31 \pm 0.82) \text{ km s}^{-1}\text{Mpc}^{-1}, \quad (w_0, w_a \text{ varying}), \tag{1225b}$$

for the parameter combination *Planck* TT,TE,EE+lowE+lensing+BAO+Pantheon. Modifying the dark energy sector in the late universe does not resolve the discrepancy with R18.

If the difference between base ΛCDM and the R18 measurement of H_0 is caused by new physics, then it is unlikely to be through some change to the late-time distance-redshift relationship. Another possibility is a change in the sound horizon scale. If we use the R18 measurement of H_0, combined with Pantheon supernovae and BAO, the acoustic scale is $r_{\rm drag} = (136.4 \pm 3.5)$ Mpc. The difficulty is to find a model that can give this much smaller

value of the sound horizon (compared to $r_{\rm drag} = (147.05 \pm 0.3)$ Mpc from *Planck* TT,TE,EE +lowE in ΛCDM), while preserving a good fit to the CMB power spectra and a baryon density consistent with BBN. We discuss some extensions to ΛCDM in Sect. **??** that allow larger H_0 values (e.g., $N_{\rm eff} > 3.046$); however, these models are not preferred by the *Planck* data, and tend to introduce other tensions, such as a higher value of σ_8.[153]

The tension between base ΛCDM and the SH0ES H_0 measurement is intriguing and emphasizes the need for independent measurements of the distance scale. It will be interesting in the future to compare the Cepheid distance scale in more detail with other distance indicators, such as the tip of the red giant branch [138], and with completely different techniques such as gravitational-lensing time delays [352] and gravitational-wave standard sirens [180, 3, 74, 128].

27.5.5 Weak gravitational lensing of galaxies

The distortion of the shapes of distant galaxies by lensing due to large-scale structure along the line of sight is known as galaxy lensing or cosmic shear [see e.g., 35, for a review]. It constrains the gravitational potentials at lower redshift than CMB lensing, with tomographic information and completely different systematics, so the measurements are complementary. Since the source galaxy shapes and orientations are in general unknown, the lensing signal is a small effect that can only be detected statistically. If it can be measured robustly it is a relatively clean way of measuring the Weyl potential (and hence, in GR, the total matter fluctuations); however, the bulk of the statistical power comes from scales where the signal is significantly nonlinear, complicating the cosmological interpretation. The measurement is also complicated by several other issues. Intrinsic alignment between the shape of lensed galaxies and their surrounding potentials means that the galaxy shape correlation functions actually measure a combination of lensing and intrinsic alignment effects [176]. Furthermore, to get a strong statistical detection, a large sample of galaxies is needed, so most current results use samples that rely mainly on photometric redshifts; accurate calibration of the photometric redshifts and modelling of the errors are required in order to use the observed lensing signal for cosmology.

Cosmic shear measurements are available from several collaborations, including CFHTLenS [171, 121, which we discussed in PCP15], `DLS` [192], and more recently the Dark Energy Survey [DES, 106], Hyper Suprime-Cam [HSC, 172, 158], and KiDS [174, 200, 175]. The CFHTLenS and KiDS results found a modest tension with the *Planck* ΛCDM cosmology, preferring lower values of $\Omega_{\rm m}$ or σ_8. A combined analysis of KiDS with GAMA [358] galaxy clustering has found results consistent with *Planck*, whereas a similar analysis combining KiDS lensing measurements with spectroscopic data from the 2-degree Field Lensing Survey and BOSS claims a $2.6\,\sigma$ discrepancy with *Planck* [194]. [356] have shown that a more accurate treatment of the intrinsic galaxy shape noise, multiplicative shear calibration uncertainty, and angular scale of each bin can significantly change earlier KiDS results (by about $1\,\sigma$), making them more consistent with *Planck*. At the time of running our chains the DES lensing results had been published and included this improved modelling, while an updated analysis from KiDS was not yet available; we therefore only consider the DES results in detail here. [356] reports consistent results from DES and their new analysis of KiDS, and HSC also report results consistent with *Planck*. However, the more recent KiDS analysis by [175] still finds a $2.3\,\sigma$ discrepancy with *Planck*, and [195] claim that a recalibration of the DES redshifts gives results compatible with KiDS and a combined $2.5\,\sigma$ tension with *Planck*.

The DES collaboration analysed 1321 deg^2 of imaging data from the first year of DES. They analysed the cosmic shear correlation functions of 26 million source galaxies in four redshift bins [355], and also considered the auto- [120] and cross-spectrum [302] of 650 000 lens galaxies in five redshift bins. To be conservative they restricted their parameter analysis to scales that are only weakly affected by nonlinear modelling (at the expense of substantially reducing the statistical power of the data). To account for modelling uncertainties, the cosmic shear analysis marginalizes over 10 nuisance parameters, describing uncertainties in the photometric redshift

[153] To obtain simultaneously higher values of H_0, lower values of σ_8, and consistent values of $\Omega_{\rm m}$ it is necessary to invoke less common extensions of the ΛCDM model, such as models featuring non-standard interactions in the neutrino, dark-matter, dark-radiation, and/or dark-energy sector [see e.g., 267, 215, 285, 24, 209, 251, 108, 65, 301, 206, 11, 222, 25] . Such models are likely to be highly constrained by the *Planck*, BAO, and supernova data used in this paper and by future CMB observations and surveys of large-scale structure.

distributions, shear calibrations, and intrinsic alignments; the joint analysis adds an additional 10 nuisance parameters describing the bias and redshift uncertainty of the lens galaxies.

We use the first-year DES lensing (cosmic shear) likelihood, data cuts, nuisance parameters, and nuisance parameter priors, as described by Troxel et al. [355], DES Collaboration [106], Krause et al. [205]. We implement the theory model code independently, but use the same physical model and assumptions as the DES analysis,[154] treating the nuisance parameters as fast parameters for sampling in `CosmoMC`. In this section we adopt the cosmological parameter priors assumed by Troxel et al. [355], but to be consistent with our other ΛCDM analyses, we assume a single minimal-mass eigenstate rather than marginalizing over the neutrino mass, and use `HMcode` for the nonlinear corrections.[155] The shear correlation data points and parameter fits are shown in Fig. 184. Note that intrinsic alignments contribute significantly to the observed shear correlation functions (as shown by the dotted lines in the figure). This introduces additional modelling uncertainty and a possible source of bias if the intrinsic alignment model is not correct. The DES model is validated in [355, 205].

Figure 185 shows the constraints in the $\Omega_{\rm m}$–σ_8 plane from DES lensing, compared to the constraints from the CMB power spectra and CMB lensing. The DES cosmic shear constraint is of comparable statistical power to CMB lensing, but due to the significantly lower mean source redshift, the degeneracy directions are different (with DES cosmic shear approximately constraining $\Omega_{\rm m}\sigma_8^{0.5}$ and CMB lensing constraining $\Omega_{\rm m}\sigma_8^{0.25}$). The correlation between the DES cosmic shear and CMB lensing results is relatively small, since the sky area of the CMB reconstruction is much larger than that for DES, and it is also mostly not at high signal-to-noise ratio. Neglecting the cross-correlation, we combine the DES and *Planck* lensing results to break a large part of the degeneracy, giving a substantially tighter constraint than either alone. The lensing results separately, and jointly, are both consistent with the main *Planck* power-spectrum results, although preferring σ_8 and $\Omega_{\rm m}$ values at the lower end of those allowed by *Planck*. The DES joint analysis of lensing and clustering is also marginally consistent, but with posteriors preferring lower values of $\Omega_{\rm m}$ (see the next subsection). Overlap of contours in a marginalized 2D subspace does not of course guarantee consistency in the full parameter space. However, the values of the Hubble parameter in the region of $\Omega_{\rm m}$–σ_8 parameter space consistent with *Planck* $\Omega_{\rm m}$ and σ_8 are also consistent with *Planck*'s value of H_0. A joint analysis of DES with BAO and a BBN baryon-density constraint gives values of the Hubble parameter that are very consistent with the *Planck* power spectrum analysis [105].

27.5.6 Galaxy clustering and cross-correlation

The power spectrum of tracers of large-scale structure can yield a biased estimate of the matter power spectrum, which can then be used as a probe of cosmology. For adiabatic Gaussian initial perturbations the bias is expected to be constant on large scales where the tracers are out of causal contact with each other, and nearly constant on scales where nonlinear growth effects are small. Much more information is available if small scales can also be used, but this requires detailed modelling of perturbative biases out to $k \approx 0.3$–$0.6\,{\rm Mpc}^{-1}$, and fully nonlinear predictions beyond that. Any violation of scale-invariant bias on super-horizon scales would be a robust test for non-Gaussian initial perturbations protected by causality [91]. However, using the shape of the biased-tracer power spectrum on smaller scales to constrain cosmology requires at least a model of constant bias parameters for each population at each redshift, and, as precision is increased, or smaller scales probed, a model for the scale dependence of the bias. Early galaxy surveys provided cosmology constraints that were competitive with those from CMB power spectrum measurements [e.g., 264], but as precision has improved, focus has mainly moved away to using the cleaner BAO and RSD measurements and, in parallel, developing ways to get the quasi-linear theoretical predictions under better control. Most recent studies of galaxy clustering have focussed on investigating bias rather than background cosmology, with the notable exception of WiggleZ [253].

[154] Except for the modified-gravity models in Sect. **??** where we calculate the lensing spectrum directly from the power spectrum of the Weyl potential (rather than from the matter power spectrum assuming standard GR).

[155] The results are quite sensitive to the choice of cosmological parameter priors, see PL2018 for an analysis using the different priors assumed by the *Planck* CMB lensing analysis. Here we assume consistent (DES) priors for DES and CMB lensing results; however, the *Planck* power spectrum constraints are much less sensitive to priors and we use our default priors for those.

Here we focus on the first-year DES survey measurement of galaxy clustering [120] and the cross-correlation with galaxy lensing [302, "galaxy-galaxy lensing"]. By simultaneously fitting for the clustering, lensing, and cross-correlation, the bias parameters can be constrained empirically [106]. Similar analyses using KiDS lensing data combined with spectroscopic surveys have been performed by van Uitert et al. [358] and Joudaki et al. [194].

To keep the theoretical model under control (nearly in the linear regime), DES exclude all correlations on scales where modelling uncertainties in the nonlinear regime could begin to bias parameter constraints (at the price of substantially reducing the total statistical power available in the data). Assuming a constant bias parameter for each of the given source redshift bins, parameter constraints are obtained after marginalizing over the bias, as well as a photometric redshift window mid-point shift parameter to account for redshift uncertainties. Together with galaxy lensing parameters, the full joint analysis has 20 nuisance parameters. Although this is a relatively complex nuisance-parameter model, it clearly does not fully model all possible sources of error: for example, correlations between redshift bins may depend on photometric redshift uncertainties that are not well captured by a single shift in the mean of each window's population. However, Troxel et al. [355] estimate that the impact on parameters is below $0.5\,\sigma$ for all more complex models they considered. The DES theoretical model for the correlation functions (which we follow) neglects redshift-space distortions, and assumes that the bias is constant in redshift and k across each redshift bin; these may be adequate approximations for current noise levels and data cuts, but will likely need to be re-examined in the future as statistical errors improve.

Using the full combined clustering and lensing DES likelihood, for a total of 457 data points [106], the best-fit ΛCDM model has $\chi^2_{\rm eff} \approx 500$ or 513 with the *Planck* best-fit cosmology. Parameter constraints from the galaxy auto- and cross-correlation are shown in Fig. 186, together with the joint constraint with DES lensing (the comparison with DES galaxy lensing and CMB lensing alone is shown in Fig. 185).

Using the joint DES likelihood in combination with DES cosmological parameter priors gives (for our base-ΛCDM model with $\sum m_\nu = 0.06\,\text{eV}$)

$$\left.\begin{aligned} Planck2018CosmolParam/S_8 \equiv \sigma_8(\Omega_{\rm m}/0.3)^{0.5} &= 0.793 \pm 0.024, \\ \Omega_{\rm m} &= 0.256^{+0.023}_{-0.031}, \end{aligned}\right\} \quad 68\,\%,\ \text{DES}. \tag{1226}$$

Planck TT,TE,EE+lowE+lensing gives a higher value of $S_8 = 0.832 \pm 0.013$, as well as larger $\Omega_{\rm m} = 0.315 \pm 0.007$. As shown in the previous section, the DES lensing results are quite compatible with *Planck*, although peaking at lower $\Omega_{\rm m}$ and σ_8 values. The full joint DES likelihood, however, shrinks the error bars in the σ_8–$\Omega_{\rm m}$ plane so that only 95 % confidence contours overlap with *Planck* CMB data, giving a moderate (roughly 2 % PTE) tension, as shown in Fig. 186. The dotted contour in Fig. 186 shows the result using the `CamSpec` *Planck* likelihood, which gives results slightly more consistent with DES than the default `Plik` likelihood. The *Planck* result is therefore sensitive to the details of the polarization modelling at the $0.5\,\sigma$ level, and the tension cannot be quantified robustly beyond this level.

Combining DES with the baseline *Planck* likelihood pulls the *Planck* result to lower $\Omega_{\rm m}$ and slightly lower σ_8, giving

$$\left.\begin{aligned} Planck2018CosmolParam/S_8 &= 0.811 \pm 0.011, \\ \Omega_{\rm m} &= 0.3040 \pm 0.0060, \\ \sigma_8 &= 0.8062 \pm 0.0057, \end{aligned}\right\} \quad \begin{array}{l} 68\,\%,\ Planck \\ \text{TT,TE,EE+lowE} \\ \text{+lensing+DES}. \end{array} \tag{1227}$$

A similar shift is seen without including *Planck* lensing, and is disfavoured by *Planck* CMB with a total $\Delta\chi^2_{\rm eff} \approx 13$ for the CMB likelihoods (comparing the *Planck*-only best fit to the fit when combined with DES). The shift in parameters is also larger than would be expected for Gaussian distributions, given the small change in parameter covariance. The corresponding change in $\chi^2_{\rm eff}$ for the DES likelihood is $\Delta\chi^2_{\rm eff} \approx 10$, which is high, but less surprising given the 4–5 contribution expected from the number of parameters that are much better constrained by *Planck*. The summary consistency statistic $\chi^2_{\rm eff,joint} - \chi^2_{\rm eff,DES} - \chi^2_{\rm eff,Planck} \approx 14$, which is high at the roughly 1 % PTE level, given the expected value of 4, assuming roughly Gaussian statistics [304].

In summary, the DES combined probes of ΛCDM parameters are in moderate percent-level tension with *Planck*. Whether this is a statistical fluctuation, evidence for systematics, or new physics is currently unclear. In this paper, we follow the philosophy of PCP13 and PCP15 of making minimal use of other astrophysical data in combination with *Planck*, using BAO as our primary complementary data set. We therefore do not include DES results in most of the parameter constraints discussed in this paper. We do, however, consider the impact of the DES weak lensing results on dark-energy and modified-gravity constraints in Sect. **??** and on neutrino masses in Sect. **??**. We also include DES for a wider range of models in the *Planck* parameter tables available on the PLA.

27.5.7 Cluster counts

Counts of clusters of galaxies provide an additional way of constraining the amplitude of the power spectrum at low redshifts [e.g., 270, 204, and references therein]. *Planck* clusters, selected via the thermal Sunyaev-Zeldovich (tSZ) signature, were used to explore cosmological parameters in [278]. This analysis was revisited using a deeper sample of *Planck* clusters in [289]. We have not produced a new tSZ cluster catalogue in the 2018 *Planck* data release and so the results presented in this section are based on the 439 clusters in the MMF3 cluster cosmology sample, as analysed in [289]. Comparison with the 2018 CMB *Planck* power spectrum results show differences primarily from changes to the base-ΛCDM model parameters caused by the tighter constraints on τ. The impact of the lower value of τ reported in [297] on the interpretation of cluster counts has been discussed by [321].

We first review the main results from [289]. There has been increasing recognition that the calibration of cluster masses is the dominant uncertainty in using cluster counts to estimate cosmological parameters. In the analysis of *Planck* clusters, the cluster tSZ observable was related to the cluster mass M_{500}[156] using X-ray scaling relations [26], calibrated against a subsample of the *Planck* clusters. The X-ray masses are, however, derived assuming hydrostatic equilibrium and are expected to be biased low [e.g., 245]. This was accounted for by multiplying the true masses by a so-called "hydrostatic mass bias" factor of $(1-b)$. The strongest constraints on this bias factor come from weak gravitational lensing estimates of cluster masses. Planck Collaboration XXIV [289] considered three lensing mass calibrations[157]: $(1-b) = 0.69 \pm 0.07$ from 22 *Planck* clusters from the Weighing the Giants lensing programme [364]; $(1-b) = 0.78 \pm 0.08$ from 37 *Planck* clusters calibrated by the Canadian Cluster Comparison Project [178]; and $1/(1-b) = 0.99 \pm 0.19$ from *Planck* CMB-lensing mass estimates of the MMF3 cluster sample [289]. More recently, Sereno et al. [337] have analysed 35 *Planck* clusters with galaxy shear data from the CFHTLenS [170] and RCSLenS [173] surveys, finding $(1-b) \approx 0.77 \pm 0.11$ for all clusters and $(1-b) = 0.68 \pm 0.11$ for the 15 clusters in the cosmological sample. Additionally, Penna-Lima et al. [262] use gravitational lensing measurements from HST images of 21 *Planck* clusters finding $(1-b) = 0.73 \pm 0.10$.

The determination of cosmological parameters such as σ_8 and Ω_m from *Planck* cluster counts is strongly dependent on the prior adopted for the mass bias parameter. In this paper, we use the *Planck* TT,TE,EE+lowE+lensing likelihood in combination with the *Planck* cluster counts to derive a constraint on $(1-b)$ (following similar analyses described in Planck Collaboration XX 288 and Salvati et al. 321). This gives Planck2018CosmolParam/

$$Planck2018CosmolParam/(1-b) = 0.62 \pm 0.03 \quad \begin{array}{l}(68\,\%,\ Planck\ \text{TT,TE,EE} \\ +\text{lowE}+\text{lensing}),\end{array} \tag{1228}$$

compared to 0.58 ± 0.04 using the 2015 TT,TE,EE+lowP likelihood [288]. The roughly $1\,\sigma$ upward shift in Eq. (1228) is mainly caused by the 2018 change in the τ constraint. The mass bias of Eq. (1228) is at the lower end of the weak-lensing mass estimates, but is about $2\,\sigma$ lower that the *Planck* CMB-lensing mass calibration reported in [288].

Zubeldia & Challinor [378] have revisited the *Planck* CMB-lensing mass calibration, incorporating the CMB-lensing mass estimates within a likelihood describing the *Planck* cluster

[156] The mass contained within a sphere of radius R_{500}, centred on the cluster, where R_{500} is the radius at which the mean density is 500 times the critical density at the redshift of the cluster.

[157] See [340] for a discussion of dynamical mass estimates for SZ-selected clusters.

counts, together with a *Planck* prior on $\theta_{\rm MC}$. This study corrects for significant biases in the analysis reported in [288]. [378] find $(1-b) = 0.71 \pm 0.10$ and $\sigma_8(\Omega_{\rm m}/0.33)^{0.25} = 0.765 \pm 0.035$. These results, based entirely on *Planck* data, are consistent with the base-ΛCDM parameters from the *Planck* power spectra and with the inferred mass bias of Eq. (1228).

Since PCP15 there have been a number of new analyses of cluster counts using other surveys. Two recent studies [231, 95], with very different selection criteria, use weak gravitational lensing mass determinations from the Weighing the Giants programme to calibrate cluster scaling relations. [95] analysed a sample of 377 clusters at $z > 0.25$ identified with SPT, finding $\sigma_8(\Omega_{\rm m}/0.27)^{0.3} = 0.797 \pm 0.031$, while Mantz et al. [231] analysed an X-ray-selected sample of clusters from the ROSAT All-Sky survey, finding $\sigma_8(\Omega_{\rm m}/0.3)^{0.17} = 0.81 \pm 0.03$. These measurements can be compared to our baseline ΛCDM constraints (*Planck* TT,TE,EE+lowE +lensing) of $\sigma_8(\Omega_{\rm m}/0.27)^{0.3} = 0.849 \pm 0.010$ and $\sigma_8(\Omega_{\rm m}/0.3)^{0.17} = 0.817 \pm 0.076$.

Schellenberger & Reiprich [326] have analysed a sample of 64 of the brightest X-ray clusters using a prior on the hydrostatic mass bias from [52]. These authors find $\Omega_{\rm m} = 0.303 \pm 0.009$, $\sigma_8 = 0.790^{+0.030}_{-0.028}$, and $S_8 = \sigma_8(\Omega_{\rm m}/0.3)^{1/2} = 0.792 \pm 0.054$. Each of these numbers is within about $1\,\sigma$ of the *Planck* base-ΛCDM best-fit cosmology reported in this paper. Finally, we mention the analysis of ROSAT-observed X-ray clusters carried out by Böhringer et al. [58, 59]. These authors choose a central value for the hydrostatic mass bias of $(1-b) = 0.9$, although they allow for small variations in the slope (7 %) and normalization (14 %) of the X-ray luminosity-mass relation; they find constraints the σ_8–$\Omega_{\rm m}$ plane in tension with *Planck* at about $2.5\,\sigma$.

In summary, accurate calibrations of cluster masses are essential if cluster counts are to be used as cosmological probes. Given the uncertainties in these calibrations, we do not use cluster counts in our main parameter grid. Consistency of cluster counts with the best-fit *Planck* base-ΛCDM cosmology requires hydrostatic mass biases (Eq. 1228) that are at the lower end, but within about $1\,\sigma$ of bias factors estimated from weak-lensing cluster masses. The combined *Planck* CMB-lensing and cluster-count analysis reported by [378] is in good agreement with the *Planck* base-ΛCDM cosmology. At this time, there is no compelling evidence for a discrepancy between *Planck*-, SPT-, or X-ray-selected cluster counts and the base-ΛCDM model.

27.6 Internal consistency of ΛCDM model parameters

In this section we briefly discuss a couple of curious features of the *Planck* data that lead to moderate tensions in parameter consistency tests. We first discuss how parameters vary between high and low multipoles, and the relevant features in the power spectra that may be responsible for these shifts. We then discuss the related issue of how the full multipole range appears to prefer more lensing than predicted by ΛCDM fits. We end this section with a discussion of systematic uncertainties.

27.6.1 Consistency of high and low multipoles

The *Planck* CMB temperature power spectrum shows a conspicuous dip over the multipole range $20 \lesssim \ell \lesssim 30$ compared to ΛCDM fits, as can be seen in Fig. 167. This feature was first observed by WMAP [43], and was discussed in detail in PCP13. Since it is detected consistently by both WMAP and *Planck* at multiple frequencies, it cannot plausibly be explained by an instrumental systematic or foreground. The large-scale *Planck* temperature map is signal dominated, so the dip feature is almost identical in this final release. PCP13 also noted an approximately $2.7\,\sigma$ mismatch between the best-fit ΛCDM cosmology and the amplitude of the measured temperature power spectrum at $\ell \leq 30$. However, with the tighter optical depth constraints used in this paper and improvements in the high multipole likelihoods we find no strong evidence for an amplitude mismatch. The *Planck* TT,TE,EE+lowE+lensing best-fit ΛCDM model provides a good overall fit to the temperature multipoles at $\ell < 30$ ($\chi^2_{\rm eff} \approx 23$ for 28 data points), and because of the skewed χ^2-like distribution of the CMB spectrum estimators, it is expected that typically more than half of the data points are below the theoretical model values (see Fig. 167). The statistical significance of the dip feature is hard to quantify, since it was identified a posteriori, but PPL15 suggest a significance of about 2.8 % after maximizing over extremal ℓ ranges found in simulations. This could be an indication of new physics at large scales, for example associated with a sharp feature in the inflationary potential [as considered by 260, and many subsequent researchers]. Alternatively, it could just be a statistical fluctuation,

which is our baseline assumption. However, since the dip is a relatively unusual fluctuation and it is near one end of the multipole range, it tends to pull cosmological parameters more than would be expected in typical realizations of a ΛCDM cosmology. This needs to be borne in mind in assessing parameter shifts between low and high multipoles.

WMAP measured the CMB temperature fluctuations up to $\ell \approx 800$ [42]. The higher-resolution data from *Planck* substantially increases the multipole range of the temperature power spectrum out to $\ell \approx 2500$. Cosmological parameters are therefore expected to shift (usually towards the truth) from the mean posterior values measured by WMAP, together with a reduction in the error bars. This is what is seen, with the *Planck* values of H_0 and n_s decreasing, and Ω_m and $\Omega_m h^2$ increasing, along with substantially smaller errors. However as noted in PCP13 the magnitudes of the shifts appear to be slightly larger than might be expected statistically, assuming the base-ΛCDM cosmology. This stimulated additional work on the consistency of the *Planck* power spectra reported in PPL15 and to further investigations of the consistency of cosmological parameters measured from high and low multipoles from *Planck* [6, 299]. As noted in the introduction, there is a very good agreement between *Planck* and WMAP temperature maps on the scales observed by WMAP [282, 186], but an inconsistency with high multipoles could indicate either new physics beyond ΛCDM, or the presence of some unidentified systematics associated with the *Planck* data and/or the foreground model. Planck Collaboration Int. LI [299] find that although some cosmological parameters differ by more than $2\,\sigma$ between $\ell < 800$ and $\ell > 800$, accounting for the multi-dimensional parameter space including correlations between parameters, the shifts are at the 10 % level and hence not especially unusual. Nonetheless, parameter shifts, particularly in the fluctuation amplitude and Hubble parameter (which are directly relevant for the ΛCDM-comparison with external data, as discussed in Sect. 27.5) are worth a brief re-examination using the additional information provided by the *Planck* polarization spectra.

Constraints on cosmological parameters from power spectra at high multipoles require a foreground model. Previous studies have shown that results are not very sensitive to the specific assumptions that are made within the broad context of slowly varying foreground spectra expected on physical grounds [6, 299]. In this section, we use the `plik_lite` *Planck* likelihood, described in detail in PPL18, which has the standard `Plik` foreground and nuisance parameters marginalized out without further assumptions on the cosmology.[158] For standard model extensions `plik_lite` accurately reproduces results from the full `Plik` likelihood. It allows us to explore the high-ℓ likelihood accounting for foreground uncertainties, but with the foregrounds constrained in a sensible way from their spectra over the full multipole range. We consider the multipole ranges $\ell \leq 801$ and $\ell \geq 802$ (corresponding to the boundary of one of the `plik_lite` bins), so that the low-multipole range is roughly comparable to WMAP and the two ranges have similar statistical power on most parameters. Results splitting at $\ell \approx 1000$ are similar, but with larger errors in the high multipole range.

Figure 187 shows a comparison of the high and low multipole ranges, both for temperature (lower triangle, as previously discussed by Addison et al. 6 and Planck Collaboration Int. LI 299), and new results for the combined temperature-polarization likelihood (upper triangle). Part of the difference between the low- and high-multipole ranges is caused by the large-scale temperature dip discussed above; if we exclude multipoles $\ell < 30$ (unfilled grey contours), the contours from $\ell \leq 801$ shift towards the area of consistency with the high multipoles. This could indicate that the low-multipole results have been pulled unusually far from the truth by the large-scale power spectrum dip; if so, the WMAP temperature results would also have been pulled at a similar (but not identical) level. The region of overlap of the high- and low-multipole parameter constraints is consistent with constraints from the nearly-independent combination of EE polarization and lensing with a conservative $\Omega_b h^2$ prior (green contours). This is consistent with a statistical fluctuation pulling the low and high multipoles in opposite directions, so that their intersection is closer to the truth if ΛCDM is correct.

Figure 188 shows marginalized individual parameter constraints, and also a comparison with the results from the polarization likelihoods at high and low multipoles. The $\ell \geq 802$ temperature results pull parameters to a region of higher matter density and fluctuation amplitude (and to lower n_s and H_0) than the lower multipole range, and predict a CMB lensing amplitude parameter $\sigma_8 \Omega_m^{0.25} = 0.649 \pm 0.018$. This is in tension with the CMB lensing-reconstruction

[158] We do not attempt to quantify likelihood modelling differences in this section, but a `CamSpec`-based likelihood gives slightly less tension between high and low multipoles (especially with polarization), associated with the weaker preference for $A_L > 1$, as discussed in more detail in Sect. 27.6.2.

measurement of $\sigma_8\Omega_{\rm m}^{0.25} = 0.589 \pm 0.020$ at $2.2\,\sigma$ (as pointed out by Addison et al. 6 with 2015 data; also see the closely-related discussion in the next subsection). As shown in Fig. 188, combining the $\ell \geq 802$ CMB likelihood with the lensing reconstruction, all parameter results move back towards the same region of parameter space as combining with $\ell \leq 801$, consistent with the high-ℓ temperature result having fluctuated high along the main degeneracy direction. As discussed in Sects. 27.2.3 and 27.3.3, the combined CMB power spectrum results over the full multipole range are consistent with the lensing likelihood.

It is also interesting to compare to parameter constraints from the CMB power spectrum multipoles $\ell \leq 801$ combined with the lensing and BAO, which gives

$$\left.\begin{aligned} Planck2018CosmolParam/H_0 &= (67.85 \pm 0.52)\ \text{km s}^{-1}\text{Mpc}^{-1}, \\ \sigma_8 &= 0.8058 \pm 0.0063, \\ \Omega_{\rm m} &= 0.3081 \pm 0.0065. \end{aligned}\right\} \begin{array}{l} 68\,\%,\ \text{TT,TE,EE} \\ [\ell \leq 801] \\ +\text{lowE} \\ +\text{lensing} \\ +\text{BAO}. \end{array} \tag{1229}$$

These results are entirely independent of the cosmological parameter fit to the $\ell \geq 801$ power spectra, but agree well at the $1\,\sigma$ level with the full joint results in Table 24 (which have similar errors on these parameters). An equivalent result could be obtained using WMAP data after replacing their low-ℓ polarization with the *Planck* HFI measurement (i.e., lowE).

For the temperature likelihoods, the difference between the low- and high-multipole constraints remains evident, with $\Omega_{\rm m}h^2$ differing at the $2.8\,\sigma$ level. Adding polarization, the results from the multipole ranges are more consistent, as shown in Fig. 188, though the difference in $\Omega_{\rm m}h^2$ is still unusual at the roughly $2\,\sigma$ level. However, the shifts in the different parameters are all highly correlated, due to partial parameter degeneracies, so the significance of any individual large shift is lower after accounting for the number of parameters [299]. The internal tensions between multipole ranges appear to be consistent with moderate statistical fluctuations, related to the low-ℓ dip at large scales and correlated with the lensing amplitude on small scales. The large-scale feature is well determined by both WMAP and *Planck* and very robustly measured. The internal consistency of the *Planck* power spectra between different frequencies and detectors (PPL15, PPL18) argues against systematics driving large parameter shifts at high multipoles. Equation (1229) also demonstrates that any effect from the high-multipole spectra alone cannot be pulling our baseline parameters by more than about $1\,\sigma$. In the next subsection we describe in more detail the apparent preference for a higher lensing amplitude, and the features in the observed spectrum that could be responsible for it.

27.6.2 Lensing smoothing and $A_{\rm L}$

In addition to the direct measurement of CMB lensing described in Sect. 27.2.3 and PL2018, lensing can be seen in the *Planck* CMB power spectra via the lensing-induced smoothing of the acoustic peaks and transfer of power to the damping tail. This effect is modelled in our main parameter analysis, and can be calculated accurately from the unlensed CMB power spectra and the CMB lensing potential power spectrum in each model [332, 220]. Interesting consistency checks include testing if the amplitude of the smoothing effect in the CMB power matches expectation and whether the amplitude of the smoothing is consistent with that measured by the lensing reconstruction. To do this, the theoretical prediction for the lensing spectrum in each model is often scaled by an "$A_{\rm L}$" consistency parameter, where the theoretical expectation is that $A_{\rm L} = 1$ [66].

As shown in Fig. 169, the *Planck* lensing-reconstruction power spectrum is consistent with the amplitude expected for ΛCDM models that fit the CMB spectra, so the *Planck* lensing measurement is compatible with $A_{\rm L} = 1$. However, the distributions of $A_{\rm L}$ inferred from the CMB power spectra alone are shown in Fig. 189 for various different data combinations, and these indicate a preference for $A_{\rm L} > 1$, with

$$A_{\rm L} = 1.243 \pm 0.096 \quad (68\,\%,\ \textit{Planck}\ \text{TT+lowE}), \tag{1230a}$$
$$A_{\rm L} = 1.180 \pm 0.065 \quad (68\,\%,\ \textit{Planck}\ \text{TT,TE,EE+lowE}), \tag{1230b}$$

assuming a ΛCDM+$A_{\rm L}$ model. The TE polarization data alone slightly prefer $A_{\rm L} < 1$, with the EE data slightly preferring $A_{\rm L} > 1$; however, both are consistent with $A_{\rm L} = 1$ within $2\,\sigma$. The

joint combined likelihood shifts the value preferred by the TT data downwards towards $A_{\rm L} = 1$, but the error also shrinks, increasing the significance of $A_{\rm L} > 1$ to $2.8\,\sigma$ (99.8 % of parameter samples have $A_{\rm L} > 0$, so the one-tailed limit is almost exactly $3\,\sigma$). Moreover, combining with the lensing likelihood further pulls the constraint towards $A_{\rm L} = 1$, which is then consistent with the data to within $2\,\sigma$; we see that the preference for $A_{\rm L} > 1$ is driven by the CMB power spectra alone.

The preference for high $A_{\rm L}$ is not just a volume effect in the full parameter space (see PCP13 for discussion of such effects in multi-parameter fitting), with the best fit improved by $\Delta\chi^2_{\rm eff} = -8.7$ when adding $A_{\rm L}$ for TT+lowE and $\Delta\chi^2_{\rm eff} = -9.7$ for TT,TE,EE+lowE. The bulk of the $\Delta\chi^2_{\rm eff}$ comes from the high-ℓ likelihood (mostly in the range $600 < \ell < 1500$); however, the low-ℓ temperature `commander` likelihood fit is also improved if $A_{\rm L}$ is free, with $\Delta\chi^2_{\rm eff} = -2.3$ and $\Delta\chi^2_{\rm eff} = -1.3$ for the TT+lowE and TT,TE,EE+lowE, respectively, due to the lower amplitude of the $A_{\rm L}$ fit on large scales. The change in fit to the low-ℓ polarization is not very significant ($\Delta\chi^2_{\rm eff} = -0.2$ and $\Delta\chi^2_{\rm eff} = -0.4$).

The determination of $A_{\rm L}$ from the high-ℓ polarization data and the TT,TE,EE+lowE joint combination depends on the calibration of the polarization channels, and is affected by different ways of modelling the polarization efficiencies, as discussed in Sect. 27.2.2. The results from the `CamSpec` likelihood (which uses spectrum-based rather than map-based calibrations for TE and EE) are somewhat shifted with respect to the `Plik` likelihood, as shown by the dotted lines in Fig. 189, and have larger errors, giving

$$A_{\rm L} = 1.246^{+0.092}_{-0.100} \quad (68\,\%,\ \text{TT+lowE [CamSpec]}), \tag{1231a}$$

$$A_{\rm L} = 1.149 \pm 0.072 \ (68\,\%,\ \text{TT,TE,EE+lowE [CamSpec]}). \tag{1231b}$$

Using `CamSpec` there is still a clear preference for $A_{\rm L} > 1$, but the joint result with polarization is now only just over $2\,\sigma$ above $A_{\rm L} = 1$. The differences between these `Plik` and `CamSpec` results arise from differences in the methodologies used to create the likelihoods. Although both likelihoods clearly show a preference for $A_{\rm L} > 1$, this cannot be claimed to be a robust detection at much over $2\,\sigma$ [see also 117].

The preference for $A_{\rm L} > 1$ within the ΛCDM model is a curious feature of the *Planck* CMB power spectrum data, and has already been discussed extensively in PCP13, PCP15, and Planck Collaboration Int. LI [299], although it is now slightly more significant. In temperature, over half of the small (approximately 0.02) upward shift in $A_{\rm L}$ compared to 2015 is explained by the lower optical depth from the 2018 low-ℓ likelihood: lower τ implies lower $A_{\rm s}$ to match the high-ℓ CMB fluctuation amplitude, and hence larger $A_{\rm L}$ to yield a lensing amplitude and hence amount of smoothing at the same level as 2015. In polarization about 40 % of the shift in $A_{\rm L}$ is explained by changes in τ, with changes in the maps, modelling for beam leakage, and polarization efficiencies explaining the rest.

The high-ℓ temperature likelihood preference for more lensing smoothing than allowed by ΛCDM can be seen by eye in the smoothed data residuals plotted in Fig. 190; over almost all the allowed ΛCDM parameter space there is an oscillatory residual in the range $1100 \lesssim \ell \lesssim 2000$ that matches the shape of the lensing smoothing[159] (although in other multipole ranges it does not match at all). The residual is not obviously anomalous, with the TT ΛCDM best fit improving by $\Delta\chi^2 \approx 4$ if a best-fit oscillatory residual (with $A_{\rm L} \approx 1.1$) is added to the best-fit ΛCDM theory model. The stronger preference for $A_{\rm L} > 1$ when $A_{\rm L}$ varies arises because degeneracies between $A_{\rm L}$, cosmological parameters, and foregrounds improves the fit at both high and lower multipoles, as shown by the black dashed line in Fig. 190. In ΛCDM the lensing amplitude can be increased by increasing $\Omega_{\rm m}h^2$; however, the model then becomes a bad fit because of the poorer agreement at $\ell < 1000$). Varying $A_{\rm L}$ allows a high $A_{\rm L}$ to remove the oscillatory residual at high multipoles that appears in ΛCDM with lower $\Omega_{\rm m}h^2$, giving best fits with lower $\Omega_{\rm m}h^2$ and higher H_0 (by 1.5–2.0 σ, depending on the exact combination of data used) that are not favoured in the physical ΛCDM model. Lower values of $\Omega_{\rm m}h^2$ give higher values of $n_{\rm s}$, lowering the theory prediction on large scales, so high $A_{\rm L}$ models are also slightly

[159] Although the oscillatory pattern looks most similar to lensing at high multipoles, an increase in the foreground model amplitude can decrease the oscillation amplitude in the theory contribution to the spectrum, and hence appear as an oscillatory difference. For example $\Delta n_{\rm s} \approx -0.02$, combined with an implausibly large change in the foreground model, gives a difference in the predicted spectrum with an oscillatory component that has similar amplitude to $\Delta A_{\rm L} \approx 0.1$; see the related discussion in Planck Collaboration Int. LI [299].

preferred by the dip in the $\ell < 30$ *Planck* temperature data. The parameter degeneracies are illustrated in Fig. 191.

The $A_{\rm L}$ results appear to be robust to changes in foreground modelling in the baseline likelihood, with the `CamSpec` 545 GHz cleaned likelihood (see Appendix **??**) giving very similar results. However, the dip in the residuals at $1420 \lesssim \ell \lesssim 1480$, part of the oscillatory feature that looks like additional lensing, nearly coincides with an approximately $3\,\sigma$ discrepancy (for the best-fit foreground cosmology model) between the 143-GHz and 217-GHz power spectra at $1450 \lesssim \ell \lesssim 1510$, with the 217-GHz spectrum pulling the coadded spectrum low compared to 143 GHz by an amount comparable to the coadded residual (at $\ell = 1480$ the 217-GHz spectrum is $\mathcal{D}_\ell \approx 7\mu{\rm K}^2$ lower than 143 GHz with smoothing $\sigma_\ell = 40$; see PPL18). This may be an indication that the preference for $A_{\rm L}$ at high multipoles is partly due to unknown systematics or foregrounds. However, tightly cutting the ℓ range that contributes to the $3\,\sigma$ frequency difference does not in itself shift $A_{\rm L}$ to substantially lower values (though cutting all of $1420 \lesssim \ell \lesssim 1480$ does), and the significance of the oscillatory feature in the ΛCDM CMB residual is in any case not very high. If it is largely a statistical fluctuation, it would be expected to vary with changes in sky area; that is somewhat the case, with around 80 % sky area giving a substantially less oscillatory residual to the same best-fit ΛCDM model at $\ell \lesssim 1600$, but still favouring high $A_{\rm L}$. Different power spectrum analyses have also shown the preference for $A_{\rm L}$ [347, 83], though with varying significance, which could indicate that our roughly $3\,\sigma$ significance is partly an issue of analysis choices, e.g., the sky areas included and foreground priors chosen.

The dashed line in Fig. 169 shows the lensing power spectrum in the *Planck* TT,TE,EE +lowE best-fit ΛCDM+$A_{\rm L}$ model, which is clearly inconsistent with the lensing reconstruction, since it lies above almost all of the measured data points. Because the amplitude of the lensing smoothing effect can be calculated from the lensing potential power spectrum alone, which we can also empirically measure, it is impossible to increase the lensing smoothing of the CMB peaks without also increasing the measured lensing reconstruction amplitude. This remains true if the lensing power spectrum is allowed to vary in shape [242]. The actual lensing smoothing effect can also partly be removed by delensing, as shown by [212] and [70]. In PL2018 we update these delensing analyses, and show (using the internal lensing reconstruction, a *Planck* CIB map as a tracer of the lensing potential, and a combined estimate) that the amount of peak sharpening observed after delensing is consistent with theoretical expectations (e.g., for the TT spectrum, we measure a reduction in lensing smoothing of 0.411 ± 0.028, compared to the expected value of 0.375 when using a combination of CIB and *Planck* lensing reconstruction).

Although the residuals shown in Fig. 190 between the data and the ΛCDM best fit temperature spectrum show what looks like an oscillatory lensing residual at high ℓ, the fit itself is determined by the entire range of multipoles (and the low-ℓ polarization than constrains τ). The preference for $A_{\rm L} > 1$ could therefore be attributed to other scales when considering the CMB spectra alone. For example, after removing $\ell < 30$ in both temperature and polarization, $A_{\rm L}$ from TT is consistent with unity to within $1\,\sigma$. However, in this case the ΛCDM lensing amplitudes are still large, giving a value of $\Omega_{\rm m}\sigma_8^{0.25}$, in $2\,\sigma$ tension with the lensing reconstruction. This is another reflection of the tension noted in Sect. 27.6.1 between the lensing reconstruction and the lensing amplitude predicted using temperature multipoles $\ell \gtrsim 800$: the two tensions are therefore not independent and largely driven by the same features of the ΛCDM fit to the temperature and low-ℓ polarization data.

If $A_{\rm L} > 1$ is not just a statistical fluctuation, but comes from new physics changing the theoretical predictions, it could be something that mimics the smoothing effect in the CMB peaks. The *lensing* smoothing effect comes from averaging over the sky a spectrum that is locally varying (due to magnification and shear locally changing the scale and shape of the CMB peaks). Conceptually, the temperature lensing reconstruction works by looking for this spatial variation in scale and shear of the local power. Any non-lensing isotropic change in the amplitude of the small-scale peaks and troughs, either from new physics or random fluctuations, would therefore only have a small effect on the lensing reconstruction, which is sensitive to scale and shape, not amplitude.

One locally anisotropic physical effect that has been considered as a possible explanation is the presence of large-scale compensated isocurvature modes, discussed in detail in [296]. Because the large-scale isocurvature modes locally vary the baryon-to-photon ratio, they can partially mimic the lensing smoothing effect by spatially varying the acoustic scale [243, 357]. However, because they have a similar local effect to lensing, they also affect the large-scale lensing reconstruction [345]. Combining with the *Planck* 2018 lensing reconstruction, which

now extends down to $L = 8$, as shown in [296] this model therefore does not offer a significant improvement in overall fit ($\Delta\chi^2_{\rm eff} = -3.3$ with *Planck* TT,TE,EE+lowE+lensing).

If the $A_{\rm L} > 1$ preference is simply a statistical excursion (perhaps the most likely explanation), this indicates that there are random features in the spectrum that are pulling some parameters unusually far from expected values.[160] There are several theoretical models that can fit the CMB power spectra and also predict larger lensing amplitudes. These include ΛCDM models with spatial curvature, for which we find $\Omega_K < 0$ at over $3\,\sigma$ (Sect. **??**) from the CMB power spectra, and some dark energy and modified gravity models (Sect. **??**). For extensions to base-ΛCDM, parameters that decrease the lensing amplitude are more constrained by the *Planck* power spectra than might otherwise be expected; for example, higher neutrino masses lower the predicted lensing power compared to base ΛCDM, leading to surprisingly tight constraints (Sect. **??**). Adding the lensing-reconstruction information significantly reduces the parameter space of larger lensing amplitudes and partially mitigates these effects. However, the statistical power of the *Planck* power spectra is sufficiently high that the joint constraints prefer lensing amplitudes in the higher range allowed by the lensing data.

Even within ΛCDM, the fact that the data prefer more lensing leads to a preference for higher fluctuation amplitudes, hence the high-ℓ data yield higher $A_{\rm s}$ and higher τ than we infer in combination with large-scale E-mode polarization (Sect. 27.2.2) or lensing reconstruction. Since these preferences are degenerate with $\Omega_{\rm m}h^2$, $n_{\rm s}$, and H_0 (see PCP13, PCP15), these parameters are also pulled ($\Omega_{\rm m}h^2$ higher, $n_{\rm s}$ and H_0 lower). Our baseline best-fit results include both the "lowE" data and the lensing reconstruction, each of which restrict the range of allowed variation, so the remaining pulls should be modest; however, it should not perhaps be too much of a surprise if the central values of the parameters inferred from *Planck* turn out to be slightly more shifted than typical with respect to the ultimate truth if the base-ΛCDM model is correct.

[160] It is not trivial to assess how unlikely a fluctuation in a consistency parameter like this is given the number of different cosmological and consistency test parameters we might have looked at. We are only discussing $A_{\rm L}$ in detail here because it comes out high; other consistency parameters, for example the relative amplitude of ISW, Doppler, and Sachs-Wolfe contributions to the temperature spectrum, come out perfectly consistent with expectations.

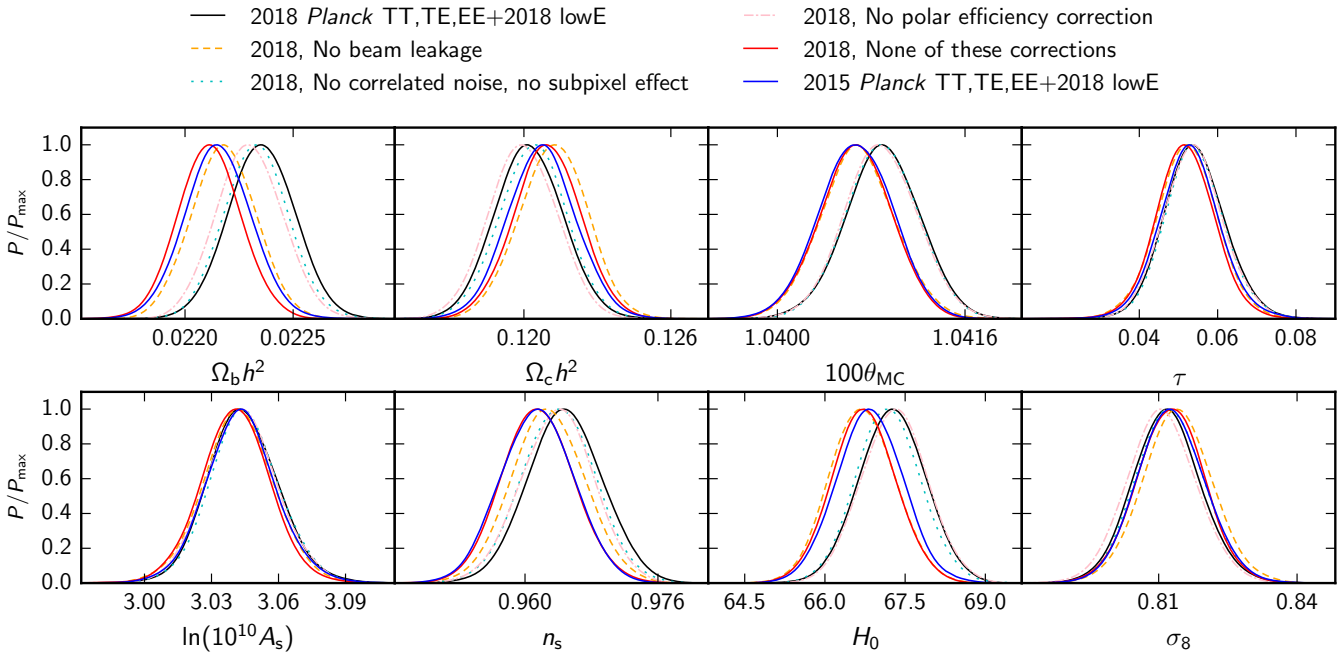


Figure 174: Impact of corrections for systematic effects on 2018 marginalized ΛCDM parameters from *Planck* TT,TE,EE+lowE. The plot shows the baseline results (black solid line), and the baseline result excluding corrections for various effects: beam leakage (dashed orange); polarization efficiencies (dot-dashed pink); and subpixel effects and correlated noise (dotted cyan). The impact of not including any of these corrections is shown in solid blue, and agree fairly well with the 2015 results if the 2015 low-ℓ polarization likelihood is replaced with 2018 lowE likelihood (2015 *Planck* TT,TE,EE+2018 lowE). This shows that corrections for polarization systematics account for most of the small changes between the 2015 and 2018 results that are not caused by the change in optical depth.

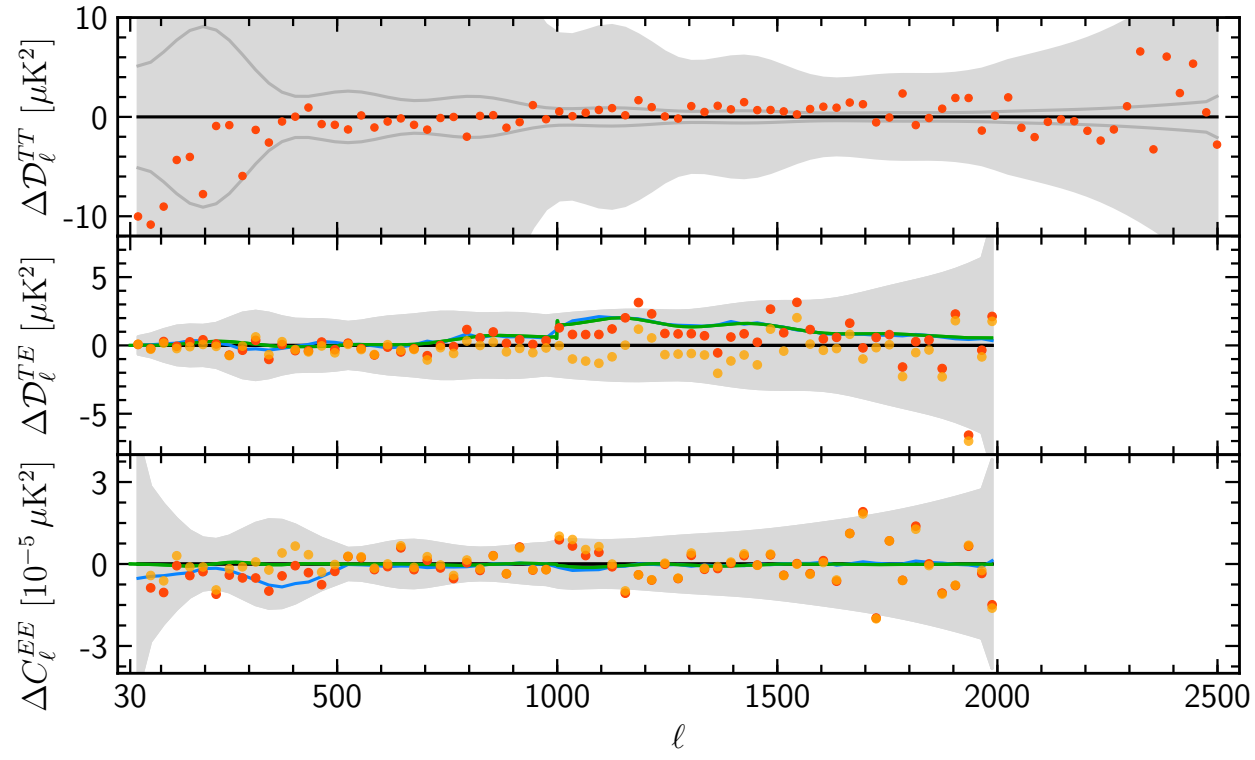
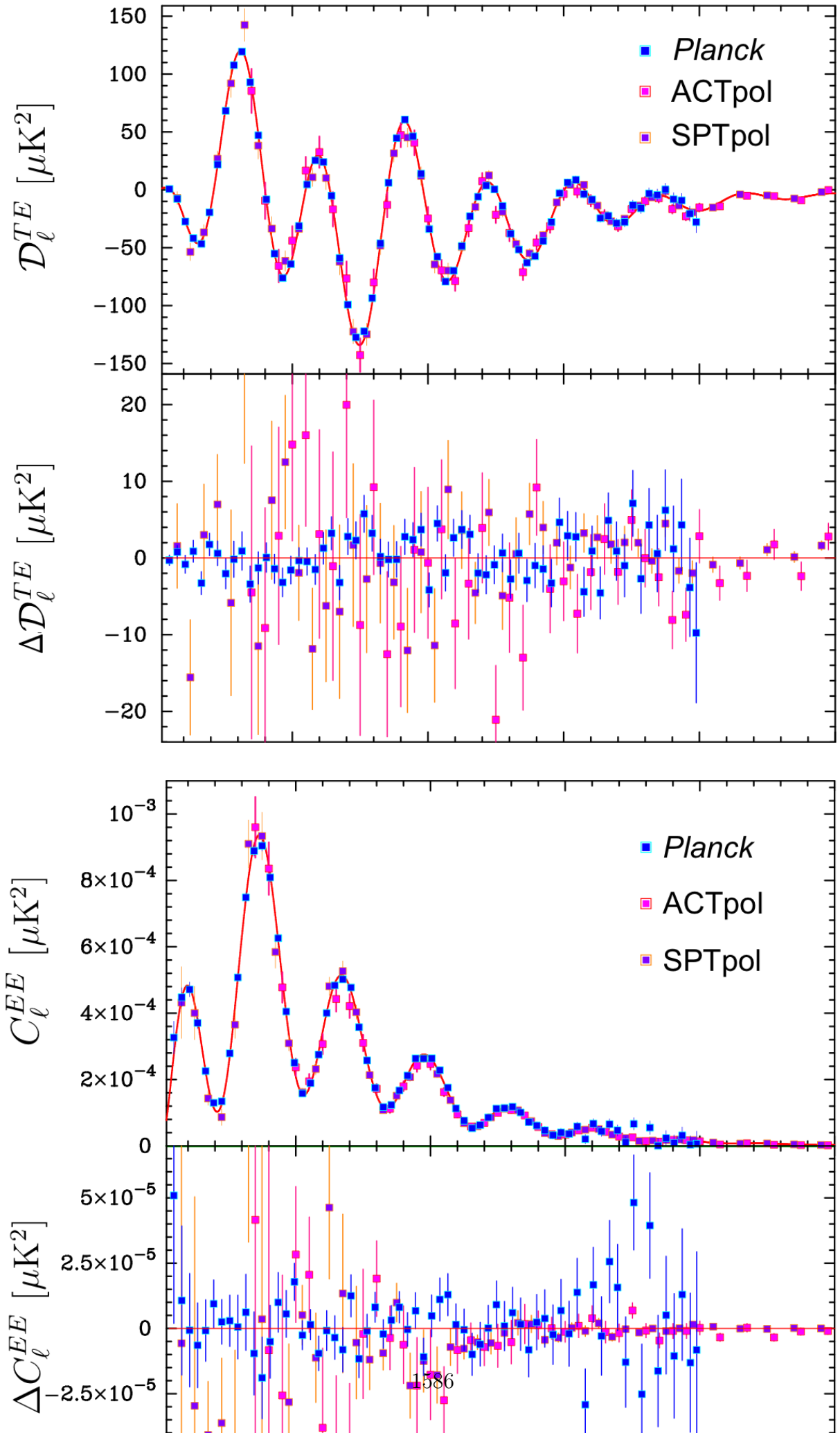


Figure 175: Differences between the 2018 and 2015 coadded power spectra at high ℓ in TT, TE, and EE from top to bottom (red points). The 2015 TT spectrum has been recalibrated by a factor of 1.00014. For TE and EE, the orange points show the same differences but without applying the polar-efficiency and beam-leakage corrections to the 2018 spectra. This shows


Planck
ACTpol
SPTpol
$\mathcal{D}_\ell^{TE}$ [μK^2]
$\Delta\mathcal{D}_\ell^{TE}$ [μK^2]
C_ℓ^{EE} [μK^2]
ΔC_ℓ^{EE} [μK^2]

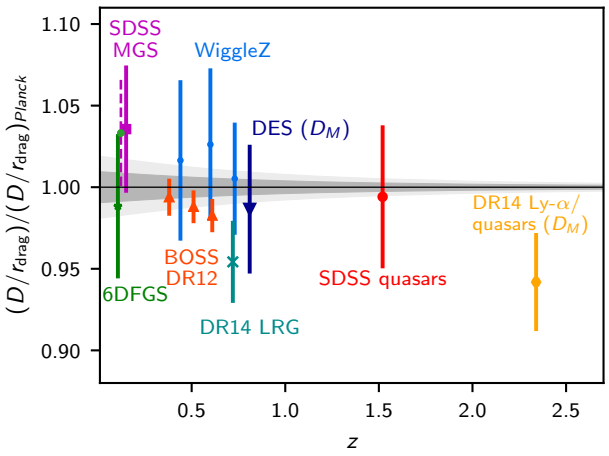


Figure 177: Acoustic-scale distance measurements divided by the corresponding mean distance ratio from *Planck* TT,TE,EE+lowE+lensing in the base-ΛCDM model. The points, with their $1\,\sigma$ error bars are as follows: green star, 6dFGS [47]; magenta square, SDSS MGS [313]; red triangles, BOSS DR12 [15]; small blue circles, WiggleZ [as analysed by 199]; large dark blue triangle, DES [107]; cyan cross, DR14 LRG [38]; red circle, SDSS quasars [28]; and orange hexagon, which shows the combined BAO constraints from BOSS DR14 Lyman-α [96] and Lyman-α cross-correlation with quasars, as cited in [57]. The green point with magenta dashed line is the 6dFGS and MGS joint analysis result of Carter et al. [71]. All ratios are for the averaged distance $D_{\rm V}(z)$, except for DES and BOSS Lyman-α, where the ratio plotted is $D_{\rm M}$ (results for $H(z)$ are shown separately in Fig. 182). The grey bands show the 68 % and 95 % confidence ranges allowed for the ratio $D_{\rm V}(z)/r_{\rm drag}$ by *Planck* TT,TE,EE+lowE+lensing (bands for $D_{\rm M}/r_{\rm drag}$ are very similar).

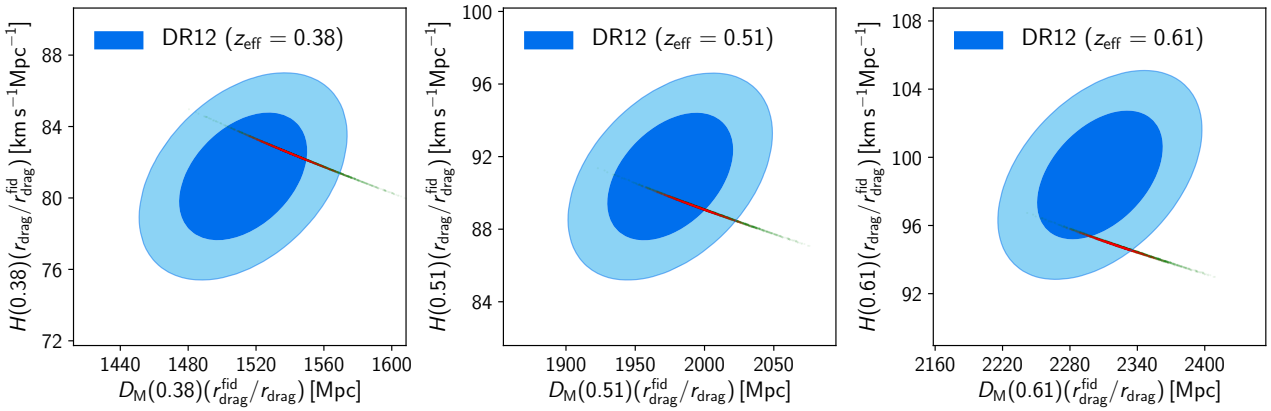


Figure 178: Constraints on the comoving angular diameter distance $D_{\rm M}(z)$ and Hubble parameter $H(z)$ at the three central redshifts of the Alam et al. [15] analysis of BOSS DR12. The dark blue and light blue regions show 68 % and 95 % CL, respectively. The fiducial sound horizon adopted by Alam et al. [15] is $r^{\rm fid}_{\rm drag} = 147.78\,$Mpc. Green points show samples from *Planck* TT +lowE chains, and red points corresponding samples from *Planck* TT,TE,EE+lowE+lensing, indicating good consistency with BAOs; one can also see the shift towards slightly lower $D_{\rm M}$ and higher H as more CMB data are added.

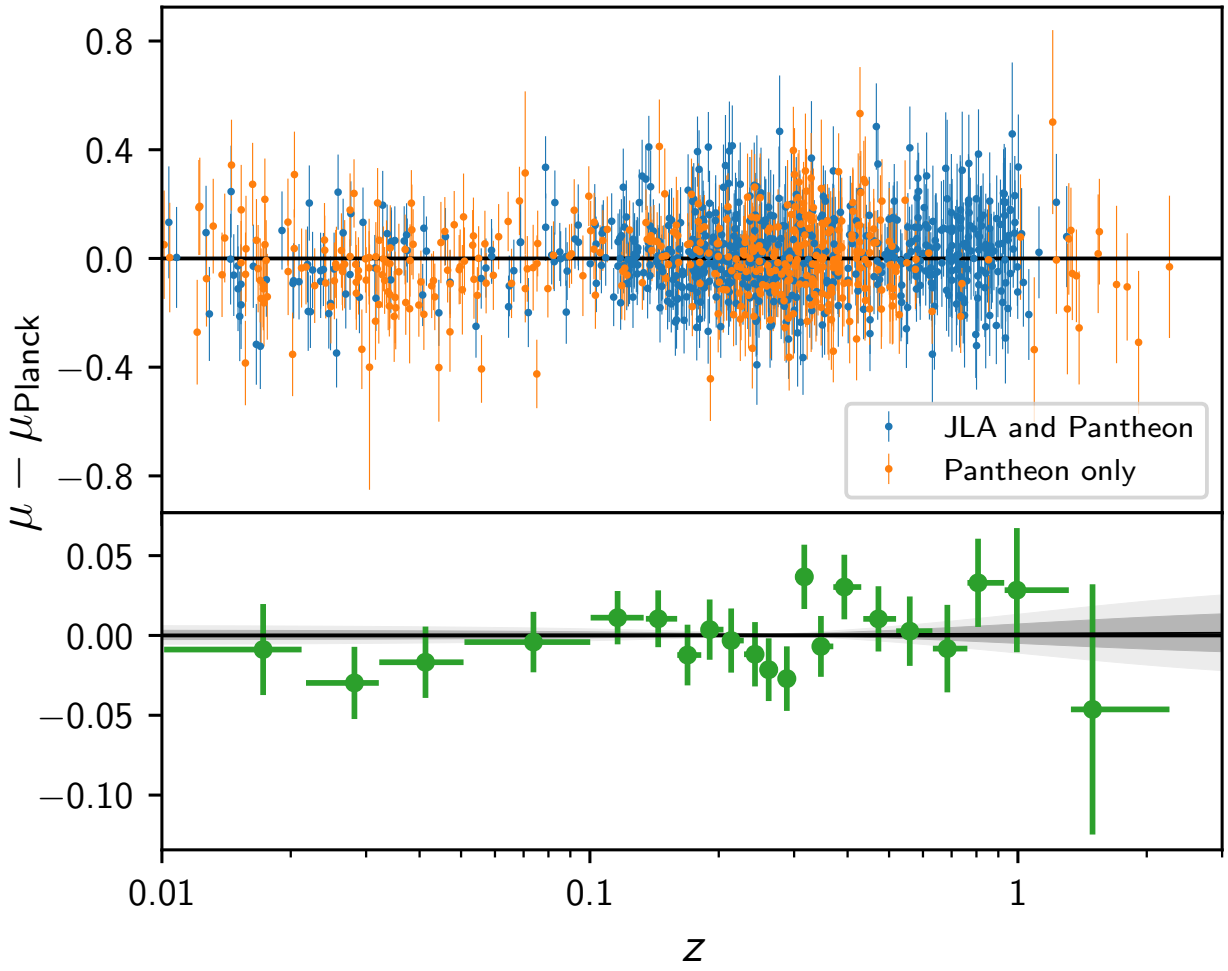


Figure 179: Distance modulus $\mu = 5\log_{10}(D_L) + \text{constant}$ (where D_L is the luminosity distance) for supernovae in the Pantheon sample [329] with $1\,\sigma$ errors, compared to the *Planck* TT,TE,EE +lowE+lensing ΛCDM best fit. Supernovae that were also in the older Joint Lightcurve Analysis [45, JLA] sample are shown in blue. The peak absolute magnitudes of the SNe, corrected for light-curve shape, colour, and host-galaxy mass correlations [see equation 3 of 329], are fixed to an absolute distance scale using the H_0 value from the *Planck* best fit. The lower panel shows the binned errors, with equal numbers of supernovae per redshift bin (except for the two highest redshift bins). The grey bands show the ± 1 and $\pm 2\,\sigma$ bounds from the *Planck* TT,TE,EE+lowE+lensing chains, where each model is calibrated to the best fit, as for the data.

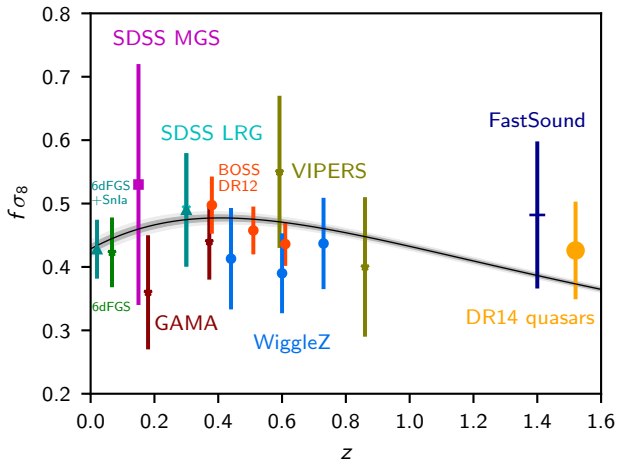


Figure 180: Constraints on the growth rate of fluctuations from various redshift surveys in the base-ΛCDM model: dark cyan, 6dFGS and velocities from SNe Ia [187]; green, 6dFGRS [48]; purple square, SDSS MGS [182]; cyan cross, SDSS LRG [249]; dark red, GAMA [54]; red, BOSS DR12 [15]; blue, WiggleZ [53]; olive, VIPERS [269]; dark blue, FastSound [250]; and orange, BOSS DR14 quasars [373]. Where measurements are reported in correlation with other variables, we here show the marginalized posterior means and errors. Grey bands show the 68 % and 95 % confidence ranges allowed by *Planck* TT,TE,EE+lowE+lensing.

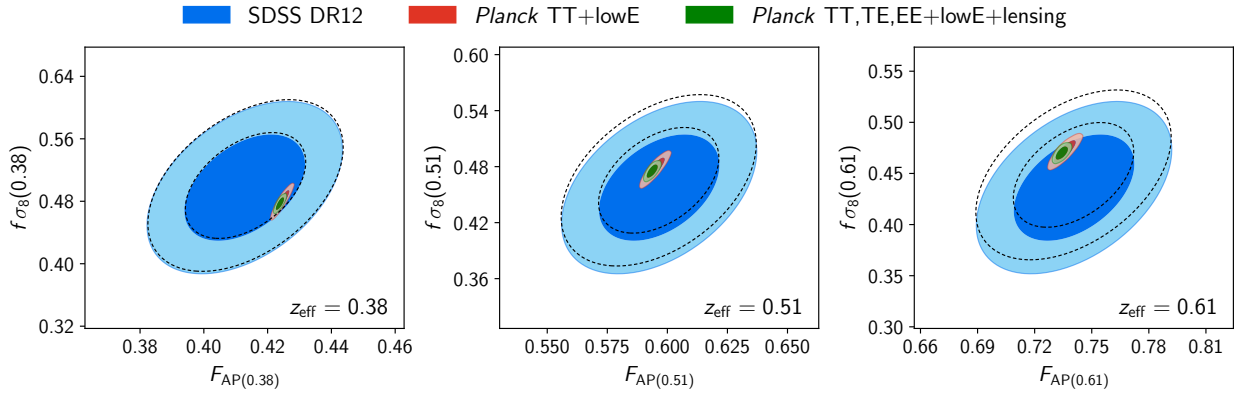


Figure 181: Constraints on $f\,\sigma_8$ and $F_{\rm AP}$ (see Eqs. 1223 and 1224) from analysis of redshift-space distortions. The blue contours show 68 % and 95 % confidence ranges on $(f\,\sigma_8, F_{\rm AP})$ from BOSS-DR12, marginalizing over D_V. Constraints from *Planck* for the base-ΛCDM cosmology are shown by the red and green contours. The dashed lines are the 68 % and 95 % contours for BOSS-DR12, conditional on the *Planck* TT,TE,EE+lowE+lensing constraints on D_V.

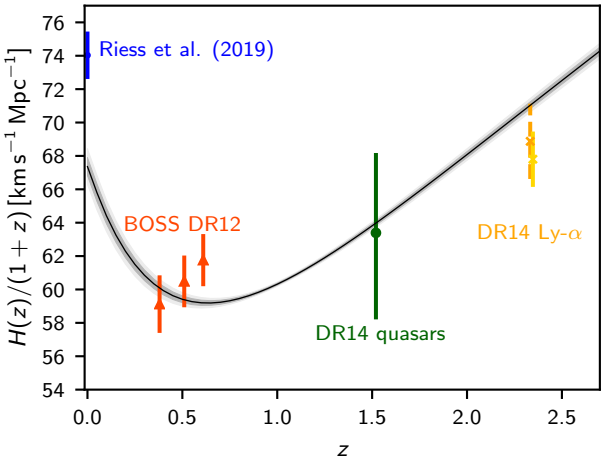


Figure 182: Comoving Hubble parameter as a function of redshift. The grey bands show the 68 % and 95 % confidence ranges allowed by *Planck* TT,TE,EE+lowE+lensing in the base-ΛCDM model, clearly showing the onset of acceleration around $z = 0.6$. Red triangles show the BAO measurements from BOSS DR12 [15], the green circle is from BOSS DR14 quasars [373], the orange dashed point is the constraint from the BOSS DR14 Ly α auto-correlation at $z = 2.34$ [96], and the solid gold point is the joint constraint from the Ly α auto-correlation and cross-correlation with quasars from Blomqvist et al. [57]. All BOSS measurements are used in combination with the *Planck* base-model measurements of the sound horizon $r_{\rm drag}$, and the DR12 points are correlated. The blue point at redshift zero shows the inferred forward-distance-ladder Hubble measurement from Riess et al. [306].

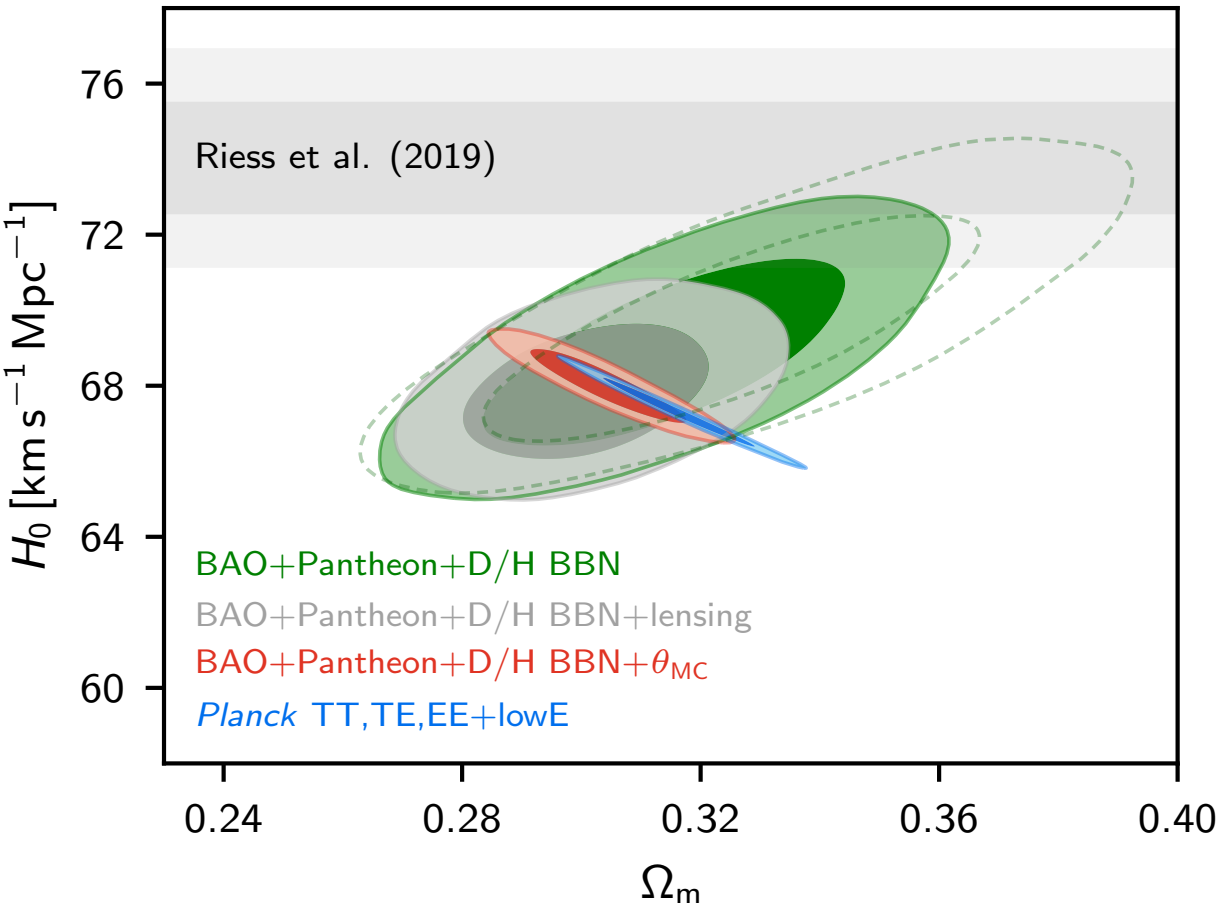


Figure 183: Inverse-distance-ladder constraints on the Hubble parameter and Ω_m in the base-ΛCDM model, compared to the result from the full *Planck* CMB power-spectrum data. BAO data constrain the ratio of the sound horizon at the epoch of baryon drag and the distances; the sound horizon depends on the baryon density, which is constrained by the conservative prior of $\Omega_b h^2 = 0.0222 \pm 0.0005$, based on the measurement of D/H by Cooke et al. [82] and standard BBN with modelling uncertainties. Adding *Planck* CMB lensing constrains the matter density, or adding a conservative *Planck* CMB "BAO" measurement ($100\theta_{MC} = 1.0409 \pm 0.0006$) gives a tight constraint on H_0, comparable to that from the full CMB data set. Grey bands show the local distance-ladder measurement of Riess et al. [306]. Contours contain 68 % and 95 % of the probability. Marginalizing over the neutrino masses or allowing dark energy equation of state parameters $w_0 > -1$ would only lower the inverse-distance-ladder constraints on H_0. The dashed contours show the constraints from the data combination BAO+JLA+D/H BBN .

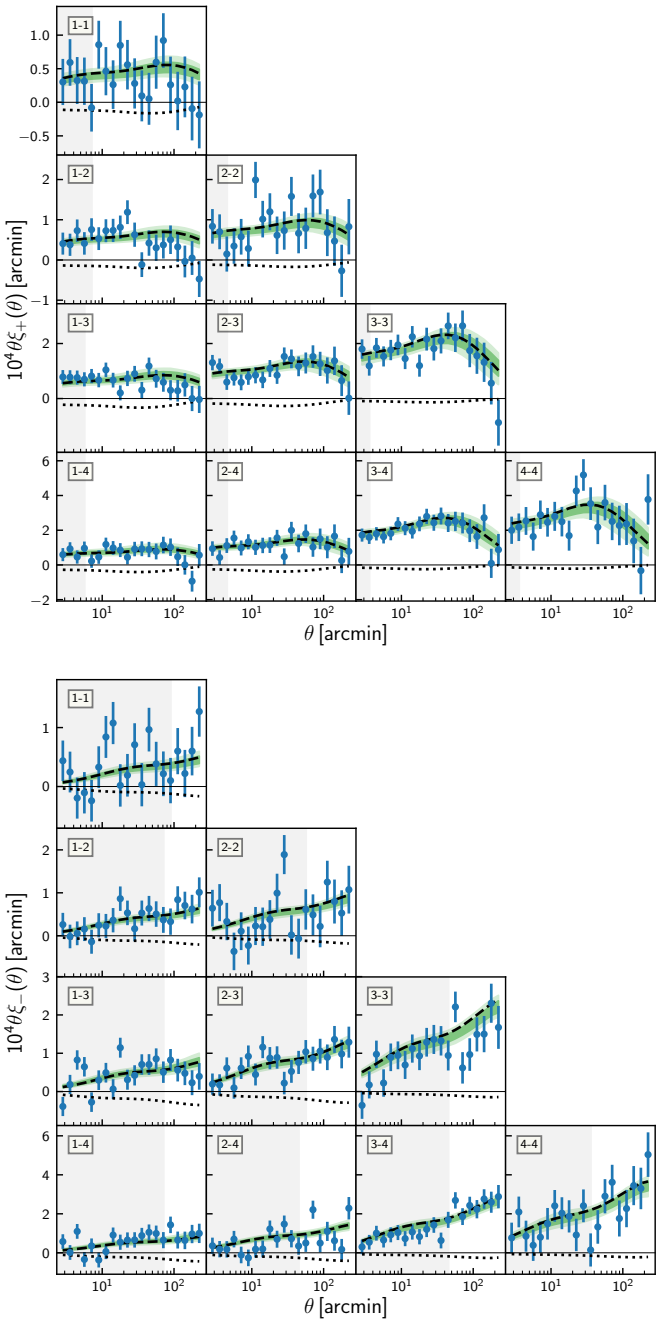


Figure 184: Dark Energy Survey (DES) shear correlations functions, ξ_+ (left) and ξ_- (right), for the auto- and cross-correlation between the four DES source redshift bins [355]. Green bands show the 68 % and 95 % distribution of model fits in the DES lensing-only base-ΛCDM parameter fits. The dashed line shows the DES lensing parameter best fit when the cosmological parameters are fixed to the best fit model for *Planck* TT,TE,EE+lowE only; dotted lines show the size of the contribution of intrinsic alignment terms to the dashed lines. Grey bands show the scales excluded from the DES analysis, in order to reduce sensitivity to nonlinear effects.

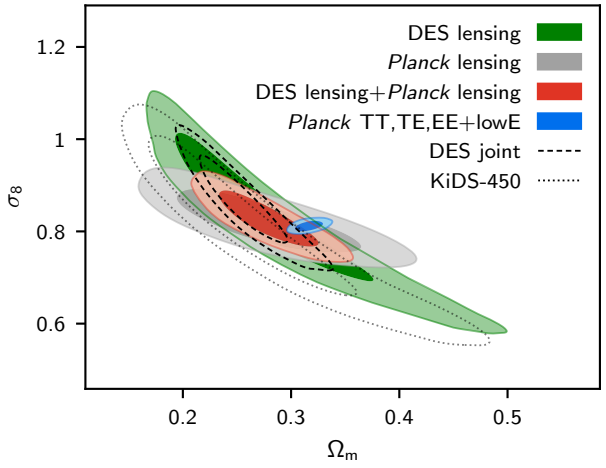


Figure 185: Base-ΛCDM model 68 % and 95 % constraint contours on the matter-density parameter Ω_m and fluctuation amplitude σ_8 from DES lensing [355, green], *Planck* CMB lensing (grey), and the joint lensing constraint (red). For comparison, the dashed line shows the constraint from the DES cosmic shear plus galaxy-clustering joint analysis [106], the dotted line the constraint from the original KiDS-450 analysis [174, without the corrections considered in Troxel et al. 356], and the blue filled contour shows the independent constraint from the *Planck* CMB power spectra.

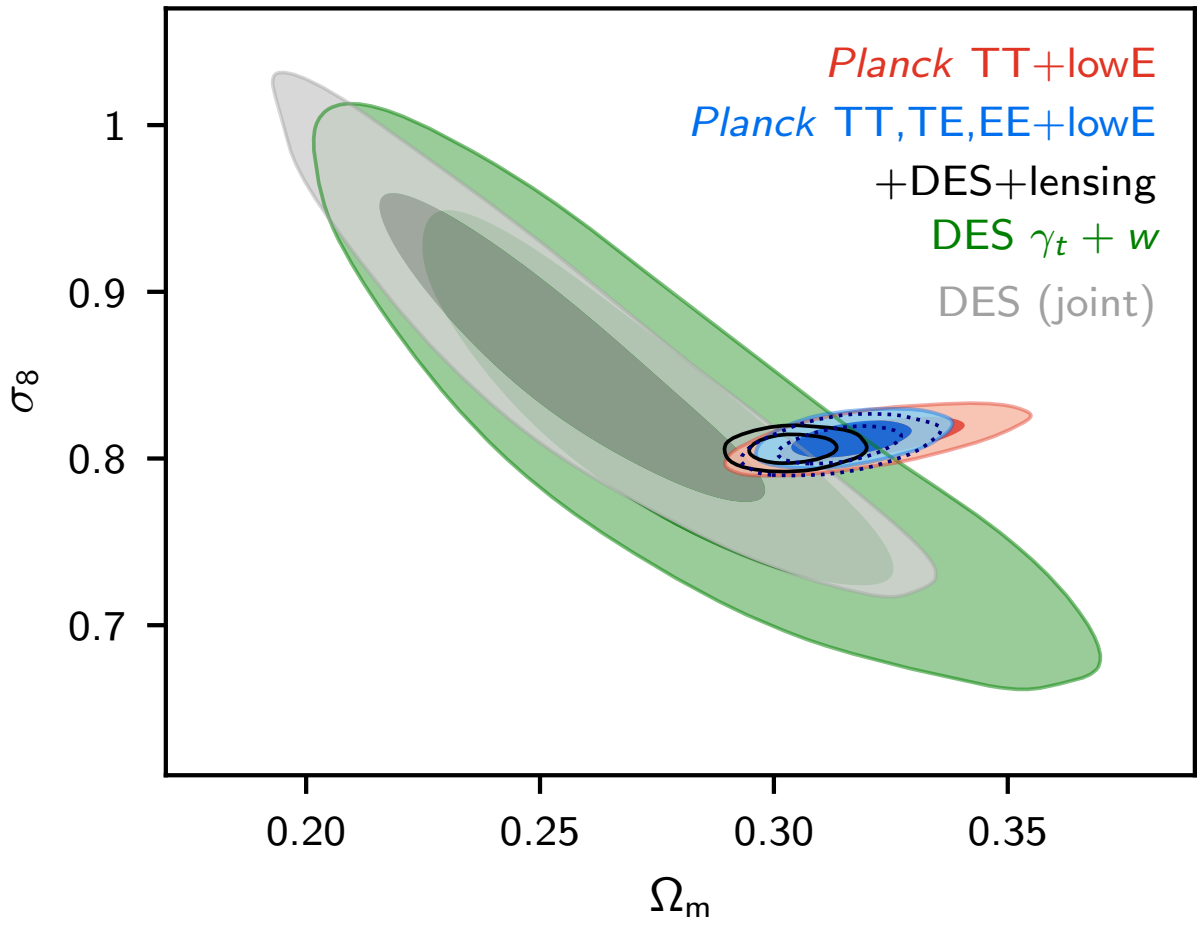


Figure 186: Base-ΛCDM model constraints from the Dark Energy Survey (DES), using the shear-galaxy correlation and the galaxy auto-correlation data (green) and the joint result with DES lensing (grey), compared with *Planck* results using TT+lowE and TT,TE,EE+lowE. The black solid contours show the joint constraint from *Planck* TT,TE,EE+lowE+lensing+DES, assuming the difference between the data sets is purely statistical. The dotted line shows the *Planck* TT,TE,EE+lowE result using the `CamSpec` likelihood, which is slightly more consistent with the DES contours than using the default `Plik` likelihood. Contours contain 68 % and 95 % of the probability.

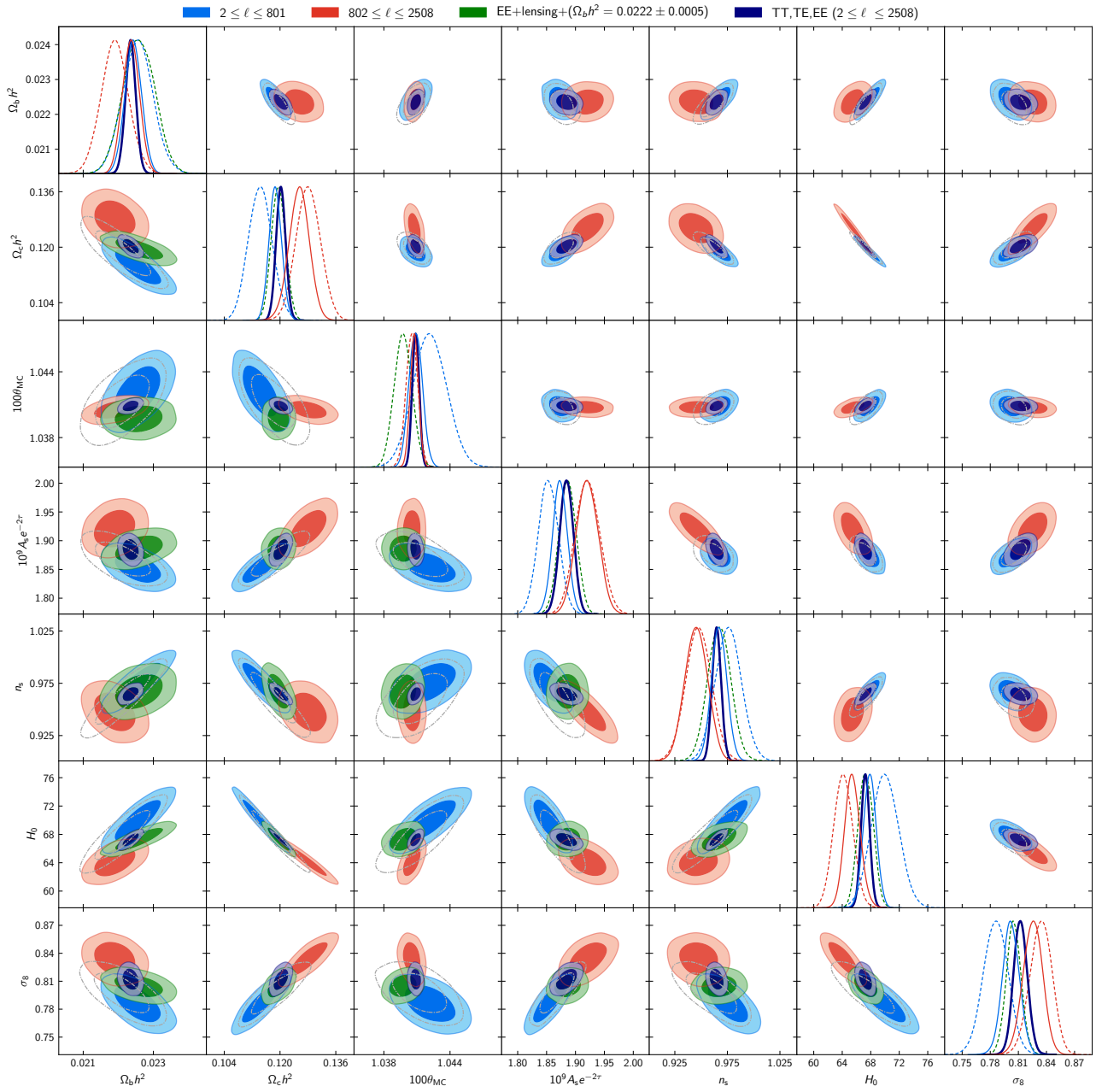


Figure 187: Base-ΛCDM 68 % and 95 % parameter constraint contours from the CMB power spectra using $\ell \leq 801$ (blue), compared to $\ell \geq 802$ (red). All results use the `plik_lite` *Planck* likelihood, and also include the low-ℓ `SimAll` "lowE" likelihood to constrain the optical depth τ; the `Commander` likelihood is used for temperature multipoles $\ell < 30$. The lower triangle contains the *Planck* temperature likelihoods, which show a moderate tension between high and low multipoles; however, they intersect in a region of parameter space consistent with the nearly-independent constraint from EE+lensing combined with a conservative prior $\Omega_{\rm b}h^2 = 0.0222 \pm 0.0005$, motivated by element-abundance observations (green). The upper triangle shows the equivalent results from *Planck* TT,TE,EE+lowE at low and high multipoles. The full combined result from *Planck* TT,TE,EE+lowE is shown as the navy contours. The unfilled grey contours show the result for multipoles $30 \leq \ell \leq 801$ (that is, removing the low-ℓ `Commander` likelihood that pulls parameters to give lower temperature power, due to the dip below $\ell \approx 30$). The diagonal plots are the marginalized parameter constraints, where results corresponding to the lower triangle are shown dashed, while the upper triangle are the solid curves.

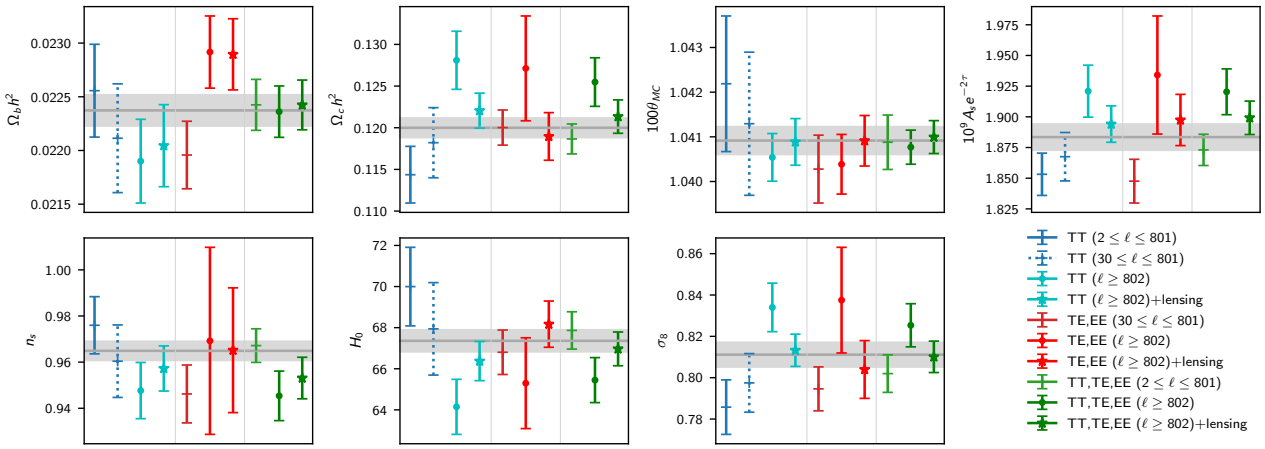


Figure 188: Base-ΛCDM 68 % marginalized parameter constraints for various combinations of power spectrum ranges (all using `plik_lite` and also including low-ℓ polarization "lowE"). Points marked with a cross are from $2 \le \ell \le 801$, while points marked with a circle are from $\ell \ge 802$. Dotted errors are the result from $30 \le \ell \le 801$, without the `Commander` large-scale temperature likelihood, showing that $\ell < 30$ pulls the low-multipole parameters further from the joint result. Points marked with a star are from $\ell \ge 802$ combined with the lensing likelihood, showing that constraining the lensing amplitude pulls all the results from high multipoles towards better consistency with the results from lower multipoles. The grey horizontal band shows the combined 68 % constraint from *Planck* TT,TE,EE+lowE+lensing.

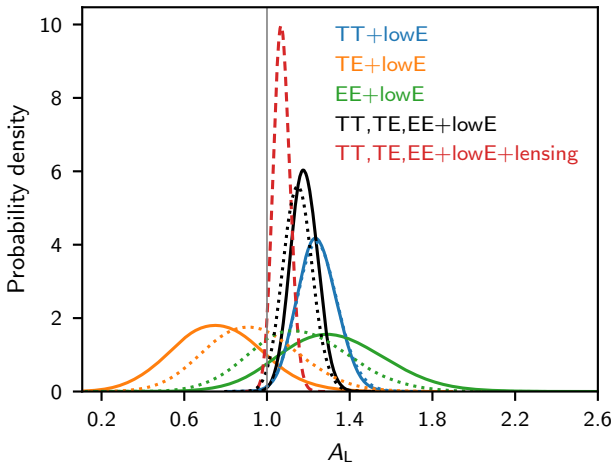


Figure 189: Constraints on the value of the consistency parameter A_L, as a single-parameter extension to the base-ΛCDM model, using various combinations of *Planck* data. When only power spectrum data are used, $A_L > 1$ is favoured at about $3\,\sigma$, but including the lensing reconstruction the result is consistent at $2\,\sigma$ with $A_L = 1$. The dotted lines show equivalent results for the `CamSpec` likelihood, which peak slightly nearer to $A_L = 1$, indicating some sensitivity of the A_L results to choices made in constructing the high-multipole likelihoods.

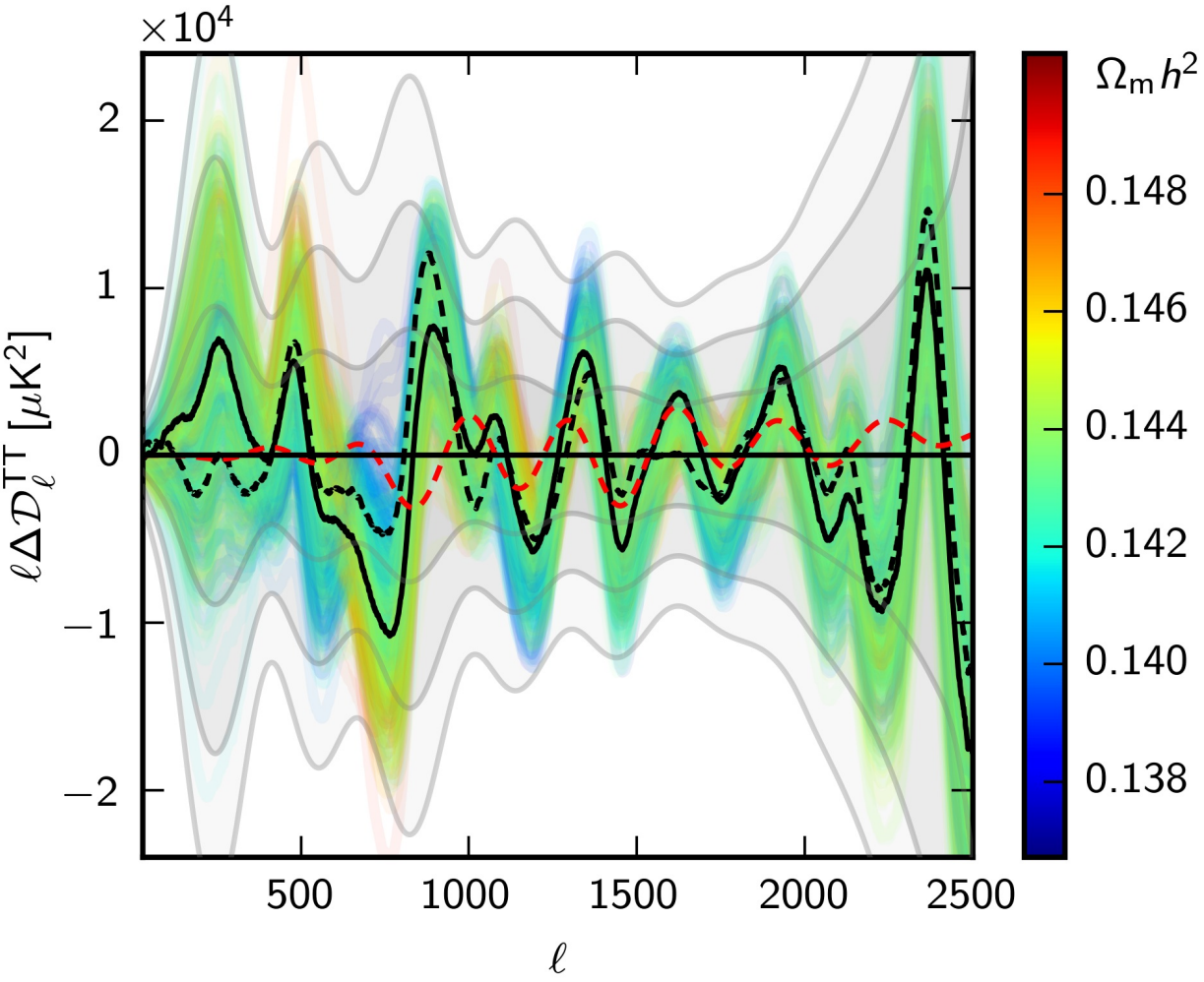


Figure 190: Base-ΛCDM model ($A_{\rm L} = 1$) TT power spectrum residuals smoothed with a Gaussian kernel of width $\sigma_\ell = 40$. The black line shows the smoothed difference between the coadded data points and the theoretical model for the *Planck* TT+lowE best-fit model, while coloured lines show the residuals for samples over the allowed parameter space coloured by the value of $\Omega_{\rm m}h^2$. Grey bands show the 1, 2, and 3 σ diagonal range expected for the smoothed residuals in the best-fit model. The red dashed line shows 10 % of the lensing-smoothing difference predicted in the best-fit model, displaying the oscillatory signal expected if there were more lensing of the acoustic peaks. The data residuals are not particularly anomalous, but the residuals have a similar pattern to the lensing smoothing difference over the approximate range $\ell = 1100$–2000, giving a preference for around 10 % more lensing at fixed cosmological parameters. Allowed models with lower $\Omega_{\rm m}h^2$ (and hence higher H_0) predict less lensing and give a larger oscillatory residual, preferring relatively more lensing smoothing than models with high matter density. The black dashed line shows the smoothed residual for the *Planck* TT +lowE best fit to ΛCDM+$A_{\rm L}$ (with $A_{\rm L} = 1.19$).

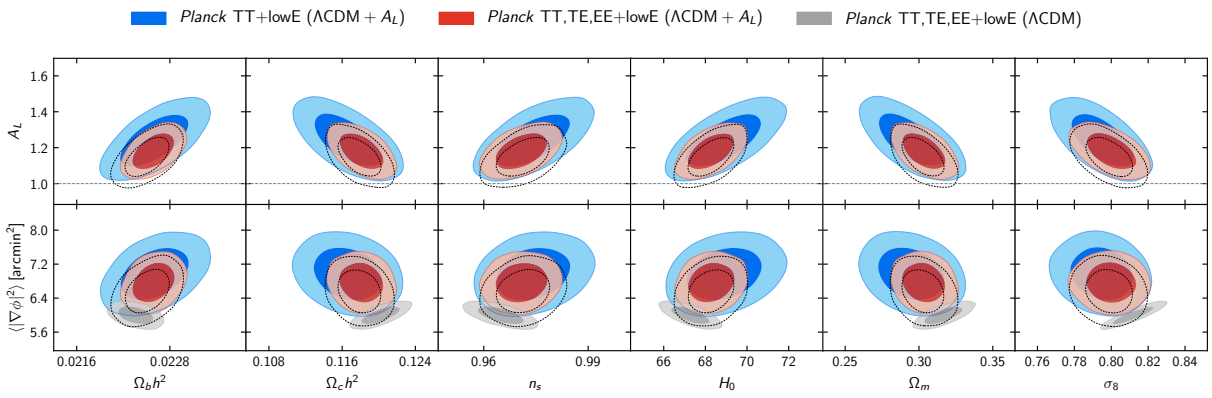


Figure 191: Marginalized 68 % and 95 % parameter constraint contours when adding $A_{\rm L}$ as a single-parameter extension to the base-ΛCDM model, with (red) and without (blue) small-scale polarization, compared to the constraints in the base-ΛCDM model (grey). The dashed contours show equivalent results for *Planck* TT,TE,EE+lowE when using the `CamSpec` likelihood, which gives results with $A_{\rm L}$ nearer unity and with slightly larger errors. The second row of subplots show, on the left axis, the predicted lensing deflection angle variance (from lensing multipoles $2 \leq L \leq 2000$), which is a measure of the amount of actual lensing: the TT,TE,EE+lowE likelihood prefers about 10 % more actual lensing power (associated with lensing smoothing), but in the (unphysical) varying-$A_{\rm L}$ case this can be achieved using cosmological parameters that predict less lensing than in ΛCDM but substantially larger $A_{\rm L}$, giving a preference for $A_{\rm L} \approx 1.2$.

28 Planck 2018 results. X. Constraints on inflation

We report on the implications for cosmic inflation of the 2018 release of the *Planck* cosmic microwave background (CMB) anisotropy measurements. The results are fully consistent with those reported using the data from the two previous *Planck* cosmological releases, but have smaller uncertainties thanks to improvements in the characterization of polarization at low and high multipoles. *Planck* temperature, polarization, and lensing data determine the spectral index of scalar perturbations to be $n_{\rm s} = 0.9649 \pm 0.0042$ at 68 % CL. We find no evidence for a scale dependence of $n_{\rm s}$, either as a running or as a running of the running. The Universe is found to be consistent with spatial flatness with a precision of 0.4 % at 95 % CL by combining *Planck* with a compilation of BAO data. The *Planck* 95 % CL upper limit on the tensor-to-scalar ratio, $r_{0.002} < 0.10$, is further tightened by combining with the BICEP2/Keck Array BK15 data to obtain $r_{0.002} < 0.056$. In the framework of standard single-field inflationary models with Einstein gravity, these results imply that: (a) the predictions of slow-roll models with a concave potential, $V''(\phi) < 0$, are increasingly favoured by the data; and (b) based on two different methods for reconstructing the inflaton potential, we find no evidence for dynamics beyond slow roll. Three different methods for the non-parametric reconstruction of the primordial power spectrum consistently confirm a pure power law in the range of comoving scales $0.005\,{\rm Mpc}^{-1} \lesssim k \lesssim 0.2\,{\rm Mpc}^{-1}$. A complementary analysis also finds no evidence for theoretically motivated parameterized features in the *Planck* power spectra. For the case of oscillatory features that are logarithmic or linear in k, this result is further strengthened by a new combined analysis including the *Planck* bispectrum data. The new *Planck* polarization data provide a stringent test of the adiabaticity of the initial conditions for the cosmological fluctuations. In correlated, mixed adiabatic and isocurvature models, the non-adiabatic contribution to the observed CMB temperature variance is constrained to 1.3 %, 1.7 %, and 1.7 % at 95 % CL for cold dark matter, neutrino density, and neutrino velocity, respectively. *Planck* power spectra plus lensing set constraints on the amplitude of compensated cold dark matter-baryon isocurvature perturbations that are consistent with current complementary measurements. The polarization data also provide improved constraints on inflationary models that predict a small statistically anisotropic quadupolar modulation of the primordial fluctuations. However, the polarization data do not support physical models for a scale-dependent dipolar modulation. All these findings support the key predictions of the standard single-field inflationary models, which will be further tested by future cosmological observations.

28.1 Introduction

[!tbp]

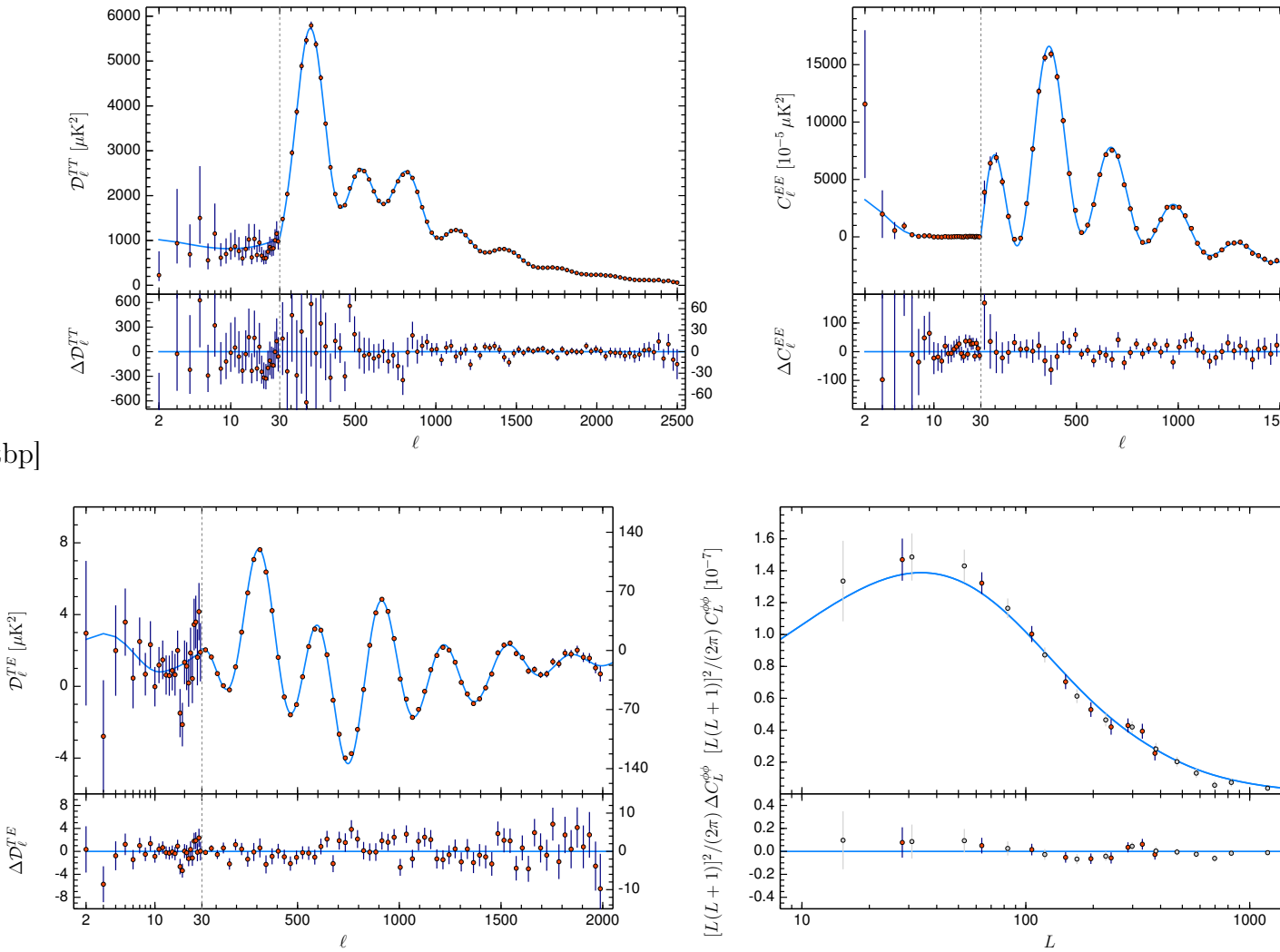


Figure 192: *Planck* 2018 CMB angular power spectra, compared with the base-ΛCDM best fit to the *Planck* TT,TE,EE+lowE+lensing data (blue curves). For each panel we also show the residuals with respect to this baseline best fit. Plotted are $\mathcal{D}_\ell = \ell(\ell+1)C_\ell/(2\pi)$ for TT and TE, C_ℓ for EE, and $L^2(L+1)^2C_L^{\phi\phi}/(2\pi)$ for lensing. For TT, TE, and EE, the multipole range $2 \leq \ell \leq 29$ shows the power spectra from `Commander` (TT) and `SimAll` (TE, EE), while at $\ell \geq 30$ we display the co-added frequency spectra computed from the `Plik` cross-half-mission likelihood, with foreground and other nuisance parameters fixed to their best-fit values in the base-ΛCDM cosmology. For the *Planck* lensing potential angular power spectrum, we show the conservative (orange dots; used in the likelihood) and aggressive (grey dots) cases. Note some of the different horizontal and vertical scales on either side of $\ell = 30$ for the temperature and polarization spectra and residuals.

This paper, one of a set associated with the 2018 release of data from the *Planck*[161] mission, presents the implications for cosmic inflation of the 2018 *Planck* measurements of the cosmic microwave background (CMB) anisotropies. In terms of data, this paper updates [279]

[161] *Planck* (`http://www.esa.int/Planck`) is a project of the European Space Agency (ESA) with instruments provided by two scientific consortia funded by ESA member states and led by Principal Investigators from France and Italy, telescope reflectors provided through a collaboration between ESA and a scientific consortium led and funded by Denmark, and additional contributions from NASA (USA).

(henceforth **(alias?)**), which was based on the temperature data of the nominal *Planck* mission ("PR1"), including the first 14 months of observations, and [288] (henceforth **(alias?)**), which used temperature data and a first set of polarization data from the full *Planck* mission ("PR2"), comprising 29 and 52 months of observations for the high frequency instrument (HFI) and low frequency instrument (LFI), respectively.

The ideas underlying cosmic inflation were developed during the late 1970s and early 1980s in order to remedy a number of defects of the hot big-bang cosmological model (e.g., the horizon, smoothness, flatness, and monopole problems) [44, 260, 153, 247, 119, 170, 15, 171]. Subsequently, it was realized that, on account of quantum vacuum fluctuations, cosmic inflation also provides a means to generate the primordial cosmological perturbations [205, 206, 130, 121, 261, 18, 204]. The development of cosmic inflation is one of the major success stories of modern cosmology, and in this paper we explore how the latest 2018 release of the *Planck* data constrains inflationary models.

Planck data currently provide the best constraints on the CMB anisotropies, except on very small angular scales beyond the resolution limit of *Planck* . The *Planck* data set has enabled a precision characterization of the primordial cosmological perturbations and has allowed cosmological parameters to be constrained at the sub-percent level. One of the main data products, described in more detail in the following section, is the *Planck* TT, TE, EE, and lensing power spectra, which are shown in Fig. 192, together with the residuals from the six-parameter concordance Λ cold dark matter (ΛCDM) model using the best-fit parameter values.

In order to provide a quantitative estimate of the improvement achieved by *Planck*, as well as to show where *Planck* stands compared to an ultimate cosmic-variance-limited survey, we consider an idealistic estimator for the number of modes [i.e., the effective number of $a_{\ell m}$'s measured [282]]:

$$N^{XY}_{\rm modes}(\ell) \equiv 2\sum_{\ell'=2}^{\ell}\left(\frac{C^{XY}_{\ell'}}{\Delta C^{XY}_{\ell'}}\right)^2 , \tag{1232}$$

where C^{XY}_ℓ (ΔC^{XY}_ℓ) is the (error on the) angular power spectrum of the XY channel [282, 249]. The number of modes measured by *Planck* is 1 430 000 and 109 000 for temperature ($XY = TT$ up to $\ell = 2500$) and polarization ($XY = EE$ up to $\ell = 2000$), respectively [230]. The number of modes measured is increased by approximately a factor of 7 (570) for temperature (polarization) with respect to the WMAP 9-year measurement, but there is still a factor of 3 (40) to gain for a cosmic-variance-limited experiment up to $\ell = 2500$ accessing 70 % of the sky. The additional modes measured by *Planck* play a key role in improving the constraints on the initial conditions for the cosmological perturbations and on models of inflation with respect to previous measurements of CMB anisotropies.

Planck data have also greatly improved the constraints on bispectral non-Gaussianity, both for the "local" pattern, as predicted by many inflationary models, and for other templates such as the equilateral one, as analysed and reported in detail in the dedicated *Planck* non-Gaussianity papers [280, 287, 295]. The constraints on the non-Gaussianity parameter $f_{\rm NL}$ are limited by a combination of cosmic variance and instrumental noise. An order-of-magnitude estimate for the signal-to-noise ratio for the local pattern (with $f^{\rm loc}_{\rm NL} = 1$) is given by

$$\left(\frac{S}{N}\right)^2 \propto \Omega_{\rm sky}\ell^2_{\rm max}\ln\left(\frac{\ell_{\rm max}}{\ell_{\rm min}}\right) . \tag{1233}$$

For the local shape, the logarithm enters because most of the signal derives from detecting the modulation of the small-scale power by the large-scale CMB anisotropy, highlighting the importance of full-sky maps for this kind of analysis. For other shapes such as equilateral, one instead has $(S/N)^2 \sim \Omega_{\rm sky}\ell^2_{\rm max}$. *Planck* has significantly sharpened the constraints on $f_{\rm NL}$, largely on account of its measurement of high multipoles with higher signal-to-noise ratio compared to past surveys. Some improvement has also been obtained from including polarization.

The *Planck* measurements have significantly constrained the physics of inflation. The hypothesis of adiabatic Gaussian scalar fluctuations with a power spectrum described by a simple power law, which is the key prediction of the standard single-field slow-roll inflationary models, has been tested to unprecedent accuracy (**(alias?)**, **(alias?)**;Planck Collaboration XXIV 280, Planck Collaboration XVII 287). *Planck* has set tight constraints on the amount of inflationary gravitational waves by exploiting the shape of the CMB temperature spectrum [**(alias?)**]. These results have inspired a resurgence of activity in inflationary model building.

For more details, see, for example, the following review articles and references therein: Linde [168]; Martin et al. [182]; Guth et al. [120]; and Burgess et al. [49]. *Planck* analysis and interpretation have also sparked a debate on the likelihood of initial conditions for some inflationary models [144, 143, 169], which is primarily of theoretical interest and is not addressed in this paper. In combination with more sensitive B-mode ground-based polarization measurements, as from BICEP-Keck Array [**(alias?)**], *Planck* has convincingly ruled out the slow-roll inflationary model with a quadratic potential [**(alias?)**]. In terms of physics beyond the simplest slow-roll inflationary models, the pre-*Planck* hints of a running spectral index [139] or of large non-Gaussianities [31] have disappeared as a result of the *Planck* measurements. How to interpret anomalies on the largest angular scales and at high multipoles is a question motivating the search for new non-standard inflationary models. We discuss how the *Planck* 2018 release data further test these ideas.

This paper is organized as follows. In Sect. 28.2 we describe the statistical methodology, the *Planck* likelihoods, and the complementary data sets used in this paper. In Sect. 28.3 we discuss the updated constraints on the spectral index of the scalar perturbations, on spatial curvature, and on the tensor-to-scalar ratio. Section 28.4 is devoted to constraining slow-roll parameters and to a Bayesian model comparison of inflationary models, taking into account the uncertainties in connecting the inflationary expansion to the subsequent big-bang thermalized era. In Sect. 28.5 the potential for standard single-field inflation is reconstructed using two different methodologies. Section 28.6 describes the primordial power spectrum reconstruction using three different approaches. In Sect. 28.7, the parametric search for features in the primordial scalar power spectrum is described, including a dedicated study of the axion monodromy model. In Sect. 28.8, the *Planck* power spectrum data are combined with information from the *Planck* bispectrum in a search for oscillations in the primordial spectra. The constraints on isocurvature modes are summarized in Sect. 28.9. Section 28.10 updates and extends the constraints on anisotropic inflationary models of inflation. We summarize our conclusions in Sect. 28.11, highlighting the key results and the legacy of *Planck* for inflation.

28.2 Methodology and data

The general theoretical background and analysis methods applied in this paper closely match those of the previous *Planck* inflation papers [**(alias?)**, **(alias?)**]. Consequently, in this section we provide only a brief summary of the methodology and focus on changes in the *Planck* likelihood relative to previous releases.

28.2.1 Cosmological models and inference

For well over a decade, the base-ΛCDM model has been established as the simplest viable cosmological model. Its six free parameters can be divided into primordial and late-time parameters. The former describe the state of perturbations on observable scales (corresponding to a wavenumber range of $10^{-4}\,\mathrm{Mpc}^{-1} \lesssim k \lesssim 10^{-1}\,\mathrm{Mpc}^{-1}$ today) prior to re-entering the Hubble radius around recombination. In base ΛCDM, the initial state of perturbations is assumed to be purely adiabatic and scalar, with the spectrum of curvature perturbations given by the power law

$$\ln \mathcal{P}_{\mathcal{R}}(k) = \ln A_{\mathrm{s}} + (n_{\mathrm{s}} - 1)\ln(k/k_*) \equiv \ln \mathcal{P}_0(k), \tag{1234}$$

where k_* denotes an arbitrary pivot scale. The late-time parameters, on the other hand, determine the linear evolution of perturbations after re-entering the Hubble radius. By default we use the basis $(\omega_{\mathrm{b}}, \omega_{\mathrm{c}}, \theta_{\mathrm{MC}}, \tau)$[162] for the late-time parameters, but occasionally also consider non-minimal late-time cosmologies. Because of the inflationary perspective of this paper, we are mainly interested in exploring modifications of the primordial sector and their interpretation in terms of the physics of the inflationary epoch.

Perturbations produced by generic single-field slow-roll models of inflation are typically well

[162] Refer to Table 27 for definitions.

Table 27: Baseline and optional late-time parameters, primordial power spectrum parameters, and slow-roll parameters. All primordial quantities are evaluated at a pivot scale of $k_* = 0.05\,\mathrm{Mpc}^{-1}$, unless otherwise stated.

Parameter	**Definition**
$\omega_\mathrm{b} \equiv \Omega_\mathrm{b} h^2$	Baryon density today
$\omega_\mathrm{c} \equiv \Omega_\mathrm{c} h^2$	Cold dark matter density today
θ_MC	Approximation to the angular size of sound horizon at last scattering
τ	Optical depth to reionization
N_eff	Effective number of neutrino species
Σm_ν	Sum of neutrino masses
Ω_K	Spatial curvature parameter
w_0	Dark energy equation of state parameter
A_s	Scalar power spectrum amplitude
n_s	Scalar spectral index
$\mathrm{d}n_\mathrm{s}/\mathrm{d}\ln k$	Running of scalar spectral index
$\mathrm{d}^2 n_\mathrm{s}/\mathrm{d}(\ln k)^2$	Running of running of scalar spectral index
r	Tensor-to-scalar power ratio
n_t	Tensor spectral index
$\epsilon_1 = -\dot{H}/H^2$	First Hubble slow-roll parameter
$\epsilon_{n+1} = \dot{\epsilon}_n/(H\epsilon_n)$	$(n+1)$st Hubble slow-roll parameter $(n \geq 1)$
$\epsilon_V = M_\mathrm{pl}^2 V_\phi^2/(2V^2)$	First potential slow-roll parameter, where $_\phi \equiv \mathrm{d}/\mathrm{d}\phi$
$\eta_V = M_\mathrm{pl}^2 V_{\phi\phi}/V$	Second potential slow-roll parameter
$\xi_V^2 = M_\mathrm{pl}^4 V_\phi V_{\phi\phi\phi}/V^2$	Third potential slow-roll parameter
$\varpi_V^3 = M_\mathrm{pl}^6 V_\phi^2 V_{\phi\phi\phi\phi}/V^3$	Fourth potential slow-roll parameter

approximated by the following form of the adiabatic scalar and tensor components:

$$\ln \mathcal{P}_\mathcal{R}(k) = \ln \mathcal{P}_0(k) + \frac{1}{2}\frac{d\ln n_\mathrm{s}}{d\ln k}\ln(k/k_*)^2 + \frac{1}{6}\frac{d^2\ln n_\mathrm{s}}{d\ln k^2}\ln(k/k_*)^3 + \ldots, \tag{1235}$$

$$\ln \mathcal{P}_\mathrm{t}(k) = \ln(rA_\mathrm{s}) + n_\mathrm{t}\ln(k/k_*) + \ldots, \tag{1236}$$

which allows for a weak scale dependence of the scalar spectral index, n_s, modelled by a running, $d\ln n_\mathrm{s}/d\ln k$, or a running of the running, $d^2\ln n_\mathrm{s}/d(\ln k)^2$.[163] The power spectrum parameterization in Eq. 1235 can be extended to address wider classes of inflation-related questions, (e.g., the search for isocurvature perturbations, specific primordial features in the spectra, etc.), as described in subsequent sections. We also go beyond simple functions to parameterize the primordial power spectrum. In the spirit of reconstructing the primordial spectrum from the data, we consider some general parameterizations (e.g., taking the power spectrum as an interpolation between knots of freely varied amplitudes at fixed or varying wavenumbers).

One could argue that the primordial power spectra are merely intermediate quantities and assess theories directly from more fundamental parameters. By using the slow-roll approximation, or by evolving the mode equations to obtain exact numerical predictions for the spectra without resorting to the slow-roll approximation, we can relate the primordial perturbations to the dynamics of the Hubble parameter during inflation or to the inflaton potential and its derivatives, thus constraining these quantities directly.

For any given model, theoretical predictions of CMB-related and other cosmological observables are calculated using appropriately modified versions of the Boltzmann codes `CAMB` [166] or `CLASS` [56]. As in [**(alias?)**, **(alias?)**], we compare models $\mathcal{M}_1$ and $\mathcal{M}_2$ by the difference in the logarithm of the likelihood of their best fits, or *effective* $\Delta\chi^2 \equiv 2\,[\ln \mathcal{L}_\mathrm{max}(\mathcal{M}_1) - \ln \mathcal{L}_\mathrm{max}(\mathcal{M}_2)]$.

[163] Unless explicitly stated otherwise, we adopt a default pivot scale $k_* = 0.05\,\mathrm{Mpc}^{-1}$ in this work. As in previous *Planck* releases, we will also quote the tensor-to-scalar ratio $r_{0.002}$ at $k_* = 0.002\,\mathrm{Mpc}^{-1}$ in order to facilitate comparison with earlier primordial tensor-mode constraints.

We apply Bayesian statistical methods to infer the posterior probability distributions of the model parameters and select between competing models [270], using either the Metropolis-Hastings Markov-chain Monte Carlo (MCMC) sampling algorithm, as implemented in `CosmoMC` [164] and `Monte Python` [30], or software based on nested sampling [253], such as `MultiNest` [88, 89] or `PolyChord` [126, 127]. The latter can simultaneously estimate the Bayesian evidence $\mathcal{E}_i$ of a model $\mathcal{M}_i$, allowing the comparison between different models via the Bayes factor, $B = \mathcal{E}_2/\mathcal{E}_1$, where $|\ln B| > 5$ is commonly considered "strong" evidence for or against the respective model [147, 268].

28.2.2 Data

Planck data

The *Planck* data processing has improved in a number of key aspects with respect to the previous 2015 cosmological release. We briefly summarize the main points here, referring the interested reader to [290] and [291] for details.

The flagging procedure in the LFI 2018 pipeline has been made more aggressive, in particular for the first 200 operational days. However, the most important improvement in the LFI pipeline is in the calibration approach. Whereas in the 2015 release, the main calibration source for LFI was the *Planck* orbital dipole (i.e., the amplitude modulation of the CMB dipole induced by the satellite orbit) of each single radiometer model, the 2018 procedure also includes the Galactic emission along with the orbital dipole in the calibration model and becomes iterative [290].

The HFI data for the 2018 release have also been made more conservative, cutting approximately 22 days of observations under non-stationary conditions with respect to 2015. The main change in the HFI data processing is the use of a new map making and calibration algorithm called `SRoll`, whose first version was introduced in [297] for the initial analysis of HFI polarization on large angular scales. This algorithm employs a generalized polarization destriper which uses the redundancy in the data to extract several instrumental systematic-effect parameters directly from the data [291].

These improvements have a minor impact on *Planck* temperature maps, but are much more important for polarization, particularly on large angular scales, allowing, for instance, the removal of the high-pass filtering adopted in the 2015 study of isotropy and statistics [226].

In the following, we summarize the essentials of the *Planck* inputs used in this paper (i.e., the *Planck* likelihood approach to the information contained in the 2-point statistics of the temperature and polarization maps and the *Planck* CMB lensing likelihood). As for previous cosmological releases, the *Planck* likelihood approach is hybridized between low- and high-multipole regions, which therefore are summarized separately below. We refer the interested reader to the relevant papers Planck Collaboration V [293] (henceforth **(alias?)**) and Planck Collaboration VIII [294] (henceforth **(alias?)**) for a more complete description of these inputs.

Planck low-ℓ likelihood

As in the *Planck* 2015 release, several options are available for evaluating the temperature likelihood on large angular scales, each with its own computational complexity and approximations. One option is based on the `Commander` framework and implements full Bayesian sampling of an explicit parametric model that includes both the cosmological CMB signal and non-cosmological astrophysical signals, such as thermal dust, CO, and low-frequency foregrounds. This framework is described in earlier papers [see Planck Collaboration XI [283] and Planck Collaboration Int. XLVI [297] and references therein for details]. The only changes since the 2015 implementation concern the data and model selection. As described in Planck Collaboration IV [292], we only use the *Planck* 2018 data in the current data release, whereas the previous 2015 version additionally included WMAP [33] and Haslam [129] observations. With fewer frequencies available, this requires a simpler model, and in particular we now fit for only a single low-frequency foreground component, rather than individual synchrotron, free-free, and spinning dust emission components, and we only fit a single CO component, rather than for individual CO line components at 100, 217, and 353 GHz. On the one hand, this results in fewer internal foreground degeneracies compared to the 2015 version, and a likelihood that only depends on *Planck* data, but at the same time the simpler foreground modelling also requires a slightly larger Galactic mask. Overall, the two versions are very compatible in terms

of the recovered CMB power spectra, as discussed in **(alias?)**. For additional details on the `Commander` temperature analysis, see Planck Collaboration IV [292].

The HFI low-ℓ polarization likelihood is based on the full-mission HFI 100-GHz and 143-GHz Q and U low-resolution maps cleaned through a template-fitting procedure with LFI 30 GHz and HFI 353 GHz information[164] used as tracers of polarized synchrotron and thermal dust, respectively (see **(alias?)** for details about the cleaning procedure). The likelihood method, called `SimAll`, represents a follow-up of the `SimBaL` algorithm presented in Planck Collaboration Int. XLVI [297] and uses the FFP10 simulations to construct empirically the probability for the EE and BB spectra. The method is based on the quadratic maximum likelihood estimation of the cross-spectrum between 100 and 143 GHz, and its multipole range spans from $\ell = 2$ to $\ell = 29$. We only built the likelihood for EE and BB and not for TE, due to the poor statistical consistency of the TE spectrum for $\ell > 10$, and due to the difficulty of describing accurately the correlation with TT and EE, given the limited number of simulations available; see discussions in section 2.2.6 of **(alias?)**. Further details about the method and consistency tests are presented in **(alias?)**. When combined with the low-ℓ temperature likelihood (based on the `Commander` CMB solution), the low-ℓ polarization likelihood implies $\tau = 0.051 \pm 0.009$ and $r_{0.002} < 0.41$ at 95 % CL.

As an alternative to the `Commander` and `SimAll` low-ℓ likelihood, an update of the joint temperature and polarization pixel-based low-ℓ LFI likelihood used in 2015 is part of this *Planck* data release. Its methodology (see **(alias?)** for details) is similar to that of 2015 [283], i.e., a pixel-based approach in TQU at $N_{\rm side} = 16$, and employs the `Commander` solution in temperature along with the LFI 70-GHz linear polarization maps, foreground cleaned using the *Planck* 30-GHz and 353-GHz channels as tracer templates for synchrotron and dust, respectively. This 2018 version allows for a larger sky fraction in polarization (66.4 %, compared to the previous 46 %) and retains the sky surveys 2 and 4 that were excluded in 2015. By performing a two-parameter estimate for $A_{\rm s}$ and τ restricted to $\ell < 30$, we find using this likelihood that $\tau = 0.063 \pm 0.020$ and $\ln(10^{10} A_{\rm s}) = 2.975 \pm 0.056$ at 68 % CL. The latter value has been derived by varying the TT, EE, and TE CMB spectra.

Planck high-ℓ likelihood

The 2018 baseline high-ℓ likelihood (`Plik`) is an update of the 2015 baseline version. The `CamSpec` likelihood is also used to explore alternative data cuts and modelling of the data and is described below. Both approaches implement a Gaussian likelihood approximation using cross-spectra between the 100-, 143-, and 217-GHz maps. `Plik` covers the multipoles $30 \leq \ell \leq 2509$ in temperature and $30 \leq \ell \leq 1997$ in polarization (i.e., for TE and EE). In order to avoid noise bias, the high-ℓ likelihood relies only on half-mission map cross-spectra, which have been demonstrated to be largely free of correlated noise. The spectra are computed on masked maps in order to reduce the anisotropic Galactic contamination (dominated by dust emission), and in the case of TT also strong point sources and CO emission. The `Plik` masks, identical to the 2015 masks, are tailored to each frequency channel and differ in temperature and polarization to take into account differing foreground behaviour and channel beams. The `Plik` intensity (polarization) masks effectively retain 66, 57, and 47 % (70, 50, and 41 %) of the sky after apodization for the 100, 143, and 217 GHz channels, respectively. Unlike in 2015, when the map beams were computed for an average sky fraction, they are now computed for the exact sky fraction used at each frequency. The data vector used in the likelihood approximation discards multipoles that are highly contaminated by foregrounds or have low signal-to-noise ratios.

The `Plik` power spectra are binned using the same scheme as in 2015. Unbinned likelihoods are also available. When forming the data vector, individual cross-frequency spectra are not co-added. This allows for independent exploration of the calibration, nuisance, and foreground parameter space for each cross-spectrum using dedicated templates in the theory vector.

The `Plik` (and `CamSpec`) covariance matrices are computed for a fixed fiducial CMB including the latest estimate of the foreground and systematics, which are all assumed to be Gaussian. As verified in 2015, after the masks have been applied this is a reasonable assumption. The covariance matrix computation uses an approximation to account for mask-induced correla-

[164] The polarized synchrotron component is fitted only at 100 GHz, being negligible at 143 GHz. For the polarized dust component, following the prescription in Planck Collaboration III [291], the low-ℓ HFI polarization likelihood used the 353-GHz map constructed only from polarization-sensitive bolometers.

tions. Plik uses only the large Galactic mask in the analytic computation and then takes extra correlations due to the point-source mask into account using a Monte Carlo estimate of the extra variance induced. Missing pixels are ignored in the covariance. In 2015 its was shown that this approach (i.e., Gaussian approximation and approximate covariance) induced only a less than 0.1σ bias on $n_{\rm s}$ (from the $30 < \ell < 100$ modes).

The Plik noise model has been re-estimated on the latest HFI maps using the same methodology as in 2015, based on a comparison between noise-biased auto-spectra and cross-spectra. This procedure avoids correlated glitch residuals, which had biased previous noise estimates [283], particularly in polarization at $\ell \lesssim 500$.

The 2018 HFI data processing pipeline has refined the maps used in the likelihood relative to 2015. For example, an improved destriping procedure reduced the residual scatter in the polarization maps, in particular at 143 GHz (yielding about 12 % lower noise on the half-mission cross-spectrum). More stringent selection cuts resulted in the discarding of the last 1000 rings of data, increasing the noise in the temperature half-mission spectrum by about 3 %. Also, a higher threshold was imposed on the conditioning of the TQU intra-pixel noise covariance matrix for a pixel to be considered well-measured, resulting in more missing pixels relative to 2015.

The data modelling has also significantly improved, in particular for polarization, making cosmological constraints from polarization more reliable. In 2015, the polarization spectra (TE and EE) displayed relatively large inter-frequency disagreements. A plausible explanation (at least for TE) was the temperature-to-polarization leakage induced by beam and calibration differences (so-called "beam leakage"). The beam-leakage modelling has improved substantially in 2018 [137] so that we can now propagate the beams, gain differences, polarization angles, etc. to compute a reliable template for the beam leakage and thus remove these leakage effects. These improvements substantially reduce the TE inter-frequency disagreements.

We also reassessed the estimates of the polarization efficiency for the polarized channels. Comparing different data-based estimates demonstrates that the ground-based polarization efficiency uncertainty estimates (of the order of a fraction of a percent) were too optimistic by a factor of 5 to 10. Correcting for the observed polarization efficiency errors (at the percent level) very significantly reduces the EE inter-frequency disagreements. This calibration correction relies on cosmological priors (using the TT best-fit cosmology). Calibrating using either the TE or the EE spectra yields generally consistent results, except at 143 Ghz where there is disagreement at more than 2σ. At this level, this discrepancy can be caused either by a statistical fluctuation, or by an unknown residual.

The Plik baseline likelihood implements a map-based calibration. The TE calibration is deduced from the TT and EE calibrations, including at 143 GHz. Other improvements over the 2015 version are the following. The dust model has been improved in temperature and polarization, using also the latest version of the 353-GHz maps. The level of synchrotron contamination in the 100-GHz and 143-GHz maps has been estimated and shown to be negligible. Sub-pixel noise has been included in TT and EE (and demonstrated to have a negligible effect on the cosmological parameters). Finally, a correlated component of the noise has been observed in the end-to-end HFI simulation, affecting the large scales and very small scales of the EE auto-frequency spectra. The large-scale contribution affects the dust correction and the $n_{\rm s}$ constraints. We constructed an empirical model of this correlated noise from our simulations, which is included in the Plik likelihood.

CamSpec was the baseline for the 2013 release and was described in detail in Planck Collaboration XV [275], and used cross-spectra formed from detector-set temperature maps using data from the nominal mission period. It was extended for the 2015 release to include both polarization and temperature-polarization cross-spectra and to use the data from the full mission period. Similarly to Plik, CamSpec switched from detector-set cross-spectra to cross-spectra formed from frequency maps constructed from separate halves of the full mission data in order to mitigate the effects of noise correlated between detectors. In 2015, the foreground modelling was also modified and the sky fraction retained at each frequency was increased, using common masks with Plik in temperature. CamSpec used a more conservative mask in polarization than Plik.

Differently from Plik, CamSpec corrects each TE and EE cross-frequency spectrum with a fixed dust and temperature-to-polarization leakage template before co-adding them to form the EE and TE components of its data vector and bases its noise estimate on differences between maps constructed using alternating pointing periods. Note also that CamSpec uses an

individual-spectrum-based calibration scheme, where the TE calibrations are not fixed to be those inferred from the TT and EE ones.

In the 2018 release further improvements in the `CamSpec` foreground modelling have been implemented. The dust model in temperature has been updated in a way similar to `Plik`. `CamSpec` now uses a richer model of the cosmic infrared background, allowing for the exploration of any impact on the cosmological parameters. As explained above, the noise modelling was also modified. Further modifications of the masking have been made for polarization, still using the same masks for each frequency channel, but different from the `Plik` mask. As we discussed above, beam-leakage and polarization-efficiency corrections are applied to the individual polarization spectra before their addition for inclusion in the likelihood. More details on the `Plik` and `CamSpec` likelihoods can be found in **(alias?)**.

As in 2015, the high-ℓ `Plik` and `CamSpec` likelihoods are in excellent agreement for temperature. The different assumptions for the polarization-efficiency parameters and the masks [283] propagate into differences in cosmological parameter estimates. For the baseline ΛCDM model, the difference in cosmological parameters between the `Plik` likelihood and the `CamSpec` likelihood (using joint TT,TE,EE in combination with `Commander`, `SimAll`, and lensing) is at most 0.5σ (for $\Omega_b h^2$) (Planck Collaboration XIII 284, henceforth **(alias?)**). Similar differences in cosmological parameters occur in extended cosmological models. The differences between the `Plik` and `CamSpec` parameters are dominated by calibration-model differences for the joint TT,TE,EE and TE-only cases, and by mask differences for the EE-only case. To a large extent, the `CamSpec` results can be reproduced within the `Plik` framework simply by changing in `Plik` the calibration model (for TE) and the polarization mask (for EE). Below we use `Plik` as the *Planck* baseline high-ℓ likelihood. `CamSpec` results are used to assess the residual uncertainty from modelling and mask choices. We quote values obtained with `CamSpec` only for a few cases.

We use the following conventions for naming the *Planck* likelihoods: (i) *Planck* TT+lowE denotes the combination of the high-ℓ TT likelihood at multipoles $\ell \geq 30$ and the low-ℓ temperature-only `Commander` likelihood, plus the `SimAll` low-ℓ EE-only likelihood in the range $2 \leq \ell \leq 29$; (ii) *Planck* TE and *Planck* EE denote the TE and EE likelihood at $\ell \geq 30$, respectively; (iii) *Planck* TT,TE,EE+lowE denotes the combination of the combined likelihood using TT, TE, and EE spectra at $\ell \geq 30$, the low-ℓ temperature `Commander` likelihood, and the low-ℓ `SimAll` EE likelihood; and (iv) *Planck* TT,TE,EE+lowP denotes the combination of the likelihood using TT, TE, and EE spectra at $\ell \geq 30$ and the alternative joint temperature-polarization likelihood at $2 \leq \ell \leq 29$, based on the temperature `Commander` map and the 70-GHz polarization map. Unless otherwise stated, high-ℓ results are based on the `Plik` likelihood and low-ℓ polarization information is based on `SimAll`.

Planck CMB lensing likelihood

The *Planck* 2018 lensing likelihood, presented in **(alias?)**, uses the lensing trispectrum to estimate the power spectrum of the lensing potential $C_L^{\phi\phi}$. This signal is extracted using a minimum-variance combination of a full set of temperature- and polarization-based quadratic lensing estimators [210] applied to the `SMICA` CMB map over approximately 70 % of the sky using CMB multipoles $100 \leq \ell \leq 2048$, as described in **(alias?)**. We use the lensing bandpower likelihood, with bins spanning lensing multipoles $8 \leq L \leq 400$, which has been validated with numerous consistency tests. Because its multipole range has been extended down to $L = 8$ (compared to $L = 40$ for the *Planck* 2015 analysis), the statistical power of the lensing likelihood used here is slightly greater.

Non-*Planck* data

While the data derived exclusively from *Planck* observations are by themselves already extremely powerful at constraining cosmology, external data sets can provide helpful additional information. The question of consistency between *Planck* and external data sets is discussed in detail in Planck Collaboration VI [235] (henceforth **(alias?)**). Here we focus on two data sets that are particularly useful for breaking degeneracies and whose errors can be assessed reliably. We consider the measurement of the CMB B-mode polarization angular power spectrum by the BICEP2/Keck Array collaboration and measurements of the baryon acoustic oscillation

(BAO) scale. The supplementary B-mode data provide independent constraints on the tensor sector, which are better than those that can be derived from the *Planck* data alone (based on the shape of the scalar power spectrum). The BAO data, on the other hand, do not directly constrain the primordial perturbations. These data, however, provide invaluable low-redshift information that better constrain the late-time cosmology, especially in extensions of ΛCDM, and thus allow degeneracies to be broken.

BICEP2/Keck Array 2015 B-mode polarization data

Although *Planck* measured the CMB polarization over the full sky, its polarization sensitivity in the cosmological frequency channels is not sufficient to compete with current suborbital experiments surveying small, particularly low-foreground patches of the sky very deeply using many detectors. In **(alias?)**, constraints on r using the joint BICEP2/Keck Array and *Planck* (BKP) analysis [51] were reported. Here we make use of the most recent B-mode polarization data available from the analysis of the BICEP2/Keck field (Ade et al. 11, henceforth **(alias?)**), unless otherwise stated. The BK15 likelihood draws on data from the new Keck array at 220 GHz in addition to those already in use for the **(alias?)** [10] likelihood, i.e., the 95 and 150-GHz channel, as well as from *Planck* and WMAP to remove foreground contamination. The BK15 observations measure B-mode polarization using 12 auto- and 56 cross-spectra between the BICEP2/Keck maps at 95, 150, and 220 GHz, the WMAP maps at 23 and 33 GHz, and the *Planck* maps at 30, 44, 70, 100, 143, 217, and 353 GHz, using nine bins in multipole number. By using B-mode information only within the BK15 likelihood, a 95 % upper limit of $r < 0.07$ is found [**(alias?)**], which improves on the corresponding 95 % CL $r < 0.09$ [**(alias?)**] based on the BK14 likelihood.

Baryon acoustic oscillations

Acoustic oscillations of the baryon-photon fluid prior to recombination are responsible for the acoustic peak structure of the CMB angular power spectra. The counterpart to the CMB acoustic peaks in the baryon distribution are the BAOs, which remain imprinted into the matter distribution to this day. In the position-space picture, the BAOs of the power spectrum correspond to a peak in the correlation function, defining a characteristic, cosmology-dependent length scale that serves as a standard ruler and can be extracted (e.g., from galaxy redshift surveys). The transverse information of a survey constrains the ratio of the comoving angular diameter distance and the sound horizon at the drag epoch (i.e., when the baryon evolution becomes unaffected by coupling to the photons), D_M/r_d, whereas the line-of-sight information yields a measurement of $H(z)r_\mathrm{d}$. Sometimes, these two observables are combined to form the direction-averaged quantity $D_V/r_\mathrm{d} \equiv \left[czD_M^2(z)H^{-1}(z)\right]^{1/3}/r_\mathrm{d}$.

For our BAO data compilation, we use the measurements of D_V/r_d from the 6dF survey at an effective redshift $z_\mathrm{eff} = 0.106$ [47], and the SDSS Main Galaxy Sample at $z_\mathrm{eff} = 0.15$ [313], plus the final interpretation of the SDSS III DR12 data [15], with separate constraints on $H(z)r_\mathrm{d}$ and D_M/r_d in three correlated redshift bins at $z_\mathrm{eff} = 0.38$, 0.51, and 0.61. In [9], the same set of BAO data combined with either non-*Planck* CMB data or measurements of the primordial deuterium fraction was shown to favour a cosmology fully consistent with, but independent of, *Planck* data.

28.3 *Planck* 2018 results for the main inflationary observables

As in **(alias?)** and **(alias?)**, we start by describing *Planck* measurements of the key inflationary parameters. Some of the results reported in this section can be found in the Planck Legacy Archive.[165]

Parameter	TT+lowE	EE+lowE	TE+lowE	TT,TE,EE+lowE	TT,TE,EE
$\Omega_b h^2$	0.02212 ± 0.00022	0.0240 ± 0.0012	0.02249 ± 0.00025	0.02236 ± 0.00015	0.0223
$\Omega_c h^2$	0.1206 ± 0.0021	0.1158 ± 0.0046	0.1177 ± 0.0020	0.1202 ± 0.0014	0.120
$100\theta_{MC}$	1.04077 ± 0.00047	1.03999 ± 0.00089	1.04139 ± 0.00049	1.04090 ± 0.00031	1.0409
τ	0.0522 ± 0.0080	0.0527 ± 0.0090	0.0496 ± 0.0085	$0.0544^{+0.0070}_{-0.0081}$	0.054
$\ln(10^{10} A_s)$	3.040 ± 0.016	3.052 ± 0.022	$3.018^{+0.0020}_{-0.0018}$	3.045 ± 0.016	3.04
n_s	0.9626 ± 0.0057	0.980 ± 0.015	0.967 ± 0.011	0.9649 ± 0.0044	0.964
H_0	66.88 ± 0.92	69.9 ± 2.7	68.44 ± 0.91	67.27 ± 0.60	67.3
Ω_m	0.321 ± 0.013	$0.289^{+0.026}_{-0.033}$	0.301 ± 0.012	0.3166 ± 0.0084	0.315
σ_8	0.8118 ± 0.0089	0.796 ± 0.018	0.793 ± 0.011	0.8120 ± 0.0073	0.811

Table 28: Confidence limits for the cosmological parameters in the base-ΛCDM model from *Planck* temperature, polarization, and temperature-polarization cross-correlation separately and combined, in combination with the EE measurement at low multipoles.

28.3.1 Results for the scalar spectral index

Planck temperature data in combination with the *EE* measurement at low multipoles determine the scalar spectral tilt in the ΛCDM model as

$$n_s = 0.9626 \pm 0.0057 \quad (68\,\%\ \text{CL},\ \textit{Planck}\ \text{TT+lowE})\,. \tag{1237}$$

This result for n_s is compatible with the *Planck* 2015 68 % CL value $n_s = 0.9655 \pm 0.0062$ for *Planck* TT+lowP [**(alias?)**]. The slightly lower value for n_s is mainly driven by a corresponding shift in the average optical depth τ, now determined as

$$\tau = 0.052 \pm 0.008 \quad (68\,\%\ \text{CL},\ \textit{Planck}\ \text{TT+lowE})\,, \tag{1238}$$

which is to be compared with the *Planck* 2015 value $\tau = 0.078 \pm 0.022$ [**(alias?)**]. This more precise determination of τ is due to better noise sensitivity of the HFI 100- and 143-GHz channels employed in the low-ℓ `SimAll` polarization likelihood, compared to the joint temperature-polarization likelihood based on the LFI 70-GHz channel in 2015. Because of the degeneracy between the average optical depth and the amplitude of the primordial power spectrum, A_s and σ_8 are also lower than in the *Planck* 2015 release. These shifts from the *Planck* 2015 values for the cosmological parameters have been anticipated with the first results from the HFI large-angular polarization pattern [297, 298].

The trend toward smaller values for (n_s,τ) with respect to the *Planck* 2015 release also occurs for different choices for the low-ℓ likelihood. By substituting `Commander` and `SimAll` with the updated joint temperature-polarization pixel likelihood coming from the LFI 70-GHz channel, we obtain in combination with high-ℓ temperature data:

$$n_s = 0.9650 \pm 0.0061 \quad (68\,\%\ \text{CL},\ \textit{Planck}\ \text{TT+lowP})\,; \tag{1239}$$

$$\tau = 0.072 \pm 0.016 \quad (68\,\%\ \text{CL},\ \textit{Planck}\ \text{TT+lowP})\,. \tag{1240}$$

Although with larger errors, these latter results are consistent with the shifts induced by a determination of a lower optical depth than in the *Planck* 2015 release,[166] as found in Eqs. (1237) and (1238). Given this broad agreement and the consistency in the values of τ derived with `SimAll` separately from the three cross-spectra 70 × 100, 70 × 143, and 100 × 143 [**(alias?)**], we will mainly use the baseline low-ℓ likelihood in the rest of the paper.

As anticipated in 2015, the information in the high-ℓ polarization *Planck* data is powerful for breaking degeneracies in the parameters and to further decrease parameter uncertainties

[165] http://www.cosmos.esa.int/web/planck/pla

[166] As in 2015, the combination with high-ℓ data pulls τ to larger values than the low-ℓ pixel likelihood alone, i.e. $\tau = 0.063 \pm 0.020$ at 68 % CL; see Sect. 28.2. This effect is less pronounced for the `SimAll` likelihood.

compared to temperature data alone. The addition of high-ℓ polarization leads to a tighter constraint on $n_{\rm s}$:

$$n_{\rm s} = 0.9649 \pm 0.0044 \quad (68\,\%\ \text{CL},\ \textit{Planck}\ \text{TT,TE,EE+lowE}). \tag{1241}$$

This is in good agreement with the *Planck* 2015 TT,TE,EE+lowP 68 % CL result, $n_{\rm s} = 0.9645 \pm 0.0049$. In this 2018 release the mean value of $n_{\rm s}$ is approximately 0.5σ larger than the temperature result in Eq. (1237). This pull is mainly due to a higher value for the scalar tilt preferred by *Planck* 2018 polarization and temperature-polarization cross-correlation data only:

$$n_{\rm s} = 0.969 \pm 0.009 \quad (68\,\%\ \text{CL},\ \textit{Planck}\ \text{TE,EE+lowE})\,. \tag{1242}$$

This pull is then mitigated in combination with temperature due to the larger uncertainty in the determination by TE,EE only. Similar considerations hold for the alternative `CamSpec` high-ℓ likelihood, which leads to a 68 % CL result $n_{\rm s} = 0.9658 \pm 0.0045$, consistent with the baseline `Plik` reported in Eq. (1241). Overall, the cosmological parameters from *Planck* baseline temperature, polarization, and temperature-polarization cross-correlation separately and combined are very consistent, as can be seen from Table 28 and Fig. 193 for the ΛCDM model.

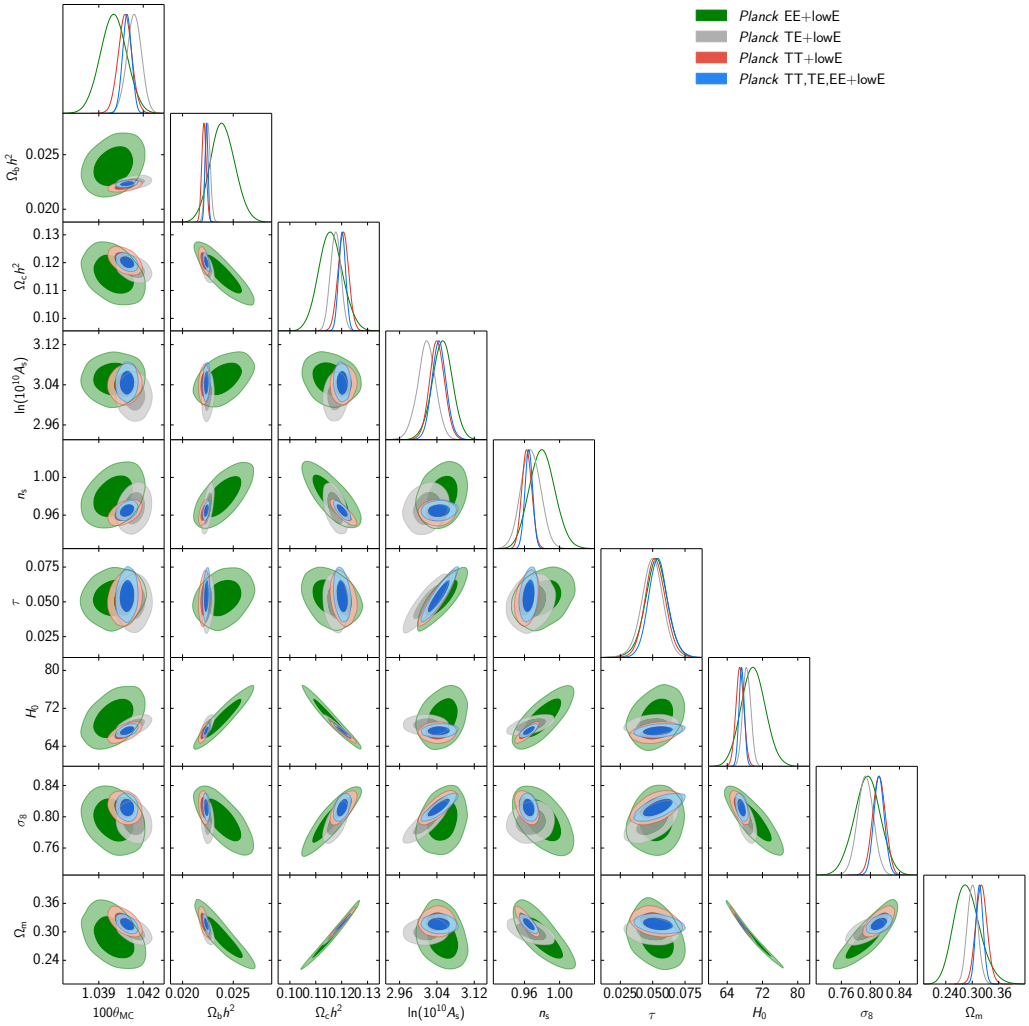


Figure 193: Marginalized joint 68 % and 95 % CL regions for the cosmological parameters in ΛCDM with *Planck* TT, EE, TE, and joint TT,TE,EE, all in combination with the EE likelihood at low multipoles.

After combining with *Planck* lensing, we obtain

$$n_s = 0.9634 \pm 0.0048 \quad (68\,\%\ \text{CL},\ \textit{Planck}\ \text{TT+lowE+lensing}), \tag{1243}$$

$$n_s = 0.9649 \pm 0.0042 \quad (68\,\%\ \text{CL},\ \textit{Planck}\ \text{TT,TE,EE+lowE+lensing}). \tag{1244}$$

The shift in n_s (and, more generally, in the cosmological parameters of the base-ΛCDM model) obtained when *Planck* lensing is combined with TT,TE,EE+lowE is smaller than in 2015 because of the improved polarization likelihoods. The combination with lensing is, however, powerful for breaking parameter degeneracies in extended cosmological models, and, therefore, for this 2018 release we will consider the full information contained in temperature, polarization, and lensing, i.e., TT,TE,EE+lowE+lensing, as the baseline *Planck* data set. Figure 194 shows a comparison of the *Planck* 2018 baseline results with those from alternative likelihoods and from the 2015 baseline for the ΛCDM cosmological parameters.

As in 2013 and 2015, BAO measurements from galaxy surveys are consistent with *Planck*. When BAO data are combined, we obtain for the base-ΛCDM cosmology:

$$n_s = 0.9665 \pm 0.0038 \quad (68\,\%\ \text{CL},\ \textit{Planck}\ \text{TT,TE,EE+lowE+lensing+BAO})\,. \tag{1245}$$

The combination with BAO data decreases (increases) the marginalized value of $\Omega_c h^2$ ($\Omega_b h^2$) obtained by *Planck*, and this effect is compensated for by a shift in n_s towards slightly larger values.

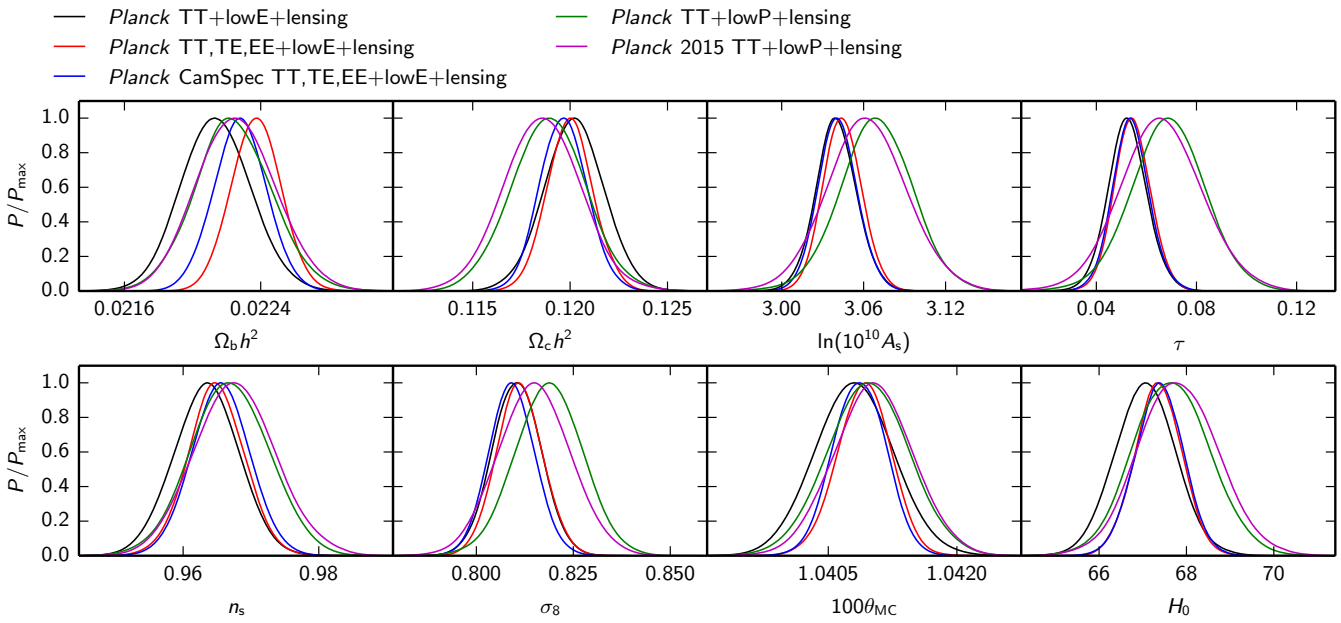


Figure 194: Comparison of the marginalized probability density of the primary parameters and $(\sigma_8\,, H_0)$ for the baseline cosmological model from *Planck* TT+lowE+lensing (black curves), TT,TE,EE+lowE+lensing (red curves), and the alternative likelihood `Camspec`. For comparison we also display the *Planck* 2018 TT+lowP+lensing (blue curves) and the corresponding *Planck* 2015 TT+lowP+lensing (green curves) results.

28.3.2 Ruling out $n_s = 1$

One of the main findings drawn from previous *Planck* releases was that the scale-independent Harrison-Zeldovich (HZ) spectrum [128, 282, 214] is decisively ruled out. This conclusion is reinforced in this release: in standard ΛCDM late-time cosmology, the scalar spectral index from Table 28 lies 6.6, 8.0, and 8.4σ away from $n_s = 1$, for *Planck* TT+lowE, *Planck* TT,TE,EE+lowE,

and *Planck* TT,TE,EE+lowE+lensing, respectively. The corresponding effective $\Delta\chi^2$ between the power-law spectrum and the best-fit HZ model are $\Delta\chi^2 = 43.9$, 66.9, and 72.4.

Simple one-parameter modifications of the cosmological model are not sufficient to reconcile a scale-invariant power spectrum with *Planck* data. For instance, when the effective number of neutrino species $N_{\rm eff}$ is allowed to float for a cosmology with a scale-invariant spectrum, the effective $\Delta\chi^2$ with respect to the power-law spectrum are $\Delta\chi^2 = 12.9$, 27.5, and 30.2, respectively.

When instead the assumption of flat spatial sections is relaxed,[167] we obtain effective $\Delta\chi^2$ values of $\Delta\chi^2 = 11.8$, 28.8, and 40.9, respectively, for the same data sets. Therefore, the corresponding closed cosmological models fitting *Planck* TT+lowE ($\Omega_K = -0.122^{+0.039}_{-0.029}$, $H_0 = 44.2^{+3.1}_{-4.3}\,{\rm km\,s^{-1}\,Mpc^{-1}}$ at 68 % CL), *Planck* TT,TE,EE+lowE ($\Omega_K = -0.095^{+0.029}_{-0.019}$, $H_0 = 47.1 \pm 3.2\,{\rm km\,s^{-1}\,Mpc^{-1}}$ at 68 % CL), and *Planck* TT,TE,EE+lowE+lensing ($\Omega_K = -0.032^{+0.006}_{-0.007}$, $H_0 = 58.9 \pm 2.0\,{\rm km\,s^{-1}\,Mpc^{-1}}$ at 68 % CL) provide a worse fit compared to the tilted flat ΛCDM model.[168]

28.3.3 Constraints on the scale dependence of the scalar spectral index

The *Planck* 2018 data are consistent with a vanishing running of the scalar spectral index. Using *Planck* TT,TE,EE+lowE+lensing we obtain

$$\frac{dn_{\rm s}}{d\ln k} = -0.0045 \pm 0.0067 \quad (68\,\%\ {\rm CL})\,. \tag{1246}$$

These results are consistent with, and improve on, the *Planck* 2015 result, $dn_{\rm s}/d\ln k = -0.008 \pm 0.008$ [**(alias?)**].

As discussed in **(alias?)** and **(alias?)**, a better fit to the temperature low-ℓ deficit was found in 2015, thanks to a combination of non-negative values for the running and the running of the running. The *Planck* 2018 release has significantly reduced the parameter volume of this extension of the base-ΛCDM model. The *Planck* 2018 TT(TT,TE,EE)+lowE+lensing constraints for the model including running of running are

$$n_{\rm s} = 0.9587 \pm 0.0056 \ (0.9625 \pm 0.0048)\,, \tag{1247}$$

$$dn_{\rm s}/d\ln k = 0.013 \pm 0.012 \ (0.002 \pm 0.010)\,, \tag{1248}$$

$$d^2n_{\rm s}/d\ln k^2 = 0.022 \pm 0.012 \ (0.010 \pm 0.013)\,, \tag{1249}$$

all at 68 % CL. It is interesting to note that the high-ℓ temperature data still allow a sizable value for the running of the running, although slightly decreased with respect to the *Planck* 2015 results [**(alias?)**]. However, when high-ℓ *Planck* 2018 polarization data are also included, $dn_{\rm s}/d\ln k$ and $d^2n_{\rm s}/d\ln k^2$ are tightly constrained.

The model including a scale-dependent running can produce a better fit to the low-ℓ deficit at the cost of an increase of power at small scales; this latter effect is constrained in this release. As an example of a model with suppression only on large scales, we also reconsider the phenomenological model with an exponential cutoff:

$$\mathcal{P}_{\mathcal{R}}(k) = \mathcal{P}_0(k)\left\{1 - \exp\left[-\left(\frac{k}{k_{\rm c}}\right)^{\lambda_{\rm c}}\right]\right\}, \tag{1250}$$

which can be motivated by a short stage of inflation [63, 62] (see also Kuhnel & Schwarz [158], Hazra et al. [131], and Gruppuso et al. [118] for other types of large-scale suppressions). We

[167] For non-flat models the power spectra encode the eigenvalues of the corresponding Laplacian operator of the spatial sections and scale invariance holds at scales much smaller than the curvature radius.

[168] This is not a new result based on the *Planck* 2018 release, but just an update of a similar conclusion also reached with the *Planck* 2015 data. Compared to the flat ΛCDM tilted model, we obtain $\Delta\chi^2 = 12.3$, 34.8, and 45 with *Planck* 2015 TT+lowP, *Planck* 2015 TT,TE,EE+ lowP, and *Planck* 2015 TT,TE,EE+lowP+lensing, respectively. Therefore, even with *Planck* 2015 data, a closed model with $n_{\rm s} = 1$ provides a worse fit than tilted ΛCDM and is not compelling as claimed in [211].

do not find any statistically significant detection of k_c using either logarithmic or linear priors and for different values of λ_c, with any combination of *Planck* baseline likelihoods. We have also checked that these results depend only weakly on the exclusion of the EE quadrupole in `SimAll` and are stable to the substitution of `Commander` and `SimAll` with the joint temperature-polarization likelihood based on the 70-GHz channel.

28.3.4 Constraints on spatial curvature

Cosmological model ΛCDM+r	Parameter	*Planck* TT,TE,EE +lowEB+lensing	*Planck* TT,TE,EE +lowE+lensing+BK15	*Planck* TT,TE,EE +lowE+lensing+BK15
	r	< 0.11	< 0.061	< 0.063
	$r_{0.002}$	< 0.10	< 0.056	< 0.058
	n_s	0.9659 ± 0.0041	0.9651 ± 0.0041	0.9668 ± 0.0037
	r	< 0.16	< 0.067	< 0.068
	$r_{0.002}$	< 0.16	< 0.065	< 0.066
$+dn_s/d\ln k$	n_s	0.9647 ± 0.0044	0.9639 ± 0.0044	0.9658 ± 0.0040
	$dn_s/d\ln k$	-0.0085 ± 0.0073	-0.0069 ± 0.0069	-0.0066 ± 0.0070
	r	< 0.092	< 0.061	< 0.064
	$r_{0.002}$	< 0.085	< 0.055	< 0.059
$+N_{\rm eff}$	n_s	$0.9607^{+0.0086}_{-0.0084}$	0.9604 ± 0.0085	0.9660 ± 0.0070
	$N_{\rm eff}$	2.92 ± 0.19	2.93 ± 0.19	3.02 ± 0.17
	r	< 0.097	< 0.061	< 0.061
	$r_{0.002}$	< 0.091	< 0.056	< 0.056
$+m_\nu$	n_s	0.9654 ± 0.0044	0.9649 ± 0.0044	0.9668 ± 0.0036
	$\sum m_\nu$ [eV]	< 0.24	< 0.23	< 0.11
	r	< 0.12	< 0.066	< 0.062
	$r_{0.002}$	< 0.12	< 0.062	< 0.057
$+\Omega_K$	n_s	$0.9703^{+0.0045}_{-0.0046}$	0.9697 ± 0.0046	0.9663 ± 0.0044
	Ω_K	$-0.012^{+0.007}_{-0.006}$	$-0.012^{+0.006}_{-0.007}$	0.0006 ± 0.0019
	r	< 0.11	< 0.064	< 0.062
	$r_{0.002}$	< 0.10	< 0.059	< 0.057
$+w_0$	n_s	0.9675 ± 0.0042	0.9669 ± 0.0042	0.9659 ± 0.0040
	w_0	$-1.58^{+0.14}_{-0.34}$	$-1.58^{+0.14}_{-0.34}$	-1.04 ± 0.05

Table 29: Constraints on the tensor-to-scalar ratio r and scalar tilt n_s for the ΛCDM+r model and some important extensions and different data sets. For each model we quote 68 % confidence limits on measured parameters and 95 % upper bounds on other parameters.

Since the vast majority of inflation models predict that the Universe has been driven towards spatial flatness, constraints on the spatial curvature provide an important test of the standard scenario. Therefore in this subsection we extend the base-ΛCDM model with the addition of the spatial curvature parameter, Ω_K. For the case of *Planck* TT,TE,EE+lowE+lensing, we find a constraint of

$$\Omega_K = -0.011^{+0.013}_{-0.012} \quad (95\,\%\ \mathrm{CL})\,. \tag{1251}$$

The inclusion of *Planck* lensing information only weakly breaks the geometrical degeneracy [76] which results in the same primary fluctuations while varying the total matter density parameter, Λ, and H_0. The degeneracy can be effectively broken with the addition of BAO data, in which case *Planck* TT,TE,EE+lowE+lensing+BAO gives

$$\Omega_K = 0.0007 \pm 0.0037 \quad (95\,\%\ \mathrm{CL})\,. \tag{1252}$$

Although Ω_K is one of the cosmological parameters exhibiting some differences between `Plik` and `Camspec`, the constraints in Eqs. (1251) and (1252) are quite robust due to the inclusion of lensing (and BAO) information.

A constraint on the curvature parameter can be translated into a constraint on the radius of curvature, R_K, via

$$R_K = \left(a_0 H_0 \sqrt{|\Omega_K|}\right)^{-1}, \tag{1253}$$

in units such that $c = 1$. For the case of *Planck* TT,TE,EE+lowE+lensing+BAO we find

$$R_K > 67\,\mathrm{Gpc} \quad \text{(open)}, \tag{1254}$$

$$R_K > 81\,\mathrm{Gpc} \quad \text{(closed)}, \tag{1255}$$

both at 95 % confidence. These lengths are considerably greater than our current (post-inflation) particle horizon, at 13.9 Gpc.

Our tightest constraint, Eq. (1252), tells us that our observations are consistent with spatial flatness, with a precision of about 0.4 %. However, even if inflation has driven the background curvature extremely close to zero, the presence of fluctuations implies a fundamental "cosmic variance" for measurements of curvature confined to our observable volume. In particular, the known amplitude of fluctuations implies a standard deviation for Ω_K of roughly 2×10^{-5} [280]. Therefore our best constraint is still a factor of roughly 10^2 above the cosmic variance limit for a flat universe. A future measurement of negative curvature above the cosmic variance floor would point to open inflation [112, 113, 46, 281, 242, 175], while a measurement of positive curvature could pose a problem for the inflationary paradigm due to the difficulty of producing closed inflationary models [155].

Alternatively, excess spatial curvature might be evidence for the intriguing possibility that there was "just enough" inflation to produce structure on the largest observable scales. Indeed an upper limit on spatial curvature implies a lower limit on the total number of e-folds of inflation (see, e.g., Komatsu et al. 157). We can relate these limits to the number of e-folds of inflation, $N_* = N(k_*)$, after scale k_* left the Hubble radius during inflation, to be given explicitly in Eq. (1278). We define the (constant) curvature scale, k_K, as the inverse of the comoving radius of curvature, i.e.,

$$k_K \equiv aH\sqrt{|\Omega_K|}. \tag{1256}$$

In the absence of special initial conditions, inflation will begin with a curvature parameter of order unity. Equation (1256) then implies that $k_K \sim aH$ at the start of inflation, i.e., the curvature scale is "exiting the horizon" at that time. Then the lower limit on the number of e-folds of inflation will simply be $N_K \equiv N(k_K)$, i.e., the number of e-folds after scale k_K left the Hubble radius during inflation. With Eq. (1278) this gives[169]

$$N_K = \ln\frac{k_*}{a_0 H_0} - \frac{1}{2}\ln|\Omega_K| + N_*. \tag{1257}$$

With the pivot scale of $k_* = 0.002\,\mathrm{Mpc}^{-1}$ (for comparison with the values in Sect. 28.4.2) and our tightest upper limit on Ω_K from Eq. (1252), this becomes

$$N_K \gtrsim 4.9 + N_*. \tag{1258}$$

That is, our constraint on spatial curvature implies that inflation must have lasted at least about 5 e-folds longer than required to produce the pivot scale k_*. Equation (1258) provides a model-independent comparison between the e-folds required to solve the flatness problem (to current precision) and to produce large-scale fluctuations (at scale k_*). We stress that this limit assumes a unity curvature parameter at the start of inflation (although the dependence on this assumption, being logarithmic, is weak).

For comparison with the result of Komatsu et al. [157], we can simplify to the case of instantaneous thermalization and constant energy density during inflation. Then we find

$$N_K \gtrsim 34.2 + \ln\frac{T_{\mathrm{end}}}{1\,\mathrm{TeV}}, \tag{1259}$$

where T_{end} is the reheating temperature.

[169] This expression ignores a negligible correction, $(1/2)\ln V_{k_K}/V_{k_*}$, due to the different inflationary potential scales at k_* and k_K exit.

28.3.5 Constraints on the tensor-to-scalar ratio

This subsection updates constraints on the tensor-to-scalar ratio r assuming that the tensor tilt satisfies the consistency relation, $n_{\rm t} = -r/8$, which is the case for slow-roll inflation driven by a single scalar field with a canonical kinetic term.

By combining *Planck* temperature, low-ℓ polarization, and lensing we obtain

$$r_{0.002} < 0.10 \quad (95\,\%\ \text{CL},\ \textit{Planck}\ \text{TT+lowE+lensing})\,. \tag{1260}$$

This constraint slightly improves on the corresponding *Planck* 2015 95 % CL bound, i.e., $r_{0.002} < 0.11$ [**(alias?)**], and is unchanged when high-ℓ polarization data are also combined. Note that by using `CAMspec` instead of `Plik` as the high-ℓ joint temperature-polarization likelihood, we obtain a slightly looser bound, i.e., $r_{0.002} < 0.14$ at 95 % CL. By including the *Planck* B-mode information at $2 < \ell < 30$ in the low-ℓ polarization likelihood, the 95 % CL constraint is essentially unchanged.

Since inflationary gravitational waves contribute to CMB temperature anisotropies mostly at $\ell \lesssim 100$, the low-ℓ temperature deficit contributes in a nontrivial way to the *Planck* bound on r. By excising the $2 \leq \ell \leq 29$ temperature data, the constraint on r with *Planck* TT,TE,EE+lensing+lowEB relaxes to

$$r_{0.002} < 0.16 \quad (95\,\%\ \text{CL})\,. \tag{1261}$$

This result improves on the 2015 95 % CL result, i.e., $r \lesssim 0.24$ [**(alias?)**], because of the inclusion of high-ℓ polarization and of the improved determination of τ.

Since this *Planck* constraint on r relies on temperature and E-mode polarization, the *Planck*-only limit depends somewhat on the underlying cosmological model. Table 29 shows the constraints on $n_{\rm s}$ and r for a few important extensions of ΛCDM plus tensors, which include a non-zero running, a non-zero spatial curvature, and a non-minimal neutrino sector. We observe that the bound on r is relaxed by at most 30 % when the scale dependence of the scalar tilt is allowed to vary. In all the other extensions the *Planck* r bound is modified at most by 10 %, demonstrating the constraining power of the *Planck* 2018 release in reducing the degeneracy of the tensor-to-scalar ratio with other cosmological parameters. As far as the scalar tilt is concerned, we find the largest shift (by roughly 1σ higher) when the assumption of spatial flatness is relaxed.

A B-mode polarization measurement can further tighten the constraint on r and help in reducing its degeneracies with other cosmological parameters that may appear when using only temperature and E-mode polarization data. After the release of the first BICEP-Keck Array-*Planck* [**(alias?)**] joint cross-correlation, constraints on r from B-mode polarization data alone have become tighter than those based on *Planck* data alone, thanks to the inclusion of the 95-GHz channel [**(alias?)**] and of the 220-GHz channel [**(alias?)**]. By combining the *Planck* 2018 and BK15 data we obtain

$$\begin{aligned} r_{0.002} < 0.056 \quad & (95\,\%\ \text{CL},\ \textit{Planck}\ \text{TT,TE,EE} \\ & +\text{lowE+lensing+BK15})\,. \end{aligned} \tag{1262}$$

This bound improves on the corresponding one obtained in combination with **(alias?)**, i.e., $r_{0.002} < 0.064$ at 95 % CL. Note that by using `CAMspec` instead of `Plik` as high-ℓ TT,TE,EE likelihood, we obtain a slightly looser bound, i.e., $r_{0.002} < 0.069$ at 95 % CL. The effectiveness of the combination with the **(alias?)** likelihood in constraining r is also remarkable in the extensions of ΛCDM plus tensors, as can be seen from Table 29. By further combining with BAO the limits for r are only slightly modified.

The *Planck* 2018 baseline plus BK15 constraint on r is equivalent to an upper bound on the energy scale of inflation when the pivot scale exits the Hubble radius of

$$V_* = \frac{3\pi^2 A_{\rm s}}{2}\, r\, M_{\rm Pl}^4 < (1.6 \times 10^{16}\ \text{GeV})^4 \quad (95\,\%\ \text{CL})\,. \tag{1263}$$

Equivalently, this last result implies an upper bound on the Hubble parameter during inflation of

$$\frac{H_*}{M_{\rm Pl}} < 2.5 \times 10^{-5} \quad (95\,\%\ \text{CL})\,. \tag{1264}$$

28.3.6 Beyond the tensor-to-scalar ratio consistency condition

The increasing constraining power of B-mode polarization data allows us to set upper bounds on r without imposing the consistency condition for the tensor tilt, $n_t = -r/8$, which is motivated by standard slow-roll single-scalar-field inflation. Deviations can occur in multifield inflation [21, 279, 52], in the models with generalized Lagrangians [103, 156], in gauge inflation [178], or in a more radical way in alternative models to inflation [104, 41, 43].

As the current data do not lead to a detection of a non-zero tensor amplitude, virtually any value of n_t would give a good fit as long as r is close enough to zero. Therefore, as in **(alias?)**, we characterize the tensor perturbations by two well-constrained parameters that we choose to be r at two different scales, (r_{k_1}, r_{k_2}), with $k_1 = 0.002\,\mathrm{Mpc}^{-1}$ and $k_2 = 0.02\,\mathrm{Mpc}^{-1}$, and assume a power-law power spectrum. We call this two-parameter extension of the ΛCDM model the "ΛCDM+$r_{0.002}$+$r_{0.02}$" model. We also quote our results in terms of $(r_{\tilde{k}}, n_t)$, calculated from the primary parameters as $n_t = [\ln(r_{k_2}/r_{k_1}) \,/\, \ln(k_2/k_1)] + n_s - 1$ and $r_{\tilde{k}} = r_{k_1}(\tilde{k}/k_1)^{n_t - n_s + 1}$. For $\tilde{k}$ we choose $0.01\,\mathrm{Mpc}^{-1}$, which corresponds roughly to the decorrelation scale of r and n_t when using the *Planck* and BK15 data.

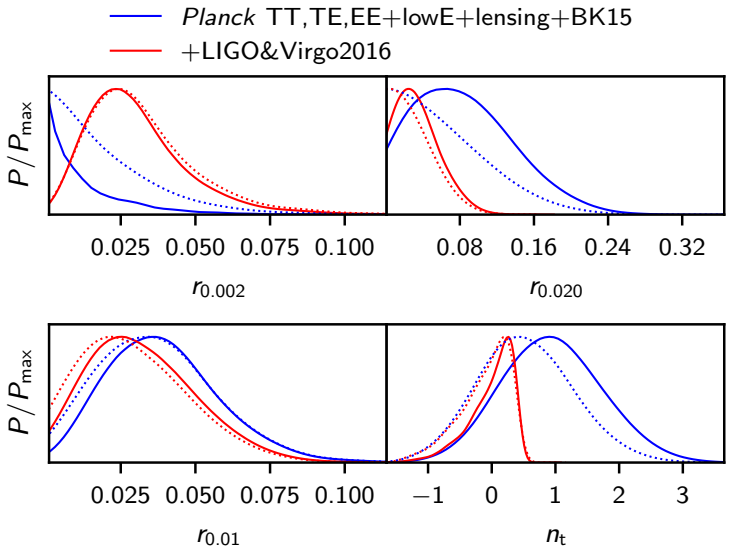


Figure 195: Posterior probability density of the tensor-to-scalar ratio at two different scales in the ΛCDM+$r_{0.002}$+$r_{0.02}$ model, i.e., when the inflationary consistency relation is relaxed (top panels). The solid contours show the results when $r_{0.002}$ and $r_{0.02}$ are used as sampling parameters with uniform priors, which leads to non-uniform priors for the derived parameters $r_{0.01}$ and n_t (bottom panels). The dotted contours indicate the results after weighting the posterior by the Jacobian $J = r_{0.01}/[r_{0.002} r_{0.02} \ln(0.02/0.002)]$ of the transformation $(r_{0.002}, r_{0.02}) \to (r_{0.01}, n_t)$, giving the result we would have obtained had we assigned uniform priors on $r_{0.01}$ and n_t.

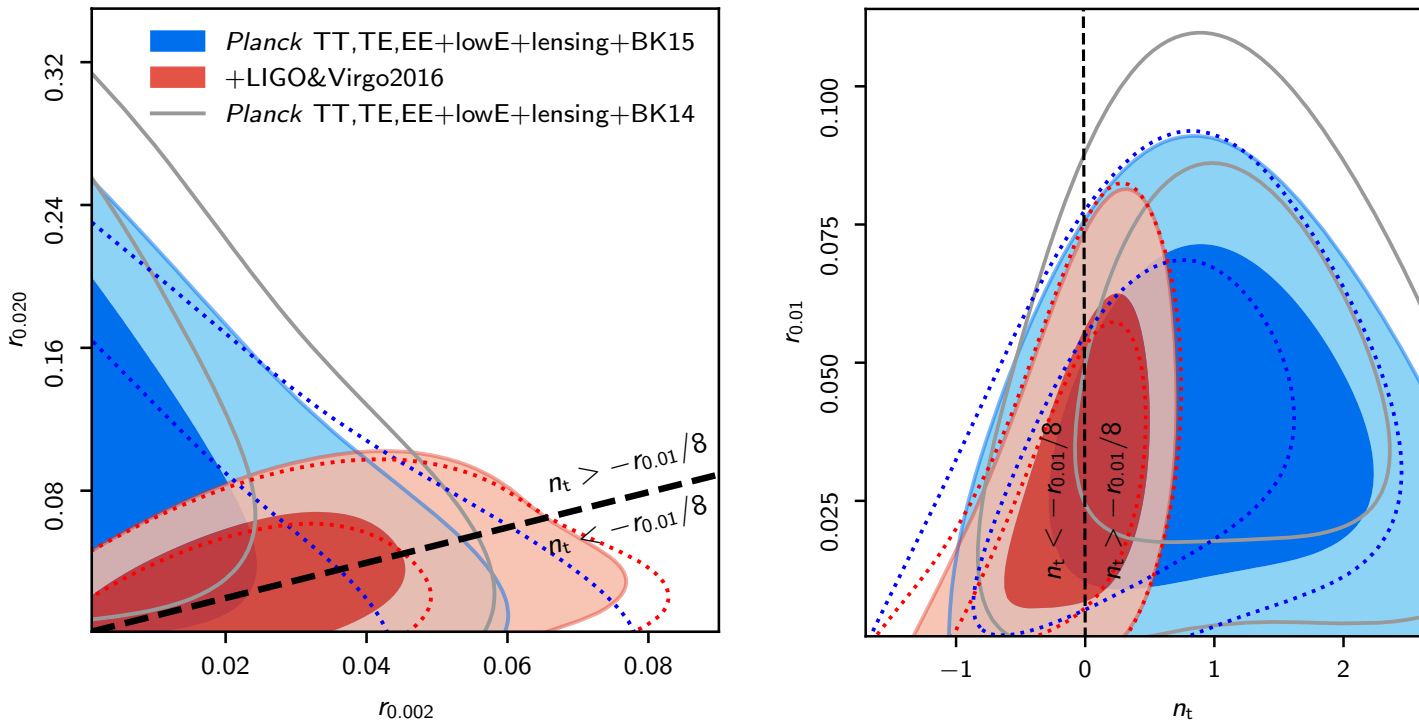


Figure 196: 68 % and 95 % CL constraints on tensor perturbations in the ΛCDM+$r_{0.002}$+$r_{0.02}$ model, i.e., when the inflationary consistency relation is relaxed. Filled contours in the left panel show the results for our independent primary parameters $r_{0.002}$ and $r_{0.02}$, which have uniform priors, and in the right panel for the derived parameters n_t and $r_{0.01}$, which have non-uniform priors. The dotted lines assume uniform priors on $r_{0.01}$ and n_t, calculated as in Fig. 195. The scale $k = 0.01\,\mathrm{Mpc}^{-1}$ is near the decorrelation scale of (n_t, r) for the *Planck*+BK15 data. In both panels the dashed black line indicates the inflationary consistency condition, $n_t = -r_{0.01}/8$. (The grey contours follow if we use the older BK14 data instead of the BK15 data.)

The one-dimensional posteriors are displayed in Fig. 195 (which also shows an additional data set, "LIGO&Virgo2016," discussed at the end of this subsection). We obtain for the ΛCDM+$r_{0.002}$+$r_{0.02}$ model:

$$\left.\begin{aligned} r_{0.002} &< 0.044 \\ r_{0.02} &< 0.184 \end{aligned}\right\} \begin{array}{l}(95\,\%\ \mathrm{CL},\ \textit{Planck} \\ \mathrm{TT,TE,EE} \\ +\mathrm{lowE}+\mathrm{lensing}+\mathrm{BK15}).\end{array} \tag{1265}$$

The constraints on the derived tensor parameters are $r_{0.01} < 0.076$ and $-0.55 < n_t < 2.54$ at 95 % CL.

The left and right panels of Fig. 196 show the two-dimensional contours for the primary parameters $(r_{0.002}, r_{0.02})$ and the derived ones $(r_{0.01}, n_t)$, respectively. The consistency condition, $n_t = -r/8$, denoted by the dashed lines, is fully compatible with *Planck*+BK15 data. However, a very blue tensor tilt with $n_t \simeq 2$ and $r_{0.01} \simeq 0.05$ is still within the 68 % CL region. Indeed, despite the larger amplitude of the primordial tensor power spectrum at small scales for blue n_t, the tensor modes are suppressed when re-entering the Hubble radius, which leads to damping of the observational signal at high k. This explains why the 68 % CL CMB constraint on $r_{0.02}$ is about a factor of four weaker than the one on $r_{0.002}$. Fig. 196 also shows a slight improvement of constraints by BK15 compared to the older BK14 data.

A stochastic background of gravitational waves (GWs) with a blue tensor tilt could be further constrained at much shorter wavelengths, such as those probed by ground-based interferometers dedicated to the direct detection of GWs. For example, assuming a scale-invariant tensor spectrum and using the frequency range (20–85.8) Hz, which corresponds to the wavenumbers

$k = 2\pi f = (1.3\text{–}5.5) \times 10^{16}\,\mathrm{Mpc}^{-1}$, LIGO and Virgo set an upper bound on the GW density parameter of $\Omega_{\rm GW}(f) \leq 1.7 \times 10^{-7}$ at 95 % CL [2]. While these scales are likely to be dominated by astrophysical sources, such as GWs from binary mergers, we next examine what constraints the LIGO&Virgo upper bound sets on primordial tensor perturbations, if we assume that they had a power-law spectrum all the way from CMB scales to ultra-short scales. We refer the interested reader to Meerburg et al. [187] and Cabass et al. [53] for the use of alternative data on short scales or of additional constraints on the effective energy-momentum tensor of the stochastic background of GWs averaged over wavelengths.

We obtain a conservative upper limit on the *primordial* contribution by demanding that the GW density from our scale-dependent primordial tensor perturbations [187, 2, 53],

$$\Omega_{\rm GW}(k) = \frac{k}{\rho_{\rm critical}} \frac{d\rho_{\rm GW}}{dk} = \frac{A_{\rm t}(k)}{24 z_{\rm eq}} = \frac{A_{\rm t1}(k/k_1)^{n_{\rm t}}}{24 z_{\rm eq}}, \tag{1266}$$

stays below the above-quoted limit at least at $k = 1.3 \times 10^{16}\,\mathrm{Mpc}^{-1}$ ($f = 20\,\mathrm{Hz}$). The posterior probability densities when this constraint is included in the analysis as a half-Gaussian prior are compared with those obtained by *Planck*+BK15 alone in Figs. 195 and 196. LIGO&Virgo sets a very high upper bound[170] on r at ultra-high k, separated from CMB scales by a factor of 10^{18} in k. Due to the long arm length, this effectively provides a cutoff for $n_{\rm t}$ and excludes the bluest spectra that were allowed by the CMB alone, leading to

$$\left.\begin{aligned} r_{0.002} &< 0.064 \\ r_{0.02} &< 0.081 \end{aligned}\right\} \begin{array}{l} (95\,\%\ \text{CL, } \textit{Planck} \\ \text{TT,TE,EE+lowE} \\ \text{+lensing+BK15+LIGO\&Virgo2016}), \end{array} \tag{1267}$$

or $r_{0.01} < 0.066$ and $-0.76 < n_{\rm t} < 0.52$. The consistency condition $n_{\rm t} = -r/8$ is also compatible with these tighter constraints, as can be seen by comparing the red contours and dashed black lines in Fig. 196. As LIGO&Virgo pushes $r_{0.02}$ down (and we assume a power-law tensor spectrum), the upper bound on $r_{0.002}$ becomes weaker than with the CMB alone. This is not surprising, since the system is analogous to a see-saw with a pivot point at $k \sim 0.01\,\mathrm{Mpc}^{-1}$, where the data are the most sensitive to the tensor perturbations (taking into account also the transfer function from primordial tensor perturbations to the observable B-mode signal). Once one end of the see-saw is pushed down the other end can go up without disturbing the spectrum too much at the middle scales. We will observe analogous behaviour with isocurvature perturbations, for which we also assume a power-law spectrum and have only an upper bound (not a detection); see Sect. 28.9.3.

28.4 Implications for single-field slow-roll inflationary models

In this section we discuss the implications of the *Planck* 2018 likelihood for standard single-field slow-roll inflation. We first update the results for the Hubble flow functions (HFFs) ϵ_i and the potential slow-roll parameters obtained by the analytic perturbative expansion in terms of the HFFs for the primordial spectra of fluctuations. For definitions of the HFF hierarchy and the potential slow-roll parameters see Table 27. We then present a Bayesian comparison for a representative selection of standard slow-roll inflationary models.

[170] Using Eq. (1266), the upper bound $\Omega_{\rm GW}(f = 20\,\mathrm{Hz}) \leq 1.7 \times 10^{-7}$ corresponds to a tensor perturbation amplitude $A_{\rm t}(k = 1.3 \times 10^{16}\,\mathrm{Mpc}^{-1}) \leq 24 z_{\rm eq} \times 1.7 \times 10^{-7} = 1.4 \times 10^{-2}$, where we used $z_{\rm eq} \simeq 3400$. Assuming further for the scalar perturbations that $n_{\rm s} = 0.9659$ and $\ln(10^{10} A_{\rm s}) = 3.044$ at $k = 0.05\,\mathrm{Mpc}^{-1}$, this can be converted into an upper bound $r \leq 2.6 \times 10^{7}$ at $k = 1.3 \times 10^{16}\,\mathrm{Mpc}^{-1}$.

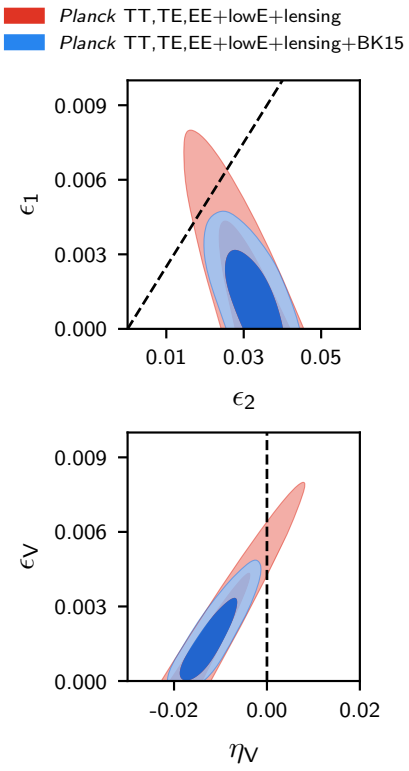


Figure 197: Marginalized joint two-dimensional 68 % and 95 % CL regions for $(\epsilon_1\,,\epsilon_2)$ (top panel) and $(\epsilon_V\,,\eta_V)$ (bottom panel) for *Planck* TT,TE,EE+lowE+lensing (red contours), compared with *Planck* TT,TE,EE+lowE+lensing+BK15 (blue contours). The dashed lines divide between convex and concave potentials.

28.4.1 Constraints on slow-roll parameters

Exploiting the approximate analytic expressions for the primordial power spectrum of scalar and tensor fluctuations obtained by the Green's function method [264, 107, 162], we can construct constraints on the slow-roll parameters.

When restricting to parameters first order in the HFFs, we obtain with *Planck* TT,TE,EE+lowE+lensing

$$\epsilon_1 < 0.0063 \qquad (0.0039) \qquad (95\,\%\ \mathrm{CL})\,, \tag{1268}$$

$$\epsilon_2 = 0.030^{+0.007}_{-0.005} \qquad (0.031 \pm 0.005) \qquad (68\,\%\ \mathrm{CL})\,. \tag{1269}$$

The *Planck* TT,TE,EE+lowE+lensing(+BK15) constraints on the slow-roll potential parameters ϵ_V and η_V can be obtained by an exact remapping of the constraints on the HFF parameters [162, 92] given above:

$$\epsilon_V < 0.0063 \qquad (0.0039) \qquad (95\,\%\ \mathrm{CL})\,, \tag{1270}$$

$$\eta_V = -0.010^{+0.004}_{-0.008} \qquad \left(-0.012^{+0.004}_{-0.005}\right) \qquad (68\,\%\ \mathrm{CL})\,. \tag{1271}$$

As can be seen from Fig. 197, the 95 % CL allowed contours are in the region of concave potentials when BK15 is combined with *Planck* 2018 data.

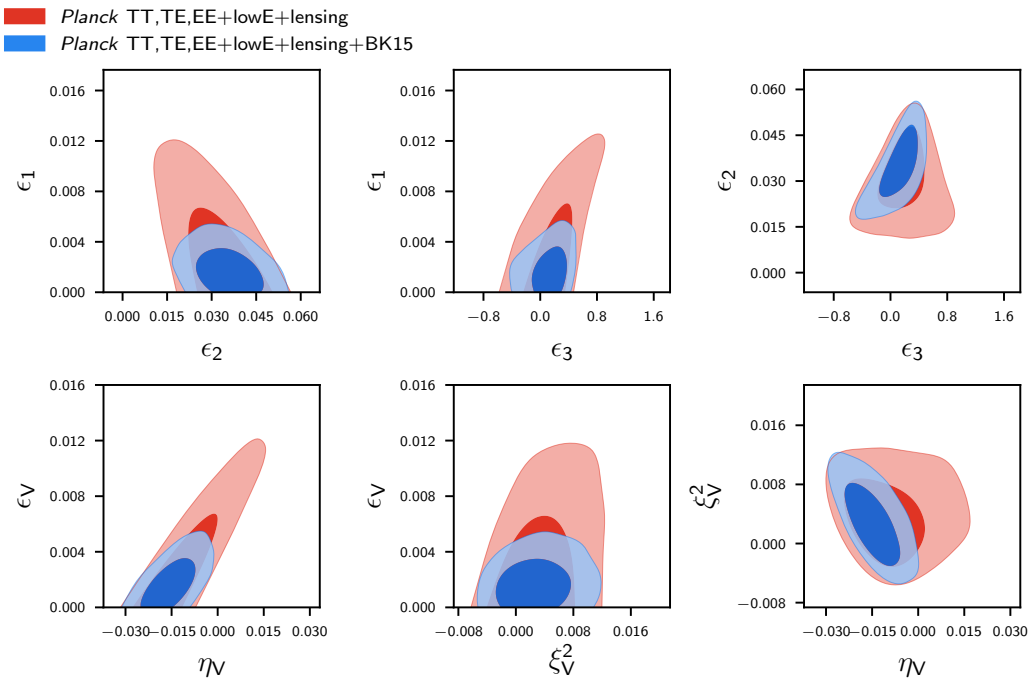


Figure 198: Marginalized joint two-dimensional 68 % and 95 % CL regions for combinations of $(\epsilon_1\,,\epsilon_2\,,\epsilon_3)$ (upper panels) and $(\epsilon_V\,,\eta_V,\xi^2_V)$ (lower panels) for *Planck* TT,TE,EE+lowE+lensing (red contours), compared with *Planck* TT,TE,EE+lowE+lensing+BK15 (blue contours).

When contributions to the primordial power spectra that are second-order in the HFFs are included, for *Planck* TT,TE,EE+lowE+lensing(+BK15) we obtain the following constraints on the slow-roll HFFs:

$$\epsilon_1 < 0.0097 \qquad (0.0044) \qquad (95\,\%\ \mathrm{CL})\,, \tag{1272}$$
$$\epsilon_2 = 0.032^{+0.009}_{-0.008} \qquad (0.035 \pm 0.008) \qquad (68\,\%\ \mathrm{CL})\,, \tag{1273}$$
$$\epsilon_3 = 0.19^{+0.55}_{-0.53} \qquad \left(0.12^{+0.36}_{-0.42}\right) \qquad (95\,\%\ \mathrm{CL})\,, \tag{1274}$$

and on the slow-roll potential parameters we obtain:

$$\epsilon_V < 0.0097 \qquad (0.0044) \qquad (95\,\%\ \mathrm{CL})\,, \tag{1275}$$
$$\eta_V = -0.010^{+0.007}_{-0.011} \qquad (-0.015 \pm 0.006) \qquad (68\,\%\ \mathrm{CL})\,, \tag{1276}$$
$$\xi^2_V = 0.0035^{+0.0078}_{-0.0072} \qquad \left(0.0029^{+0.0073}_{-0.0069}\right) \qquad (95\,\%\ \mathrm{CL})\,. \tag{1277}$$

The marginalized 68 % and 95 % CLs for the slow-roll HFF and potential parameters, allowing $\epsilon_3 \neq 0$, with *Planck* data alone or in combination with BK15, are displayed in Fig. 198.

Table 30: Priors for cosmological parameters used in the Bayesian comparison of inflationary models.

Parameter range	Prior type
$0.019 < \Omega_{\rm b}h^2 < 0.025$	uniform
$0.095 < \Omega_{\rm c}h^2 < 0.145$	uniform
$1.03 < 100\,\theta_{\rm MC} < 1.05$	uniform
$0.01 < \tau < 0.4$	uniform

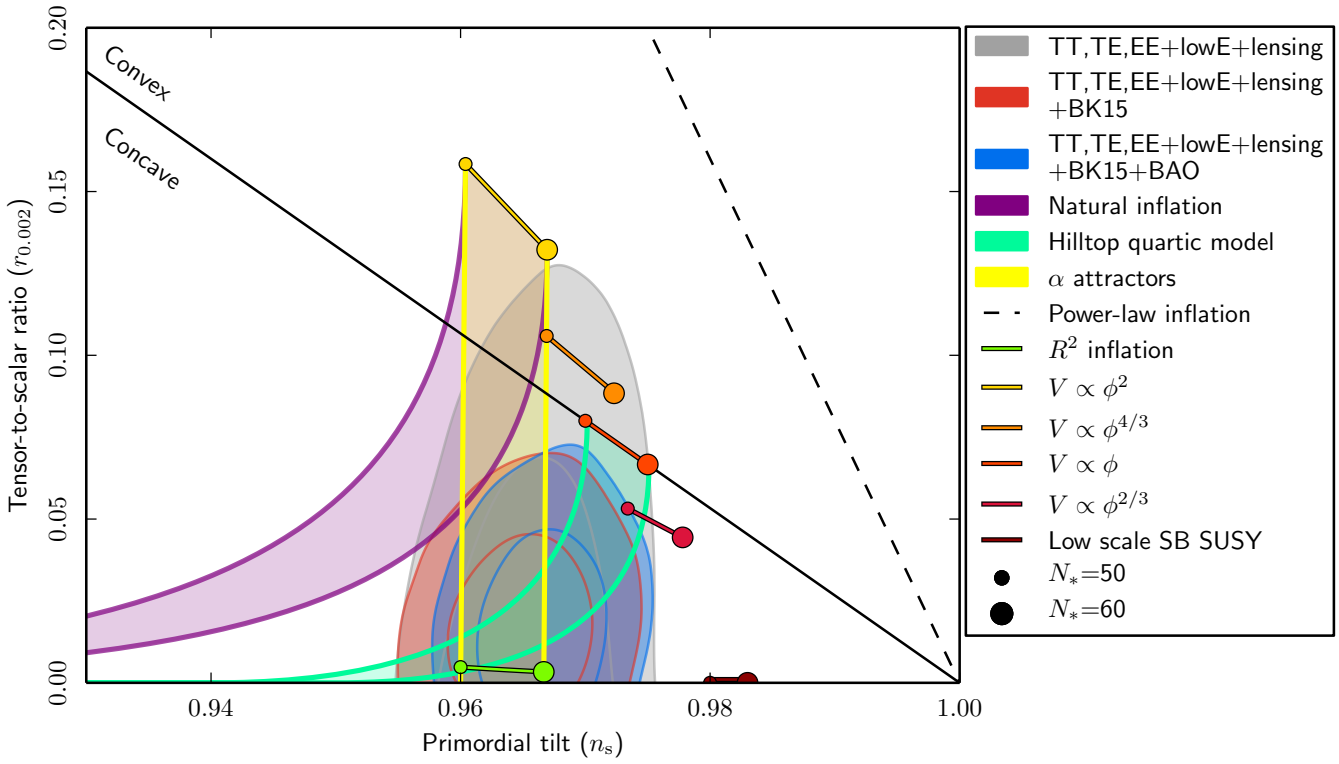


Figure 199: Marginalized joint 68 % and 95 % CL regions for $n_{\rm s}$ and r at $k = 0.002\,{\rm Mpc}^{-1}$ from *Planck* alone and in combination with BK15 or BK15+BAO data, compared to the theoretical predictions of selected inflationary models. Note that the marginalized joint 68 % and 95 % CL regions assume $dn_{\rm s}/d\ln k = 0$.

28.4.2 Implications for selected slow-roll inflationary models

The predictions for $(n_{\rm s}, r)$ to first order in the slow-roll approximation for a few inflationary models are shown in Fig. 199, which updates figure 12 of **(alias?)** and figure 1 of **(alias?)** with the same notation. These predictions are calculated for scale $k = 0.002\,{\rm Mpc}^{-1}$ and include an uncertainty in the number of e-folds of $50 < N_* < 60$.

In the following we discuss the implications of the *Planck* 2018 data release by taking into account the uncertainties in the entropy generation stage for a selection of representative standard single-field slow-roll inflationary models, updating the analysis reported in **(alias?)** and **(alias?)**. As in **(alias?)**, we use the primordial power spectra of cosmological fluctuations generated during slow-roll inflation parameterized by the HFFs, ϵ_i, to second order, which can be expressed in terms of the parameters of the inflationary model and the number of e-folds to the end of inflation, N_* [167, 181], given by [**(alias?)**]

$$N_* \simeq 67 - \ln\left(\frac{k_*}{a_0 H_0}\right) + \frac{1}{4}\ln\left(\frac{V_*^2}{M_{\rm pl}^4\,\rho_{\rm end}}\right)$$

$$+\frac{1-3w_{\rm int}}{12(1+w_{\rm int})}\ln\left(\frac{\rho_{\rm th}}{\rho_{\rm end}}\right)-\frac{1}{12}\ln(g_{\rm th})\,, \tag{1278}$$

where $\rho_{\rm end}$ is the energy density at the end of inflation, a_0H_0 is the present Hubble scale, V_* is the potential energy when k_* left the Hubble radius during inflation, $w_{\rm int}$ characterizes the effective equation of state between the end of inflation and the thermalization energy scale $\rho_{\rm th}$, and $g_{\rm th}$ is the number of effective bosonic degrees of freedom at the energy scale $\rho_{\rm th}$. We fix $g_{\rm th}=10^3$ and $\epsilon_{\rm end}=1$, and we use modified routines of the public code `ASPIC`[171] [183]. In order to make contact with Fig. 199, we consider the pivot scale $k_*=0.002\,{\rm Mpc}^{-1}$ in this subsection. We assume the uniform priors for the cosmological parameters listed in Table 30, and logarithmic priors on $10^{10}A_{\rm s}$ (over the interval $[e^{2.5},e^{3.7}]$) and $\rho_{\rm th}$ (over the interval $[(1\,{\rm TeV})^4,\rho_{\rm end}]$). Prior ranges for additional parameters in the inflationary models considered are listed in Table 31. In this paper we consider the implications of the *Planck* 2018 data for the selection of representative models studied in **(alias?)** by restricting ourselves to $w_{\rm int}=(p-2)/(p+2)$, when the potential can be approximated as $V(\phi)\propto\phi^p$ during the coherent oscillation regime after inflation, or simply $w_{\rm int}=0$ when the potential considered describes only the inflationary stage.[172] For data we use the full constraining power of *Planck*, i.e., *Planck* TT,TE,EE+lowE+lensing, in combination with BK15.

[171] http://cp3.irmp.ucl.ac.be/~ringeval/aspic.html

[172] Note that some inflationary potentials in this selection are a valid model for all stages, from the slow-roll phase all the way to coherent oscillations around the minimum during reheating, while others are "incomplete" in the sense that they only describe the slow-roll regime. The hilltop, D-brane, potential with exponential tails, and spontaneously broken SUSY models fall into the latter category and rely on additional terms, denoted by the ellipses, to complete the potential at the end of inflation. With the increasing precision of CMB data and accompanying accuracy requirements for theoretical predictions, the precise form of the additional terms may affect the scientific interpretation of some incomplete models, as pointed out for the case of quadratic hilltop and double-well inflationary models in **(alias?)**.

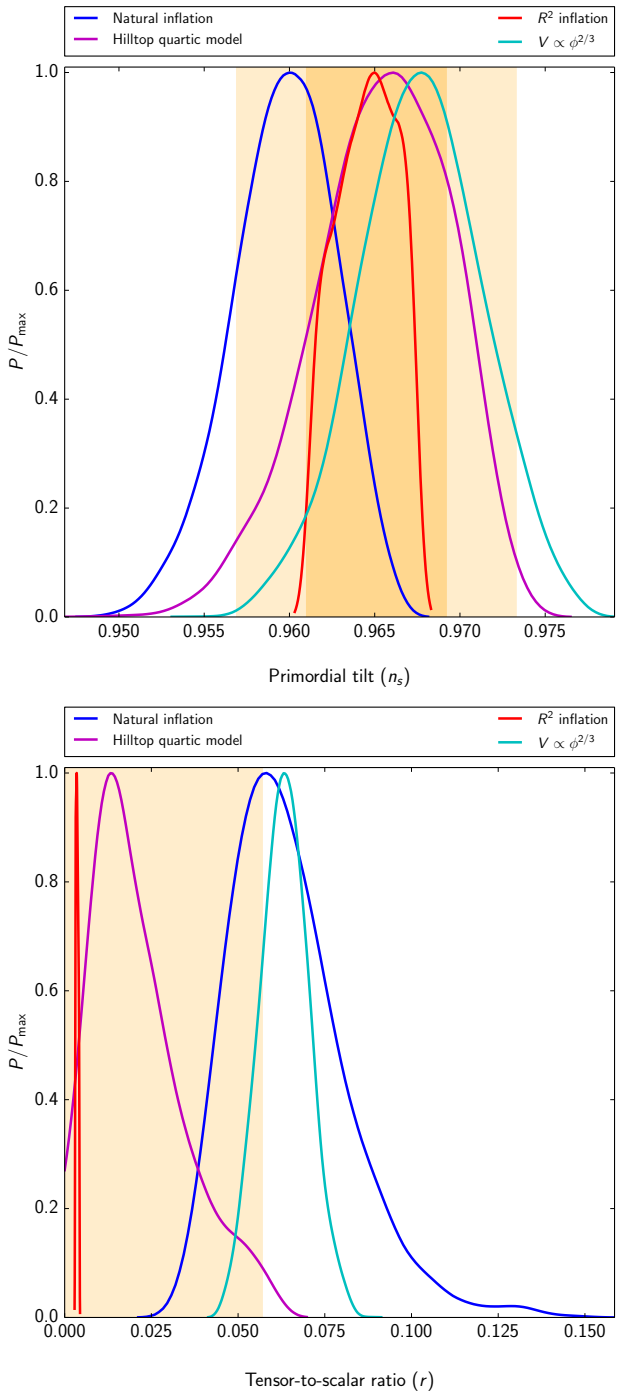


Figure 200: Marginalized probability densities of the scalar tilt n_s (top panel) and r (bottom panel) at $k = 0.002\,\mathrm{Mpc}^{-1}$ for natural, R^2, hilltop quartic, and $V(\phi) \propto \phi^{2/3}$ inflation, obtained by marginalizing over the uncertainties in the entropy generation stage, compared to the corresponding 68 % and 95 % CL limits obtained from a ΛCDM-plus-tensor fit.

The $\Delta\chi^2$ and the Bayesian evidence values for a selection of inflationary models with respect to the R^2 model [260, 205, 259] are shown in Table 31. Figure 200 shows the resulting marginalized probability densities of n_s and r at $k = 0.002\,\mathrm{Mpc}^{-1}$ for a few inflationary models with the above specified priors, compared to the corresponding 68 % and 95 % CL limits obtained from a ΛCDM-plus-tensor fit. We refer the interested reader to **(alias?)** for a concise description of the inflationary models studied here and we limit ourselves here to a summary of the main results of this analysis.

- The inflationary predictions [205, 259] originally computed for the R^2 model [260] to

Inflationary model	Potential $V(\phi)$	Parameter range
$R + R^2/(6M^2)$	$\Lambda^4 \left(1 - e^{-\sqrt{2/3}\phi/M_{\rm Pl}}\right)^2$	...
Power-law potential	$\lambda M_{\rm Pl}^{10/3} \phi^{2/3}$	...
Power-law potential	$\lambda M_{\rm Pl}^3 \phi$	...
Power-law potential	$\lambda M_{\rm Pl}^{8/3} \phi^{4/3}$	...
Power-law potential	$\lambda M_{\rm Pl}^2 \phi^2$	...
Power-law potential	$\lambda M_{\rm Pl} \phi^3$	...
Power-law potential	$\lambda \phi^4$	...
Non-minimal coupling	$\lambda^4 \phi^4 + \xi \phi^2 R/2$	$-4 < \log_{10} \xi < 4$
Natural inflation	$\Lambda^4 \left[1 + \cos\left(\phi/f\right)\right]$	$0.3 < \log_{10}(f/M_{\rm Pl}) < 2.5$
Hilltop quadratic model	$\Lambda^4 \left(1 - \phi^2/\mu_2^2 + \dots\right)$	$0.3 < \log_{10}(\mu_2/M_{\rm Pl}) < 4.85$
Hilltop quartic model	$\Lambda^4 \left(1 - \phi^4/\mu_4^4 + \dots\right)$	$-2 < \log_{10}(\mu_4/M_{\rm Pl}) < 2$
D-brane inflation ($p = 2$)	$\Lambda^4 \left(1 - \mu_{\rm D\,2}^2/\phi^p + \dots\right)$	$-6 < \log_{10}(\mu_{\rm D\,2}/M_{\rm Pl}) < 0.3$
D-brane inflation ($p = 4$)	$\Lambda^4 \left(1 - \mu_{\rm D\,4}^4/\phi^p + \dots\right)$	$-6 < \log_{10}(\mu_{\rm D\,4}/M_{\rm Pl}) < 0.3$
Potential with exponential tails	$\Lambda^4 \left[1 - \exp\left(-q\phi/M_{\rm Pl}\right) + \dots\right]$	$-3 < \log_{10} q < 3$
Spontaneously broken SUSY	$\Lambda^4 \left[1 + \alpha_h \log\left(\phi/M_{\rm Pl}\right) + \dots\right]$	$-2.5 < \log_{10} \alpha_h < 1$
E-model ($n = 1$)	$\Lambda^4 \left\{1 - \exp\left[-\sqrt{2}\phi \left(\sqrt{3\alpha_1^{\rm E}} M_{\rm Pl}\right)^{-1}\right]\right\}^{2n}$	$-2 < \log_{10} \alpha_1^{\rm E} < 4$
E-model ($n = 2$)	$\Lambda^4 \left\{1 - \exp\left[-\sqrt{2}\phi \left(\sqrt{3\alpha_2^{\rm E}} M_{\rm Pl}\right)^{-1}\right]\right\}^{2n}$	$-2 < \log_{10} \alpha_2^{\rm E} < 4$
T-model ($m = 1$)	$\Lambda^4 \tanh^{2m}\left[\phi \left(\sqrt{6\alpha_1^{\rm T}} M_{\rm Pl}\right)^{-1}\right]$	$-2 < \log_{10} \alpha_1^{\rm T} < 4$
T-model ($m = 2$)	$\Lambda^4 \tanh^{2m}\left[\phi \left(\sqrt{6\alpha_2^{\rm T}} M_{\rm Pl}\right)^{-1}\right]$	$-2 < \log_{10} \alpha_2^{\rm T} < 4$

Table 31: Bayesian comparison for a selection of slow-roll inflationary models with $w_{\rm int}$ fixed (see text for more details). We quote 0.3 as the error on the Bayes factor. Models are strongly disfavoured when $\ln B < -5$.

lowest order,

$$n_{\rm s} - 1 \simeq -\frac{2}{N}\,, \quad r \simeq \frac{12}{N^2}, \tag{1279}$$

are in good agreement with *Planck* 2018 data, confirming the previous 2013 and 2015 results. The 95 % CL allowed range $49 < N_* < 59$ is compatible with the R^2 basic predictions $N_* = 54$, corresponding to $T_{\rm reh} \sim 10^9$ GeV [37]. A higher reheating temperature $T_{\rm reh} \sim 10^{13}$ GeV, as predicted in Higgs inflation [36], is also compatible with the *Planck* data.

- Monomial potentials [171] $V(\phi) = \lambda M_{\rm Pl}^4 \left(\phi/M_{\rm Pl}\right)^p$ with $p \geq 2$ are strongly disfavoured with respect to the R^2 model. For these values the Bayesian evidence is worse than in 2015 because of the smaller level of tensor modes allowed by **(alias?)**. Models with $p = 1$ or $p = 2/3$ [251, 184, 185] are more compatible with the data.

- There are several mechanisms which could lower the predictions for the tensor-to-scalar ratio for a given potential $V(\phi)$ in single-field inflationary models. Important examples are a subluminal inflaton speed of sound due to a non-standard kinetic term [103], a non-minimal coupling to gravity [258, 174, 245, 82], or an additional damping term for the inflaton due to dissipation in other degrees of freedom, as in warm inflation [34, 24].

In the following we report on the constraints for a non-minimal coupling to gravity of the type $F(\phi)R$, with $F(\phi) = M_{\rm Pl}^2 + \xi\phi^2$, and a quartic potential. For this model we compute the theoretical predictions in terms of HFFs and number of e-folds to the end of inflation in the Einstein frame as for the R^2 model above, but we omit the technical details for the sake of brevity.[173] Our results show that a quartic potential, which would be excluded at high statistical significance for a minimally-coupled scalar inflaton as seen from Table 31, can be reconciled with the *Planck* and BK15 data for $\xi > 0$: we obtain a 95 % CL lower limit $\log_{10}\xi > -1.5$ with $\ln B = -2.4$.

- Natural inflation [99, 7] is strongly disfavoured by the *Planck* 2018 plus BK15 data with a Bayes factor $\ln B = -6.6$.
- Within the class of hilltop inflationary models [40] we find that a quartic potential provides a better fit than a quadratic one. In the quartic case we find the 95 % CL lower limit $\log_{10}(\mu_2/M_{\rm Pl}) > 1.0$.
- D-brane inflationary models [148, 69, 101] provide a good fit to *Planck* and BK15 data for a large portion of their parameter space.
- For the simple class of inflationary potentials with exponential tails [106, 263, 70, 48, 61] we find $\ln B = -1.0$.
- *Planck* 2018 and BK15 data strongly disfavour the hybrid model driven by logarithmic quantum corrections in spontaneously broken supersymmetric (SB SUSY) theories [71], with $\ln B = -6.8$.
- *Planck* and BK15 data set tight constraints on α attractors [149, 91]. We obtain $\log_{10}\alpha_1^{\rm E} < 1.3$ and $\log_{10}\alpha_2^{\rm E} < 1.1$ at 95 % CL for the E-model. We obtain slightly tighter 95 % CL bounds for the T-model, i.e., $\log_{10}\alpha_1^{\rm T} < 1.0$ and $\log_{10}\alpha_2^{\rm T} < 1.0$. Given the relation $|R_K| = 2/(3\alpha)$ between the curvature of the Kähler geometry R_K and α in some of the T-models motivated by supergravity, *Planck* and BK15 data imply a lower bound on $|R_K|$, which is still in the low-curvature regime. The discrete set of values $\alpha = i/3$ with an integer i in the range $[1, 7]$ motivated by maximal supersymmetry [90, 150] is compatible with the current data.

28.5 Reconstruction of the inflaton potential

28.5.1 Taylor expansion of $V(\phi)$ in the observable region

In this section, as in section 6 of **(alias?)** and section 7.1 of **(alias?)**, we try to reconstruct the inflaton potential only in its observable window, making no assumptions about the end of inflation. The motivation for being so conservative is that what happens after the inflaton rolls down beyond this range might not be captured by the simplest descriptions. More elaborate treatments would be required, for instance, in the case of a non-trivial potential shape before the end of inflation, a waterfall transition involving extra scalar fields, or several short inflationary stages between the time at which CMB scales exit the Hubble radius and the nucleosynthesis epoch. The analysis of this section relies, however, on the assumption that the potential is smooth enough inside the observable window to be described by a Taylor expansion up to order four. Note that this assumption is much weaker than assuming that a Taylor expansion is valid up to the end of inflation. However, it excludes from the analysis potentials with sharp features in the observable window, such as those studied in the next sections.

We perform the Taylor expansion around the value ϕ_* of the inflaton field evaluated precisely at the time t_* when the pivot scale $k_* = 0.05\,{\rm Mpc}^{-1}$ fulfills the relation $k_* = a(t_*)H(t_*)$. We separately study the cases where the expansion is performed at order $n = 2$, $n = 3$, or $n = 4$. We compute the primordial spectrum with a full integration of the Fourier mode evolution, using the inflationary module of the `CLASS` code. Although this method assumes no slow-roll approximation at any point, we speed up the convergence of the Markov Chain by taking flat

[173] In this model the potential in the Einstein frame is known only in implicit form (see, for instance, Garcia-Bellido et al. [100]) and the algebra is therefore more complicated.

priors not directly on the five Taylor coefficients $\{V, V_\phi, \ldots, V_{\phi\phi\phi\phi}\}$, but on combinations of them matching the definitions of the potential slow-roll parameters $\{\epsilon_V, \eta_V, \xi_V^2, \varpi_V^3\}$ presented in table 2 of **(alias?)**. Even beyond the slow-roll approximation, these combinations provide nearly linear contributions to the tilt, running, running of the running, etc., of the scalar and tensor spectrum. Hence, they are directly related to observable quantities and well constrained by the data. Instead, if we ran with flat priors on $\{V, V_\phi, \ldots, V_{\phi\phi\phi\phi}\}$, the convergence would be plagued by complicated parameter degeneracies.

The results of this analysis are shown in the panels of Fig. 201 and Table 32 for $n = 2$, 3, and 4, using two data sets for each: *Planck* TT,TE,EE+lowE alone; or *Planck* TT,TE,EE+lowE+lensing+BK1
The plot in Fig. 201 deliberately has a lot of white space because, for the sake of comparison, we plotted it over the same parameter ranges as the same plot in **(alias?)**. We notice some significant improvement. Comparing *Planck* TT,TE,EE+lowE results from 2015 and 2018, we find that error bars on individual parameters have typically been reduced by 30 % thanks to improved polarization data. Including BK data provides further constraining power. Comparing *Planck* 2015 TT+lowP+BAO and *Planck* 2018 TT,TE,EE+lowE+lensing+BK15, we find that the error bars on $\{\epsilon_V, \eta_V, \xi_V^2, \varpi_V^3\}$ shrink by factors of 2 to 4. The new data tend to resolve degeneracies which previously appeared in the $n = 4$ case and could be understood as a compensation mechanism between potentials with large running of the tilt, running of the running, tensor contribution, etc. The parameters ξ_V^2 and ϖ_V^3 are perfectly compatible with zero (see Fig. 201 and Table 32), and so are $V_{\phi\phi\phi}$ and $V_{\phi\phi\phi\phi}$ (see the contours on the parameters $\{V, V_\phi, \ldots, V_{\phi\phi\phi\phi}\}$ in Fig. 202). This is consistent with the fact that the new data set brings no evidence for running or running of the running. It also explains why the results of this section are close to those of Sect. 28.4.1, obtained under the slow-roll approximation. Similar to 2015, the best-fit value of running for $n = 3$ is negative, but has moved down from -0.013 to -0.007, and remains compatible with zero at the 1.0σ level. For $n = 4$, the trend observed in 2015 to fit the data slightly better with a non-zero tensor contribution has disappeared. The decrease of the minimum effective χ^2 when moving from $n = 2$ to $n = 3$ is insignificant and even smaller than in 2015, showing that the data do not require anything more complicated than an approximately parabolic shape for the inflaton potential within the observable window.

This can be checked by considering the random sample of well-fitting potentials presented in Fig. 203. Actually, for $n = 4$, a few of the plotted potentials have a non-parabolic "spoon-like" shape (with a kink and a plateau), because non-negligible values of $|V_{\phi\phi\phi\phi}|$ are still allowed. However, this sub-class of models is by no means preferred over simpler parabolic-like potentials with a negligible $|V_{\phi\phi\phi\phi}|$; otherwise, we would have obtained a better $\chi^2_{\rm eff}$ for $n = 4$. Hence one should *not* take from Fig. 203 the message that special potentials with a kink and a plateau are favoured by the *Planck* data. Comparing this plot to figure 15 of **(alias?)**, we see that the models with the largest $V(\phi)$ amplitude are excluded by stronger bounds on the tensor modes.

Finally, it is interesting to notice that the predictions for the parameters of the minimal ΛCDM model, such as $n_{\rm s}$ or τ, remain extremely stable when increasing the freedom in the inflaton potential.

n	2	3	4
ϵ_V	< 0.0042	< 0.0045	< 0.0048
η_V	$-0.0124^{+0.0033}_{-0.0052}$	$-0.0163^{+0.0057}_{-0.0063}$	$-0.0082^{+0.0096}_{-0.0120}$
ξ^2_V	...	$0.0036^{+0.0035}_{-0.0037}$	$-0.004^{+0.011}_{-0.009}$
ϖ^3_V	...	...	$0.0048^{+0.0052}_{-0.0067}$
τ	$0.0546^{+0.0072}_{-0.0075}$	$0.0559^{+0.0074}_{-0.0080}$	$0.0571^{+0.0077}_{-0.0085}$
n_s	$0.9650^{+0.0042}_{-0.0042}$	$0.9639^{+0.0043}_{-0.0048}$	$0.9623^{+0.0047}_{-0.0050}$
$10^3 \frac{dn_\mathrm{s}}{d\ln k}$	$-0.37^{+0.29}_{-0.19}$	$-7.3^{+7.0}_{-6.7}$	$-1.9^{+9.0}_{-9.2}$
$r_{0.002}$	< 0.060	< 0.063	< 0.069
$\Delta\chi^2_\mathrm{eff}$	...	$\Delta\chi^2_{3/2} = -0.22$	$\Delta\chi^2_{4/3} = -0.82$

Table 32: Numerical reconstruction of the potential slow-roll parameters *beyond* any slow-roll approximation, when the potential is Taylor-expanded to nth order, using *Planck* TT,TE,EE+lowE+lensing+BK15. We also show the corresponding bounds on some related parameters (here n_s, $dn_\mathrm{s}/d\ln k$, and $r_{0.002}$ are derived from the numerically computed primordial spectra). All error bars are at the 68 % CL and all upper bounds at the 95 % CL. The effective χ^2 value of model n is given relative to model $n-1$.

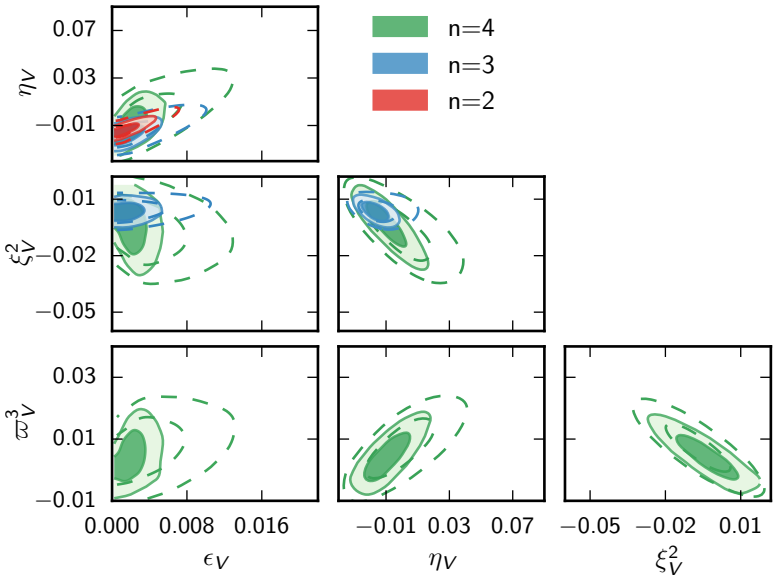


Figure 201: Taylor expansion of $V(\phi)$ at order $n = 2$, 3, and 4 in the observable region, making no assumption about the end of inflation. The parameters are combinations of Taylor coefficients with flat priors. Dashed contours are *Planck* TT,TE,EE+lowE, while solid contours are *Planck* TT,TE,EE+lowE+lensing+BK15. The scales are the same as in **(alias?)**.

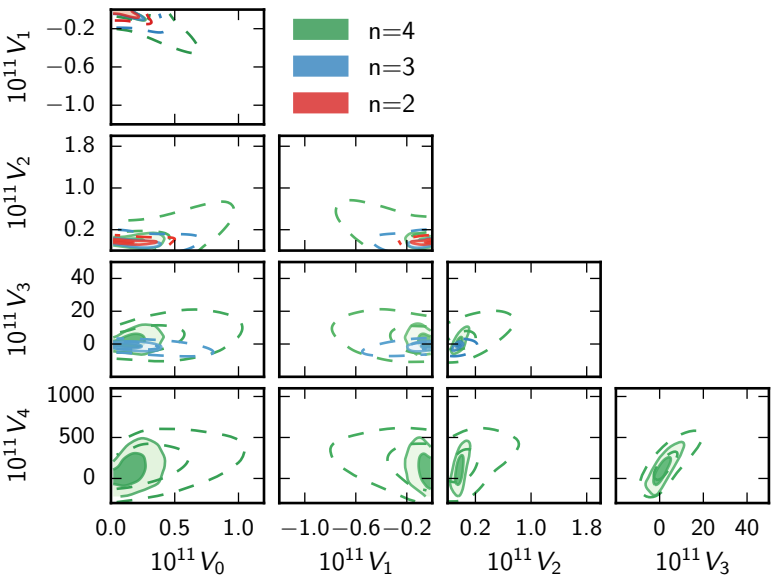


Figure 202: Taylor expansion of $V(\phi)$ at order $n = 2$, 3, and 4 in the observable region, making no assumption about the end of inflation. In natural units (where $\sqrt{8\pi}M_{\rm Pl} = 1$). The parameters are the Taylor coefficients, obtained here as derived parameters with non-flat priors. Dashed contours are *Planck* TT,TE,EE+lowE, while solid contours are *Planck* TT,TE,EE+lowE+lensing+BK15. The scales are the same as in **(alias?)**.

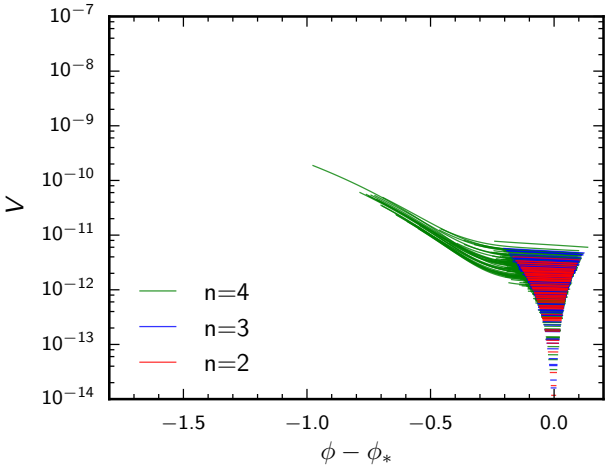


Figure 203: Representative sample of the observable region of inflaton potentials allowed at the 95 % CL, when the potential is Taylor-expanded at order $n = 2$, 3, and 4 in the observable region, making no assumption about the end of inflation, and using *Planck* TT,TE,EE+lowE+lensing+BK15. In natural units (where $\sqrt{8\pi}M_{\rm Pl} = 1$). We use the same scales as in **(alias?)**. Note that there is another branch of solutions that is symmetric under $(\phi - \phi_*) \to -(\phi - \phi_*)$.

28.5.2 Taylor expansion of $H(\phi)$ in the observable region

To assess the robustness of our method, in this section we repeat the analysis with a Taylor expansion of the Hubble function $H(\phi)$ in the observable window, as we did in 2015. We refer the reader to section 7.2 of **(alias?)** for a precise description of this analysis, and we recall that the difference with respect to the $V(\phi)$ reconstruction is more than a mere change of priors. For each value of n, the new parameterization covers a slightly different range of potentials, and, more importantly, it naturally includes a marginalization over the uncertainty in the initial value of the derivative $\dot{\phi}$ when the inflaton enters the observable window. Instead,

n	2	3	4
ϵ_H	< 0.0041	< 0.0046	< 0.0041
η_H	$-0.0139^{+0.0026}_{-0.0038}$	$-0.0170^{+0.0044}_{-0.0048}$	$-0.0158^{+0.0057}_{-0.0056}$
ξ^2_H	...	$0.046^{+0.043}_{-0.045}$	$0.021^{+0.071}_{-0.076}$
ϖ^3_H	...	...	$0.16^{+0.64}_{-0.37}$
τ	$0.0548^{+0.0075}_{-0.0074}$	$0.0556^{+0.0076}_{-0.0078}$	$0.0563^{+0.0073}_{-0.0078}$
n_s	$0.9651^{+0.0040}_{-0.0044}$	$0.9637^{+0.0042}_{-0.0046}$	$0.9637^{+0.0042}_{-0.0048}$
$10^3 \frac{dn_\mathrm{s}}{d\ln k}$	$-0.25^{+0.20}_{-0.12}$	$-7.5^{+7.0}_{-6.7}$	$-5.1^{+7.8}_{-8.1}$
$r_{0.002}$	< 0.059	< 0.065	< 0.057
$\Delta\chi^2_\mathrm{eff}$	...	$\Delta\chi^2_{3/2} = -1.60$	$\Delta\chi^2_{4/3} = -2.32$

Table 33: Numerical reconstruction of the Hubble slow-roll parameters *beyond* any slow-roll approximation, using *Planck* TT,TE,EE+lowE+lensing+BK15. We also show the corresponding bounds on some related parameters (here n_s, $dn_\mathrm{s}/d\ln k$, and $r_{0.002}$ are derived from the numerically computed primordial spectra). All error bars are at the 68 % CL and all upper bounds at the 95 % CL. The effective χ^2 value of model n is given relative to model $n-1$.

in the previous analysis, $\dot\phi$ was assumed to have reached the inflaton attractor solution, i.e., there was an implicit assumption that inflation started well before that time. In the analysis based on $H(\phi)$, inflation models with a minimal duration are not excluded by the priors.

The improvement with respect to the 2015 results is even more impressive in this case. Bounds on the $n = 4$ parameters are typically 3 to 4 times stronger compared with 2015, as can be checked from Table 33 and Fig. 204. We found that a factor of 2 improvement comes from switching to the new set of low-ℓ likelihoods, and another factor of 2 from adding the BK likelihood. On the other hand, the use of more recent high-ℓ and lensing likelihoods has a modest impact.

A consequence of these improved constraints can be seen in Fig. 205, when we compare it to its counterpart from 2015 (figure 20 in **(alias?)**). Again, for a better comparison Fig. 205 uses the same scale as figure 20 of **(alias?)**. For $n = 4$, the previously best-fitting models included many scenarios starting with a fast-roll stage, producing a tail with large $V(\phi)$ before pivot-scale crossing. These models are now excluded by better polarization data and tensor constraints.

Going beyond the parabolic approximation for $H(\phi)$ does not improve the goodness-of-fit: as in the potential-based analysis of Sect. 28.5.1, the $\Delta\chi^2$s between $n = 2$, $n = 3$, and $n = 4$ are negligible, and the parameters ξ^2_H and ϖ^3_H related to $H_{\phi\phi\phi}$ and $H_{\phi\phi\phi\phi}$ are compatible with zero.

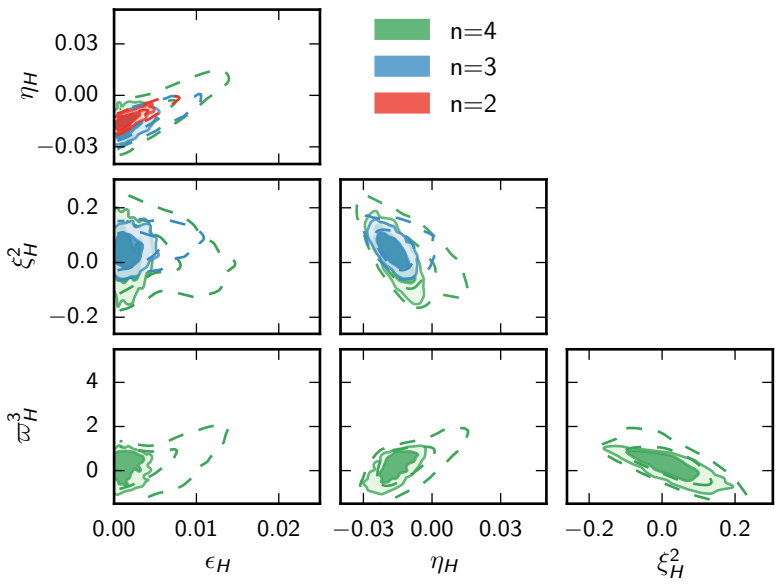


Figure 204: Taylor expansion of $H(\phi)$ at order $n = 2$, 3, and 4 in the observable region, making no assumption about the end of inflation. The parameters are combinations of Taylor coefficients with flat priors. Dashed contours are *Planck* TT,TE,EE+lowE, while solid contours are *Planck* TT,TE,EE+lowE+lensing+BK15. The scales are the same as in **(alias?)**.

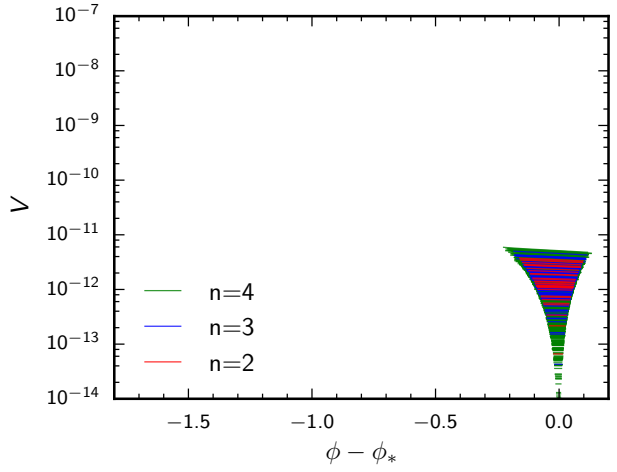


Figure 205: Representative sample of the observable region of inflaton potentials allowed at the 95 % CL, inferred from $H(\phi)$ when that function is Taylor-expanded at order $n = 2$, 3, and 4 in the observable region, making no assumption about the end of inflation, and using *Planck* TT,TE,EE+lowE+lensing+BK15. In natural units (where $\sqrt{8\pi}M_{\rm Pl} = 1$). The scales are the same as in **(alias?)**. Note that there is another branch of solutions symmetric under $(\phi - \phi_*) \to -(\phi - \phi_*)$.

28.5.3 Taylor expansion of full $V(\phi)$

We now present a new analysis with less conservative assumptions than in the previous subsections. We switch to the hypothesis that the inflaton potential is very smooth not only within its observable window, but also until the end of inflation, such that its whole shape can be captured by a Taylor expansion. We further assume that inflation ends when the first slow-roll condition is violated ($\epsilon_V = 1$), without invoking any other field. Finally, we fix the number of e-folds between Hubble crossing of the pivot scale and the end of inflation to $N_* = 55$, which implicitly relies on the hypothesis that no further inflationary stage took place at a later epoch.

Technically, the analysis pipeline for this case is similar to that of Sect. 28.5.1, except for an extra step in which the `CLASS` inflationary module integrates the background equations until the end of inflation, goes backwards in time by 55 e-folds, and imposes that the Hubble crossing for the pivot scale $k_* = 0.05\,\mathrm{Mpc}^{-1}$ matches that time.

These models are much more constrained than those of Sects. 28.5.1 and 28.5.2, since the e-fold condition is imposed in addition to having a potential with a good shape within the observable window. The constraining power is then sufficient for running the MCMC chains directly with flat priors on $\{V, V_\phi, \ldots, V_{\phi\phi\phi\phi}\}$.

Our results are presented in Figs. 206 and 207 and in Table 34. For models with a purely quadratic potential, the numerically computed tilt and tensor-to-scalar ratio depend almost exclusively on N_*, thus they remain fixed to $n_s = 0.963$ and $r_{0.002} = 0.136$. Such a large r is in tension with the *Planck* data, and even more so with the *Planck*+BK data. Thus the effective χ^2 is poor in the $n = 2$ case and improves considerably when adding some freedom in going to $n = 3$. Indeed, the presence of an additional cubic term allows us to reach smaller values of the tensor-to-scalar ratio for roughly the same scalar tilt, and lowers $\chi^2_{\rm eff}$ by more than 13. Instead, when also adding a quartic term, we find no significant improvement in the goodness of fit, and the coefficient of the ϕ^4 term is consistent with zero.

These findings are consistent with the global picture that *Planck* data prefer potentials which are concave in the observable window. The blue and green curves in the lower left panel of Fig. 207 illustrate the preference of the *Planck*+lensing+BK15 data for potentials with an inflection point, appearing qualitatively similar to scalar field potentials associated with spontaneous symmetry breaking models, hilltop models, new inflation, natural inflation, etc.

In these runs, the value of the scalar tilt running is always very precisely constrained around a value of $dn_{\rm s}/d\ln k \simeq -6 \times 10^{-4}$. This does not come as a surprise if we keep in mind that these bounds are not imposed directly by the data, but rather by the class of inflationary potentials considered here, with potential parameters fixed by observational bounds on the *amplitude* and *tilt* of the scalar and tensor spectra. In other words, the running is not directly measured, but rather predicted as a function of the scalar/tensor amplitudes and scalar tilt. Interestingly, if combinations of future CMB and large-scale structure data with a wide lever arm in wavenumber space could become directly sensitive to such tiny values (which would require a factor of around 10 improvement in sensitivity compared to current CMB+BAO data), a very large class of currently successful inflationary models could be either confirmed or ruled out.

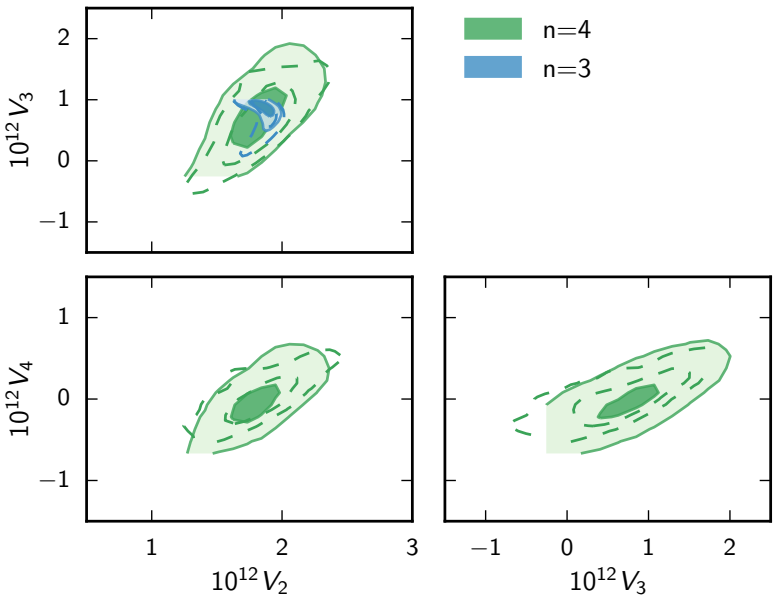


Figure 206: Taylor expansion of the full $V(\phi)$ at order $n = 3$ and 4, trusted until the end of inflation, in natural units (where $\sqrt{8\pi}M_{\rm Pl} = 1$). The parameters are the Taylor coefficients with flat priors. Dashed contours are *Planck* TT,TE,EE+lowE, while solid contours are *Planck* TT,TE,EE+lowE+lensing+BK15.

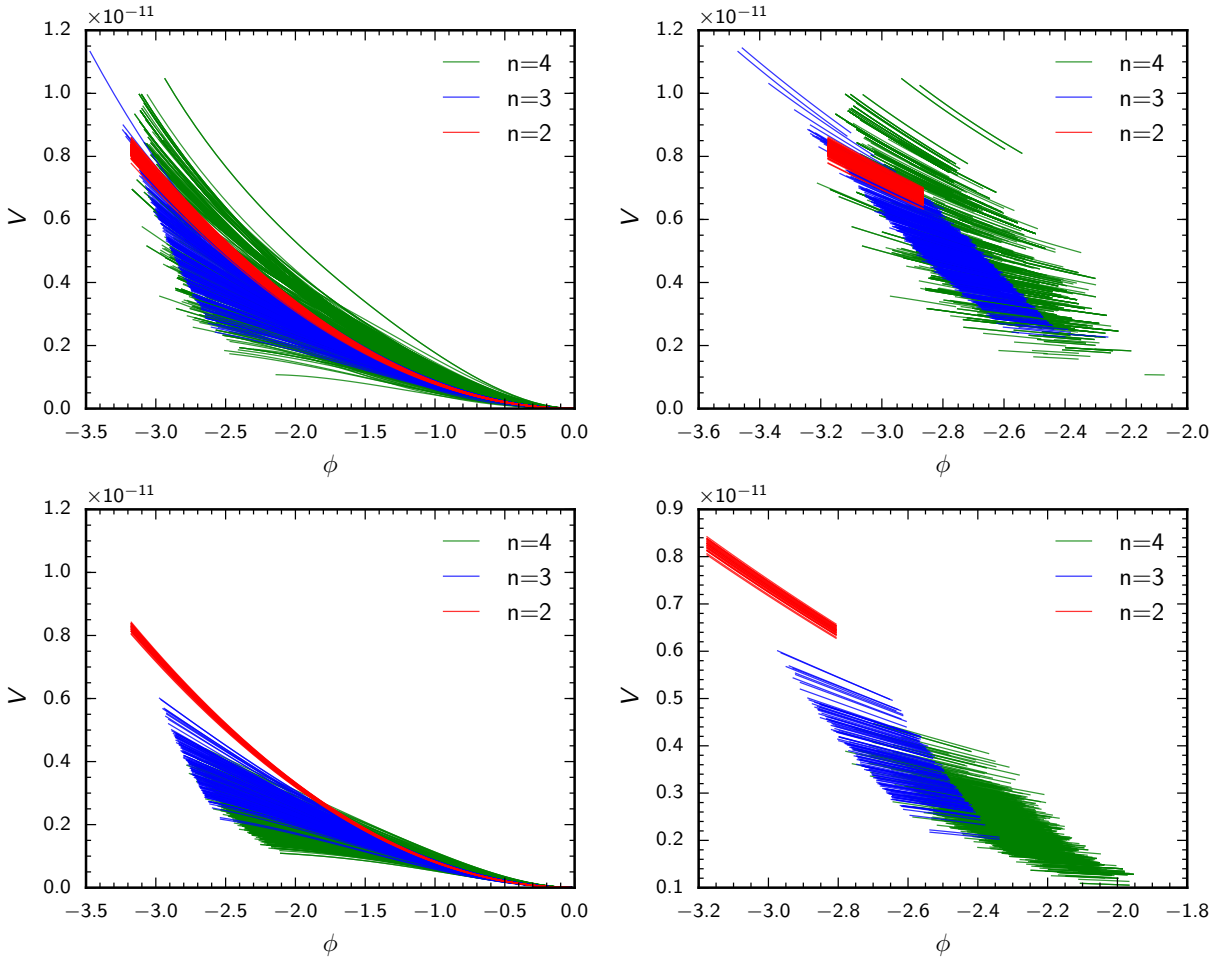


Figure 207: Representative sample of the inflaton potentials allowed at the 95 % CL, when the potential is Taylor-expanded at order $n = 2$, 3, and 4 and trusted until the end of inflation, and under the assumption of $N_* = 55$ e-folds of inflation between Hubble-radius crossing for the pivot scale and the end of inflation. In natural units (where $\sqrt{8\pi}M_{\rm Pl} = 1$). *Left panels:* Full potential from the beginning of the observable window till the end of inflation. *Right:* Zoom on the observable window directly constrained by inflation. *Top: Planck* TT,TE,EE+lowE. *Bottom: Planck* TT,TE,EE+lowE+lensing+BK15. Note that there is another branch of solutions that is symmetric under $\phi \to -\phi$.

28.5.4 Free-form potential reconstruction

As a complementary analysis to the previous three subsections, we next perform a free-form reconstruction of the inflationary potential with cubic splines, in a manner akin to the reconstructions of **(alias?)** and Sect. 28.6.2. Further plots and theoretical detail can be found in Handley et al. [125].

A free-form reconstruction usually proceeds by parameterizing the function of interest via a spline and taking the locations of the interpolation knots as free parameters in a posterior distribution. These are then varied along with any other model parameters, and then marginalized out to yield a model-independent reconstruction of the function of interest. The analysis is run for differing numbers of knots, N, and the Bayesian evidence is computed to allow for model comparison to determine how many knots are appropriate from the perspective of the data.

To reconstruct the inflationary potential $V(\phi)$, one cannot take a linear interpolating spline (as in Sect. 28.6.2), since the equations of motion in general depend on first (and sometimes second) derivatives of V. We therefore choose to parameterize the *second derivative* of the log-potential as a linear spline. The log-potential is computed by integrating this function twice,

n	2	3	4
$10^{12}V_2$	$1.631^{+0.022}_{-0.022}$	$1.81^{+0.12}_{-0.06}$	$1.86^{+0.25}_{-0.23}$
$10^{12}V_3$	...	$0.89^{+0.10}_{-0.03}$	$0.85^{+0.46}_{-0.64}$
$10^{12}V_4$	...	...	$0.044^{+0.26}_{-0.35}$
τ	$0.0518^{+0.0066}_{-0.0066}$	$0.0501^{+0.0078}_{-0.0069}$	$0.05628^{+0.0075}_{-0.0087}$
n_s	0.963	$0.9599^{+0.0034}_{-0.0018}$	$0.9656^{+0.0035}_{-0.0043}$
$10^3\frac{dn_\mathrm{s}}{d\ln k}$	$-0.6731^{+0.0005}_{-0.0005}$	$-0.534^{+0.079}_{-0.096}$	$-0.74^{+0.16}_{-0.13}$
$r_{0.002}$	0.136	$0.066^{+0.010}_{-0.016}$	$0.042^{+0.009}_{-0.014}$
$\Delta\chi^2_\mathrm{eff}$	...	$\Delta\chi^2_{3/2} = -13.18$	$\Delta\chi^2_{4/3} = -3.50$

Table 34: Numerical reconstruction of the potential parameters *beyond* any slow-roll approximation, when the potential is Taylor-expanded to nth order, trusted until the end of inflation, and using *Planck* high-ℓ TT,TE,EE+lowE+lensing+BK15. We also show the corresponding bounds on some related parameters (here n_s, $dn_\mathrm{s}/d\ln k$, and $r_{0.002}$ are derived from the numerically computed primordial spectra). All error bars are at the 68 % CL and all upper bounds at the 95 % CL. The effective χ^2 value of model n is given relative to model $n-1$.

yielding a function with two additional free parameters—a global offset and a gradient. Our reconstruction function is therefore

$$\ln V = \ln V_* + (\phi-\phi_*)\frac{d\ln V_*}{d\phi} + \int_{\phi_*}^{\phi} d\phi' \int_{\phi_*}^{\phi'} d\phi''\, \mathrm{LS}(\phi'';\theta), \tag{1280}$$

$$\theta = \left(\phi_1,\ldots,\phi_N, \frac{d^2\ln V_1}{d\phi^2},\ldots,\frac{d^2\ln V_N}{d\phi^2}\right), \tag{1281}$$

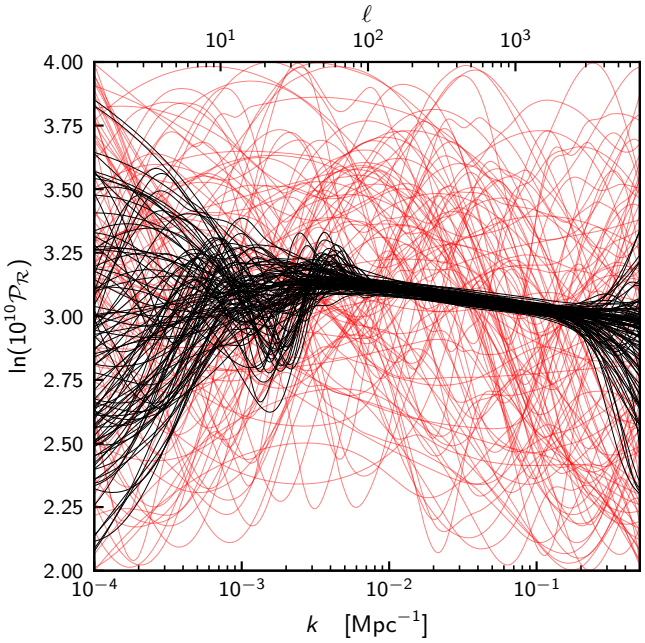
$$\mathrm{LS}(\phi;\theta) = \left\{ \frac{d^2\ln V_i}{d\phi^2}\frac{\phi-\phi_{i+1}}{\phi_i-\phi_{i+1}} + \frac{d^2\ln V_{i+1}}{d\phi^2}\frac{\phi-\phi_i}{\phi_{i+1}-\phi_i} \quad : \phi_i < \phi < \phi_{i+1} \right. . \tag{1282}$$

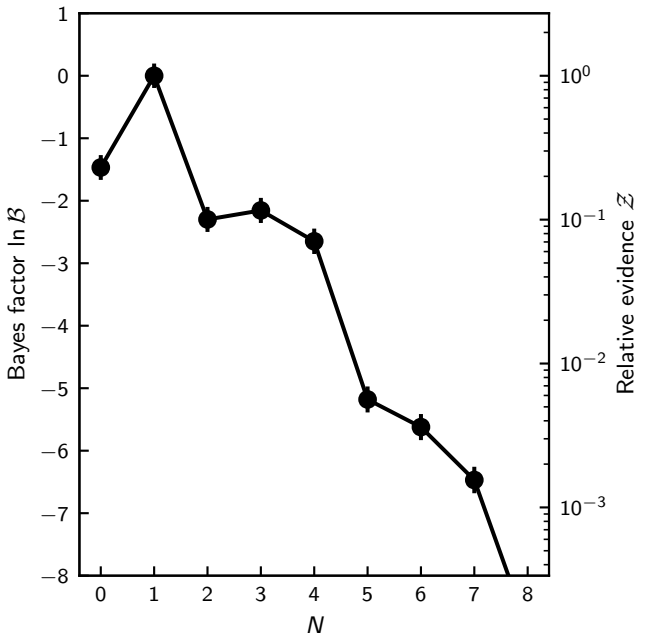
Here $\mathrm{LS}(\phi;\theta)$ is a standard linear spline dependent on N knots, $\ln V_*$ is the potential at the pivot scale, and $d\ln V_*/d\phi$ is the gradient of the log-potential at the pivot scale.

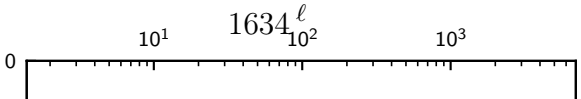
In general, any reconstruction of the potential will be sensitive only to the observable window of inflation in $\phi \in [\phi_\mathrm{min}, \phi_\mathrm{max}]$, where ϕ_min and ϕ_max are defined as the field values when the largest and smallest observable scales k_min and k_max exit the Hubble radius during inflation. Regions of the potential outside these ϕ values are unconstrained by current CMB data. In our analysis, we take $k_\mathrm{min} = 10^{-4}\,\mathrm{Mpc}^{-1}$ and $k_\mathrm{max} = 10^{-0.3}\,\mathrm{Mpc}^{-1}$, which encompasses the multipole range constrained by *Planck* (see Sect. 28.6.2). The locations $\phi_1,\ldots,\phi_N$ of the reconstruction knots should be distributed throughout this observable window. Whilst the locations $\phi_1,\ldots\phi_N$ and heights $d^2\ln V_1/d\phi^2,\ldots,d^2\ln V_N/d\phi^2$ themselves influence the size of the observable window, a reasonable approach is to first estimate it using the unperturbed potential (i.e., setting $N=0$). This gives an alternative range $\phi \in [\tilde{\phi}_\mathrm{min}, \tilde{\phi}_\mathrm{max}]$. The priors on all our variables are indicated in Table 35.

Parameters	Prior type	Prior range
N	Discrete uniform	$[0, 8]$
$\ln V_*$	Uniform	$[-25, -15]$
$d\ln V_*/d\phi$	Log-uniform	$[10^{-3}, 10^{-0.3}]$
$d^2\ln V_1/d\phi^2, \ldots, d^2\ln V_N/d\phi^2$	Uniform	$[-0.5, 0.5]$
$\phi_1, \ldots, \phi_N$	Sorted uniform	$[\tilde{\phi}_{\rm min}\,, \tilde{\phi}_{\rm max}]$
$\ln 10^{10}\mathcal{P}_\mathcal{R}(k)$	Indirect constraint	$[2, 4]$

Table 35: Parameters of the free-form potential reconstruction analysis and details of the priors. There is a further prior constraint in that we require that the inflaton should evolve in an inflating phase throughout the observable window, that the inflaton should be rolling downhill from negative to positive ϕ throughout, and that any primordial power spectra generated sit in the range $2 < \ln 10^{10}\mathcal{P}_\mathcal{R}(k) < 4$.







Our results are detailed in Fig. 208. The Bayesian evidence shows that the reconstruction preferred by the data is that using $N = 1$, corresponding to a constant non-zero $d^2 \ln V/d\phi^2$. This indicates that the *Planck* data do not significantly constrain the inflationary potential within the window any further than up to a quadratic term in a Taylor expansion. It is illuminating, however, to consider adding further structure to the potential, and Fig. 208 shows reconstructions for $N = 8$.

Considering the predictive posterior of the primordial power spectrum, we see that our parameterization is sufficient to exhibit the deficit at $\ell \simeq 30$, cosmic variance at low ℓ, and the loss of resolution at high ℓ, as seen in Sect. 28.6.2. Consistent with the rest of the analyses, ε_V is unconstrained, whilst the *Planck* data provide relatively powerful constraints on η_V within the observable window of inflation.

28.6 Primordial power spectrum reconstruction

This section reports results for the non-parametric reconstruction of the primordial scalar power spectrum using the new *Planck* 2018 likelihoods, as well as comparisons with the previously reported results for the *Planck* 2013 and 2015 releases. The objective here is to search for deviations from a simple power-law primordial power spectrum (i.e., $\mathcal{P}_\mathcal{R}(k) = A_\mathrm{s}(k/k_*)^{n_\mathrm{s}-1}$) in a manner that does not presuppose any particular theoretical model giving rise to such deviations. This work is complementary to the searches considered in Sect. 28.7, where particular functional forms for such deviations motivated by theory are investigated.

Here we apply three distinct nonparametric methods. In 2013 only the first method was used to reconstruct the primordial power spectrum, the so-called "penalized likelihood" method, for which the 2018 results are presented in Sect. 28.6.1. In 2015 two additional methods were also used: a linear spline method (discussed in Sect. 28.6.2) for which both the number of knots and their positions were allowed to vary, and ideas from Bayesian model selection were applied to determine the appropriate number of knots; and a method using cubic splines (discussed in Sect. 28.6.3). Although the discussion below includes some description of each method in order to make the paper self-contained, for details the reader is referred to the 2013 and 2015 papers. Here we specify only those details specific to the 2018 analysis or different from the choices in the 2013 and 2015 analyses. See references in **(alias?)** and Hunt & Sarkar [141], Hazra et al. [133], and Hunt & Sarkar [142] for other approaches to non-parametric reconstruction of the primordial power spectrum.

28.6.1 Penalized likelihood

The underlying idea of the penalized likelihood approach is to add a term to the log-likelihood that penalizes deviations from a perfect power-law spectrum. We parameterize the power spectrum as

$$\mathcal{P}_\mathcal{R}(k) = \mathcal{P}_0(k) \exp\bigl[f(k)\bigr], \tag{1283}$$

where $\mathcal{P}_0(k) = A_\mathrm{s}(k/k_*)^{n_\mathrm{s}-1}$, and add the following term to $-2\ln\mathcal{L}$:

$$\begin{aligned}\mathbf{f}^\mathrm{T}\,\mathbf{R}(\lambda,\alpha)\,\mathbf{f} \equiv \lambda \int_{\kappa_\mathrm{min}}^{\kappa_\mathrm{max}} d\kappa\, \left(\frac{\partial^2 f(\kappa)}{\partial\kappa^2}\right)^2 \\ + \alpha \int_{-\infty}^{\kappa_\mathrm{min}} f^2(\kappa) + \alpha \int_{\kappa_\mathrm{max}}^{+\infty} f^2(\kappa),\end{aligned} \tag{1284}$$

where $\kappa = \ln k$. The interval $[\kappa_\mathrm{min}, \kappa_\mathrm{max}]$ is chosen to cover the range over which the likelihood is able to constrain the data. The two α terms serve to pin the reconstruction to the simple power law where the data have almost no constraining power. One may imagine that $\alpha > 0$ should be infinite, but for numerical reasons a large but finite value is used to simplify the numerics. Numerically, for each λ the dimension of $\mathbf{f}$ is chosen to be so large that the continuum version of the penalty given in Eq. (1284) has been accurately approximated. For more details see [105] and the extensive references therein to prior literature, as well as **(alias?)** and **(alias?)**.

In Fig. 209 we show the results using *Planck* TT+lowE and in Fig. 210 we show the results for *Planck* TT,TE,EE+lowE. In both cases we have assumed the usual base-ΛCDM model specified in **(alias?)**, except that the power spectrum is now parameterized by a set of spline

points. In addition to these spline points, we also maximize the likelihood with respect to the dimensionless Hubble parameter, h, and the baryon, $\Omega_{\rm b}h^2$, and CDM, $\Omega_{\rm c}h^2$, densities. All other cosmological and nuisance parameters are the same as those quoted in **(alias?)**.

For the TT-only case, the maximum deviations are 1.55σ, 2.10σ, 1.80σ, and 1.65σ for $\lambda = 10^3$, 10^4, 10^5, and 10^6, respectively, for which the probabilities to exceed are 13 %, 28 %, 31 %, and 23 % (where we have taken into account the look-elsewhere effect). Similarly, for the TT,TE,EE case, the maximum deviations are 2.07σ, 1.77σ, 1.77σ, and 1.08σ for $\lambda = 10^3$, 10^4, 10^5, and 10^6, respectively, for which the probabilities to exceed are 29 %, 23 %, 32 %, and 25 %. We consequently find no statistically significant evidence for a deviation from the simple power-law hypothesis. This result is consistent with the results previously reported for the *Planck* 2013 and 2015 releases using essentially the same method. It is likewise consistent with the results below in Sects. 28.6.2 and 28.6.3, which use different methods.

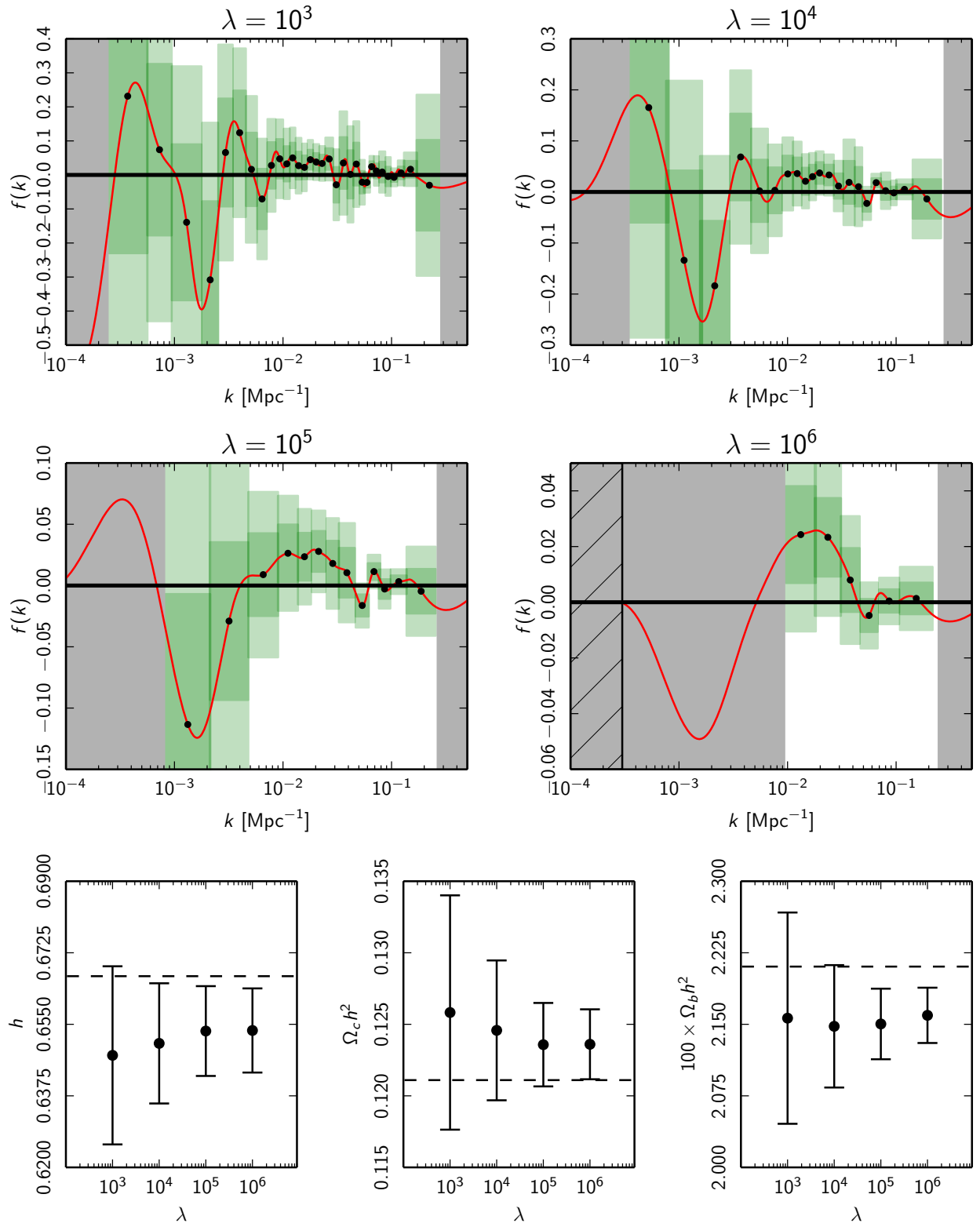


Figure 209: *Planck* TT+lowE penalized likelihood primordial power spectrum reconstruction. *Top four panels*: The deviation $f(k)$ for four different roughness penalties. The red curves indicate the best-fit deviation, while the vertical extents of the dark and light green error bars indicate the $\pm 1\sigma$ and $\pm 2\sigma$ errors, respectively. The width of the error bars indicates the minimum reconstructible width (the minimum width for a Gaussian feature such that the mean square deviation of the reconstruction is less than 10 %). The grey regions display where the minimum reconstructible width is undefined, meaning that the reconstruction in these regions is untrustworthy. The hatched region in the $\lambda = 10^6$ plot indicates where the fixing penalty has been applied. *Lower three panels:* $\pm 1\sigma$ error bars for the three non-primordial-specctrum cosmological parameters included in the reconstruction. The respective best-fit fiducial model values are indicated by the dashed lines.

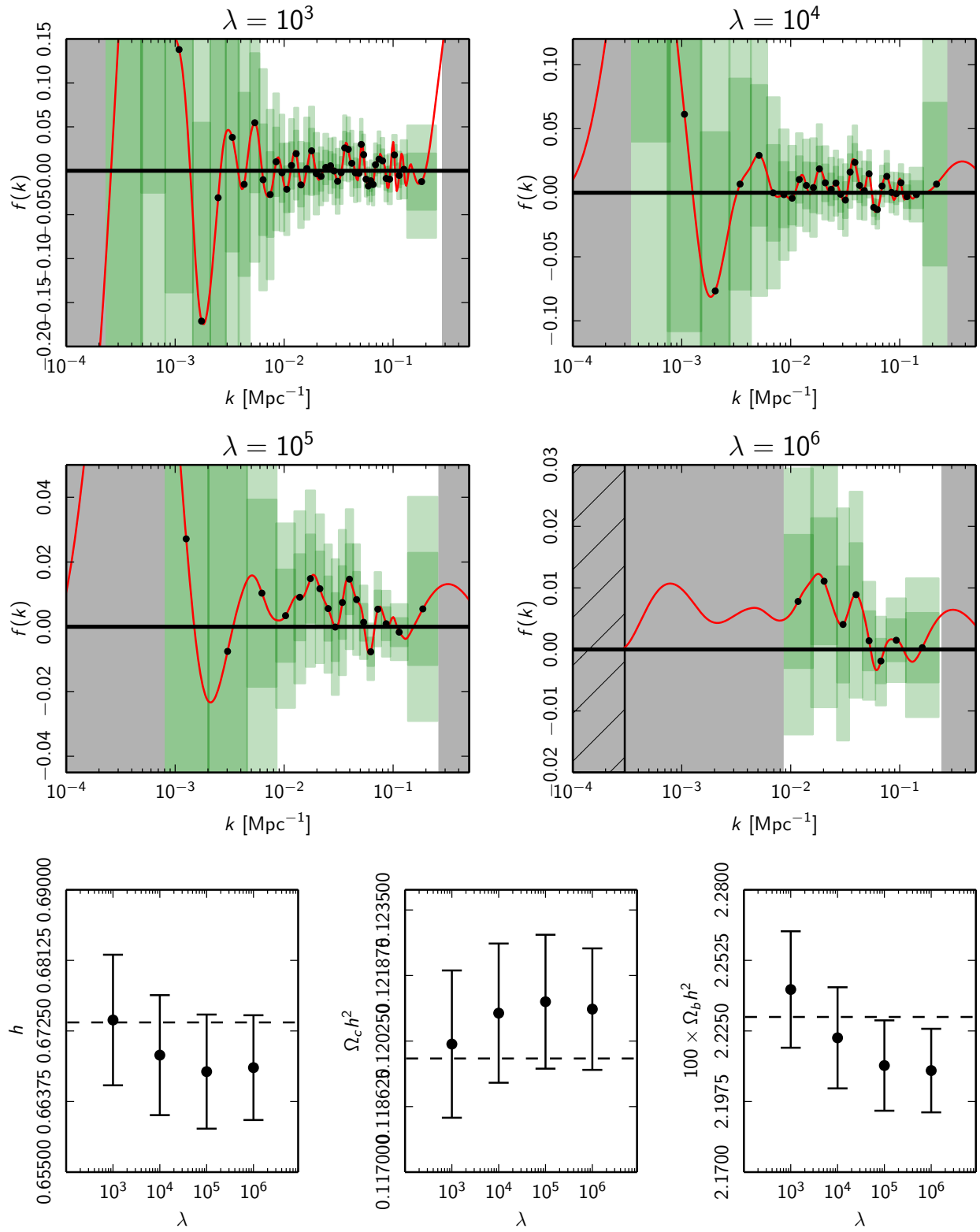


Figure 210: Penalized likelihood reconstruction, as Fig. 209 but for *Planck* TT,TE,EE+lowE.

28.6.2 Bayesian reconstruction

To reconstruct the primordial power spectrum of curvature perturbations, we follow the methodology of section 8.2 of **(alias?)**, using an N-point interpolating logarithmic spline with the positions of the knots considered as free parameters in the full posterior distribution. The positions of the points in the $(k, \mathcal{P})$ plane are treated as likelihood parameters with log-uniform priors. Further, the k-positions are sorted a priori such that $k_1 < k_2 < \cdots < k_N$, with k_1 and k_N fixed. We compute posteriors and evidence values (conditioned on N) using `PolyChord` [126, 127], also varying all cosmological and nuisance parameters. We then use evidence values for each model to correctly marginalize out the number of knots N.

To plot our reconstructions of $\mathcal{P}(k)$, we compute the marginalized posterior distribution of $\ln \mathcal{P}$ conditioned on k. The iso-probability confidence intervals are then plotted in the $(k, \mathcal{P})$ plane (see, e.g., Fig. 211), using code recorded in [124]. To quantify the constraining power of a given experiment, we use the conditional Kullback-Leibler (KL) divergence as exemplified by [135]. For two distributions $P(\theta)$ and $Q(\theta)$, the KL divergence is defined as

$$D_{KL}(P|Q) = \int \ln \left[\frac{P(\theta)}{Q(\theta)}\right] P(\theta) d\theta, \tag{1285}$$

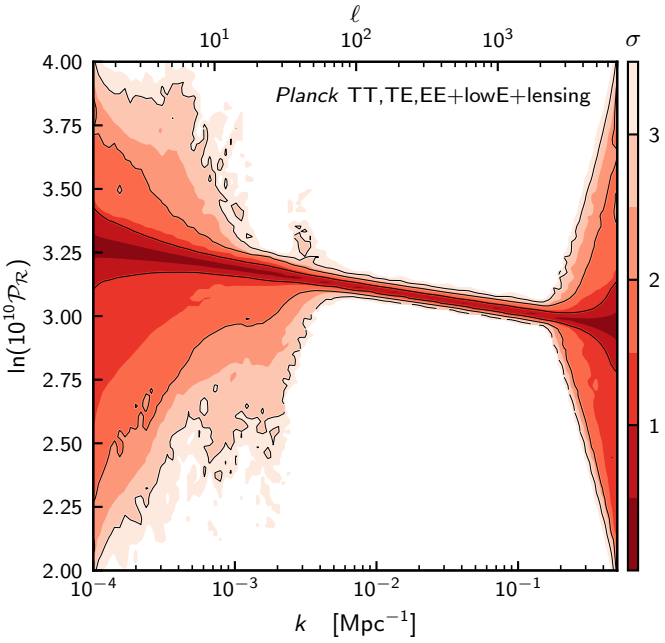
and may be interpreted as the information gain in moving from a prior Q to a posterior P [243]. For our reconstructions, we compute the KL divergence of each distribution conditioned on k and N, and then marginalize over N using evidence values to produce a k-dependent number which quantifies the compression or information that each data set provides at each value of k. Further plots and theoretical detail can be found in Handley et al. [125].

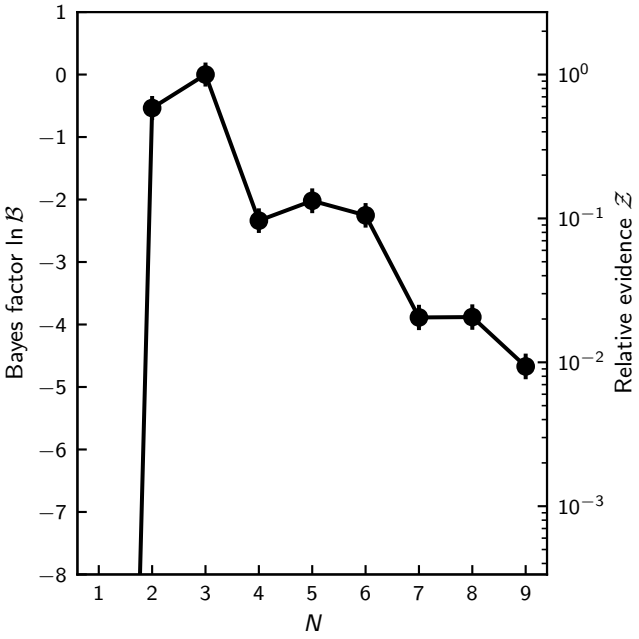
Update on *Planck* 2015

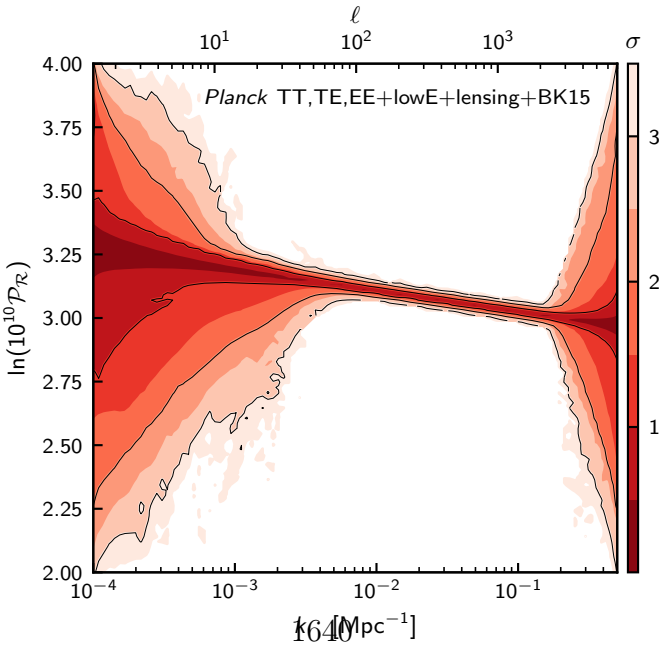
In **(alias?)**, our analysis focussed predominantly on the TT+lowTEB data set. Here we present results for TT,TE,EE+lowE+lensing. First, in updating to the lowE likelihood, we find that there is a marked tightening in the constraint on the amplitude of the reconstructed spectrum at all values of k. The improvement in the constraint can be seen directly in the predictive posterior plots (Fig. 211, top-left panel, and Fig. 212), and is quantified in Fig. 211 (bottom-right) via the KL divergence. The reason for the high-ℓ constraint provided by a low-ℓ likelihood change is due to the reduced uncertainty on τ that `SimAll` EE provides. This can be seen by examining the shifts in the underlying cosmological parameters in Fig. 193.

Upon adding TE and EE data, we find that the hint of a feature at $\ell \simeq 30$ is still present, in spite of the additional constraining power provided by polarization. Using polarization data, the $N = 3$ case is now the most strongly favoured model by the evidence criterion. This indicates that there is some scope for models which account for low-ℓ cosmic variance to be preferred in a Bayesian sense. The other underlying cosmological parameters are unaffected by the additional degrees of freedom in the primordial power spectrum provided by the reconstruction.

In order to combine *Planck* polarization data with BK15, we also allow the tensor-to-scalar ratio r to vary, and fix the tensor tilt n_t via the inflationary consistency condition. As can be seen in the bottom-left panel of Fig. 211, upon adding BK15, the effect of the low-ℓ deficit is softened, but with otherwise little change to the reconstruction. We repeated our analysis with `CamSpec` in place of `Plik` and found our results to be qualitatively and quantitatively unchanged.


ℓ
10^1
10^2
10^3
σ
Planck TT,TE,EE+lowE+lensing
ln(10^10 P_R)
k [Mpc^−1]


Bayes factor ln B
Relative evidence Z
N


ℓ
10^1
10^2
10^3
σ
Planck TT,TE,EE+lowE+lensing+BK15
ln(10^10 P_R)
k [Mpc^−1]

ℓ

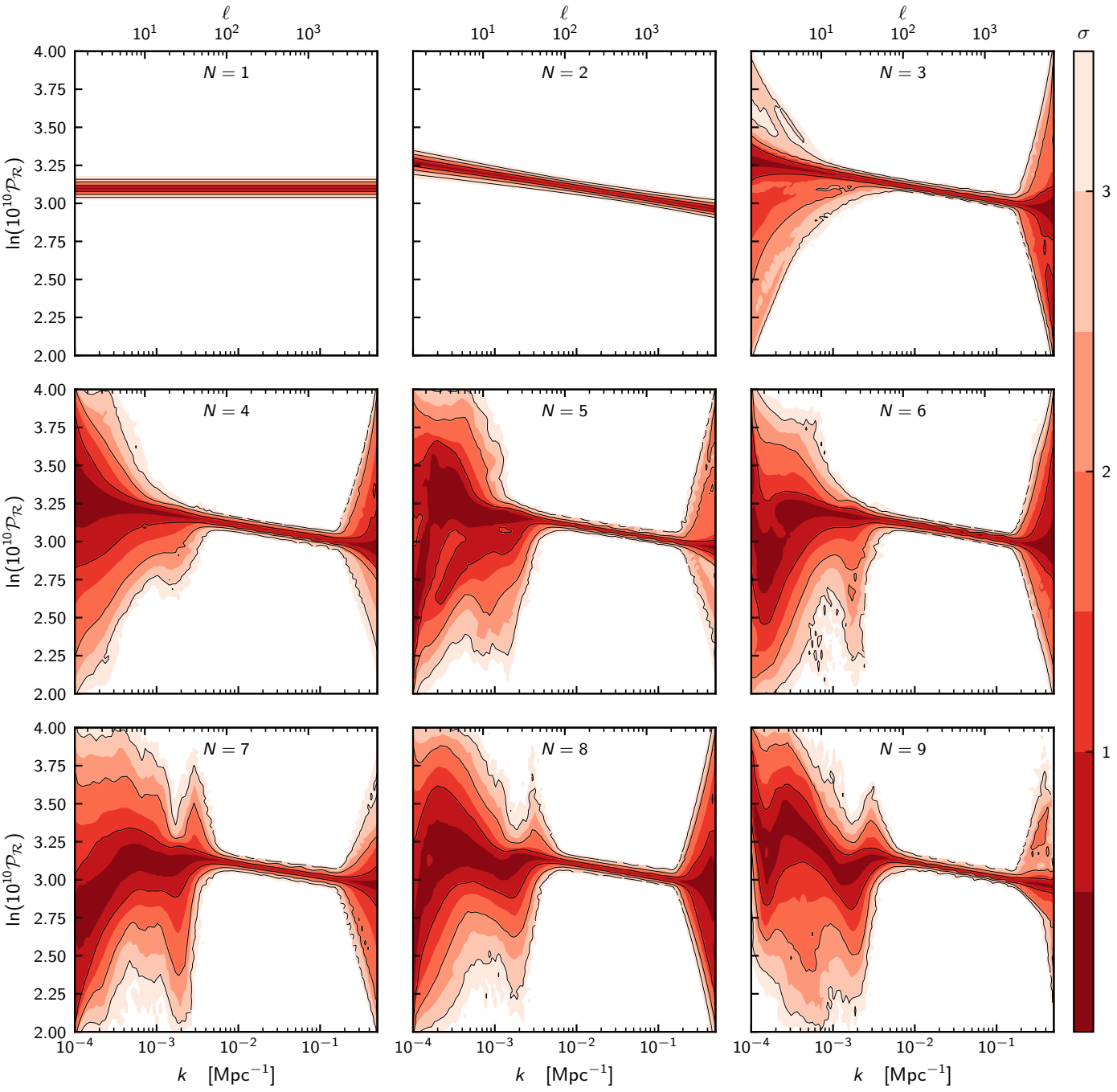


Figure 212: Free-form Bayesian reconstruction of the primordial power spectrum for varying numbers of knots (Sect. 28.6.2) using TT,TE,EE+lowE+lensing. The amplitude and tilt are consistent with the rest of the results with the same combination of likelihoods. As more knots are added, the $\ell \simeq 30$ feature in the C_ℓ temperature spectrum is visible as a dip to lower power.

Parameter	Prior	Range
$k_i, k_{\max}$	Sorted uniform	$0 < k_1 < \cdots < k_N < 0.2\,\mathrm{Mpc}^{-1}$
d_i	Uniform	$0 < d_1, \ldots, d_N < 0.01\,\mathrm{Mpc}^{-1}$
$k_i, k_{\max}$	Sorted log-uniform	$-4 < \log_{10} k_1 \cdots \log_{10} k_N < -0.3$
d_i	Log-uniform	$0 < \log_{10} d_1, \cdots, \log_{10} d_N < 1$
h_i	Uniform	$-1 < h_1, \ldots, h_N < 1$
$\ln(10^{10} A_\mathrm{s})$	Uniform	$2 < \ln(10^{10} A_\mathrm{s}) < 4$
n_s	Uniform	$0.8 < n_\mathrm{s} < 1.2$

Table 36: Priors for the search for sharp features in the primordial power spectrum. Units for k and d are Mpc^{-1}.

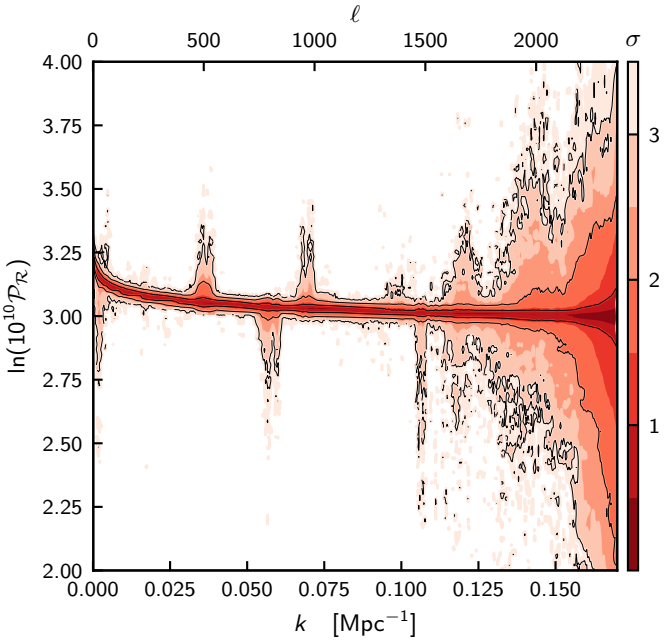
Free-form search for features

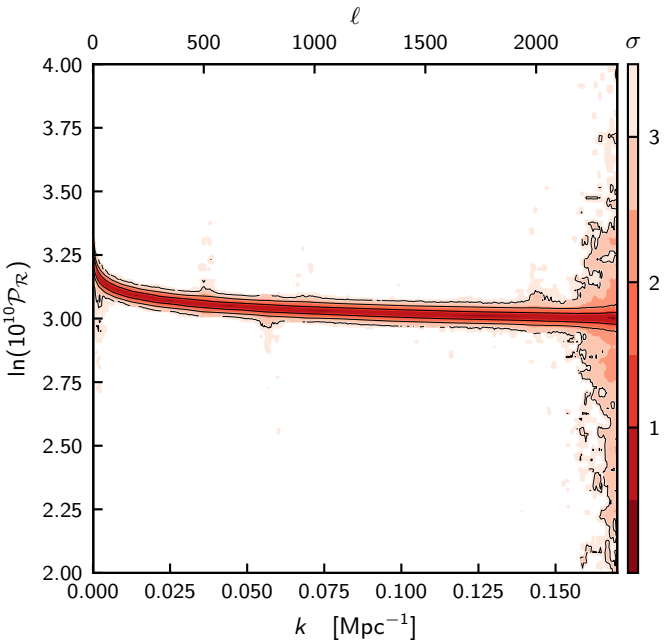
Next we examine the effect that sharp features in the primordial power spectrum can have on cosmological parameters. We model sharp features in the spectrum as a variable number of top-hat functions with varying widths, heights, and locations. On top of the traditional $A_\mathrm{s}, n_\mathrm{s}$ parameterization of the power spectrum, we place N sharp top-hat features into the spectrum at locations k_i with widths d_i and heights h_i $(i = 1, \ldots, N)$. That is, we set

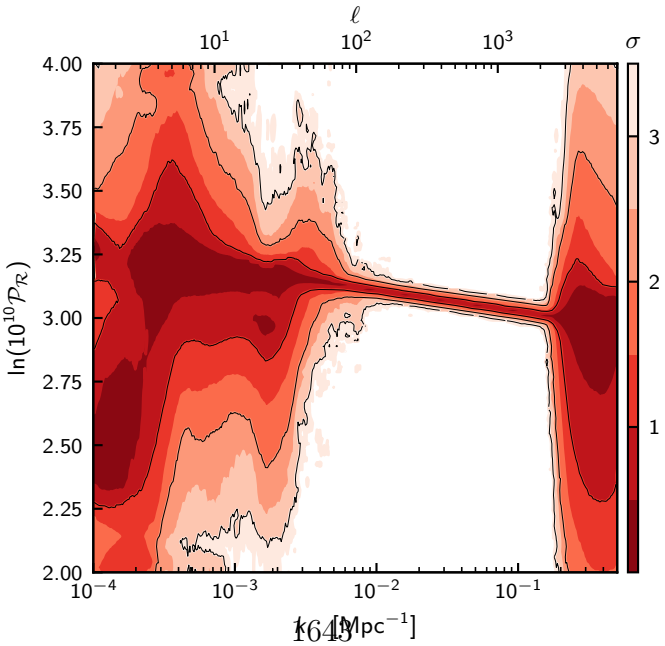
$$\ln \mathcal{P}_\mathcal{R}(k) = \ln A_\mathrm{s} + (n_\mathrm{s} - 1) \ln \left(\frac{k}{k_*} \right) + \sum_{i=1}^{N} h_i \left[|k - k_i| < \frac{d_i}{2} \right], \tag{1286}$$

where the square brackets in the summation denote a logical truth function as introduced by Graham et al. [114]. For values of $N = 0, \ldots, 8$, we treat the variables in parameterization (1286) as parameters in a posterior distribution along with the traditional cosmological and *Planck* nuisance parameters, with priors as detailed in Table 36. We run with both linear and logarithmic priors on the k-locations of the features, as this alters the sensitivity to the type of features uncovered. We sample the posteriors using `PolyChord` [126, 127].

Figures 213 and 214 show our results. With the linear priors case, there are statistically insignificant features corresponding to the peaks of the TT spectrum, which arise due to the enhanced cosmic variance at these locations. With the logarithmic priors case, a stronger but still statistically insignificant feature is detected at $\ell \simeq 30$, with a small deficit and surrounding enhancement of power. This case reproduces the results found in Sect. 28.6.2. In both cases, the Bayesian evidence shows preference for a no-features spectrum, and steadily declines as more features are added. The cosmological parameters remain unperturbed despite the introduction of features.


ℓ
0
500
1000
1500
2000
σ
4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.00
ln(10^10 P_R)
0.000
0.025
0.050
0.075
0.100
0.125
0.150
k [Mpc^-1]
3
2
1


ℓ
0
500
1000
1500
2000
σ
4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.00
ln(10^10 P_R)
0.000
0.025
0.050
0.075
0.100
0.125
0.150
k [Mpc^-1]
3
2
1


ℓ
10^1
10^2
10^3
σ
4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.00
ln(10^10 P_R)
10^-4
10^-3
10^-2
10^-1
k [Mpc^-1]
3
2
1

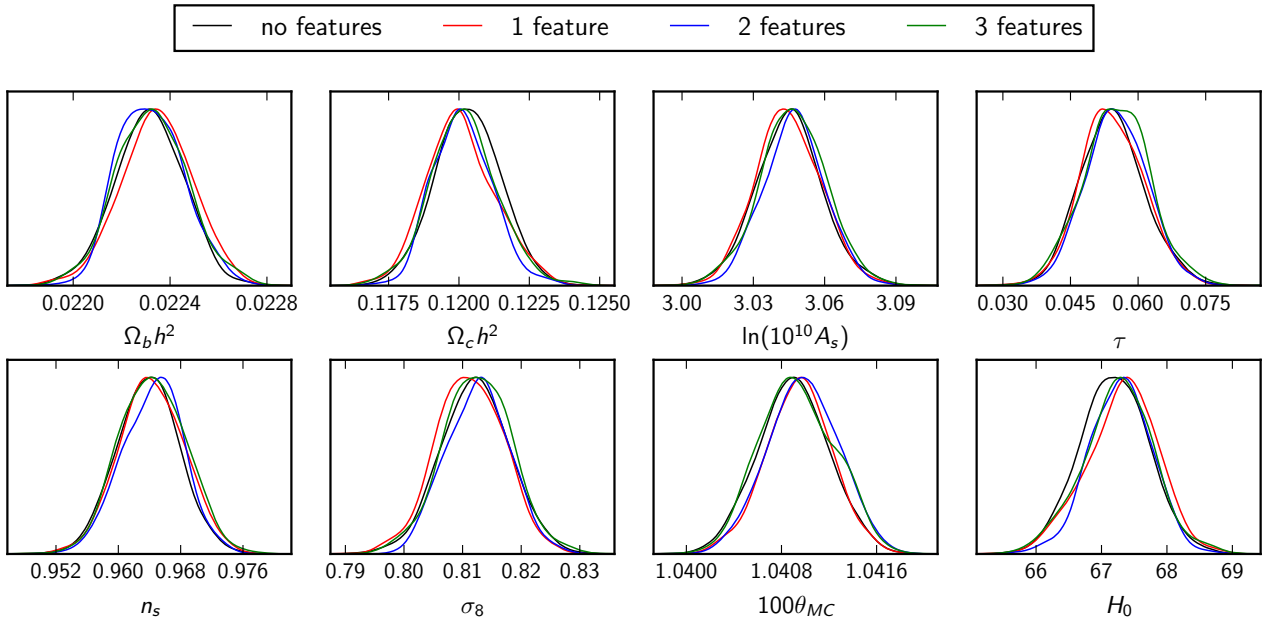


Figure 214: The effect on the underlying cosmological parameters of the free-form Bayesian search for features (Sect. 28.6.2), for $N = 0, \ldots, 3$ features with linear k-priors. The parameters remain stable up to $N = 8$ features, and when changing to logarithmic k-priors.

28.6.3 Cubic spline reconstruction

In this subsection we update the third method of reconstruction used in **(alias?)**, in which $\ln \mathcal{P}_\mathcal{R}(\ln(k))$ was expanded in cubic splines localized in $\ln(k)$ about uniformly spaced "knots," $\{\ln(k_b), b = 1, \ldots, N\}$, whose range was chosen to cover all relevant cosmological scales, from $10^{-4}\,\mathrm{Mpc}^{-1}$ to $O(1)\,\mathrm{Mpc}^{-1}$. We single out the standard scalar power spectrum pivot scale as a "pivot knot" $b{=}p$, with $k_p = k_* = 0.05\,\mathrm{Mpc}^{-1}$. Its associated power $\ln A_\mathrm{s} = \ln \mathcal{P}_\mathcal{R}(k_*)$ is assigned a uniformly distributed prior. A tilted primordial power spectrum $\mathcal{P}_{\mathcal{R},\mathrm{fid}} \equiv A_\mathrm{s}(k/k_*)^{n_{\mathrm{s,fid}}-1}$, with fixed spectral index $n_{\mathrm{s,fid}}$ is used as the fiducial baseline from which deviations are measured, expressed in terms of $N-1$ relative spectral shape parameters: $q_b = \ln\left(\mathcal{P}_\mathcal{R}(k_b)/\mathcal{P}_{\mathcal{R},\mathrm{fid}}(k_b)\right)$ for $b \neq p$. For the results presented here, $n_{\mathrm{s,fid}} = 0.967$ was chosen. As in **(alias?)** we continue to use cubic splines for the k-space modes we expand in, with natural boundary conditions (i.e., vanishing second derivatives at the first and last knots). The treatment here is therefore quite analogous to that in Sect. 28.5.4, where the inflaton potential rather than the curvature power spectrum is expanded in cubic splines. Knot numbers up to 18 were reported in **(alias?)**, and it was shown that 12 were sufficient to capture the variations desired by the *Planck* CMB data. The mode functions were also varied. For example, linear interpolation leads to similar reconstructions as long as enough knots are used. A weak uniform prior $(-1 \leq q_b \leq 1)$ was imposed on q_b. Outside of the spline coverage region $[k_1, k_N]$ we set $\ln\left(\mathcal{P}_\mathcal{R}(k)/\mathcal{P}_{\mathcal{R},\mathrm{fid}}(k)\right)$ to be q_1 for $k < k_1$ and q_N for $k > k_N$. The prior on q_b and boundary condition choices have little impact on the reconstructions over most of the k-range.

The current *Planck* TT,TE,EE+lowE+lensing+BK15 data give only upper limits to the allowed values of the tensor amplitude, $r < 0.06$. Consequently, adding shape degrees of freedom to the tensor power spectrum would yield a completely prior-driven result. Instead we adopt the standard model power-law parameterization for tensors, $\mathcal{P}_\mathrm{t}(k) = rA_\mathrm{s}(k/k_*)^{n_\mathrm{t}}$, with the tensor spectral index constrained by the consistency relation $n_\mathrm{t} = -r/8$. Without B-mode constraints and with enough knots one could deform the primordial scalar spectrum to mimic a tensor contribution to the CMB power. However, this near-degeneracy is broken with direct B-mode observations, effectively so even if there are only upper limits as for the BK15 data. Our reconstructions here focus on letting r float over a prior range $0 \leq r \leq 1$, but the posterior is strongly constrained by the BK15 data.

The joint probability distributions of $\{q_b, b \neq p\}$, $\ln A_\mathrm{s}$, and the other cosmic and nuisance parameters are determined by `CosmoMC` modified to incorporate the N-knot parameterization

for fixed knot number N. Figure 215 shows the reconstruction. Apart from the mean and 1σ and 2σ limits on the ensemble of trajectories allowed by the posterior probability, we also show a set of individual trajectories with parameters taken from 1σ samples to illustrate the knot-to-knot coherence (dashed curves). The tensor trajectories are straight lines, as required by the adopted tensor power model.

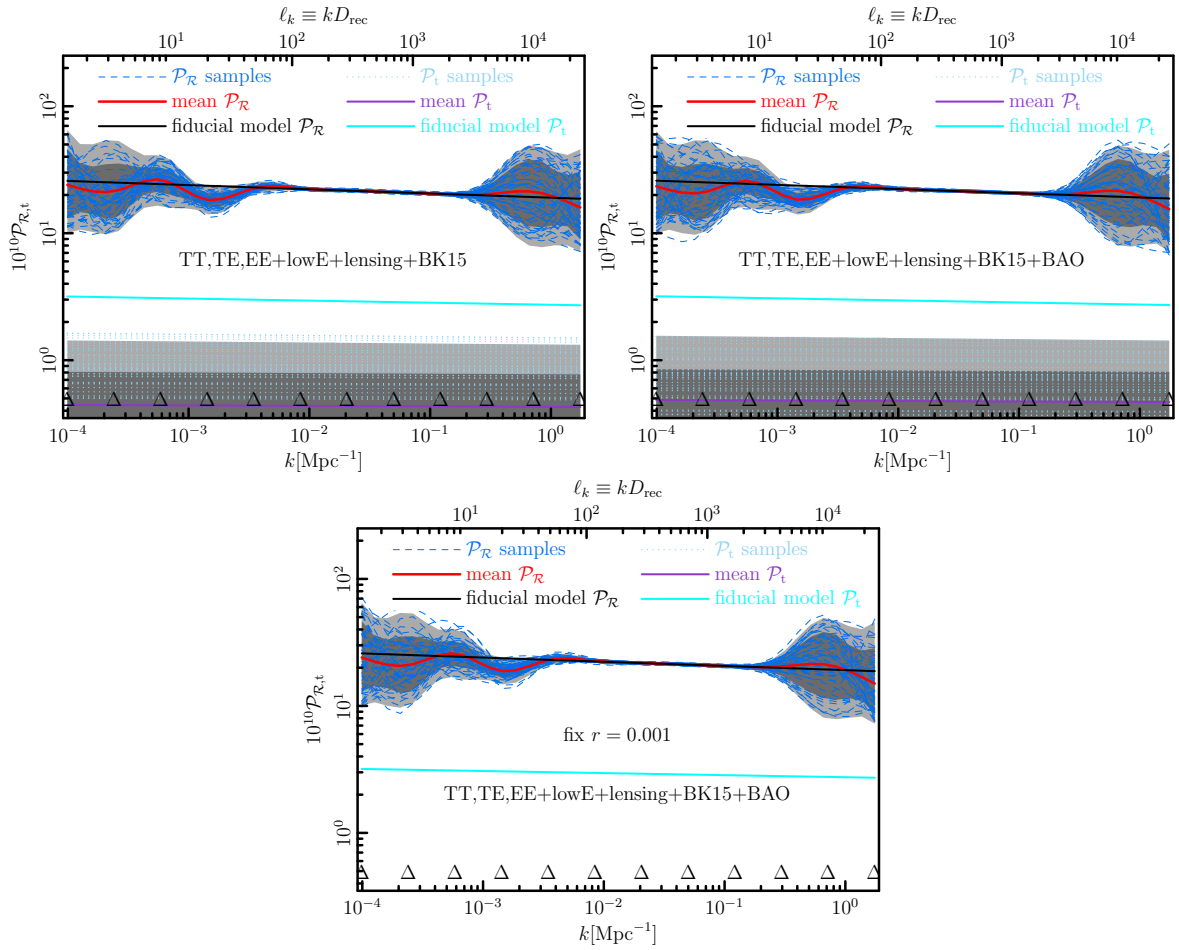


Figure 215: Reconstructed primordial scalar power spectrum derived using *Planck* TT,TE,EE+lowE+lensing+BK15 data and 12 knots for the cubic spline interpolation (with positions marked as Δ at the bottom of each panel). Mean (ensemble-averaged) spectra are heavy lines, allowed $\pm1\sigma$ and $\pm2\sigma$ regions for trajectories are the shaded regions, and the dashed lines denote selected trajectories with parameters sampled within the $\pm1\sigma$ posterior. Below the scalar power is the tensor power reconstruction. The addition of the BAO likelihood shown in the middle panel makes almost no visual difference to the reconstructions. In the bottom panel, fixing the tensor-to-scalar ratio to $r = 0.001$ also produces only small differences in reconstruction. Knot positions in k roughly translate to multipoles through $kD_{\rm rec}$, where $D_{\rm rec}$ is the comoving distance to recombination.

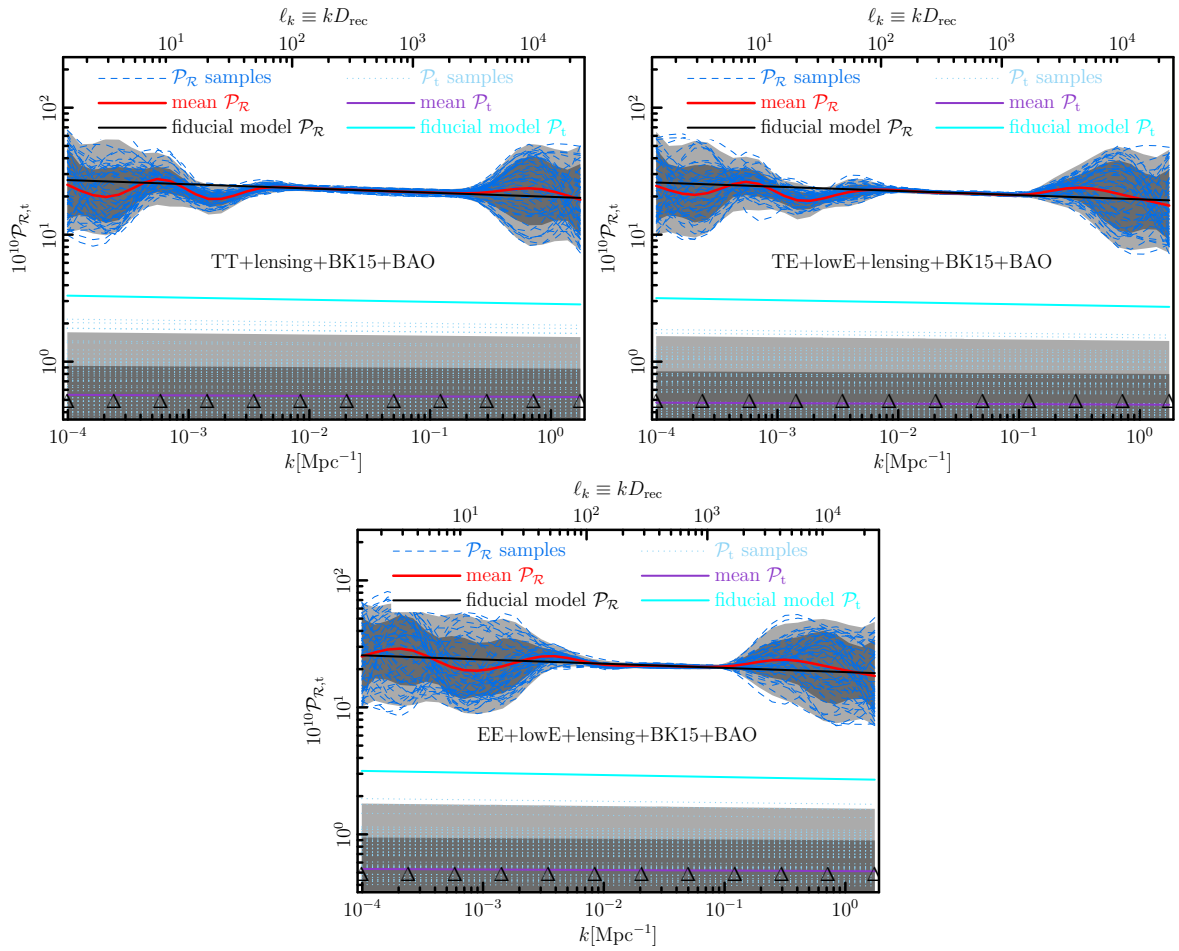


Figure 216: Reconstructed 12-knot power spectra. The robustness of the reconstruction is apparent when sub-selections of the *Planck* data are used: *Planck* TT+lowE+lensing+BK15 (top); *Planck* TE+lowE+lensing+BK15 (middle); and *Planck* EE+lowE+lensing+BK15 (bottom).

In spite of the extra scalar shape freedom in the k-space region over which the tensor modes affect the CMB, the 12 knot reconstruction still leads to a strong constraint of $r < 0.069$, rather close to the $r < 0.06$ limit obtained if the only shape parameter is $n_{\rm s}$. In fact we find that the current limits on r are such that the scalar-power reconstructions are not sensitive to the details of the r distribution. To illustrate this, the lower panel of Fig. 215 shows the spectrum when r is fixed at the tiny value of $r = 0.001$. One could regard this as a theoretical prior for low-energy inflation models or a forecast for a future in which r is measured or tightly constrained by B-mode experiments.

In **(alias?)**, the main cubic spline reconstruction included non-CMB data to help pin down cosmological parameters such as H_0, τ, and the late-time expansion history. The improvements in the data from 2015 to 2018, especially the decreased errors on τ, result in no non-CMB data being needed. Although τ and $\ln A_{\rm s}$ are about 90 % correlated, as they are in the standard power-law model, neither are very correlated with the q_b. The strongest is about 40 % for q_3 at $k \simeq 0.0006\,{\rm Mpc}^{-1}$, corresponding to $\ell \sim 10$ where reionization is kicking in. The second strongest is about 30 % for $k \simeq 0.02\,{\rm Mpc}^{-1}$, similar to the correlation of τ and $n_{\rm s}$ in the standard power-law model. (The correlations among the q_b are also relatively small, except for the high k bands $b = 10$ and 11, where the data are not constraining.)

The middle panel in Fig. 215 shows the effect of adding the BAO constraint. Although apparently visually identical, there are slight differences. For example, the 1σ error on q_b at $k \simeq 0.02\,{\rm Mpc}^{-1}$ decreases by about 7 %, from 0.0090 to 0.0084, while at $k \simeq 0.1\,{\rm Mpc}^{-1}$ the decrease is about 5 %. The restriction to $r = 0.001$ does not change the error bars over the

floating r case. At intermediate k for modes 5 to 8 the errors on q_b are so close to zero that the reconstruction is quite compatible with a simple power law, corresponding to a straight line in Fig. 215. This was also a main result of the 2015 *Planck* reconstructions.

The errors on the q_b grow above ± 0.1 for $b = 4$ as a consequence of increased cosmic variance, giving more freedom in the allowed trajectories. Unfortunately this is also the region of relevance to the TT power spectrum deficit in the $\ell \simeq 20$–30 range. The most significant deviation from zero occurs for q_4 at $k \simeq 0.0014\,\mathrm{Mpc}^{-1}$: -0.254 ± 0.127, -0.255 ± 0.125, and -0.235 ± 0.128 for the three cases. Thus the anomaly in terms of deviation from the power law of the standard model hovers at around the 2σ level. More precisely, the 95 % upper confidence limits on q_4 are -0.011, -0.018, and $+0.017$, for the respective cases. This 2σ level of the anomaly was also the conclusion of the 2015 *Planck* reconstructions. Therefore, even though the low-k deficit is robust against the various choices for the reconstruction, we conclude that it is not statistically significant. The associated TT, TE, EE, and BB power spectra responses to the allowed primordial power variations are derived from the mode expansion, and match the $\mathcal{D}_\ell^{XY}$ data well, in particular following the dip in TT at $\ell \simeq 20$–30.

In Fig. 216 we show the reconstructed power spectra using only the TT, TE, and EE data in conjunction with BK15. The fixed $r = 0.001$ cases look very similar. Except at high k, the polarization data using either EE or TE also enforce a nearly uniform $n_\mathrm{s}(k)$ over a broad range in k, with values in excellent agreement with those obtained from TT alone, from TE and EE used in combination, and from the combined TT,TE,EE results. For example, the ± 0.0087 and ± 0.0060 1σ errors at $k \simeq 0.02\,\mathrm{Mpc}^{-1}$ and $k \simeq 0.1\,\mathrm{Mpc}^{-1}$, respectively, increase only slightly for TT only, to ± 0.012 and ± 0.0068, but, more significantly, to ± 0.017 and ± 0.069 with EE alone. The deficit region remains about the same, with the TT,TE,EE result for q_4 of -0.255 ± 0.125 quoted above changing slightly for TT alone, to -0.252 ± 0.130, but with no hint of an anomaly for EE alone, at -0.126 ± 0.460. If just the TE cross data are used, the values are closer to the TT case, namely, -0.232 ± 0.163, now under a 2σ excursion from the tilted fiducial model.

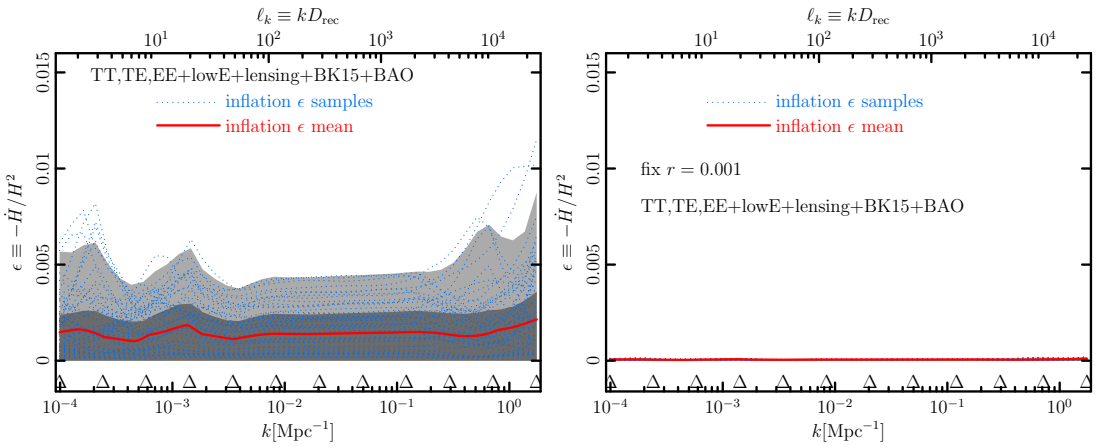


Figure 217: Acceleration history $\epsilon(k)$ for reconstructed trajectories using 12 knots (marked as Δ at the bottom of the figure), with cubic-spline interpolation and the *Planck* TT,TE,EE+lowE+lensing+BK15+BAO data for the two cases of floating r and r fixed at 0.001. Sample 1σ trajectories for the floating r case allow wide variability, which is naturally greatly diminished if r is fixed to $r = 0.001$.

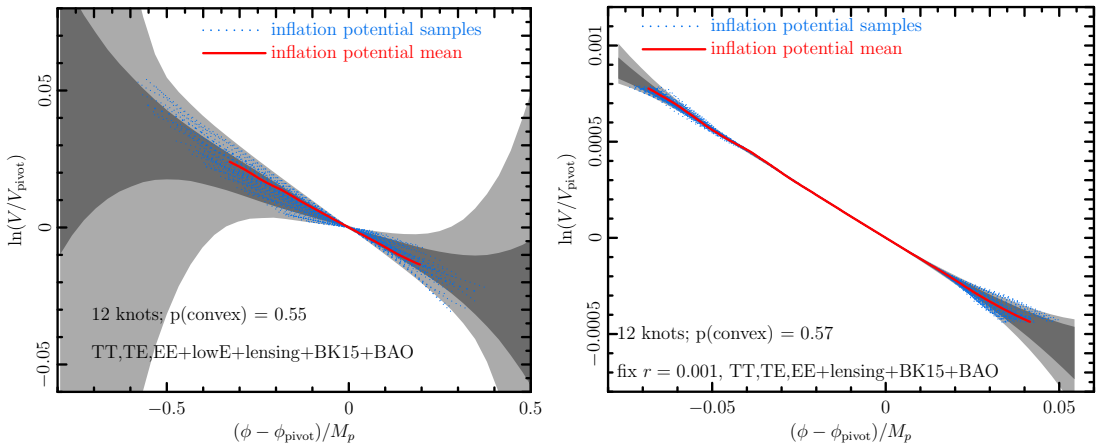


Figure 218: *Top:* Reconstructed shape of the single-field inflaton potential from the cubic-spline power spectra mode-expansion using 12 knots and the *Planck* TT,TE,EE+lowE+lensing+BK15+BAO data. *Bottom:* Result when r is fixed at 0.001. Instead of plotting as a function of wavenumber k we plot $\ln V(\phi)/V_{\rm pivot}$ about a pivot field value $\phi_{\rm pivot}$. Note that the range on the ϕ axis is quite different for the small r case than the floating case. The probability of local convexity evaluated at $\phi_{\rm pivot}$ is denoted as p(convex).

As in **(alias?)**, we can use the $\mathcal{P}_\mathcal{R}(k) \propto H^2/\epsilon$ and $\mathcal{P}_t(k)/\mathcal{P}_\mathcal{R}(k) \simeq 16\epsilon$ reconstructions to get an idea of the history of the acceleration of the Universe as a function of time over the significant number of e-folds of the cosmic expansion that the CMB data probe, codified by the dynamical slow-roll parameter $\epsilon(k) = -\dot{H}/H^2|_{k=aH}$, considered as a function of aH, the value of the wavenumber at Hubble crossing. Results with floating r and r fixed to 0.001 are shown in Fig. 217. For the dynamical time variable we use $k = aH$ for the horizontal axis for ease in comparing with the $\mathcal{P}_\mathcal{R}(k)$ curves of Fig. 215. The wide spread in the ϵ trajectories for the floating r case is a consequence of being able to fit the $n_{\rm s}(k)$ shape by a combination of $\epsilon(k)$ and $d\ln\epsilon(k)/d\ln k$. When r is small, $n_{\rm s}(k)$ is almost entirely determined by $d\ln\epsilon/d\ln k$, and the $\epsilon(k)$ values cluster near $r/16$.

The Hamilton-Jacobi energy constraint equation relates the potential to ϵ and H via $V = 3M_P^2H^2(1-\epsilon/3)$. Figure 218 shows the reconstructed inflationary potential shapes in the region over which the allowed inflationary potentials are constrained by the data for the floating r and fixed r cases. Instead of using k for the horizontal axis, we translate into inflaton-field ϕ-space using the relation between ϕ and $\sqrt{\epsilon}$, referenced to the pivot position $\phi_{\rm pivot}$. For the vertical axis we plot $\ln V/V_{\rm pivot}$, with the overall normalization $V_{\rm pivot}$ removed. Its value is set by r, hence there is a distribution of constant $V_{\rm pivot}$ amplitudes to superimpose if we want the total V. The radically different visual appearance for the floating r and fixed r cases is due to the observable k range being compressed through the smallness of ϵ into a small precisely determined field range, whereas this range has a distribution in the floating r case. One can monitor whether the shapes of the individual realizations of the potential trajectories bend upwards or downwards or do both, an indication of convexity. The sample trajectories shown are not exclusively convex or concave, and a measure of the probability that they are convex can be made from the ensemble. As indicated in Fig. 218 for the 12 knot case, the ensemble-averaged potentials are roughly exponential, with individual trajectories bending away from the mean, but with no strong tendency for convexity or concavity. (The roughly 50 % probability changes somewhat depending upon the combination of data used, whether TT,TE,EE or the individual data sets.)

The standard cosmological parameter determinations are highly robust to the addition of these spline shape degrees of freedom. The mean values change little and the error bars grow slightly, by around 10 % for $\ln A_{\rm s}$, τ, and H_0. The largest error increase is for σ_8, with $\sigma_8 = 0.812 \pm 0.0058$ becoming 0.814 ± 0.0096. The main conclusions of this section on ϵ and V, and $\mathcal{P}_\mathcal{R}(k)$, remain as in **(alias?)**, but the results have been noticeably sharpened by the improvements in the *Planck* 2018 data sets.

28.7 Search for primordial features in the *Planck* power spectrum

The "bottom-up" power spectrum reconstruction methods of the previous section are an excellent way to search for coarse features in the spectrum, but lack the resolution to detect the higher-frequency features generically predicted by various physical mechanisms [see, e.g., [59] for a review]. It is therefore useful to complement power spectrum reconstruction with a "top-down" approach by fitting specific feature models to the data. In this section we will analyse a representative range of power spectrum templates which parameterize features in terms of a handful of new parameters.

With *Planck*'s temperature and polarization data, we have two essentially independent probes of features at our disposal and will pay particular attention to examining the consistency between the two [197].

28.7.1 Power spectrum templates with features

	Log osc	Running log osc	Lin osc	Step	Kin cutoff	Rad cutoff	Kink cuto
$\mathcal{A}_X$	[0,0.5]	[0,0.5]	[0,0.5]	...	...	...	...
$\log_{10}\omega_X$	[0,2.1]	[0,2.1]	[0,2]	...	...	...	...
$\varphi_X/(2\pi)$	[0,1]	[0,1]	[0,1]	...	...	...	...
$\alpha_{\rm rf}$	...	[−0.1,0.1]	...	...	...	...	...
$\mathcal{A}_{\rm s}$	...	...	...	[0,1]	...	...	...
$\log_{10}\left(k_{\rm s}/{\rm Mpc}^{-1}\right)$	...	...	...	[−5,−1]	...	...	...
$\ln x_{\rm s}$	...	...	...	[−1,5]	...	...	...
$\log_{10}\left(k_Y^{\rm c}/{\rm Mpc}^{-1}\right)$	...	...	...	...	[−5,−3]	[−5,−3]	[−5,−3]
$R_{\rm c}$	...	...	...	...	...	...	[−1,0.7]

Table 37: Prior ranges for the parameters of the feature model templates of Sect. 28.7.1.

Global oscillation models

Periodic or quasi-periodic modulations of the power spectrum which extend over the entire observable range of wavenumbers can occur in a variety of models [cf., e.g., Danielsson [66], Martin & Brandenberger [180], Bozza et al. [42], Chen [56], Jackson & Shiu [145]]. A general parameterization of models with a sinusoidal modulation of the primordial power spectrum reads

$$\mathcal{P}_{\mathcal{R}}^{X}(k) = \mathcal{P}_{\mathcal{R}}^{0}(k)\left[1 + \mathcal{A}_X \cos\left(\omega_X \Xi_X(k) + \varphi_X\right)\right], \tag{1287}$$

where $X \in \{\log, \mathrm{lin}, \mathrm{rf}\}$. Defining $\kappa \equiv k/k_*$,[174] we consider the logarithmic oscillation model, given by $\Xi_{\log} \equiv \ln\kappa$, and the linear oscillation model, $\Xi_{\rm lin} \equiv \kappa$. In addition, we investigate a logarithmic model with running frequency, $\Xi_{\rm rf} \equiv \ln\kappa\,(1 + \alpha_{\rm rf}\ln\kappa)$. For $0 \leq \alpha_{\rm rf} \lesssim 0.01$, this is a good approximation for the scalar power spectrum in the axion monodromy model [96], which will be analysed in more detail below, but here we allow for a wider range of the running parameter $\alpha_{\rm rf}$, including negative values (i.e., decreasing frequency with increasing k).

Localized oscillatory features: inflation with a step

A sudden transient event in the evolution of the inflation field, triggered by a sharp feature in the inflaton potential, or a sharp turn in field space, generically leads to a localized oscillatory feature in the power spectrum [8, 57, 5, 196, 19]. As an example of this class of feature models,

[174] Throughout this section, we take $k_* = 0.05\,{\rm Mpc}^{-1}$.

we consider here the case of a tanh-step in an otherwise smooth inflaton potential [8], whose power spectrum can be parameterized as [195]

$$\ln \mathcal{P}^{\rm s}_{\mathcal{R}}(k) = \ln \mathcal{P}^{0}_{\mathcal{R}}(k) + \mathcal{I}_0(k) + \ln\left(1 + \mathcal{I}_1^2(k)\right), \tag{1288}$$

where the first- and second-order terms are given by

$$\mathcal{I}_0 = \mathcal{A}_{\rm s}\, \mathcal{W}_0(k/k_{\rm s})\, \mathcal{D}\left(\frac{k/k_{\rm s}}{x_{\rm s}}\right), \tag{1289}$$

$$\mathcal{I}_1 = \frac{1}{\sqrt{2}}\left[\frac{\pi}{2}(1-n_{\rm s}) + \mathcal{A}_{\rm s}\, \mathcal{W}_1(k/k_{\rm s})\, \mathcal{D}\left(\frac{k/k_{\rm s}}{x_{\rm s}}\right)\right], \tag{1290}$$

with window functions

$$\mathcal{W}_0(x) = \frac{1}{2x^4}\left[\left(18x - 6x^3\right)\cos 2x + \left(15x^2 - 9\right)\sin 2x\right], \tag{1291}$$

$$\mathcal{W}_1(x) = -\frac{3}{x^4}(x\cos x - \sin x)\left[3x\cos x + \left(2x^2 - 3\right)\sin x\right], \tag{1292}$$

and damping function

$$\mathcal{D}(x) = \frac{x}{\sinh x}. \tag{1293}$$

In this model, the parameter $\mathcal{A}_{\rm s}$ determines the amplitude of the oscillatory feature, the step scale $k_{\rm s}$ sets the position of the step in k-space, and the damping parameter $x_{\rm s}$ determines the width of the envelope function.

Models with suppressed power at large scales

The apparent lack of power at the largest scales in the temperature power spectrum with respect to the expectation of base ΛCDM serves as a motivation for models with a suppression of primordial perturbations below a cutoff scale $k^{\rm c}$. Physically, this effect may be due to fluctuations at the largest observable scales being generated at the onset of the inflationary phase after a prior era of, e.g., kinetic or radiation domination [278, 63], or due to an isolated event such as a kink in the inflaton potential [262].

In these scenarios, the primordial spectrum can generally be analytically approximated by an expression of the form

$$\ln \mathcal{P}^{Y}_{\mathcal{R}}(k) = \ln \mathcal{P}^{0}_{\mathcal{R}}(k) + \ln \Upsilon_Y(k/k^{\rm c}_Y), \tag{1294}$$

with $Y \in \{\mathrm{kin}, \mathrm{rad}, \mathrm{kink}\}$, where Υ_Y is a function with $\ln \Upsilon_Y \to 0$ in the limit $k \gg k^{\rm c}_Y$ that describes the shape of the cutoff and the transition to a power-law spectrum at smaller scales.

Initial kinetic domination

If inflation is preceded by an era dominated by the kinetic energy of the inflaton field (i.e., fast roll), we have

$$\Upsilon_{\rm kin}(y) = \frac{\pi}{16}\, y\, |C_{\rm c}(y) - D_{\rm c}(y)|^2, \tag{1295}$$

with

$$C_{\rm c}(y) = e^{-iy}\left[H_0^{(2)}\left(\frac{y}{2}\right) - \left(\frac{1}{y} + i\right) H_1^{(2)}\left(\frac{y}{2}\right)\right], \tag{1296}$$

$$D_{\rm c}(y) = e^{iy}\left[H_0^{(2)}\left(\frac{y}{2}\right) - \left(\frac{1}{y} - i\right) H_1^{(2)}\left(\frac{y}{2}\right)\right], \tag{1297}$$

where $H_n^{(2)}$ denotes the Hankel function of the second kind [63].

Initial radiation domination

If inflation begins immediately after a radiation-dominated phase, the cutoff function reads [278]

$$\Upsilon_{\rm rad}(y) = \frac{1}{4y^4}\left|e^{-2iy}\,(1+2iy) - 1 - 2y^2\right|^2 . \tag{1298}$$

Kink in the inflaton potential (Starobinsky model)

A kink in the inflaton potential, first discussed by [262], leads to a spectrum approximately given by

$$\begin{aligned}\Upsilon_{\rm kink}(y) = 1 - 3(R_{\rm c}-1)\frac{1}{y}\left[\left(1-\frac{1}{y^2}\right)\sin 2y + \frac{2}{y}\cos 2y\right] \\ + \frac{9}{2}(R_{\rm c}-1)^2\frac{1}{y^2}\left(1+\frac{1}{y^2}\right)\left[1+\frac{1}{y^2}+\left(1-\frac{1}{y^2}\right)\cos 2y - \frac{2}{y}\sin 2y\right],\end{aligned} \tag{1299}$$

with the parameter $R_{\rm c}$ expressing the ratio of the slopes of the inflaton potential before and after the kink [252].

28.7.2 Data analysis

We employ a modified version of `CAMB` with suitably increased numerical precision settings to calculate the CMB angular power spectra for the feature models. Since variations of the primordial spectrum may be degenerate with late-time cosmology parameters [209], we explore a parameter space consisting of the base-ΛCDM parameters and the respective additional free parameters of the feature models (see Table 37 for the prior ranges). Note that we take primordial tensor perturbations to be absent in our analysis. In the results presented in Sect. 28.7.3, nuisance parameters are assumed to be uncorrelated with the feature parameters and kept fixed to their base-ΛCDM best-fit values.

In order to maximize sensitivity to narrow features, we use only the *unbinned* versions of the *Planck* high-ℓ likelihoods in the following combinations: (i) temperature data, *Planck* TT(unbinned)+lowE; (ii) E-polarization data only, *Planck* EE(unbinned)+lowE; and (iii) temperature plus polarization data, *Planck* TT,TE,EE(unbinned)+lowE.

For all combinations of feature models and data, the parameter space is sampled with the nested sampling algorithm as implemented in `MultiNest`. The improvement in the fit due to the introduction of a feature is quantified by the effective $\Delta\chi^2 \equiv -2(\ln\mathcal{L}^{\rm best\,fit}_{\Lambda\rm CDM} - \ln\mathcal{L}^{\rm best\,fit}_{\rm feature})$. Being more complex than a power-law spectrum, feature models will in general have a negative $\Delta\chi^2$. However, determining whether the improvement in fit is due to overfitting scatter in the data or due to an actual feature is not straightforward and requires model-dependent simulations [**(alias?)**] or analytic estimates [83] to determine the expected $\Delta\chi^2$ under the null-hypothesis of an underlying power-law spectrum. In the Bayesian approach, a feature model's general performance relative to base ΛCDM can be expressed in terms of the Bayesian evidence $\mathcal{E}$ [268], which is also evaluated by `MultiNest`.

28.7.3 Feature candidates and their evidence

We list the best-fit effective $\Delta\chi^2$ and Bayes factors with respect to a power-law spectrum in Tables 38 and 39. Examining the effective $\Delta\chi^2$ for the feature models previously considered in **(alias?)** reveals only minor differences, with a general trend towards smaller improvements due to features. The $\Delta\chi^2$ of the oscillation and step models fall well within the expected range of $\Delta\chi^2 \sim 10$ found in **(alias?)**. Of note are the relatively high values of the radiation and kink cutoff models for polarization-only data, partially driven by the high quadrupole of the EE data. However, the best-fit parameters and spectra (see Fig. 221) do not match their counterparts in the temperature data at all, which strongly suggests that this is not a physical effect. The same observation can also be made for the step model: the best fit to the EE data is clearly out of phase with the temperature best fit.

	Step			Kin cutoff			Rad cutoff			
	TT	EE	TT,TE,EE	TT	EE	TT,TE,EE	TT	EE	TT,TE,EE	T
$\Delta\chi^2_{\rm eff}$	−7.0	−5.2	−5.4	−1.2	0.0	−0.9	−0.2	−4.7	−0.0	−2
$\ln B$	0.0	−0.2	0.1	0.0	−0.2	0.0	−0.8	0.0	−0.7	−0
$\mathcal{A}_{\rm s}$	0.29	0.19	0.38	...	...	...	...	...	...	..
$\log_{10}(k_{\rm s})$	−3.11	−3.47	−3.09	...	...	...	...	...	...	..
$\ln x_{\rm s}$	0.57	2.17	0.15	...	...	...	...	...	...	..
$\log_{10}(k_Y^{\rm c})$	...	...	...	−3.70	−4.98	−3.72	−4.87	−3.48	−4.86	−3
$R_{\rm c}$	...	...	...	...	...	...	...	...	...	−0

Table 38: Best-fit effective $\Delta\chi^2$ and logarithm of the Bayes factors with respect to a featureless power spectrum, as well as best-fit feature parameters, for the step and cutoff models. Negative values of $\ln B$ indicate a preference for a power-law spectrum, while positive ones prefer the feature model. Wavenumbers are in units of Mpc^{-1}.

	Log osc			Running log osc			Lin osc		
	TT	EE	TT,TE,EE	TT	EE	TT,TE,EE	TT	EE	TT,TE,EE
$\Delta\chi^2_{\rm eff}$	−8.5	−13.5	−11.0	−9.3	−16.5	−11.4	−4.2	−9.0	−10.8
$\ln B$	−1.5	−0.2	−0.9	−1.3	0.2	−0.5	−1.8	−1.3	−0.8
$\mathcal{A}_X$	0.024	0.073	0.014	0.028	0.082	0.016	0.024	0.046	0.015
$\log_{10}\omega_X$	1.51	1.72	1.26	1.50	1.71	1.26	1.74	1.84	1.05
$\varphi_X/(2\pi)$	0.60	0.07	0.07	0.68	0.62	0.11	0.34	0.81	0.56
$\alpha_{\rm rf}$	...	...	...	−0.028	0.022	−0.021	...	...	...

Table 39: Same as Table 38, but for the oscillatory feature models.

We find a similar conclusion for the three oscillation models. As can be seen from the profile likelihood of the frequency parameters in Fig. 222, the likelihood peaks in the modulation frequencies do not match up between the TT and EE data sets. Furthermore, the preferred modulation amplitude for the EE data is in all cases much larger than that for the TT or TT,TE,EE data—given that the polarization data are noisier, this behaviour would be expected for a procedure that is overfitting the data.

Consequently, the Bayesian evidence for all combinations of models and data lies between barely worth mentioning and substantial evidence against the feature model on the Jeffreys scale. This implies that, currently, the *Planck* data do not show a preference for the feature models considered here.

Conversely, within the frequency ranges given by our priors, the relative modulation of the power spectrum is constrained to not exceed roughly 3 %, as shown in Fig. 219 for the logarithmic and linear oscillation models.

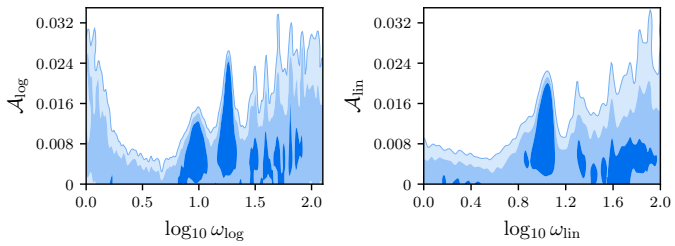


Figure 219: Marginalized joint 68 %, 95 %, and 99 % CL regions of the modulation amplitude versus frequency parameter using the TT,TE,EE data set for the logarithmic (*left*) and linear (*right*) oscillation models.

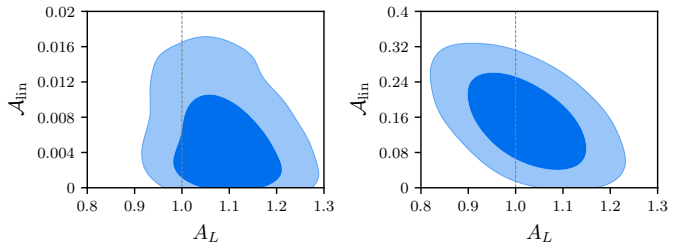


Figure 220: Marginalized joint 68 % and 95 % CL regions for the lensing parameter $A_{\rm L}$ and the modulation amplitude parameter $\mathcal{A}_{\rm lin}$ using the TT data set. *Left:* Linear oscillation model with $\log_{10}\omega_{\rm lin} = 1.158$ and $\varphi_{\rm lin} = \pi$. *Right:* Modified linear oscillation model with a Gaussian envelope function (see text) and $\log_{10}\omega_{\rm lin} = 1.158$, $\varphi_{\rm lin} = \pi$, $\mu_{\rm env} = 0.2\,{\rm Mpc}^{-1}$, and $\sigma_{\rm env} = 0.057\,{\rm Mpc}^{-1}$.

It may also be worth pointing out that in models with oscillations linear in k, the wavelength of the corresponding modulation of the angular power spectra matches that of the CMB's acoustic oscillations, $\Delta\ell \simeq 300$, if $\log_{10}\omega_{\rm lin} \simeq 1.158$. One might therefore suspect that features with frequencies around this value and carefully tuned amplitudes and phases could in principle mimic the (unphysical) effect of a lensing parameter, $A_{\rm L} \neq 1$. However, for a model with a modulation at the BAO frequency and a k-independent modulation amplitude $\mathcal{A}_{\rm lin}$, it can be seen in the left panel of Fig. 220 that we find no correlation between $A_{\rm L}$ and $\mathcal{A}_{\rm lin}$. This is due to a different ℓ-dependence of the respective $\Delta\mathcal{D}_\ell$'s. Explaining the lensing discrepancy would thus require a model with a carefully arranged scale-dependent linear modulation of the primordial spectrum. We demonstrate this possibility for a shaped modulation with a Gaussian envelope of the form $\mathcal{P}_{\mathcal{R}}(k) = \mathcal{P}^0_{\mathcal{R}}(k)\,[1 + \mathcal{A}_{\rm lin}\,\exp(-(k-\mu_{\rm env})^2/2\sigma^2_{\rm env})\,\cos\,(\omega_{\rm lin}k/k_* + \varphi_{\rm lin})]$ in the right panel of Fig. 220, but it should be noted that this particular example is of course highly tuned to produce the desired effect.

Additionally, while the phenomenology of $A_{\rm L}$ and linear modulation models is similar for temperature and polarization spectra individually, the two scenarios are in principle distinguishable by a combination of temperature and polarization data. This is due to the phase difference of the acoustic peaks in TT, TE, and EE, which leads to similar phase differences for the residuals when varying $A_{\rm L}$—unlike modifications of the primordial spectrum which do not shift phase in the same way. However, for features with an amplitude chosen to resemble the apparent lensing excess in the *Planck* TT data, the *Planck* TE and EE data are not sensitive enough to make this distinction.

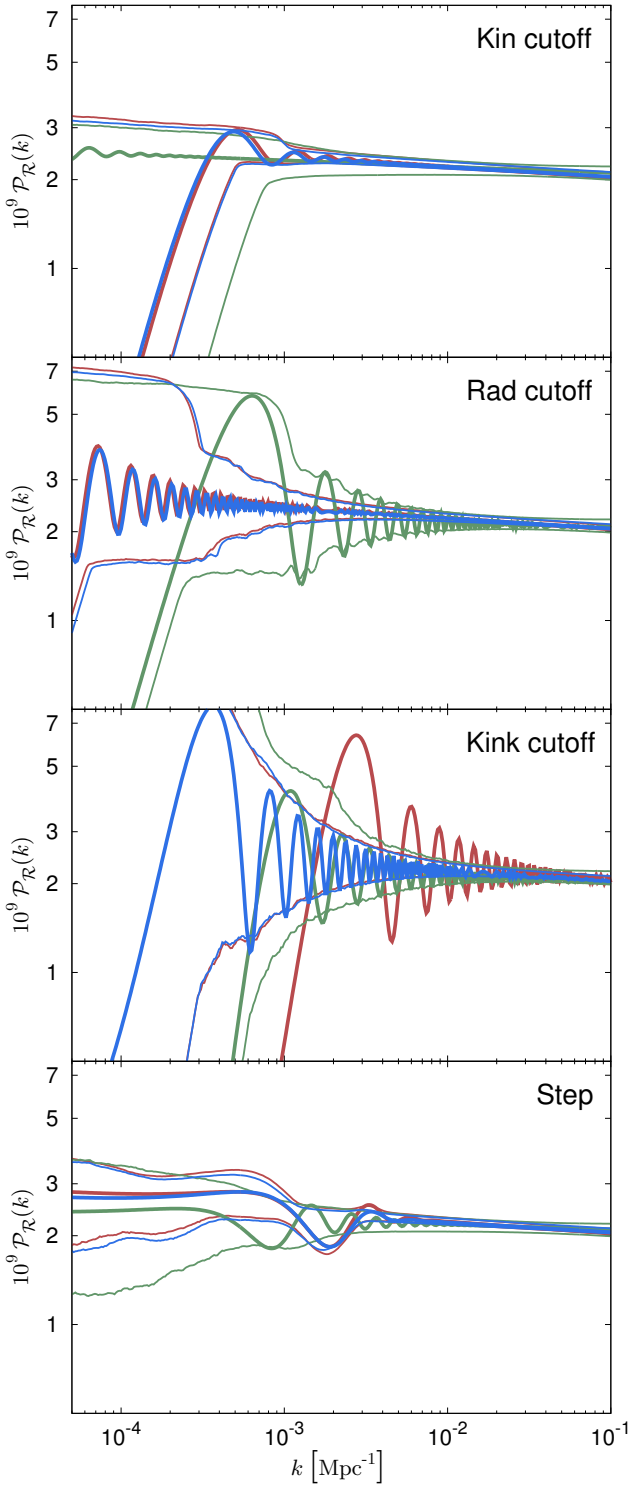


Figure 221: Best-fit and central 95 % CL regions for the primordial power spectrum in the three cutoff and the step models for TT data (red curves), EE data (green), and TT,TE,EE data (blue). Note that for the combination of kink cutoff model and TT data, the best-fit value for the cutoff scale $k^c_{\rm kink}$ lies close to the prior boundary, and therefore the best-fit spectrum does not fall within the *central* 95 %-credible band.

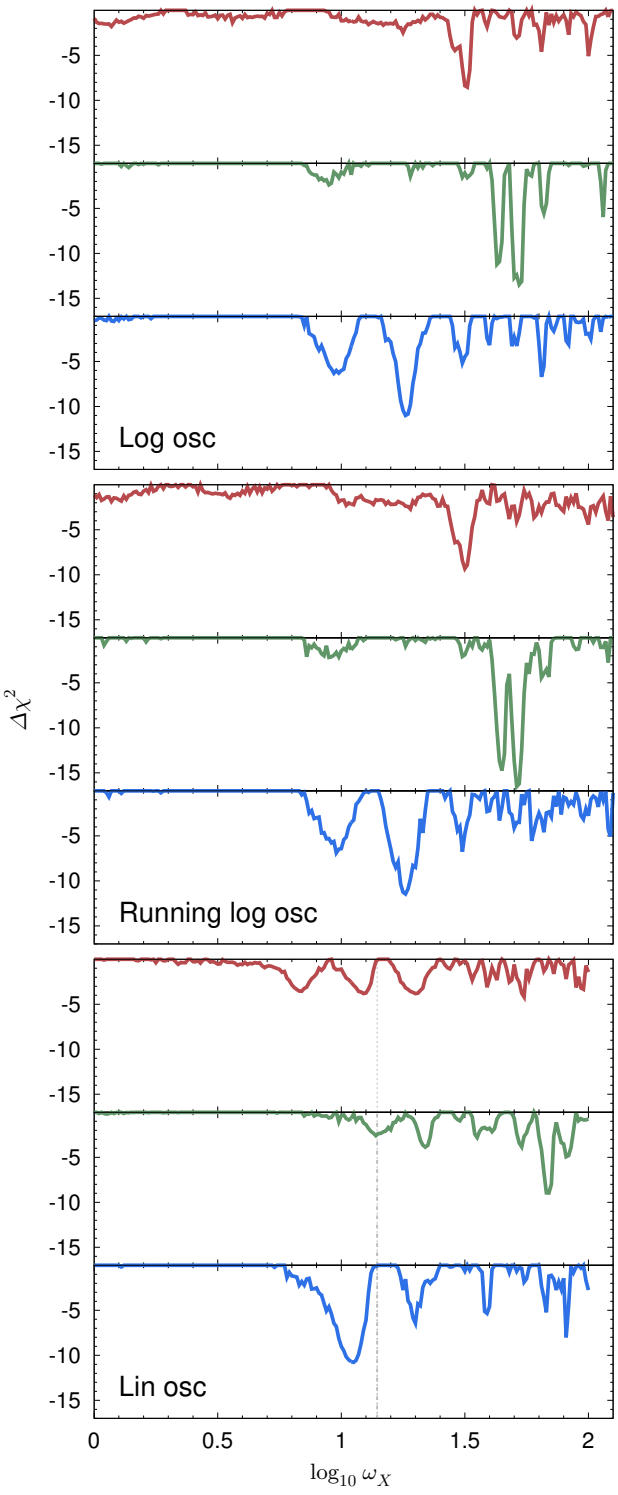


Figure 222: Profile likelihood of the frequency parameter in the three oscillatory feature models for TT (red curves), EE (green), and TT,TE,EE data (blue). The dotted grey line in the bottom panels marks the frequency for which the linear oscillation model leads to a modulation of the angular power spectra whose wavelength roughly matches that of the CMB's acoustic oscillations. Note the lack of alignment between the temperature and polarization likelihood peaks in the vicinity of this frequency.

28.7.4 Axion monodromy

As in section 10.3 of **(alias?)**, we next derive constraints on the underlying parameters in axion monodromy inflation [251, 184, 151, 96], which within string theory motivates a broad class of inflationary potentials of the form

$$V(\phi) = \mu^{4-p}\phi^p + \Lambda_0^4 e^{-C_0(\phi/\phi_0)^{p_\Lambda}} \cos\left(\gamma_0 + \frac{\phi_0}{f}\left(\frac{\phi}{\phi_0}\right)^{p_f+1}\right), \tag{1300}$$

where μ, Λ_0, f, and ϕ_0 are constants which have dimensions of mass, while C_0, p, p_Λ, p_f, and γ_0 are dimensionless. In the literature, one can find theoretically motivated models with $p = 3$, 2, 4/3, 1, and 2/3 [251, 184, 185]. In the following, we neglect a possible amplitude drift in the modulation amplitude by fixing $C_0 = p_\Lambda = 0$, focussing instead on a possible frequency drift p_f, as was done in previous analyses [215, 72, 146, 190, 193, 191, 192].

Due to its oscillating nature, a numerical study of this model is restrictive [215]. As such, we employ the semi-analytic template [96] used in previous analyses, namely

$$\mathcal{P}_\mathcal{R}(k) = \mathcal{P}_\mathcal{R}(k_*)\left(\frac{k}{k_*}\right)^{n_\mathrm{s}-1}\left\{1 + \delta n_\mathrm{s}\cos\left[\frac{\phi_0}{f}\left(\frac{\phi_k}{\phi_0}\right)^{p_f+1} + \Delta\phi\right]\right\}. \tag{1301}$$

We neglect the effect of small oscillations in the tensor primordial spectrum, and approximate it as a power law with a very small spectral index n_t (fixed by the single-field slow-roll self-consistency condition). The most well studied case to date is for $p = 4/3$, but given the high tensor-to-scalar ratio predicted by this model and the current upper bounds on r given in Sect. 28.3.5, we extend our study to the cases of $p = 1$ and $p = 2/3$. Furthermore, to completely specify this template, we assume instantaneous reheating, which, for a pivot scale of $k_* = 0.05\,\mathrm{Mpc}^{-1}$, corresponds to $N_* \approx 57.5$, and $\phi_0 = 12.38 M_\mathrm{Pl}$ with $\phi_\mathrm{end} = 0.59 M_\mathrm{Pl}$. This leads to definite predictions for (r, n_s); namely, $(0.0922, 0.971)$ for $p = 4/3$, $(0.0692, 0.974)$ for $p = 1$, and $(0.0462, 0.977)$ for $p = 2/3$.

To constrain this model, we carry out a Bayesian analysis using a modified version of `CLASS` [163, 56], which has been adapted to allow for a full parameter exploration, using the aforementioned template. As part of these modifications, special care needs to be taken to ensure that a correct sampling Δk in wavenumber space is chosen, at two different levels in the Boltzmann code: when computing an interpolation table for the primordial spectrum of scalars and tensors; and when performing the integral over the squared photon transfer functions multiplied by the primordial spectra to get the multipoles C_ℓ. This sampling needs to be fine enough to guarantee that no features are smoothed out or lost in this convolution, and we checked carefully that this is the case in our runs. The grid of ℓ values at which the C_ℓ's are actually computed and not just interpolated also needs to be refined.

We fit to the data the five cosmological parameters $\{\omega_\mathrm{b}, \omega_\mathrm{c}, \theta, A_\mathrm{s}, \tau\}$ plus the frequency f of the underlying axion decay constant, the frequency drift p_f, and the oscillation amplitude δn_s. We adopt the same priors used in previous analyses: $-4 \leq \log_{10}(f/M_\mathrm{Pl}) \leq -1$ for the frequency; $-0.75 < p_f < 1$ for the frequency drift; and an upper bound on the amplitude of $\delta n_\mathrm{s} < 0.5$. Furthermore, for the phase parameter $\Delta\phi$ we take a uniform prior of $-\pi < \Delta\phi < \pi$.

In Figs. 223, 224, and 225 we show the joint posterior constraints on pairs of primordial parameters for the semi-analytic template, for $p = 4/3$, $p = 1$, and $p = 2/3$, respectively.

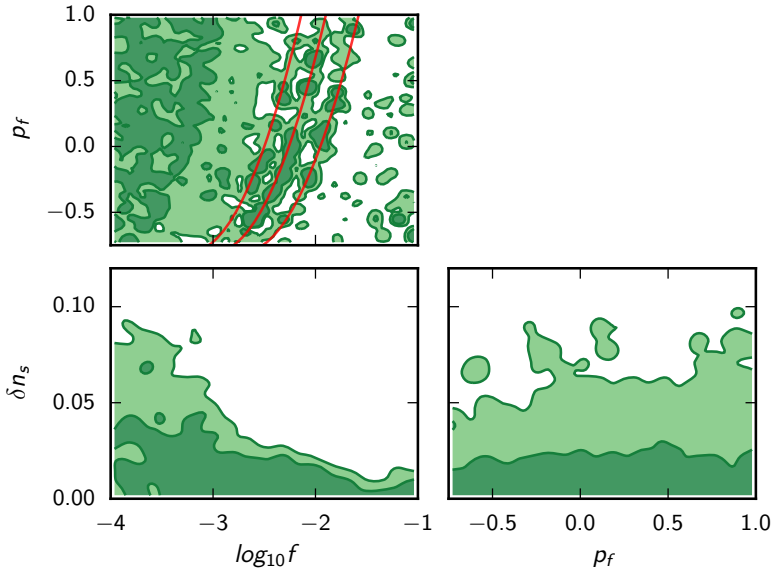


Figure 223: Joint 68 % and 95 % CL constraints on the axion monodromy parameters using *Planck* (unbinned) TT,TE,EE+lowE+BK14, for the case of $p = 4/3$. All smoothing has been turned down in the $p_f - \log_{10}(f/M_{\rm Pl})$ posterior to avoid smoothing the features highlighted in red.

In all three cases, we find two expected asymptotic behaviours. First, when the frequency is very high (which means that f is small in our parameterization), the oscillations in the primordial spectrum are smoothed out in the angular power spectrum, and the oscillation amplitude parameter $\delta n_{\rm s}$ becomes irrelevant and unconstrained. Second, in the limit of a very small amplitude parameter $\delta n_{\rm s}$, the oscillations become undetectable and the parameter f is also unconstrained. In all cases, no preferred frequency drift is found, which is compatible with previous analyses.

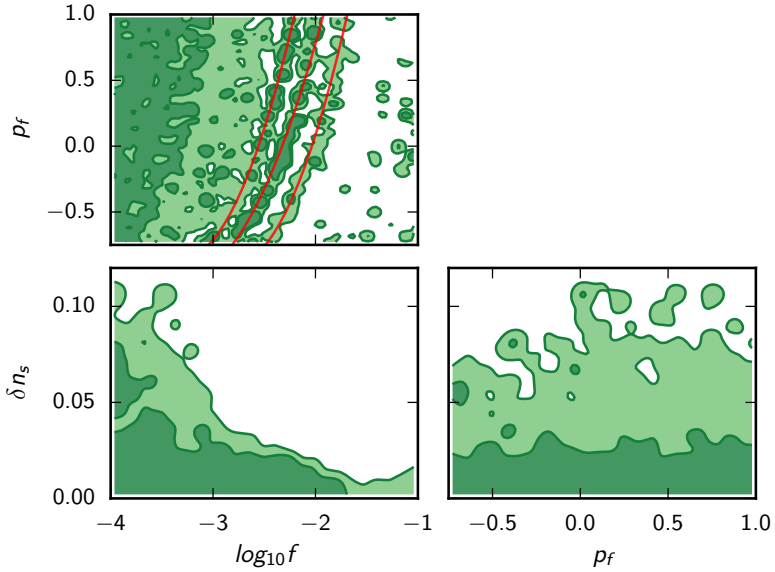


Figure 224: Same as Fig. 223, but for the case of $p = 1$.

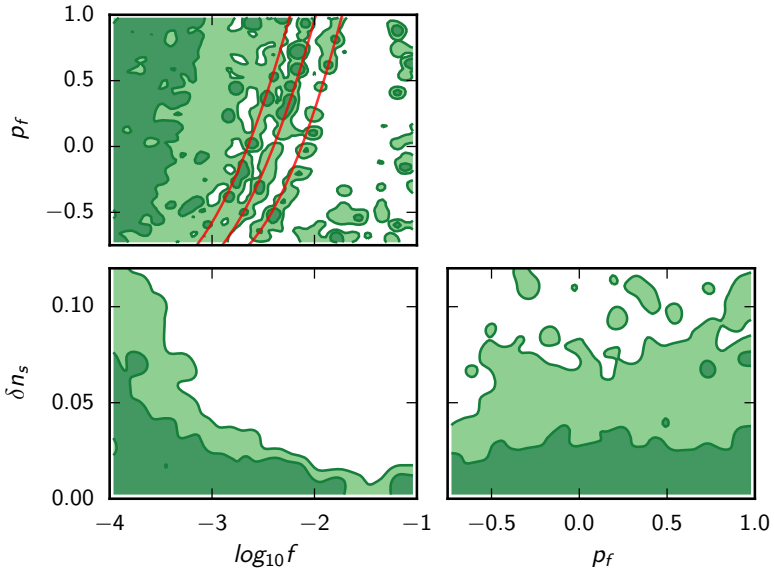


Figure 225: Same as Fig. 223, but for the case of $p = 2/3$.

We recover the complex structures (highlighted in red in Figs. 223, 224, and 225) found in previous analyses in the frequency-frequency drift parameter space, which, as was discussed in **(alias?)**, arise due to underlying modulations in the data and the model [73]. These structures become more apparent as we reduce the index p.

We perform a χ^2 comparison with the minimal 6-parameter ΛCDM model, and find $\Delta\chi^2_{(p=4/3)/\Lambda\mathrm{CDM}} = 0.4$, $\Delta\chi^2_{(p=1)/\Lambda\mathrm{CDM}} = 0.6$ and $\Delta\chi^2_{(p=2/3)/\Lambda\mathrm{CDM}} = 1.2$. The reason for higher χ^2 in the axion monodromy models, despite the addition of extra parameters, is that the predicted r values are in tension with the CMB data. This shows that, overall, axion monodromy models are disfavoured due to their high tensor-mode amplitude.

In order to check specifically whether the data give any hint of oscillatory patterns in the primordial spectrum matching the axion monodromy template, as well as to compare with the results discussed in the previous subsection, we fitted the data with ΛCDM+r models in which r and n_s were fixed to the same values as in the axion monodromy model with $p = 2/3$, 1, and 4/3. In each case, the comparison between axion monodromy and ΛCDM+r with the same (r, n_s) gives $\Delta\chi^2_{(p=4/3)/\Lambda\mathrm{CDM}+r} = -7.8$, $\Delta\chi^2_{(p=1)/\Lambda\mathrm{CDM}+r} = -7.6$ and $\Delta\chi^2_{(p=2/3)/\Lambda\mathrm{CDM}+r} = -8$. That is, in all cases we find $\Delta\chi^2 \sim 10$, which is compatible with the general results shown in Table 39. With three more free parameters, these improvements are statistically insignificant, and we conclude that the data show no preference for axion monodromy models.

28.8 Combined power spectrum and bispectrum analysis for oscillatory features

28.8.1 Approach

This section establishes constraints on oscillatory models using the power spectrum and the bispectrum simultaneously. Oscillatory features can appear in multiple correlation functions [58, 194, 94, 98, 5, 12, 4, 96, 97] (see, e.g., [59] for a recent review). More powerful constraints result when spectra of various orders are combined [212, 200, 108]. Past work has suggested that the statistical weight of the oscillations in the bispectrum (or higher-order correlation functions) is less than that in the power spectrum [26]; however, counterexamples exist as well (see, e.g., Behbahani & Green 27). The analysis in Sect. 28.7 used the *Planck* data to establish stringent constraints on the presence of features in the power spectrum. The 2015 *Planck* data were analyzed to constrain non-Gaussianties containing features [287], where, as in the power spectrum analysis, several candidate features were identified at low statistical significance. The analysis here focuses on the *location,* or frequency, of the feature. Joint analyses of the power spectrum and bispectrum were discussed in several studies [83, 84, 189]. We apply some of the tools developed there to the *Planck* temperature and polarization data.

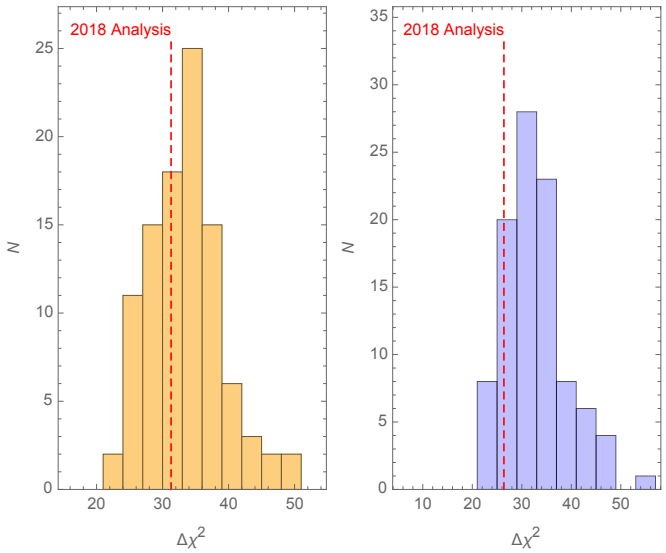


Figure 226: Typical best-fit improvement in units of $\Delta\chi^2$ in 100 simulations compared to the real data (red dashed lines) for the log feature (left) and the linear feature (right) models.

The analysis here is incomplete and limited in several respects. First, we analyse the bispectrum keeping all cosmological parameters fixed. Second, the parameters varied in the bispectrum are not varied in the Bayesian sense. The bispectrum is analyzed using a best-fit analysis based on how well a template shape fits the data. Third, the data suggest the primary bispectrum is close to zero and its covariance dominated by the scalar contributions in the power spectrum. We find that an ideal Bayesian analysis is not computationally feasible (see, e.g., Verde et al. 277).

The output from the bispectrum analysis for features provides us with a map that specifies the significance of a feature in units of σ, given the location (frequency) and the phase of the feature. We can turn this map into a likelihood, which we can simply add to that of the power spectrum; in other words, we take

$$\begin{aligned}\ln \mathcal{L}_{\rm tot} &= \ln \mathcal{L}_{\rm PS}(c, f, \omega_{\rm P}, A_{\rm P}, \phi_{\rm P}|{\rm dat}) \\ &+ \ln \mathcal{L}_{\rm BS}(A_{\rm B}, \omega_{\rm B}, \phi_{\rm B}|{\rm dat}),\end{aligned} \tag{1302}$$

where c represents the standard cosmological parameters, f the foregrounds, $\omega_{\rm P,B}$ the frequency, $A_{\rm P,B}$ the amplitude, and $\phi_{\rm P,B}$ the phase of the modulation in, respectively, the power spectrum and the bispectrum. We assume vanishing covariance between the power spectrum and the bispectrum, which has been shown to be a good approximation [83, 84, 189]. Furthermore, the likelihood $\ln \mathcal{L}_{\rm BS}$ is not normalized (more precisely $\ln \mathcal{L}_{\rm tot}$ is not normalized in a universe with a non-zero bispectrum).

Strictly speaking, we do not have a likelihood that measures $A_{\rm B}$ with a certain probability. Furthermore, several studies have shown that the frequency parameter, in combination with the amplitude and the phase, does not obey a χ^2 fitting to the data [123, 192, 193, 191, 186, 72]. Removing the frequency from the search results in a χ^2 distribution with two degrees of freedom. Since $\ln \mathcal{L}_{\rm tot}$ is rather large (of order 10^4 when combining all data), we can change the equation above by limiting ourselves only to improvements that are driven by $\omega_{\rm B} \sim \omega_{\rm P} \equiv \omega$, that is,

$$\begin{aligned}\ln \mathcal{L}_{\rm tot} &= \ln \mathcal{L}_{\rm PS}(c, f, \omega, A_{\rm P}, \phi_{\rm P}|{\rm dat}) \\ &+ \Delta \ln \mathcal{L}_{\rm BS}(A_{\rm B}, \omega, \phi_{\rm B}|{\rm dat}).\end{aligned} \tag{1303}$$

Assuming that $\phi_{\rm B}$ and $A_{\rm B}$ are well described by a two-parameter χ^2 distribution, we can now convert our σ map into a χ^2 improvement via

$$\chi^2 = -2 \log \left[{\rm Erf}\left(\sigma/\sqrt{2}\right) + 1 \right], \tag{1304}$$

or, in terms of the likelihoods,

$$-2\ln\mathcal{L}_{\rm tot} = -2\ln\mathcal{L}_{\rm PS}\left(c, f, \omega, A_{\rm P}, \phi_{\rm P}|{\rm dat}\right) + 2\log\left\{{\rm Erf}\left[\sigma(\omega, A_{\rm B}, \phi_{\rm B})/\sqrt{2}\right] + 1\right\}. \tag{1305}$$

We will use the above expression to derive the posterior of the joint fit.

28.8.2 Models

We will focus on two models: the local or linear feature model and the log feature model. For the log model we set $\{A_{\rm P}, \omega, \phi_{\rm P}, A_{\rm B}, \phi_{\rm B}\} = \{A_{\log}, \omega_{\log}, \phi_{\log}, B_{\log}, \tilde{\phi}_{\log}\}$ and use for the power spectrum

$$\mathcal{P}_{\log}(k) = \mathcal{P}_0(k)\left[1 + A_{\log}\cos\left(\omega_{\log}\log\frac{k}{\tilde{k}_0} + \phi_{\log}\right)\right], \tag{1306}$$

with $\mathcal{P}_0(k) = A_{\rm s}(k/k_*)^{n_{\rm s}-1}$. For the bispectrum we use [55]

$$B_{\log}(k_1, k_2, k_3) = \frac{B_{\log}A_{\rm s}}{k_1^2k_2^2k_3^2}\cos\left(\omega_{\log}\log\sum\frac{k_i}{\tilde{k}_0} + \tilde{\phi}_{\log}\right). \tag{1307}$$

The above parameterized spectra are examples that could be generated in axion monodromy inflation [95, 98], but generally are expected to appear in models where there exists an oscillatory potential.

For the linear model we follow [84] with $\{A_{\rm P}, \omega_{\rm P}, \phi_{\rm P}, A_{\rm B}, \phi_{\rm B}\} = \{A_{\rm lin}, \omega_{\rm lin}, \phi_{\rm lin}, B_{\rm lin}, \tilde{\phi}_{\rm lin}\}$ and write

$$\mathcal{P}_{\rm lin}(k) = \mathcal{P}_0(k)\left[1 + A_{\rm lin}\sin\left(2\omega_{\rm lin}\frac{k}{\tilde{k}_0} + \phi_{\rm lin}\right)\right] \tag{1308}$$

and [57]

$$B_{\rm lin}(k_1, k_2, k_3) = \frac{B_{\rm lin}A_{\rm s}}{k_1^2k_2^2k_3^2}\cos\left[\omega_{\rm lin}\left(\sum\frac{k_i}{\tilde{k}_0}\right) + \tilde{\phi}_{\rm lin}\right]. \tag{1309}$$

In both models we choose $\tilde{k}_0 = 1\,{\rm Mpc}^{-1}$, *which is different from the choice in Sect. 28.7 for the linear model.* As a result the linear frequencies can be related using $\omega_{\rm lin,Sect8} = 10\omega_{\rm lin,Sect7}$. The pivot scale is set to the usual value $k_* = 0.05\,{\rm Mpc}^{-1}$. The above parameterization is a proxy for models that contain sharp oscillatory features [55, 140, 13], although typically such effects would generate decaying features, which will not be considered here.

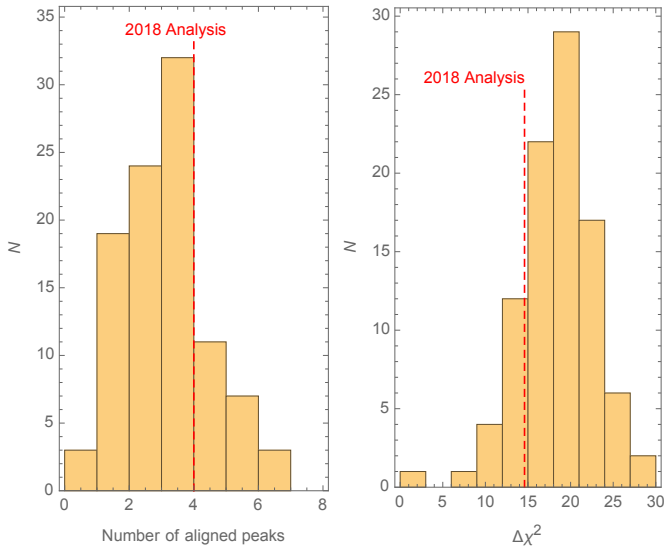


Figure 227: *Left:* Number of aligned peaks in the power spectrum and the bispectrum for the log feature model. *Right:* Mean improvement of those same peaks in 100 simulated bispectra combined with the unbinned high-ℓ likelihood.

28.8.3 Data analysis

Power spectrum

Our analysis uses a modified version of `CAMB` [166], which is capable of adaptively changing the sampling in both k and ℓ depending on the frequency of the feature, allowing us to scan a wide range of frequencies. As in the previous section, we use the unbinned versions of the *Planck* high-ℓ likelihoods for temperature plus polarization, in combination with lensing and large-scale temperature and polarization (i.e., lowE). We compared the power spectrum results for the limited frequency range considered in Sect. 28.7 for the log and linear model, and find excellent agreement (sampled with `Multinest`). We developed a bispectrum likelihood module based on Eq. (1305) using the 2015 data analysis [287] for both the log and linear feature models, obtained using optimal estimators following Münchmeyer et al. [207, 208] and Meerburg & Münchmeyer [188]. For the log model, the frequency range is set to $10 \leq \omega_{\rm log} \leq 1000$, while for the linear model we consider the frequency range $10 \leq \omega_{\rm lin} \leq 3000$. This joint analysis excludes the very low frequencies known to (weakly) correlate with cosmological parameters. The cosmological parameters are held fixed in the bispectrum analysis. We consider amplitudes $0 \leq A_{\rm log,lin} \leq 0.9$; the highest amplitudes will only be allowed for high frequencies where projection suppresses the power of the oscillating part in the power spectrum significantly. The phase is varied and marginalized over in the joint analysis. We use the `PolyChord` sampler [126, 127], which is powerful enough to include foregrounds (with $n_{\rm live} = 512$).

Bispectrum

The bispectrum likelihood is derived from the posterior distributions generated in [295]. Although the linear bispectrum of Eq. (1309) can easily be factorized, the log bispectrum of Eq. (1307) is not of the factorized form. Using modal techniques developed by Fergusson & Shellard [87] and Fergusson et al. [85, 86], any shape can be factorized, with a close-to-optimal estimator. The modal method converts the angular-average bispectrum into a set of factorizable orthogonal mode functions. These functions can be directly constrained using foreground-cleaned CMB maps. From these measurements, a large number of bispectra can be reconstructed and constrained by appropriately weighting the mode functions. The convergence of this method, in terms of how many mode functions are required to accurately reconstruct the shape of interest, depends on the choice of the mode functions. In the 2015 analysis, two different mode functions were used: a polynomial-based reconstruction; and a trigonometric-based reconstruction. The latter was developed by [207, 208] and relies on expanding around linear oscillations. The polynomial-based reconstruction is extremely powerful for most bispectra, but is non-optimal for oscillatory bispectra, which require a large number of modes (e.g., more than 2000 for $\omega_{\rm log} = 50$). Trigonometric modes allow for faster convergence and provide good reconstruction for much higher frequencies, both for linear- and log-type modulations. For low frequencies, both methods can be compared and results show excellent agreement [287]. In addition, both methods were developed independently, which provides further confidence in the results. In the analysis presented here, we use the results obtained using the trigonometric mode functions. Further details can be found in Planck Collaboration IX [295].

28.8.4 Estimating significance

Next we will estimate the significance of the improvements driven by the joint analysis. For this purpose we generate 100 mock spectra as in [189] *without features* and perform an analysis jointly with true CMB power spectrum data, i.e., we use the same power spectrum likelihood (real data) in combination with the simulated bispectrum likelihoods (mock data). We will do this for both the linear and the log models, with 100 simulations in total. Each analysis requires a similar amount of time as does the real data analysis, using about 12 000 CPU hours for the linear feature and about 40 000 CPU hours for the log feature *per simulation*. More details on the simulated spectra can be found in [295].

These simulations help us assess the statistical significance of our results. Improvement in fit is given in units of χ^2 compared to a no-feature model as defined the previous section [i.e., $\Delta\chi^2 \simeq -2(\ln \mathcal{L}^{\rm best\,fit}_{\Lambda{\rm CDM}} - \ln \mathcal{L}^{\rm best\,fit}_{\rm feature})$]. The left panel of Fig. 226 shows the typical best-fit improvement from a set of simulations for the log feature model. This first analysis shows that the best fit in the data is perfectly consistent with a standard ΛCDM universe, without features,

with $P(\Delta\chi^2 \geq \Delta\chi^2_{\rm data}) = 28\,\%$. This outcome is not unexpected, given earlier analyses for the power spectrum (see, e.g., Meerburg et al. 193, Easther & Flauger 72, Benetti 30, Miranda & Hu 195, Fergusson et al. 83, Planck Collaboration XX 288, Hazra et al. 132) and the significance of features in the bispectrum alone [284]. The look-elsewhere effect lowers the significance of features and by jointly constraining features in the power spectrum and the bispectrum it is possible to alleviate some of this suppression. To quantify this, we consider the following two questions: 1) considering the various frequencies with $\Delta\chi^2$ improvement over no features in the joint analysis, how many of those were present in the power spectrum analysis only; and 2) what is the mean improvement, in units of $\Delta\chi^2$, of these fits? We compare the results of the simulations, which do not contain any real features, to the data.

Before we answer these questions we need a criterion to decide if two frequencies will be considered the same or not, i.e., we need a frequency correlation measure. We will consider a simple ansatz, which will have an analytical solution and will serve to estimate the correlation between frequencies, by defining

$$f_{\log}[\omega_1, \omega_2, \phi] \simeq \int dx \cos[\omega_1 \log x + \phi] \cos[\omega_2 \log x + \phi]. \tag{1310}$$

Next we marginalize over phase, defining

$$\begin{aligned} g_{\log}[\omega_1, \omega_2] &= \int d\phi f_{\log}[\omega_1, \omega_2, \phi] \\ &= \frac{\pi}{1 + \Delta\omega_{12}^2} \{x_{\rm max} \cos\left[\Delta\omega_{12} \log\left(x_{\rm max}\right)\right] \\ &\quad - \cos\left[\Delta\omega_{12} \log\left(x_{\rm min}\right)\right] x_{\rm min} \\ &\quad + \Delta\omega_{12} \left[\sin\left[\Delta\omega_{12} \log\left(x_{\rm max}\right)\right] x_{\rm max}\right. \\ &\quad \left. - \sin\left[\Delta\omega_{12} \log\left(x_{\rm min}\right)\right] x_{\rm min}\right]\}\,, \end{aligned} \tag{1311}$$

where $\Delta\omega_{12} = \omega_1 - \omega_2$. For linear oscillations we can derive a similar measure, with

$$\begin{aligned} g_{\rm lin}[\omega_1, \omega_2] &= \frac{\pi}{2\Delta\omega_{12}} \left[\sin\left(2\Delta\omega_{12} x_{\rm max}\right)\right. \\ &\quad \left. - \sin\left(2\Delta\omega_{12} x_{\rm min}\right)\right]. \end{aligned} \tag{1312}$$

The correlation is given by

$$\mathrm{Cor}(\omega_1, \omega_2) \quad = \frac{g[\omega_1, \omega_2]}{\sqrt{g[\omega_1, \omega_1] g[\omega_2, \omega_2]}}. \tag{1313}$$

The parameters $x_{\rm min}$ and $x_{\rm max}$ play a role in determining the correlation length. Although strictly speaking they correspond to the minimum and maximum scales observable in the CMB, they can be used to model the correlator to allow for shifts in the frequency coming from a non-optimal analysis. We argue that this is reasonable given the low number of peaks in the analysis. We tested the above on various nearby peaks in the data and found that demanding $\mathrm{Cor}(\omega_1, \omega_2) \leq 0.1$ is generally sufficient to effectively identify independent peaks. We explored the sensitivity of the results to the correlation criterion of 0.1. First we increased it to 0.3 and found that in this case many peaks were missed when counting the number of aligned peaks (with little effect on determining the peaks). When we lowered the criterion to 0.01, we obtained many aligned peaks that should not be aligned. Small changes in the correlation criterion have minimal effect on the results presented here. Ideally, more simulations should be generated, which would help to establish the best choice for the correlation criterion. We found that the choice of $x_{\rm min}$ does not affect the correlator as long as $x_{\rm min} \ll 1$. We set $x_{\rm max} = 0.05$ for linear oscillations, which roughly correlates peaks with $\Delta\omega_{\rm lin} \sim 10$, which is within the tails of the observed widths of the peaks in the power and bispectrum analysis. For log oscillations we choose $x_{\rm max} = 1$, which has the advantage that the correlator has no zero-crossings near the peak (but hardly effects the correlation length). We find $\Delta\omega_{\log} \sim 1$ at $\omega_{\log} = 100$, which seems reasonable in light of the power and bispectrum peak widths (see Fig. 222).

In Fig. 227 we show the number of peaks in the joint analysis that have improved (left panel) as well as their mean improvement (right) over a no-feature analysis. We find $P(\#\text{peaks} \geq$

$\#\text{peaks}_{\text{data}}) = 16\,\%$ and those peaks do not lead to significant improvements in the joint χ^2, with $P(\Delta\chi^2_{\text{peaks}} \geq \Delta\chi^2_{\text{peaks,data}}) = 83\,\%$. Assuming that these 100 simulations provide a fair sample of the noisy data, we conclude that there are no significant features present in both the power spectrum and the bispectrum for the models considered within the chosen range of feature parameters.

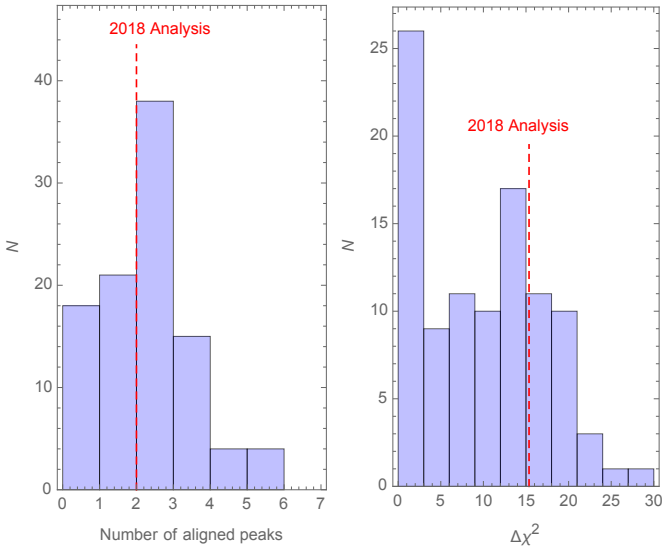


Figure 228: *Left:* Number of aligned peaks in the power spectrum and the bispectrum for the linear feature model. *Right:* Mean improvement of those same peaks in 100 simulated bispectra combined with the unbinned high-ℓ likelihood.

We carry out the same analysis for linear features and show the results in Fig. 226 (right panel), deriving a typical best fit from 100 simulated noisy spectra. The true best fit, as derived from the joint analysis of the 2018 bispectrum and the 2018 power spectrum, shows a relatively small improvement with $P(\Delta\chi^2 \geq \Delta\chi^2_{\text{data}}) = 85\,\%$. Further correlated-features searches find 2 features within the frequency window which may be considered aligned, with a mean $\Delta\chi^2$ of 12.3, as illustrated in Fig. 228. Compared to 100 simulated noisy spectra, we obtain $P(\#\text{peaks} \geq \#\text{peaks}_{\text{data}}) = 42\,\%$ and $P(\Delta\chi^2_{\text{peaks}} \geq \Delta\chi^2_{\text{peaks,data}}) = 26\,\%$. Since the overall improvement from fitting these aligned peaks does not exceed the 3σ threshold, we conclude that there is no statistically significant evidence for any of these features.

We conclude that the simple parameterization considered in this analysis does not provide any evidence for features. The two models analyzed are representive of a broad class and have well-studied phenomenological spectra; however, other classes of models exist. Features, for example, can have scale dependence [i.e., an "envelope" [55, 3, 267]]. Likewise, more realistic modelling of axion models shows that the frequency could depend on scale [i.e., "running" [96, 97]]. Both these possibilities could substantially change the spectra and likely the joint analysis and significance.

Here we have not imposed an explicit relation between the amplitude of the bispectrum and the frequency of the power spectrum. In the simplest form of axion monodromy, one has $f_{\text{NL}} = A_{\log}\omega^2_{\log}/8$. On account of the quadratic scaling with frequency and the fact that the constraint on the amplitude tends to become poorer as the frequency increases due to projection (see, e.g., Fig. 223), it was already pointed out in [288] that there is no evidence for this relation in the data, and with the current data set this situation remains unchanged.

Model and data	Δn	$100\beta_{\rm iso}$ at $k_{\rm low}$	$k_{\rm mid}$	$k_{\rm high}$	$100\cos\Delta$	$100\alpha_{\rm non-adi}$	$\Delta\chi^2$	$\ln B$
General models (*three* isocurvature parameters):								
CDI *Planck* 2015 TT+lowP	3	4.1	37	57	[−30 : 20]	[−1.48 : 1.91]	−2.1	
CDI *Planck* TT+lowE	3	3.6	38	61	[−23 : 27]	[−0.76 : 2.05]	−0.7	−12.6
CDI `CamSpec` TT+lowE	3	3.8	35	56	[−22 : 23]	[−0.62 : 2.12]	−0.7	−13.4
CDI *Planck* TT+lowP	3	4.2	35	56	[−25 : 23]	[−1.03 : 1.98]	−0.5	−12.6
CDI *Planck* TT + τ prior	3	8.4	27	40	[−21 : 29]	[−0.83 : 5.35]		
CDI *Planck* 2015 TT+lowP+lensing	3	4.5	[1 : 40]	[1 : 62]	[−28 : 17]	[−1.05 : 1.86]	−1.2	
CDI *Planck* TT+lowE+lensing	3	4.0	35	57	[−28 : 23]	[−1.20 : 2.04]	−0.6	−12.3
CDI `CamSpec` TT+lowE+lensing	3	3.7	34	55	[−24 : 24]	[−0.96 : 2.10]	−0.5	−12.8
CDI *Planck* 2015 TT,TE,EE+lowP	3	2.0	[3 : 28]	[5 : 52]	[−6 : 20]	[0.09 : 1.51]	−5.3	
CDI *Planck* TT,TE,EE+lowE	3	2.1	[1 : 31]	58	[−11 : 15]	[−0.18 : 1.24]	−3.0	−12.8
CDI `CamSpec` TT,TE,EE+lowE	3	2.8	21	38	[−12 : 20]	[−0.20 : 1.67]	−1.2	−14.0
CDI *Planck* TT,TE,EE+lowP	3	2.4	27	50	[−11 : 17]	[−0.16 : 1.45]	−2.3	−13.4
CDI *Planck* TT,TE,EE + τ prior	3	6.2	17	30	[−13 : 14]	[−0.48 : 3.94]		
CDI *Planck* TT,TE,EE+lowE+lensing	3	2.5	[1 : 26]	47	[−12 : 15]	[−0.25 : 1.31]	−2.8	−12.8
CDI `CamSpec` TT,TE,EE+lowE+lensing	3	3.0	19	33	[−16 : 18]	[−0.38 : 1.54]	−0.9	−14.1
CDI *Planck* TT,TE,EE+lowP+lensing	3	2.2	[1 : 27]	50	[−11 : 16]	[−0.16 : 1.36]		
CDI WMAP-9	3	20.1	[2 : 50]	66	[−38 : 34]	[−1.79 : 6.46]	−0.2	−9.6
NDI *Planck* 2015 TT+lowP+lensing	3	15.8	[2 : 24]	[2 : 29]	[−32 : 0]	[−4.04 : 1.37]	−2.8	
NDI *Planck* TT+lowE+lensing	3	15.3	17	21	[−36 : 4]	[−4.20 : 1.53]	−1.9	−10.8
NDI *Planck* TT,TE,EE+lowE+lensing	3	7.4	[3 : 17]	[2 : 23]	[−13 : 8]	[−0.76 : 1.74]	−5.3	−10.9
NVI *Planck* 2015 TT+lowP+lensing	3	9.8	[1 : 12]	14	[−23 : 7]	[−2.03 : 2.95]	−2.5	
NVI *Planck* TT+lowE+lensing	3	7.1	10	12	[−36 : 3]	[−3.34 : 1.71]	−2.5	−12.6
NVI *Planck* TT,TE,EE+lowE+lensing	3	6.8	[1 : 8]	10	[−20 : 0]	[−1.66 : 1.29]	−5.2	−12.0
CDI+$A_{\rm L}$ *Planck* TT+lowE	4	9.4	28	41	[−41 : 10]	[−2.32 : 2.29]	−9.2	−10.1
CDI+$A_{\rm L}$ *Planck* TT+lowE+lensing	4	6.0	36	57	[−27 : 18]	[−1.16 : 2.19]	−4.1	−13.1
CDI+$A_{\rm L}$ *Planck* TT,TE,EE+lowE	4	3.3	20	36	[−12 : 19]	[−0.24 : 1.89]	−10.6	−11.2
CDI+$A_{\rm L}$ *Planck* TT,TE,EE+lowE+lensing	4	2.7	[1 : 27]	49	[−10 : 16]	[−0.12 : 1.53]	−8.1	−13.5
Special CDI cases (*one* isocurvature parameter):								
Uncorrelated, $n_{\mathcal{II}} = 1$								
"axion I" *Planck* 2015 TT+lowP+lensing	1	3.9	4.3	4.4	0	[0 : 1.70]	0	
"axion I" *Planck* TT+lowE+lensing	1	3.5	3.9	3.9	0	[0 : 1.58]	0	−5.7
"axion I" *Planck* TT,TE,EE+lowE+lensing	1	3.5	3.8	3.9	0	[0 : 1.55]	0	−5.5
Fully correlated, $n_{\mathcal{II}} = n_{\mathcal{RR}}$								
"curvaton I" *Planck* 2015 TT+lowP+lensing	1	0.2	0.2	0.2	+100	[0.30 : 2.70]	0	
"curvaton I" *Planck* TT+lowE+lensing	1	0.2	0.2	0.2	+100	[0.09 : 2.91]	0	−8.9
"curvaton I" *Planck* TT,TE,EE+lowE+lensing	1	0.1	0.1	0.1	+100	[0.07 : 1.81]	0	−9.7
Fully anti-correlated, $n_{\mathcal{II}} = n_{\mathcal{RR}}$								
"curvaton II" *Planck* 2015 TT+lowP+lensing	1	0.5	0.5	0.5	−100	[−4.40 : −0.40]	−0.6	
"curvaton II" *Planck* TT+lowE+lensing	1	0.5	0.5	0.5	−100	[−4.40 : −0.35]	−0.4	−7.2
"curvaton II" *Planck* TT,TE,EE+lowE+lensing	1	0.1	0.1	0.1	−100	[−2.04 : −0.13]	0	−9.2
Special CDI cases (*two* isocurvature parameters):								
Uncorrelated, $n_{\mathcal{II}}$ free								
"axion II" *Planck* TT+lowE+lensing	2	2.3	[3 : 43]	[6 : 75]	0	[0.03 : 1.24]	−0.5	−6.7
"axion II" *Planck* TT,TE,EE+lowE+lensing	2	1.1	[5 : 38]	[10 : 77]	0	[0.07 : 0.66]	−2.8	−6.3
Arbitrarily correlated, $n_{\mathcal{II}} = n_{\mathcal{RR}}$								
"curvaton III" *Planck* TT+lowE+lensing	2	4.7	4.7	4.7	[−75 : 28]	[−3.38 : 1.99]	−0.4	−10.0
"curvaton III" *Planck* TT,TE,EE+lowE+lensing	2	3.9	3.9	3.9	[−41 : 31]	[−1.30 : 2.10]	0	−10.5
Fully correlated, $n_{\mathcal{II}}$ free								
Planck TT+lowE+lensing	2	0.1	4.6	16.0	+100	[0.28 : 2.15]	−0.5	−13.2
***Planck* TT,TE,EE+lowE+lensing**	2	0.02	1.5	6.2	+100	[0.14 : 0.99]	−0.3	−16.0
Fully anti-correlated, $n_{\mathcal{II}}$ free								
Planck TT+lowE+lensing	2	0.6	0.9	1.3	−100	[−5.56 : −0.53]	−1.3	−13.8
***Planck* TT,TE,EE+lowE+lensing**	2	0.3	0.2	0.2	−100	[−4.56 : −0.16]	−0.6	−16.7

Table 40: Constraints on mixed adiabatic and isocurvature models. We report 95 % CL intervals or upper bounds on the isocurvature fraction $\beta_{\rm iso}$ at three scales ($k_{\rm low} = 0.002\,{\rm Mpc}^{-1}$, $k_{\rm mid} = 0.050\,{\rm Mpc}^{-1}$, and $k_{\rm high} = 0.100\,{\rm Mpc}^{-1}$), the scale-independent correlation fraction, $\cos\Delta$, and the non-adiabatic contribution to the CMB temperature variance, $\alpha_{\rm non\text{-}adi}$. Here $\Delta\chi^2$ is the difference between the χ^2 of the best-fit mixed and pure adiabatic models. In the last column we give the difference between the log of Bayesian evidences. (A negative $\ln B$ means that Bayesian model comparison disfavours the mixed model.) The number of extra parameters compared with ΛCDM is denoted by Δn in the first column. Note that the uniform priors on the primordial powers at two scales lead to non-uniform priors on the parameters reported in this table. This is particularly significant for $\beta_{\rm iso}(k_{\rm mid})$, where the prior peaks at a non-zero value. The baseline *Planck* 2018 TT,TE,EE+lowE+lensing results are highlighted in bold.

28.9 Constraints on isocurvature fluctuations

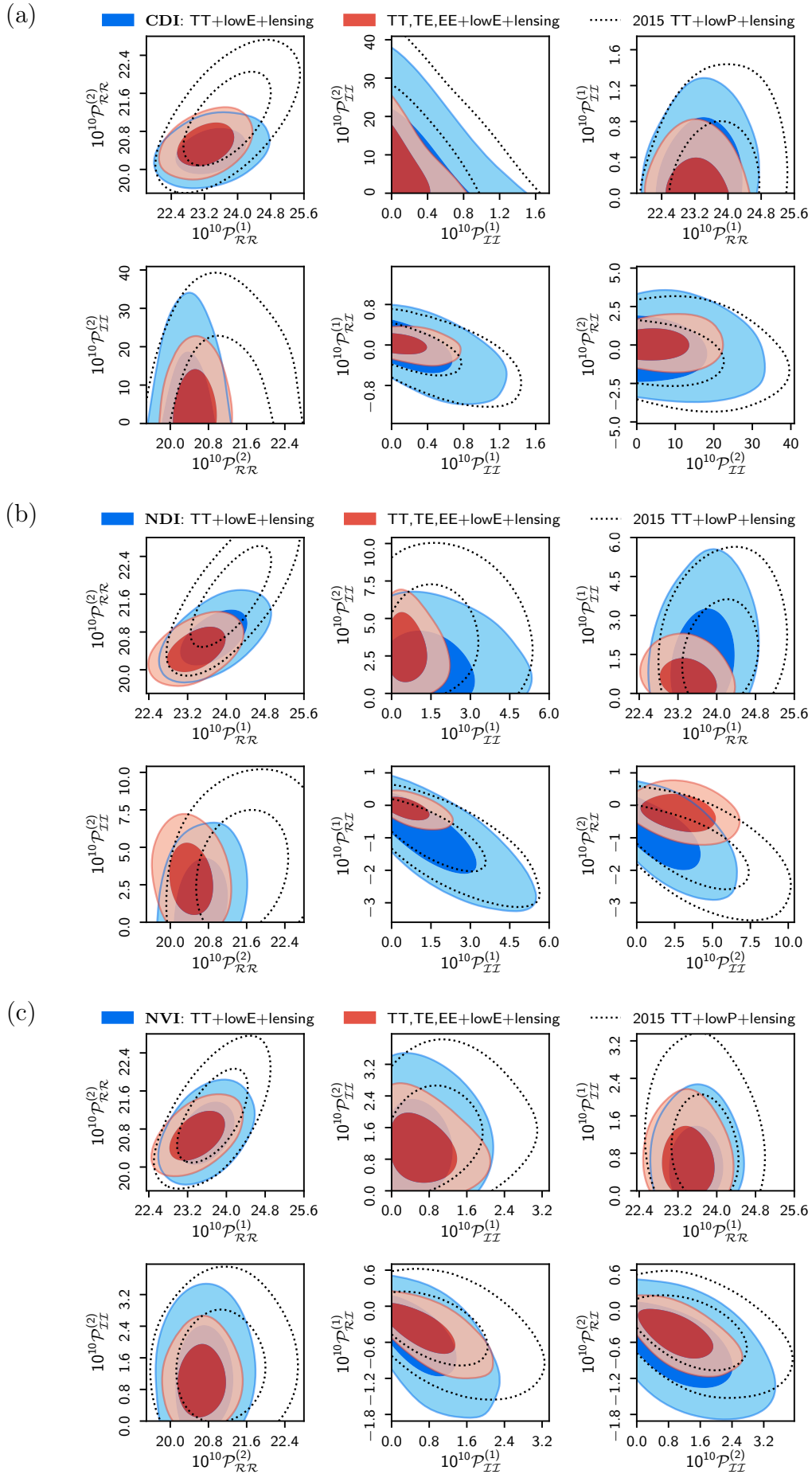


Figure 229: Constraints on the primordial perturbation power in generally correlated ADI+CDI (a), ADI+NDI (b), and ADI+NVI (c) models at two scales, $k_1 = 0.002\,\mathrm{Mpc}^{-1}$ (1) and $k_2 = 0.100\,\mathrm{Mpc}^{-1}$ (2). Note that in our modelling $\mathcal{P}^{(2)}_{\mathcal{RI}}$ is not an independent parameter.

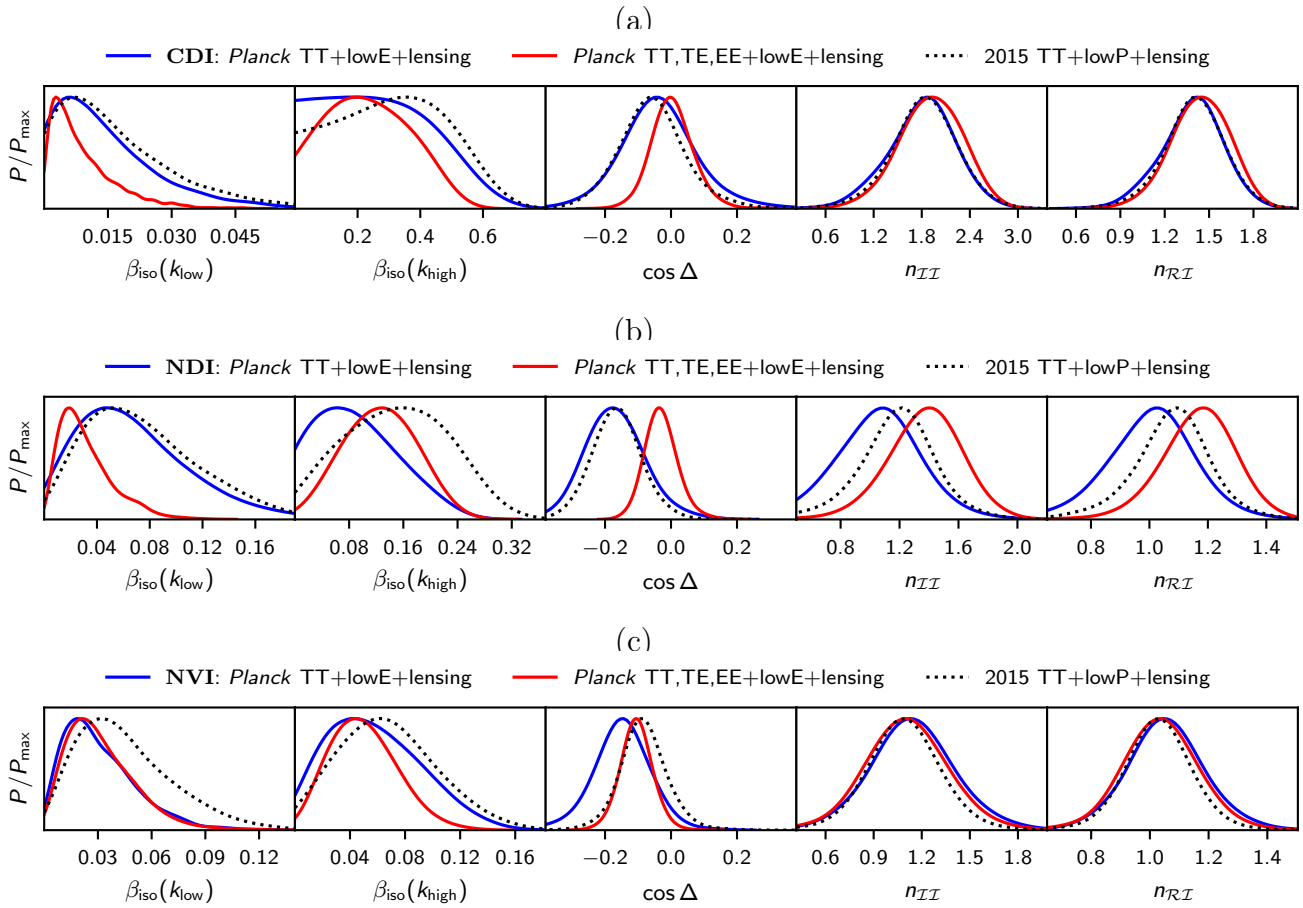


Figure 230: Constraints on the primordial isocurvature fraction, $\beta_{\rm iso}$, at $k_{\rm low} = 0.002\,{\rm Mpc}^{-1}$ and $k_{\rm high} = 0.100\,{\rm Mpc}^{-1}$; the primordial correlation fraction, $\cos\Delta$; the isocurvature spectral index, $n_{\mathcal{II}}$; and the correlation spectral index, $n_{\mathcal{RI}} = (n_{\mathcal{RR}}+n_{\mathcal{II}})/2$, for the generally correlated mixed ADI+CDI model (a), for the ADI+NDI model (b), and for the ADI+NVI model (c). All these parameters are derived, and the distributions shown here result from a uniform prior on the primary parameters shown in Fig. 229. However, the effect of the non-flat derived-parameter priors is negligible for all parameters except for $n_{\mathcal{II}}$ (and $n_{\mathcal{RI}}$) where the prior biases the distribution toward unity. Note that these spectral indices are not well constrained, since we do not have a detection of non-zero isocurvature or correlation amplitude. With a sufficiently small isocurvature or correlation amplitude, an arbitrarily small or large spectral index leads to a very good fit to the data, since the model is then practically adiabatic over the range covered by the *Planck* data.

28.9.1 Background and modeling

Single-field inflation with a canonical kinetic term gives rise to primordial super-Hubble comoving curvature perturbations, $\mathcal{R}$. In this case the relative number densities of the various particle species are spatially constant, i.e., the perturbations are adiabatic. Typically photons are chosen as a reference species. Then adiabaticity implies that for every particle species with number density n_i the quantity $\delta(n_i/n_\gamma)$ vanishes. However, in addition to $\mathcal{R}$, multi-field inflation can stimulate isocurvature modes, $\mathcal{I}_i$, where at primordial times n_i/n_γ varies spatially [172, 241, 173, 102]. In this section we consider all possible non-decaying modes of this type [47]: cold dark matter density isocurvature (CDI); baryon density isocurvature (BDI); and neutrino density isocurvature (NDI) modes. For completeness, we also constrain the fourth non-decaying mode, neutrino velocity isocurvature (NVI), although there are no known mechanisms to excite it. Finally, we consider compensated isocurvature perturbations (CIP) between baryons and CDM [115, 116]. In this case, opposite BDI and CDI perturbations cancel in such a way that the total matter isocurvature perturbation vanishes and there is no first-order isocurvature

signal in the CMB. However, we utilize a higher-order lensing-like effect from this mode to obtain constraints on CIP from *Planck* temperature and polarization power spectra. We find the most powerful power-spectra-based constraints on this mode by exploiting the cosmological information in the low-L lensing potential reconstruction in Sect. 28.9.5, but leave the use of *Planck* trispectra in constraining CIP for future work.

As the positions of the peaks and dips of the CMB angular power spectra in the density isocurvature models are roughly in opposite phase compared to the pure adiabatic (ADI) spectrum, the primordial CDI, BDI, and NDI modes leave a very distinctive observational imprint on the CMB, whereas the imprint of the NVI mode more closely resembles the pure ADI mode; see, e.g., figure 43 in **(alias?)**. Prior to the detection of CMB anisotropies, studies such as Peebles & Yu [213] and Efstathiou & Bond [74, 75] discussed the possibility that isocurvature perturbations were the sole source of cosmological fluctuations. However, at least after the detection of the first acoustic peak in TT, it became clear that the density isocurvature mode(s) had to be subdominant [77, 78], while the adiabatic mode led to a good agreement with observations. Several pre-*Planck* isocurvature constraints were obtained [265, 217, 161, 16, 216, 274, 45, 199, 28, 159, 67, 25, 269, 154, 157, 273].

The mixture of curvature and isocurvature perturbations can be uncorrelated, but typically an arbitrary amount of correlation arises between them if the trajectory in field space is curved between Hubble radius exit and the end of multi-field inflation [111]. In extreme cases, such as the simplest curvaton models, there is full correlation or full anticorrelation between $\mathcal{R}$ and $\mathcal{I}$. In the following subsections, we start with the generic case of generally correlated adiabatic and CDI, NDI, or NVI perturbations. Then we deal with various special CDI (or BDI) cases with no correlation or full (anti)correlation.

We parameterize the primordial perturbations as in **(alias?)**, following the notation described there. The primary perturbation parameters scanned by `MultiNest` (in addition to the four standard ΛCDM background cosmological parameters and the *Planck* nuisance parameters) are the primordial abiabatic perturbation power and isocurvature perturbation power at two scales, corresponding to $k_1 = k_{\rm low} = 0.002\,{\rm Mpc}^{-1}$ and $k_2 = k_{\rm high} = 0.1\,{\rm Mpc}^{-1}$, namely, $\mathcal{P}^{(1)}_{\mathcal{RR}}$, $\mathcal{P}^{(2)}_{\mathcal{RR}}$, $\mathcal{P}^{(1)}_{\mathcal{II}}$, $\mathcal{P}^{(2)}_{\mathcal{II}}$, and the correlation power between $\mathcal{R}$ and $\mathcal{I}$ at k_1, i.e., $\mathcal{P}^{(1)}_{\mathcal{RI}}$. We assume a power-law form for the adiabatic and isocurvature power spectra and denote the spectral indices that can be calculated from the primary parameters by $n_{\mathcal{RR}}$ and $n_{\mathcal{II}}$. The correlation spectrum is also assumed to obey a power law, with spectral index $n_{\mathcal{RI}} = (n_{\mathcal{RR}} + n_{\mathcal{II}})/2$. Thus $\mathcal{P}^{(2)}_{\mathcal{RI}}$ is not an independent parameter. This ensures that the correlation fraction $\cos\Delta = \mathcal{P}_{\mathcal{RI}}/(\mathcal{P}_{\mathcal{RR}}\mathcal{P}_{\mathcal{II}})^{1/2}$ stays inside the interval $(-1, 1)$ at every k, as long as we reject any $\mathcal{P}^{(1)}_{\mathcal{RI}}$ which does not obey this requirement. While the correlation fraction is k-independent in our modelling, the primordial isocurvature fraction $\beta_{\rm iso}(k) = \mathcal{P}_{\mathcal{II}}(k)/\left[\mathcal{P}_{\mathcal{RR}}(k) + \mathcal{P}_{\mathcal{II}}(k)\right]$ depends on k, unless $n_{\mathcal{II}} = n_{\mathcal{RR}}$. We also report $\beta_{\rm iso}$ at an intermediate scale, $k_{\rm mid} = 0.05\,{\rm Mpc}^{-1}$.

We do not separately quote constraints on BDI or the total matter density isocurvature (MDI), since these modes are observationally indistinguishable from the CDI case.[175]

Numerical results for various isocurvature models and selected derived parameters are reported in Table 40, utilizing various data combinations. The table is divided into three main sections: generally correlated models (discussed in Sect. 28.9.2); one-isocurvature-parameter CDI models (discussed in Sects. 28.9.4 and 28.9.4); and, finally, two-isocurvature-parameter CDI models (discussed in Sects. 28.9.4, 28.9.4, and 28.9.4). For generally correlated CDI we study the stability of constraints (see Sect. 28.9.3) by using several different subsets of the *Planck*

[175] If we assume no NVI or NDI perturbations, then the MDI perturbation (i.e., the spatial perturbation in the relative number densities of matter particles and photons) is

$$\mathcal{I}_{\rm MDI} = \frac{\Omega_{\rm c}}{\Omega_{\rm m}}\mathcal{I}_{\rm CDI} + \frac{\Omega_{\rm b}}{\Omega_{\rm m}}\mathcal{I}_{\rm BDI}\,. \tag{1314}$$

As we will see, the posteriors for $\Omega_{\rm c}h^2$ and $\Omega_{\rm b}h^2$ are insensitive to the assumed initial conditions. Thus it is a good approximation to use the mean values obtained in the generally correlated mixed adiabatic and CDI model with TT,TE,EE+lowE+lensing data, namely $\Omega_{\rm c}/\Omega_{\rm m} \simeq 0.842$, $\Omega_{\rm b}/\Omega_{\rm m} \simeq 0.158$, $\Omega_{\rm c}/\Omega_{\rm b} \simeq 5.33$, and $(\Omega_{\rm c}/\Omega_{\rm b})^2 \simeq 28.4$. For example, to convert our CDI upper bound on $\mathcal{P}_{\mathcal{II}}$ to a BDI bound, we should multiply the constraint by $(\Omega_{\rm c}/\Omega_{\rm b})^2 = 28.4$, and to convert the CDI $\mathcal{P}_{\mathcal{RI}}$ to BDI, we should multiply the constraint by $\Omega_{\rm c}/\Omega_{\rm b} \simeq 5.33$. If $\beta_{\rm iso} \ll 1$, then this also can be converted to a BDI constraint by multiplying the CDI constraint by 28.4. The constraint on $\cos\Delta$ will be the same for the CDI and BDI cases, since the conversion factor cancels out.

data: (1) only high-ℓ TT; (2) high-ℓ TT+lensing; (3) TT,TE,EE; and (4) TT,TE,EE+lensing. For comparison, some *Planck* 2015 and WMAP results are also cited. Table 40 also includes comparisons to the pure adiabatic model in terms of the difference in the best-fit χ^2 and the natural logarithm of the Bayesian evidence ("model probability") ratios $\ln B$, negative $\ln B$ being evidence against the mixed models.[176]

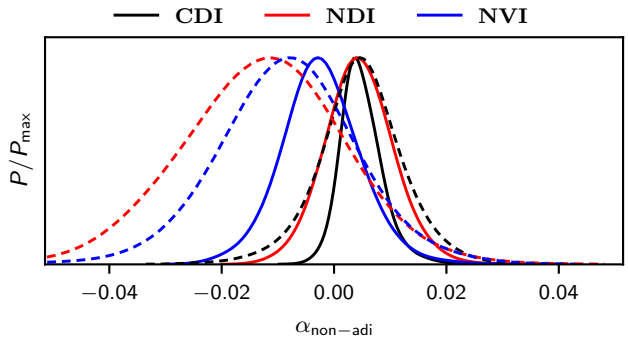


Figure 231: Posterior probability density of the observable non-adiabatic fraction of the CMB temperature variance, assuming a generally correlated mixed adiabatic and isocurvature model. These results used *Planck* TT+lowE+lensing data (dashed lines) and TT,TE,EE+lowE+lensing data (solid lines).

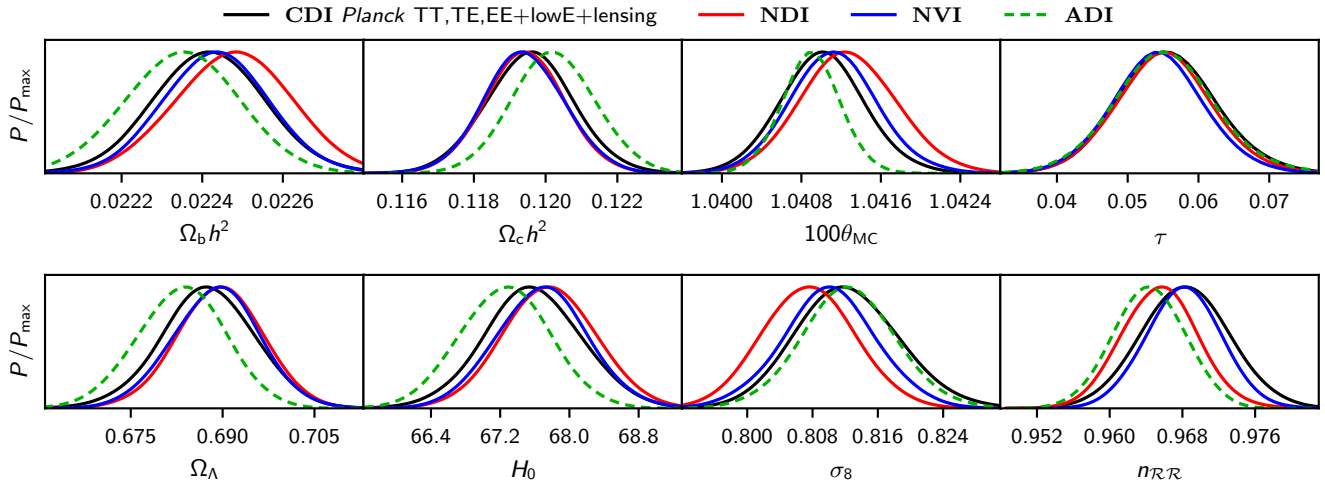


Figure 232: Comparison of posterior probability density for the standard cosmological parameters in mixed adiabatic and isocurvature models (solid lines) to those in pure adiabatic ΛCDM model (ADI, dashed green lines), using *Planck* TT,TE,EE+lowE+lensing data.

[176]The values of $\ln B$ depend on the priors. We adopt uniform priors in the range $(15, 40) \times 10^{-10}$ for the adiabatic, $(0, 100) \times 10^{-10}$ for the isocurvature, and $(-100, 100) \times 10^{-10}$ for the primordial correlation power parameter.

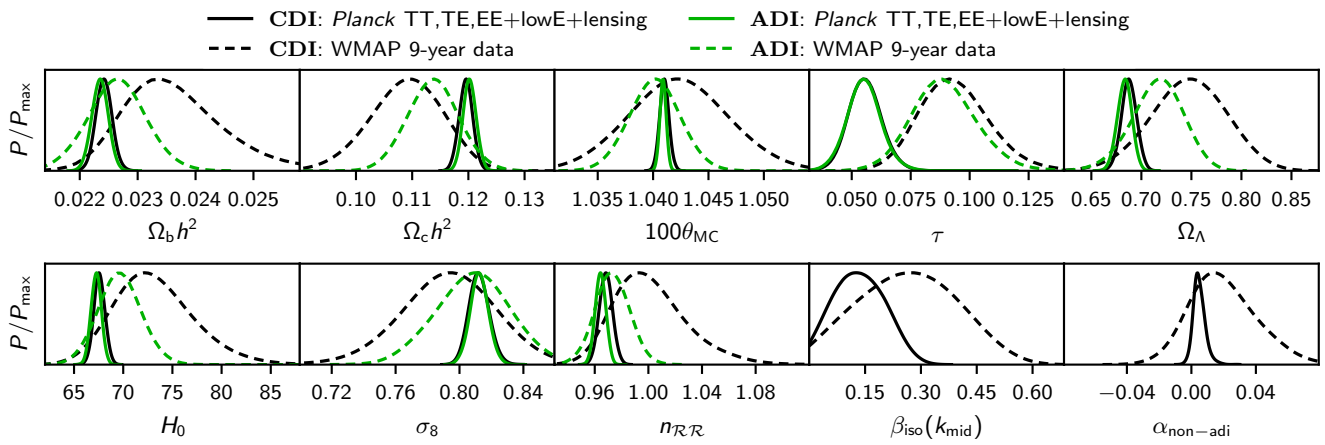


Figure 233: Comparison of posterior probability density for selected cosmological parameters with *Planck* data and pre-*Planck* (i.e., WMAP 9-year) data. Black lines indicate the results obtained for the generally correlated mixed CDI+ADI model, and green lines for the pure adiabatic model.

28.9.2 Results for generally correlated adiabatic and isocurvature modes

This subsection explores mixed adiabatic and isocurvature models where only one isocurvature mode at a time is considered. We consider the CDI, NDI, and NVI modes using the *Planck* 2018 TT(,TE,EE)+lowE(+lensing) data. All five primordial perturbation power amplitudes (of which three describe the isocurvature perturbations) are free parameters. It follows that $n_{\mathcal{II}}$ and $n_{\mathcal{RR}}$ are independent and $\cos\Delta$ varies between -1 and $+1$. The constraints for the primary perturbation parameters and the derived parameter $\mathcal{P}^{(2)}_{\mathcal{RI}}$ are shown in Fig. 229.

In all three cases the *Planck* TT+lowE+lensing results are very similar to the previous results from the *Planck* 2015 TT+lowP+lensing likelihood. As expected, the lower value of τ preferred by the 2018 (lowE) data is reflected in the adiabatic amplitudes $\mathcal{P}^{(1)}_{\mathcal{RR}}$ and $\mathcal{P}^{(2)}_{\mathcal{RR}}$. For CDI and NDI there is no significant shift in the constraints on isocurvature parameters, but we find slightly tighter constraints than in 2015. For NVI, a minor shift towards more negative correlations is observed (see the last two panels of Fig. 229c). As in 2015, adding the high-ℓ TE,EE data significantly tightens the constraints in all three cases.

When fitting the generally correlated three-isocurvature-parameter models, the *Planck* data are consistent with null detection, i.e., with the pure adiabatic model, $\left(\mathcal{P}^{(1)}_{\mathcal{II}}, \mathcal{P}^{(2)}_{\mathcal{II}}\right) = (0,0)$ and $\mathcal{P}^{(1)}_{\mathcal{RI}} = 0$ (and $\mathcal{P}^{(2)}_{\mathcal{RI}} = 0$). The natural logarithm of the ratio of model probabilities [i.e., the Bayes factor $\ln B = \ln(P_{\rm ISO}/P_{\rm ADI})$] is below -10.9, corresponding to odds of less than $1:54\,000$ for all three (CDI, NDI, NVI) models. If there were an undetected subdominant isocurvature contribution to the primordial perturbations, a negative correlation between $\mathcal{R}$ and $\mathcal{I}$ would be favoured, in particular for NDI and NVI (see the last two panels of Figs. 229a,b,c). With our sign convention, this leads to a negative contribution to the Sachs-Wolfe effect and hence reduces the amplitude of the temperature angular power spectrum at low multipoles.

Figure 230 updates the 2015 *Planck* constraints on the derived primordial fractions and spectral indices. At large scales we find with *Planck* TT,TE,EE+lowE+lensing that $\beta_{\rm iso}(k_{\rm low}) < 2.5\,\%$ for the CDI, $7.4\,\%$ for the NDI, and $6.8\,\%$ for the NVI model, all at $95\,\%$ CL. Figure 231 shows the non-adiabatic fraction in the observed CMB temperature variance, defined as

$$\alpha_{\rm non-adi} = 1 - \frac{(\Delta T)^2_{\mathcal{RR}}(\ell = 2, 2500)}{(\Delta T)^2_{\rm tot}(\ell = 2, 2500)}, \tag{1315}$$

where

$$(\Delta T)^2_{\rm X}(\ell = 2, 2500) = \sum_{\ell=2}^{2500} (2\ell+1) C^{TT}_{{\rm X},\ell}. \tag{1316}$$

The non-adiabatic fraction $|\alpha_{\rm non-adi}|$ is below 1.7 % with *Planck* TT,TE,EE+lowE+lensing data for all three cases at 95 % CL.

Since the *Planck* data do not allow a significant isocurvature contribution, the determination of standard cosmological parameters depends only very weakly on the assumed initial conditions, as seen in Fig. 232. We place this result in historical perspective in Fig. 233 (and Table 40) where the parameter determinations of the mixed CDI model and the pure adiabatic model are compared to the pre-*Planck* constraints set by the WMAP 9-year data.[177] *Planck* has dramatically tightened the constraint on the adiabatic spectral index. Its value is now 8.4σ below unity (scale-invariance) in the pure ADI case. Allowing for generally correlated CDI reduces the significance of this detection only slightly, to 7σ, whereas the WMAP 9-year data were consistent with a blue tilt as large as $n_{\mathcal{RR}} = 1.06$ at 95 % CL. The non-adiabatic contribution to the CMB temperature variance is constrained (about zero) 5 times more tightly than by WMAP. Finally, the allowed range for the sound horizon angle has shrunk by a factor of 10 in the CDI case, thanks to the *Planck* data covering more acoustic peaks beyond the first three peaks detected by WMAP.

[177] The pivot scales $k_{\rm low}$ and $k_{\rm high}$ used here to parameterize the primary perturbation amplitudes are not optimal for WMAP, since the WMAP data extend only to $k \simeq k_{\rm mid}$. Nevertheless, an analysis tailored to WMAP [see Savelainen et al. [248], who used $k_{\rm mid}$ as an upper pivot k] gives a similar posterior range for $\alpha_{\rm non-adi}$.

28.9.3 Role of lensing parameter $A_{\rm L}$ and likelihood choices

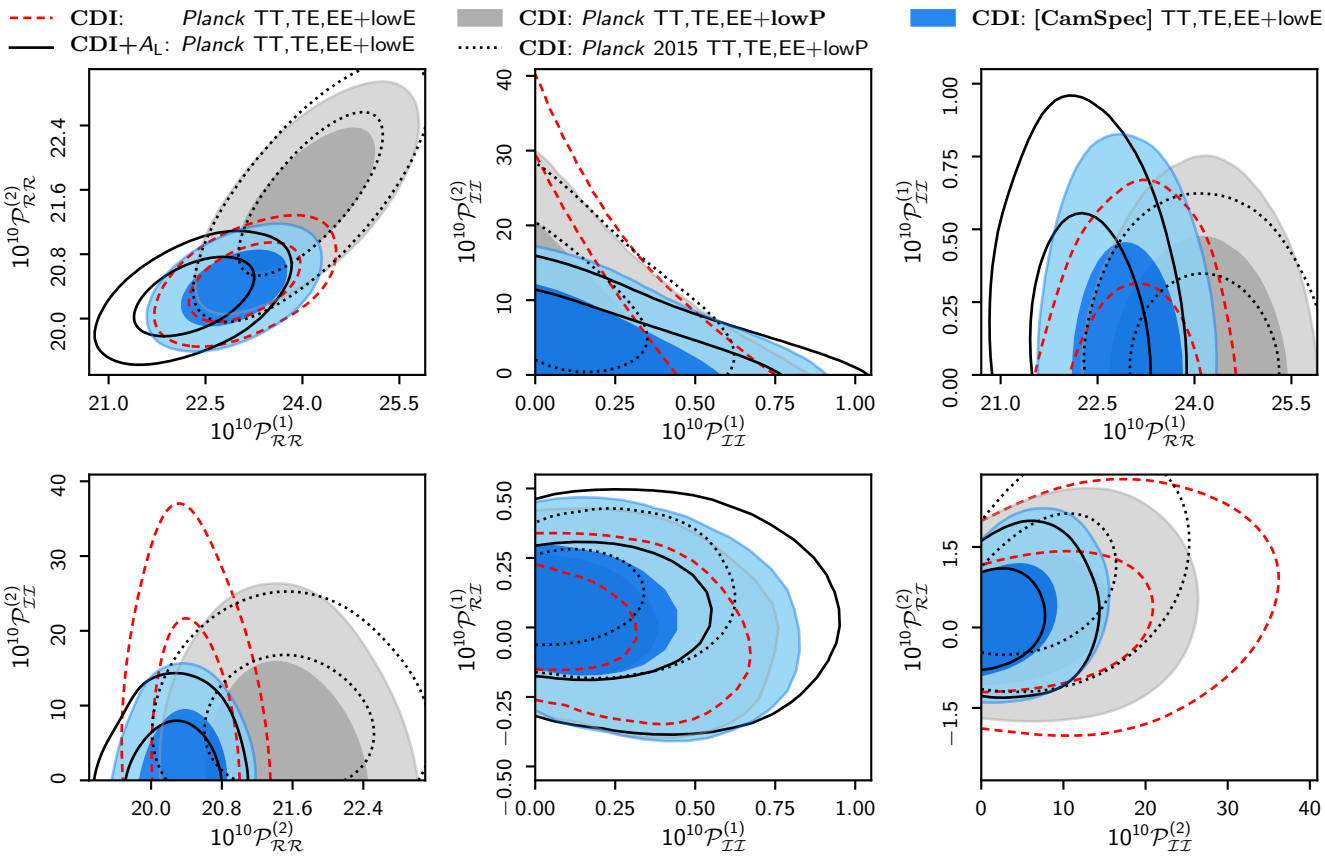


Figure 234: Comparison of the effect of *Planck* 2018 likelihood choices and phenomenological lensing amplitude, $A_{\rm L}$, on the constraints on the generally correlated mixed adiabatic and CDI model. The reference case, indicated by the red dashed curves, is for the 2018 baseline `Plik` high-ℓ likelihood supplemented by the low-ℓ `Commander` TT likelihood and the low-ℓ `SimAll` EE likelihood. This combination is the same as in Fig. 230a (red curves), except now without the lensing likelihood, which to some extent hides the differences between other likelihoods and the effect of $A_{\rm L}$. Black solid contours show the results using the same likelihood as in the reference case, but now for the mixed adiabatic and CDI model when simultaneously allowing $A_{\rm L}$ to vary. The remaining two curves are for the mixed adiabatic and CDI model (with $A_{\rm L} = 1$), but now changing the low-ℓ likelihood from `Commander` TT+`SimAll` EE to LFI 70 GHz T,E,B (grey), or high-ℓ likelihood from `Plik` to `CamSpec` (blue).

The small-scale primordial CDI amplitude is extremely sensitive to the details of the high-ℓ temperature and polarization power spectra and to choices made in constructing the likelihoods. Therefore the general CDI model serves as a robustness test of the *Planck* data and likelihoods. We now discuss a few curious aspects related to CMB lensing and likelihoods.

Lensing smooths the peaks of the CMB power spectra. This effect is taken into account in our theoretical predictions for the mixed adiabatic and isocurvature models by first calculating the total unlensed CMB spectra as a sum of adiabatic, isocurvature, and correlation C_ℓ's, and performing a similar summation for the lensing potential power spectrum [332, 220]. The total lensing potential is then used to lens the total CMB spectra. Starting with the WMAP data, accounting for CMB lensing became necessary for calculating constraints on isocurvature models, as [275] showed that there is a strong degeneracy between the lensing effect and the CDI contribution in the generally correlated mixed models. Fixing $n_{\mathcal{II}} = 1$ or $n_{\mathcal{II}} = n_{\mathcal{RR}}$ (as is done in the next subsection) makes this degeneracy disappear. This is because in these models the CDI contribution modifies only the low-ℓ part of the angular power spectra. The transfer

function mapping the primordial CDI mode to the TT (and EE) angular power is suppressed by a factor $(k/k_{\rm eq})^{-2} \sim (\ell/\ell_{\rm eq})^{-2}$ relative to the adiabatic mode. Therefore, to be observable at high ℓ, the CDI mode must be blue tilted ($n_{\mathcal{II}} > 1$). A blue-tilted CDI mode affects the total angular power spectra in a manner somewhat similar to lensing. Since the acoustic peaks of the CDI mode have the opposite phase compared to the adiabatic mode, a CDI admixture can "smooth" the peaks and dips of adiabatic acoustic oscillations. The NDI mode does not have precisely the opposite phase and is not damped relative to the adiabatic mode (see figure 43 in **(alias?)**). Thus we expect a weaker impact of lensing on the *primordial* NDI amplitude than in the CDI case. Therefore in this subsection we explore the general CDI model as an example.

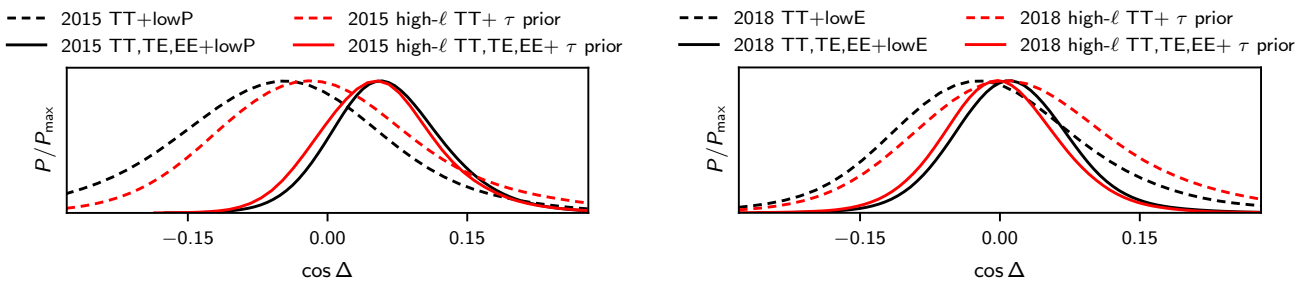


Figure 235: Scale-independent primordial correlation fraction in the mixed adiabatic and CDI model. The black curves are with *Planck* high-ℓ and low-ℓ data, while the red ones result from using only the high-ℓ `Plik` likelihood supplemented with a Gaussian prior on the optical depth. For the 2015 data (top panel) this prior was $\tau = 0.078 \pm 0.019$, whereas for the 2018 case (bottom panel) we have adopted $\tau = 0.055 \pm 0.007$ from the *Planck* 2018 TT,TE,EE+lowE+lensing CDI chain.

Starting with the *Planck* 2013 release, the consistency of the smoothing effect with the adiabatic ΛCDM model has been routinely tested by multiplying the lensing power spectrum by a phenomenological lensing consistency parameter, $A_{\rm L}$, prior to lensing the unlensed CMB spectra [66]. The expectation is that $A_{\rm L} = 1$. However, the *Planck* temperature and polarization data prefer a higher level of lensing-like smoothing ($A_{\rm L} > 1$) than expected in the adiabatic ΛCDM model. In the 2018 release [**(alias?)**] we have

$$A_{\rm L} = 1.243 \pm 0.096 \quad (68\,\%\ {\rm CL},\ \textit{Planck}\ {\rm TT{+}lowE}), \tag{1317}$$
$$A_{\rm L} = 1.180 \pm 0.065 \quad (68\,\%\ {\rm CL},\ \textit{Planck}\ {\rm TT,TE,EE{+}lowE}). \tag{1318}$$

Adding the *Planck* CMB lensing likelihood pulls these constraints towards $A_{\rm L} = 1$ (see also Table 41). The measurement of $A_{\rm L}$ when TT,TE,EE data are included depends on the calibration of the polarization channels. This procedure and the details of the sky masks differ between the *Planck* baseline `Plik` TT,TE,EE and the alternative *Planck* `CamSpec` TT,TE,EE likelihood, as discussed in **(alias?)** and **(alias?)**. The *Planck* `CamSpec` TT,TE,EE likelihood prefers a smaller value of $A_{\rm L}$ than `Plik`, but still lying about 2σ above unity.

Given the above motivation, we check the response of the generally correlated CDI model to the various possible choices of likelihoods available in the *Planck* 2018 release, and, on the other hand, we gauge how the baseline `Plik` likelihood reacts when allowing $A_{\rm L}$ to vary. For clarity, in Fig. 234 we restrict the analysis to high-ℓ TT,TE,EE and low-ℓ TT,EE(,BB) data without the lensing reconstruction data, but in Table 40 we also report TT+lowE and TT+lowP results, and include *Planck* lensing in some cases.

We notice a considerable variation in the constraints on the isocurvature power at high k, $\mathcal{P}^{(2)}_{\mathcal{II}}$, which corresponds to the high-ℓ region in the observed power spectra. In Fig. 234 we can compare the red dashed reference contours (obtained with the *Planck* baseline `Plik` TT,TE,EE+lowE likelihood) for the CDI model (where $A_{\rm L} = 1$) with the solid black contours (obtained with the same data) for the CDI+$A_{\rm L}$ model, where $A_{\rm L}$ is allowed to vary. In many cases, adding an extra free parameter is expected to weaken the constraints on the other parameters, but in this case adding $A_{\rm L}$ tightens the 95 % CL constraint on $\mathcal{P}^{(2)}_{\mathcal{II}}$ by a factor of 2.5 from 28.6×10^{-10} to 11.4×10^{-10}. This is reflected in the derived primordial isocurvature fraction

$\beta_{\rm iso}(k_{\rm high})$, whose upper bound changes from 0.58 to 0.36, according to Table 40. Therefore, we conclude that, when $A_{\rm L} = 1$, the CDI mode partially accounts for the extra lensing-like smoothing effect required by the *Planck* TT(,TE,EE) data. Once we allow the lensing amplitude to vary, there is not much need for the CDI contribution at high ℓ, which should be kept in mind when interpreting the results. In Table 40 we report four cases where $A_{\rm L}$ is allowed to vary. In all the other cases we have fixed $A_{\rm L} = 1$. In these cases the constraints at high k are "conservative", i.e., weaker than the *Planck* data were expected to be capable of [93], due to CDI (or NDI) partially fitting the lensing anomaly. Furthermore, again comparing the red dashed and black solid contours, we observe a slight weakening of the constraint for $\mathcal{P}^{(1)}_{\mathcal{II}}$ in the CDI+$A_{\rm L}$ model. This is due to the rigidity of the assumed power-law spectrum. When the high-ℓ data allow much less CDI, the low-ℓ (low-k) CDI amplitude can be larger without much affecting the middle-ℓ range of the CMB power spectra between the first and third acoustic peaks, which is the most sensitive region to departures from adiabaticity. (This is the same see-saw effect discussed in the end of Sect. 28.3.6 in the case of tensor perturbations.) Finally, in the $\left(\mathcal{P}^{(1)}_{\mathcal{RR}}, \mathcal{P}^{(2)}_{\mathcal{RR}}\right)$ panel we see a minor shift toward smaller amplitudes, which is an indication of the well-known degeneracy between $A_{\rm L}$ and the overall primordial perturbation amplitude.

From Table 40 it is obvious that adding the lensing data reduces the differences discussed above. This is again as expected, since the lensing data favour values of $A_{\rm L}$ only mildly above unity. For example, with *Planck* TT,TE,EE+lowE+lensing we obtain $\beta_{\rm iso}(k_{\rm high}) < 0.49$ for CDI+$A_{\rm L}$ and 0.47 for the CDI model.

We now proceed to a comparison of the likelihoods. The grey shaded contours in Fig. 234 indicate the results for the same CDI model as the red dashed contours, but changing the low-ℓ likelihood from the combination `Commander` TT+`SimAll` EE to the LFI 70 GHz pixel-based low T,E,B, which is by its methodology and construction very similar to the 2015 baseline low-ℓ likelihood (dotted contours). Indeed, this can be seen in the results: most of the isocurvature parameters follow more closely the 2015 results with this likelihood combination than with the 2018 baseline. This implies that when it comes to isocurvature, not much has changed in high-ℓ TT. 2018 lowP favours slightly smaller values of the optical depth τ than the 2015 version, hence the small shift towards smaller values of the adiabatic amplitude in the $\left(\mathcal{P}^{(1)}_{\mathcal{RR}}, \mathcal{P}^{(2)}_{\mathcal{RR}}\right)$ panel. With respect to the red dashed contours, the grey contours prefer higher adiabatic amplitudes and have a long degeneracy line in the $\left(\mathcal{P}^{(1)}_{\mathcal{RR}}, \mathcal{P}^{(2)}_{\mathcal{RR}}\right)$ plane. This is due to lowP having a higher central value and larger uncertainty on τ.

Finally, the blue shaded contours in Fig. 234 represent the results when using the `CamSpec` likelihood, to be compared to the red dashed contours obtained by the baseline `Plik` likelihood. All the other parameters shown are relatively stable against the high-ℓ likelihood, but $\mathcal{P}^{(2)}_{\mathcal{II}}$ stands out. `CamSpec` leads to an upper bound of 12.6×10^{-10}, whereas the baseline `Plik` result was 28.6×10^{-10}, or for $\beta_{\rm iso}(k_{\rm high})$ 0.38 versus 0.58 at 95 % CL, according to Table 40. This difference is not surprising, given the different responses of these likelihoods to $A_{\rm L}$ in the adiabatic ΛCDM+$A_{\rm L}$ model, and keeping in mind the $A_{\rm L}$–CDI degeneracy in the CDI model. However, this difference is not as concerning as it might appear at first sight: all the cases shown in Fig. 234 are fully consistent with zero isocurvature. It is only the upper bound that varies, with the baseline `Plik` likelihood and CDI model with $A_{\rm L} = 1$ leading to the most conservative (i.e., weakest or safest) upper bounds.

For CDI the 2015 release *Planck* high-ℓ TT data favoured a negative correlation fraction but the preliminary high-ℓ TT,TE,EE data favoured a slightly positive correlation. This was confirmed using only the high-ℓ `Plik` likelihood (and a prior on τ), as shown by the red curves in the top panel of Fig. 235. Including the low-ℓ data (black curves) did not significantly alter this tension between TT and TT,TE,EE results. In the present 2018 *Planck* release this tension has disappeared. Both high-ℓ TT and TT,TE,EE data lead to a correlation fraction posterior peaking at zero, as demonstrated in the bottom panel of Fig. 235. Including the low-ℓ TT data (black dashed curve) still shifts the posterior slightly towards negative values, due to the low TT power at low multipoles in the data.

28.9.4 Specific CDI models

In this subsection we constrain CDI models with only one or two isocurvature parameters. The two-parameter cases were not studied in the 2013 and 2015 *Planck* releases.

First we fix $n_{\mathcal{II}}$ to unity and assume no correlation between the CDI and adiabatic modes ("axion"), or we fix $n_{\mathcal{II}} = n_{\mathcal{RR}}$ and assume full (anti)correlation between the CDI and adiabatic modes ("curvaton I/II"). These models are less sensitive to any residual systematic effects in the high-ℓ data (such as the determination of polarization efficiencies or foreground modeling) than the generally correlated models, since CDI now modifies the angular power spectra insignificantly at $\ell \gtrsim 200$ (see figure 43 in **(alias?)**). As seen in the middle section of Table 40, the Bayesian evidence values for the one-parameter extensions of the adiabatic ΛCDM model are higher than for the three-parameter extensions, but all Bayes factors fall below −5. None of the one-parameter extensions improve χ^2 over the adiabatic ΛCDM model. The two-parameter extensions in the bottom section of Table 40 are even more strongly disfavoured, except for the uncorrelated case with free $n_{\mathcal{II}}$ ("axion II"), which is actually the only model that improves the best-fit χ^2 by slightly more than the number of extra parameters.

Uncorrelated ADI+CDI ("axion I")

Particularly insensitive to any $\ell \gtrsim 30$ data is the "axion I" case, since the CDI transfer function has a $(k/k_{\rm eq})^{-2}$ suppression and there is no correlation component whose amplitude would be higher than that of the isocurvature alone and hence would modify the adiabatic spectrum. The "axion I" case is achieved in our parameterization by setting $\mathcal{P}_{\mathcal{RI}} = 0$ and $\mathcal{P}^{(2)}_{\mathcal{II}} = \mathcal{P}^{(1)}_{\mathcal{II}}$. Thus the only varied isocurvature parameter is $\mathcal{P}^{(1)}_{\mathcal{II}}$. This uncorrelated case with $n_{\mathcal{II}} = 1$ is a good approximation for many multi-field inflationary models where the slow-roll parameter (in the isocurvature field perturbation direction) η_{ss} is negligible and the background trajectory in field space is straight between Hubble radius exit and the end of inflation. The predictions for the spectral indices (to first order in the slow-roll parameters) are $n_{\mathcal{RR}} = 1-6\epsilon+2\eta_{\sigma\sigma}$ and $n_{\mathcal{II}} = 1-2\epsilon+2\eta_{ss}$, where $\epsilon \geq 0$ and $\eta_{\sigma\sigma}$ is the second slow-roll parameter in the "adiabatic" direction (i.e., along the trajectory) in the field space. (An exact match with our model would require $\eta_{ss} = \epsilon$.) The axion model [see, e.g., a recent review by Marsh [179] and references therein], which was originally proposed to solve the strong CP problem and provides a dark matter candidate, can produce this type of isocurvature modes with $n_{\mathcal{II}} \simeq 1$ under the following assumptions (**(alias?)**, **(alias?)**): the Peccei-Quinn symmetry should be broken before inflation; it should not be restored by quantum fluctuations of the inflaton nor by thermal fluctuations when the Universe reheats; and axions produced through the misalignment angle should form a significant fraction of the dark matter.

Table 40 indicates a slight tightening of the "axion I" constraints using TT+lowE+lensing with respect to 2015 TT+lowP+lensing. This is due to the change of the baseline low-ℓ data from the 2015 LFI 70-GHz pixel-based T,E,B to the 2018 combination of `Commander` TT and `SimAll` EE, which in the generally correlated cases also gave tighter constraints at low k. As expected, the addition of high-ℓ TE,EE data only marginally improves the constraints, since the standard (non-isocurvature) parameters are better constrained now. For $n_{\mathcal{II}} = 1$ uncorrelated CDI, we obtain

$$\left.\begin{array}{r} \beta_{\rm iso}(k_{\rm mid}) < 0.038 \\ 0 \leq \alpha_{\rm non-adi} < 1.55\,\% \end{array}\right\} \begin{array}{l} (95\,\%\ \text{CL, } \textit{Planck} \\ \text{TT,TE,EE} \\ \text{+lowE+lensing}). \end{array} \tag{1319}$$

Using equation (73) of **(alias?)**, we convert the constraint on the primordial isocurvature fraction to a bound on the inflationary energy scale. If all the dark matter is in axions, the above $\beta_{\rm iso}(k_{\rm mid})$ constraint corresponds to the same limit we quoted in 2015, that is,

$$H_{\rm inf} < 0.86 \times 10^7\,{\rm GeV} \left(\frac{f_a}{10^{11}\,{\rm GeV}}\right)^{0.408} \qquad (95\,\%\ \text{CL})\,, \tag{1320}$$

where $H_{\rm inf}$ is the expansion rate at Hubble radius exit of the scale corresponding to $k_{\rm mid}$ and f_a is the Peccei-Quinn symmetry-breaking energy scale.

Fully (anti)correlated ADI+CDI ("curvatonI/II")

If $n_{\mathcal{II}} = n_{\mathcal{RR}}$, the low-$\ell$ data are maximally sensitive to the fully correlated isocurvature perturbations. In this case the correlation component is a geometric average of the adiabatic and isocurvature components, and hence much larger than the isocurvature component alone. We achieve this case in our parameterization by setting $\mathcal{P}^{(2)}_{\mathcal{II}} = \left(\mathcal{P}^{(2)}_{\mathcal{RR}}/\mathcal{P}^{(1)}_{\mathcal{RR}}\right)\mathcal{P}^{(1)}_{\mathcal{II}}$ and $\mathcal{P}^{(1)}_{\mathcal{RI}} = \pm\left(\mathcal{P}^{(1)}_{\mathcal{RR}}\mathcal{P}^{(1)}_{\mathcal{II}}\right)^{1/2}$, i.e., $\cos\Delta = \pm 1$. The only isocurvature parameter to be varied is again $\mathcal{P}^{(1)}_{\mathcal{II}}$. Since $n_{\mathcal{II}} = n_{\mathcal{RR}}$, the derived isocurvature fraction $\beta_{\rm iso}$ is independent of k. A physically motivated example of this type of model is the simplest curvaton model, where a light scalar field χ that is subdominant (and hence irrelevant for the inflationary dynamics) starts to oscillate at the bottom of its potential after the end of inflation, causing its average energy density to evolve like non-relativistic matter. Once fully (or almost fully) dominating the energy density of the Universe, this curvaton field decays either to CDM or to other species [198, 173, 79, 201, 177, 20, 176]. The amount of isocurvature and non-Gaussianity present after curvaton decay depends on the "curvaton decay fraction," $r_D = 3\bar\rho_\chi/(3\bar\rho_\chi + 4\bar\rho_{\rm radiation})$, evaluated at curvaton decay time. Under a number of (very) restrictive assumptions discussed in **(alias?)**, the curvaton model can lead to fully (anti)correlated CDI (or BDI) and adiabatic perturbations.

Not surprisingly, both in the fully correlated and anticorrelated cases, the constraint on $\beta_{\rm iso}$ is much (about 40 times) stronger than in the uncorrelated case. At 95 % CL, *Planck* TT,TE,EE+lowE+lensing leads to

$$\beta_{\rm iso} < 0.00095 \quad \text{(Fully correlated, "curvaton I")}, \tag{1321}$$

$$\beta_{\rm iso} < 0.00107 \quad \text{(Fully anticorrelated, "curvaton II")}, \tag{1322}$$

both rounded to 0.001 in Table 40. As in 2015, the TT data favour anticorrelation, due to the low power in the low-ℓ temperature compared to the expectation of the adiabatic ΛCDM model. But when the TE,EE data (which do not particularly favour negative correlation) are added, a very tight (one part per thousand) constraint on the primordial isocurvature fraction results.

Fully correlated perturbations are obtained, e.g., in case 4 described in Gordon & Lewis [109]. Many models giving anticorrelation produce too large an isocurvature fraction to be consistent with the above limit, but case 9 of Gordon & Lewis [109] survives. After the curvaton decay, the primordial isocurvature fraction in these models will be $\beta_{\rm iso} \simeq 9(1-\tilde r)^2/[r_D^2 + 9(1-\tilde r)^2]$, where $\tilde r = r_D$ for the fully correlated CDI case and $\tilde r = r_D/R_{\rm c} \geq 1$ for the fully anticorrelated CDI case, and $R_{\rm c} = \bar\rho_{\rm c}/(\bar\rho_{\rm c} + \bar\rho_{\rm b})$ is the CDM fraction of the total non-relativistic matter.

On the other hand, the nonlinearity parameter describing non-Gaussianity is [246]

$$f_{\rm NL}^{\rm local} = \left(1 + \Delta_s^2\right)\frac{5}{4r_D} - \frac{5}{3} - \frac{5r_D}{6}, \tag{1323}$$

where $\Delta_s^2 = \langle\delta\chi^2\rangle_s/\bar\chi^2$ is the small-scale variance of the curvaton perturbations, or the ratio of the energy density carried by the curvaton particles to the energy density of the curvaton field (if there is significant production of curvaton particles). The parameter $f_{\rm NL}^{\rm local}$ cannot be smaller than $-5/4$, which is obtained when $r_D = 1$ and $\Delta_s^2 = 0$, as implicitly assumed, for example, in Bartolo et al. [22, 23]. The above $\beta_{\rm iso}$ limits correspond to the following r_D and $f_{\rm NL}^{\rm local}$ constraints (assuming $\Delta_s^2 = 0$):[178]

$$0.98982 < r_D \leq 1 \qquad \text{("curvaton I")}$$
$$\Rightarrow -1.2500 \leq f_{\rm NL}^{\rm local} < -1.2287, \tag{1324}$$
$$0.84347 \leq r_D < 0.85129 \qquad \text{("curvaton II")}$$
$$\Rightarrow -0.9077 < f_{\rm NL}^{\rm local} \leq -0.8876. \tag{1325}$$

[178] The quoted precision of these constraints does not reflect the precision of the sampling of the likelihood surface, but we report several digits here since the constraints are so tight that rounding to, e.g., two significant digits would give empty or almost empty ranges.

Even with the maximal allowed isocurvature fraction, the local non-Gaussianity in the curvaton model is well within the observational *Planck* limits presented in Planck Collaboration IX [295]. The residual isocurvature peturbations in the two studied curvaton models set much tighter constraints on the curvaton decay fraction than do constraints on the observed (consistent with zero) non-Gaussianity.

Uncorrelated ADI+CDI with free $n_{\mathcal{II}}$ ("axion II")

Axion models do not necessarily produce nearly scale-invariant isocurvature perturbations. In particular, even highly blue-tilted spectra (in the observable CMB range) are possible. For example, Kasuya & Kawasaki [152] construct a model with $n_{\mathcal{II}} = 2$–4. This motivates studying a two-isocurvature-parameter model, where adiabatic and isocurvature modes are uncorrelated, but the isocurvature fraction and spectral index are free to vary. In our parameterization this is achieved by setting $\mathcal{P}^{(1)}_{\mathcal{RI}} = 0$, and varying $\mathcal{P}^{(1)}_{\mathcal{II}}$ and $\mathcal{P}^{(2)}_{\mathcal{II}}$ independently. The results for this model are presented in the first two rows of the third section of Table 40. The low-ℓ temperature data do not favour any extra contribution beyond the (already too high) abiabatic contribution, whereas the fit to the high-ℓ temperature and polarization data can be improved slightly by the "smoothing" caused by the CDI mode. This leads to a very blue isocurvature spectrum. *Planck* TT,TE,EE+lowE+lensing gives at 95 % CL $1.55 < n_{\mathcal{II}} < 3.67$, consistent with the recent findings of Chung & Upadhye [60]. Even the very large upper bound $\beta_{\rm iso}(k_{\rm high}) < 77\,\%$ corresponds to a contribution of less than order 1% to the observable CMB TT (or EE) power spectra at $\ell \simeq 1400$. The uncertainty in the *Planck* TT spectrum at these high multipoles is $\Delta\mathcal{D}^{TT}_\ell \sim 10\,\mu{\rm K}^2$ and the actual spectrum is $\mathcal{D}^{TT}_\ell \sim 1000\,\mu{\rm K}^2$. Thus the allowed CDI contribution is only of the same 1 % order as the observable uncertainty. Consequently the non-adiabatic contribution to the observed CMB temperature variance, $\alpha_{\rm non-adi}$, is also vanishingly small, between 7×10^{-4} and 7×10^{-3}.

Arbitrarily correlated ADI+CDI with $n_{\mathcal{II}} = n_{\mathcal{RR}}$ ("curvaton III")

Apart from the extremes of ±100 % correlation, some curvaton models predict an arbitrary degree of correlation. The generic feature of most curvaton models is that the isocurvature and adiabatic spectral indices are equal. This is because both perturbations typically arise from the same source. In the next-to-simplest models, the correlation fraction can be written as $\cos\Delta = \sqrt{\lambda/(1+\lambda)}$, where $\lambda = (8/9)r_D^2\epsilon_*(M_{\rm Pl}/\bar\chi_*)^2$. Therefore, the model is fully correlated only if $\lambda \gg 1$, in which case the results of "curvaton I" apply. If the slow-roll parameter ϵ_* is very close to zero or the curvaton field value $\bar\chi_*$ is large compared to the Planck mass, this model leads to almost uncorrelated perturbations and the constraints are well approximated by "axion I." Any other case leads to an arbitrary degree of positive correlation between the CDI and adiabatic modes.

Modulated reheating with thermal or non-thermal production of gravitinos can lead to positive or negative correlation, respectively [266]. While the correlation could in principle be arbitrarily large, the observational constrains on $\beta_{\rm iso}$ favour only small correlations.

Arbitrarily correlated ADI+CDI with $n_{\mathcal{II}} = n_{\mathcal{RR}}$ is also a good approximation for those two-field (or multi-field) slow-roll models [e.g., double quadratic inflation [160, 29]] where the trajectory in field space is curved between the Hubble radius exit of perturbations during inflation and the end of inflation. The fraction of isocurvature perturbations converted to adiabatic depends on how the trajectory is curved and this part of the adiabatic perturbations will be fully (anti)correlated with the isocurvature modes, whereas the adiabatic perturbations already present at Hubble radius exit are uncorrelated with isocurvature modes to first order in the slow-roll parameters, and only slightly correlated to second order. [See, e.g., Gordon et al. [111], Amendola et al. [16], van Tent [276], and Byrnes & Wands [52].] The result is a non-zero correlation between isocurvature and total adiabatic perturbations. The spectral indices of both components are typically $1 - \mathcal{O}$(slow-roll parameters), which is well approximated by $n_{\mathcal{II}} = n_{\mathcal{RR}}$ since the data indicate $n_{\mathcal{RR}} \simeq 0.965$.

As expected, the *Planck* data favour negative correlations, since these $n_{\mathcal{II}} = n_{\mathcal{RR}}$ models modify only the low-ℓ part of the CMB spectra, where TT power is lower than predicted by the adiabatic ΛCDM model. With TT,TE,EE+lowE+lensing we find, at 95 % CL, $\beta_{\rm iso} < 0.039$ and $-0.41 < \cos\Delta < 0.31$.

Fully (anti)correlated ADI+CDI with free $n_{\mathcal{II}}$

The remaining two-parameter CDI extensions of the adiabatic ΛCDM model are those where the perturbations are fully (anti)correlated, as in the simplest curvaton models, but the isocurvature spectral index is not fixed to the adiabatic one. In this case the free isocurvature parameters are $\mathcal{P}^{(1)}_{\mathcal{II}}$ and $\mathcal{P}^{(2)}_{\mathcal{II}}$, while $\mathcal{P}^{(1)}_{\mathcal{RI}} = \pm\left(\mathcal{P}^{(1)}_{\mathcal{RR}}\mathcal{P}^{(1)}_{\mathcal{II}}\right)^{1/2}$. These models are somewhat difficult to motivate, since full (anti)correlation typically implies that the curvature and isocurvature perturbations have their origin in (the decay products of) the same field. Then one would expect equal spectral indices, as in the curvaton model. The conversion of isocurvature perturbations to adiabatic ones (e.g., between Hubble radius exit and the end of inflation, or by curvaton-type decay, or by reheating/thermalization) should be scale dependent in order to obtain $n_{\mathcal{II}} \neq n_{\mathcal{RR}}$. Slow-roll two-field inflation leads to an exact match, $n_{\mathcal{II}} = n_{\mathcal{RR}}$, in the case where $\cos^2\Delta = 1$. [See, e.g., Byrnes & Wands [52]]. Nevertheless, for completeness we report constraints on these phenomenological models in the last four rows of Table 40. Since the low-ℓ TT data favour negative correlation, a larger isocurvature fraction is allowed in the fully anticorrelated case at low k. This leads to scale-invariant isocurvature perturbations being in the favoured region of parameter space, namely $-0.28 < n_{\mathcal{II}} < 1.86$ with TT,TE,EE+lowE+lensing at 95 % CL. In contrast, in the fully correlated case the low-ℓ TT data disfavour any isocurvature contribution, and hence prefer a blue spectrum, with $1.37 < n_{\mathcal{II}} < 3.65$.

28.9.5 Compensated BDI–CDI mode

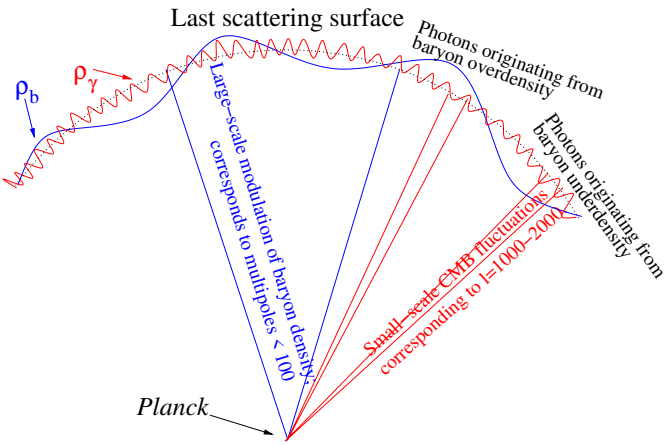


Figure 236: Illustration of how the large-scale modulation of the baryon density by CIP gets converted into a small-scale "smoothing" effect of the temperature and polarization anisotropies.

This subsection presents constraints on uncorrelated adiabatic and scale-invariant CIP modes and discusses the strong degeneracy between the phenomenological lensing parameter $A_{\rm L}$ and the CIP amplitude [357]. Assuming that there are no NVI or NDI perturbations, the total matter density isocurvature perturbation $\mathcal{I}_{\rm MDI}$, given by Eq. (1314), vanishes if

$$\mathcal{I}_{\rm CDI} = -\frac{\Omega_{\rm b}}{\Omega_{\rm c}}\mathcal{I}_{\rm BDI}. \tag{1326}$$

This mode, where the anticorrelated CDI and BDI perturbations cancel even though their individual amplitudes can be large, is called a compensated baryon and cold dark matter isocurvature mode. The CIP mode does not leave a linear-order isocurvature signal in the CMB or matter power spectra [109], although it modifies the trispectrum [115, 117]. However, at the next order there is a smoothing effect on the high-ℓ TT, TE, and EE spectra. A formal derivation can be found, for example, in Smith et al. [345]. Here we summarize the heuristic arguments of Muñoz et al. [243].

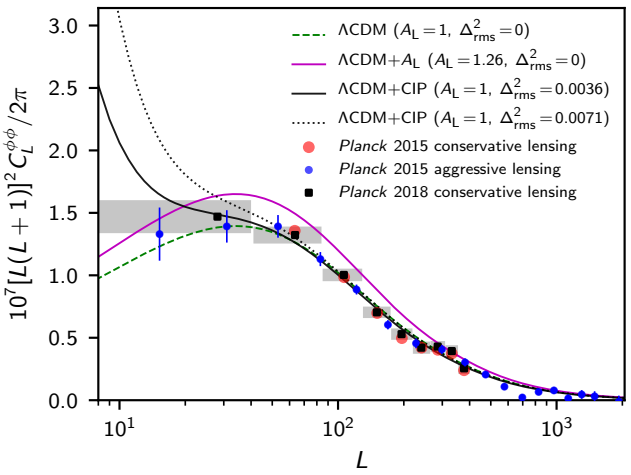


Figure 237: Conservative *Planck* 2015 lensing data (red points), aggressive *Planck* 2015 lensing data (blue points with error bars), and conservative *Planck* 2018 lensing data (black squares in grey boxes), along with the best-fit models to the *Planck* data: the best-fit adiabatic ΛCDM model to 2018 TT+lowE (green dashed line); the best-fit ΛCDM+A_L model to 2018 TT+lowE (magenta solid line, $A_L = 1.26$); and the best-fit ΛCDM+CIP model to 2018 TT,TE,EE+lowE and conservative lensing data (black solid line, $\Delta^2_{rms} = 0.0036$) and to 2015 TT,TE,EE+lowP and conservative lensing data (black dotted line, $\Delta^2_{rms} = 0.0071$). As CIP modifies only the very low-L part of the lensing power spectrum, the conservative 2015 lensing data ($40 \leq L \leq 400$) are insensitive to CIP even when $\Delta^2_{rms} = 0.0071$. On the other hand, the first two data points of the 2015 aggressive lensing data disfavour the large CIP amplitude [345], which gives a very good fit to all the other data. *Planck* 2018 conservative lensing data cover the range $8 \leq L \leq 400$ and consequently disfavour CIP variances $\Delta^2_{rms} \gtrsim 0.004$.

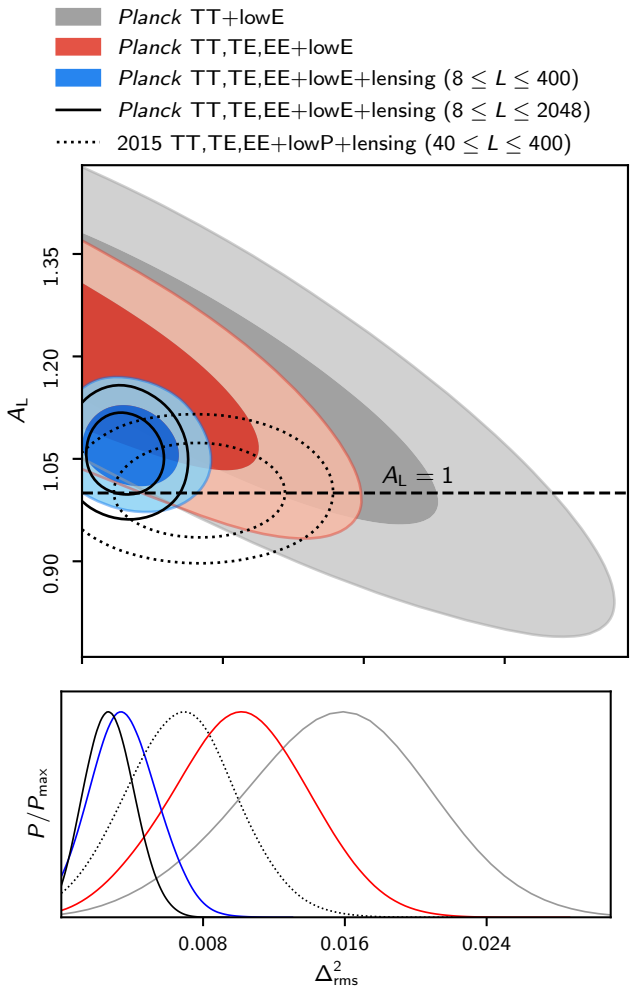


Figure 238: Degeneracy between $A_{\rm L}$ and $\Delta^2_{\rm rms}$ in the ΛCDM+$A_{\rm L}$+CIP model (top panel) and constraints on $\Delta^2_{\rm rms}$ in the ΛCDM+CIP model, where $A_{\rm L} = 1$ (bottom panel).

On scales larger than the sound horizon, condition (1326) is preserved until last scattering and can be written as

$$\delta\rho_{\rm c}(\vec{x}) \simeq -\delta\rho_{\rm b}(\vec{x})\,. \tag{1327}$$

Consequently, CIP can be described as a large scale modulation of the baryon and CDM density [243, 136, 357], with

$$\Omega_{\rm b}(\vec{\hat{n}}) = [1+\Delta(\vec{\hat{n}})]\bar{\Omega}_{\rm b}\,, \qquad \Omega_{\rm c}(\vec{\hat{n}}) = \bar{\Omega}_{\rm c} - \Delta(\vec{\hat{n}})\bar{\Omega}_{\rm b}\,. \tag{1328}$$

Here the overbar denotes an average over the whole sky and $\Delta(\vec{\hat{n}})$ a small perturbation about this average in the direction $\vec{\hat{n}}$, as illustrated in Fig. 236. In patches of sky where the CMB photons originate from baryon-overdense regions, the odd acoustic peaks at high-ℓ are more pronounced relative to the even peaks compared to the patches where the photons originate from baryon-underdense regions. Averaging over the sky leads to a lensing-like smoothing of the high-ℓ peaks.

A convenient measure of CIP is the variance $\Delta^2_{\rm rms} \equiv \langle|\Delta(\hat{n})|^2\rangle \simeq \mathcal{P}_{\mathcal{I}_{\rm BDI}\mathcal{I}_{\rm BDI}}$. If Δ is a Gaussian random variable, the observed angular power of TT, TE, or EE will be

$$\begin{aligned} &C_\ell^{\rm obs}(\bar{\Omega}_{\rm b}, \bar{\Omega}_{\rm c}, \tau, H_0, n_{\rm s}, A_{\rm s}) \\ &= \frac{1}{\sqrt{2\pi\Delta^2_{\rm rms}}} \int C_\ell\big(\Omega_{\rm b}(\Delta), \Omega_{\rm c}(\Delta), \tau, H_0, n_{\rm s}, A_{\rm s}\big)\, e^{-\Delta^2/(2\Delta^2_{\rm rms})} d\Delta\,, \end{aligned} \tag{1329}$$

where $\Omega_{\rm b}(\Delta) = (1+\Delta)\bar{\Omega}_{\rm b}$ and $\Omega_{\rm c}(\Delta) = \bar{\Omega}_{\rm c} - \bar{\Omega}_{\rm b}\Delta$. For brevity, we will denote the power spectrum in the integrand by $C_\ell|_{\Delta=\delta}$. For each δ it can be calculated by assuming adiabatic

initial conditions. Approximating the integrand by the first three terms of its Taylor series about $\Delta = 0$, we end up with

$$C_\ell^{\rm obs} \simeq C_\ell|_{\Delta=0} + \frac{1}{2}\left.\frac{d^2 C_\ell}{d\Delta^2}\right|_{\Delta=0} \Delta^2_{\rm rms}\,. \tag{1330}$$

In the following we describe parameter scans where we vary the six standard (adiabatic) ΛCDM parameters, the *Planck* nuisance parameters, and the CIP variance $\Delta^2_{\rm rms}$, calling this one-parameter extension of the ΛCDM model the "ΛCDM+CIP" model. We evaluate the right-hand side of Eq. (1330) at each point in parameter space using a finite-difference approximation for the second derivative:

$$\frac{1}{2}\left.\frac{d^2 C_\ell}{d\Delta^2}\right|_{\Delta=0} \Delta^2_{\rm rms} \simeq \frac{\frac{1}{2}C_\ell|_{\Delta=\delta} - C_\ell|_{\Delta=0} + \frac{1}{2}C_\ell|_{\Delta=-\delta}}{\delta^2}\,, \tag{1331}$$

where δ should be "sufficiently small." In practice, good numerical accuracy is achieved if δ is of order $\sqrt{\Delta^2_{\rm rms}}$. So at each point in our `MultiNest` scan we set $\delta = \sqrt{\Delta^2_{\rm rms}}$ for the point currently under evaluation, and thus the result of Eq. (1330) simplifies to

$$C_\ell^{\rm obs} \simeq \frac{C_\ell\big|_{\Delta=\sqrt{\Delta^2_{\rm rms}}} + C_\ell\big|_{\Delta=-\sqrt{\Delta^2_{\rm rms}}}}{2}\,. \tag{1332}$$

With this method each angular power spectra evaluation takes twice as long as for the pure adiabatic case since the spectra are now an average of two spectra, resulting from different values of $\Omega_{\rm b}$ and $\Omega_{\rm c}$.[179]

Unlike the high-ℓ TT, TE, and EE spectra, the high-L lensing potential power spectrum is virtually unaffected by CIP. Instead, CIP modifies the low multipoles of $[L(L+1)]^2 C_L^{\phi\phi}/(2\pi)$ by, approximately, adding a term $\Delta^2_{\rm rms} \times (L/0.053)^{-2}$. For details, see table II in Smith et al. [345]. As illustrated in Fig. 237, when using the *Planck* 2015 conservative lensing data ($40 \le L \le 400$) this term does not affect the results. In contrast, the *Planck* 2018 conservative lensing data also contain the range $8 \le L < 40$ and thus CIP variances $\Delta^2_{\rm rms} \gtrsim 0.004$ fit the first data point of the 2018 lensing power spectrum ($8 \le L \le 400$) worse than in ΛCDM. However, even in this case the joint fit of the ΛCDM+CIP model to the TT, TE, EE, and lensing data is better than that of the ΛCDM model, the improvement being of the same order as for the ΛCDM+$A_{\rm L}$ model.

The top panel of Fig. 238 shows the $A_{\rm L}$–$\Delta^2_{\rm rms}$ degeneracy in the ΛCDM+$A_{\rm L}$+CIP model and how it can be broken by the lensing data. The value $A_{\rm L} = 1$ provides a good fit to the TT+lowE data, if $\Delta^2_{\rm rms} \simeq 0.016$, and to the TT,TE,EE+lowE data, if $\Delta^2_{\rm rms} \simeq 0.010$. The ΛCDM+CIP model (where $A_{\rm L} = 1$) with $\Delta^2_{\rm rms} \simeq 0.008$ provides a better simultaneous fit to the *Planck* 2015 TT,TE,EE and conservative lensing data ($40 \le L \le 400$) than does the ΛCDM+$A_{\rm L}$ model. When using the *Planck* 2018 conservative lensing data ($8 \le L \le 400$), the best-fit value of $\Delta^2_{\rm rms}$ decreases to 0.0036. This is due to the extra term $\propto L^{-2}$ brought by CIP to the lensing power estimator, as discussed above and shown in Fig. 237.

Since *Planck* TT,TE,EE+lowE and lensing data can be fit well by $A_{\rm L} = 1$ in the CIP model, we show in the bottom panel of Fig. 238 the one-dimensional posterior of $\Delta^2_{\rm rms}$ in the ΛCDM+CIP model. A non-zero value of $\Delta^2_{\rm rms}$ is preferred at the 2.9σ (2.5σ) level by *Planck* 2018 (2015) TT+lowE(lowP) data and at the 2.6σ (1.8σ) level by the TT,TE,EE+lowE(lowP) data. Without lensing the 2018 data thus more strongly favour the non-zero CIP amplitude than the 2015 data, which is as we would expect, since the favoured $A_{\rm L}$ value in the ΛCDM+$A_{\rm L}$ model has also increased. When using 2018 TT,TE,EE+lowE and the 2018 conservative lensing data the significance decreases to 2.0σ, while switching to the aggressive lensing data ($8 \le L \le 2048$) leads to 2.1σ. The 68 % CL ranges of $\Delta^2_{\rm rms}$ in the ΛCDM+CIP model, obtained with the baseline high-ℓ `Plik` likelihood in combination of other *Planck* data, are highlighted in Table 41. Replacing `Plik` with `CamSpec` (in particular `CamSpec` TT,TE,EE) leads to somewhat lower

[179] Actually, we call `CAMB` three times (i.e., also with $\bar{\Omega}_{\rm b}$ and $\bar{\Omega}_{\rm c}$, or $\Delta = 0$), in order to obtain some auxiliary parameters such as σ_8 correctly, although we do not report them here.

Data and model	Constraints $1000\Delta^2_{\rm rms}$	Constraints $A_{\rm L}$	Best fit $1000\Delta^2_{\rm rms}$	Best fit $A_{\rm L}$	Best fit $\Delta\chi^2$	$\ln B$
TT+lowE						
ΛCDM+CIP	$\mathbf{15.5^{+5.3}_{-5.4}}$		15.5	1.00	−6.9	1.7
ΛCDM+$A_{\rm L}$		$1.24^{+0.10}_{-0.10}$	0.0	1.26	−8.7	2.1
ΛCDM+$A_{\rm L}$+CIP	< 24.4	$1.12^{+0.14}_{-0.12}$	3.4	1.23	−8.8	0.6
TT,TE,EE+lowE						
ΛCDM+CIP	$\mathbf{10.1^{+3.9}_{-3.9}}$		10.0	1.00	−5.7	0.9
ΛCDM+$A_{\rm L}$		$1.18^{+0.07}_{-0.07}$	0.0	1.19	−9.7	2.5
ΛCDM+$A_{\rm L}$+CIP	< 12.7	$1.13^{+0.09}_{-0.08}$	0.3	1.19	−9.7	0.2
TT,TE,EE+lowE+lensing (conserv.)						
ΛCDM+CIP	$\mathbf{3.7^{+1.6}_{-2.1}}$		3.6	1.00	−3.3	−1.4
ΛCDM+$A_{\rm L}$		$1.07^{+0.04}_{-0.04}$	0.0	1.07	−3.4	−1.2
ΛCDM+$A_{\rm L}$+CIP	$3.1^{+1.4}_{-2.0}$	$1.07^{+0.04}_{-0.04}$	2.9	1.07	−6.4	−2.8
TT,TE,EE+lowE+lensing (aggr.)						
ΛCDM+CIP	$\mathbf{2.8^{+1.2}_{-1.5}}$		2.7	1.00	−4.0	−1.2
ΛCDM+$A_{\rm L}$		$1.06^{+0.04}_{-0.04}$	0.0	1.05	−2.2	−1.8
ΛCDM+$A_{\rm L}$+CIP	$2.6^{+1.2}_{-1.6}$	$1.06^{+0.04}_{-0.04}$	2.4	1.06	−6.2	−3.0

Table 41: Comparison of ΛCDM+CIP, ΛCDM+$A_{\rm L}$, and ΛCDM+$A_{\rm L}$+CIP models with various *Planck* datasets, when using the baseline `Plik` likelihood at high ℓ. The first two columns ("Constraints") are the 68% CL ranges or the 95% CL upper bounds on $1000\Delta^2_{\rm rms}$ (highlighted in bold for ΛCDM+CIP) and $A_{\rm L}$. The remaining columns give the best-fit pameter values, and the difference of the best-fit χ^2 and the difference of the log of the Bayesian evidence with respect to the pure adiabatic ΛCDM model. A negative $\Delta\chi^2$ means that the quoted model fits the data better than ΛCDM, while a positive $\ln B$ means that the Bayesian model comparison favours the quoted model, when adopting the uniform priors $0 \le 1000\Delta^2_{\rm rms} < 75$ and $0.3 \le A_{\rm L} \le 1.7$.

values,

$$1000\Delta^2_{\rm rms} = \begin{cases} 14.1^{+5.2}_{-5.2} \text{ TT+lowE}, \\ 6.5^{+3.0}_{-4.2} \text{ TT,TE,EE+lowE}, \\ 2.8^{+1.2}_{-2.2} \text{ TT,TE,EE+lowE+lensing (conserv.)}, \\ 2.2^{+1.0}_{-1.5} \text{ TT,TE,EE+lowE+lensing (aggr.)}, \end{cases} \tag{1333}$$

and reduced significance above zero: 2.7σ, 1.9σ, 1.7σ, and 1.9σ, respectively.

In order to check that the preference for $\Delta^2_{\rm rms} > 0$ or $A_{\rm L} > 1$ is not just a parameter-space volume effect upon marginalization over other parameters, we also report in Table 41 the difference of χ^2 between the best fit in extended models and the base adiabatic ΛCDM model. With all data sets, all three extended models lead to an improvement of χ^2 which clearly exceeds the number of extra parameters of the model (1 for ΛCDM+CIP and ΛCDM+$A_{\rm L}$, and 2 for ΛCDM+$A_{\rm L}$+CIP). Although the inclusion of lensing data reduces this improvement of fit, the ΛCDM+CIP model gives a rather impressive $\Delta\chi^2 = -4$ with *Planck* TT,TE,EE+lowE and aggressive lensing data.

Since we observe a moderate preference for a non-zero CIP amplitude, it might be tempting to "solve" the *Planck* lensing anomaly by using CIP. However, this explanation seems quite unlikely, since in our treatment the CIP and adiabatic perturbations should be uncorrelated with each other, whereas CDI and BDI should be fully anticorrelated (and have a few orders of magnitude larger amplitude than the adiabatic modes while keeping the perturbations nearly Gaussian). It is difficult to imagine a physical model that could lead to this situation. For example, some variants of curvaton model would naturally lead to anticorrelated CDI and BDI, but in these models there would be a correlation with the adiabatic mode too [109, 134].

The above-studied compensated BDI-CDI mode falls into a similar category to NVI: it is an interesting theoretical setup, but a compelling early-Universe model for stimulating this mode has still to be discovered.

Nevertheless, the baseline *Planck* `Plik` TT,TE,EE+lowE plus conservative lensing result, $\Delta^2_{\rm rms} = 0.0037^{+0.0016}_{-0.0021}$, is fully compatible with current complementary observations, in particular, the WMAP 95 % CL trispectrum constraint, $\Delta^2_{\rm rms} \lesssim 0.012$ [117], and the upper bound, $\Delta^2_{\rm rms} \lesssim 0.006$, following from the direct measurements of the variation of the baryon fraction in galaxy clusters [138, 117]. It will be interesting to learn what other future CMB anisotropy [1, 357, 93] and complementary measurements, such as observations of the distribution of neutral hydrogen using 21 cm absorption lines [110], BAO [256, 257], or CMB spectral distortion anisotropies [122], will tell us about the possible contribution of CIP to the primordial perturbations.

28.10 Constraints on anisotropic models of inflation

In this section we will test specific physical models for statistical anisotropy in the primordial fluctuations. More phenomenological multipole- or map-space tests are performed in the companion paper, Planck Collaboration VII [236]. Here we update the results of the 2015 release [**(alias?)**] with polarization and new temperature analyses. Incorporating polarization into these tests is particularly important, due to the mild statistical significance of temperature anomalies such as the dipolar asymmetry. Polarization offers the potential to confirm or refute a physical origin for such anomalies via the measurement of independent fluctuation modes. We perform such a new test with k-space dipolar modulation models. In cases such as quadrupolar asymmetry, where no detection has been claimed with temperature, polarization offers the prospect of tightening existing constraints.

Some asymmetry models predict a modification to the *isotropic* power spectra, in addition to a dipolar or quadrupolar asymmetry. In other words, for these models, as well as non-zero *off-diagonal* multipole covariance elements, we expect departures in the *diagonal* elements relative to the standard ΛCDM prediction. Therefore the isotropic spectra can provide independent tests of such models even using temperature data alone [64]. The curvaton dipole modulation model we examine in Sect. 28.10.1 exhibits this property, and can be constrained via its predictions for isotropic isocurvature power. Similarly, some versions of the quadrupolar modulation model we study in Sect. 28.10.2 modify the isotropic spectra via a monopole term. In both cases these isotropic constraints will be important in narrowing the viable parameter space.

28.10.1 Dipolar asymmetry

A dipolar temperature power asymmetry has long been observed at the largest scales in the CMB [81], although its statistical significance is not high and is subject to a posteriori (look-elsewhere) corrections [32, 220, 226]. Nevertheless, its large-scale character suggests potential links with inflationary physics and various models have been proposed to explain it. In this subsection we examine several physical models for a dipolar modulation. Some models where a generic CDM density isocurvature (CDI) or tensor component is dipole modulated have already been ruled out due to their isotropic predictions [64], so we do not consider these further here.

Curvaton model

First we update our 2015 study [**(alias?)**] of a specific inflationary model for the dipolar asymmetry: namely, the modulated curvaton model of Erickcek et al. [80]. In that study we showed that that model could not explain the observed asymmetry. Here, we generalize the curvaton model to allow for a non-scale-invariant uncorrelated CDI component. In addition, we treat the power spectrum (isotropic) constraints in a fully unified way with the asymmetry likelihood. Finally, we incorporate polarization.

The modulated curvaton model employs a gradient in a background curvaton field to explain the observed large-scale power asymmetry. The curvaton, via coupling κ, produces nearly scale-invariant CDI fluctuations, as well as a fraction, ξ, of the adiabatic fluctuations. Both of these components will be modulated. Up to a sign, ξ is equal to the correlation parameter, and is also

a measure of the amplitude of dipolar modulation. The isocurvature fraction can be written in terms of these two parameters as

$$\beta_{\rm iso} = \frac{9\kappa^2\xi}{1+9\kappa^2\xi}. \tag{1334}$$

Full details of this model and our treatment of it can be found in Erickcek et al. [80] and **(alias?)**.

Using the dipolar asymmetry estimator from **(alias?)** we find the posteriors for the dipolar modulation parameters κ and ξ; the results are presented in Fig. 239 (red contours). We see that a substantial amount of asymmetry (as measured by amplitude ξ) can be captured by the model. This preference for asymmetry simply means that the curvaton model can explain the well-known dipolar asymmetry in temperature. However, isocurvature constraints from the power spectra via Eq. (1334), which we refer to as the *isotropic* constraints, can provide independent information [64]. This is also shown in Fig. 239, with the blue contours. Here we see that the asymmetry and isotropic posteriors only weakly overlap, and the independent isotropic data do not support the presence of asymmetry for this model. No evidence for asymmetry (i.e., no preference for $\xi > 0$) is present in the joint constraints, which treat the isotropic and asymmetry data as independent. In other words, we have no reason to prefer this model over base ΛCDM.

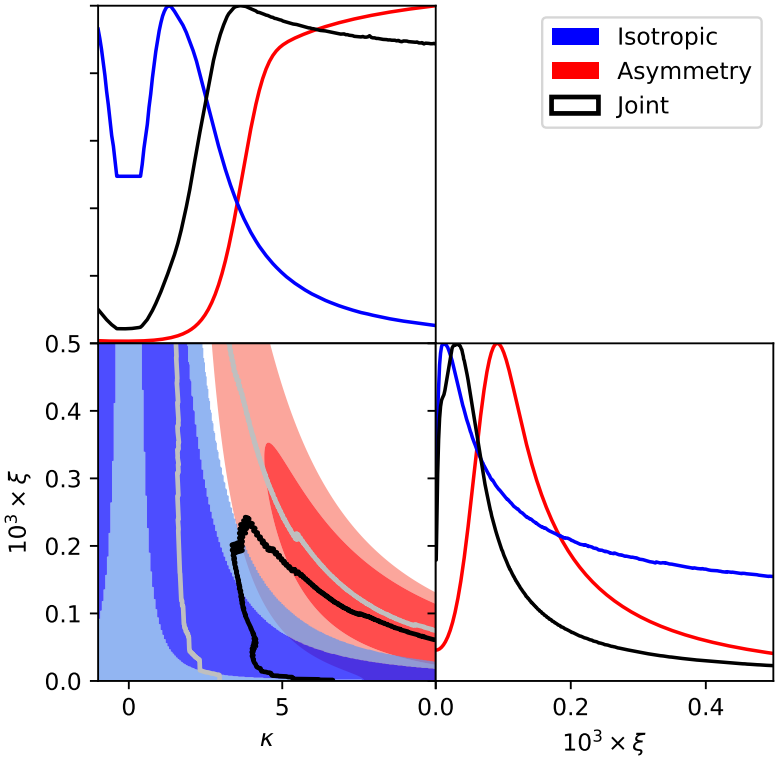


Figure 239: Posteriors for the curvaton dipolar modulation model parameters κ and ξ. Contours enclose 68 % and 95 % of the posteriors. The model can explain the well-known dipolar asymmetry: note the preference for $\xi > 0$ in the asymmetry constraint (red contours and curves). However, the modulation preferred by the asymmetry constraint is reduced substantially when the isotropic constraint (blue) is added (black). The asymmetry constraint here uses `SMICA`, while the isotropic constraint uses *Planck* TT,TE,EE+lowE+lensing. Resolution is reduced at very small κ due to the sampling in $\beta_{\rm iso}$.

Adiabatic models

In the presence of a sufficiently large bispectrum it is possible that a long-wavelength mode can induce a dipolar asymmetry in the two-point function across our observable volume, although such scenarios appear to require fine tuning [51]. Nevertheless, examples have been constructed which satisfy the *Planck* $f_{\rm NL}$ constraints [50]. In this subsection we consider adiabatic

models of this type, in which the isotropic power spectra agree with standard ΛCDM, while a scale-dependent dipolar asymmetry is present in the off-diagonal multipole covariance [65, 64]. As proposed in Contreras et al. [65], we fit the asymmetry model parameters to the temperature data and then use those parameters to predict the asymmetry in polarization. We then compare those predictions with the *Planck* polarization data as a test for a physical modulation. Importantly, a position- (or k-) space model for the modulation is needed for reliable polarization predictions—it is not enough to restrict considerations to multipole space [65].[180]

As discussed in detail in Contreras et al. [65], we take a portion $\widetilde{\mathcal{R}}^{\rm lo}(x)$ of the adiabatic primordial fluctuations to be spatially linearly modulated according to

$$\widetilde{\mathcal{R}}^{\rm lo}(\vec{x}) = \mathcal{R}^{\rm lo}(\vec{x})\left(1 + A\frac{\vec{x}\cdot\hat{\vec{d}}}{r_{\rm LS}}\right), \tag{1335}$$

where $\mathcal{R}^{\rm lo}(\vec{x})$ is statistically isotropic with power spectrum $\mathcal{P}^{\rm lo}_{\mathcal{R}}(k)$, $A \leq 1$ and $\hat{\vec{d}}$ are the amplitude and direction of modulation, respectively, and $r_{\rm LS}$ is the comoving radius to last scattering. This leads, to a good approximation, to the total temperature or polarization multipole covariance

$$C_{\ell m\ell' m'} \equiv \langle a_{\ell m} a^*_{\ell' m'}\rangle \tag{1336}$$

$$= C_\ell \delta_{\ell\ell'}\delta_{mm'} + \frac{\delta C_{\ell\ell'}}{2}\sum_M \Delta X_M \xi^M_{\ell m\ell' m'}, \tag{1337}$$

to first order in A. Here C_ℓ is the usual ΛCDM anisotropy power spectrum; $\delta C_{\ell\ell'} \equiv 2(C^{\rm lo}_\ell + C^{\rm lo}_{\ell'})$, where $C^{\rm lo}_\ell$ is the power spectrum calculated in the usual way from $\mathcal{P}^{\rm lo}_{\mathcal{R}}(k)$; ΔX_M is the multipole decomposition of $A\hat{\vec{n}}\cdot\hat{\vec{d}}$; and the $\xi^M_{\ell m\ell' m'}$ coefficients couple ℓ to $\ell \pm 1$ via

$$\xi^M_{\ell m\ell' m'} \equiv \sqrt{\frac{4\pi}{3}}\int Y_{\ell' m'}(\hat{\vec{n}}) Y_{1M}(\hat{\vec{n}}) Y^*_{\ell m}(\hat{\vec{n}}) d\Omega. \tag{1338}$$

In principle the scale dependence of the asymmetry spectrum $\mathcal{P}^{\rm lo}_{\mathcal{R}}(k)$ is completely free, but here we take three phenomenological forms which are capable of producing a large-scale asymmetry with a small number of parameters. First, we consider a simple power-law modulation,

$$\mathcal{P}^{\rm lo}_{\mathcal{R}}(k) = \mathcal{P}^0_{\mathcal{R}}\left(k^{\rm lo}_0\right)\left(\frac{k}{k^{\rm lo}_0}\right)^{n^{\rm lo}_{\rm s}-1}, \tag{1339}$$

where $\mathcal{P}^0_{\mathcal{R}}(k)$ is the usual ΛCDM spectrum, and $n^{\rm lo}_{\rm s}$ and $k^{\rm lo}_0$ are the tilt and pivot scale of the modulation. We consider only red asymmetry tilts with $n^{\rm lo}_{\rm s} \leq n_{\rm s}$, and choose $k^{\rm lo}_0 = 1.5 \times 10^{-4}\,{\rm Mpc}^{-1}$. We also consider a tanh model, defined according to

$$\mathcal{P}^{\rm lo}_{\mathcal{R}}(k) = \frac{1}{2}\mathcal{P}^0_{\mathcal{R}}(k)\left[1 - \tanh\left(\frac{\ln k - \ln k_{\rm c}}{\Delta\ln k}\right)\right]. \tag{1340}$$

This spectrum approaches that of ΛCDM on scales larger than $k_{\rm c}$, with a width determined by $\Delta\ln k$. That is, scales well above the cutoff $k_{\rm c}$ will be modulated with amplitude A, and scales below will be unmodulated. Finally, we consider a model with a linear gradient in the scalar tilt, $n_{\rm s}$, across our volume. In this case the asymmetry spectrum can be written as

$$C^{\rm lo}_\ell = -\frac{\Delta n_{\rm s}}{2}\frac{dC_\ell}{dn_{\rm s}}, \tag{1341}$$

with modulation amplitude $\Delta n_{\rm s}$. There will be an implicit dependence on the pivot scale k_* for this model.

[180] E.g., if we observe a dipolar modulation in T at, say, 5 % to $\ell = 65$, there is no reason to expect a modulation of the same amplitude and to the same scale in E, due to the different T and E transfer functions [65].

Given the multipole covariance, Eq. (1337), we can construct a maximum likelihood estimator for the modulation, ΔX_M. In the noise-free, full-sky case this takes the form [202, 226]

$$\Delta\hat{X}_M = \frac{1}{4}\sigma_X^2 \sum_{\ell m \ell' m'} \frac{\delta C_{\ell\ell'}}{C_\ell C_{\ell'}} \xi^M_{\ell m \ell' m'} a^*_{\ell m} a_{\ell' m'}, \tag{1342}$$

where the cosmic variance of the estimator is given by

$$\sigma_X^2 = 12 \left(\sum_\ell (\ell+1) \frac{\delta C^2_{\ell\ell+1}}{C_\ell C_{\ell+1}} \right)^{-1}. \tag{1343}$$

The modifications we use to deal with realistic skies are described in detail in Planck Collaboration XVI [226] and Contreras et al. [65].

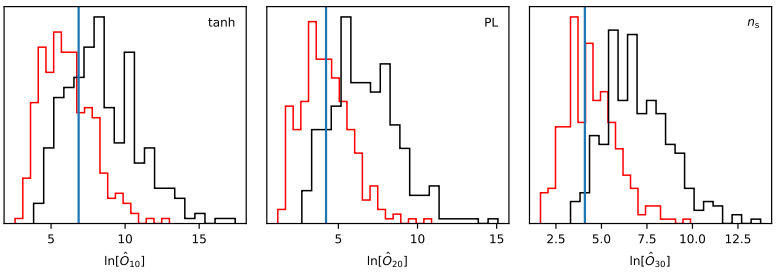


Figure 240: Histograms of the quantity $\hat{O}_{j0}$ for the tanh, power-law, and n_s gradient modulation models using *Planck* temperature data combined with 300 statistically isotropic polarization simulations (red outlines) or 300 polarization simulations modulated according to the best-fit parameters from the temperature data (black). The blue lines indicate the values for the actual SMICA polarization data. A large value relative to the isotropic (red) simulations would indicate that the modulation model is preferred over ΛCDM.

To decide whether the polarization data support the modulation model or not, we consider the quantity $\hat{O}_{j0}$, which is the ratio of the maximum likelihood for modulation model j to that of ΛCDM [65]. In Fig. 240 we plot for the three adiabatic models histograms of $\hat{O}_{j0}$ calculated for 300 statistically isotropic polarization simulations (sharing the required TE correlation with the real T data) added to the *Planck* temperature data (red outlines). This indicates our expectation for $\hat{O}_{j0}$ for the scenario that the temperature asymmetry is due to a statistical fluctuation and not to a physical modulation. We also plot in Fig. 240 histograms for 300 polarization simulations modulated with the best-fit parameters from the *Planck* temperature data (black outlines), to represent the scenario that the asymmetry is due to a physical modulation. In both cases the polarization simulations contain realistic levels of noise for *Planck*. By comparing the isotropic and modulated histograms, we can see that the quantity $\hat{O}_{j0}$ can serve to distinguish the two scenarios, but only relatively weakly for *Planck* noise [65]. The blue lines indicate the values using the actual SMICA polarization data (the results for the other component-separation methods are similar). We see that for these models the data do not help to decide whether we have a physical modulation or not, with p-values of 43 %, 30 %, and 57 % for the power-law, tanh, and n_s gradient models, respectively, relative to the isotropic simulations.

28.10.2 Quadrupolar asymmetry

We will next explore models that predict a quadrupolar direction dependence in the primordial power spectrum. In **(alias?)** we found no evidence for such a modulation, but several inflationary models have been constructed which predict this effect [6, 255, 271]. Therefore it is important to extend those results with the improved polarization data. We now attempt to reduce the effect of unresolved point sources using the bias-hardened estimator approach of

Planck Collaboration XV [286]. In **(alias?)** we pointed out that some models of quadrupolar asymmetry predict a modification to the angular power spectra as well. Here we will account for such modifications in our analysis, increasing the constraining power of temperature data, in particular for tilted models with non-scale-invariant modulation spectral index. Note that independent searches relaxing our approximation of power-law spectra have also been carried out [68].

We assume a modulation of the primordial comoving curvature power spectrum of the form

$$\mathcal{P}_{\mathcal{R}}(\vec{k}) = \mathcal{P}^0_{\mathcal{R}}(k)\left[1 + g(k)\left(\hat{\vec{k}}\cdot\hat{\vec{d}}\right)^2\right], \tag{1344}$$

which can be rewritten as

$$\mathcal{P}_{\mathcal{R}}(\vec{k}) = \mathcal{P}^0_{\mathcal{R}}(k)\left[1 + \frac{1}{3}g(k) + \sum_m g_{2m}(k)\,Y_{2m}(\hat{\vec{k}})\right]. \tag{1345}$$

Here

$$g_{2m}(k) \equiv \frac{8\pi}{15}g(k)\,Y^*_{2m}(\hat{\vec{d}}), \tag{1346}$$

with $g_{2m}(k)$ satisfying $g_{2,-m}(k) = (-1)^m\, g^*_{2m}(k)$. We parameterize the scale dependence of the modulation as $g(k) = g_*(k/k_*)^q$, with pivot scale $k_* = 0.05\,\mathrm{Mpc}^{-1}$. For $q \neq 0$, in addition to producing a quadrupolar modulation of the anisotropies, this model affects the CMB isotropic power spectra via the term $g(k)/3$ in Eq. (1345). We therefore consider a joint constraint with the isotropic power spectra likelihood to improve constraints over the modulation alone.

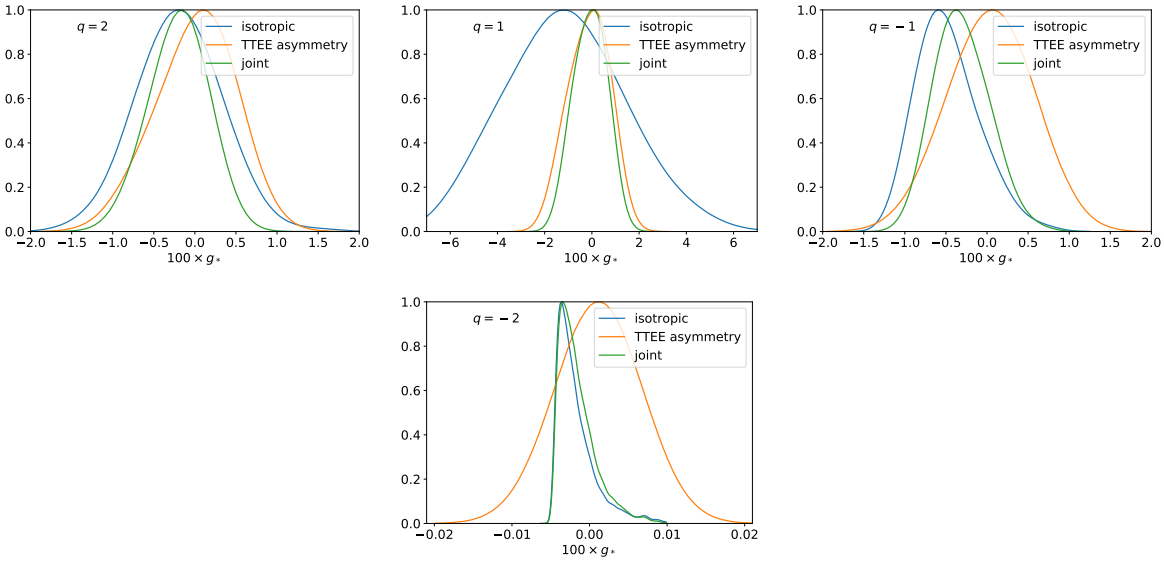


Figure 241: Marginalized posteriors for quadrupolar modulation parameter g_*, using `SMICA` data for the TT+EE asymmetry constraints (orange curves) and *Planck* TT,TE,EE+lowE for the isotropic constraints (blue curves), which probe the modification to the power spectrum via Eq. (1345). *Top:* constraints for $q = 2$ and 1 (left and right, respectively). *Bottom:* constraints for $q = -1$ and -2 (left and right, respectively). Strongly negative g_* is suppressed for $q < 0$, due to the unphysical prediction of negative power.

As in **(alias?)**, we obtain constraints on the modulation parameters by forming quadratic maximum-likelihood estimates, $\hat{g}_{2m}$, for the data and simulations. For this we use the component-separated data and 300 simulations provided by `NILC`, `SEVEM`, `SMICA`, and `COMMANDER`. For brevity we only show the `SMICA` results. We can then compute a covariance G and likelihood as

$$\mathcal{L} \propto |G|^{-1/2}\exp\left[-\frac{1}{2}M^{\mathsf{T}}G^{-1}M\right],$$

Table 42: Minimum-χ^2 g_* values for quadrupolar modulation, determined from the SMICA foreground-cleaned maps. Also given are p values, defined as the fraction of isotropic simulations with larger $|g_*|$ than the data. The TT results use $\ell_{\rm min} = 2$ and $\ell_{\rm max} = 1200$, while EE uses $\ell_{\rm min} = 2$ and $\ell_{\rm max} = 850$. These results indicate that the data are consistent with cosmic variance in statistically isotropic skies.

	TT		EE		$TT+EE$	
q	g_*	p value [%]	g_*	p value [%]	g_*	p value [%]
-2	-6.83×10^{-5}	75.7	1.23×10^{-4}	54.7	-6.90×10^{-5}	75.0
-1	-8.56×10^{-3}	64.7	1.44×10^{-2}	30.0	-6.15×10^{-3}	86.0
0	1.08×10^{-2}	82.7	3.17×10^{-2}	55.3	1.07×10^{-2}	83.0
1	7.77×10^{-3}	82.7	5.09×10^{-2}	24.0	7.75×10^{-3}	82.3
2	4.92×10^{-3}	78.3	5.62×10^{-2}	17.0	4.92×10^{-3}	78.7

Table 43: As for Table 42, but for the quantity $g_2 \equiv \sqrt{\sum_m |g_{2m}|^2/5}$ for a completely general quadrupolar modulation.

	TT		EE		$TT+EE$	
q	g_2	p value [%]	g_2	p value [%]	g_2	p value [%]
-2	3.30×10^{-5}	82.3	9.00×10^{-5}	27.0	3.32×10^{-5}	81.0
-1	4.34×10^{-3}	66.0	6.81×10^{-3}	40.0	3.24×10^{-3}	87.0
0	7.65×10^{-3}	51.7	1.79×10^{-2}	45.7	7.62×10^{-3}	51.7
1	5.39×10^{-3}	58.0	2.95×10^{-2}	12.7	5.38×10^{-3}	58.3
2	3.15×10^{-3}	58.7	3.39×10^{-2}	16.0	3.15×10^{-3}	58.7

where

$$M \equiv \hat{g}_{2m} - g_{2m}(g_*, \vec{\hat{d}}). \tag{1347}$$

We then evaluate the marginalized (over the angles) posterior for g_*. For the isotropic constraints we simply include the modulation parameters in a CosmoMC run using *Planck* TT,TE,EE+lowE data, and then evaluate the marginalized (over all other ΛCDM parameters) posterior for g_*.

For $q > 0$ the effect of g_* on the isotropic spectra occurs mainly at high ℓ, and is highly degenerate with $n_{\rm s}$. This degeneracy leads to slightly less stringent constraints than what one would achieve with a fixed $n_{\rm s}$. We show the marginalized posteriors for this case in the top panels of Fig. 241, where we see that the isotropic constraints are roughly comparable in strength to (and fully consistent with) the constraints from the asymmetry data.

For $q < 0$ the isotropic constraints are much more constraining than the modulation constraints, as seen in the bottom panels of Fig. 241. This is because for large scales the factor k_*/k can become large and a negative g_* will decrease isotropic power on those scales, which is compensated for by increasing $A_{\rm s}$ and τ. Strongly negative g_* values are disallowed by predicting unphysical negative power spectra at low ℓ. Note that even the parameter ranges in which the power spectra are reduced to close to zero are likely beyond the perturbative regime for the models in question, and so should be approached with caution. The isotropic constraints still prefer a slightly negative g_*, likely due to being able to fit the power deficit at large scales. The joint constraint in this case is then greatly improved by the isotropic data.

Minimum-χ^2 and p values (relative to isotropic simulations) for g_* are presented in Table 42. The addition of polarization does not affect the temperature results greatly.

Finally, when allowing the completely general form of quadrupolar modulation, i.e.,

$$\mathcal{P}_{\mathcal{R}}(\vec{k}) = \mathcal{P}^0_{\mathcal{R}}(k)\left[1 + \sum_m g_{2m}\,(k/k_*)^q\; Y_{2m}(\vec{\hat{k}})\right], \tag{1348}$$

with no restriction on the g_{2m}, we present results for the quantity $g_2 \equiv \sqrt{\sum_m |g_{2m}|^2/5}$ in Table 43. In all cases there is no significant detection of quadrupolar modulation, as quantified by the p values.

28.11 Conclusions

This paper summarizes the status of cosmic inflation in light of the *Planck* 2018 release. The main improvements are in the *Planck* polarization likelihoods. The 2018 release now includes a low-ℓ HFI polarization likelihood based on the 100- and 143-GHz channels. This likelihood is now the baseline, whereas the *Planck* 2015 likelihood was based only on the LFI 70-GHz channel data, which also have been updated in this release. Corrections for beam-leakage effects, which had been flagged in the 2015 release as the main limitation of the TE and EE data at that time, have improved the accuracy of the high-ℓ polarization likelihoods. Our analyses focus on the results obtained using the *Planck* baseline likelihoods alone, but results supplemented by the BK15 likelihood (when tensors are included) and a compilation of BAO likelihoods are also given in order to help break cosmological parameter degeneracies. We summarize the main results of this paper in the form of responses to a number of key questions.

1. *What is the value of the scalar tilt?*

   Using a characterization of polarization anisotropy better at all multipoles in this release, we find that $n_{\rm s} = 0.9649 \pm 0.0042$ at 68 % CL, including the full information provided by *Planck* (TT,TE,EE+lowE+lensing). The 2018 uncertainty is approximately 2/3 of that obtained with the *Planck* 2015 baseline likelihood. Importantly, this determination rules out perfect scale invariance (i.e., $n_{\rm s} = 1$) at 8.4σ. From an inflationary perspective, this result is consistent with slow-roll inflation evolving towards a natural exit.

2. *Does $n_{\rm s}$ depend on the wavelength?*

   We investigated the possibility of a running spectral index, as well as a running of the running [i.e., the next two (subleading) terms in a power series expansion of $\ln(\mathcal{P}_{\mathcal{R}})$ in $\ln(k)$], corresponding to non-negligible third- and fourth-order derivatives of the inflationary potential. Starting with its first 2013 cosmological release, *Planck* has removed any hint of a running spectral index, which had been suggested by pre-*Planck* data and would have pointed to inflationary models beyond the slow-roll approximation. *Planck* 2018 sets $dn_{\rm s}/d\ln k = -0.005 \pm 0.013$ as the tightest 95 % CL constraint, when $d^2n_{\rm s}/d\ln k^2 = 0$. No hints of further extensions, such as running of the running, are found with *Planck* 2018 data. These results are consistent with the simplest slow-roll dynamics for the inflaton. A detection of running at the level predicted by slow-roll models will require a combination of future ambitious CMB anisotropy experiments and galaxy surveys.

3. *Is the Universe spatially flat?*

   Most simple models of inflation predict a spatially flat universe, although inflationary models with a minimum degree of fine tuning producing a hyperbolic universe have been constructed. *Planck* has been the first experiment to constrain the spatial curvature at the percent level without any external information, thanks to the CMB lensing likelihood. Although negative values of $\Omega_K \sim -0.01$ provide a non-statistically significant improvement to the fit of *Planck* temperature and polarization data (compared to the minimal ΛCDM model), *Planck* 2018 data including lensing constrain $\Omega_K = -0.011^{+0.013}_{-0.012}$ at 95 % CL. Combining with BAO data further tightens the uncertainty, constraining Ω_K to lie within 0.4 % of a flat spatial geometry (at 95 % CL).

4. *Are tensor modes required?*

   Inflationary models predict that tensor modes were also excited during the nearly exponential expansion, with a power spectrum amplitude proportional to the energy scale of inflation. Using the measurement of CMB temperature and E-mode polarization anisotropies from the quadrupole into the acoustic peak region, *Planck* has reduced the degeneracy between the tensor-to-scalar ratio r and $n_{\rm s}$, establishing the bound $r_{0.002} < 0.10$ at 95 % CL, assuming $n_{\rm t} = -r/8$ as predicted by the simplest inflationary models. When the *Planck* likelihood is combined with the B-mode polarization likelihood of the BICEP2-Keck Array experiment, a tight 95 % CL upper limit of $r_{0.002} < 0.056$ is obtained, corresponding to a 95 % CL bound on the energy scale of inflation of $V_*^{1/4} < 1.6 \times 10^{16}$ GeV. *Planck* 2018 and BK15 data also set tight bounds on gravitational waves generated in the early Universe when r and $n_{\rm t}$ are varied independently, complementary to the results obtained by the direct-detection interferometers LIGO and VIRGO at much higher frequencies.

5. *Which inflationary models are best able to account for the data?*

   Starting with the 2013 release using only a part of the data, *Planck* has substantially tightened the constraints on slow-roll inflationary models, ruling out hybrid models with $n_s > 1$ and power-law inflation [**(alias?)**]. In combination with the BK15 data, *Planck* 2018 now strongly disfavours monomial models with $V(\phi) \propto \phi^p$ and $p > 1$, natural inflation, and low-scale SUSY models. Within the representative cases studied in this paper, inflationary models such as R^2, T and E α-attractor models, D-brane inflation, and those having a potential with exponential tails provide good fits to *Planck* and BK15 data. We used two methods to reconstruct the inflaton potential beyond the slow-roll approximation: by Taylor expanding the inflaton potential or Hubble parameter in the observable region; and through a free-form reconstruction of the potential with cubic splines. No statistically significant detection beyond the second derivative of the potential was found, suggesting that the slow-roll approximation is adequate for the *Planck* 2018 likelihood in combination with the BK15 data.

6. *What model-independent constraints can be placed on the primordial power spectrum?*

   We reported on three different methods for the non-parametric reconstruction of the primordial power spectrum (penalized likelihood, a Bayesian spline reconstruction. and a method based on cubic splines). All three methods give broadly consistent results. In no case is any statistically significant evidence for a deviation from a pure power law found. The constraints on the deviations are at the few-percent level for wavenumbers in the range $0.005\,\mathrm{Mpc}^{-1} \lesssim k \lesssim 0.2\,\mathrm{Mpc}^{-1}$ probed by the CMB, the precise constraint depending on the level of smoothing allowed.

7. *Is there evidence for features in the primordial power spectrum?*

   We explored several classes of theoretically motivated parametric models with strong departures from a power law for the primordial power spectra and tested their predictions using combinations of *Planck* temperature and polarization power spectra. We also carried out an analysis using bispectrum data as well. No statistically significant evidence for features was found.

8. *Were the primordial cosmological perturbations solely adiabatic?*

   A key question is whether the primordial cosmological fluctuations consisted exclusively of adiabatic growing-mode perturbations or whether isocurvature perturbations, possibly correlated with the adiabatic mode and with each other, were also excited. The new polarization data has helped to sharpen constraints on the allowed isocurvature fraction compared to the *Planck* 2015 results. In correlated mixed adiabatic and isocurvature models, the 95 % CL upper bound for the non-adiabatic contribution to the observed CMB temperature variance is $|\alpha_{\text{non-adi}}| < 1.3\,\%$, $1.7\,\%$, and $1.7\,\%$ for CDM, neutrino density, and neutrino velocity isocurvature, respectively. For this release we also report constraints on a scale-invariant compensated baryon-CDM isocurvature mode, which is uncorrelated with the adiabatic mode. This mode would cause an additional lensing-like smoothing at high ℓ and modify the lensing potential at $\ell \lesssim 40$. By using the temperature, polarization, and lensing data, we obtain the constraint $\Delta^2_{\text{rms}} = 0.0037^{+0.0016}_{-0.0021}$ at 68 % CL for the variance of the baryon isocurvature density perturbation. A detection of isocurvature modes would suggest the need for a theory beyond single-field inflation, which is able to excite only one mode.

9. *Were the primordial fluctuations statistically isotropic?*

   The *Planck* analysis has confirmed evidence at low statistical significance of anomalies in the CMB temperature anisotropies on large angular scales that are not alleviated in models with nontrivial topology or an anisotropic expansion [228]. This motivates an exploration of inflation-based models giving such violation of statistical isotropy. We have found no statistically significant evidence in favour of a curvaton model for dipolar asymmetry (compared to the base-ΛCDM model), nor any evidence for a quadrupolar asymmetry in the temperature or polarization anistropies. Theoretical models producing the observed temperature dipolar asymmetry make a prediction for the polarization dipolar asymmetry. We tested whether the fit to the temperature dipolar asymmetry gives a prediction for the polarization asymmetry consistent with the data. We found

no statistically significant evidence that the pattern seen in temperature is repeated in polarization. However, the discriminating power of this test is weak, due to the low polarization signal-to-noise ratio on large angular scales.

The *Planck* 2013, 2015, and 2018 releases have substantially improved the constraints on the space of inflationary models, as described above. Future CMB polarization data will be crucial for further constraining those inflationary models that currently provide an adequate fit to *Planck* and other data. Forthcoming E-mode polarization data will be decisive for determining whether the intriguing features in the temperature power spectrum, such as the deficit at $\ell \simeq 20$–30, the smaller average amplitude at $\ell \lesssim 40$, and other anomalies at higher multipoles require new physics or whether these features are simply the result of statistical fluctuations plus instrumental noise. Improved measurements of the B modes promise to constrain inflation even more tightly and it will be interesting to see how the search for B modes evolves. One possibility would be a convincing detection of inflationary gravitational waves, but a tighter upper limit of $r \lesssim 10^{-3}$ is also an achievable outcome. Either case would substantially advance our understanding of inflation and the constraints on the physics of the very early Universe.

Part VIII
Consequences of High Energy Physics on Cosmology

29 Inflation in String Theory by Baumann's 5th lecture

Abstract

We end this lecture series with a discussion of a slightly more advanced topic: inflation in string theory. We provide a pedagogical overview of the subject based on a recent review article with Liam McAllister [16]. The central theme of the lecture is the sensitivity of inflation to Planck-scale physics, which we argue provides both the primary motivation and the central theoretical challenge for realizing inflation in string theory. We illustrate these issues through two case studies of inflationary scenarios in string theory: warped D-brane inflation and axion monodromy inflation. Finally, we indicate opportunities for future progress both theoretically and observationally.

29.1 Why Combine Two Speculative Ideas?

In the previous lectures we have seen that inflation is remarkably successful as a phenomenological model for the dynamics of the very early universe. However, a detailed understanding of the physical origin of the inflationary expansion has remained elusive. Inflation and string theory are both ambitious attempts to understand the physical universe at the highest energies. Both inflation and string theory are speculative theories that still await experimental confirmation. One may therefore wonder why it is timely to address the problem of inflation in string theory.

In this lecture we will highlight specific aspects of inflation that depend sensitively on the ultraviolet (UV) completion of quantum field theory and gravity, *i.e.* on the field content and interactions at energies approaching the Planck scale. Such issues are most naturally addressed in a theory of Planck-scale physics, for which string theory is arguably the best-developed candidate. This motivates understanding the physics of inflation in string theory.

Readers less interested in the details of the string theory constructions in §29.3 might still find the generic effective field theory arguments in §29.2 informative.

29.2 UV Sensitivity of Inflation

29.2.1 Effective Field Theory and Inflation

As a phenomenon in Quantum Field Theory coupled to General Relativity, inflation does not appear to be natural. In particular, the set of Lagrangians suitable for inflation is a minute subset of the set of all possible Lagrangians. Moreover, in wide classes of models, inflation emerges only for rather special initial conditions, *e.g.* initial configurations with tiny kinetic energy, in the case of small-field scenarios. Although one would hope to explore and quantify the naturalness both of inflationary Lagrangians and of inflationary initial conditions, the question of initial conditions appears inextricable from the active yet incomplete program of understanding measures in eternal inflation (see §7.7.2 for a critical evaluation). In this lecture we will focus on the question of how (un)natural it is to have a Lagrangian suitable for inflation.

For a single inflaton field with a canonical kinetic term, the necessary conditions for inflation can be stated in terms of the inflaton potential (see **Lecture 1**). Inflation requires a potential that is quite flat in Planck units

$$\epsilon_{\rm v} = M_{\rm pl}^2 \left(\frac{V_{,\phi}}{V}\right)^2 \ll 1\,, \qquad \eta_{\rm v} = \frac{M_{\rm pl}^2}{2}\frac{V_{,\phi\phi}}{V} \ll 1\,. \tag{1349}$$

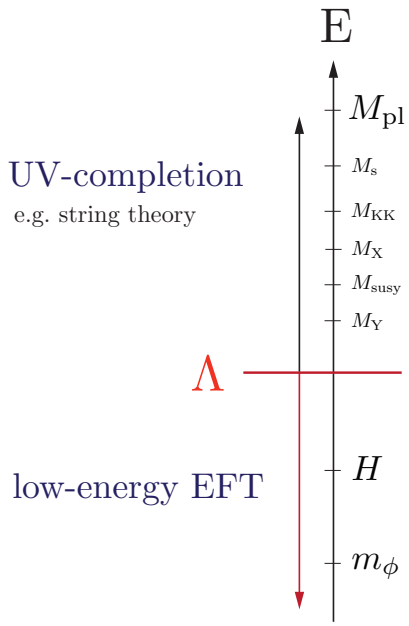


Figure 242: The Effective Field Theory (EFT) of Inflation. The cut-off Λ of the EFT is defined by the mass of the lightest particle that is not included in the spectrum of the low-energy theory. Particles with masses above the cut-off are integrated out, correcting the Lagrangian for the light fields such as the inflaton.

As we now argue, this condition is sensitive to Planck-scale physics.

Let us recall that the presence of some form of new physics at the Planck scale is required in order to render graviton-graviton scattering sensible, just as unitarity of W-W scattering requires new physics at the TeV scale. Although we know that new degrees of freedom must emerge, we cannot say whether the physics of the Planck scale is a finite theory of quantum gravity, such as string theory, or is instead simply an effective theory for some unimagined physics at yet higher scales. However, the structure of the Planck-scale theory has meaningful – and, in very favorable cases, testable – consequences for the form of the inflaton potential.

As usual, the effects of high-scale physics above some cutoff Λ are efficiently described by the coefficients of operators in the low-energy effective theory (see Fig. 242). Integrating out particles of mass $M \geq \Lambda$ gives rise to operators of the form

$$\boxed{\frac{\mathcal{O}_\delta}{M^{\delta-4}}}\,, \tag{1350}$$

where δ denotes the mass dimension of the operator.

Sensitivity to such operators is commonplace in particle physics: for example, bounds on flavor-changing processes place limits on physics above the TeV scale, and lower bounds on the proton lifetime even allow us to constrain GUT-scale operators that would mediate proton decay. However, particle physics considerations alone do not often reach beyond operators of dimension $\delta = 6$, nor go beyond $M \sim M_{\rm GUT}$. (Scenarios of gravity-mediated supersymmetry breaking are one exception.) Equivalently, Planck-scale processes, and operators of very high dimension, are irrelevant for most of particle physics: they decouple from low-energy phenomena.

In inflation, however, the flatness of the potential in Planck units introduces sensitivity to $\delta \leq 6$ *Planck-suppressed* operators, such as

$$\frac{\mathcal{O}_6}{M_{\rm pl}^2}\,. \tag{1351}$$

An understanding of such operators is required to address the smallness of the eta parameter, *i.e.* to ensure that the theory supports at least 60 e-folds of inflationary expansion. This sensitivity to dimension-six Planck-suppressed operators is therefore common to all models of inflation.

For large-field models of inflation the UV sensitivity of the inflaton action is dramatically enhanced. In this important class of inflationary models the potential becomes sensitive to an *infinite* series of operators of arbitrary dimension (see §29.2.3).

29.2.2 The Eta Problem

In the absence of any specific symmetries protecting the inflaton potential, contributions to the Lagrangian of the general form

$$\boxed{\frac{\mathcal{O}_6}{M_{\rm pl}^2} = \frac{\mathcal{O}_4}{M_{\rm pl}^2}\,\phi^2} \tag{1352}$$

are allowed. If the dimension-four operator $\mathcal{O}_4$ has a vacuum expectation value (vev) comparable to the inflationary energy density, $\langle \mathcal{O}_4 \rangle \sim V$, then this term corrects the inflaton mass by order H, or equivalently corrects the eta parameter by order one, leading to an important problem for inflationary model-building. Let us reiterate that contributions of this form may be thought of as arising from integrating out Planck-scale degrees of freedom. In this section we discuss this so-called *eta problem* first in effective field theory, §29.2.2, and then illustrate the problem in a supergravity example, §29.2.2.

Radiative Instability of the Inflaton Mass

In a generic effective theory with cutoff Λ (see Fig. 242), the mass of a scalar field runs to the cutoff scale unless it is protected by some symmetry. Since the cutoff for an effective theory of inflation is at least the Hubble scale, $\Lambda \geq H$, this implies that a small inflaton mass ($m_\phi \ll H$) is radiatively unstable. Equivalently, the eta parameter receives radiative corrections,

$$\Delta\eta_{\rm v} = \frac{\Delta m_\phi^2}{3H^2} \geq 1\,, \tag{1353}$$

preventing prolonged inflation.

The difficulty here is analogous to the Higgs hierarchy problem, but supersymmetry does not suffice to stabilize the inflaton mass: the inflationary energy necessarily breaks supersymmetry, and the resulting splittings in supermultiplets are of order H, so that supersymmetry does not protect a small inflaton mass $m_\phi \ll H$.

In §29.3.3 we discuss the natural proposal to protect the inflaton potential via a shift symmetry $\phi \to \phi + const.$, which is equivalent to identifying the inflaton with a pseudo-Nambu-Goldstone-boson. In the absence of such a symmetry the eta problem seems to imply the necessity of fine-tuning the inflationary action in order to get inflation.

Supergravity Example

An important instance of the eta problem arises in locally-supersymmetric theories, *i.e.* in supergravity [102]. This case is relevant for many string theory models of inflation because four-dimensional supergravity is the low-energy effective theory of supersymmetric string compactifications [103, 104].

In $\mathcal{N} = 1$ supergravity, a key term in the scalar potential is the F-term potential,

$$V_F = e^{K/M_{\rm pl}^2}\left[K^{\varphi\bar{\varphi}} D_\varphi W \overline{D_\varphi W} - \frac{3}{M_{\rm pl}^2}|W|^2\right], \tag{1354}$$

where $K(\varphi, \bar{\varphi})$ and $W(\varphi)$ are the Kähler potential and the superpotential, respectively; φ is a complex scalar field which is taken to be the inflaton; and we have defined $D_\varphi W \equiv \partial_\varphi W + M_{\rm pl}^{-2}(\partial_\varphi K)W$. For simplicity of presentation, we have assumed that there are no other light degrees of freedom, but generalizing our expressions to include other fields is straightforward.

The Kähler potential determines the inflaton kinetic term, $-K_{,\varphi\bar{\varphi}}\,\partial\varphi\partial\bar{\varphi}$, while the superpotential determines the interactions. To derive the inflaton mass, we expand K around

some chosen origin, which we denote by $\varphi \equiv 0$ without loss of generality, *i.e.* $K(\varphi, \bar{\varphi}) = K_0 + K_{,\varphi\bar{\varphi}}|_0\, \varphi\bar{\varphi} + \cdots$. The inflationary Lagrangian then becomes

$$\mathcal{L} \approx -K_{,\varphi\bar{\varphi}}\, \partial\varphi\partial\bar{\varphi} - V_0\Big(1 + K_{,\varphi\bar{\varphi}}|_0\, \frac{\varphi\bar{\varphi}}{M_{\rm pl}^2} + \dots\Big) \tag{1355}$$

$$\equiv -\partial\phi\partial\bar{\phi} - V_0\Big(1 + \frac{\phi\bar{\phi}}{M_{\rm pl}^2}\Big) + \dots\,, \tag{1356}$$

where we have defined the canonical inflaton field $\phi\bar{\phi} \approx K_{\varphi\bar{\varphi}}|_0\, \varphi\bar{\varphi}$ and $V_0 \equiv V_F|_{\varphi=0}$. We have retained the leading correction to the potential originating in the expansion of $e^{K/M_{\rm pl}^2}$ in Eqn. (1354), which could plausibly be called a universal correction in F-term scenarios. The omitted terms, some of which can be of the same order as the terms we keep, arise from expanding

$$\left[K^{\varphi\bar{\varphi}} D_\varphi W \overline{D_\varphi W} - \frac{3}{M_{\rm pl}^2}|W|^2\right] \tag{1357}$$

in Eqn. (1354) and clearly depend on the model-dependent structure of the Kähler potential and the superpotential.

The result is of the form of Eqn. (1351) with

$$\mathcal{O}_6 = V_0\, \phi\bar{\phi} \tag{1358}$$

and implies a large model-independent contribution to the eta parameter

$$\Delta\eta_{\rm v} = 1\,, \tag{1359}$$

as well as a model-dependent contribution which is typically of the same order. It is therefore clear that in an inflationary scenario driven by an F-term potential, eta will generically be of order unity.

Under what circumstances can inflation still occur, in a model based on a supersymmetric Lagrangian? One obvious possibility is that the model-dependent contributions to eta (1357) approximately cancel the model-independent contribution (1356), so that the smallness of the inflaton mass is a result of fine-tuning. In the case study of §29.3.2 we will provide a concrete example in which the structure of all relevant contributions to eta can be computed, so that one can sensibly pursue such a fine-tuning argument.

Clearly, it would be far more satisfying to exhibit a mechanism that *removes* the eta problem by ensuring that all corrections are small, $\Delta\eta_{\rm v} \ll 1$. This requires either that the F-term potential is negligible, or that the inflaton does not appear in the F-term potential. The first case does not often arise, because F-term potentials play an important role in presently-understood models for stabilization of the compact dimensions of string theory [105]. However, in §29.3.3 we will present a scenario in which the inflaton is an axion and does not appear in the Kähler potential, or in the F-term potential, to any order in perturbation theory. This evades the particular incarnation of the eta problem that we have described above.

29.2.3 The Lyth Bound

In **Lecture 2** we derived the Lyth bound [106]:

$$\frac{\Delta\phi}{M_{\rm pl}} \simeq \mathcal{O}(1)\left(\frac{r}{0.01}\right)^{1/2}\,, \tag{1360}$$

where r is the value of the tensor-to-scalar ratio on CMB scales. In any model with $r > 0.01$ one must therefore ensure that $\epsilon_{\rm v}, |\eta_{\rm v}| \ll 1$ over a super-Planckian range $\Delta\phi > M_{\rm pl}$. This result implies two necessary conditions for *large-field inflation*:

i) an obvious requirement is that large field ranges are *kinematically* allowed, *i.e.* that the scalar field space (in canonical units) has diameter $> M_{\rm pl}$. This is nontrivial, as in typical string compactifications many fields are not permitted such large excursions.

ii) the flatness of the inflaton potential needs to be controlled *dynamically* over a super-Planckian field range. We discuss this challenge in effective field theory in §29.2.4 and in string theory in §29.3.3.

29.2.4 Super-Planckian Fields and Flat Potentials

To begin, let us consider super-Planckian field excursions in the context of Wilsonian effective field theory.

No Shift Symmetry

In the absence of any special symmetries, the potential in large-field inflation becomes sensitive to an infinite series of Planck-suppressed operators. The physical interpretation of these terms is as follows: as the inflaton expectation value changes, any other fields χ to which the inflaton couples experience changes in mass, self-coupling, etc. In particular, any field coupled with at least gravitational strength to the inflaton experiences significant changes when the inflaton undergoes a super-Planckian excursion. These variations of the χ masses and couplings in turn feed back into changes of the inflaton potential and therefore threaten to spoil the delicate flatness required for inflation. Note that this applies not just to the light degrees of freedom, but even to fields with masses near the Planck scale: integrating out Planck-scale degrees of freedom generically (*i.e.*, for couplings of order unity) introduces Planck-suppressed operators in the effective action. For nearly all questions in particle physics, such operators are negligible, but in inflation they play an important role.

The particular operators which appear are determined, as always, by the symmetries of the low-energy action. As an example, imposing only the symmetry $\phi \to -\phi$ on the inflaton leads to the following effective action:

$$\boxed{\mathcal{L}_{\rm eff}(\phi) = -\frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{1}{4}\lambda\phi^4 - \sum_{p=1}^{\infty}\left[\lambda_p\phi^4 + \nu_p(\partial\phi)^2\right]\left(\frac{\phi}{M_{\rm pl}}\right)^{2p} + \cdots}\,. \tag{1361}$$

Unless the UV theory enjoys further symmetries, one expects that the coefficients λ_p and ν_p are of order unity. Thus, whenever ϕ traverses a distance of order $M_{\rm pl}$ in a direction that is not protected by a suitably powerful symmetry, the effective Lagrangian receives substantial corrections from an infinite series of higher-dimension operators. In order to have inflation, the potential should of course be approximately flat over a super-Planckian range. If this is to arise by accident or by fine-tuning, it requires a conspiracy among infinitely many coefficients, which has been termed 'functional fine-tuning' (compare this to the eta problem which only requires tuning of one mass parameter).

Shift Symmetry

There is a sensible way to control this infinite series of corrections: one can invoke an approximate symmetry that forbids the inflaton from coupling to other fields in any way that would spoil the structure of the inflaton potential. Such a shift symmetry,

$$\phi \to \phi + const.\,, \tag{1362}$$

protects the inflaton potential in a natural way.

In the case with a shift symmetry, the action of chaotic inflation [171]

$$\mathcal{L}_{\rm eff}(\phi) = -\frac{1}{2}(\partial\phi)^2 - \lambda_p\,\phi^p\,, \tag{1363}$$

with small coefficient λ_p is 'technically natural'. However, because we require that this symmetry protects the inflaton even from couplings to Planck-scale degrees of freedom, it is essential that the symmetry should be approximately respected by the Planck-scale theory – in other words, the proposed symmetry of the low-energy effective action should admit a UV-completion. Hence, large-field inflation should be formulated in a theory that has access to information about approximate symmetries at the Planck scale. Let us remark that in effective field theory in general, UV-completion of an assumed low-energy symmetry is rarely an urgent question. The present situation is different because we do *not* know whether all reasonable effective actions can in fact arise as low-energy limits of string theory, and indeed it has been conjectured that many effective theories do not admit UV-completion in string theory [108, 1389, 110]. Therefore, it

is important to verify that any proposed symmetry of Planck-scale physics can be realized in string theory.

To construct an inflationary model with detectable gravitational waves, we are therefore interested in finding, in string theory, a configuration that has both a large kinematic range, and a potential protected by a shift symmetry that is approximately preserved by the full string theory.

29.3 Inflation in String Theory

29.3.1 From String Compactifications to the Inflaton Action

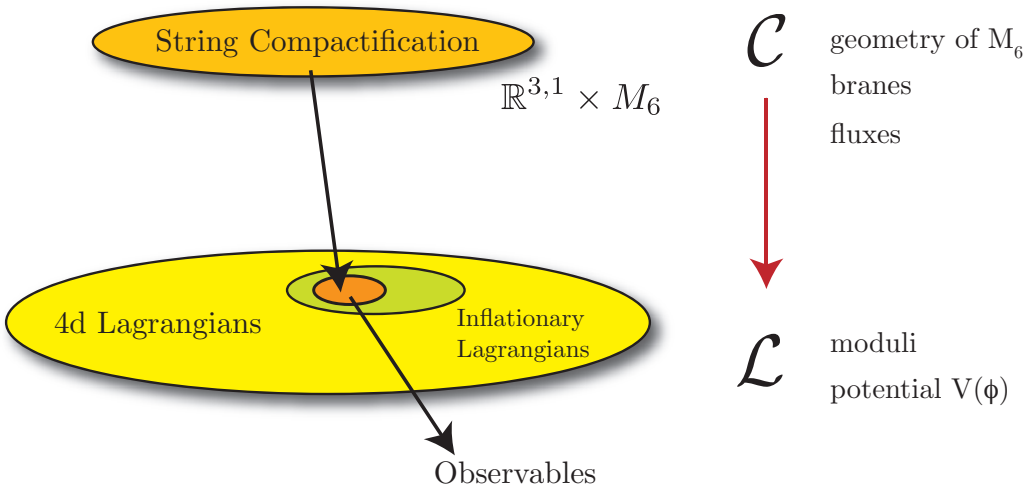


Figure 243: From 10d Compactification Data to 4d Action.

Elements of String Compactifications

It is a famous fact that the quantum theory of strings is naturally defined in more than four spacetime dimensions, with four-dimensional physics emerging upon compactification of the additional spatial dimensions. For concreteness, we will focus on compactifications of the critical ten-dimensional type IIB string theory on six-dimensional Calabi-Yau spaces.[181]

The vast number of distinct compactifications in this class are distinguished by their topology, geometry, and discrete data such as quantized fluxes and wrapped D-branes. A central task in string theory model-building is to understand in detail how the ten-dimensional sources determine the four-dimensional effective theory (see Fig. 243). If we denote the ten-dimensional compactification data by $\mathcal{C}$, the procedure in question may be written schematically as

$$\mathcal{S}_{10}[\mathcal{C}] \quad \rightarrow \quad \mathcal{S}_4\,. \tag{1364}$$

Distinct compactification data $\mathcal{C}$ give rise to a multitude of four-dimensional effective theories $\mathcal{S}_4$ with varied field content, kinetic terms, scalar potentials, and symmetry properties (this is the *landscape* of solutions to string theory). By understanding the space of possible data $\mathcal{C}$ and the nature of the map in Eqn. (1364), we can hope to identify, and perhaps even classify, compactifications that give rise to interesting four-dimensional physics.

The Effective Inflaton Action

For our purposes, the most important degrees of freedom of the effective theory are four-dimensional scalar fields. Scalar fields known as *moduli* arise from deformations of the compactification manifold, typically numbering in the hundreds for the Calabi-Yau spaces under

[181] Readers unfamiliar with this terminology may find a useful Stringlish-to-English dictionary in [111].

Figure 244: Elements of Flux Compactifications: Fluxes and Wrapped Branes.

consideration, and from the positions, orientations, and gauge field configurations of any D-branes. From given compactification data one can compute the kinetic terms and scalar potentials of the moduli; in turn, the expectation values of the moduli determine the parameters of the four-dimensional effective theory. In the presence of generic ten-dimensional sources of stress-energy, such as D-branes and quantized fluxes, there is an energy cost for deforming the compactification, and many (though not always all) of the moduli fields become massive [112].

It is useful to divide the scalar fields arising in $\mathcal{S}_4$ into a set of light fields ϕ, ψ with masses below the Hubble scale ($m_\phi, m_\psi \ll H$) and a set of heavy fields χ with masses much greater than the Hubble scale ($m_\chi \gg H$). Here one of the light fields, denoted ϕ, has been identified as the inflaton candidate.

To understand whether successful inflation can occur, one must understand all the scalar fields, both heavy and light. First, sufficiently massive moduli fields are effectively frozen during inflation, and one should integrate them out to obtain an effective action for the light fields only,

$$\mathcal{S}_4(\phi, \psi, \chi) \quad \rightarrow \quad \mathcal{S}_{4,\text{eff}}(\phi, \psi)\,. \tag{1365}$$

Integrating out these heavy modes generically induces contributions to the potential of the putative inflaton: that is, *moduli stabilization contributes to the eta problem.* This is completely analogous to the appearance of corrections from higher-dimension operators in our discussion of effective field theory in §29.2.1.

Next, if scalar fields in addition to the inflaton are light during inflation, they typically have important effects on the dynamics, and one should study the evolution of all fields ψ with masses $m_\psi \ll H$. Moreover, even if the resulting multi-field inflationary dynamics is suitable, light degrees of freedom can create problems for late-time cosmology. Light scalars absorb energy during inflation and, if they persist after inflation, they can release this energy during or after Big Bang nucleosynthesis, spoiling the successful predictions of the light element abundances. Moreover, light moduli would be problematic in the present universe, as they mediate fifth forces of gravitational strength. To avoid these late-time problems, it suffices to ensure that $m_\psi \gg 30$ TeV, as in this case the moduli decay before Big Bang nucleosynthesis. A simplifying assumption that is occasionally invoked is that all fields aside from the inflaton should have $m \gg H$, but this is not required on physical grounds: it serves only to arrange that the effective theory during inflation has only a single degree of freedom.

29.3.2 *Case Study:* Warped D-brane Inflation

In string theory models of inflation the operators contributing to the inflaton potential can be enumerated, and in principle even their coefficients can be computed in terms of given compactification data. To illustrate these issues, it is useful to examine a concrete model in detail. In the following we therefore present a case study of a comparatively well-understood model of small-field inflation, *warped D-brane inflation.*

D3-branes in Warped Throat Geometries

In this scenario inflation is driven by the motion of a D3-brane in a warped throat region of a stabilized compact space [46]. To preserve four-dimensional Lorentz (or de Sitter) invariance, the D3-brane fills our four-dimensional spacetime and is pointlike in the extra dimensions (see Figure 245). The global compactification is assumed to be a warped product of four-dimensional

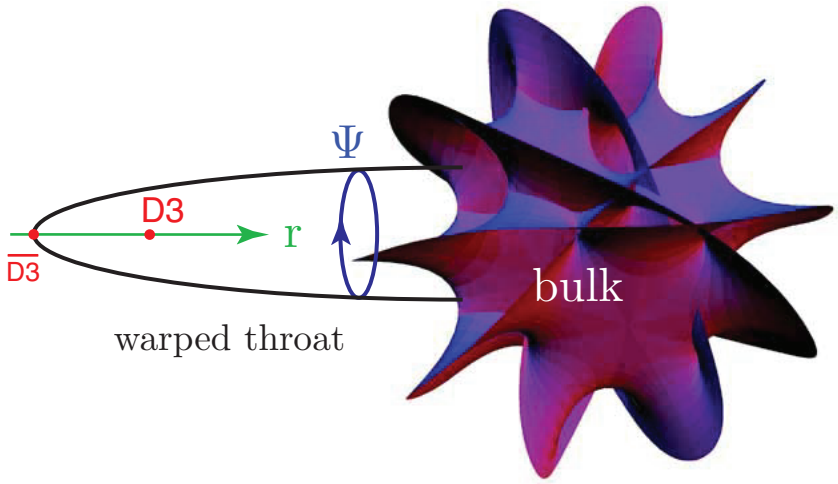


Figure 245: D3-brane inflation in a warped throat geometry. The D3-branes are spacetime-filling in four dimensions and therefore pointlike in the extra dimensions. The circle stands for the base manifold X_5 with angular coordinates Ψ. The brane moves in the radial direction r. At $r_{\rm UV}$ the throat attaches to a compact Calabi-Yau space. Anti-D3-branes minimize their energy at the tip of the throat, $r_{\rm IR}$.

spacetime (with metric $g_{\mu\nu}$) and a conformally-Calabi-Yau space,

$$ds^2 = e^{2A(y)} g_{\mu\nu} dx^\mu dx^\nu + e^{-2A(y)} g_{mn} dy^m dy^n \,, \tag{1366}$$

with g_{mn} a Calabi-Yau metric that can be approximated in some region by a cone over a five-dimensional Einstein manifold X_5,

$$g_{mn} dy^m dy^n = dr^2 + r^2 ds^2_{X_5} \,. \tag{1367}$$

A canonical example of such a throat region is the Klebanov-Strassler (KS) geometry [113], for which X_5 is the $\Big(SU(2) \times SU(2)\Big)/U(1)$ coset space $T^{1,1}$, and the would-be conical singularity at the tip of the throat, $r = 0$, is smoothed by the presence of appropriate fluxes. The tip of the throat is therefore located at a finite radial coordinate $r_{\rm IR}$, while at $r = r_{\rm UV}$ the throat is glued into an unwarped bulk geometry. In the relevant regime $r_{\rm IR} \ll r < r_{\rm UV}$ the warp factor may be written as [114]

$$e^{-4A(r)} = \frac{R^4}{r^4} \ln \frac{r}{r_{\rm IR}} \,, \qquad R^4 \equiv \frac{81}{8} (g_s M \alpha')^2 \,, \tag{1368}$$

where

$$\ln \frac{r_{\rm UV}}{r_{\rm IR}} \approx \frac{2\pi K}{3 g_s M} \,. \tag{1369}$$

Here, M and K are integers specifying the flux background [113, 27].

Warping sourced by fluxes is commonplace in modern compactifications, and there has been much progress in understanding the stabilization of the moduli of such a compactification [112]. Positing a stabilized compactification containing a KS throat therefore seems reasonable given present knowledge.

The Field Range Bound

Before addressing the complicated problem of the shape of the inflationary potential let us ask if these models can ever source a large gravitational wave amplitude. It turns out that this question can be phrased in purely geometrical terms and does not depend on the details of inflationary dynamics [116]. By the Lyth bound we know that a large gravitational wave signal requires super-Planckian field variation. As a minimal requirement we therefore ask if super-Planckian field values are accessible in warped D-brane inflation.

The inflaton kinetic term is determined by the Dirac-Born-Infeld (DBI) action for a probe D3-brane, and leads to an identification of the canonical inflaton field with a multiple of the radial coordinate, $\phi^2 \equiv T_3 r^2$. Here, $T_3 \equiv [(2\pi)^3 g_s \alpha'^2]^{-1}$ is the D3-brane tension, with g_s the string coupling and $2\pi\alpha'$ the inverse string tension. The length of the throat, $\Delta r = r_{\rm UV} - r_{\rm IR} \approx r_{\rm UV}$ provides an upper limit on the inflaton field variation

$$\Delta\phi^2 \;<\; T_3\, r_{\rm UV}^2 \,. \tag{1370}$$

Naively, it seems that this could be made arbitrarily large by simply increasing the length of the throat. However, this changes the volume of the compact space which affects the four-dimensional Planck mass, the unit in which we should measure the inflaton variation. To take this effect into account, we notice that dimensional reduction relates the four-dimensional Planck mass, $M_{\rm pl}$, to the ten-dimensional gravitational coupling, $\kappa_{10}^2 = \frac{1}{2}(2\pi)^7 g_s^2 (\alpha')^4$,

$$M_{\rm pl}^2 = \frac{V_6}{\kappa_{10}^2} \,, \tag{1371}$$

where $V_6 \equiv \int {\rm d}^6 y \sqrt{g} e^{2A(y)}$ is the (warped) volume of the internal space. Since we are interested in an upper limit on $\Delta\phi/M_{\rm pl}$ we bound V_6 from below by the volume of the throat region (including an estimate of the bulk volume would only strengthen our conclusions)

$$V_6 \;>\; (V_6)_{\rm throat} = {\rm Vol}(X_5) \int_0^{r_{\rm UV}} {\rm d}r\, r^5 e^{2A(r)} = 2\pi^4 g_s N (\alpha')^2 r_{\rm UV}^2 \,, \tag{1372}$$

where $N \equiv MK$ measures the background flux. For control of the supergravity approximation (and to achieve sufficient warping of the background) we require $N \gg 1$. Combining the above results we find [116]

$$\boxed{\frac{\Delta\phi}{M_{\rm pl}} \;<\; \frac{2}{\sqrt{N}}} \,. \tag{1373}$$

Since $N \gg 1$, this implies that the inflaton variation will always be sub-Planckian, $\Delta\phi \ll M_{\rm pl}$, and the gravitational wave amplitude is necessarily small. We emphasize again that this argument was purely geometrical and didn't depend on the complicated details of the inflationary potential which we discuss next.

The D3-brane Potential

Inflation proceeds as a D3-brane moves radially inward in the throat region, towards an anti-D3-brane that is naturally situated at the tip of the throat. The exit from inflation occurs when open strings stretched between the approaching pair become tachyonic and condense, annihilating the branes.

In this simplified picture, inflation is driven by the extremely weak (warping-suppressed) Coulomb interaction of the brane-antibrane pair [46]. The true story, however, is more complex, as moduli stabilization introduces new terms in the inflaton potential which typically overwhelm the Coulomb term and drive more complicated dynamics [46, 117, 118, 47, 119, 48]. This pattern is precisely what we anticipated in our effective field theory discussion: integrating out moduli fields can be expected to induce important corrections to the potential.

An important correction induced by moduli stabilization is the inflaton mass term arising from the supergravity F-term potential, §29.2.2. In a vacuum stabilized by an F-term potential,

i.e. by superpotential terms involving the moduli, one finds the mass term $H_0^2\phi^2 = \frac{1}{3}V_0(\phi_\star)\frac{\phi^2}{M_{\rm pl}^2}$ [46], where $\phi_\star$ is an arbitrary reference value for the inflaton field and the parameter H_0 should not be confused with the present-day Hubble constant.

However, one expects additional contributions to the potential from a variety of other sources, such as additional effects in the compactification that break supersymmetry [48]. Let us define $\Delta V(\phi)$ to encapsulate all contributions to the potential aside from the Coulomb interaction $V_0(\phi)$ and the mass term $H_0^2\phi^2$; then the total potential and the associated contributions to the eta parameter may be written as

$$V(\phi) = V_0(\phi) + H_0^2\phi^2 + \Delta V(\phi) \tag{1374}$$

$$\eta_{\rm v}(\phi) = \eta_0 + \frac{2}{3} + \Delta\eta_{\rm v}(\phi) = ? \tag{1375}$$

where $\eta_0 \ll 1$ because the Coulomb interaction is very weak. (More generally, $V_0(\phi)$ can be *defined* to be all terms in $V(\phi)$ with negligible contributions to η. Besides the brane-antibrane Coulomb interaction, this can include any other sources of nearly-constant energy, *e.g.* bulk contributions to the cosmological constant.)

Clearly, $\eta_{\rm v}$ can only be small if ΔV can cancel the mass term in Eqn. (1374). We must therefore enumerate all relevant contributions to ΔV, and attempt to understand the circumstances under which an approximate cancellation can occur. Note that identifying a subset of contributions to ΔV while remaining ignorant of others is insufficient.

Warped D-brane inflation has received a significant amount of theoretical attention in part because of its high degree of computability. Quite generally, if we had access to the full data of an explicit, stabilized compactification with small curvatures and weak string coupling, we would in principle be able to compute the potential of a D-brane inflaton to any desired accuracy, by performing a careful dimensional reduction. This is not possible at present for a generic compact Calabi-Yau, for two reasons: for general Calabi-Yau spaces hardly any metric data is available, and examples with entirely explicit moduli stabilization are rare.

However, a sufficiently long throat is well-approximated by a *noncompact* throat geometry (*i.e.*, a throat of infinite length), for which the Calabi-Yau metric can often be found, as in the important example of the Klebanov-Strassler solution [113], which is entirely explicit and everywhere smooth. Having complete metric data greatly facilitates the study of probe D-brane dynamics, at least at the level of an unstabilized compactification. Furthermore, we will now explain how the effects of moduli stabilization and of the finite length of the throat can be incorporated systematically. The method involves examining perturbations to the supergravity solution that describes the throat in which the D3-brane moves. For concreteness we will work with the example of a KS throat, but the method is far more general. Our treatment will allow us to give explicit expressions for the correction terms ΔV in Eqn. (1374), and hence to extract the characteristics of inflation in the presence of moduli stabilization.

Sketch of the Supergravity Analysis

In the following we describe the computation of the inflaton potential for warped D3-brane inflation. This is only meant to give a flavor of the challenges involved in understanding the full potential. For more details we refer the reader to [46, 117, 47, 48, 118].

Type IIB string theory contains a good dozen of fields (going by names such as dilaton, p-form fluxes, warp factors, metric perturbations, etc.). In principle, we would have to worry that all those fields could couple to the inflaton degree of freedom and hence have to be considered when computing the inflaton potential to the desired accuracy. However, D3-branes are special in that they only couple to a specific combination of the warp factor and the five-form flux and are blind to perturbations in all other fields

$$V_{\rm D3}(\phi) = T_3(e^{4A} - \alpha) \equiv T_3\Phi_-\,, \tag{1376}$$

where the scalar function $\alpha(\phi)$ is related to the five-form flux F_5. We are therefore interested in perturbations of the object $\Phi_- = e^{4A} - \alpha$. In the KS background Φ_- vanishes, but coupling of the throat to the bulk geometry and interaction with moduli-stabilizing degrees of freedom like wrapped D7-branes, induces a non-zero Φ_-. To study the induced Φ_- perturbations, we

investigate the supergravity equation of motion

$$\boxed{\nabla^2\Phi_- = \frac{1}{24}|\mathcal{G}_-|^2 + R}\,, \tag{1377}$$

where $\mathcal{G}_-$ is a special (imaginary anti-self-dual) combination of 3-form fluxes and R is the 4-dimensional Ricci scalar. During inflation R is given by the square of the Hubble parameter H.

All fields are expressed as harmonic expansions on the five-dimensional base manifold $X_5 = T_{1,1}$, *e.g.*

$$\Phi_-(\phi, \Psi) = \sum_{LM} \Phi_{LM} \left(\frac{\phi}{\phi_{\rm UV}}\right)^{\Delta(L)} Y_{LM}(\Psi) + c.c.\,, \tag{1378}$$

where Ψ parameterizes five angles on $T_{1,1}$ and the scaling dimension Δ is determined by the eigenvalues of the angular Laplacian. The spectrum of eigenvalues hence determines the radial scaling of correction terms.

1. *Homogeneous solution*

   The solution to the homogeneous equation

   $$\nabla^2\Phi_- = 0\,, \tag{1379}$$

   was found in Ref. [48]. The leading corrections have the following radial scalings

   $$\Delta \;=\; \frac{3}{2}\,,\, 2\,,\, \cdots\,. \tag{1380}$$

2. *Inhomogeneous solution*

   (a) *Curvature-induced correction*

   The Ricci scalar of the four-dimensional de Sitter spacetime couples to the inflaton. This is reflected by a source term in Φ_- equation of motion

   $$\nabla^2\Phi_- = R\,. \tag{1381}$$

   For constant $R = 12H^2$ this induces a correction to the inflaton mass. This is precisely the Hubble scale inflaton mass term found by KKLMMT [46].

   (b) *Flux-induced corrections*

   Imaginary anti-self dual 3-form fluxes[182], $\star_6\mathcal{G}_- = -i\mathcal{G}_-$, also source corrections of the D3-brane potential [121]

   $$\nabla^2\Phi_- = \frac{1}{24}|\mathcal{G}_-|^2\,. \tag{1382}$$

   Consistently also solving the $\mathcal{G}_-$ equation of motion, $\mathrm{d}\mathcal{G}_- = 0$, we find the following leading corrections

   $$\Delta \;=\; 1\,,\frac{5}{2}\,,\, \cdots\,. \tag{1383}$$

In summary, solving Eqn. (1377) we found [48]

$$V_{\rm D3} = T_3\Phi_- = \sum_\Delta \phi^\Delta f_\Delta(\Psi)\,, \tag{1384}$$

where

$$\boxed{\Delta \;=\; 1\,,\, \frac{3}{2}\,,\, 2\,,\, \cdots}\,. \tag{1385}$$

The discrete *spectrum* (1385) of corrections to the inflaton potential determines the phenomenology of the model.

[182] Here, $\star_6$ is the six-dimensional Hodge star operator, see *e.g.* [120].

Phenomenological Implications

Two different scenarios arise depending on whether the $\Delta = \frac{3}{2}$ or the $\Delta = 2$ correction is the dominant contribution to $\Delta\eta_{\rm v}$ at small ϕ (note that $\Delta = 1$ doesn't contribute to $\eta_{\rm v}$):

1. *Quadratic case*

   If the $\Delta = \frac{3}{2}$ mode is projected out of the spectrum (this can be achieved by imposing discrete symmetries on the UV boundary conditions, see Ref. [48]), the effective radial potential is

   $$V(\phi) = V_0(\phi) + \beta H^2 \phi^2 \,. \tag{1386}$$

   The phenomenology of these types of potentials was first studied analytically by [46] and [122], and numerically by [123].

2. *Fractional case*

   If the fractional mode $\Delta = \frac{3}{2}$ is present, it leads to inflection point models [48, 47, 118, 119] (see Fig. 246).

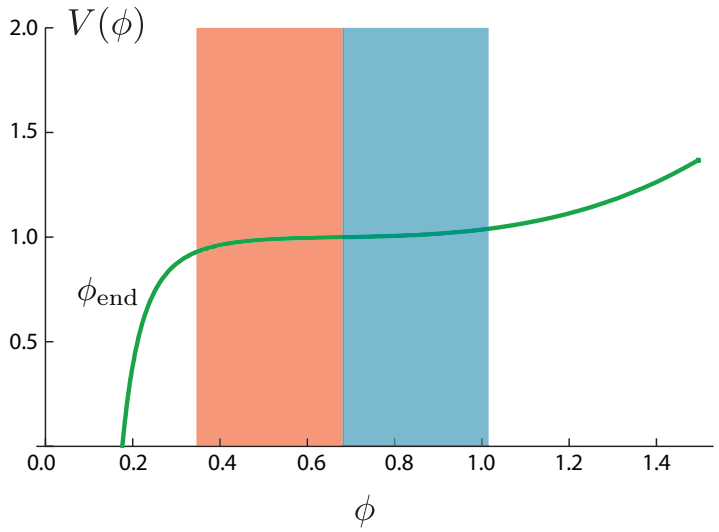


Figure 246: Inflection Point Inflation.

Summary and Perspective

In §29.2.1 we explained how the eta problem is sensitive to dimension-six Planck suppressed operators. In effective field theory models of inflation one can of course always *assume* a solution to the eta problem by a cancellation of the contributing correction terms; in other words, one can postulate that a flat potential $V(\phi)$ arises after an approximate cancellation among dimension-six Planck-suppressed corrections. In string theory models of inflation, to follow this path would be to abdicate the opportunity to use Planck-suppressed contributions as a (limited) window onto string theory. Moreover, once ϕ is identified with a physical degree of freedom of a string compactification, the precise form of the potential is in principle fully specified by the remaining data of the compactification. (Mixing conjecture into the analysis at this stage would effectively transform a 'string-derived' scenario into a 'string-inspired' scenario; the latter may be interesting as a cosmological model, but will not contribute to our understanding of string theory.) Thus, overcoming the eta problem becomes a detailed computational question. One can in principle compute the full potential from first principles, and in practice one can often classify corrections to the leading-order potential.

In this section, we have enumerated the leading corrections for warped D-brane inflation and showed that an accidental cancellation (or fine-tuning) allows small eta over a limited range of inflaton values. This gives a non-trivial existence proof for inflationary solutions in warped throat models with D3-branes.

29.3.3 *Case Study:* Axion Monodromy Inflation

We now turn to our second case study, an example of large-field inflation in string theory. As we have discussed in §29.2.4, the particular challenge in these models is the need to control an infinite series of contributions to the inflaton potential, arising from couplings of the inflaton to degrees of freedom with masses near the Planck scale. Direct enumeration and fine-tuning of such terms (as in the small-field example in §29.3.2) is manifestly impractical, and it appears essential to develop a symmetry argument controlling or forbidding these terms.

An influential proposal in this direction is Natural Inflation [124] (see **Lecture 1**), in which a pseudo-Nambu-Goldstone boson (*i.e.*, an axion) is the inflaton. At the perturbative level, the axion field a enjoys a continuous shift symmetry $a \to a + const$ which is broken by non-perturbative effects to a discrete symmetry $a \to a + 2\pi$. The nonperturbative effects generate a periodic potential

$$V(\phi) = \frac{\Lambda^4}{2}\left[1 - \cos\left(\frac{\phi}{f}\right)\right] + \dots , \tag{1387}$$

where Λ is a dynamically-generated scale, f is known as the axion decay constant, $\phi \equiv af$, and the omitted terms are higher harmonics.

As explained above, an important question, in any proposed effective theory in which a super-Planckian field range is protected by a shift symmetry, is whether this structure can be UV-completed. We should therefore search in string theory for an axion with decay constant $f > M_{\rm pl}$.

Axions in String Theory

Axions from p-Forms

Axions are plentiful in string compactifications, arising from p-form gauge potentials integrated on p-cycles of the compact space. For example, in type IIB string theory, there are axions $b_i = 2\pi \int_{\Sigma_i} B$ arising from integrating the Neveu-Schwarz (NS) two-form B over two-cycles Σ_i, as well as axions $c_i = 2\pi \int_{\Sigma_i} C$ arising from the Ramond-Ramond (RR) two-form C. In the absence of additional ingredients such as fluxes and space-filling wrapped branes, the potential for these axions is classically flat and has a continuous shift symmetry which originates in the gauge invariance of the ten-dimensional action. Instanton effects break this symmetry to a discrete subgroup, $b_i \to b_i + 2\pi$ ($c_i \to c_i + 2\pi$). This leads to a periodic contribution to the axion potential whose periodicity we will now estimate. We will find that the axion decay constants are smaller than $M_{\rm pl}$ in known, computable limits of string theory [125, 126]. Readers less familiar with string compactifications can accept this assertion and skip to §29.3.3 without loss of continuity.

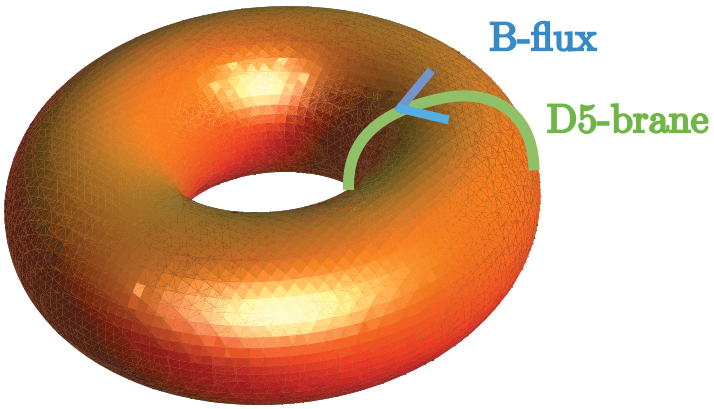


Figure 247: Axion Monodromy

Axion Decay Constants in String Theory

Let ω^i be a basis for $H^2(X, Z)$, the space of two-forms on the compact space X, with $\int_{\Sigma_i} \omega^j = \alpha' \delta_i{}^j$. The NS two-form potential B may be expanded as

$$B = \frac{1}{2\pi} \sum_i b_i(x)\, \omega^i \,, \tag{1388}$$

with x the four-dimensional spacetime coordinate. The axion decay constant can be inferred from the normalization of the axion kinetic term, which in this case descends from the ten-dimensional term

$$\frac{1}{(2\pi)^7 g_s^2 \alpha'^4} \int \mathrm{d}^{10}x\, \frac{1}{2} |\mathrm{d}B|^2 \quad \supset \quad \frac{1}{2} \int \mathrm{d}^4x \sqrt{-g}\, \gamma^{ij} (\partial^\mu b_i \partial_\mu b_j) \,, \tag{1389}$$

where

$$\gamma^{ij} \equiv \frac{1}{6(2\pi)^9 g_s^2 \alpha'^4} \int_X \omega^i \wedge \star_6\, \omega^j \tag{1390}$$

and $\star_6$ is the six-dimensional Hodge star operator. By performing the integral over the internal space X and diagonalizing the field space metric as $\gamma^{ij} \to f_i^2 \delta_{ij}$, one can extract the axion decay constant f_i.

It is too early to draw universal conclusions, but a body of evidence suggests that the resulting axion periodicities are always smaller than $M_{\rm pl}$ in computable limits of string theory [125, 126]. As this will be essential for our arguments, we will illustrate this result in a simple example. Suppose that the compactification is isotropic, with typical length-scale L and volume L^6. Then using

$$\alpha' M_{\rm pl}^2 = \frac{2}{(2\pi)^7} \frac{L^6}{g_s^2 \alpha'^3} \tag{1391}$$

we find from Eqn. (1390) that

$$f^2 \approx M_{\rm pl}^2 \frac{\alpha'^2}{6(2\pi)^2 L^4} \,. \tag{1392}$$

In controlled compactifications we require $L \gg \sqrt{\alpha'}$, so that $f \ll M_{\rm pl}$. Qualitatively similar conclusions apply in much more general configurations [125, 126].

Axion Inflation in String Theory

The above result would seem to imply that Natural Inflation from a single axion field cannot be realized in known string compactifications: string theory provides many axions, but none of these has a sufficiently large field range. However, there are at least two reasonable proposals to circumvent this obstacle.

N-flation

The first suggestion was that a collective excitation of many hundreds of axions could have an effective field range large enough for inflation [43, 127]. The role of the inflaton is played by the collective field

$$\phi^2 = \sum_{i=1}^{N} \phi_i^2 \,. \tag{1393}$$

Even if each individual field has a sub-Planckian field range, $\phi_i < M_{\rm pl}$, for sufficiently large number of fields N, the effective field ϕ can have a super-Planck excursion. This 'N-flation' proposal is a specific example of assisted inflation [128], but, importantly, one in which symmetry helps to protect the axion potential from corrections that would impede inflation. Although promising, this scenario still awaits a proof of principle demonstration, as the presence of a large number of light fields leads to a problematic renormalization of the Newton constant, and hence to an effectively reduced field range. For recent studies of N-flation see [129, 130].

Axion Monodromy

We will instead describe an elementary mechanism, *monodromy*, which allows inflation to persist through multiple circuits of a single periodic axion field. A system is said to undergo monodromy if, upon transport around a closed loop in the (naive) configuration space, the system reaches a new configuration. A spiral staircase is a canonical example: the naive configuration space is described by the angular coordinate, but the system changes upon transport by 2π. (In fact, we will find that this simple model gives an excellent description of the potential in axion monodromy inflation.) The idea of using monodromy to achieve controlled large-field inflation in string theory was first proposed by Silverstein and Westphal [251], who discussed a model involving a D4-brane wound inside a nilmanifold. In this section we will focus instead on the subsequent axion monodromy proposal of Ref. [44], where a monodromy arises in the four-dimensional potential energy upon transport around a circle in the field space parameterized by an axion.

Monodromies of this sort are possible in a variety of compactifications, but we will focus on a single concrete example. Consider type IIB string theory on a Calabi-Yau orientifold, *i.e.* a quotient of a Calabi-Yau manifold by a discrete symmetry that includes worldsheet orientation reversal and a geometric involution. Specifically, we will suppose that the involution has fixed points and fixed four-cycles, known as O3-planes and O7-planes, respectively. If in addition the compactification includes a D5-brane that wraps a suitable two-cycle Σ and fills spacetime, then the axion $b = 2\pi \int_\Sigma B$ can exhibit monodromy in the potential energy. (Similarly, a wrapped NS5-brane produces monodromy for the axion $c = 2\pi \int_\Sigma C$.) In other words, a D5-brane wrapping Σ carries a potential energy that is *not* a periodic function of the axion, as the shift symmetry of the axion action is broken by the presence of the wrapped brane; in fact, the potential energy increases without bound as b increases.

In the D5-brane case, the relevant potential comes from the Dirac-Born-Infeld action for the wrapped D-brane,

$$S_{\rm DBI} = \frac{1}{(2\pi)^5 g_s \alpha'^3} \int_{\mathcal{M}_4 \times \Sigma} \mathrm{d}^6\xi \, \sqrt{\det(G+B)} \tag{1394}$$

$$= \frac{1}{(2\pi)^6 g_s \alpha'^2} \int_{\mathcal{M}_4} \mathrm{d}^4 x \, \sqrt{-g} \, \sqrt{(2\pi)^2 \ell_\Sigma^4 + b^2} \,, \tag{1395}$$

where ℓ_Σ is the size of the two-cycle Σ in string units. The brane energy, Eqn. (1395), is clearly not invariant under the shift symmetry $b \to b + 2\pi$, although this is a symmetry of the corresponding compactification without the wrapped D5-brane. Thus, the DBI action leads directly to monodromy for b. Moreover, when $b \gg \ell_\Sigma^2$, the potential is asymptotically *linear* in the canonically-normalized field $\varphi_b \propto b$.

The qualitative inflationary dynamics in this model is as follows: One begins with a D5-brane wrapping a curve Σ, upon which $\int_\Sigma B$ is taken to be large. In other words, the axion b has a large initial vev. Inflation proceeds by the reduction of this vev, until finally $\int_\Sigma B = 0$ and the D5-brane is nearly 'empty', *i.e.* has little worldvolume flux. During this process the D5-brane does not move, nor do any of the closed-string moduli shift appreciably. For small axion vevs, the asymptotically linear potential we have described is inaccurate, and the curvature of the potential becomes non-negligible; see Eqn. (1395). At this stage, the axion begins to oscillate around its origin. Couplings between the axion and other degrees of freedom, either closed string modes or open string modes, drain energy from the inflaton oscillations. If a sufficient fraction of this energy is eventually transmitted to visible-sector degrees of freedom – which may reside, for example, on a stack of D-branes elsewhere in the compactification – then the hot Big Bang begins. The details of reheating depend strongly on the form of the couplings between the Standard Model degrees of freedom and the inflaton, and this is an important open question, both in this model and in string inflation more generally.

Compactification Considerations

Having explained the essential idea of axion monodromy inflation, we must still ensure that the proposed inflationary mechanism is compatible with moduli stabilization and can be realized in a consistent compactification. An immediate concern is whether there are additional contributions to the potential, beyond the linear term identified above, that could have important effects during inflation. As we have emphasized throughout this review, one expects that in the absence of a symmetry protecting the inflaton potential, generic corrections due to moduli stabilization will contribute $\Delta\eta \sim \mathcal{O}(1)$. It is therefore essential to verify that the continuous shift symmetry which protects the inflaton potential is preserved to an appropriate degree by the stabilized compactification. For the special case of moduli stabilization in which nonperturbative effects play a role, ensuring that the shift symmetry is not spoiled can be quite subtle. This is described in detail in Ref. [44].

Summary and Perspective

The Lyth bound shows that an observable gravitational wave signal correlates with the inflaton field moving over a super-Planckian distance during inflation. Effective field theory models of large-field inflation then require a shift symmetry to protect the flatness of the potential over a super-Planckian range. It has therefore become an important question whether such shift symmetries arise in string theory and can be used to realize large-field inflation.

In this section, we argued that the first examples of shift symmetries in string theory that protect the potential over a super-Planckian range are becoming available. We explained the dual role of the monodromy: i) it results in a large kinematic field range $\Delta\phi > M_{\rm pl}$ by allowing a small fundamental domain to be traversed repeatedly, and ii) in combination with the shift symmetry it controls corrections to the potential over a super-Planckian range. The shift symmetry, only weakly broken by V, controls corrections ΔV within a fundamental domain, and furthermore relates corrections in one fundamental domain to those in any other. Monodromy therefore effectively reduces a large-field problem to a small-field problem [251].

Although more work is required to understand these models and the compactifications in which they arise, monodromy appears to be a robust and rather promising mechanism for realizing large-field inflation, and hence an observable gravitational wave signal, in string theory.

29.4 Outlook

29.4.1 Theoretical Prospects

As we hope this lecture has illustrated, theoretical progress in recent years has been dramatic. A decade ago, only a few proposals for connecting string theory to cosmology were available, and the problem of stabilizing the moduli had not been addressed. We now have a wide array of inflationary models motivated by string theory, and the best-studied examples among these incorporate some information about moduli stabilization. Moreover, a few mechanisms for inflation in string theory have been shown to be robust, persisting after full moduli stabilization with all relevant corrections included.

Aside from demonstrating that inflation is possible in string theory, what has been accomplished? In our view the primary use of explicit models of inflation in string theory is as test cases, or toy models, for the sensitivity of inflation to quantum gravity. On the theoretical front, these models have underlined the importance of the eta problem in general field theory realizations of inflation; they have led to mechanisms for inflation that might seem unnatural in field theory, but are apparently natural in string theory; and they have sharpened our understanding of the implications of a detection of primordial tensor modes.

It is of course difficult to predict the direction of future theoretical progress, not least because unforeseen fundamental advances in string theory can be expected to enlarge the toolkit of inflationary model-builders. However, it is safe to anticipate further gradual progress in moduli stabilization, including the appearance of additional explicit examples with all moduli stabilized; entirely explicit models of inflation in such compactifications will undoubtedly follow. At present, few successful models exist in M-theory or in heterotic string theory, and under mild assumptions, inflation can be shown to be impossible in certain classes of type IIA compactifications [131, 132, 133]. It would be surprising if it turned out that inflation is much

more natural in one weakly-coupled limit of string theory than in the rest, and the present disparity can be attributed in part to the differences among the moduli-stabilizing tools presently available in the various limits. Clearly, it would be useful to understand how inflation can arise in more diverse string vacua.

The inflationary models now available in string theory are subject to stringent theoretical constraints arising from consistency requirements (*e.g.*, tadpole cancellation) and from the need for some degree of computability. In turn, these limitations lead to correlations among the cosmological observables, *i.e.* to predictions. Some of these constraints will undoubtedly disappear as we learn to explore more general string compactifications. However, one can hope that some constraints may remain, so that the set of inflationary effective actions derived from string theory would be a proper subset of the set of inflationary effective actions in a general quantum field theory. Establishing such a proposition would require a far more comprehensive understanding of string compactifications than is available at present.

29.4.2 Observational Signatures?

The theoretical aspects of inflation described in this lecture are interesting largely because they can be *tested* experimentally using present and future cosmological data (see **Lecture 3**).

As we have repeatedly emphasized throughout these lectures, the most dramatic confirmation of inflation would come from a detection of B-mode polarization, which would establish the energy scale of inflation and would indicate that the inflaton traversed a super-Planckian distance. As we have argued in this lecture, super-Planckian displacements are a key instance in which the inflaton effective action is particularly sensitive to the physics of the Planck scale. As a concrete example of the discriminatory power of tensor perturbations, any detection of primordial gravitational waves would exclude the warped D3-brane inflation scenario of §29.3.2 [116], while an upper bound $r < 0.07$ (or a detection with $r \gg 0.07$) would exclude the axion monodromy scenario of §29.3.3 [44].

A further opportunity arises because single-field slow-roll inflation predicts null results for many cosmological observables, as the primordial scalar fluctuations are predicted to be scale-invariant, Gaussian and adiabatic to a high degree. A detection of non-Gaussianity, isocurvature fluctuations or a large scale-dependence (running) would therefore rule out single-field slow-roll inflation. Inflationary effective actions that do allow for a significant non-Gaussianity, non-adiabaticity or scale-dependence often require higher-derivative interactions and/or more than one light field, and such actions arise rather naturally in string theory. Although we have focused in this lecture on the sensitivity of the inflaton potential to Planck-scale physics, the inflaton kinetic term is equally UV-sensitive, and string theory provides a promising framework for understanding the higher-derivative interactions that can produce significant non-Gaussianity [92, 77].

Finally, CMB temperature and polarization anisotropies induced by relic cosmic strings or other topological defects provide probes of the physics of the end of inflation or of the post-inflationary era. Cosmic strings are automatically produced at the end of brane-antibrane inflation [134, 135], and the stability and phenomenological properties of the resulting cosmic string network are determined by the properties of the warped geometry. Detecting cosmic superstrings via lensing or through their characteristic bursts of gravitational waves is an exciting prospect.

30 Scalar Speed Limits and Cosmology: Acceleration from D-cceleration by Silverstein, Tong

Causality on the gravity side of the AdS/CFT correspondence restricts motion on the moduli space of the $\mathcal{N} = 4$ super Yang Mills theory by imposing a speed limit on how fast the scalar field may roll. This effect can be traced to higher derivative operators arising from integrating out light degrees of freedom near the origin. In the strong coupling limit of the theory, the dynamics is well approximated by the Dirac-Born-Infeld Lagrangian for a probe D3-brane moving toward the horizon of the AdS Poincare patch, combined with an estimate of the (ultimately suppressed) rate of particle and string production in the system. We analyze the

motion of a rolling scalar field explicitly in the strong coupling regime of the field theory, and extend the analysis to cosmological systems obtained by coupling this type of field theory to four dimensional gravity. This leads to a mechanism for slow roll inflation for a massive scalar at subPlanckian VEV without need for a flat potential (realizing a version of k-inflation in a microphysical framework). It also leads to a variety of novel FRW cosmologies, some of which are related to those obtained with tachyon matter.

30.1 Introduction

It is now almost 100 years since Einstein introduced the concept of a universal speed limit for all physical systems propagating in spacetime. However, in general, motion on the configuration space of a physical system is not constrained to obey a speed limit. For example, the vacuum expectation values (VEVs) of light scalar fields naturally move in an internal moduli space where motion is uninhibited; within classical, relativistic field theory there is no restriction on the rate of change of the VEV. The purpose of this paper is to show that this situation does not necessarily continue to hold in the quantum theory. We exhibit situations where the strong coupling dynamics do impose a speed limit on an internal moduli space, and examine applications to cosmology, including a mechanism for slow roll inflation.

We focus on the $\mathcal{N} = 4$ supersymmetric Yang Mills (SYM) theory. Although the (supersymmetry-protected) metric on the moduli space is flat, the quantum induced speed limit ensures that a rolling scalar field slows down as it approaches the origin. This fact can be seen immediately from the gravity side of the AdS/CFT correspondence, where the process corresponds to a D3-brane domain wall in AdS_5 moving toward the horizon. The familiar causal speed-limit in the bulk translates into a speed limit on moduli space which becomes more pronounced as the brane approaches the origin. Indeed, from the perspective of a boundary observer, the probe brane takes an infinite time to cross the horizon (though as we will discuss, the probe approximation breaks down due to back reaction at very late times). This speed limit on the moduli space, arising from causality in the bulk, was first stressed by Kabat and Lifschytz [1].

On the field theory side of the correspondence, this result reflects the breakdown of the moduli space σ-model approximation as a scalar field approaches a locus with new light degrees of freedom where higher derivative terms become important. In AdS/CFT dual pairs, we can use the gravity side of the correspondence to determine the net effect of these higher-derivative corrections where they are summed into a Dirac-Born-Infeld (DBI) action. The resulting dynamics is dramatically different from naive expectation based on the supersymmetric moduli space metric.

In this paper, we will study the dynamics explicitly, both in quantum field theory in its own right and in the cosmological context arising from quantum field theory coupled to four dimensional gravity. We use the dual picture of $D3$-branes and anti-$D3$-branes moving in an AdS-like throat and we exhibit late time solutions describing the physics as the scalar field approaches the origin. Among our results, we find that the slowing down of the scalar field (relative to the behavior predicted by the two derivative action) can lead to new regimes exhibiting inflationary behavior.

The idea of obtaining inflation from $\overline{D3}$ and $D3$ branes in a warped throat was studied recently in [2]. In particular, the authors observe that one can gain extra control from the warping, but some obstacles to obtaining explicit inflationary models in string compactifications were identified. Our results here concern the effects of the crucial higher derivative terms in the effective action and allow us to probe a different regime from that studied in [2] which may

help address the challenges of [2]. In particular the slow motion of the scalar field, enforced by higher derivative terms, leads to a new mechanism for slow roll inflation.

A scalar field mass is required to obtain our simplest inflationary solution. The required mass scale is not finely tuned in ordinary four dimensional effective field theory terms, which provides an interesting distinction from the usual inflationary models. In the case of the $\mathcal{N} = 4$ super Yang Mills theory coupled to gravity, we propose some effective field theory couplings to other sectors generating such a mass while leaving intact the form of the crucial kinetic corrections to the moduli space approximation of the field theory. However, we should note that the couplings of Kaluza Klein modes in the corresponding brane throat to other sectors may be important and may require tuned coefficients to avoid destabilization of the throat; this is a little understood aspect of current compactification technology. Because of the plethora of independent ingredients (and types of domain wall branes) available in string compactifications, we expect the combination of approximate AdS-metric induced kinetic terms and scalar masses generated from other couplings is likely to be available in some set of examples. It remains the main weakness of our results however that we will not exhibit an explicit example here. The other field theoretic and cosmological phases we will exhibit are not tied to this subtlety.

Our results have further relations to previous work. The use of higher derivative terms to change scalar field dynamics in ways interesting for cosmology has been dubbed "k-inflation" [3] or "k-essence" [4]. A particularly well-studied example occurs in the effective field theory describing the decay of branes and antibranes leading to "tachyon matter" [5, 6, 7] and the associated cosmology [8, 9]. While our system shares many features with the rolling tachyon story, including the existence of a pressureless dust equation of state, it also differs in several ways important for cosmology. In particular, we shall argue that particle and string production is suppressed in our system and the dynamics is governed to good approximation by our equations of motion. For a few examples of previous works studying the application of rolling moduli to string cosmology see [10, 11, 12] and the review [13]. The crucial physics of our model is ultimately extracted from the dynamics of D-branes; there are of course many interesting investigations in this area, for example [14, 15, 16].

The manner in which higher derivative terms can drastically affect the dynamics of a system raises many questions concerning potential applications to real models of cosmology as well as more theoretical issues. Among the latter set is the question of whether or not motion toward other finite distance singularities in field theory and string theory moduli spaces (such as the conifold singularity) also exhibits similar slowing down effects for some regime of the couplings. Noticeably, recent studies of the flop transition, which explore the effects of the states which become light at the singularity, suggest that the rolling scalar field does indeed become stuck in the region of the singularity [17, 18].

The paper is organized as follows. We start in Section 2 by describing our basic set-up, stressing the appearance of a speed limit on the moduli space. In Section 3 we study the consequences of this speed limit for the dynamics in the global conformal field theory. We then couple our system to gravity in Section 4. We take particular care to describe the possible deformations of the system arising from other sectors in a string compactification, including the generation of a potential on the moduli space and the effects this has on the AdS geometry seen by the probe brane. In Section 5 we study cosmologies arising from our low-energy effective actions. In Section 6, we check that our solutions are not destabilised by perturbations or particle production.

30.2 The Basic Setup

Typically studies of scalar field dynamics consider an effective field theory Lagrangian containing a kinetic term (up to two derivatives) together with a potential energy on the space of scalar fields:

$$S_{two\ deriv} = \int d^4x\sqrt{g}\left(G_{ij}(\phi)g^{\mu\nu}\partial_\mu\phi^i\partial_\nu\phi^j - V(\phi)\right) \tag{30.1396}$$

In supersymmetric situations, the moduli space metric G_{ij} and the form of the potential energy V are highly constrained, and much of the recent work on such quantum field theories has focused on determining these quantities and the BPS spectrum of states exactly.

It is well known however that the moduli space approximation (in which the physics is governed by the action (30.1396) along flat directions of the potential $V(\phi)$) can break down due to the presence of new light degrees of freedom arising on special loci of the space of scalar field VEVs.

In the $U(N)$ $\mathcal{N} = 4$ SYM theory, one has scalar fields in the adjoint representation of the gauge group which can be represented by $N \times N$ matrices Φ. The moduli space of the field theory is parameterized by diagonal (commuting) matrices, with eigenvalues $\phi_1, \ldots, \phi_N$ whose moduli space metric, including quantum corrections, is flat. Away from the origin $\phi_i = 0$ of this moduli space, the low energy gauge symmetry is generically $U(1)^N$ and off diagonal matrix elements of the scalar fields, fermions, and gauge bosons obtain masses. We shall refer to these modes collectively as "W bosons". In particular, we will consider a configuration in which $\langle\phi_1\rangle \equiv \langle\phi\rangle \neq 0$ but all the other ϕ_i, $i \neq 1$ have vanishing VEVs. This means that the theory has a low energy unbroken gauge symmetry $U(N-1) \times U(1)$ with massive W bosons in the $(\mathbf{N-1}, +) \oplus (\overline{\mathbf{N-1}}, -)$ representation. The W bosons are BPS protected states with mass

$$m_W = \phi \tag{30.1397}$$

We will work expanding about the large N limit of the theory, in which the natural parameters in the field theory are the rank N and the 't Hooft coupling $\lambda = g_{YM}^2 N$ in terms of the Yang Mills coupling g_{YM}.

30.2.1 The System at Weak Coupling

At weak coupling in the field theory, the effective action for ϕ gets contributions from virtual W bosons. For large $\langle\phi\rangle$, these contributions scale like powers of $\lambda\dot{\phi}^2/\phi^4$ in the planar limit (see [19] for a comprehensive discussion of these higher derivative corrections). If we send the scalar field rolling toward the origin from a finite point on the Coulomb branch, the classical action (30.1396) would predict a constant velocity $\dot{\phi} = v_0$, but the form of the corrections just noted shows that the action (30.1396) becomes unreliable at the distance

$$\phi^2 = \sqrt{\lambda}\, v_0 \tag{30.1398}$$

The question then arises about how to take into account these corrections. However, in the weak coupling regime, the point is moot as we have another issue to confont before we get this close to the origin. Since the W bosons become light, they may be produced on shell during the evolution. The time dependent mass leads to particle production controlled by the parameter $\dot{m}_W/m_W^2$, which therefore becomes important at the distance,

$$\phi^2 = v_0 \tag{30.1399}$$

For $\lambda \ll 1$, we first reach the production point (30.1399).

Naively extrapolating the above perturbative results to the strong coupling regime, we reach the point (30.1398) first and therefore expect that the dynamics will be governed by the effective action for ϕ taking into account the $\lambda\dot{\phi}^2/\phi^4$ corrections arising from virtual W bosons. In the following section we shall confirm this expectation and show that the scale (30.1398) is where the speed limit on the moduli space first becomes apparent.

30.2.2 The System at Strong Coupling and the Speed Limit

At strong coupling $\lambda \gg 1$, the effective description of the theory is via gravity and string theory using the AdS/CFT correspondence [20]. A point on the Coulomb branch is described on the gravity side in the Poincare patch via a D3-brane at fixed radial position r in the metric

$$ds^2 = \frac{r^2}{R^2}\left(-dt^2 + dx^2\right) + \frac{R^2}{r^2}\,dr^2 \tag{30.1400}$$

and at a point on the S^5. The field theory coordinate ϕ on the Coulomb branch translates into r/α' in the gravity variables. We further have the relations $R = (g_s N)^{1/4}\sqrt{\alpha'}$ and $g_{YM}^2 = 4\pi g_s$.

As in other applications of AdS/CFT [20] and Randall-Sundrum [21], the warp factor in (30.1400) plays an important role in understanding the energy spectrum of the system. The string mass scale at position r is related to that at position r' by the ratio of warp factors r/r'. The open string oscillator modes on the brane at position $r = \langle\phi\rangle\alpha'$ have masses of order

$$m_s(\langle\phi\rangle) = \frac{\phi}{\lambda^{1/4}} \tag{30.1401}$$

As in the Randall Sundrum scenario, the effective cutoff for modes on the brane is at this warped string energy scale (30.1401) rather than the UV string scale $1/\sqrt{\alpha'}$.

The W bosons are strings stretched from the brane to the horizon at $r = 0$. Although this is an infinite proper distance for the string to stretch, the warp factor r/R reduces the effective tension of the string enough to produce a finite total mass $m_W = \phi$ in accord with the BPS formula (30.1397).

We would now like to highlight the simple feature of our system which lies behind most of our detailed results to follow. It is immediately clear from (30.1400) that, for a boundary observer using coordinates (t, x), a probe brane takes infinite time to reach the origin of the Coulomb branch at $r = 0$. (In our solutions it will turn out that the proper time for the probe to fall to the origin will be finite.) The radial velocity of the $D3$-brane in the AdS space is limited by the speed of light which, translated in field theory variables, becomes

$$\dot{\phi} \le \dot{\phi}_c = \frac{\phi^2}{\sqrt{\lambda}} \tag{30.1402}$$

This restriction was noted previously by Kabat and Lifschytz [1], who discussed some interesting aspects of the phenomenon. Note that the probe brane takes infinite time despite the fact that the moduli space metric in (30.1396) for the $\mathcal{N} = 4$ SYM theory is uncorrected quantum mechanically; the distance to the origin in the field theory moduli space metric is finite for any value of the coupling. As discussed above, the fact that it takes an infinite time to reach the origin arises from crucial corrections to the moduli space approximation. These corrections apply to *any* physical process in which the scalar rolls toward the origin of the Coulomb branch.

Note that the $\lambda\dot{\phi}^2/\phi^4$ corrections, while they are higher derivative corrections, are not suppressed by powers of the Planck mass; as we discussed above they are suppressed only by the W boson mass ϕ. Let us compare this situation to a brane in flat space: such an object has a Lagrangian proportional to $\sqrt{1-v^2}$ where v is the proper velocity of the brane. Written in terms of canonically scaled brane fields ϕ, this becomes (for a D3-brane) $\sqrt{1-\dot{\phi}^2\alpha'^2}$ which expands to a series of higher derivative terms which are suppressed by powers of the string mass m_s. In the global limit $m_s \to \infty$, these corrections die and the motion on ϕ space is unconstrained. It is the warp factor in (30.1400) that produces higher derivative effects that are crucial at the field theory level, as we will see in detail in our analysis.

30.3 The Global CFT

The corrections to the moduli space approximation, and the resulting dynamics, can be understood rather explicitly in a controlled analysis on the gravity side in which the rolling scalar VEV is modelled by a moving D3-brane probe in AdS_5. In this section we examine the resulting dynamics of the scalar field.

30.3.1 Effective Action and Approximation Scheme

On the gravity side, the effective Lagrangian appropriate for a probe D-brane at arbitrary velocity (less than or equal to the speed of light) but low proper acceleration is the Dirac-Born-Infeld Lagrangian. This Lagrangian describes the effects of virtual open strings at the planar level; it includes the effects of the background geometry and field strengths but does not include production of on shell closed strings or W bosons or loops of closed strings. We will analyze the motion of the 3-brane starting from this Lagrangian, and check for self consistency of the resulting solutions. This requires checking that the proper acceleration is small and taking the string coupling to be small so that the DBI action is a good approximation to the effective action of the probe. It further requires determining the range of parameters and times for which the energy in the probe does not back react significantly on the geometry, so that the probe approximation remains valid. Finally, we must also check that fluctuations about the solution, density perturbations, and particle and string production are not too large. In Section 4 we will consider generalizations of this system to include coupling to four dimensional gravity and other sectors, including effects of a cutoff throat in the IR; we will take into account new contributions to the action at the level of its most relevant terms at low energy.

In this section we start by writing down the effective action applicable to the pure conformal field theory (CFT) without gravity (infinite AdS space). The DBI action for a probe D3-brane moving in $AdS_5 \times S^5$ can be found in [20] and we will use their conventions. We concentrate only on radial fluctuations, and set the field strength on the brane to vanish. Working in field theory variables $\phi \equiv r/\alpha'$, we have

$$S = -\frac{1}{g_{YM}^2} \int dt\, d^3x\, f(\phi)^{-1} \left[-\det\left(\eta_{\mu\nu} + f(\phi)\,\partial_\mu\phi\,\partial_\nu\phi\right)^{1/2} - 1 \right] \tag{30.1403}$$

where for now we take the background brane metric to be flat $\eta_{\mu\nu} = \mathrm{diag}(-1,1,1,1)$, and $f(\phi)$ is the harmonic function

$$f(\phi) = \frac{\lambda}{\phi^4} \tag{30.1404}$$

Expanding the action out in derivatives leads to a canonical kinetic term for $\phi \equiv r/\alpha'$ and a series of higher order derivative interactions. The potential cancels out, reflecting the BPS

nature of the D3-brane. In fact, for a single excited scalar field ϕ, this form of the higher derivative action can be shown to hold even at weak coupling [22] without recourse to a dual gravitational description: it is fixed by the requirement of a non-linearly realised conformal invariance, together with known non-renormalisation theorems.

So in general the dynamics of ϕ in our strongly coupled CFT differs in two important ways from a naive scalar field system governed by (30.1396). Firstly, the kinetic term is corrected to that in (30.1403); as discussed above we will find that the proper velocity approaches the speed of light so that the quantity in the square root approaches zero, resulting in a system far away from the regime where the two derivative action suffices. Secondly, there is no potential $V(\phi)$. In later sections we shall remedy the latter issue by generalizing to systems in which potentials are generated, both by anti-branes and by considering a field theory coupled to other sectors. The speed limit will remain a crucial ingredient in our generalizations.

30.3.2 Dynamics

Let us now study more explicitly the approach of the $\mathcal{N} = 4$ SYM theory to the origin of its Coulomb branch, using the action (30.1403). Since we are interested only in the time dependence of the solution, we shall ignore the spatial derivatives, leaving us with the action,

$$S = -\frac{N}{\lambda^2}\int d^4x\phi^4\left(\sqrt{1-\lambda\dot{\phi}^2/\phi^4}-1\right) \tag{30.1405}$$

To determine the late time behaviour[183] of this system, we first compute the conserved energy density E, given by (up to a factor of $1/g_{YM}^2$)

$$E = \frac{1}{\lambda}\phi^4\left(\frac{1}{\sqrt{1-\lambda\dot{\phi}^2/\phi^4}}-1\right) \approx \tfrac{1}{2}\dot{\phi}^2 + \tfrac{1}{8}\frac{\lambda\dot{\phi}^4}{\phi^4} + \dots \tag{30.1406}$$

where, in the second line, we have expanded around large ϕ. The first term is the canonical kinetic energy for a rolling scalar field, with subsequent terms becoming important at small ϕ, capturing the effect of virtual W bosons.

As mentioned above, the Born-Infeld action is valid for arbitrarily high velocities $\dot{\phi}$ but only for small proper accelaration a,

$$a\sqrt{\alpha'} = \frac{\lambda^{1/4}}{\phi}\frac{d}{dt}\left(\frac{\sqrt{\lambda}\,\dot{\phi}}{\phi^2}\right)$$

So in order for our analysis to be self-consistent, we must check that the proper acceleration is much smaller than string scale in our solution. Inverting (30.1406) to solve for $\dot{\phi}$, and subsequently taking the time derivative, it is simple to see that we can trust the DBI analysis when we are in the strong coupling regime $\lambda \gg 1$.

[183] Readers who prefer $u = 1$ units may make the substitution in all expressions of this type.

Integrating once, the trajectory of the scalar ϕ is given by,

$$t - t_0 = \frac{1}{\sqrt{E}} \int_{\phi_0}^{\phi} d\varphi \, \frac{1}{\varphi^2} \frac{\lambda E + \varphi^4}{\sqrt{\lambda E + 2\varphi^4}}$$

We are interested in the dynamics of the scalar field after we hit the critical point on moduli space defined by (30.1398). Using the classical expectation $E \sim v_0^2$, this means we are interested in the regime

$$\phi^4 \ll \lambda E \tag{30.1407}$$

Here we find the late time behaviour,

$$\phi(t) \to \frac{\sqrt{\lambda}}{t} \tag{30.1408}$$

saturating the speed limit (30.1402). We therefore find that from the perspective of the field theory observer, Xeno is vindicated and the scalar field takes an infinite time to reach the origin as advertised.

Background Check

The above calculation treated the $D3$-brane as a probe and is valid only when the back reaction can be neglected. Since the brane travels at almost the speed of light at late times on the gravity side, it carries a lot of energy and we must determine the conditions under which its back reaction does not destabilize the AdS background.[184] The gravitational field surrounding a highly boosted object in locally flat space is given by the Aichelburg-Sexl metric [23], which in our case of codimension 6 is given by

$$h_{--} \sim \frac{l_s^8 g_s E_p}{r_\perp^3} \delta(x^-) \tag{30.1409}$$

Here $E_p = E(R/r)^4$ is the proper energy density of the probe (related to the Poincare observer's energy density E by the appropriate powers of the warp factor r/R), $r_\perp$ is the distance from the probe in the transverse dimensions, and x^- is the light cone coordinate with respect to which the brane trajectory is localized in locally flat coordinates. If we insist that the corresponding curvature $\mathcal{R}_{10} \sim h_{--}/r_\perp^2$ smeared over a string scale distance across the brane be smaller than the ambient AdS curvature $1/R^2$ at a distance R from the probe, this gives the condition

$$E < \frac{\phi^4}{\lambda^{1/4} g_s} \tag{30.1410}$$

If we impose this condition, we still obtain a window

$$\lambda^{1/4} g_s E < \phi^4 < \lambda E$$

in which our back reaction constraint (30.1410) intersects with our regime (30.1407) of limiting speed.

[184] This constraint was obtained via discussions with M. Fabinger.

This constraint (30.1410) may be too strong, since the novel behavior we found for ϕ based on the DBI action depends only on the AdS geometry. We can estimate the back reaction of our probe on the AdS geometry as follows. The probe forms a domain wall of energy density E_p in the five dimensional gravity theory obtained by dimensional reduction on the S^5. Such a wall will jump the warp factor accross it. By dimensional analysis, this gives the relation

$$\frac{1}{R'} - \frac{1}{R} \sim E_p l_5^3 \tag{30.1411}$$

in terms of the five dimensional Planck length l_5, where R' is the curvature radius on the IR side of the domain wall. Using $l_5^3 \sim l_{10}^8/R^5 = l_s^8 g_s^2/R^5$, the condition that the jump in warp factor is smaller than the original AdS warp factor is

$$E < \frac{\phi^4}{g_s}$$

Again, this leads to a window at strong coupling in which the rolling scalar field saturates its speed limit (30.1408) in the regime

$$g_s E < \phi^4 < \lambda E \tag{30.1412}$$

Similar back reaction criteria apply to all the cases discussed in the paper, and we will derive a related bound in our interesting inflationary phase. It would be interesting to explore what happens when this condition is violated. We will discuss the other consistency conditions for our background (including ruling out significant back reaction from particle production) in Section 6.

30.3.3 Antibranes

Here we introduce the first of our generalizations: anti-D3-branes moving in the AdS background (30.1400). The action for the $\overline{D3}$-brane probe in AdS_5 differs from the $D3$-brane case (30.1405) merely by a change of sign in the final term,

$$S = -\frac{N}{\lambda^2} \int d^4x \phi^4 \left(\sqrt{1 - \lambda \dot{\phi}^2/\phi^4} + 1 \right) \tag{30.1413}$$

Upon expanding the square-root in powers of $\lambda\dot{\phi}^2/\phi^4$, this gives rise to a potential which is quartic in ϕ. There is no quadratic $m^2\phi^2$ term. This may be at first sight surprising since the system has broken supersymmetry in the presence of the antibrane, and loops of open strings which probe the susy breaking are included at the level of the action (30.1413). However this result is to be expected using simple Randall-Sundrum ideas. The local string scale at the position of the brane is $\phi/\lambda^{1/4}$ (30.1401). Similarly all hard masses are warped down by a factor of $r/R = \phi\sqrt{\alpha'}/\lambda^{1/4}$ at position ϕ in the throat. Applying this to the susy breaking scale m_{SUSY}, a potential $V \sim m_{SUSY}^2\phi^2$ really scales like ϕ^4. Said differently, there is no independent scale in the CFT on the Coulomb branch beyond the position ϕ.

The ϕ^4 potential implies that the $\overline{D3}$-brane experiences a force towards the horizon at $r = 0$. Nonetheless, it is a simple matter to check that the late time dynamics is dominated by the speed limit and is identical to that of the $D3$-brane (30.1408).

In some sense, the $\overline{D3}$-brane probe in AdS space can be thought of as a strong coupling limit of the usual tachyon matter system. Indeed, we shall see in Section 5 that the resulting dynamics bears some similarity to the results that have been obtained for the much-studied open string tachyon decay in the weakly coupled brane-antibrane system (see [24] for recent review of this system). However, there are some differences which we stress here. At weak coupling, the $\overline{D3}$-brane system has tachyons from strings stretching between the antibrane and the $N-1$ D3-branes described by the AdS throat. However, at strong coupling $\lambda \gg 1$, it is easy to see from (30.1397) and (30.1401) that this mode is not tachyonic. The negative contribution to the would-be tachyon mass squared, scaling like the (warped) string scale $m_s(\langle\phi\rangle)^2$, is dwarfed by the positive contribution m_W^2 to the mass squared from the stretching of the string from the antibrane to the AdS horizon. Thus the brane-antibrane system at strong 't Hooft coupling (obtained by taking many branes and a single antibrane) has no tachyon. The system *is* unstable – as we will discuss in detail the antibrane falls towards the horizon. However this evolution is not condensation of a brane-antibrane tachyon since none exists; as in [25] the annihilation of the antibrane against the unit of $D3$-brane charge encoded in the RR flux is a nonperturbative effect.

30.4 Coupling to Gravity

In both the brane and antibrane cases, we will be interested in two further important generalizations. One is coupling to four dimensional gravity, as well as to other sectors suppressed by higher dimension operators that may arise in the corresponding string compactifications. The other is a generalization to a throat which is capped off in the region corresponding to the IR behavior of the field theory – i.e. the system with a mass gap. In a string theoretic setup, these effects can be achieved by gluing an AdS-throat like solution onto a compactification geometry which acts as a UV-brane as in [26, 27].

In this section we shall discuss the effective actions which result from these generalizations, focusing on the relative strength of corrections to the potential energy and DBI kinetic energy terms in the effective action. Our motivation for considering these theories is the application of the speed limit mechanism to cosmology. We will postpone a full discussion of this until Section 5. To avoid undue suspense, let us here summarize some of the important results.

Because our matter sector itself slows the scalar field ϕ in its progress toward the origin (in the field theory variables), the conditions for slow roll inflation are modified in a useful way. In a situation with a strong enough additional potential energy for ϕ (which may arise for example from a setup with a brane or antibrane in a cutoff AdS throat coupled to other sectors) we will indeed find a phase of slow roll inflation. This can be obtained at subPlanckian VEV for ϕ, with the slow roll provided by the DBI kinetic term corrections rather than from the usual gravitational damping or from an unnaturally flat potential. This allows us to evade some of the problems discussed in [2]. While we have not yet produced a complete model of real-world inflation from this mechanism, it seems a very promising ingredient.

More generally, we will find some familiar behaviors for the FRW scale factor $a(t)$ but arising in unusual ways from our matter sector. For example, as in "tachyon matter" we will find a dust equation of state for motion of a D3-brane toward the origin of the Coulomb branch, though in our case production of massive matter is suppressed. With other potentials, we also find a novel steady state late-time behaviour for the universe that does not involve fine tuning of initial conditions.

The motion of the moduli themselves is an important aspect of string cosmology, and the basic result of the previous sections leads to interesting novelties in this area. In particular, if the moduli field ϕ corresponding to the motion of the threebrane heads toward the origin of the Coulomb branch, it gets slowed down and, at least in the probe limit, takes an infinite time to reach the origin. This is radically different from the naive treatment based on the moduli space metric in which it would shoot past the origin without pausing. Furthermore, if the scalar relaxes into a potential well surrounding the origin, it does not oscillate around the minimum in the same manner as moduli treated with the usual action (30.1396). However in the cases when the ϕ sector with all the DBI corrections leads to a dust equation of state, it can cause similar problems to the usual moduli problem.

30.4.1 CFT Coupled to Gravity

We firstly ask what becomes of the low-energy effective action (30.1403) when the gauge theory is coupled to gravity by introducing a dynamical background metric $g_{\mu\nu}$. At the same time, we add a potential term $V(\phi)$ to the action that may arise when the system is coupled to four dimensional gravity and other sectors involved in a full string compactification. A simple four dimensional covariantization of (30.1403) is

$$\mathcal{L}_0 = -\frac{1}{g_{YM}^2}\sqrt{-g}\left(f(\phi)^{-1}\sqrt{1+f(\phi)\,g^{\mu\nu}\partial_\mu\phi\,\partial_\nu\phi}+V(\phi)\mp f(\phi)^{-1}\right) \tag{30.1414}$$

where the $\pm$ refers to the $D3$-brane and $\overline{D3}$-brane respectively while $f(\phi)=\lambda/\phi^4$ as before.

In the presence of four dimensional gravity, there are in general further corrections to the action (30.1414) coming from the following considerations. Dynamical $4d$ gravity leads generically to nontrivial four dimensional curvature $\mathcal{R}$. In effective field theory, this leads to series of contributions of the form

$$\mathcal{R}\phi^2(1+c_1\mathcal{R}/\phi^2+\dots). \tag{30.1415}$$

More generally each contribution in the action (30.1414) could be corrected by terms suppressed by powers of $\mathcal{R}/\phi^2$. In particular, there will be corrections depending on $\mathcal{R}$ and $\partial\phi$, which will be outside the scope of our analysis; our results will be contingent on these corrections being subdominant (which may require tuning in appropriate circumstances). In the weakly coupled $\mathcal{N}=4$ SYM theory coupled to four dimensional gravity, such corrections are evident from diagrams containing loops of W bosons. At strong coupling, such corrections were found in [28, 29] starting from the DBI action for curved slices. From all these points of view, the coefficients c_i in (30.1415) are expected to be of order one (i.e. not parametrically large or small as a function of λ).

In our solutions, we will find that these curvature corrections are negligible. Our strategy will be to simply ignore them at this stage, analyze the dynamics, and then check that this is self consistent (in Section 5.3 below). In any case, we can collect these effects in an effective Lagrangian containing both gravity and matter:

$$S=\int\left(\tfrac{1}{2}\sqrt{-g}\,(M_p^2+\phi^2)\,\mathcal{R}+\mathcal{L}_0+\dots\right) \tag{30.1416}$$

where the $\dots$ refers to corrections to the terms in (30.1416) which are down by powers of $\mathcal{R}/\phi^2$ from the leading terms. The final action is a special case of those considered in k-inflation and k-essence scenarios [3, 4]. In our case, the higher derivative terms in $\mathcal{L}_0$ follow from the strong coupling dynamics of the field theory near the origin of its moduli space.

Let us now discuss the reliability and plausibility of this action including the coupling to four dimensional gravity. In string theory, such a coupling is obtained by embedding the AdS geometry into a compactification as a throat emanating from a Calabi Yau compactification of type IIB string theory [26, 27]. The first question that arises is whether this geometrical combination exists or if instead the coupling of the AdS to the Calabi Yau produces a large deformation of the AdS space. If such a geometry exists, we may move onto the second question: whether the low energy effective field theory action is given by (30.1414) to a good approximation for the system we wish to study, and what is the order of magnitude of the parameters in the potential V. Thirdly, in the end we must check that the time dependent solutions we find are stable against small fluctuations, particle production, and density perturbations. (This is in addition to the checks for small acceleration and small back reaction of the probe energy of the sort we completed in Section 3 for the global case.) This last question will be addressed in section 6.

As far as the first question goes, as discussed in [26, 27], the AdS throat is obtained from a collection of many D3-branes at a smooth point on the Calabi Yau. At the level of the no scale potential obtained for CY compactifications of type IIB at large radius in the absence of α' corrections and nonperturbative effects [27], the scalars on the D3-branes have a potential identical to that in the global quantum field theory on their worldvolume. They are mutually BPS and, in this sense, are also mutually BPS with the rest of the compactification. This follows from the analysis in [27], and may be related to the fact that in the $\mathcal{N} = 4$ SYM at strong coupling, the operator $tr\Phi^2$, which is relevant at weak coupling, is highly irrelevant at strong coupling (something exploited by e.g. [32]). The quantum mechanical corrections to the Kahler potential and the superpotential of the model, which are important for fixing the Kahler moduli, will in general also determine the positions of the D3-branes. Since the configuration corresponding to their being on top of each other at a smooth point in the Calabi Yau is an enhanced symmetry point, it is likely that an order one fraction of the models have a minimum containing to a good approximation an AdS throat arising from such a collection of D3-branes. We will momentarily estimate the size of corrections to this geometry as seen by our probe.

Now let us discuss the second question about the reliability of our effective action and the parameters in the potential. Attached to the Calabi Yau, our AdS throat is coupled to four dimensional gravity and other sectors. We will start by discussing the effects of the coupling of these sectors to ϕ in effective field theory. Then we will discuss the geometrical description of these effects. We will argue that (consistently in both descriptions) significant corrections to the mass term in the potential can naturally arise from these couplings while corrections to the kinetic terms arising from them are subleading to those in the original action. This ensures that we can preserve our slow roll effect for the scalar field coming from the kinetic terms while introducing potential energy sources for gravity which lead to interesting, and in one case accelerating, cosmologies. As we will also discuss, it is an interesting open question whether this setup requires tuning couplings of throat Kaluza Klein modes to other sectors to avoid further effects on the kinetic terms for ϕ.

Effective Field Theory Description of Corrections

As well as curvature couplings such as $\mathcal{R}\phi^2$ arising from the introduction of gravity, we will generically have further couplings of ϕ to other sectors. These could be of the form,

$$\phi^2\eta^2 \quad \text{or} \quad \frac{\phi^2(\partial\eta)^2}{M_*^2} \tag{30.1417}$$

where η is a field from another sector and M_* is a mass scale (such as the GUT, string, or Planck scale) in the system above the energies we wish to consider. In the next subsection we will explain why we expect such couplings to be available naturally in string compactifications, but

we have not constructed an explicit example and will largely treat these couplings by effective field theory in our analysis of the cosmology.

We should emphasize that the effective field theory description we use in this subsection makes use of the weak coupling expansion available on the gravity side of AdS/CFT; it is just the couplings of the ϕ fields to the new sectors involved in the compactification which we will treat here via effective field theory. A full treatment of the 4d effective field theory on the gravity side also requires understanding the couplings and effects of the Kaluza Klein modes. This involves understanding what tuning is required to preserve an (approximate) AdS throat in a Calabi Yau compactification in the absence of a probe brane. This is not well understood. In any case, as we will discuss in the next subsection and in Section 5, the probe approximation for the brane in a large-radius approximate AdS throat is consistent with the presence of a large enough mass on the probe brane for interesting cosmological effects such as inflation.

More generally, we may view (30.1417) as coupling the strongly coupled CFT to the η sector by hand in effective field theory. The action (30.1414) contains no closed string (gravity) loops, or loops of η from couplings such as (30.1417). We must estimate contributions to our effective actions from these loop effects. We will assume a weak enough string coupling for perturbative string theory to be valid on the gravity side.

Out first important task is to determine the scale of any relevant or marginal corrections to the action that may arise from the couplings involving ϕ. Let us start by discussing the potential $V(\phi)$, considering a power series expansion

$$V \equiv V_0 + V_2\phi^2 + V_4\phi^4 + \dots \tag{30.1418}$$

The most relevant term in the Lagrangian at low energies is the hard cosmological constant V_0. This gets contributions in principle from all sectors in the system. In effective field theory (and approximately in string theory using the Bousso Polchinski mechanism [30]) we may tune this (close) to the value of interest for a given application. We will make use of this freedom in our analysis.

Now let us move on to the ϕ-dependent couplings, starting with the mass term $V_2\phi^2$. As we discussed in the global case, in an exact conformal field theory (which is either not coupled to gravity and other sectors or is not destabilized by their presence), a mass coupling $m_*^2\phi^2$ with m_* a constant mass parameter is ruled out by the conformal invariance encoded in the effects of the warping in the AdS throat. Thus, with an exact AdS throat geometry, the potential is expected to receive corrections starting with the quartic $V_4\phi^4$ coupling. However, in general the couplings to gravity and other sectors can generate corrections to the theory (30.1416) which violate conformal invariance. For example, corrections to the scalar mass coming from loops containing virtual gravitons yield a contribution,

$$m_*^2 \sim \Lambda_{UV}^4/M_p^2 \tag{30.1419}$$

where the fourth power of the UV loop momentum cutoff Λ_{UV} can be traced to the fact that the Feynman diagrams contain derivative couplings. Similar contributions arise from couplings to other sectors (30.1417). A mass for ϕ can be generated either from a VEV for η, or from loop corrections involving the derivative coupling, giving rise to respective contributions of the form

$$m_*^2 \sim \langle\eta\rangle^2 \quad \text{or} \quad m_*^2 \sim \Lambda_{UV}^4/M_*^2 \tag{30.1420}$$

We should also note that the quartic and higher couplings $V_4\phi^4 + \dots$ in the effective potential will also get corrections from interactions such as (30.1417). In our strategy for obtaining inflation below, we will require the net quartic and higher couplings to be small enough relative to the quadratic coupling. This will not require significant tuning of parameters in our solutions.

Now let us consider the strength of corrections to the crucial generalized kinetic terms in the DBI action (i.e. the series in $v_p^2 = \lambda\dot{\phi}^2/\phi^4$) appearing in the action. Importantly, the corrections to these terms coming from (30.1417) are suppressed relative to the existing terms that generate our slow roll effect (in contrast to the situation for the soft mass term just described, for which the (30.1420) constitutes the leading effect.) For example, if we compute the effect of the coupling (30.1417) on the $\dot{\phi}^4$ term in the effective action via a loop of ηs, we obtain a contribution scaling like

$$\left(\frac{\dot{\phi}}{M_*}\right)^4 \log(\Lambda_{UV}/M_*)$$

This is much smaller than the original contribution $N\dot{\phi}^4/\phi^4$ in the small ϕ regime of interest: $\phi^4 \ll NM_*^4$.

The fact that the DBI kinetic terms are robust against large corrections from (30.1417) arises from the fact that they are marginal and get logarithmic corrections suppressed by the coupling constants in terms like (30.1417). There are no soft contributions to the derivative terms in the action–they would have to be of the form $(m_*/\phi)^n(\dot{\phi}^2/\phi^4)$. Such corrections would be more infrared divergent as $\phi \to 0$ than the existing ones in (30.1403). There are no perturbative diagrams producing such effects, and we find it implausible also in the strong coupling regime that coupling to gravity could worsen the IR behavior of the system.

Although these corrections to the DBI kinetic terms are small, they are nonzero in general. Since the kinetic terms arose from the DBI action given by the volume of the brane embedded in the ambient geometry, corrections to them correspond to corrections to the AdS geometry at least as seen by our probe brane. Indeed once we generate (or add by hand in effective field theory) a deformation of the theory by a mass term for ϕ, we expect deviations from exact AdS geometry in the IR. However, as above these deviations are small effects on our probe evolution relative to the effects of the original kinetic terms. In particular, if we deform the Lagrangian to introduce an $m^2\phi^2$ term, we can estimate its effects on the higher derivative terms in the DBI action. These came from integrating out "W bosons", with the importance of the higher derivative terms arising from the fact that these modes become light at the origin $\phi \to 0$. The mass term for ϕ will at some order induce new contributions to the mass of the W bosons; up to logarithms these will at most scale like

$$m_W^2 \to \phi^2 + g_{YM}^2 m^2 \tag{30.1421}$$

In this manner, the contributions of virtual W bosons to the DBI Lagrangian for ϕ become a series in $\lambda\dot{\phi}^2/(\phi^2 + g_{YM}^2 m^2)^2$. This means that for $\phi \gg g_{YM}^2 m$, the evolution is well approximated by that given by the original DBI Lagrangian. As we will see, this constraint preserves a window of interesting behavior for ϕ dictated by,

$$\phi^2 > g_s m^2 \tag{30.1422}$$

In Section 5.1, we shall see that the same constraint arises on the gravity side by requiring small back reaction so that the probe approximation holds.

In summary, the same couplings which give rise to a mass m for ϕ do not badly alter the generalised kinetic terms for $\phi \gg m$. In fact, the mass term itself is a small effect on the overall probe evolution as we will see from the ϕ equation of motion in which the DBI kinetic terms dominate. However, the mass term in the potential will be an important source for the four dimensional spacetime geometry in our cosmological solutions. The corrections on the kinetic terms for $\phi < m$ do mean that ultimately the geometry is probably better approximated by a cut off throat. We will turn to this analysis in subsection 4.2.

As emphasized above, the four dimensional effective field theory however contains more than just ϕ and the BPS W bosons, but also Kaluza Klein modes in the AdS. It is important to check whether couplings similar to (30.1417) and curvature couplings appear with ϕ replaced by Kaluza Klein modes. If so, and if the induced masses m_{KK} of these Kaluza Klein modes were as big as m, the corresponding deformation of the geometry in the IR region would remove our inflationary solution to be described in section 5. In the context of effective field theory, we may tune away these couplings if necessary, and as discussed in the next section we generically expect such tuning to be possible in string theory. It would be interesting to determine whether such tuning is necessary or if $m_{KK} \ll m$ appears naturally.

Gravity Description of Corrections

In the previous subsection, we gave arguments based on simple effective field theory couplings that the (soft) mass term for ϕ generated by (30.1417) will generically be affected more strongly (relative to the original mass term for ϕ in the Lagrangian) than the generalized kinetic terms (relative to the original generalized kinetic terms in the Lagrangian). At first sight, this may appear surprising from the point of view of the geometrical picture of a brane probe. Such a probe has kinetic terms determined by the metric of the ambient spacetime, and potential terms introduced via couplings to other background fields (such as the five form RR field strength in the $AdS_5 \times S^5$ solution of type IIB supergravity). There is no general relation between these two effects, though in particular examples they are related in particular ways via the coupled equations of motion for the metric curvature and the other low energy fields. In familiar examples, such as the $AdS_5 \times S^5$ solution, or the Klebanov Strassler solution [31], the strength of the potential is smaller than the effects we found in the previous subsection.

These classical solutions however do not include the effects introduced by the other sectors (mocked up by η in the above analysis) located in the bulk of a Calabi Yau to which the throat attaches. In general, the relation between the spacetime metric and the other background fields is different from that obtained in the familiar classical solutions. More general flux backgrounds (including AdS solutions) with different types of brane probes will have a different balance between fluxes felt by the probe and metric curvature due to the contribution of other ingredients. Our results in the above subsection reflect an aspect of this in the context of a coupling of ϕ in the familiar solutions to other sectors via, for example, an embedding in a Calabi Yau.

To make this more concrete, let us use known scales of couplings between different sectors of a Calabi Yau geometry to estimate the effective field theory parameters M_* and Λ_{UV} in the effective field theory analysis of the last subsection. In [33] couplings between brane throats were computed. The results were consistent with couplings between a brane throat and another sector (which could be a much smaller throat) suppressed by powers of $M_* \sim 1/R$. Taking the UV cutoff Λ_{UV} of the second sector to also be of order $1/R$, one obtains an estimate for the mass m of order $m \sim 1/l_s\lambda^{1/4}$. As discussed in the previous section, the same coupling yields

a suppressed correction to the kinetic terms if we lie within the window

$$\frac{1}{l_s\lambda^{1/4}} < \phi < \frac{N^{1/4}}{l_s\lambda^{1/4}} \tag{30.1423}$$

To summarise: the important corrections to the mass term for ϕ we derived via effective field theory couplings in the last subsection fit with known properties of geometrical embedding of brane probes, taking into account effects generated by sectors outside the brane throat corresponding to our original CFT. Therefore we find it likely that such a mass term is available (and potentially generic given the presence of a large brane throat) in explicit string models. In any case, a mass for ϕ is physically consistent with a probe approximation for ϕ. This follows from the above considerations regarding the many independent ways of assembling ingredients such as the target space metric and other background fields, combined with the self consistency against back reaction (to be checked in section 5) of the brane energy carried in the ϕ mass term.

However, we have not constructed an explicit example of this in a full compactification model. It is a subtle problem to determine the couplings among sectors in a full string compactification [33][2], and we plan to pursue it systematically in future work.

30.4.2 QFT with Mass Gap Coupled to Gravity

Motivated by the discussion above, we here consider geometries dual to theories with a mass gap, where the AdS throat is cut off at the IR end. The prototypical example of such a geometry is the Klebanov-Strassler solution of [31]. The effects of such an IR cut-off include the following. The warp factor $f(\phi)$ in (30.1404) is changed, and further corrections to the action may arise if we start from a cutoff throat at the classical level. Here we shall consider only a toy model in which we focus on the effect on $f(\phi)$; similar comments to those in the previous subsection apply to the question of further corrections to the action.

For the purposes of our discussion, we replace the full Klebanov-Strassler solution with a simple toy model which reproduces the relevant features. We motivate this by considering a theory in which the formula for the W boson mass (30.1397) is deformed to

$$m_W^2 = \phi^2 + \mu^2 \tag{30.1424}$$

where, in the effective field theory discussion above, we had $\mu^2 \sim g_{YM}^2 m^2$. In general μ will be related to m in a similar fashion, but for the purposes of this section we keep it arbitrary. To model the W mass (30.1424) from the holographic perspective, we keep the action (30.1414), but replace the warp factor f with the appropriate function. For simplicity we shall consider,

$$f(\phi) = \frac{\lambda}{(\phi^2 + \mu^2)^2} \tag{30.1425}$$

Such a geometry does not satisfy Einstein's equations by itself, but nonetheless exhibits the important features of a capped off throat.

In the DBI Lagrangian (30.1414) for the cutoff throat (30.1425) in the antibrane case, there is a hard mass of order

$$m_{DBI}^2 \sim \mu^2/\lambda \tag{30.1426}$$

Loop corrections arising from the ϕ^4 interactions drive this up to $m_{DBI}^2 \sim \Lambda_\phi^2/\lambda$ where Λ_ϕ is the ultra-violet cut-off pertaining to the ϕ loop.

In our cutoff geometry we also have corrections to the masses and other couplings as discussed in the last subsection (where now there may be some new sectors associated with the physics at scale μ). Again the effects on the kinetic terms are subleading but the mass corrections are important.

30.5 Cosmological Solutions

In Section 3, we saw that the scalar field motion in our system is radically different from the naive expectation based on the finite distance to the origin of the moduli space in the metric appearing in the two derivative action (30.1396). Here we shall ask what effects this novel matter sector has on cosmological solutions. We will investigate this in detail by studying the possible FRW cosmologies solutions which follow from (30.1414). We consider only spatially flat cosmologies,

$$ds^2 = -dt^2 + a(t)^2 dx^2$$

Since spatially inhomogeneous terms are redshifted away during inflation, we consider the scalar field ansatz $\phi = \phi(t)$. With this ansatz, the equations of motion can be concisely expressed by first defining the analog of the Lorentz contraction factor in special relativity,

$$\gamma \equiv \frac{1}{\sqrt{1 - f(\phi)\dot{\phi}^2}} \tag{30.1427}$$

Our stategy is to analyze the solutions following from (30.1414) by first ignoring the effects of the curvature coupling $\mathcal{R}\phi^2$, and later showing that its effects are self consistently negligible on our solutions (unlike the situation in usual inflationary models based on (30.1396)). The energy density ρ and pressure p following from (30.1414) (taking the + sign, corresponding to a D3-brane, for definiteness) are given by,

$$\rho = \frac{\gamma}{f} + (V - f^{-1}) \tag{30.1428}$$

$$p = -\frac{1}{f\gamma} - (V - f^{-1}) \tag{30.1429}$$

These definitions do not include the overall coefficient of $1/g_{YM}^2$ in (30.1414) which instead combines with M_p in the Einstein equations so that the scale $(M_p\sqrt{g_s}) = M_p g_{YM}$ appears in all equations. The Friedmann equations read,

$$3H^2 = \frac{1}{g_s M_p^2}\,\rho \tag{30.1430}$$

$$2\frac{\ddot{a}}{a} + H^2 = -\frac{1}{g_s M_p^2}\,p \tag{30.1431}$$

where $H = \dot{a}/a$ is the usual Hubble parameter. Finally, the equation of motion for ϕ reads,

$$\ddot{\phi} + \frac{3f'}{2f}\dot{\phi}^2 - \frac{f'}{f^2} + \frac{3H}{\gamma^2}\dot{\phi} + \left(V' + \frac{f'}{f^2}\right)\frac{1}{\gamma^3} = 0 \tag{30.1432}$$

where prime denotes a derivative with respect to ϕ. Note that at as we approach the speed limit on moduli space, so $\gamma \to \infty$, both the friction term and the potential term in the equation of motion for ϕ become subdominant. It is a simple matter to check that the second Friedmann equation (30.1431) follows from the first (30.1430) together with the equation of motion (30.1432). The effective equation of state of our system, defined by $p = \omega\rho$, is therefore given by

$$\omega = \frac{-\gamma^{-2} - (Vf - 1)\gamma^{-1}}{1 + (Vf - 1)\gamma^{-1}} \tag{30.1433}$$

The Hamilton-Jacobi Formalism

As occurred in the global case, we will find solutions in which the brane asymptotes to the speed of light at late times in the gravity side background. This means that the quantity γ grows as t gets large, and the behavior of ρ will be substantially different from the usual case of $\rho = \frac{1}{2}\dot{\phi}^2 + V$ to which it reduces for small proper velocity.

Firstly, we want to re-write the Friedmann equations (30.1430) and (30.1432) in a more tractable form. In fact, they can be integrated once. In the inflation literature this is referred to as the "Hamilton-Jacobi" formalism [34]. Another perspective can be obtained by viewing the resulting cosmology as a Wick rotation of a BPS domain wall; the first order Friedmann equations are related to the Bogomoln'yi equations derived in [35].

To derive the Hamilton-Jacobi equations the important step is to view the scalar field ϕ as the time variable. In practice, this means we consider $H = H(\phi)$ with $\phi = \phi(t)$. This immediately puts a limitation on the dynamics since it assumes ϕ is monotonic and H is a single valued function of ϕ and therefore does not allow for oscillatory behaviour in ϕ. We start by taking the time derivative of (30.1430) and, using (30.1432) along the way, we find

$$6HH'\dot{\phi} = -\frac{1}{g_s M_p^2}\, 3H\gamma\dot{\phi}^2$$

which can clearly be solved simply by

$$\dot{\phi} = -2(g_s M_p^2)\,\frac{H'}{\gamma}$$

Since (γ^{-1}) depends on $\dot{\phi}$, it's useful to invert this to give us

$$\dot{\phi} = \frac{-2H'}{\sqrt{1/(g_s^2 M_p^4) + 4fH'^2}} \tag{30.1434}$$

Finally, we can substitute this expression for $\dot{\phi}$ into (30.1430), using the full expression for ρ from (30.1428) to get an expression for the potential $V(\phi)$ in terms of the Hubble parameter $H(\phi)$,

$$V = 3(g_s M_p^2)\,H^2 - \frac{(g_s M_p^2)}{f}\sqrt{1/(g_s^2 M_p^4) + 4fH'^2} + \frac{1}{f} \tag{30.1435}$$

The advantage of these equations is that they can be solved sequentially. Given a potential V, we solve (30.1435) to find $H(\phi)$. This can then be plugged into (30.1434) to find $\phi(t)$. Finally, this can be substituted back into $H(\phi)$ to find the dynamics of the universe. We shall now solve for the late time behaviour of these equations, both in the AdS throat where $f = \lambda/\phi^4$, and in our toy model for the cut-off geometry with $f = \lambda/(\phi^2 + \mu^2)^2$.

30.5.1 Cosmology in the AdS Throat

We start by considering cosmology in the CFT, corresponding to a brane moving in the full AdS throat. In this case, the first order Friedmann equations (30.1434) and (30.1435) read

$$\dot{\phi} = \frac{-2H'}{\sqrt{1/(g_s^2 M_p^4) + 4\lambda H'^2/\phi^4}} \tag{30.1436}$$

and

$$\lambda V = 3\lambda(g_s M_p^2)\, H^2 - (g_s M_p^2)\phi^4\sqrt{1/(g_s^2 M_p^4) + 4\lambda H'^2/\phi^4} + \phi^4 \tag{30.1437}$$

We will consider potentials of the form

$$V = V_0 + V_2\phi^2 + V_4\phi^4 + \dots \tag{30.1438}$$

and determine the behaviour of the scale factor as $\phi \to 0$. For now, we will simply introduce these parameters by hand to analyze the equations but, as mentioned above, there are constraints on these coefficients in the $\mathcal{N} = 4$ SYM and in this theory coupled to gravity which we should include in the final analysis. For now however, we work with the ad hoc potential above. There there are three possible cases, depending on which of V_0, V_2 or V_4 is the first non-vanishing coefficient. (We shall in fact see that if the first two vanish, and the potential is positive, then the result is universal).

Case A: $V_0 \neq 0$

If V is simply a constant then equation (30.1437) is solved by $H = \sqrt{V_0/3(g_s M_p^2)}$, in which case equation (30.1436) shows that $\phi = \text{const}$, and we get a standard de Sitter phase. Of course, our kinetic terms aren't playing a role here since ϕ isn't moving. Note that equation (30.1436) requires $V_0 > 0$ here.

Suppose instead that we take $V = V_0 + V_4\,\phi^4$ which suffices to get ϕ moving. Then for small ϕ, we may choose the simple ansatz $H = h_0 + h_4\phi^4 + \dots$ where, from (30.1437) we must choose the coefficients $h_0 = \sqrt{V_0/3(g_s M_p^2)}$ and $h_4 = V_4/6(g_s M_p^2)\, h_0$. Substituting this into the first of the Friedmann equations, we have the late time behaviour for ϕ,

$$\phi \to \frac{1}{4(M_p\sqrt{g_s})\sqrt{h_4}}\,\frac{1}{\sqrt{t}}$$

We see that, as in the case of the global CFT considered in Section 3, the scalar field takes infinite time to reach the origin. However, in this case the friction from the expanding universe means that ϕ does not saturate the speed limit (30.1402). From the late time behaviour of ϕ, we may immediately determine the late time behaviour of the scale factor. We have,

$$H = \sqrt{\frac{V_0}{3(g_s M_p^2)}}\left(1 + \frac{3}{2^7 V_4 (g_s M_p^2)}\,\frac{1}{t^2}\right)$$

This result highlights the peculiar cosmologies that can arise due to strong coupling effects of the gauge theory. Here we find a standard de Sitter phase, superposed with a cosmology in which the scale factor has the form $a(t) \sim exp(-c/tM_p)$. We shall see more cosmologies with this characteristic exponential behaviour in Case D.

Case B: $V_0 = 0$

Let us now consider the interesting case that will give rise to inflation. We suppose that $V_2 \neq 0$ so that this quadratic mass term dominates at late times when ϕ is small.

We work with the simple ansatz,

$$H = h_1\phi + \dots \tag{30.1439}$$

Substituting this into (30.1437), we get contributions to V_2 from both the H^2 term as well the $\sqrt{\cdots}$ term. We find the potential takes the form,

$$V = \left(3h_1^2 - \frac{2h_1}{\sqrt{\lambda}}\right)(g_s M_p^2)\,\phi^2 + \mathcal{O}(\phi^4) \equiv V_2\phi^2 + \mathcal{O}(\phi^4) \tag{30.1440}$$

which, solving for h_1 in terms of V_2 gives

$$h_1 = \frac{1}{3\sqrt{\lambda}}\left(1 + \sqrt{1 + 3V_2\lambda/(g_s M_p^2)}\right) \tag{30.1441}$$

Substituting our ansatz for H into the equation (30.1436), we find

$$\dot{\phi} = \frac{-2h_1\phi^2}{\sqrt{\phi^4/(g_s^2 M_p^4) + 4\lambda h_1^2}}$$

As $\phi \to 0$, we simply throw away the ϕ^4 term in the denominator to find the late time behaviour

$$\phi \to \frac{\sqrt{\lambda}}{t} + \dots$$

This coincides with the result (30.1408) from the global analysis of Section 3; in this theory the slowing down of the scalar field is driven by the speed limit rather than friction from the expanding spacetime. We now substitute this back into our ansatz (30.1439) to find the late time behaviour of the scale factor,

$$a(t) \to a_0\, t^{h_1\sqrt{\lambda}} \tag{30.1442}$$

Using (30.1441), the exponent is

$$h_1\sqrt{\lambda} = \frac{1}{3}\left(1 + \sqrt{1 + \frac{3V_2\lambda}{(g_s M_p^2)}}\right) \approx \sqrt{\frac{V_2\lambda}{3(g_s M_p^2)}} \tag{30.1443}$$

where the approximation is true only for $V_2\lambda \gg g_s M_p^2$. We see that we can orchestrate an inflationary universe if the potential V_2 is sufficiently large. Specifically

$$\frac{V_2\lambda}{g_s M_p^2} > 1 \tag{30.1444}$$

It is amusing to recall that usual slow roll inflation occurs only if the potential is suitably flat. Here we find the opposite result: with a speed limit on the scalar field, inflation occurs only if V_2 corresponds to a high enough mass scale (we see from (30.1444) that this mass scale can be much lower than M_p, but the acceleration gets stronger for larger V_2).

Interestingly, we have obtained a form of power law inflation, since the scale factor behaves as a power of t. However it arises in the case of a polynomial (quadratic) potential, which under the usual circumstances (30.1396) leads to exponential inflation $a \sim e^{Ht}$. (The potential required to obtain power law inflation from (30.1396) is exponential in ϕ.)

Naively it appears that this result suggests an alternative to the usual quintessence scenario where a scalar field rolls towards an asymptotic regime of moduli space. Here the scalar field could roll a finite distance in the field theory moduli space metric in the interior of the moduli space and still lead to asymptotic acceleration. However, we are limited to a finite period of controlled speed of light evolution because of the back reaction of the probe on the geometry. In the global case we determined this backreaction in (30.1412). It is a simple matter to repeat the calculation here where the energy of the brane is dominated by the potential $V_2\phi^2$. Once again, translating from the proper energy E_p to the observers energy density $E = E_p(r/R)^4$, we have the constraint

$$E_p l_5^3 \sim \frac{V_2\phi^2}{g_{YM}^2}\left(\frac{R}{r}\right)^4 \frac{\alpha'^4 g_s^2}{R^5} < \frac{1}{R}$$

which gives us the constraint that our solution is only to be trusted for ϕ greater than

$$\phi^2 > g_s V_2$$

Notice that this is identical to the lower bound on ϕ arising derived in Section 4 by calculating one-loop corrections to the W-boson mass using effective field theory methods (30.1422). Despite this infra-red cut-off to our inflationary phase, we shall shortly argue that we still retain a viable window of inflation in this scenario.

Case C: $V_0 = 0$ **and** $V_2 = 0$

Let us now suppose that the potential has only $V_4 \neq 0$. As we have seen, such a potential is generated by a $\overline{D3}$-brane moving in the AdS throat. The late-time analysis of the Hamilton-Jacobi equations (30.1436) and (30.1437) is actually just a special case of the above discussion, where we put $h_1 = 2/3\sqrt{\lambda}$. We can read of the final result from (30.1442): we have

$$a \to a_0\, t^{2/3}$$

which looks like dust. In fact, for any potential for which $V_0 = V_2 = 0$ with higher order terms non-vanishing, the speed limit on the moduli space ensures that the rolling scalar field looks has equation of state $\omega \approx 0$: a kinetic dust.

Case D: $V < 0$

To find equations of state with $\omega > 0$, we require negative V_2. From (30.1441), we see that for $0 < h_1 < 2/3\sqrt{\lambda}$ we have

$$-\lambda(g_s M^2)^2\,\frac{\lambda}{3} < V_2 < 0$$

The late time behaviour of the scale factor is once again given by (30.1442). During this expansion, the scalar field is rolling up the inverted harmonic oscillator potential. Without the higher kinetic terms, the scalar would overshoot the top and roll down the other side. However, the speed limit does not allow this to happen, and slows the field to prevent it reaching the top at $\phi = 0$ in finite time.

Let us now consider the case $V_2 = 0$ with $V_4 < 0$. We can generate such a potential by considering, for example, $H = h_3\phi^3$ which gives rise to the potential $V = V_4\phi^4 + V_6\phi^6$ where

$$\lambda V_4 = -\sqrt{1 + 36\lambda h_3^2\,(g_s^2 M_p^4)} + 1 \qquad \lambda V_6 = 3\lambda h_3^2\,(g_s M_p^2)$$

At late times the scalar field rolls like $\phi \sim 1/t$, and the equation of state parameter diverges: $\omega \to \infty$. Meanwhile, the scale factor a has the peculiar behaviour

$$a = a_0 \exp\left(-\frac{(1/(g_s M_p^2) + 36\lambda h_3^2)^{3/2}}{2^4\, 3^3\, h_3^2} \frac{1}{t^2} \right)$$

which undergoes a period of accelerated expansion before settling down to a steady-state cosmology at late times. The period of acceleration coincides with the period when the $V_6\phi^6$ term dominates over the $V_4\phi^4$ term in the potential; in the solution this also correlates with a period in which the potential dominates in the gravitational dynamics determined by ω (30.1433). As in case B, the effect becomes stronger with a stronger potential (so a flat potential is not required for slow roll since this is provided by the DBI kinetic terms).

Scale factors with the characteristic $\exp(-1/t^2)$ behaviour also occur in standard cosmology, but only for very finely tuned inital conditions. The novelty here is that such dynamics is generic for a large enough intitial velocity. A similar steady state dynamics also arises from $H = h_4\phi^4$, this time with $a \sim \exp(1/M_p t)$. Each of these scenarios finds Xeno exonerated and Hoyle happy.

30.5.2 Cosmology in the Cut-Off Throat

The interesting Case B of the previous section has $V_2 \neq 0$ which is not compatible with the conformal symmetry implied by an exact AdS throat. Instead, one ultimately expects a geometry in which the throat is cut off in the infra-red. We described such geometries in Section 4. Here we will work with the toy model in which the warp factor is taken to be

$$f(\phi) = \frac{\lambda}{(\phi^2 + \mu^2)^2}$$

For $\phi \gg \mu$, the geometry is essentially that of the AdS throat and the analysis of the previous sections holds. However, for $\phi \leq \mu$, the speed limit becomes simply

$$\dot{\phi} \leq \frac{\mu^2}{\sqrt{\lambda}}$$

So the scalar field can now reach the origin in finite time and happily sail right through at finite speed. This has the advantage that at late times the field oscillates around the minimum of the potential as usual and reheating may occur (depending on the couplings to the Standard Model). However, we must check that we can still arrange for a period of acceleration. The condition for acceleration requires large enough V_2 as we saw in (30.1444) Generally we have

$$V_2 = m^2$$

where masses m^2 which could be generated in various ways were listed in Section 4 and include contributions from graviton loop (30.1419), couplings to other sectors (30.1420) and a hard mass from the deformed geometry (30.1426).

Let us start by examining the upper limit for ϕ. In fact, we are quite entitled to take ϕ as large as we wish and, for $\phi \gg M_p$, one finds the standard slow roll expansion of chaotic inflation. However, as we shall discuss in Section 5.3, models with superPlanckian VEVs suffer from destabilization from a slew of quantum corrections involving for example gravitational curvature couplings $\mathcal{R}\phi^2$ [2]. To avoid this, we shall instead restrict ourselves to *subPlanckian* VEVs

$$\phi \ll M_p, \tag{30.1445}$$

where, as the speed limit is saturated, we may still obtain inflationary behaviour.[185]

In our cut-off geometry we also need ϕ to be suitably large so that the probe brane is experiencing the AdS throat-like background before the cut-off is reached. It can be checked that the solution described in Case B of Section 5.1 continues to hold in the cut-off geometry provided $\phi \gg \mu$. We therefore find our window of inflation when the scalar field lives in the regime,

$$\mu \ll \phi \ll M_p \tag{30.1446}$$

during which time we have the cosmological solution $a(t) = a_0 t^{h_1\sqrt{\lambda}}$ while the scalar field slows down as $\phi = \sqrt{\lambda}/t$. Let us assume that this solution actually holds all the way to the end points (30.1446). This may be overly optimistic, but it gives us a quick and ready way to calculate the number of e-foldings. We require that at the initial time $t = t_0$ we have $\phi(t_0 = 0) = M_p$, while at the final time $t = t_f$ we have $\phi(t = t_f) = \mu$. Then,

$$t_0 = \frac{\sqrt{\lambda}}{M_p} \quad \text{and} \quad t_f = \frac{\sqrt{\lambda}}{\mu}$$

The number of e-foldings is therefore given by,

$$n = \log\left(\frac{a(t = t_f)}{a(t = 0)}\right) = h_1\sqrt{\lambda}\log\left(M_p/\mu\right)$$

The argument of the logarithm is roughly the distance travelled by the scalar field. Any large number of efoldings comes from the prefactor. Using (30.1443) we have

$$n \sim h_1\sqrt{\lambda} \sim \frac{m\sqrt{\lambda}}{\sqrt{g_s}M_p} \geq 100 \tag{30.1447}$$

This criterion is the same as the criterion that the potential V dominate in the equation of state (in particular in ρ).

This inflationary phase is novel in several ways. Most importantly, the inflation can occur on a steeper potential hill than works for ordinary slow roll inflation. Moreover, at least before considering observational constraints from density perturbations, we find inflation to occur at sub-Planckian field strength for ϕ. We will show in section 5.3 that, when combined with the novel form of the equation of motion for ϕ, this allows us to avoid destabilizing effects from a curvature coupling $\mathcal{R}\phi^2$, circumventing some of the difficulties involved in placing inflation within a string compactification [2]. Moreover, as noted in Section 5.1, we have obtained *power-law* inflation from a quadratic potential.

At later times, the brane reaches the end of the cutoff throat and the novelties of the DBI action wear off. At this stage, we expect to reduce to an ordinary matter or radiation dominated phase, though we have not yet controlled the details of the exit from our inflationary stage.

[185] In our subsequent analysis of density perturbations [39], we discovered that to satisfy all the observational constraints including those on non-Gaussianity, we require starting inflation at Planck scale ϕ, which may require fine tuning.

The number of e-foldings (30.1447) can be made large if we can generate a large mass term $m^2\phi^2$ for our scalar field. In Section 4, we discussed possible mechanisms for generating such a term and argued that corrections to our crucial higher derivative terms were sub-leading. The simplest of the mass corrections was the hard contributions (30.1426) from the DBI action itself: $m_{DBI} \sim \mu/\sqrt{\lambda}$. Comparing with (30.1447) we immediately see that in our strong coupling regime $\lambda \gg 1$ there is no inflationary window with such a mass: we need another mechanism. As discussed at length in Section 4, such a mechanism may arise from coupling to other sectors which (depending on the strength of the couplings and cutoff scales for the other sectors in the system) may produce a large enough mass. More simply, as discussed above we may consider deforming the theory by a mass term for ϕ and note that this induces only subdominant corrections to the DBI kinetic terms. Again, it would be very interesting to determine the magnitude of mass m arising in specific string compactifications.

A Comment on V_4 Terms

In the above discussion, we concentrated on the possible mass terms that could be generated, and the associated window of applicability of the inflationary solution described in Case B of Section 5.1. However, the solution itself is valid only if the quadratic term in the potential dominates over the quartic term; otherwise we obtain the dust-like evolution of Case C. We therefore require,

$$V_4\phi^2 < m^2 \Rightarrow \phi^2 < m^2/V_4 \tag{30.1448}$$

If V_4 is too big, then this constraint on ϕ would be much stronger than (30.1445) and potentially ruin our window of inflation. Let us therefore examine the possible V_4 couplings.

In the DBI action, V_4 scales like $1/\lambda$. However, quantum corrections change this. Let's focus on the virtual effects of other sectors, since these were successful in giving us an inflationary window. The derivative coupling (30.1417), leads to a contribution of order

$$\Delta V_4 \sim \Lambda_{UV}^4/M_*^4 \tag{30.1449}$$

while at the same time inducing a mass $m_* \sim \Lambda_{UV}^2/M_*$ (30.1420) as we discussed earlier. Plugging this into (30.1448) leads to the constraint $\phi < M_*$ which is stronger than (30.1445). At the level of our effective field theory analysis (in terms of an unknown UV cutoff parameter Λ_{UV} and a coupling scale $1/M_*$), this is consistent with our inflationary window for appropriate values of these parameters.

30.5.3 Curvature Corrections

Let us now restore into our analysis the curvature coupling $\mathcal{R}\phi^2$ and the other curvature corrections scaling as powers of $\mathcal{R}/\phi^2$. We will consider the regime of our solutions where $\phi \ll M_p$, and show that the solutions are not destabilized by the addition of these curvature couplings (in other regimes more fine tuning would be required [39]). We should note here that the *conformal* coupling required to render the stress energy tensor traceless in the case of large γ will be more complicated than a $\mathcal{R}\phi^2$ coupling.

Let us first consider the $\mathcal{R}/\phi^2$ corrections to the terms in the Lagrangian (30.1416) which scale like H^2/ϕ^2. In our inflating solution, these terms are of the order m^2/M_p^2. As long as this quantity is sufficiently smaller than 1, then the correction terms are negligible compared to the terms in (30.1416) that we did include in the initial analysis. In the noninflating solutions, $H \sim 1/t$ and these corrections are suppressed by powers of λ which makes them very safe from affecting the results above.

Now let us turn to the curvature coupling $\mathcal{R}\phi^2$. First, let us consider the Einstein equations for the metric. The effect of the conformal coupling is to replace M_p^2 with $M_p^2 + \phi^2$ in these equations, changing the effective Planck mass. Since $\phi \ll M_p$, this does not change the solutions for $a(t)$ significantly. Similar comments apply to the $H^2\phi^2$ terms which plague F-term inflation in supergravity models.

Secondly, we must check whether the scalar field equation of motion is self consistently solved by the original solution in the presence of a curvature coupling. The coupling $\xi\mathcal{R}\phi^2$ contributes to the equation of motion for ϕ, (30.1432), as

$$\frac{\lambda\ddot{\phi} - 6\lambda\dot{\phi}^2/\phi + 4\phi^3}{(1-\lambda\dot{\phi}^2/\phi^4)^{3/2}} + \frac{3\lambda H\dot{\phi}}{(1-\lambda\dot{\phi}^2/\phi^4)^{1/2}} + (V' + 2\xi\mathcal{R}\phi - 4\phi^3) = 0 \tag{30.1450}$$

Because $\gamma(t)^{-1} = \sqrt{1-\lambda\dot{\phi}^2/\phi^4}$ is going to zero at late times, the terms involving V and $\mathcal{R}$ are subdominant in this regime. This is reflected in the fact that, for Cases B and C of Section 5.1, the late time behaviour of the scalar field saturates the speed limit and is independent of the potential.

In summary, the curvature couplings we have considered here do not change our results. This is in contrast to the usual situation (30.1396), where for example the slow roll conditions for inflation in a polynomial potential are impossible to satisfy in the presence of the coupling $\mathcal{R}\phi^2$. This point was emphasized for the case of 3-branes in warped throats in [2]. Here see that near the origin, where the DBI velocity corrections are crucial, the structure of the scalar equation of motion is dramatically different due to the factors of $\gamma(t)$ in the denominator in (30.1450).

30.5.4 Relation to other Works

The behavior of our system, especially in Case C in Section 5.1, bears a strong resemblance to the tachyon matter system studied in recent years [5, 8]. Indeed, the anti-brane moving in the *AdS* throat is, in some sense, a strong coupling limit of the tachyon matter system. There are important differences between our case and the weakly coupled tachyon matter system, most notably the fact that as we discussed above in our situation the tachyon has been lifted and, as we shall see, particle production is suppressed. Moreover, the spectra of particles and strings in the two backgrounds differ.

A simple field redefinition can be used to relate the effective field theories in the two cases as follows. In tachyon cosmology, the action for the tachyon is of the form [36]

$$\mathcal{L} = -F(T)\sqrt{1-\dot{T}^2} \tag{30.1451}$$

which, upon expanding the square-root, has a potential term $F(T)$. We can simply generalise this action to include a further potential term $G(T)$, which we could use either to cancel F or simply to make the potential different from the coefficient of the kinetic terms. Such a potential may be generated in situations where the brane antibrane system is embedded in a compactification. We therefore consider,

$$\mathcal{L} = -F(T)\sqrt{1-\dot{T}^2} - G(T)$$

The stress-energy tensor and pressure can be easily calculated to yield energy density

$$\rho = \frac{F}{\sqrt{1-\dot{T}^2}} + G \qquad p = -F\sqrt{1-\dot{T}^2} - G$$

Tachyon cosmologists work with this Lagrangian imposing $G = 0$. Of particular relevance for the present work is the observation [9] that the power-law inflationary phase of Case B can be obtained by the choice $F(T) \sim 1/T^2$ in the Lagrangian (30.1451). It was further shown that higher order (inverse) potentials give rise to dust behaviour as we saw in Case C. To return to our favourite Lagrangian and compare these results with those above, we employ the field redefinitions

$$T = \frac{\sqrt{\lambda}}{\phi} \qquad F(T) = \frac{\lambda^2}{T^4} \qquad G(T) = V(\phi) - \phi^4$$

It is very interesting to ask how generally motion on internal scalar field configuration spaces can be imbued with a geometric interpretation (and therefore a speed limit induced by causality). It is intriguing that the tachyon Lagrangian (30.1451) has a structure reminiscent that of a relativistic particle moving in spacetime.

The motion of the probe brane in the cut-off geometry is also somewhat similar to the discussion of [37]. These authors consider the induced mirage cosmology [38] on a probe brane as it moves in the Klebanov-Strassler geometry [31]. This differs from our analysis in two important respects: firstly, in [37] the probe brane was taken to move slowly through the background geometry, so that the DBI action could be approximated by two-derivative terms. Secondly, the cosmology was viewed from an observer on the probe brane, rather than from the perspective of a boundary observer as is relevant for our discussion.

30.6 Perturbations and Particle Production

We now come to the promised self consistency checks of our time dependent evolution. In order to trust our results in the previous sections, we must check that particle and string production and fluctuations about the solution do not destabilize it. We have already checked that acceleration is small enough to avoid closed string production, and that back reaction from the probe's large kinetic energy is small in the window (30.1412).

30.6.1 W boson and String Production

Firstly, we consider the production of massive strings on the brane and W-bosons. Our effective DBI action is blind to the possible on-shell creation of these modes and we must check by hand that the production is suppressed in the $\lambda \gg 1$ regime. This is rather simple. The time dependent solution we found entails time dependent masses for both W-bosons (30.1397) and massive string modes on the brane (30.1401) since these masses are proportional to $\phi \sim \sqrt{\lambda}/t$.

The strength of particle production for a time dependent frequency $\omega(t)$ is controlled by $\dot{\omega}/\omega^2$. This is a rough estimate which can be obtained by starting from the requirement for adiabatic evolution for a frequency which changes over a time period δt:

$$\delta t \gg 1/\omega_{min} \tag{30.1452}$$

where ω_{min} is the smallest value of ω obtained in the evolution. Dividing by the change $\delta\omega$ arising in the process yields

$$\delta t/\delta\omega \gg 1/(\omega_{min}\delta\omega) \tag{30.1453}$$

Consider a long enough evolution so that $\delta\omega$ is of order ω_{max}. Replacing $\delta\omega$ on the right hand side by ω_{max} and replacing $\omega_{max}\omega_{min}$ by ω^2 yields a condition under which particle production may be consistently neglected:

$$\dot{\omega} \ll \omega^2 \tag{30.1454}$$

This can be thought of as expressing the fact that (on average) the energy from the time dependence of the frequency is smaller than the jump in energy needed to produce an on shell particle. We find in our situation that the inequality (30.1454) is satisfied parametrically in λ for large λ. In particular for the W boson zero modes we find

$$\frac{\dot{m}_W}{m_W^2} \sim \frac{1}{\sqrt{\lambda}} \tag{30.1455}$$

which is suppressed in our large λ regime. Nonzero modes are more suppressed. We also need to check the production of open string oscillator modes on the brane. The masses for these modes are given by the warped value $m_s(\phi) \sim \phi/\lambda^{1/4}$. If they were created at rest, a similar calculation to the one above would yield again parametric suppression of $\dot{m}_s/m_s^2 \sim 1/\lambda^{1/4}$. Since the brane is moving with a large velocity $\sqrt{\lambda}\dot{\phi}/\phi^2$, the strings created on the brane have energies boosted to

$$\omega \sim \gamma\sqrt{N_{osc}m_s(\phi)^2 + \vec{p}^2} \tag{30.1456}$$

where N_{osc} is an integer coming from the oscillator level of the string. These energies are large, and we find that again the time dependent background does not inject enough energy to create these states. Therefore our solution is stable against string and W boson production in the strong coupling regime.

30.6.2 Perturbations of ϕ

In the remainder of this section, we will analyze the fluctuations of ϕ itself about the solutions we have found. Our main goal will be to determine in what regime the perturbations are not dangerous for our solutions above. These results also pertain to the spectrum of density perturbations produced in our inflationary phase. We will defer a detailed discussion of any predictions of that model for the CMBR to a later investigation. We perturb our solutions as,

$$\phi(t) \to \phi(t) + \alpha(x,t)$$

Expanding in Fourier modes, $\alpha(x,t) = \alpha_k(t)e^{ikx}$, the equation of motion for α obtained by expanding the action (30.1414) is,

$$\ddot{\alpha}_k + \left(\frac{6}{t} + 3H\right)\dot{\alpha}_k + \left(\frac{6}{t^2} + \frac{6H}{t} + \frac{k^2}{\gamma_0^2 a^2 t^4}\right)\alpha_k = 0$$

where we have used the late time behavior of the function γ defined in (30.1427) which, for most of our solutions, is $\gamma \to t^2/\gamma_0$ for constant γ_0. In particular, for the inflationary solution of Case B we have $\gamma_0 \sim mM_p/\sqrt{\lambda}$.

If we analyze this equation of motion in the global case, where $H = 0$, we obtain the results $\delta\phi \propto t^{-2}$ and $\delta\phi \propto t^{-3}$ for the $k = 0$ mode. For the $k \neq 0$ modes, the perturbations will be further suppressed. So we see that the perturbations do not grow relative to the background solution $\phi \sim \sqrt{\lambda}/t$; i.e. these perturbations are non-tachyonic.

Let us now analyze the equation with gravity turned on, focusing on the inflationary solution. We would like to estimate the density perturbations that result from the fluctuation $\delta\phi$ in the field. In our inflationary phase, we have

$$H = \frac{1}{\epsilon_0 t} \tag{30.1457}$$

where, from (30.1442), we have $\epsilon_0 = (h_1\sqrt{\lambda})^{-1}$. Since we have $\epsilon_0 \ll 1$, in the inflationary phase the terms involving H and H/t dominate over the terms involving $1/t$ and $1/t^2$ respectively. The relative importance of the k-terms depends on the wavelength under consideration. In general, we therefore have

$$\ddot{\alpha}_k + \frac{3}{\epsilon_0 t}\dot{\alpha}_k + \left(\frac{6}{\epsilon_0 t^2} + \frac{ck^2\lambda}{M_p^2 m^2}\frac{1}{a^2 t^4}\right)\alpha_k = 0 \tag{30.1458}$$

where c is a constant of order one, parameterising our ignorance of γ_0.

For superhorizon fluctuations, defined by $k^2 \ll a^2H^2$, where aH is the comoving horizon size, we may neglect the spatial kinetic terms in (30.1458). This leads to a simple pair of solutions to the classical equation (30.1458) for the modes:

$$\alpha^{(0)} \sim \frac{\alpha_0(k)}{t^2} \tag{30.1459}$$

and

$$\alpha^{(1)} \sim \frac{\alpha_1(k)}{t^{3/\epsilon_0}} \tag{30.1460}$$

These modes decay faster at large t than the perturbations of the scalar field in ordinary slow roll inflation based on (30.1396), which go to a constant as the modes cross the horizon. There are reasonable regimes of parameters for which all the subPlanckian k/a modes have the gradient term in (30.1458) suppressed relative to the effective mass squared term there.

To obtain a prediction for the observed density perturbations, we must be careful to follow them as they re-enter the horizon at later times. For now, let us simply check that the perturbations do not destabilize our inflationary phase.

During inflation, the density perturbations are simply

$$\left.\frac{\delta\rho}{\rho}\right|_{k,\ during\ inflation} \sim \frac{V'\delta\phi_k}{V} \tag{30.1461}$$

For our polynomial potentials, this yields $\delta\phi/\phi$. As we have seen, our perturbations die faster than in ordinary inflation due to the stronger role for the mass term in (30.1458), and they certainly do not compete with ϕ in magnitude. Thus our density perturbations during inflation are small enough to avoid back reaction. The question of whether one obtains eternal inflation depends on the magnitude of $H\delta\phi/\dot{\phi}$, which can be small in our speed of light phase. We plan to pursue in future work a detailed study of the predictions of our inflationary models, as well as an analysis of reheating and other required features.

Part IX
Appendix

обсудим математические методы, которые используются в тех или иных задачах.

A Introduction and Overview of Cosmology

A.1 Another Motivation for Cosmology

A.1.1 Other General Motivation

Если мы изучаем глубоко любую теорию, то очень быстро можно задаться вопросом про то, что же было во Вселенной?

Поэтому это достаточно типичный предмет, и знать его очень прямо полезно, потому что про многое можно сказать, зная его основы. Не следует считать, что это особенный предмет, до которого сложно дойти. На самом деле задаться вопросами про космологию можно очень часто.

современное представление о мире основано на космологии

Просто интересно разные модели Вселенной рассматривать

Конечно, такой подход физически никакого не имеет смысла, потому что чем вертеть теории, можно задаваться вопросами, какая из них верная? Вот - гораздо важнее вопросы. Тем не менее, если бы какие-то параметры Вселенной были бы другими, то что бы было? Такие задачи тоже интересно порешать.

после ОТО запросто взять, узнать интересные модели

то есть это не сложно.

Круто изучать раннюю Вселенную

типа интересно, что там было?

Множество приложений

Укажем приложения
(потом раскрою)

Польза в плане развития как профессионала

(потом раскрою)

Польза в плане саморазвития

(потом раскрою)

Результаты

Обсудим, что полезного и важного человечество получило, поняв этот предмет.

Мотивация к разделам

(потом раскрою в параграфах)

A.1.2 Motivation for Exotic Theories in Cosmology

(пока хз, какие разделы есть)

К квантовой космологии

Будет что ответить эзотерикам

а то альтернативных версий, что там со вселенной происходит полно, и как же хочется четко ответить, что в какой степени верно, в какой нет! и свою версию четко рассказывать. Максимально круто ведь!

A.1.3 Amazing Facts

(????)

A.2 Mindset of a Professional in Cosmology

A.2.1 Isn't 21st Century Cosmology Bullshit? (????)

Что-то я подозреваю, что является, как минимум потому что образование в ней такое, что все самые принципиальные вопросы совершенно плохо раскрыты.
(не знаю, постоянно этот вопрос есть, потом отвечу развернуто на него)

Очень часто приходится останавливаться на том, что такая-то есть модель и так-то с ней работаем. Нормально понять, почему именно ее используем, как она экспериментально подтверждается, как про нее догадаться времени никогда нет во время первого отличного прохождения курса космологии. Это минус, но, видимо, потом это понимание может прийти

Это тоже сильно указывает на то, что космология может быть херней. Потому что мало кто обсуждает очень важные вопросы. А может быть, просто в образовании по космологии проблемы и через 100 лет будут эти же вещи рассказывать, но в другом совершенно формате?

Очень неприятно, когда люди используют утверждения, которые и близко не могут доказать. И в космологии это почти везде

То встречается отсылка к стандартной модели, то к квантовой теории поля при конечной температуре, то еще к чему-то. И откуда-то мы должны знать и понимать какую-то формулу, которая вообще-то в их курсе и не встречается. Вероятность, что формула неверная может быть большой, потому что проверить верность их нет времени, очень много времени скорее всего ушло бы.
Поэтому такая космология сейчас, может быть это все и херня, фундаментально на вопросы ответов скорее всего у профессоров не будет, будут отсылки к каким-то работам кого-то, или к другим книгам. Или может быть нужно просто лучшего космолога встретить?

A.2.2 A Cosmologist's View of the World (???)

(потом сформулирую, я еще крайне далек, чтобы из формул и теории вывести хотя бы какой-то вывод и написать ответ на вопросы ниже)

Что такое гармония в мире, если ты понимаешь историю Вселенно?

Какие основные законы функционирования мира, если ты понимаешь историю Вселенно?

По каким законам лучше всего жить людям, если ты понимаешь историю Вселенно?

A.2.3 How to Derive and Solve Equations in Cosmology?

(отдельный раздел про это, потому что часы ушли только на этот вывод. Аккуратно смотреть, где какие буковки и что они обозначают. так только будет движение вепред, космология отлично этот навык тренирует, без него ничего не докажется.)
(general discussion how to get to the very important equations in cosmology)

Без четкости и аккуратности в космологии ничего не достичь. Можно даже сказать, что тренировки по обычным задачам в космологии - уже стоят того, чтобы ее изучать, потому что это очень хороший способ прокачать эти навыки.

(потом раскрою эту мысль)

О задачах, где есть пара уравнений и нужно в каком-то приближении найти решения (!!!!!)

(задача про квантовую космологию така!!!! укажу тут, как ее решить, хорошая ведь задача!!!)

A.2.4 Equations in Cosmology: Important Features of Particular Cases

(напишу потом, тоже много про это соображений будет с опытом)

Всегда решение начинается с того, что убеждаешься в том, правильно ли ты понимаешь обозначения, особенно, по какому параметру от сложной ($f(g(x))$) или простой ($f(x)$) функции чего производные!

Никогда решение задач в космологии не начинается с другого!
Самое важное - всегда смотреть, в каком времени взяты производные! Иначе минус дни только лишь на то, чтобы осознать, что на самом деле производные в уравнениях в конформном времени записаны, оказывается.
Поэтому имеет смысл вообще напротив каждого уравнения это напоминать себе. Так лучше, чем понадеяться, что на прошлой странице написано, так что можно восстановить.

О задачах про инфляцию

(!! скоро их прорешаю, одно из первых занятий в космологии в планах!)

В разделе про инфляцию всегда следим, по чему именно берем производные, по ϕ, τ или t, не забываем, что часто нужно брать именно сложные производные!

(раскрою потом эту важную мысль. поэтому я всегда и указываю с боку, какое время?)

Что делать, если типичная задача про инфляцию не решается?

(еще потренируюсь, пока тут, потом мб в дополнения перенесу)

- Подумай, по какому времени у нас производная: по конформному или по обычному?

  Может, всю инфляцию нужно переписать, чтобы путаницы в этом не было? Потому что столько путаницы в этих временах, столько часов уходит только лишь на то, что оказывается, там именно от конформного времени производная! Пока важно просто это помнить, это очень неприятно.

  (уже писал в дополнениях, но пусть уж лучше пока 2 раза будет записано, потом мб удалю отсюда)

- Если ты дифференцируешь, то следует помнить, какая функция от какого аргумента или функции зависит, а какой штрих или точка что обозначает.

  Часто от этого возникают дополнительные множители, без которых выражение будет не верным.

A.2.5 Ways to Earn Money by Knowing Cosmology and Most Used Topics (??)

(?????)

A.2.6 Overview of Misconceptions About Cosmology (??)

Examples of misconceptions or easily misinterpreted statements in the literature by Davis Lineweaver

(хорошие авторы их собрали, спасибо им, может, посмотрю когда-то)

In text books and works of popular science it is often standard practice to simplify arguments for the reader. Some of the quotes below fall into this category. We include them here to point out the difficulty encountered by someone starting in this field and trying to decipher what is really meant by 'the expansion of the Universe'.

[1] Feynman, R. P. 1995, *Feynman Lectures on Gravitation (1962/63)*, (Reading, Mass.: Addison-Wesley) p. 181, "It makes no sense to worry about the possibility of galaxies receding from us faster than light, whatever that means, since they would never be observable by hypothesis."
[2] Rindler, W. 1956, MNRAS, **6**, 662-667, *Visual Horizons in World-Models*, Rindler acknowledged that faster than c expansion is implicit in the mathematics, but expresses discomfort with the concept: "... certain physical difficulties seem to be inherent in models possessing a particle-horizon: if the model postulates point-creation we have material particles initially separating at speeds exceeding those of photons."
[31] [3] McVittie, G. C. 1974, QJRAS, **15**, 246-263, *Distances and large redshifts*, Sect. 4, "These fallacious arguments would apparently show that many quasars had 'velocities of recession' greater than that of light, which contradicts one of the basic postulates of relativity theory."
[?] [4] Weinberg, S. 1977, *The First Three Minutes*, (New York: Bantum Books), p. 27, "The conclusion generally drawn from this half century of observation is that the galaxies are receding from us, with speeds proportional to the distance (at least for speeds not too close to that of light).", see also p. 12 and p. 25. Weinberg makes a similar statement in his 1972 text *Gravitation and Cosmology* (New York: Wiley), p. 417, "a *relatively close* galaxy will move away from or toward the Milky Way, with a radial velocity $[v_{\rm rec} = \dot{R}(t_0)\chi]$." (emphasis ours). Shortly thereafter he adds a caution about SR and distant sources: "it is neither useful nor strictly correct to interpret the frequency shifts of light from very distant sources in terms of a special-relativistic Döppler shift alone. [The reader should be warned though, that astronomers conventionally report even large frequency shifts in terms of a recessional velocity, a "red shift" of v km/sec meaning that $z = v/(3 \times 10^5)$.]"
[?] [5] Field, G. 1981, *This Special Galaxy*, in Section II of *Fire of life, the book of the Sun*, (Washington, DC: Smithsonian Books) "The entire universe is only a fraction of a kilometer across [after the first millionth of a second], but it expands at huge speeds — matter quite close to us being propelled at almost the speed of light."
[?] [6] Schutz, B. F. 1985, *A first course in General Relativity*, (Cambridge: Cambridge U. Press) p. 320, "[v=HD] cannot be exact since, for $D > 1.2 \times 10^{26} m = 4000$ Mpc, the velocity exceeds the velocity of light! These objections are right on both counts. Our discussion was a *local* one (applicable for recession velocity

<< 1) and took the point of view of a particular observer, ourselves. Fortunately, the cosmological expansion is slow..."
[?] [7] Peebles, P. J. E., Schramm, D. N., Turner, E. L. and Kron, R. G. 1991, Nature **352**, 769, *The case for the relativistic hot Big Bang cosmology*, "There are relativistic corrections [to Hubble's Law, $v = H_0 D$,] when v is comparable to the velocity of light c." However, Peebles, in his 1993 text *Principles of Physical Cosmology*, (Princeton: Princeton University Press), p. 98, explains: "Since equation [$D = R\chi$] for the proper distance [D] between two objects is valid whatever the coordinate separation, we can apply it to a pair of galaxies with separation greater than the Hubble length... Here the rate of change of the proper separation, [$\dot{D} = HD$], is greater than the velocity of light. This is not a violation of special relativity;" Moreover, in the next paragraph Peebles makes it clear that, dependent upon the cosmological parameters, we can actually observe objects receding faster than the speed of light.
[?] [8] Peacock, J. A. 1999, *Cosmological Physics*, (Cambridge: Cambr Univ. pr.) p. 6, "...objects at a vector distance $\mathbf{r}$ appear to recede from us at a velocity $\mathbf{v} = H_0\mathbf{r}$, where H_0 is known as Hubble's constant (and is not constant at all as will become apparent later.) This law is only strictly valid at small distances, of course, but it does tell us that objects with $r \simeq c/H_0$ recede at a speed approaching that of light. This is why it seems reasonable to use this as an upper cutoff in the radial part of the above integral." However, Peacock makes it very clear that cosmological redshifts are not due to the special relativistic Doppler shift, p72, "it is common but misleading to convert a large redshift to a recession velocity using the special-relativistic formula $1 + z = [(1 + v/c)/(1 - v/c)]^{1/2}$. Any such temptation should be avoided"
[26] [?] [9] Davies, P. C. W. 1978, *The Runaway Universe* (London: J. M. Dent & Sons Ltd) p. 26, "...galaxies several billion light years away seem to be increasing their separation from us at nearly the speed of light. As we probe still farther into space the redshift grows without limit, and the galaxies seem to fade out and become black. When the speed of recession reaches the speed of light we cannot see them at all, for no light can reach us from the region beyond which the expansion is faster than light itself. This limit is called our horizon in space, and separates the regions of the universe of which we can know from the regions beyond about which no information is available, however powerful the instruments we use."
[?] [10] Berry, M. 1989, *Principles of Cosmology and Gravitation*, (Bristol, U.K.: IOP Publishing) p. 22 "...if we assume that Euclidean geometry may be employed, ...galaxies at a distance $D_{max} = c/H \sim 2 \times 10^{10}$ light years $\sim 6 \times 10^9$ pc are receding as fast as light. Light from more distant galaxies can never reach us, so that D_{max} marks the limit of the observable universe; it corresponds to a sort of *horizon*."
[?] [11] Raine, D. J. 1981, *The Isotropic Universe*, (Bristol: Adam Hilber Ltd) p. 87, "One might suspect special relativistic effects to be important since some quasars are observed to exhibit redshifts, z, in excess of unity. This is incompatible with a Newtonian interpretation of the Doppler effect, since one would obtain velocities $v = cz$ in excess of that of light. The special relativistic Doppler formula $1 + z = (c + v)/(c - v)^{1/2}$ always leads to sub-luminal velocities for objects with arbitrarily large redshifts, and is at least consistent. In fact we shall find that the strict special relativistic interpretation is also inadequate. Nevertheless, at the theoretical edge of the visible Universe we expect at least in principle to see bodies apparently receding with the speed of light."
[?] [12] Liddle, A. R. 1988, *An introduction to Modern Cosmology*, (Sussex: John Wiley & Sons Ltd) p. 23, Sect 3.3, "...ants which are far apart on the balloon could easily be moving apart at faster than two centimetres per second if the balloon is blown up fast enough. But if they are, they will never get to tell each other about it, because the balloon is pulling them apart faster than they can move together, even at full speed."
[?] [13] Krauss, L. M. and Starkman, G. D. 1999, ApJ, 531(1), 22–30, *Life, the universe and nothing: Life and death in an ever-expanding universe*, "Equating this recession velocity to the speed of light c, one finds the physical distance to the so-called de Sitter horizon... This horizon, is a sphere enclosing a region, outside of which no new information can reach the observer at the center". This would be true if only applied to empty universes with a cosmological constant - de Sitter universes. However this is not its usage: "the universe became Λ-dominated at about 1/2 its present age. The 'in principle' observable region of the Universe has been shrinking ever since.... Objects more distant than the de Sitter horizon [Hubble Sphere] now will forever remain unobservable."
[?] [14] Harrison, E. R. 1991, ApJ, 383, 60–65, *Hubble spheres and particle horizons*, "All accelerating universes, including universes having only a limited period of acceleration, have the property that galaxies at distances $L < L_H$ are later at $L > L_H$, and their subluminal recession in the course of time becomes superluminal. Light emitted outside the Hubble sphere and traveling through space toward the observer recedes and can never enter the Hubble sphere and approach the observer. Clearly, there are events that can never be observed, and such universes have event horizons." The misleading part of this quote is subtle – there will be an event horizon in such universes (accelerating universes), but it needn't coincide with the Hubble sphere. Unless the universe is accelerating so quickly that the Hubble sphere does not expand (exponential expansion) we will still observe things from beyond the Hubble sphere, even though there is an event horizon (see Fig. 17).
[?] [15] Harwit, M. 1998, *Astrophysical Concepts*, 3rd Ed., (New York: Springer-Verlag) p. 467, "Statement (i) [In a model without an event horizon, a fundamental observer can sooner or later observe any event.] depends

on the inability of particles to recede at a speed greater than light when no event horizon exists."
[?] [16] Hubble, E. and Humason, M. L. 1931, ApJ, 74, 443–480, *The Velocity-Distance relation among Extra-Galactic Nebulae*, pp. 71–72, "If an actual velocity of recession is involved, an additional increment, equal to that given above, must be included in order to account for the difference in the rates at which the quanta leave the source and reach the observer[1].", Footnote 1: "The factor is $\sqrt{\frac{1+v/c}{1-v/c}}$ which closely approximates $1 + d\lambda/\lambda$ for red-shifts as large as have been observed. A third effect due to curvature, negligible for distances observable at present, is discussed by R. C. Tolman..."
[?] [17] Lightman, A. P., Press, W. H., Price, R. H. and Teukolsky, S. A. 1975, *Problem book in relativity and gravitation*, (Princeton, U.S.A.: Princeton U. Press) Prob. 19.7
[18] Halliday, D., Resnick, R. and Walker, J. 1974, *Fundamentals of Physics*, (USA: John Wiley & Sons) 4th Ed., Question 34E, "Some of the familiar hydrogen lines appear in the spectrum of quasar 3C9, but they are shifted so far toward the red that their wavelengths are observed to be three times as large as those observed for hydrogen atoms at rest in the laboratory. (a) Show that the classical Doppler equation gives a relative velocity of recession greater than c for this situation. (b) Assuming that the relative motion of 3C9 and the Earth is due entirely to recession, find the recession speed that is predicted by the relativistic Doppler equation." See also Questions 28E, 29E and 33E.
[19] Seeds, M. A. 1985, *Horizons - Exploring the Universe*, (Belmont, California: Wadsworth Publishing) pp. 386–387, "If we use the classical Doppler formula, a red shift greater than 1 implies a velocity greater than the speed of light. However, as explained in Box 17-1, very large red shifts require that we use the relativistic Doppler formula.... *Example:* Suppose a quasar has a red shift of 2. What is its velocity? *Solution:* [uses special relativity]"
[20] Kaufmann, W. J. and Freedman, R. A. 1988, *Universe*, (New York: W. H. Freeman & Co.) Box 27-1, p. 675, "... quasar PKS2000-330 has a redshift of $z = 3.78$. Using this value and applying the full, relativistic equation to find the radial velocity for the quasar, we obtain $v/c = (4.78^2 - 1)/(4.78^2 + 1) = ... = 0.92$. In other words, this quasar appears to be receding from us at 92% of the speed of light."
[21] Taylor, E. F. and Wheeler, J. A. 1991, *Spacetime Physics: introduction to special relativity*, (New York: W. H. Freeman & Co.) p. 264, Ex. 8-23
[22] Hu, Y., Turner, M. S. and Weinberg, E. J. 1993, Phys. Rev. D, 49(8), 3830–3836, *Dynamical solutions to the horizon and flatness problems*, "... many viable implementations of inflation now exist. All involve two key elements: a period of superluminal expansion..." They define superluminal expansion later in the paper as "Superluminal expansion might be most naturally defined as that where any two comoving points eventually lose causal contact." Their usage of superluminal conforms to this definition, and as long as the reader is familiar with this definition there is no problem. Nevertheless, we should use this definition with caution because even if the recession velocity between two points is $\dot{D} > c$ this does not mean those points will eventually lose causal contact.
[23] Khoury, J., Ovrut, B. A., Steinhardt, P. J. and Turok, N. 2001, Phys. Rev. D, 64(12), 123522, *Ekpyrotic universe: Colliding branes and the origin of the hot big bang*, "The central assumption of any inflationary model is that the universe underwent a period of superluminal expansion early in its history before settling into a radiation-dominated evolution." They evidently use a definition of 'superluminal' that is common in inflationary discussions, but again, a definition that should be used with caution.
[24] Lovell, B. 1981, *Emerging Cosmology*, (New York: Columbia U. Press) p. 158, "... observations with contemporary astronomical instruments transfer us to regions of the universe where the concept of distance loses meaning and significance, and the extent to which these penetrations reveal the past history of the universe becomes a matter of more sublime importance"
[25] McVittie, G. C. 1974, QJRAS, 15(1), 246–263, *Distance and large redshifts*,"... conclusions derived from the assertion that this or that object is 'moving with the speed of light' relative to an observer must be treated with caution. They have meaning only if the distance and time used are first carefully defined and it is also demonstrated that the velocity so achieved has physical significance."

A.2.7 General Approach to a Research in Cosmology

Ожидай, что много не очень умных людей этим занимаются, так что лучше иметь отличную подготовку по другим предметам, только с ней заниматься космологией

Без этого слишком большая вероятность, что будешь фигней страдать.

A.2.8 Ways to Guess the Main Ideas (???)

незаменимая часть нормального понимания предмета.

A.2.9 Effective Studying of Cosmology

Не трать особо время на то, в чем ты не понимаешь обозначения and also if you don't understand mathematical transformations

В космологии очень много обозначений, и часто один параметр может значить многое разное.

Так что я всегда и указываю после уравнения, что чем является!! (проверю потом, так ли это!!!)

И это огромная трата времени думать про что-то, что не определено рядом с уравнением!!! Не делай так, и можешь просто недели сэкономить. Не читай такие места в книгах или записях лекций!!!

Огромная трата времени думать про что-то, в чем ты не понимаешь параметры, так что обходи это стороной.

(I'll write later here, that there are not so hard mathematical transformations, one just needs to understand, what is really meant by what is written. since sometimes it is explained very bad, there could be days wasted. I'd recommend that if the formula is not reproduced in some reasonable time by all reasonable methods, then just assume it and don't be stuck on derivations! this can save even 5 days during 30 days of research in cosmology!)

чтобы начать заниматься космологией, собери просто основы теории поля и начинай.

как начать заниматься космологией?

казалось бы, в космосе может происходить что угодно, как вообще подступиться к описанию чего-то там?

подробно подумаю, как начать

основы дифференциальной геометрии должны быть взяты

думаю, без этого никак.

A.3 Acknowledgements

Currently, no one except me has worked on the sections of this note (with the exception of sections taken from books).

A.4 Literature on Cosmology

A.4.1 Main Literature

Basic educational literature

[?] D. Baumann Cosmology Part III Mathematical Tripos (Cambridge University)

165 страниц (с решениями). Хорошие лекции для введения, многое очень компактно и достаточно понятно написано, но некоторые темы совсем мало раскрыты там. Не уверен, что этих лекций достаточно для понимания темы.

David Tong Cosmology http://www.damtp.cam.ac.uk/user/tong/cosmo.html

Вроде хорошие компактные лекции, многое уже выгружено и скорее всего все темы отсюда у меня тоже будут написаны.

[?] J. Lesgourgues Cosmological perturbations

Вроде хорошие лекции по возмущениям в космологии, точно не на пару дней изучения. Обещаются качественные обсуждения возмущений, что очень круто.

Edward L. Wright. course for Ucla university https://www.astro.ucla.edu/~wright/cosmolog.htm (2005)
Очень крутые онлайн страницы с очень интересной информацией. Есть секция ответов на вопросы и очень много ссылок.

Курс университета Калифорнии https://phys.libretexts.org/Courses/University_of_California_Davis/UCD%3A_Physics_156_-_A_Cosmology_Workbook/Workbook
Кроме того, что интересные и важные темы написаны, там есть некоторые интерактивные игры-квизы, что просто прекрасно. Может, когда-нибудь такое везде будет обучение? Очень хорошая идея - почитать их.

K. Zhou Lecture notes on cosmology https://knzhou.github.io/notes/cos.pdf
Хороший обзор космологии от студента Стенфорда, примерно такой объем тем и входит в минимальную подготовку по космологии.

Основная профессиональная литература

[2] Scott Dodelson Modern Cosmology
Книга, где очень много всего написано, многие на нее ссылаются, так что и я пока считаю ее основной, скорее всего многое из нее выгружу.

Books With Many Solved Problems

Лайтман Пресс Прайс Тюкольски Сборник задач по ОТО и гравитации
Шикарный задачник, в котором также есть раздел про космологию.

A.4.2 Articles on Different Effects

Статьи про явления во Вселенной

[?] Perlmutter, S. and Aldering, G. and Goldhaber, G. Measurements of Ω and Λ from 42 High-Redshift Supernovae
Статья о том, что все-таки Вселенная расширяется по средствоим наблюдения дальних звезд, за которую дали Нобелевскую премию.

[?] Type Ia Supernova Discoveries at z > 1 from the Hubble Space Telescope: Evidence for Past Deceleration and Constraints on Dark Energy EvolutionAdam G. Riess and Louis-Gregory Strolger and John Tonry and Stefano Casertano et al.
Известная статья о том, что где-то что ли Вселенная замедляется? Не особо разбирался, но многие ее цитируют.

Max Tegmark's website https://space.mit.edu/home/tegmark/movies.html
Красивая анимация, которая показывает, как изменится спектр СМБ и разное другое, если поменять космологические параметры. Самый быстрый способ ответить на вопросы про это - нажать туда на кнопку и посмотреть эти анимации.

О квантовых эффектах в начале Вселенной

[?], Coleman, Sidney and De Luccia, Frank Gravitational effects on and of vacuum decay
Там во время фазовых переходов образуются пузыри и с ними происходили фазовые переходы в начале Вселенной. Вот описание этого. Отдельная заморочка на самом деле, но интересная теория. Слушал про это лекцию.

Углубленная литература и литература с дополнительными темами

[?] Viatcheslav Mukhanov Physical Fundations of Cosmology https://sites.astro.caltech.edu/~george/ay21/readings/Mukhanov_PhysFoundCosm.pdf
Тоже интересная, большая книга, многое из нее не знаю, наверное, много хороших идей, которых в других книгах не прочитать.

[?] Landau L. D. Field theory
A couple of simple and interesting models have been discussed, uploaded, but I'm not sure if these models are used with notebooks at all? Then I'll continue to sort it out, somehow these topics turned out to be not very necessary.

Обзорные статьи сложных явлений или многих моделей

The Best Inflationary Models After Planck Jerome Martin, Christophe Ringeval, Roberto Trotta, Vincent Vennin https://arxiv.org/abs/1312.3529v3
Шикарная статья с обзором инфляционных моделей, очень интересно потом почитать ее будет.

Daniele Bertacca, Nicola Bartolo, Sabino Matarrese Unified Dark Matter Scalar Field Models https://arxiv.org/abs/1008.0614
Статья с обзором моделей темной материи, полистал, пройду основы - будет смысл добавить разные модели, подумать про них. Должна быть полезной, но сейчас слишком слабая теория.

Статьи про основы космологии

Я. В. Зельдович Космологическая Постоянная И Теория Элементарных Частиц
Выглядит как очень хорошая статья, так что если захочу про основы космологии узнать лучше, дня 2 ее только и буду читать.

Shawn Westmoreland Energy Conditions And Scalar Field Cosmology
Там много условий на энергии есть и разные детали, интересно будет посмотреть на то, какие условия что говорят. Пока забил, потому что отдельно рыться долго в этих основах, займусь только после прохождения базовых курсов и понимании основ космологии.

Статьи про инфляцию и раннюю Вселенную

Robert Brandenberger, Jurg Frohlich, Ryo Namba Unifed Dark Matter, Dark Energy and Baryogenesis via a Cosmological Wetting Transition arXiv:1907.06353v2
Насколько понял, порешали динамику двух скалярных полей в спонтанно нарушенном симметрией потенциале, вывели фазовый переход. Звучит интересно, не уверен, насколько модель полезная, но как проработаю обычные разделы по космологии и теории поля - поизучаю, если захочу.

Статьи про CMB

[?] Kogut, A. et al. Dipole Anisotropy in the COBE Differential Microwave Radiometers First-Year Sky Maps
По идее тут ответ про то, что такое диполь СМБ. пока не знаю, но чуть что скорее всего ее и буду читать.

Статьи про другие эффекты

Job Feldbruggea Rien van de Weygaertb Johan Hiddingb,c Joost Feldbrugged Caustic Skeleton & Cosmic Web
крутая статья про микроволновой фон и теорию катастроф, скорее всего потом добавлю из неё темы и сюда и в математику. Второстепенно, но выглядит очень круто.

[?] Husain, Viqar and Tasic, Vladimir An Indeterminate Universe: Dark Energy and Norton's Dome
Статья про неопределенность вселенной, темную материю. звучит интересно, в школе по космологии на нее ссылка была.

[?] Martin, Jerome Everything you always wanted to know about the cosmological constant problem (but were afraid to ask
Очень много обзора космологической постоянной, возможно, показан ее вывод из стандартной модели.

[?] Job Feldbrugge, Jean-Luc Lehners, Neil Turok Lorentzian Quantum Cosmology
Что-то написано про квантовую космологию, особо ничего не понял, не уверен, что статья полезная, но тем не менее, может быть когда-то почитаю и пойму её.

По квантовой космологии

(вроде их много, мб дойду до этого, но очень не скоро)

A.4.3 Additional, Popular, Literature for Help

Другая обучающая и методическая

[?] Вайнберг С. Космология
много всего, потом им дополню когда-то то, что уже тут написал.

Tom Theuns Physical Cosmology http://icc.dur.ac.uk/~tt/Lectures/UA/L4/cosmology.pdf
Еще одни лекции, где другим деталям уделено внимание, так что думаю, что посидеть над ними хотя бы 4 дня точно стоит того. Выглядит, как очень неплохой вариант изложения космологии.

Galsxies and Cosmology Ay 21 lectures George Djorgovski, Caltech https://sites.astro.caltech.edu/~george/ay21/ https://www.youtube.com/watch?v=pUbLsA6g1us&list=PLdoUUpjpM9818MBicKyIuXebavZggG&index=2&ab_channel=ArindamKumarChatterjee
Очень крутые лекции, объясняющие, что же происходит в обычной теории Баумана например. Интересные, и очень углубленные в приложения и астронию лекции, теоретики такого вообще не проходят. Разве что это отдельный большой курс, так что за выходные его не пройти, даже если прошел обычную космологию. Также на сайте курса много хороших ссылок на статьи и т.п.

Theoretical AstroPhysics Including Relativity and Cosmology at Caltech Physics 217bc: The Standard Model - Cosmology. http://www.tapir.caltech.edu/~chirata/ph217/
Крутые компактные лекции, где много интересных тем, которых нет в других курсах, полезно тоже досмотреть их.

[?] Davis, Tamara M. and Lineweaver, Charles H. Expanding Confusion: Common Misconceptions of Cosmological Horizons and the Superluminal Expansion of the Universe
25 страниц, обсуждается расширение Вселенной, много нетривиального рассказывается, очень полезная для понимания этого важного вопроса.

Wannie Hu's lectures http://background.uchicago.edu/~whu/Courses/
Тоже много хорошего изложения типичных тем.

[?] PhD thesis Wandering in the Background: A Cosmic Microwave Background Explorer Wannie Hu http://background.uchicago.edu/~whu/thesis/thesis.pdf
200 страниц углубленной теории, я даже не выгружал, потому что мне такие детали не нужны, основы еще доучивать нужно, ну да и нет пока проектов по космологии.

Stackexchange questions https://physics.stackexchange.com/questions/tagged/cosmology
It can be a good practice and it was a source of some questions in my section of problems.

Физическая в помощь

Мизнер Торн Уиллер т. 1
Наверное, крутая книга, очень подробная, очень многое прописать нужно, прежде чем будет потребность в этих деталях.

Мизнер Торн Уиллер т. 2
Наверное, крутая книга, очень подробная, очень многое прописать нужно, прежде чем будет потребность в этих деталях.

Мизнер Торн Уиллер т. 3
Наверное, крутая книга, очень подробная, очень многое прописать нужно, прежде чем будет потребность в этих деталях.

Популярные и почти популярные книги

[**?**] Хокинг С., Пенроуз Природа пространства и времени
Название крутое, темы интересные, не уверен, что что-то полезное будет написано. После основных тем только посмотрю.
Barbara Ryden Introduction to Cosmology
Наверное, хорошая книга, но не знаю, зачем мне она, потому что вроде все это уже в других книгах обсуждается. Может, когда-то почитаю, дополню ей свою запись.

A.4.4 Suggested Readings from Good Books

SUGGESTED READING EINSTEIN EQUATIONS

Most cosmology books offer some treatment of the perturbed Einstein equations in cosmology. Again The Large Scale Structure of the Universe (Peebles) is a useful reference, especially for synchronous gauge. Cosmological Inflation and Large Scale Structure (Liddle and Lyth) has a very nice treatment which, among other virtues, explains the physics of gauge choices. Probably the two most comprehensive works are the review articles by Mukhanov, Feldman, and Brandenberger (1992) and Kodama and Sasaki (1984), with the former slightly more accessible and the latter more general. These are both based on the seminal Bardeen (1980) article which is remarkable for its clarity and conciseness in its treatment of gauge invariant variables.

The general relativity books mentioned in Chapter 2 all have good discussions of gravity waves. Before turning to any of the technical literature, though, you must read Black Holes and Time Warps (Thorne), a wonderful mixture of the history, science, and personalities associated with 20th-century general relativity. It is the best popular science book I have ever read.

SUGGESTED READING SCALAR PERTURBATIONS

The 30 or so pages on inflation in this chapter, which were heavily slanted toward production of perturbations, offer but a glimpse into the many facets of this remarkable theory. Recently, Guth wrote a popular account of his discovery of inflation, The Inflationary Universe. One of the other originators of the theory, Linde, has a more technical book, Inflation and Quantum Cosmology, which emphasizes model building much more than I have here. As I mentioned earlier, The Early Universe (Kolb and Turner) has an excellent chapter on inflation. The recent Cosmological Inflation and Large Scale Structure (Liddle and Lyth) is most similar in spirit to this book, with a heavy emphasis on perturbations. The discussion there of the perturbation spectrum is laden with less algebra than the one in Section 6.5 so is worth reading. (Beware that their Planck mass is our $m_{\rm Pl}/\sqrt{8\pi}$.)

An extremely clear and deep look into inflation is given in 300 Years of Gravitation (ed. Hawking and Israel) in the article by Blau and Guth. Many other articles in that thick compilation volume are also fascinating. The initial article by Guth (1981) is completely accessible and as clear a statement possible of the problems that led to inflation and the initial attempt (old inflation) to solve them. Indeed, I would recommend reading Guth's initial article because this chapter motivates inflation with the horizon problem, while Guth had several different problems in mind, including the monopole problem and the flatness problem (Exercise 4).

There have been many papers reviewing the production of perturbations during inflation. Two clear reviews are Lidsey et al. (1997) and Lyth and Riotto (1999). The former focuses on methods for going beyond the predictions elucidated here, which are accurate only to first order in the slow-roll parameters ϵ and δ, and on extracting the potential V from observations. The latter summarizes efforts to tie inflation to realistic particle physics models. The eight-page paper of Stewart and Lyth (1993) is a remarkably concise treatment of the techniques used to go beyond the first-order slow-roll approximation. Hollands and Wald (2002) have written a thoughtful critique of inflation, which is a refreshing antidote to some of the euphoria emanating from the discoveries of the late 1990s. Besides the importance of this critique in its own right, the paper has one of the clearest qualitative descriptions of perturbation generation during inflation that I have ever read.

The initial conditions relating the various perturbations described in Section 6.1 are perhaps most clearly discussed in the review article by Efstathiou (1990). Isocurvature perturbations, for the most part ignored here, are treated in detail there.

I have ignored the possibility of perturbations produced by topological defects. These theories, while fascinating, have not succeeded in making robust predictions; to the extent that predictions can be extracted from them, they are wrong. Nonetheless the numerics involved in their study is sufficiently complicated that I would not be shocked to see them make a comeback some day. There exist many books with comprehensive discussions of topological defects. Amorig them are Cosmic Strings and Other Topological Defects (Vilenkin and Shellard) and The Formation and Evolution of Cosmic Strings (ed. Gibbons, Hawking, and Vachaspati).

SUGGESTED READING PROBES OF INHOMOGENEITIES

The Large Scale Structure of the Universe (Peebles) has a good description and derivation of the angular correlation function, using Limber's (1953) original derivation. The derivation given in Section 9.1 is based on the Appendix of Kaiser's (1992) work on weak lensing. This derivation has the advantage of being physically intuitive and also generally applicable. It will allow us to easily compute the weak lensing correlation functions we encounter in Chapter 10.

The data discussed in Section 9.1 come from the APM Survey (Maddox et al., 1990). Other recent angular results of note include the Edinburgh/Durham Southern Galaxy Catalogue (EDSGC, Collins, Nichol, and Lumsden, 1992) and the Sloan Digital Sky Survey (York et al., 2000), the first analysis of which is presented in Scranton et al. (2002). As I hinted in the text, analysis of these data sets has gotten progressively more sophisticated over time. Baugh and Efstathiou (1993) first inverted the APM angular correlation function and extracted the 3D power spectrum, confirming that it was quite different from the standard CDM power spectrum on large scales. Dodelson and Gaztanaga (2000) pointed out that the resultant errors on the power spectrum are correlated and that an accurate treatment would also account for the fact that the errors in $w(\theta)$ are also correlated. Eisenstein and Zaldarriaga (2001) and then Efstathiou and Moody (2001) accounted for these correlations leading to the softened conclusions summarized in Figure 9.6. The Eisenstein and Zaldarriaga paper also contains a clear discussion of the relation between $w(\theta)$ and the 2D power spectrum, or the C_l 's. They observed that errors on the C_l 's are much less correlated, which leads me to believe that C_l 's will ultimtely replace $w(\theta)$ as the statistic of choice for angular surveys. Indeed, Huterer, Knox, and Nichol (2001) have analyzed the EDSGC survey with C_l 's, and Tegmark et al. (2002) obtained C_l 's from early SDSS data.

Cosmological Physics (Peacock) is a good resource for peculiar velocities and their effect on redshift surveys. Two important and informative review articles are Strauss and Willick (1995) and Hamilton (1998), the former particularly good for experimental issues involved in determining peculiar velocities and the latter for analyzing galaxy surveys in the presence of redshift space distortions. Another leader in the field, Dekel (1997), has written a good review of the cosmological implications of the peculiar velocity field.

The seminal work on redshift space distortions is by Kaiser (1987), who solved the problem working in Fourier space for linear distortions in the distant-observer, low redshift approximation. Hamilton (1992) found the analogue of this solution in real space. Recently, with large-area, relatively deep surveys coming on line, the generalization for cosmological corrections (i.e., distance is not equal to cz/H_0 as z gets large) and all-sky analysis has been carried out by a number of authors. As examples of the work currently going on in the field, Szalay, Matsubara, and Landy (1998) generalized Kaiser's work to large angles while Magira, Ying, and Suto (2000) accounted for cosmological distortions and also nonlinearities and evolution.

Peacock and Dodds (1994) analyzed a variety of surveys, arguing that - accounting for redshift space distortions, nonlinearities, and bias properly - the power spectrum has been well measured. Undoubtedly, with the upcoming Sloan Digital Sky Survey, Two Degree Field, and others, more will be learned about the power spectrum in coming years. The semianalytic work referred to on Page 283 is based on the seminal papers of White and Rees (1978) and White and Frenk (1991). Related, but separate form these semianalytic models, is the halo model (reviewed by Cooray and Sheth, 2002) which postulates that all the dark matter is in halos, thereby reducing the clustering problem to (i) the clustering of the halos and (ii) the distribution of matter and galaxies within the halo. Good descriptions of what the halo model is and how it can be used to compare theories with redshift surveys can be found in White (2001); Seljak (2000); and Berlind and Weinberg (2001).

The prediction of the cluster abundance is treated nicely in Cosmological Physics (Peacock) and Cosmological Inflation and Large Scale Structure (Liddle and Lyth). Structure Formation in the Universe (Padmanabhan) has a detailed section on the spherical collapse model, which is the source of the numbers 1.686 and 178 in Section 9.5. In addition to the papers cited in the text, some important cluster normalization papers are Viana and Liddle (1996, 1999) and Pierpaoli, Scott, and White (2001).

SUGGESTED READING WEAK LENSING AND POLARIZATION

Gravitational lensing is described in exquisite detail in Gravitational Lenses (Schneider, Ehlers, and Falco). An excellent, comprehensive review of weak lensing is in Bartelmann and Schneider (2001). Electromagnetic polarization is a textbook subject, covered in, for example, Classical Electrodynamics (Jackson) and Radiative Processes in Astrophysics (Rybicki and Lightman). Initial papers on gravitational lensing by large scale structure include Blandford et al. (1991), Miralda-Escude (1991), and Kaiser (1992).

Recent theoretical work connecting lensing observations to cosmological parameters includes Jain and Seljak (1997); Bernardeau, van Waerbecke, and Mellier (1997); and Hu and Tegmark (1999). The onset of the new millenium saw the first detections of lensing by large scale structure in van Waerbecke et al. (2000); Wittman et al. (2000); Bacon, Refrgier, and Ellis (2000); Kaiser, Wilson and Luppino (2000); and Maoli et al. (2001). Active work continues. The future will undoubtedly bring observations of weak lensing on large fields. Two proposal for such observations are the SuperNova Acceleration Probe (SNAP; http://snap.Ibl.gov), which presently plans to devote $\sim$ 20% of its time to weak lensing, and the Large Scale Synaptic Telescope (LSST; http://ww. dmtelescope.org).

Polarization of the CMB was studied in the seminal papers of the 1980 s by Bond and Efstathiou (1984) and Polnarev (1985). Kosowsky's thesis (1996) is a lucid Boltzmann-esque discussion of this work. The first papers to recognize the importance of the E/B decomposition were Stebbins (1996); Seljak (1997);

Kamionkowski, Kosowsky, and Stebbins (1997a,b), and Zaldarriaga and Seljak (1997). I've followed the treatment of Seljak who worked in the small angle limit because the algebra is simpler. The review article of Hu and White (1997b) is perhaps the most accessible introduction into the recent literature on polarization of the CMB, but it is a difficult subject. Even this lucid review with its illuminating pictures requires a lot of effort to understand.

DASI (Kovac et al., 2002) detected polarization at the 5-sigma level. Previous stringent upper limits, which are still valuable on large scales, were obtained by Hedman et al. (2000) and Keating et al. (2001).

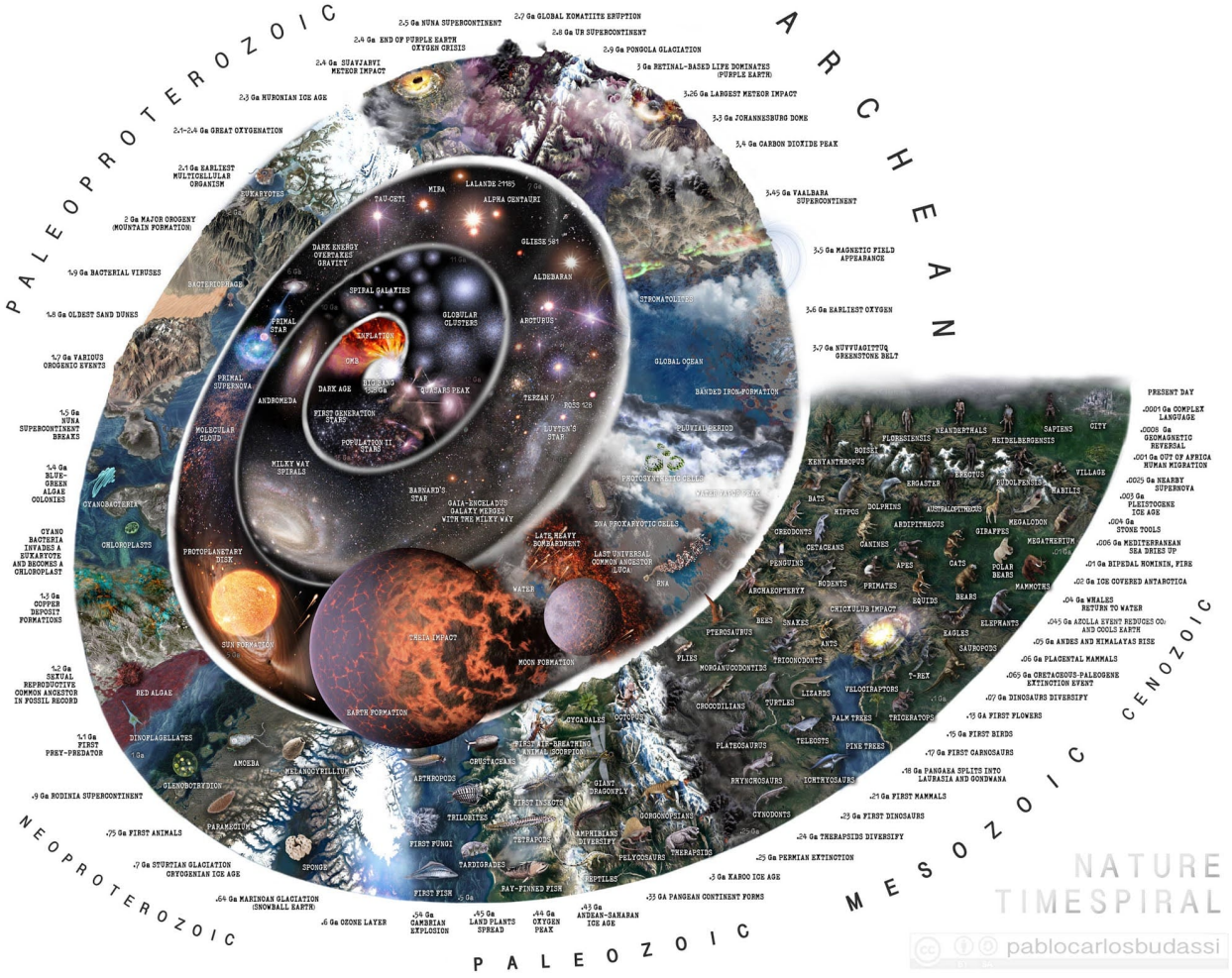
A.5 History of Cosmology in a Nutshell (???)

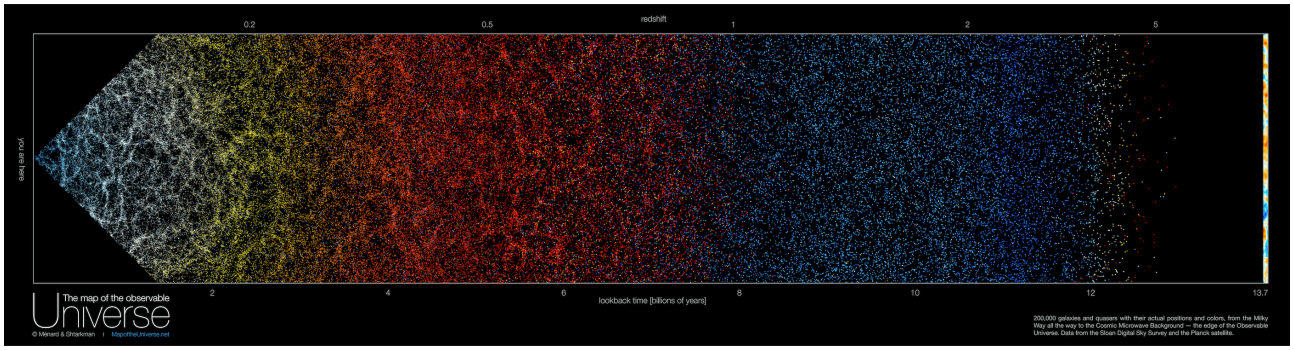
(хз пока что)

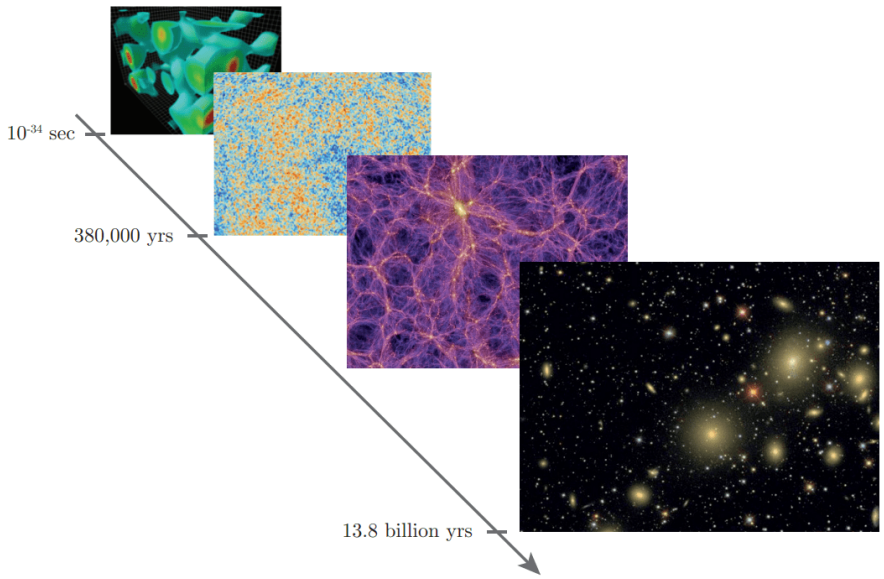
A.6 Overview of Cosmology

A.6.1 Popular Overview of Cosmology

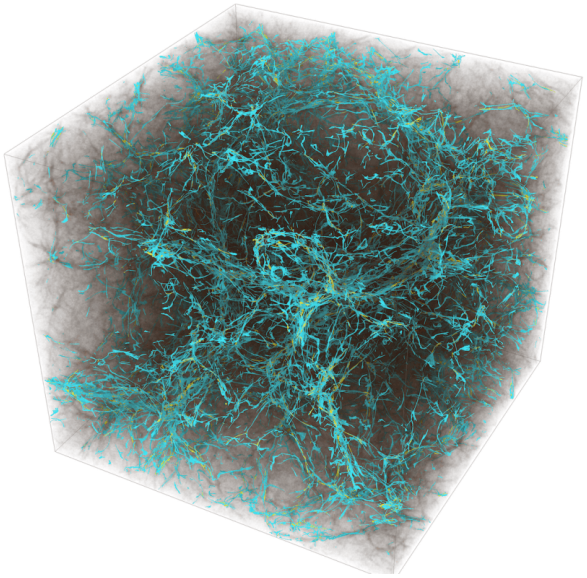
Общие закономерности



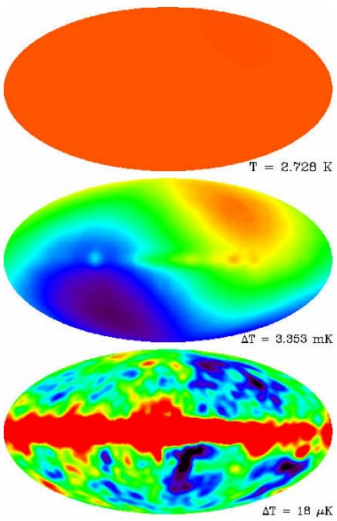




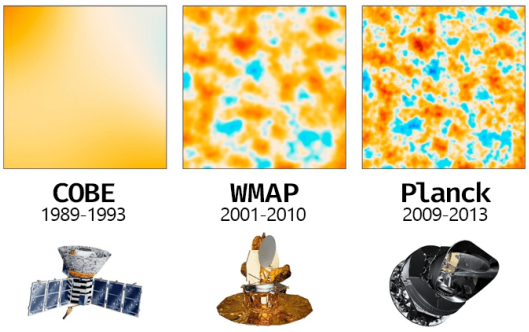
(абзац с пояснениями)



Популярно о CMB



Популярно об экспериментах для изучения космологии



A.6.2 Results of Cosmology

(??? summarize it later)

A.6.3 Applications of Cosmology in Other Branches of Physics

(?? a good question)

A.6.4 Review of Further Developments in Cosmology

(??? which questions people want to answer and why???)

A.6.5 What Is Going on in Typical Sections of Cosmology

Описание квантовой космологии

мотивация

обзор квантовой космологии

отличие от классической

A.6.6 Some Potential Applications

обсудим, какие темы на самом деле самые важные тут.

Какую можно игру компьютерную сделать про космологию (!?!??!??!?!?)

(это ведь интересно, дать возможность автору как-то влиять на вселенную, чтобы он что-то чуть менял - и смотреть, как там динамика становится другой. каике еще игры можно придумать?)

A.7 Description of the Note

A.7.1 Description of Chapters and Sections

(??? I'll write later)

описание записи в целом

первая часть

вторая часть

приложения

какие вообще приложения я разбирал?

A.7.2 Designations and Constants

Константы

1 s	$= 9.7157 \times 10^{-15}$Mpc
1yr	$= 3.1558 \times 10^{7}$ s
1Mpc	$= 3.0856 \times 10^{24}$ cm
1AU	$= 1.4960 \times 10^{13}$ cm
1 K	$= 8.6170 \times 10^{-5}$eV
$1M_{\odot}$	$= 1.989 \times 10^{33}$ g
1GeV	$= 1.6022 \times 10^{-3}$erg
	$= 1.7827 \times 10^{-24}$ g
	$= (1.9733 \times 10^{-14}$ cm$)^{-1}$
	$= (6.5821 \times 10^{-25}$ s$)^{-1}$

Planck's constant	$\hbar = 1.0546 \times 10^{-27}$ cm^2 g s^{-1}
Speed of light	$c = 2.9979 \times 10^{10}$ cm s^{-1}
Boltzmann's constant	$k_B = 1.3807 \times 10^{-16}$ergK^{-1}
Fine structure constant	$\alpha = 1/137.036$
Gravitational constant	$G = 6.6720 \times 10^{-8}$ cm^3 g^{-1} s^{-2}
Stefan-Boltzmann constant	$\sigma = ac/4 = \pi^2 k_B^4/60\hbar^3 c^2$
	$a = 7.5646 \times 10^{-15}$ergcm^{-3} K^{-4}
Thomson cross section	$\sigma_T = 8\pi\alpha^2/3m_e^2 = 6.6524 \times 10^{-25}$ cm^2
Electron mass	$m_e = 0.5110$MeV
Neutron mass	$m_n = 939.566$MeV
Proton mass	$m_p = 938.272$MeV

A.8 Cosmological Puzzles for Different Cases

сбалансирую и наберу потом по интересности

A.8.1 Puzzles about Construction of Cosmology

Идейно, как описать Вселенную?

Есть метрика, записываю то есть уравнения Фридамана.

A.8.2 General Questions about the Universe

является ли классическая физика и космология детерминированной?

(а вот я не знаю, потом займусь ответом, тут пока выгрузка из школы по космологии)
Let us consider a simple mechanical model: Norton's Dome.
A classical point particle finds itself at some maximum of the potential

$$V = V_0 - V_1 x^{\frac{3}{2}}, \quad V_0 > 0, \quad V_1 > 0$$

To such a potential corresponds a surface which is not spherical but has some kind of cusp at the top.

The second Newton law for the particle is

$$\ddot{x} = \frac{3}{2} V_1 x^{\frac{1}{2}}$$

...

The Friedmann equations for a flat universe filled with a perfect fluid with

$$p = w\rho, -1 > w > -\frac{2}{3}$$

are written in such a form that they are nonsingular at $a = 0$ and permit both the solutions $a(t) = 0$ and

Что есть настоящее расширение Вселенной?

Говорят, это расширение метрики, но о многом тут можно подумать.

Почему вселенная расширяется?

где граница вселенной?

A.8.3 General Questions about What Cosmologists are Doing

Почему космология как наука вообще возможна

(Кант утверждал, что её нет, потом порассуждаю про это)

О возникновении всего из ничего

(тоже типичный вопрос)

A.8.4 Questions for Those Who Say They Have Mastered Cosmology

В чем состоит интегральный эффект Сакса-Вульфа? (?????)

(не знаю, там много нужно знать, чтобы до него дойти в принципе)

B Some Physics for Cosmology

B.0.1 Some General Theory of Relativity

(??)

B.0.2 Some Statistical Physics

(??)

B.0.3 Some Nuclear Physics

(???)

B.0.4 Some Particle Physics

(???)

B.0.5 Some Other Physics

(????)

C Some Mathematics for Cosmology

(???? I know some methods, but I am not sure, how used are they, so I'll write this only later)

D Bibliography

References

D.0.1 From Lectures

References from Lesgourgues TASI Lectures on perturbations

References

[1] J. Lesgourgues, G. Mangano, G. Miele and S. Pastor, *Neutrino Cosmology* (Cambridge Univ. Press, Cambridge, 2013).

[2] U. Seljak and M. Zaldarriaga, *Astrophys.J.* **469**, 437 (1996).

[3] S. Weinberg, *Astrophys.J.* **581**, 810 (2002).

[4] S. Weinberg, *Cosmology* (Oxford Univ. Press, Oxford, 2008).

[5] D. J. Eisenstein and W. Hu, *Astrophys.J.* **496**, p. 605 (1998).

[6] C.-P. Ma and E. Bertschinger, *Astrophys.J.* **455**, 7 (1995).

[7] S. Dodelson, *Modern cosmology* (Academic Press, San Diego, CA, 2003).

[8] V. Mukhanov, *Physical Foundations of Cosmology* (Cambridge Univ. Press, Cambridge, 2005).

[9] W. T. Hu, *Wandering in the Background: A CMB Explorer*, astro-ph/9508126, (1995).

[10] R. Durrer, *The Cosmic Microwave Background* (Cambridge Univ. Press, Cambridge, 2008).

References from Baumann TASI lectures

References

[1] A. H. Guth, Phys. Rev. **D23**, 347 (1981).

[2] A. D. Linde, Phys. Lett. **B108**, 389 (1982).

[3] A. Albrecht and P. J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).

[4] E. Hubble, Proc. Nat. Acad. Sci. **15**, 168 (1929).

[5] R. A. Alpher, H. Bethe, and G. Gamow, Phys. Rev. **73**, 803 (1948).

[6] R. Dicke, P. J. E. Peebles, P. Roll, and D. Wilkinson, Astrophys. J. **142**, 414 (1965).

[7] CMBPol Study Team, D. Baumann *et al.*, (2008), 0811.3919.

[8] S. Dodelson, *Modern Cosmology* (Amsterdam, Netherlands: Academic Press, 2003).

[9] V. Mukhanov, *Physical Foundations of Cosmology* (Cambridge University Press, 2005).

[10] S. Weinberg, *Cosmology* (Oxford University Press, 2008).

[11] WMAP, E. Komatsu *et al.*, (2008), 0803.0547.

[12] M. Kowalski *et al.*, (2008), 0804.4142.

[13] B. A. Bassett, S. Tsujikawa, and D. Wands, Rev. Mod. Phys. **78**, 537 (2006), astro-ph/0507632.

[14] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, JHEP **03**, 014 (2008), 0709.0293.

[15] S. Weinberg, Phys. Rev. **D77**, 123541 (2008), 0804.4291.

[16] D. Baumann and L. McAllister, (2009), 0901.0265.

[17] Q. Shafi and V. N. Senoguz, Phys. Rev. **D73**, 127301 (2006), astro-ph/0603830.

[18] D. H. Lyth and A. Riotto, Phys. Rept. **314**, 1 (1999), hep-ph/9807278.

[19] D. Wands, Lect. Notes Phys. **738**, 275 (2008), astro-ph/0702187.

[20] J. M. Bardeen, Phys. Rev. **D22**, 1882 (1980).

[21] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, Phys. Rev. **D28**, 679 (1983).

[22] D. Wands, K. A. Malik, D. H. Lyth, and A. R. Liddle, Phys. Rev. **D62**, 043527 (2000), astro-ph/0003278.

[23] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, Nucl. Phys. **B667**, 119 (2003), astro-ph/0209156.

[24] J. M. Maldacena, JHEP **05**, 013 (2003), astro-ph/0210603.

[25] N. D. Birrell and P. C. W. Davies, Cambridge, Uk: Univ. Pr.)1982) 340p.

[26] T. Jacobson, gr-qc/0308048.

[27] D. H. Lyth, Phys. Rev. Lett. **78**, 1861 (1997), hep-ph/9606387.

[28] W. Hu, `http://background.uchicago.edu/ whu/`.

[29] U. Seljak and M. Zaldarriaga, astro-ph/9603033, `http://www.cmbfast.org`.

[30] A. Lewis and A. Challinor, `http://camb.info`.

[31] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, Phys. Rept. **402**, 103 (2004), astro-ph/0406398.

[32] M. Kamionkowski, A. Kosowsky, and A. Stebbins, Phys. Rev. **D55**, 7368 (1997), astro-ph/9611125.

[33] M. Zaldarriaga and U. Seljak, Phys. Rev. **D55**, 1830 (1997), astro-ph/9609170.

[34] E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

[35] D. N. Spergel and M. Zaldarriaga, Phys. Rev. Lett. **79**, 2180 (1997).

[36] S. Dodelson, AIP Conf. Proc. **689**, 184 (2003), hep-ph/0309057.

[37] J. Bock *et al.*, (2008), 0805.4207.

[38] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, Astrophys. J. **304**, 15 (1986).

[39] D. J. Eisenstein and W. Hu, Astrophys. J. **496**, 605 (1998), astro-ph/9709112.

[40] P. J. Steinhardt and N. Turok, Science **296**, 1436 (2002).

[41] P. J. Steinhardt and N. Turok, Phys. Rev. **D65**, 126003 (2002), hep-th/0111098.

[42] E. I. Buchbinder, J. Khoury, and B. A. Ovrut, Phys. Rev. **D76**, 123503 (2007), hep-th/0702154.

[43] K. M. Smith, L. Senatore, and M. Zaldarriaga, (2009), 0901.2572.

[44] S. Dimopoulos, S. Kachru, J. McGreevy, and J. G. Wacker, (2005), hep-th/0507205.

[45] L. McAllister, E. Silverstein, and A. Westphal, (2008), 0808.0706.

[46] E. Silverstein and A. Westphal, (2008), 0803.3085.

[47] S. Kachru *et al.*, JCAP **0310**, 013 (2003), hep-th/0308055.

[48] D. Baumann, A. Dymarsky, I. R. Klebanov, and L. McAllister, JCAP **0801**, 024 (2008), 0706.0360.

[49] D. Baumann, A. Dymarsky, S. Kachru, I. R. Klebanov, and L. McAllister, JHEP **03**, 093 (2009), 0808.2811.

[50] J. P. Conlon and F. Quevedo, JHEP **01**, 146 (2006), hep-th/0509012.

[51] J. J. Blanco-Pillado *et al.*, JHEP **09**, 002 (2006), hep-th/0603129.

[52] D. Baumann and M. Zaldarriaga, (2009), 0901.0958.

[53] D. Babich, P. Creminelli, and M. Zaldarriaga, JCAP **0408**, 009 (2004), astro-ph/0405356.

[54] N. Bartolo, S. Matarrese, and A. Riotto, Phys. Rev. **D65**, 103505 (2002), hep-ph/0112261.

[55] F. Bernardeau and J.-P. Uzan, Phys. Rev. **D66**, 103506 (2002), hep-ph/0207295.

[56] F. Bernardeau and J.-P. Uzan, Phys. Rev. **D67**, 121301 (2003), astro-ph/0209330.

[57] M. Sasaki, Prog. Theor. Phys. **120**, 159 (2008), 0805.0974.

[58] A. Naruko and M. Sasaki, (2008), 0807.0180.

[59] C. T. Byrnes, K.-Y. Choi, and L. M. H. Hall, (2008), 0807.1101.

[60] C. T. Byrnes and D. Wands, Phys. Rev. **D74**, 043529 (2006), astro-ph/0605679.

[61] D. Langlois, F. Vernizzi, and D. Wands, (2008), 0809.4646.

[62] J. Valiviita, H. Assadullahi, and D. Wands, (2008), 0806.0623.

[63] H. Assadullahi, J. Valiviita, and D. Wands, Phys. Rev. **D76**, 103003 (2007), 0708.0223.

[64] J. Valiviita, M. Sasaki, and D. Wands, (2006), astro-ph/0610001.

[65] F. Vernizzi and D. Wands, JCAP **0605**, 019 (2006), astro-ph/0603799.

[66] L. E. Allen, S. Gupta, and D. Wands, JCAP **0601**, 006 (2006), astro-ph/0509719.

[67] A. D. Linde and V. F. Mukhanov, Phys. Rev. **D56**, 535 (1997), astro-ph/9610219.

[68] D. H. Lyth, C. Ungarelli, and D. Wands, Phys. Rev. **D67**, 023503 (2003), astro-ph/0208055.

[69] G. Dvali, A. Gruzinov, and M. Zaldarriaga, Phys. Rev. **D69**, 023505 (2004), astro-ph/0303591.

[70] L. Kofman, (2003), astro-ph/0303614.

[71] P. Creminelli and L. Senatore, JCAP **0711**, 010 (2007), hep-th/0702165.

[72] K. Koyama, S. Mizuno, F. Vernizzi, and D. Wands, JCAP **0711**, 024 (2007), 0708.4321.

[73] E. I. Buchbinder, J. Khoury, and B. A. Ovrut, Phys. Rev. Lett. **100**, 171302 (2008), 0710.5172.

[74] J.-L. Lehners and P. J. Steinhardt, Phys. Rev. **D77**, 063533 (2008), 0712.3779.

[75] J.-L. Lehners and P. J. Steinhardt, Phys. Rev. **D78**, 023506 (2008), 0804.1293.

[76] K. Koyama, S. Mizuno, and D. Wands, Class. Quant. Grav. **24**, 3919 (2007), 0704.1152.

[77] K. Koyama and D. Wands, JCAP **0704**, 008 (2007), hep-th/0703040.

[78] M. Alishahiha, E. Silverstein, and D. Tong, Phys. Rev. **D70**, 123505 (2004), hep-th/0404084.

[79] N. Arkani-Hamed, P. Creminelli, S. Mukohyama, and M. Zaldarriaga, JCAP **0404**, 001 (2004), hep-th/0312100.

[80] L. Senatore, Phys. Rev. **D71**, 043512 (2005), astro-ph/0406187.

[81] P. Creminelli, JCAP **0310**, 003 (2003), astro-ph/0306122.

[82] X. Chen, M.-x. Huang, S. Kachru, and G. Shiu, JCAP **0701**, 002 (2007), hep-th/0605045.

[83] C. Gordon, D. Wands, B. A. Bassett, and R. Maartens, Phys. Rev. **D63**, 023506 (2001), astro-ph/0009131.

[84] D. H. Lyth, C. Ungarelli, and D. Wands, Phys. Rev. **D67**, 023503 (2003), astro-ph/0208055.

[85] S. Weinberg, Phys. Rev. **D70**, 083522 (2004), astro-ph/0405397.

[86] E. Komatsu *et al.*, (2009), 0902.4759.

[87] J. S. Schwinger, J. Math. Phys. **2**, 407 (1961).

[88] E. Calzetta and B. L. Hu, Phys. Rev. **D35**, 495 (1987).

[89] R. D. Jordan, Phys. Rev. **D33**, 444 (1986).

[90] S. Weinberg, Phys. Rev. **D72**, 043514 (2005), hep-th/0506236.

[91] E. Komatsu and D. N. Spergel, Phys. Rev. **D63**, 063002 (2001), astro-ph/0005036.

[92] J. R. Fergusson and E. P. S. Shellard, (2008), 0812.3413.

[93] E. Silverstein and D. Tong, Phys. Rev. **D70**, 103505 (2004), hep-th/0310221.

[94] R. Holman and A. J. Tolley, JCAP **0805**, 001 (2008), 0710.1302.

[95] P. Creminelli and M. Zaldarriaga, JCAP **0410**, 006 (2004), astro-ph/0407059.

[96] C. Cheung, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, JCAP **0802**, 021 (2008), 0709.0295.

[97] L. Senatore, K. M. Smith, and M. Zaldarriaga, (2009), 0905.3746.

[98] N. Bartolo, S. Matarrese, and A. Riotto, JCAP **0606**, 024 (2006), astro-ph/0604416.

[99] N. Bartolo, S. Matarrese, and A. Riotto, JCAP **0701**, 019 (2007), astro-ph/0610110.

[100] L. Senatore, S. Tassev, and M. Zaldarriaga, (2008), 0812.3658.

[101] N. Dalal, O. Dore, D. Huterer, and A. Shirokov, Phys. Rev. **D77**, 123514 (2008), 0710.4560.

[102] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, (2008), 0805.3580.

[103] E. J. Copeland, A. R. Liddle, D. H. Lyth, E. D. Stewart, and D. Wands, Phys. Rev. **D49**, 6410 (1994), astro-ph/9401011.

[104] J. Polchinski, *String Theory. Vol. 1: An Introduction to the Bosonic String* (Cambridge University Press, 1998).

[105] J. Polchinski, *String Theory. Vol. 2: Superstring Theory and Beyond* (Cambridge University Press, 1998).

[106] S. Kachru, R. Kallosh, A. Linde, and S. P. Trivedi, Phys. Rev. **D68**, 046005 (2003), hep-th/0301240.

[107] D. H. Lyth, Phys. Rev. Lett. **78**, 1861 (1997), hep-ph/9606387.

[108] A. D. Linde, Phys. Lett. **B129**, 177 (1983).

[109] C. Vafa, (2005), hep-th/0509212.

[110] H. Ooguri and C. Vafa, Nucl. Phys. **B766**, 21 (2007), hep-th/0605264.

[111] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, JHEP **10**, 014 (2006), hep-th/0602178.

[112] M. P. Hertzberg, M. Tegmark, S. Kachru, J. Shelton, and O. Ozcan, Phys. Rev. **D76**, 103521 (2007), 0709.0002.

[113] M. R. Douglas and S. Kachru, Rev. Mod. Phys. **79**, 733 (2007), hep-th/0610102.

[114] I. R. Klebanov and M. J. Strassler, JHEP **08**, 052 (2000), hep-th/0007191.

[115] I. R. Klebanov and A. A. Tseytlin, Nucl. Phys. **B578**, 123 (2000), hep-th/0002159.

[116] S. B. Giddings, S. Kachru, and J. Polchinski, Phys. Rev. **D66**, 106006 (2002), hep-th/0105097.

[117] D. Baumann and L. McAllister, Phys. Rev. **D75**, 123508 (2007), hep-th/0610285.

[118] D. Baumann *et al.*, JHEP **11**, 031 (2006), hep-th/0607050.

[119] D. Baumann, A. Dymarsky, I. R. Klebanov, L. McAllister, and P. J. Steinhardt, Phys. Rev. Lett. **99**, 141601 (2007), 0705.3837.

[120] A. Krause and E. Pajer, JCAP **0807**, 023 (2008), 0705.4682.

[121] M. Nakahara, Bristol, UK: Hilger (1990) 505 p. (Graduate student series in physics).

[122] D. Baumann, A. Dymarsky, S. Kachru, I. R. Klebanov, and L. McAllister.

[123] H. Firouzjahi and S. H. H. Tye, JCAP **0503**, 009 (2005), hep-th/0501099.

[124] S. E. Shandera and S. H. H. Tye, JCAP **0605**, 007 (2006), hep-th/0601099.

[125] K. Freese, J. A. Frieman, and A. V. Olinto, Phys. Rev. Lett. **65**, 3233 (1990).

[126] T. Banks, M. Dine, P. J. Fox, and E. Gorbatov, JCAP **0306**, 001 (2003), hep-th/0303252.

[127] P. Svrcek and E. Witten, JHEP **06**, 051 (2006), hep-th/0605206.

[128] R. Easther and L. McAllister, JCAP **0605**, 018 (2006), hep-th/0512102.

[129] A. R. Liddle, A. Mazumdar, and F. E. Schunck, Phys. Rev. **D58**, 061301 (1998), astro-ph/9804177.

[130] R. Kallosh, N. Sivanandam, and M. Soroush, Phys. Rev. **D77**, 043501 (2008), 0710.3429.

[131] T. W. Grimm, Phys. Rev. **D77**, 126007 (2008), 0710.3883.

[132] M. P. Hertzberg, S. Kachru, W. Taylor, and M. Tegmark, JHEP **12**, 095 (2007), 0711.2512.

[133] R. Flauger, S. Paban, D. Robbins, and T. Wrase, (2008), 0812.3886.

[134] C. Caviezel *et al.*, JHEP **04**, 010 (2009), 0812.3551.

[135] S. Sarangi and S. H. H. Tye, Phys. Lett. **B536**, 185 (2002), hep-th/0204074.

[136] E. J. Copeland, R. C. Myers, and J. Polchinski, JHEP **06**, 013 (2004), hep-th/0312067.

[137] C. P. Burgess, PoS **P2GC**, 008 (2006), 0708.2865.

[138] R. Kallosh, Lect. Notes Phys. **738**, 119 (2008), hep-th/0702059.

[139] A. Linde, ECONF **C040802**, L024 (2004), hep-th/0503195.

[140] L. McAllister and E. Silverstein, Gen. Rel. Grav. **40**, 565 (2008), 0710.2951.

[141] J. M. Cline, (2006), hep-th/0612129.

[142] R. Flauger, L. McAllister, E. Pajer, A. Westphal, and G. Xu, (2009), 0907.2916.

[143] Planck, (2006), astro-ph/0604069.

[144] `http://cmbpol.uchicago.edu/workshops/path2009/`.

[145] A. Kosowsky, New Astron. Rev. **47**, 939 (2003), astro-ph/0402234.

[146] SPT, J. E. Ruhl *et al.*, Proc. SPIE Int. Soc. Opt. Eng. **5498**, 11 (2004), astro-ph/0411122.

[147] Clover, A. C. Taylor, New Astron. Rev. **50**, 993 (2006), astro-ph/0610716.

[148] D. Samtleben, Nuovo Cim. **122B**, 1353 (2007), 0802.2657.

[149] K. W. Yoon *et al.*, (2006), astro-ph/0606278.

[150] P. Oxley *et al.*, Proc. SPIE Int. Soc. Opt. Eng. **5543**, 320 (2004), astro-ph/0501111.

[151] C. J. MacTavish *et al.*, (2007), 0710.0375.

[152] EPIC, J. Bock *et al.*, (2009), 0906.1188.

[153] A. Vilenkin, Phys. Rev. **D27**, 2848 (1983).

[154] A. D. Linde, Phys. Lett. **B175**, 395 (1986).

[155] A. D. Linde and A. Mezhlumian, Phys. Lett. **B307**, 25 (1993), gr-qc/9304015.

[156] A. D. Linde, D. A. Linde, and A. Mezhlumian, Phys. Rev. **D49**, 1783 (1994), gr-qc/9306035.

[157] J. Garcia-Bellido, A. D. Linde, and D. A. Linde, Phys. Rev. **D50**, 730 (1994), astro-ph/9312039.

[158] A. Vilenkin, Phys. Rev. Lett. **74**, 846 (1995), gr-qc/9406010.

[159] J. Garriga, D. Schwartz-Perlov, A. Vilenkin, and S. Winitzki, JCAP **0601**, 017 (2006), hep-th/0509184.

[160] R. Bousso, Phys. Rev. Lett. **97**, 191302 (2006), hep-th/0605263.

[161] A. Linde, JCAP **0706**, 017 (2007), 0705.1160.

[162] R. Bousso, B. Freivogel, and I.-S. Yang, Phys. Rev. **D79**, 063513 (2009), 0808.3770.

[163] A. De Simone *et al.*, (2008), 0808.3778.

[164] J. Garriga and A. Vilenkin, JCAP **0901**, 021 (2009), 0809.4257.

[165] S. Winitzki, Phys. Rev. **D78**, 123518 (2008), 0810.1517.

[166] A. Linde, V. Vanchurin, and S. Winitzki, JCAP **0901**, 031 (2009), 0812.0005.

[167] K. A. Malik and D. Wands, (2008), 0809.4944.

[168] L. Verde, (2007), 0712.3028.

[169] R. L. Arnowitt, S. Deser, and C. W. Misner, (1962), gr-qc/0405109.

[170] D. Seery, JCAP **0711**, 025 (2007), 0707.3377.

[171] P. Adshead, R. Easther, and E. A. Lim, (2009), 0904.4207.

References from Baumann TASI another lectures

References

[1] D. Baumann, *Cosmology* (http://cosmology.amsterdam).

[2] S. Dodelson, *Modern Cosmology*. Academic Press, 2003.

[3] P. Peter and J.-P. Uzan, *Primordial Cosmology*. Oxford University Press, 2013.

[4] B. Follin, L. Knox, M. Millea, and Z. Pan, "First Detection of the Acoustic Oscillation Phase Shift Expected from the Cosmic Neutrino Background," *Phys. Rev. Lett.* **115** no. 9, (2015) 091301, arXiv:1503.07863 [astro-ph.CO].

[5] D. Baumann, D. Green, J. Meyers, and B. Wallisch, "Phases of New Physics in the CMB," *JCAP* **1601** (2016) 007, arXiv:1508.06342 [astro-ph.CO].

[6] D. Baumann, "Inflation," in *Physics of the Large and the Small, TASI 09*, pp. 523–686. 2011. arXiv:0907.5424 [hep-th].

[7] D. Baumann, *Advanced Cosmology* (http://cosmology.amsterdam).

[8] R. Durrer, *The Cosmic Microwave Background*. Cambridge University Press, Cambridge, 2008.

[9] W. Hu, *Wandering in the Background: A CMB Explorer*. PhD thesis, UC, Berkeley, 1995. arXiv:astro-ph/9508126 [astro-ph].

[10] W. Hu and S. Dodelson, "Cosmic Microwave Background Anisotropies," *Ann. Rev. Astron. Astrophys.* **40** (2002) 171–216, arXiv:astro-ph/0110414 [astro-ph].

[11] W. Hu, "Lecture Notes on CMB Theory: From Nucleosynthesis to Recombination," arXiv:0802.3688 [astro-ph].

[12] W. Hu, "Concepts in CMB Anisotropy Formation," in *The Universe at High-z, Large-Scale Structure and the Cosmic Microwave Background*, pp. 207–239. Springer, 1996.

[13] R. Peccei and H. Quinn, "CP Conservation in the Presence of Pseudoparticles," *Phys. Rev. Lett.* **38** (1977) 1440.

[14] S. Weinberg, "A New Light Boson?," *Phys. Rev. Lett.* **40** (1978) 223.

[15] F. Wilczek, "Problem of Strong P and T Invariance in the Presence of Instantons," *Phys. Rev. Lett.* **40** (1978) 279.

[16] A. Arvanitaki, S. Dimopoulos, S. Dubovsky, N. Kaloper, and J. March-Russell, "String Axiverse," *Phys. Rev.* **D81** (2010) 123530, arXiv:0905.4720 [hep-th].

[17] B. Holdom, "Two $U(1)$'s and Epsilon Charge Shifts," *Phys. Lett.* **166B** (1986) 196–198.

[18] K. Abazajian *et al.*, "Light Sterile Neutrinos: A White Paper," arXiv:1204.5379 [hep-ph].

[19] K. Abazajian *et al.* (CMB-S4 Collaborations), "CMB-S4 Science Book, First Edition," arXiv:1610.02743 [astro-ph.CO].

[20] B. Wallisch, *Cosmological Probes of Light Relics*. PhD thesis, University of Cambridge, 2018.

[21] Z. Hou, R. Keisler, L. Knox, M. Millea, and C. Reichardt, "How Massless Neutrinos Affect the Cosmic Microwave Background Damping Tail," *Phys. Rev.* **D87** (2013) 083008, arXiv:1104.2333 [astro-ph.CO].

[22] S. Bashinsky and U. Seljak, "Neutrino Perturbations in CMB Anisotropy and Matter Clustering," *Phys. Rev. D* **69** (2004) 083002, arXiv:astro-ph/0310198 [astro-ph].

[23] D. Baumann, F. Beutler, R. Flauger, D. Green, M. Vargas-Magaña, A. Slosar, B. Wallisch, and C. Yèche, "First Measurement of Neutrinos in the BAO Spectrum," arXiv:1803.10741 [astro-ph.CO].

[24] C. Brust, D. E. Kaplan, and M. Walters, "New Light Species and the CMB," *JHEP* **12** (2013) 058, arXiv:1303.5379 [hep-ph].

[25] F. Wilczek, "Axions and Family Symmetry Breaking," *Phys. Rev. Lett.* **49** (1982) 1549.

[26] D. Reiss, "Can the Family Group be a Global Symmetry?," *Phys. Lett. B* **115** (1982) 217.

[27] J. Kim, "Light Pseudoscalars, Particle Physics and Cosmology," *Phys. Rep.* **150** (1987) 1.

[28] Y. Chikashige, R. Mohapatra, and R. Peccei, "Are There Real Goldstone Bosons Associated with Broken Lepton Number?," *Phys. Lett. B* **98** (1981) 265.

[29] Y. Chikashige, R. Mohapatra, and R. Peccei, "Spontaneously Broken Lepton Number and Cosmological Constraints on the Neutrino Mass Spectrum," *Phys. Rev. Lett.* **45** (1980) 1926.

[30] D. Baumann, D. Green, and B. Wallisch, "New Target for Cosmic Axion Searches," *Phys. Rev. Lett.* **117** no. 17, (2016) 171301, arXiv:1604.08614 [astro-ph.CO].

[31] D. Baumann, D. Green, and B. Wallisch, "Searching for Light Relics with Large-Scale Structure," arXiv:1712.08067 [astro-ph.CO].

[32] A. Salvio, A. Strumia, and W. Xue, "Thermal Axion Production," *JCAP* **1401** (2014) 011.

[33] G. Carosi, A. Friedland, M. Giannotti, M. Pivovaroff, J. Ruz, and J. Vogel, "Probing the Axion-Photon Coupling: Phenomenological and Experimental Perspectives,". arXiv:1309.7035 [hep-ph].

[34] P. Graham and S. Rajendran, "New Observables for Direct Detection of Axion Dark Matter," *Phys. Rev. D* **88** (2013) 035023.

[35] K. Blum, R. D'Agnolo, M. Lisanti, and B. Safdi, "Constraining Axion Dark Matter with Big Bang Nucleosynthesis," *Phys. Lett. B* **737** (2014) 30.

[36] G. Raffelt, *Stars as Laboratories for Fundamental Physics*. University of Chicago Press, 1996.

[37] M. Pospelov and A. Ritz, "Theta-Induced Electric Dipole Moment of the Neutron via QCD Sum Rules," *Phys. Rev. Lett.* **83** (1999) 2526.

[38] D. Baumann and L. McAllister, *Inflation and String Theory*. Cambridge University Press, 2015. arXiv:1404.2601 [hep-th].

[39] R. Arnowitt, S. Deser, and C. Misner, "The Dynamics of General Relativity," *Gen. Rel. Grav.* **40** (2008) 1997–2027, arXiv:gr-qc/0405109 [gr-qc].

[40] J. Maldacena, "Non-Gaussian Features of Primordial Fluctuations in Single-Field Inflationary Models," *JHEP* **05** (2003) 013, arXiv:astro-ph/0210603 [astro-ph].

[41] P. A. R. Ade *et al.* (Planck and BICEP2 Collaborations), "Joint Analysis of BICEP2 *KeckArray* and *Planck* Data," *Phys. Rev. Lett.* **114** (2015) 101301, arXiv:1502.00612 [astro-ph.CO].

[42] S. Weinberg, "Quantum Contributions to Cosmological Correlations," *Phys. Rev.* **D72** (2005) 043514, arXiv:hep-th/0506236 [hep-th].

[43] X. Chen, "Primordial Non-Gaussianities from Inflation Models," *Adv. Astron.* **2010** (2010) 638979, arXiv:1002.1416 [astro-ph.CO].

[44] Y. Wang, "Inflation, Cosmic Perturbations and Non-Gaussianities," *Commun. Theor. Phys.* **62** (2014) 109–166, arXiv:1303.1523 [hep-th].

[45] E. Lim, *Part III Advanced Cosmology: Primordial Non-Gaussianities*.

[46] E. Komatsu, "Hunting for Primordial Non-Gaussianity in the Cosmic Microwave Background," *Class. Quant. Grav.* **27** (2010) 124010, arXiv:1003.6097 [astro-ph.CO].

[47] J. Schwinger, "Brownian Motion of a Quantum Oscillator," *J. Math. Phys.* **2** (1961) 407–432.

[48] E. Calzetta and B. Hu, "Closed Time Path Functional Formalism in Curved Space-Time: Application to Cosmological Back Reaction Problems," *Phys. Rev.* **D35** (1987) 495.

[49] R. Jordan, "Effective Field Equations for Expectation Values," *Phys. Rev.* **D33** (1986) 444–454.

[50] D. Seery, "One-Loop Corrections to a Scalar Field during Inflation," *JCAP* **0711** (2007) 025, arXiv:0707.3377 [astro-ph].

[51] P. Adshead, R. Easther, and E. Lim, "The 'in-in' Formalism and Cosmological Perturbations," *Phys. Rev.* **D80** (2009) 083521, arXiv:0904.4207 [hep-th].

[52] P. Creminelli and M. Zaldarriaga, "Single-Field Consistency Relation for the Three-Point Function," *JCAP* **0410** (2004) 006, arXiv:astro-ph/0407059 [astro-ph].

[53] X. Chen and Y. Wang, "Quasi-Single-Field Inflation and Non-Gaussianities," *JCAP* **1004** (2010) 027, arXiv:0911.3380 [hep-th].

[54] D. Baumann and D. Green, "Signatures of Supersymmetry from the Early Universe," *Phys. Rev.* **D85** (2012) 103520, arXiv:1109.0292 [hep-th].

[55] X. Chen and Y. Wang, "Quasi-Single-Field Inflation with Large Mass," *JCAP* **1209** (2012) 021, arXiv:1205.0160 [hep-th].

[56] V. Assassi, D. Baumann, and D. Green, "On Soft Limits of Inflationary Correlation Functions," *JCAP* **1211** (2012) 047, arXiv:1204.4207 [hep-th].

[57] T. Noumi, M. Yamaguchi, and D. Yokoyama, "EFT Approach to Quasi-Single-Field Inflation and Effects of Heavy Fields," *JHEP* **06** (2013) 051, arXiv:1211.1624 [hep-th].

[58] N. Arkani-Hamed and J. Maldacena, "Cosmological Collider Physics," arXiv:1503.08043 [hep-th].

[59] H. Lee, D. Baumann, and G. Pimentel, "Non-Gaussianity as a Particle Detector," *JHEP* **12** (2016) 040, arXiv:1607.03735 [hep-th].

[60] L. Bordin, P. Creminelli, M. Mirbabayi, and J. Noreña, "Tensor Squeezed Limits and the Higuchi Bound," arXiv:1605.08424 [astro-ph.CO].

[61] R. Flauger, M. Mirbabayi, L. Senatore, and E. Silverstein, "Productive Interactions: Heavy Particles and Non-Gaussianity," arXiv:1606.00513 [hep-th].

[62] J.-O. Gong, S. Pi, and M. Sasaki, "Equilateral Non-Gaussianity from Heavy Fields," *JCAP* **1311** (2013) 043, arXiv:1306.3691 [hep-th].

[63] D. Baumann, G. Goon, H. Lee, and G. L. Pimentel, "Partially Massless Fields During Inflation," arXiv:1712.06624 [hep-th].

[64] E. Wigner, "On Unitary Representations of the Inhomogeneous Lorentz Group," *Annals Math.* **40** (1939) 149–204. [Reprint: Nucl. Phys. Proc. Suppl.6,9(1989)].

[65] V. Bargmann and E. Wigner, "Group Theoretical Discussion of Relativistic Wave Equations," *Proc. Nat. Acad. Sci.* **34** (1948) 211.

[66] S. Deser and A. Waldron, "Arbitrary Spin Representations in de Sitter from dS/CFT with Applications to dS Supergravity," *Nucl. Phys.* **B662** (2003) 379–392, arXiv:hep-th/0301068 [hep-th].

[67] B. de Wit, "Supergravity," in *Unity from duality: Gravity, gauge theory and strings. Proceedings, NATO Advanced Study Institute, Euro Summer School, 76th session, Les Houches, France, July 30-August 31, 2001*, pp. 1–135. 2002. arXiv:hep-th/0212245 [hep-th].

[68] L. Thomas, "On Unitary Representations of the Group of De Sitter Space," *Annals of Mathematics* **42** no. 1, (1941) 113–126. http://www.jstor.org/stable/1968990.

[69] T. Newton, "A Note on the Representations of the De Sitter Group," *Annals of Mathematics* **51** no. 3, (1950) 730–733. http://www.jstor.org/stable/1969376.

[70] T. Garidi, "What is Mass in de Sitterian Physics?," arXiv:hep-th/0309104 [hep-th].

[71] S. Deser and A. Waldron, "Gauge Invariances and Phases of Massive Higher Spins in (A)dS," *Phys. Rev. Lett.* **87** (2001) 031601, arXiv:hep-th/0102166 [hep-th].

[72] A. Higuchi, "Forbidden Mass Range for Spin-2 Field Theory in De Sitter Spacetime," *Nucl. Phys.* **B282** (1987) 397.

[73] L. Singh and C. Hagen, "Lagrangian Formulation for Arbitrary Spin. 1. The Boson Case," *Phys. Rev.* **D9** (1974) 898–909.

[74] L. Singh and C. Hagen, "Lagrangian Formulation for Arbitrary Spin. 2. The Fermion Case," *Phys. Rev.* **D9** (1974) 910–920.

[75] Y. Zinoviev, "On Massive High Spin Particles in AdS," arXiv:hep-th/0108192 [hep-th].

[76] P. Creminelli, "On Non-Gaussianities in Single-Field Inflation," *JCAP* **0310** (2003) 003, arXiv:astro-ph/0306122 [astro-ph].

[77] V. Assassi, D. Baumann, D. Green, and L. McAllister, "Planck-Suppressed Operators," *JCAP* **1401** (2014) 033, arXiv:1304.5226 [hep-th].

[78] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, "The Effective Field Theory of Inflation," *JHEP* **03** (2008) 014, arXiv:0709.0293 [hep-th].

[79] M. Alvarez *et al.*, "Testing Inflation with Large Scale Structure: Connecting Hopes with Reality," arXiv:1412.4671 [astro-ph.CO].

[80] A. Loeb and M. Zaldarriaga, "Measuring the Small-Scale Power Spectrum of Cosmic Density Fluctuations through 21 cm Tomography Prior to the Epoch of Structure Formation," *Phys. Rev. Lett.* **92** (2004) 211301, arXiv:astro-ph/0312134 [astro-ph].

References from (?)

References

[1819] S. Weinberg, Rev. Mod. Phys. **61**, 1 (1989).

[451] S. M. Carroll, W. H. Press, and E. L. Turner, Ann. Rev. Astron. Astrophys. **30**, 499 (1992).

[3] A. Dolgov (1997), astro-ph/9708045.

[4] V. Sahni and A. A. Starobinsky, Int. J. Mod. Phys. **D9**, 373 (2000), astro-ph/9904398.

[5] N. Straumann, Eur. J. Phys. **20**, 419 (1999), astro-ph/9908342.

[6] S. Weinberg (2000), astro-ph/0005265.

[449] S. M. Carroll, Living Rev. Rel. **4**, 1 (2001), astro-ph/0004075.

[8] S. E. Rugh and H. Zinkernagel, Stud. Hist. Philos. Mod. Phys. (2000), hep-th/0012253.

[1393] T. Padmanabhan, Phys. Rept. **380**, 235 (2003), hep-th/0212290.

[10] J. Yokoyama (2003), gr-qc/0305068.

[11] S. Sarkar, Curr. Sci. **88**, 2120 (2005), special section of Current Science on the legacy of Albert Einstein.

[12] J. Polchinski, pp. 216–236 (2006), hep-th/0603249.

[508] E. J. Copeland, M. Sami, and S. Tsujikawa, Int. J. Mod. Phys. **D15**, 1753 (2006), hep-th/0603057.

[14] P. Brax (2009), 0912.3610.

[15] D. Sapone, Int. J. Mod. Phys. **A25**, 5253 (2010), 1006.5694.

[16] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Addison-Wesley Publishing Company, 1995).

[17] W. Greiner and J. Reinhardt, *Field Quantization* (Berlin, Germany: Springer, 1996).

[18] C. Itzykson and J. B. Zuber, *Quantum Field Theory* (Mcgraw-hill-International Series In Pure and Applied Physics, 1980).

[19] D. Bailin and A. Love, *Supersymmetric Gauge Field Theory and String Theory* (IOP (Graduate student series in physics), 1994).

[20] M. Le Bellac, *Quantum and Statistical Field Theory* (Oxford, 1991).

[21] L. H. Ryder, *Quantum Field Theory* (Cambridge Univ. Pr., 1985).

[22] F. Mandl and G. Shaw, *Quantum Field Theory* (Wiley-Interscience Publication, 1984).

[1445] S. Perlmutter et al. (Supernova Cosmology Project), Astrophys. J. **517**, 565 (1999), astro-ph/9812133.

[1524] A. G. Riess et al. (Supernova Search Team), Astron. J. **116**, 1009 (1998), astro-ph/9805201.

[260] A. A. Starobinsky, Phys. Lett. **B91**, 99 (1980).

[119] A. H. Guth, Phys. Rev. **D23**, 347 (1981).

[205] V. F. Mukhanov and G. Chibisov, JETP Lett. **33**, 532 (1981).

[170] A. D. Linde, Phys. Lett. **B108**, 389 (1982).

[206] V. F. Mukhanov and G. Chibisov, Sov. Phys. JETP **56**, 258 (1982).

[261] A. A. Starobinsky, Phys. Lett. **B117**, 175 (1982).

[121] A. H. Guth and S. Y. Pi, Phys. Rev. Lett. **49**, 1110 (1982).

[130] S. Hawking, Phys. Lett. **B115**, 295 (1982), revised version.

[15] A. Albrecht and P. J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).

[171] A. D. Linde, Phys. Lett. **B129**, 177 (1983).

[35] J. Martin, Braz. J. Phys. **34**, 1307 (2004), astro-ph/0312492.

[36] J. Martin, Lect. Notes Phys. **669**, 199 (2005), hep-th/0406011.

[37] J. Martin, Lect.Notes Phys. **738**, 193 (2008), 0704.3540.

[38] G. F. Smoot, C. Bennett, A. Kogut, E. Wright, J. Aymon, et al., Astrophys. J. **396**, L1 (1992).

[39] J. Martin and C. Ringeval, JCAP **0608**, 009 (2006), astro-ph/0605367.

[40] P. Astier and R. Pain (2012), 1204.5493.

[41] M. Kunz (2012), 1204.5482.

[42] C. Clarkson (2012), 1204.5505.

[43] C. de Rham (2012), 1204.5492.

[1563] A. Sakharov, Sov. Phys. Dokl. **12**, 1040 (1968).

[45] J. Kapusta and C. Gale, *Finite-Temperature Field Theory: Principles and Applications* (Cambrige Univ. Pr., 2006).

[46] S. Chatrchyan et al. (CMS Collaboration) (2012), 1202.1488.

[47] E. K. Akhmedov (2002), hep-th/0204048.

[48] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables* (National Bureau of Standards, Washington, US, 1970), ninth ed.

[49] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic Press, New York and London, 1965).

[50] G. Ossola and A. Sirlin, Eur. Phys. J. **C31**, 165 (2003), 28 pages, 5 figures, hep-ph/0305050.

[51] J. F. Koksma and T. Prokopec (2011), 1105.6296.

[52] L. Rosenberg, Phys. Rev. **129**, 2786 (1963).

[53] N. D. Birrell and P. C. W. Davies, *Quantum Fields In Curved Space* (Cambridge Univ. Pr., 1982).

[54] C. M. Bender and T. T. Wu, Phys. Rev. **184**, 1231 (1969).

[55] T. Abbott, Lecture notes (2006), URL web.mit.edu/tabbott/Public/8.06/current.pdf.

[56] J. Minahan, MIT Lecture notes (2011), URL www.teorfys.uu.se/people/minahan/Courses/QFT/wholeshebang.pdf.

[57] C. Cohen-Tannoudji, B. Diu, and F. Laloe, *Quantum Mechanics* (Wiley-Interscience, 2006).

[58] P. M. Stevenson, Phys. Rev. **D30**, 1712 (1984).

[59] P. M. Stevenson, Phys. Rev. **D32**, 1389 (1985).

[60] P. M. Stevenson and I. Roditi, Phys. Rev. **D33**, 2305 (1986).

[61] P. M. Stevenson, Z. Phys. **C35**, 467 (1987).

[62] P. M. Stevenson, B. Alles, and R. Tarrach, Phys. Rev. **D35**, 2407 (1987).

[63] P. M. Stevenson and R. Tarrach, Phys. Lett. **B176**, 436 (1986).

[64] G. Hajj and P. M. Stevenson, Phys. Rev. **D37**, 413 (1988).

[65] P. M. Stevenson, G. Hajj, and J. Reed, Phys. Rev. **D34**, 3117 (1986).

[66] R. Ibanez-Meier, I. Stancu, and P. M. Stevenson, Z. Phys. **C70**, 307 (1996), hep-ph/9207276.

[67] S. R. Coleman and E. J. Weinberg, Phys. Rev. **D7**, 1888 (1973).

[68] E. J. Weinberg and A.-q. Wu, Phys. Rev. **D36**, 2474 (1987).

[69] T. Curtright and C. B. Thorn, J. Math. Phys. **25**, 541 (1984).

[70] J. Wess and B. Zumino, Nucl. Phys. **B70**, 39 (1974).

[71] S. R. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2239 (1969).

[72] J. Callan, Curtis G., S. R. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2247 (1969).

[73] J. Wess and B. Zumino, Phys. Lett. **B49**, 52 (1974).

[74] I. J. Aitchison (2005), notes of Lectures for Graduate Students in Particle Physics Oxford, 2004 and 2005, hep-ph/0505105.

[75] P. Binetruy, *Supersymmetry: Theory, experiment and cosmology* (Oxford Univ. Pr., 2006).

[76] L. Parker, In Cargese 1978, Proceedings, Recent Developments In Gravitation, 219-273 (1978).

[77] T. S. Bunch and L. Parker, Phys. Rev. **D20**, 2499 (1979).

[78] E. Poisson, A. Pound, and I. Vega, Living Rev. Rel. **14**, 7 (2011), 1102.0529.

[79] R. Scranton et al. (SDSS) (2003), astro-ph/0307335.

[80] P. Solevi, R. Mainini, and S. A. Bonometto (2004), astro-ph/0412054.

[1677] D. N. Spergel et al. (WMAP), Astrophys. J. Suppl. **170**, 377 (2007), astro-ph/0603449.

[82] M. Tegmark et al. (SDSS), Phys. Rev. **D69**, 103501 (2004), astro-ph/0310723.

[83] E. Kolb and M. Turner, *The Early Universe*, vol. 69 of *Frontiers in Physics Series* (Addison-Wesley Publishing Company, 1990).

[1497] B. Ratra and P. J. E. Peebles, Phys. Rev. **D37**, 3406 (1988).

[85] P. Brax and J. Martin, Phys. Lett. **B468**, 40 (1999), astro-ph/9905040.

[86] P. Brax and J. Martin, Phys. Rev. **D61**, 103502 (2000), astro-ph/9912046.

[87] P. Brax, J. Martin, and A. Riazuelo, Phys. Rev. **D62**, 103505 (2000), astro-ph/0005428.

[88] P. Brax, J. Martin, and A. Riazuelo, Phys. Rev. **D64**, 083505 (2001), hep-ph/0104240.

[89] J. Martin and M. A. Musso, Phys. Rev. **D71**, 063514 (2005), astro-ph/0410190.

[90] J. Martin, C. Schimd, and J.-P. Uzan, Phys. Rev. Lett. **96**, 061303 (2006), astro-ph/0510208.

[91] P. Brax and J. Martin, Phys. Lett. **B647**, 320 (2007), hep-th/0612208.

[92] P. Brax and J. Martin, JCAP **0611**, 008 (2006), astro-ph/0606306.

[93] P. Brax and J. Martin, Phys. Rev. **D75**, 083507 (2007), hep-th/0605228.

[94] P. Brax, C. van de Bruck, J. Martin, and A.-C. Davis, JCAP **0909**, 032 (2009), 0904.3471.

[95] C. F. Kolda and D. H. Lyth, Phys. Lett. **B458**, 197 (1999), hep-ph/9811375.

[448] S. M. Carroll, Phys. Rev. Lett. **81**, 3067 (1998), astro-ph/9806099.

[97] M. Doran and J. Jaeckel (2002), astro-ph/0205206.

[98] M. Doran and J. Jaeckel, Lect. Notes Phys. **646**, 273 (2004).

[99] M. Garny, Phys. Rev. **D74**, 043009 (2006), hep-ph/0606120.

[100] C. Ringeval, T. Suyama, T. Takahashi, M. Yamaguchi, and S. Yokoyama, Phys. Rev. Lett. **105**, 121301 (2010), 1006.0368.

[101] I. Maor, R. Brustein, and P. J. Steinhardt, Phys. Rev. Lett. **86**, 6 (2001), astro-ph/0007297.

[102] S. Weinberg, *Gravitation and Cosmology* (John Wiley & Sons, 1972).

[103] E. L. Wright (1998), astro-ph/9805292.

[104] A. W. Kerr, J. C. Hauck, and B. Mashhoon, Class. Quant. Grav. **20**, 2727 (2003), gr-qc/0301057.

[105] P. Jetzer (2007), 0711.3723.

[106] L. Parker, Phys. Rev. **D22**, 1922 (1980).

[107] L. Parker and L. O. Pimentel, Phys. Rev. **D25**, 3180 (1982).

[108] L. Parker, Phys. Rev. **D24**, 535 (1981).

[109] S. Moradi and E. Aboualizadeh, Gen. Rel. Grav. **42**, 435 (2010).

[110] E. Fischbach, B. S. Freeman, and W.-K. Cheng, Phys. Rev. **D23**, 2157 (1981).

[111] N. Boulanger, P. Spindel, and F. Buisseret, Phys. Rev. **D74**, 125014 (2006), hep-th/0610207.

[112] T. A. Welton, Phys. Rev. **74**, 1157 (1948).

[113] G. Plunien, B. Muller, and W. Greiner, Phys. Rept. **134**, 87 (1986).

[114] K. A. Milton, J. Phys. A **A37**, R209 (2004), hep-th/0406024.

[115] K. A. Milton, pp. 333–349 (2000), hep-th/0009173.

[116] K. A. Milton, Lect. Notes Phys. **834**, 39 (2011), 1005.0031.

[117] A. Lambrecht and S. Reynaud (2011), 1112.1301.

[118] K. A. Milton, Phys. Rev. **D68**, 065020 (2003), hep-th/0210081.

[119] D. Deutsch and P. Candelas, Phys. Rev. **D20**, 3063 (1979).

[120] P. Milonni and M.-L. Shih, Phys. Rev. **A45**, 4241 (1992).

[121] R. Jaffe, Phys. Rev. **D72**, 021301 (2005), hep-th/0503158.

[122] M. Chernodub (2012), 1203.6588.

[123] E. Calloni, L. Di Fiore, G. Esposito, L. Milano, and L. Rosa, Phys. Lett. **A297**, 328 (2002), quant-ph/0109091.

[124] G. Bimonte, E. Calloni, G. Esposito, and L. Rosa, Phys. Rev. **D76**, 025008 (2007), hep-th/0703062.

[125] S. A. Fulling, K. A. Milton, P. Parashar, A. Romeo, K. Shajesh, et al., Phys. Rev. **D76**, 025004 (2007), hep-th/0702091.

[126] K. A. Milton, S. A. Fulling, P. Parashar, A. Romeo, K. Shajesh, et al., J. Phys. A **A41**, 164052 (2008), 0710.3841.

[127] K. Milton, P. Parashar, J. Wagner, K. Shajesh, A. Romeo, et al. (2008), 0810.0081.

[128] R. Caldwell (2002), astro-ph/0209312.

[129] F. Sorge, Class. Quant. Grav. **22**, 5109 (2005).

[130] E. Masso, Phys. Lett. **B679**, 433 (2009), 0902.4318.

[131] Y. Su, B. R. Heckel, E. Adelberger, J. Gundlach, M. Harris, et al., Phys. Rev. **D50**, 3614 (1994).

[132] T. Damour, Space Sci.Rev. (2009), 0906.3174.

[133] R. Colella, A. Overhauser, and S. Werner, Phys. Rev. Lett. **34**, 1472 (1975).

[134] D. M. Greenberger and A. Overhauser, Rev. Mod. Phys. **51**, 43 (1979).

[135] D. M. Greenberger, Rev. Mod. Phys. **55**, 875 (1983).

[136] H. Salecker and E. Wigner, Phys. Rev. **109**, 571 (1958).

[137] A. Peres, Am. J. Phys. **48**, 552 (1980).

[138] C. Leavens, Solid Sate Commun. **86**, 781 (1993).

[139] P. Davies, Class. Quant. Grav. **21**, 5677 (2004).

[140] P. C. W. Davies, Classical and Quantum Gravity **21**, 2761 (2004), arXiv:quant-ph/0403027.

[141] N. Grot, C. Rovelli, and R. S. Tate, Phys. Rev. **A54**, 4679 (1996), quant-ph/9603021.

[142] Y. Aharonov, J. Oppenheim, S. Popescu, B. Reznik, and W. Unruh, Phys. Rev. **A57**, 4130 (1998), quant-ph/9709031.

[143] J. Oppenheim, B. Reznik, and W. Unruh, Phys. Rev. **A59**, 1804 (1999), quant-ph/9807043.

[144] P. C. Davies, Am. J. Phys. **73**, 23 (2005).

[145] B. Mielnik and G. Torres-Vega, ArXiv e-prints (2011), 1112.4198.

[146] L. Viola and R. Onofrio, Phys. Rev. **D55**, 455 (1997), quant-ph/9612039.

[147] M. Ali, A. Majumdar, D. Home, and A. K. Pan, Class. Quant. Grav. **23**, 6493 (2006), quant-ph/0606183.

[148] P. Chowdhury, D. Home, A. Majumdar, S. Mousavi, M. Mozaffari, et al., Class. Quant. Grav. **29**, 025010 (2012).

[149] R. Onofrio, Int. J. Mod. Phys. **A25**, 2260 (2010), 1101.1484.

[150] D. Giulini (2011), 1105.0749.

[151] E. Kajari, N. L. Harshman, E. M. Rasel, S. Stenholm, G. Süßmann, and W. P. Schleich, Applied Physics B: Lasers and Optics **100**, 43 (2010), 1006.1988.

[152] V. Nesvizhevsky, H. Borner, A. Gagarski, A. Petoukhov, G. Petrov, et al., Phys.Rev. **D67**, 102002 (2003), hep-ph/0306198.

[153] H. Abele, S. Baessler, H. Borner, A. Gagarsky, V. Nesvizhevsky, et al., AIP Conf.Proc. **842**, 793 (2006).

[154] S. Nobbenhuis (2006), ph.D. Thesis (Advisor: G. t'Hooft), gr-qc/0609011.

[155] A. Dolgov (1982).

[728] L. Ford, Phys. Rev. **D35**, 2339 (1987).

[157] A. Dolgov and F. Urban, Phys. Rev. **D77**, 083503 (2008), 0801.3090.

[158] L. W. Abramo, R. H. Brandenberger, and V. F. Mukhanov, Phys. Rev. **D56**, 3248 (1997), gr-qc/9704037.

[159] V. F. Mukhanov, L. W. Abramo, and R. H. Brandenberger, Phys. Rev. Lett. **78**, 1624 (1997), gr-qc/9609026.

[160] J. Hartle and S. Hawking, Phys. Rev. **D28**, 2960 (1983).

[161] E. Baum, Phys. Lett. **B133**, 185 (1983).

[162] S. R. Coleman, Nucl. Phys. **B310**, 643 (1988).

[163] V. Rubakov and M. Shaposhnikov, Phys. Lett. **B125**, 139 (1983).

[164] C. Burgess, Annals Phys. **313**, 283 (2004), paper has been updated to include developments since the conference, hep-th/0402200.

[165] B. Zwiebach, *A first Course in String Theory* (Cambridge Univ. Pr., 2004).

[1818] S. Weinberg, Phys. Rev. Lett. **59**, 2607 (1987).

[167] A. Linde and M. Noorbala, JCAP **1009**, 008 (2010), 1006.2170.

[168] G. D. Starkman and R. Trotta, Phys. Rev. Lett. **97**, 201301 (2006), astro-ph/0607227.

[169] R. Trotta and G. D. Starkman, AIP Conf.Proc. **878**, 323 (2006), astro-ph/0610330.

[170] N. Afshordi (2008), 0807.2639.

[171] G. F. Ellis, H. van Elst, J. Murugan, and J.-P. Uzan, Class. Quant. Grav. **28**, 225007 (2011), 1008.1196.

[172] S. Aslanbeigi, G. Robbers, B. Z. Foster, K. Kohri, and N. Afshordi, Phys. Rev. **D84**, 103522 (2011), 1106.3955.

[173] G. Volovik, Phys. Rept. **351**, 195 (2001), gr-qc/0005091.

[174] G. Volovik, Annalen Phys. **14**, 165 (2005), gr-qc/0405012.

[175] L. Sindoni (2011), 54 pages. Invited review for SIGMA Special Issue 'Loop Quantum Gravity and Cosmology', 1110.0686.

[176] G. Volovik (2001), gr-qc/0101111.

[177] G. Volovik, Int. Ser. Monogr. Phys. **117**, 1 (2006).

[178] S. H. Alexander and G. Calcagni, Found. Phys. **38**, 1148 (2008), 0807.0225.

[179] S. Finazzi, S. Liberati, and L. Sindoni, Phys. Rev. Lett. **108**, 071101 (2012), revtex4, 5 pages, 1 figures, 1103.4841.

[180] S. Finazzi, S. Liberati, and L. Sindoni (2012), 1204.3039.

References from another Baumann TASI lectures

References

[1] A. H. Guth, Phys. Rev. **D23**, 347 (1981).

[2] A. D. Linde, Phys. Lett. **B108**, 389 (1982).

[3] A. Albrecht and P. J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).

[4] E. Hubble, Proc. Nat. Acad. Sci. **15**, 168 (1929).

[5] R. A. Alpher, H. Bethe, and G. Gamow, Phys. Rev. **73**, 803 (1948).

[6] R. Dicke, P. J. E. Peebles, P. Roll, and D. Wilkinson, Astrophys. J. **142**, 414 (1965).

[7] CMBPol Study Team, D. Baumann *et al.*, (2008), 0811.3919.

[8] S. Dodelson, *Modern Cosmology* (Amsterdam, Netherlands: Academic Press, 2003).

[9] V. Mukhanov, *Physical Foundations of Cosmology* (Cambridge University Press, 2005).

[10] S. Weinberg, *Cosmology* (Oxford University Press, 2008).

[11] WMAP, E. Komatsu *et al.*, (2008), 0803.0547.

[12] M. Kowalski *et al.*, (2008), 0804.4142.

[13] B. A. Bassett, S. Tsujikawa, and D. Wands, Rev. Mod. Phys. **78**, 537 (2006), astro-ph/0507632.

[14] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, JHEP **03**, 014 (2008), 0709.0293.

[15] S. Weinberg, Phys. Rev. **D77**, 123541 (2008), 0804.4291.

[16] D. Baumann and L. McAllister, (2009), 0901.0265.

[17] Q. Shafi and V. N. Senoguz, Phys. Rev. **D73**, 127301 (2006), astro-ph/0603830.

[18] D. H. Lyth and A. Riotto, Phys. Rept. **314**, 1 (1999), hep-ph/9807278.

[19] D. Wands, Lect. Notes Phys. **738**, 275 (2008), astro-ph/0702187.

[20] J. M. Bardeen, Phys. Rev. **D22**, 1882 (1980).

[21] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, Phys. Rev. **D28**, 679 (1983).

[22] D. Wands, K. A. Malik, D. H. Lyth, and A. R. Liddle, Phys. Rev. **D62**, 043527 (2000), astro-ph/0003278.

[23] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, Nucl. Phys. **B667**, 119 (2003), astro-ph/0209156.

[24] J. M. Maldacena, JHEP **05**, 013 (2003), astro-ph/0210603.

[25] N. D. Birrell and P. C. W. Davies, Cambridge, Uk: Univ. Pr.)1982) 340p.

[26] D. H. Lyth, Phys. Rev. Lett. **78**, 1861 (1997), hep-ph/9606387.

[27] W. Hu, `http://background.uchicago.edu/ whu/`.

[28] U. Seljak and M. Zaldarriaga, astro-ph/9603033, `http://www.cmbfast.org`.

[29] A. Lewis and A. Challinor, `http://camb.info`.

[30] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, Phys. Rept. **402**, 103 (2004), astro-ph/0406398.

[31] M. Kamionkowski, A. Kosowsky, and A. Stebbins, Phys. Rev. **D55**, 7368 (1997), astro-ph/9611125.

[32] M. Zaldarriaga and U. Seljak, Phys. Rev. **D55**, 1830 (1997), astro-ph/9609170.

[33] E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

[34] D. N. Spergel and M. Zaldarriaga, Phys. Rev. Lett. **79**, 2180 (1997).

[35] S. Dodelson, AIP Conf. Proc. **689**, 184 (2003), hep-ph/0309057.

[36] J. Bock *et al.*, (2008), 0805.4207.

[37] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, Astrophys. J. **304**, 15 (1986).

[38] D. J. Eisenstein and W. Hu, Astrophys. J. **496**, 605 (1998), astro-ph/9709112.

[39] P. J. Steinhardt and N. Turok, Science **296**, 1436 (2002).

[40] P. J. Steinhardt and N. Turok, Phys. Rev. **D65**, 126003 (2002), hep-th/0111098.

[41] E. I. Buchbinder, J. Khoury, and B. A. Ovrut, Phys. Rev. **D76**, 123503 (2007), hep-th/0702154.

[42] K. M. Smith, L. Senatore, and M. Zaldarriaga, (2009), 0901.2572.

[43] S. Dimopoulos, S. Kachru, J. McGreevy, and J. G. Wacker, (2005), hep-th/0507205.

[44] L. McAllister, E. Silverstein, and A. Westphal, (2008), 0808.0706.

[45] E. Silverstein and A. Westphal, (2008), 0803.3085.

[46] S. Kachru *et al.*, JCAP **0310**, 013 (2003), hep-th/0308055.

[47] D. Baumann, A. Dymarsky, I. R. Klebanov, and L. McAllister, JCAP **0801**, 024 (2008), 0706.0360.

[48] D. Baumann, A. Dymarsky, S. Kachru, I. R. Klebanov, and L. McAllister, JHEP **03**, 093 (2009), 0808.2811.

[49] J. P. Conlon and F. Quevedo, JHEP **01**, 146 (2006), hep-th/0509012.

[50] J. J. Blanco-Pillado *et al.*, JHEP **09**, 002 (2006), hep-th/0603129.

[51] D. Baumann and M. Zaldarriaga, (2009), 0901.0958.

[52] D. Babich, P. Creminelli, and M. Zaldarriaga, JCAP **0408**, 009 (2004), astro-ph/0405356.

[53] N. Bartolo, S. Matarrese, and A. Riotto, Phys. Rev. **D65**, 103505 (2002), hep-ph/0112261.

[54] F. Bernardeau and J.-P. Uzan, Phys. Rev. **D66**, 103506 (2002), hep-ph/0207295.

[55] F. Bernardeau and J.-P. Uzan, Phys. Rev. **D67**, 121301 (2003), astro-ph/0209330.

[56] M. Sasaki, Prog. Theor. Phys. **120**, 159 (2008), 0805.0974.

[57] A. Naruko and M. Sasaki, (2008), 0807.0180.

[58] C. T. Byrnes, K.-Y. Choi, and L. M. H. Hall, (2008), 0807.1101.

[59] C. T. Byrnes and D. Wands, Phys. Rev. **D74**, 043529 (2006), astro-ph/0605679.

[60] D. Langlois, F. Vernizzi, and D. Wands, (2008), 0809.4646.

[61] J. Valiviita, H. Assadullahi, and D. Wands, (2008), 0806.0623.

[62] H. Assadullahi, J. Valiviita, and D. Wands, Phys. Rev. **D76**, 103003 (2007), 0708.0223.

[63] J. Valiviita, M. Sasaki, and D. Wands, (2006), astro-ph/0610001.

[64] F. Vernizzi and D. Wands, JCAP **0605**, 019 (2006), astro-ph/0603799.

[65] L. E. Allen, S. Gupta, and D. Wands, JCAP **0601**, 006 (2006), astro-ph/0509719.

[66] A. D. Linde and V. F. Mukhanov, Phys. Rev. **D56**, 535 (1997), astro-ph/9610219.

[67] D. H. Lyth, C. Ungarelli, and D. Wands, Phys. Rev. **D67**, 023503 (2003), astro-ph/0208055.

[68] G. Dvali, A. Gruzinov, and M. Zaldarriaga, Phys. Rev. **D69**, 023505 (2004), astro-ph/0303591.

[69] L. Kofman, (2003), astro-ph/0303614.

[70] P. Creminelli and L. Senatore, JCAP **0711**, 010 (2007), hep-th/0702165.

[71] K. Koyama, S. Mizuno, F. Vernizzi, and D. Wands, JCAP **0711**, 024 (2007), 0708.4321.

[72] E. I. Buchbinder, J. Khoury, and B. A. Ovrut, Phys. Rev. Lett. **100**, 171302 (2008), 0710.5172.

[73] J.-L. Lehners and P. J. Steinhardt, Phys. Rev. **D77**, 063533 (2008), 0712.3779.

[74] J.-L. Lehners and P. J. Steinhardt, Phys. Rev. **D78**, 023506 (2008), 0804.1293.

[75] K. Koyama, S. Mizuno, and D. Wands, Class. Quant. Grav. **24**, 3919 (2007), 0704.1152.

[76] K. Koyama and D. Wands, JCAP **0704**, 008 (2007), hep-th/0703040.

[77] M. Alishahiha, E. Silverstein, and D. Tong, Phys. Rev. **D70**, 123505 (2004), hep-th/0404084.

[78] N. Arkani-Hamed, P. Creminelli, S. Mukohyama, and M. Zaldarriaga, JCAP **0404**, 001 (2004), hep-th/0312100.

[79] L. Senatore, Phys. Rev. **D71**, 043512 (2005), astro-ph/0406187.

[80] P. Creminelli, JCAP **0310**, 003 (2003), astro-ph/0306122.

[81] X. Chen, M.-x. Huang, S. Kachru, and G. Shiu, JCAP **0701**, 002 (2007), hep-th/0605045.

[82] C. Gordon, D. Wands, B. A. Bassett, and R. Maartens, Phys. Rev. **D63**, 023506 (2001), astro-ph/0009131.

[83] D. H. Lyth, C. Ungarelli, and D. Wands, Phys. Rev. **D67**, 023503 (2003), astro-ph/0208055.

[84] S. Weinberg, Phys. Rev. **D70**, 083522 (2004), astro-ph/0405397.

[85] E. Komatsu *et al.*, (2009), 0902.4759.

[86] J. S. Schwinger, J. Math. Phys. **2**, 407 (1961).

[87] E. Calzetta and B. L. Hu, Phys. Rev. **D35**, 495 (1987).

[88] R. D. Jordan, Phys. Rev. **D33**, 444 (1986).

[89] S. Weinberg, Phys. Rev. **D72**, 043514 (2005), hep-th/0506236.

[90] E. Komatsu and D. N. Spergel, Phys. Rev. **D63**, 063002 (2001), astro-ph/0005036.

[91] J. R. Fergusson and E. P. S. Shellard, (2008), 0812.3413.

[92] E. Silverstein and D. Tong, Phys. Rev. **D70**, 103505 (2004), hep-th/0310221.

[93] R. Holman and A. J. Tolley, JCAP **0805**, 001 (2008), 0710.1302.

[94] P. Creminelli and M. Zaldarriaga, JCAP **0410**, 006 (2004), astro-ph/0407059.

[95] C. Cheung, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, JCAP **0802**, 021 (2008), 0709.0295.

[96] L. Senatore, K. M. Smith, and M. Zaldarriaga, (2009), 0905.3746.

[97] N. Bartolo, S. Matarrese, and A. Riotto, JCAP **0606**, 024 (2006), astro-ph/0604416.

[98] N. Bartolo, S. Matarrese, and A. Riotto, JCAP **0701**, 019 (2007), astro-ph/0610110.

[99] L. Senatore, S. Tassev, and M. Zaldarriaga, (2008), 0812.3658.

[100] N. Dalal, O. Dore, D. Huterer, and A. Shirokov, Phys. Rev. **D77**, 123514 (2008), 0710.4560.

[101] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, (2008), 0805.3580.

[102] E. J. Copeland, A. R. Liddle, D. H. Lyth, E. D. Stewart, and D. Wands, Phys. Rev. **D49**, 6410 (1994), astro-ph/9401011.

[103] J. Polchinski, *String Theory. Vol. 1: An Introduction to the Bosonic String* (Cambridge University Press, 1998).

[104] J. Polchinski, *String Theory. Vol. 2: Superstring Theory and Beyond* (Cambridge University Press, 1998).

[105] S. Kachru, R. Kallosh, A. Linde, and S. P. Trivedi, Phys. Rev. **D68**, 046005 (2003), hep-th/0301240.

[106] D. H. Lyth, Phys. Rev. Lett. **78**, 1861 (1997), hep-ph/9606387.

[107] A. D. Linde, Phys. Lett. **B129**, 177 (1983).

[108] C. Vafa, (2005), hep-th/0509212.

[109] H. Ooguri and C. Vafa, Nucl. Phys. **B766**, 21 (2007), hep-th/0605264.

[110] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, JHEP **10**, 014 (2006), hep-th/0602178.

[111] M. P. Hertzberg, M. Tegmark, S. Kachru, J. Shelton, and O. Ozcan, Phys. Rev. **D76**, 103521 (2007), 0709.0002.

[112] M. R. Douglas and S. Kachru, Rev. Mod. Phys. **79**, 733 (2007), hep-th/0610102.

[113] I. R. Klebanov and M. J. Strassler, JHEP **08**, 052 (2000), hep-th/0007191.

[114] I. R. Klebanov and A. A. Tseytlin, Nucl. Phys. **B578**, 123 (2000), hep-th/0002159.

[115] S. B. Giddings, S. Kachru, and J. Polchinski, Phys. Rev. **D66**, 106006 (2002), hep-th/0105097.

[116] D. Baumann and L. McAllister, Phys. Rev. **D75**, 123508 (2007), hep-th/0610285.

[117] D. Baumann *et al.*, JHEP **11**, 031 (2006), hep-th/0607050.

[118] D. Baumann, A. Dymarsky, I. R. Klebanov, L. McAllister, and P. J. Steinhardt, Phys. Rev. Lett. **99**, 141601 (2007), 0705.3837.

[119] A. Krause and E. Pajer, JCAP **0807**, 023 (2008), 0705.4682.

[120] M. Nakahara, Bristol, UK: Hilger (1990) 505 p. (Graduate student series in physics).

[121] D. Baumann, A. Dymarsky, S. Kachru, I. R. Klebanov, and L. McAllister.

[122] H. Firouzjahi and S. H. H. Tye, JCAP **0503**, 009 (2005), hep-th/0501099.

[123] S. E. Shandera and S. H. H. Tye, JCAP **0605**, 007 (2006), hep-th/0601099.

[124] K. Freese, J. A. Frieman, and A. V. Olinto, Phys. Rev. Lett. **65**, 3233 (1990).

[125] T. Banks, M. Dine, P. J. Fox, and E. Gorbatov, JCAP **0306**, 001 (2003), hep-th/0303252.

[126] P. Svrcek and E. Witten, JHEP **06**, 051 (2006), hep-th/0605206.

[127] R. Easther and L. McAllister, JCAP **0605**, 018 (2006), hep-th/0512102.

[128] A. R. Liddle, A. Mazumdar, and F. E. Schunck, Phys. Rev. **D58**, 061301 (1998), astro-ph/9804177.

[129] R. Kallosh, N. Sivanandam, and M. Soroush, Phys. Rev. **D77**, 043501 (2008), 0710.3429.

[130] T. W. Grimm, Phys. Rev. **D77**, 126007 (2008), 0710.3883.

[131] M. P. Hertzberg, S. Kachru, W. Taylor, and M. Tegmark, JHEP **12**, 095 (2007), 0711.2512.

[132] R. Flauger, S. Paban, D. Robbins, and T. Wrase, (2008), 0812.3886.

[133] C. Caviezel *et al.*, JHEP **04**, 010 (2009), 0812.3551.

[134] S. Sarangi and S. H. H. Tye, Phys. Lett. **B536**, 185 (2002), hep-th/0204074.

[135] E. J. Copeland, R. C. Myers, and J. Polchinski, JHEP **06**, 013 (2004), hep-th/0312067.

[136] C. P. Burgess, PoS **P2GC**, 008 (2006), 0708.2865.

[137] R. Kallosh, Lect. Notes Phys. **738**, 119 (2008), hep-th/0702059.

[138] A. Linde, ECONF **C040802**, L024 (2004), hep-th/0503195.

[139] L. McAllister and E. Silverstein, Gen. Rel. Grav. **40**, 565 (2008), 0710.2951.

[140] J. M. Cline, (2006), hep-th/0612129.

[141] R. Flauger, L. McAllister, E. Pajer, A. Westphal, and G. Xu, (2009), 0907.2916.

[142] Planck, (2006), astro-ph/0604069.

[143] `http://cmbpol.uchicago.edu/workshops/path2009/`.

[144] A. Kosowsky, New Astron. Rev. **47**, 939 (2003), astro-ph/0402234.

[145] SPT, J. E. Ruhl *et al.*, Proc. SPIE Int. Soc. Opt. Eng. **5498**, 11 (2004), astro-ph/0411122.

[146] Clover, A. C. Taylor, New Astron. Rev. **50**, 993 (2006), astro-ph/0610716.

[147] D. Samtleben, Nuovo Cim. **122B**, 1353 (2007), 0802.2657.

[148] K. W. Yoon *et al.*, (2006), astro-ph/0606278.

[149] P. Oxley *et al.*, Proc. SPIE Int. Soc. Opt. Eng. **5543**, 320 (2004), astro-ph/0501111.

[150] C. J. MacTavish *et al.*, (2007), 0710.0375.

[151] EPIC, J. Bock *et al.*, (2009), 0906.1188.

[152] A. Vilenkin, Phys. Rev. **D27**, 2848 (1983).

[153] A. D. Linde, Phys. Lett. **B175**, 395 (1986).

[154] A. D. Linde and A. Mezhlumian, Phys. Lett. **B307**, 25 (1993), gr-qc/9304015.

[155] A. D. Linde, D. A. Linde, and A. Mezhlumian, Phys. Rev. **D49**, 1783 (1994), gr-qc/9306035.

[156] J. Garcia-Bellido, A. D. Linde, and D. A. Linde, Phys. Rev. **D50**, 730 (1994), astro-ph/9312039.

[157] A. Vilenkin, Phys. Rev. Lett. **74**, 846 (1995), gr-qc/9406010.

[158] J. Garriga, D. Schwartz-Perlov, A. Vilenkin, and S. Winitzki, JCAP **0601**, 017 (2006), hep-th/0509184.

[159] R. Bousso, Phys. Rev. Lett. **97**, 191302 (2006), hep-th/0605263.

[160] A. Linde, JCAP **0706**, 017 (2007), 0705.1160.

[161] R. Bousso, B. Freivogel, and I.-S. Yang, Phys. Rev. **D79**, 063513 (2009), 0808.3770.

[162] A. De Simone *et al.*, (2008), 0808.3778.

[163] J. Garriga and A. Vilenkin, JCAP **0901**, 021 (2009), 0809.4257.

[164] S. Winitzki, Phys. Rev. **D78**, 123518 (2008), 0810.1517.

[165] A. Linde, V. Vanchurin, and S. Winitzki, JCAP **0901**, 031 (2009), 0812.0005.

[166] K. A. Malik and D. Wands, (2008), 0809.4944.

[167] L. Verde, (2007), 0712.3028.

[168] R. L. Arnowitt, S. Deser, and C. W. Misner, (1962), gr-qc/0405109.

[169] D. Seery, JCAP **0711**, 025 (2007), 0707.3377.

[170] P. Adshead, R. Easther, and E. A. Lim, (2009), 0904.4207.

References from description of cosmological constant

References

[1819] S. Weinberg, Rev. Mod. Phys. **61**, 1 (1989).

[451] S. M. Carroll, W. H. Press, and E. L. Turner, Ann. Rev. Astron. Astrophys. **30**, 499 (1992).

[3] A. Dolgov (1997), astro-ph/9708045.

[4] V. Sahni and A. A. Starobinsky, Int. J. Mod. Phys. **D9**, 373 (2000), astro-ph/9904398.

[5] N. Straumann, Eur. J. Phys. **20**, 419 (1999), astro-ph/9908342.

[6] S. Weinberg (2000), astro-ph/0005265.

[449] S. M. Carroll, Living Rev. Rel. **4**, 1 (2001), astro-ph/0004075.

[8] S. E. Rugh and H. Zinkernagel, Stud. Hist. Philos. Mod. Phys. (2000), hep-th/0012253.

[1393] T. Padmanabhan, Phys. Rept. **380**, 235 (2003), hep-th/0212290.

[10] J. Yokoyama (2003), gr-qc/0305068.

[11] S. Sarkar, Curr. Sci. **88**, 2120 (2005), special section of Current Science on the legacy of Albert Einstein.

[12] J. Polchinski, pp. 216–236 (2006), hep-th/0603249.

[508] E. J. Copeland, M. Sami, and S. Tsujikawa, Int. J. Mod. Phys. **D15**, 1753 (2006), hep-th/0603057.

[14] P. Brax (2009), 0912.3610.

[15] D. Sapone, Int. J. Mod. Phys. **A25**, 5253 (2010), 1006.5694.

[16] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Addison-Wesley Publishing Company, 1995).

[17] W. Greiner and J. Reinhardt, *Field Quantization* (Berlin, Germany: Springer, 1996).

[18] C. Itzykson and J. B. Zuber, *Quantum Field Theory* (Mcgraw-hill-International Series In Pure and Applied Physics, 1980).

[19] D. Bailin and A. Love, *Supersymmetric Gauge Field Theory and String Theory* (IOP (Graduate student series in physics), 1994).

[20] M. Le Bellac, *Quantum and Statistical Field Theory* (Oxford, 1991).

[21] L. H. Ryder, *Quantum Field Theory* (Cambridge Univ. Pr., 1985).

[22] F. Mandl and G. Shaw, *Quantum Field Theory* (Wiley-Interscience Publication, 1984).

[1445] S. Perlmutter et al. (Supernova Cosmology Project), Astrophys. J. **517**, 565 (1999), astro-ph/9812133.

[1524] A. G. Riess et al. (Supernova Search Team), Astron. J. **116**, 1009 (1998), astro-ph/9805201.

[260] A. A. Starobinsky, Phys. Lett. **B91**, 99 (1980).

[119] A. H. Guth, Phys. Rev. **D23**, 347 (1981).

[205] V. F. Mukhanov and G. Chibisov, JETP Lett. **33**, 532 (1981).

[170] A. D. Linde, Phys. Lett. **B108**, 389 (1982).

[206] V. F. Mukhanov and G. Chibisov, Sov. Phys. JETP **56**, 258 (1982).

[261] A. A. Starobinsky, Phys. Lett. **B117**, 175 (1982).

[121] A. H. Guth and S. Y. Pi, Phys. Rev. Lett. **49**, 1110 (1982).

[130] S. Hawking, Phys. Lett. **B115**, 295 (1982), revised version.

[15] A. Albrecht and P. J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).

[171] A. D. Linde, Phys. Lett. **B129**, 177 (1983).

[35] J. Martin, Braz. J. Phys. **34**, 1307 (2004), astro-ph/0312492.

[36] J. Martin, Lect. Notes Phys. **669**, 199 (2005), hep-th/0406011.

[37] J. Martin, Lect.Notes Phys. **738**, 193 (2008), 0704.3540.

[38] G. F. Smoot, C. Bennett, A. Kogut, E. Wright, J. Aymon, et al., Astrophys. J. **396**, L1 (1992).

[39] J. Martin and C. Ringeval, JCAP **0608**, 009 (2006), astro-ph/0605367.

[40] P. Astier and R. Pain (2012), 1204.5493.

[41] M. Kunz (2012), 1204.5482.

[42] C. Clarkson (2012), 1204.5505.

[43] C. de Rham (2012), 1204.5492.

[1563] A. Sakharov, Sov. Phys. Dokl. **12**, 1040 (1968).

[45] J. Kapusta and C. Gale, *Finite-Temperature Field Theory: Principles and Applications* (Cambrige Univ. Pr., 2006).

[46] S. Chatrchyan et al. (CMS Collaboration) (2012), 1202.1488.

[47] E. K. Akhmedov (2002), hep-th/0204048.

[48] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables* (National Bureau of Standards, Washington, US, 1970), ninth ed.

[49] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic Press, New York and London, 1965).

[50] G. Ossola and A. Sirlin, Eur. Phys. J. **C31**, 165 (2003), 28 pages, 5 figures, hep-ph/0305050.

[51] J. F. Koksma and T. Prokopec (2011), 1105.6296.

[52] L. Rosenberg, Phys. Rev. **129**, 2786 (1963).

[53] N. D. Birrell and P. C. W. Davies, *Quantum Fields In Curved Space* (Cambridge Univ. Pr., 1982).

[54] C. M. Bender and T. T. Wu, Phys. Rev. **184**, 1231 (1969).

[55] T. Abbott, Lecture notes (2006), URL web.mit.edu/tabbott/Public/8.06/current.pdf.

[56] J. Minahan, MIT Lecture notes (2011), URL www.teorfys.uu.se/people/minahan/Courses/QFT/wholeshebang.pdf.

[57] C. Cohen-Tannoudji, B. Diu, and F. Laloe, *Quantum Mechanics* (Wiley-Interscience, 2006).

[58] P. M. Stevenson, Phys. Rev. **D30**, 1712 (1984).

[59] P. M. Stevenson, Phys. Rev. **D32**, 1389 (1985).

[60] P. M. Stevenson and I. Roditi, Phys. Rev. **D33**, 2305 (1986).

[61] P. M. Stevenson, Z. Phys. **C35**, 467 (1987).

[62] P. M. Stevenson, B. Alles, and R. Tarrach, Phys. Rev. **D35**, 2407 (1987).

[63] P. M. Stevenson and R. Tarrach, Phys. Lett. **B176**, 436 (1986).

[64] G. Hajj and P. M. Stevenson, Phys. Rev. **D37**, 413 (1988).

[65] P. M. Stevenson, G. Hajj, and J. Reed, Phys. Rev. **D34**, 3117 (1986).

[66] R. Ibanez-Meier, I. Stancu, and P. M. Stevenson, Z. Phys. **C70**, 307 (1996), hep-ph/9207276.

[67] S. R. Coleman and E. J. Weinberg, Phys. Rev. **D7**, 1888 (1973).

[68] E. J. Weinberg and A.-q. Wu, Phys. Rev. **D36**, 2474 (1987).

[69] T. Curtright and C. B. Thorn, J. Math. Phys. **25**, 541 (1984).

[70] J. Wess and B. Zumino, Nucl. Phys. **B70**, 39 (1974).

[71] S. R. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2239 (1969).

[72] J. Callan, Curtis G., S. R. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2247 (1969).

[73] J. Wess and B. Zumino, Phys. Lett. **B49**, 52 (1974).

[74] I. J. Aitchison (2005), notes of Lectures for Graduate Students in Particle Physics Oxford, 2004 and 2005, hep-ph/0505105.

[75] P. Binetruy, *Supersymmetry: Theory, experiment and cosmology* (Oxford Univ. Pr., 2006).

[76] L. Parker, In Cargese 1978, Proceedings, Recent Developments In Gravitation, 219-273 (1978).

[77] T. S. Bunch and L. Parker, Phys. Rev. **D20**, 2499 (1979).

[78] E. Poisson, A. Pound, and I. Vega, Living Rev. Rel. **14**, 7 (2011), 1102.0529.

[79] R. Scranton et al. (SDSS) (2003), astro-ph/0307335.

[80] P. Solevi, R. Mainini, and S. A. Bonometto (2004), astro-ph/0412054.

[1677] D. N. Spergel et al. (WMAP), Astrophys. J. Suppl. **170**, 377 (2007), astro-ph/0603449.

[82] M. Tegmark et al. (SDSS), Phys. Rev. **D69**, 103501 (2004), astro-ph/0310723.

[83] E. Kolb and M. Turner, *The Early Universe*, vol. 69 of *Frontiers in Physics Series* (Addison-Wesley Publishing Company, 1990).

[1497] B. Ratra and P. J. E. Peebles, Phys. Rev. **D37**, 3406 (1988).

[85] P. Brax and J. Martin, Phys. Lett. **B468**, 40 (1999), astro-ph/9905040.

[86] P. Brax and J. Martin, Phys. Rev. **D61**, 103502 (2000), astro-ph/9912046.

[87] P. Brax, J. Martin, and A. Riazuelo, Phys. Rev. **D62**, 103505 (2000), astro-ph/0005428.

[88] P. Brax, J. Martin, and A. Riazuelo, Phys. Rev. **D64**, 083505 (2001), hep-ph/0104240.

[89] J. Martin and M. A. Musso, Phys. Rev. **D71**, 063514 (2005), astro-ph/0410190.

[90] J. Martin, C. Schimd, and J.-P. Uzan, Phys. Rev. Lett. **96**, 061303 (2006), astro-ph/0510208.

[91] P. Brax and J. Martin, Phys. Lett. **B647**, 320 (2007), hep-th/0612208.

[92] P. Brax and J. Martin, JCAP **0611**, 008 (2006), astro-ph/0606306.

[93] P. Brax and J. Martin, Phys. Rev. **D75**, 083507 (2007), hep-th/0605228.

[94] P. Brax, C. van de Bruck, J. Martin, and A.-C. Davis, JCAP **0909**, 032 (2009), 0904.3471.

[95] C. F. Kolda and D. H. Lyth, Phys. Lett. **B458**, 197 (1999), hep-ph/9811375.

[448] S. M. Carroll, Phys. Rev. Lett. **81**, 3067 (1998), astro-ph/9806099.

[97] M. Doran and J. Jaeckel (2002), astro-ph/0205206.

[98] M. Doran and J. Jaeckel, Lect. Notes Phys. **646**, 273 (2004).

[99] M. Garny, Phys. Rev. **D74**, 043009 (2006), hep-ph/0606120.

[100] C. Ringeval, T. Suyama, T. Takahashi, M. Yamaguchi, and S. Yokoyama, Phys. Rev. Lett. **105**, 121301 (2010), 1006.0368.

[101] I. Maor, R. Brustein, and P. J. Steinhardt, Phys. Rev. Lett. **86**, 6 (2001), astro-ph/0007297.

[102] S. Weinberg, *Gravitation and Cosmology* (John Wiley & Sons, 1972).

[103] E. L. Wright (1998), astro-ph/9805292.

[104] A. W. Kerr, J. C. Hauck, and B. Mashhoon, Class. Quant. Grav. **20**, 2727 (2003), gr-qc/0301057.

[105] P. Jetzer (2007), 0711.3723.

[106] L. Parker, Phys. Rev. **D22**, 1922 (1980).

[107] L. Parker and L. O. Pimentel, Phys. Rev. **D25**, 3180 (1982).

[108] L. Parker, Phys. Rev. **D24**, 535 (1981).

[109] S. Moradi and E. Aboualizadeh, Gen. Rel. Grav. **42**, 435 (2010).

[110] E. Fischbach, B. S. Freeman, and W.-K. Cheng, Phys. Rev. **D23**, 2157 (1981).

[111] N. Boulanger, P. Spindel, and F. Buisseret, Phys. Rev. **D74**, 125014 (2006), hep-th/0610207.

[112] T. A. Welton, Phys. Rev. **74**, 1157 (1948).

[113] G. Plunien, B. Muller, and W. Greiner, Phys. Rept. **134**, 87 (1986).

[114] K. A. Milton, J. Phys. A **A37**, R209 (2004), hep-th/0406024.

[115] K. A. Milton, pp. 333–349 (2000), hep-th/0009173.

[116] K. A. Milton, Lect. Notes Phys. **834**, 39 (2011), 1005.0031.

[117] A. Lambrecht and S. Reynaud (2011), 1112.1301.

[118] K. A. Milton, Phys. Rev. **D68**, 065020 (2003), hep-th/0210081.

[119] D. Deutsch and P. Candelas, Phys. Rev. **D20**, 3063 (1979).

[120] P. Milonni and M.-L. Shih, Phys. Rev. **A45**, 4241 (1992).

[121] R. Jaffe, Phys. Rev. **D72**, 021301 (2005), hep-th/0503158.

[122] M. Chernodub (2012), 1203.6588.

[123] E. Calloni, L. Di Fiore, G. Esposito, L. Milano, and L. Rosa, Phys. Lett. **A297**, 328 (2002), quant-ph/0109091.

[124] G. Bimonte, E. Calloni, G. Esposito, and L. Rosa, Phys. Rev. **D76**, 025008 (2007), hep-th/0703062.

[125] S. A. Fulling, K. A. Milton, P. Parashar, A. Romeo, K. Shajesh, et al., Phys. Rev. **D76**, 025004 (2007), hep-th/0702091.

[126] K. A. Milton, S. A. Fulling, P. Parashar, A. Romeo, K. Shajesh, et al., J. Phys. A **A41**, 164052 (2008), 0710.3841.

[127] K. Milton, P. Parashar, J. Wagner, K. Shajesh, A. Romeo, et al. (2008), 0810.0081.

[128] R. Caldwell (2002), astro-ph/0209312.

[129] F. Sorge, Class. Quant. Grav. **22**, 5109 (2005).

[130] E. Masso, Phys. Lett. **B679**, 433 (2009), 0902.4318.

[131] Y. Su, B. R. Heckel, E. Adelberger, J. Gundlach, M. Harris, et al., Phys. Rev. **D50**, 3614 (1994).

[132] T. Damour, Space Sci.Rev. (2009), 0906.3174.

[133] R. Colella, A. Overhauser, and S. Werner, Phys. Rev. Lett. **34**, 1472 (1975).

[134] D. M. Greenberger and A. Overhauser, Rev. Mod. Phys. **51**, 43 (1979).

[135] D. M. Greenberger, Rev. Mod. Phys. **55**, 875 (1983).

[136] H. Salecker and E. Wigner, Phys. Rev. **109**, 571 (1958).

[137] A. Peres, Am. J. Phys. **48**, 552 (1980).

[138] C. Leavens, Solid Sate Commun. **86**, 781 (1993).

[139] P. Davies, Class. Quant. Grav. **21**, 5677 (2004).

[140] P. C. W. Davies, Classical and Quantum Gravity **21**, 2761 (2004), arXiv:quant-ph/0403027.

[141] N. Grot, C. Rovelli, and R. S. Tate, Phys. Rev. **A54**, 4679 (1996), quant-ph/9603021.

[142] Y. Aharonov, J. Oppenheim, S. Popescu, B. Reznik, and W. Unruh, Phys. Rev. **A57**, 4130 (1998), quant-ph/9709031.

[143] J. Oppenheim, B. Reznik, and W. Unruh, Phys. Rev. **A59**, 1804 (1999), quant-ph/9807043.

[144] P. C. Davies, Am. J. Phys. **73**, 23 (2005).

[145] B. Mielnik and G. Torres-Vega, ArXiv e-prints (2011), 1112.4198.

[146] L. Viola and R. Onofrio, Phys. Rev. **D55**, 455 (1997), quant-ph/9612039.

[147] M. Ali, A. Majumdar, D. Home, and A. K. Pan, Class. Quant. Grav. **23**, 6493 (2006), quant-ph/0606183.

[148] P. Chowdhury, D. Home, A. Majumdar, S. Mousavi, M. Mozaffari, et al., Class. Quant. Grav. **29**, 025010 (2012).

[149] R. Onofrio, Int. J. Mod. Phys. **A25**, 2260 (2010), 1101.1484.

[150] D. Giulini (2011), 1105.0749.

[151] E. Kajari, N. L. Harshman, E. M. Rasel, S. Stenholm, G. Süßmann, and W. P. Schleich, Applied Physics B: Lasers and Optics **100**, 43 (2010), 1006.1988.

[152] V. Nesvizhevsky, H. Borner, A. Gagarski, A. Petoukhov, G. Petrov, et al., Phys.Rev. **D67**, 102002 (2003), hep-ph/0306198.

[153] H. Abele, S. Baessler, H. Borner, A. Gagarsky, V. Nesvizhevsky, et al., AIP Conf.Proc. **842**, 793 (2006).

[154] S. Nobbenhuis (2006), ph.D. Thesis (Advisor: G. t'Hooft), gr-qc/0609011.

[155] A. Dolgov (1982).

[728] L. Ford, Phys. Rev. **D35**, 2339 (1987).

[157] A. Dolgov and F. Urban, Phys. Rev. **D77**, 083503 (2008), 0801.3090.

[158] L. W. Abramo, R. H. Brandenberger, and V. F. Mukhanov, Phys. Rev. **D56**, 3248 (1997), gr-qc/9704037.

[159] V. F. Mukhanov, L. W. Abramo, and R. H. Brandenberger, Phys. Rev. Lett. **78**, 1624 (1997), gr-qc/9609026.

[160] J. Hartle and S. Hawking, Phys. Rev. **D28**, 2960 (1983).

[161] E. Baum, Phys. Lett. **B133**, 185 (1983).

[162] S. R. Coleman, Nucl. Phys. **B310**, 643 (1988).

[163] V. Rubakov and M. Shaposhnikov, Phys. Lett. **B125**, 139 (1983).

[164] C. Burgess, Annals Phys. **313**, 283 (2004), paper has been updated to include developments since the conference, hep-th/0402200.

[165] B. Zwiebach, *A first Course in String Theory* (Cambridge Univ. Pr., 2004).

[1818] S. Weinberg, Phys. Rev. Lett. **59**, 2607 (1987).

[167] A. Linde and M. Noorbala, JCAP **1009**, 008 (2010), 1006.2170.

[168] G. D. Starkman and R. Trotta, Phys. Rev. Lett. **97**, 201301 (2006), astro-ph/0607227.

[169] R. Trotta and G. D. Starkman, AIP Conf.Proc. **878**, 323 (2006), astro-ph/0610330.

[170] N. Afshordi (2008), 0807.2639.

[171] G. F. Ellis, H. van Elst, J. Murugan, and J.-P. Uzan, Class. Quant. Grav. **28**, 225007 (2011), 1008.1196.

[172] S. Aslanbeigi, G. Robbers, B. Z. Foster, K. Kohri, and N. Afshordi, Phys. Rev. **D84**, 103522 (2011), 1106.3955.

[173] G. Volovik, Phys. Rept. **351**, 195 (2001), gr-qc/0005091.

[174] G. Volovik, Annalen Phys. **14**, 165 (2005), gr-qc/0405012.

[175] L. Sindoni (2011), 54 pages. Invited review for SIGMA Special Issue 'Loop Quantum Gravity and Cosmology', 1110.0686.

[176] G. Volovik (2001), gr-qc/0101111.

[177] G. Volovik, Int. Ser. Monogr. Phys. **117**, 1 (2006).

[178] S. H. Alexander and G. Calcagni, Found. Phys. **38**, 1148 (2008), 0807.0225.

[179] S. Finazzi, S. Liberati, and L. Sindoni, Phys. Rev. Lett. **108**, 071101 (2012), revtex4, 5 pages, 1 figures, 1103.4841.

[180] S. Finazzi, S. Liberati, and L. Sindoni (2012), 1204.3039.

References from "Expanding Confusion" by Davis Lineweaver

References

[1] Bennett, C. L., Halpern, M., Hinshaw, G. et al. 2003, ApJS, 148, 1

[2] Chen, H.-W., Lanzetta, K. M. and Pascarelle, S. 1999, Nature, 398, 586

[3] Davis, T. M. and Lineweaver, C. H. 2001, in Cosmology and Particle Physics 2000, ed. R. Durrer, J. Garcia-Bellido and M. Shaposhnikov, (New York: American Institute of Physics conference proceedings, Volume 555), 348

[4] Davis, T. M and Lineweaver, C. H. 2004, in preparation

[5] Davis, T. M., Lineweaver, C. H. and Webb, J. K. 2003, Am. J. Phys., 71, 358

[6] Ebert, R. and Trümper, M. 1975, Ann. NY Acad. Sci., 262, 470

[7] Ellis, G. F. R. and Rothman, T. 1993, Am. J. Phys., 61, 883

[8] Fan, X., Strauss, M. A., Schneider, D. P. et al. 2003, AJ, 125, 1649

[9] Goldhaber, G., Dustua, S., Gabi, S. et al. 1997, in ed. R. Canal, P. Ruiz-LaPuente and J. Isern, NATO ASIC Proc. 486: Thermonuclear Supernovae, 777

[10] Goldhaber, G., Groom, D. E., Kim, A. et al. 2001, ApJ, 558, 359

[11] Gudmundsson, E. H. and Björnsson, G. 2002, ApJ, 565, 1

[12] Harrison, E. R. 1981, Cosmology: the science of the universe, 1st Ed. (Cambridge: Cambridge U. Press), and 2nd Ed. (2000)

[13] Harrison, E. R. 1993, ApJ, 403, 28

[14] Kiang, T. 1991, Acta Astrophysica Sinica, 11, 197–212

[15] Kiang, T. 1997, Chinese Astronomy and Astrophysics, 21, 1

[16] Kiang, T. 2001, Chinese Astronomy and Astrophysics, 25, 141

[17] Lake, K. 1981, ApJ, 247, 17

[18] Landsberg, P. T. and Evans, D. A. 1977, Mathematical Cosmology, an introduction (Oxford: Clarendon Press)

[19] Leibundgut, B., Schommer, R., Phillips, M. et al. 1996, ApJ, 466, L21

[20] Liske, J. 2000, MNRAS, 319, 557

[21] Loeb, A. 1998, ApJ, 499, L111

[22] Murdoch, H. S. 1977, QJRAS, 18, 242

[23] Murdoch, H. S. 1993, Quasars - how far? how fast?, unpublished, Sect. 1

[24] Outram, P. J., Chaffee, F. H. and Carswell, R. F. 1999, MNRAS, 310, 289

[25] Page, D. N. 1993, unpublished, gr-qc/9303008

[26] Peacock, J. A. 1999, Cosmological Physics, (Cambridge: Cambridge U. Press)

[27] Perlmutter, S., Aldering, G., Goldhaber, G. et al. 1999, ApJ, 517, 565

[28] Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. 1989, Numerical Recipes (Fortran Version), (Cambridge: Cambridge U. Press)

[29] Riess, A. G., Filippenko, A. V., Challis, P. et al. 1998, AJ, 116, 1009

[30] Riess, A. G., Filippenko, A. V., Leonard, D. et al. 1997, AJ, 114, 722

[31] Rindler, W. 1956, MNRAS, 6, 662

[32] Rindler, W. 1977, Essential relativity, special, general and cosmological, (New York: Springer)

[33] Sandage, A. 1962, ApJ, 136, 319

[34] Silverman, A. N. 1986, Am. J. Phys., 54, 1092

[35] Stuckey, W. M., Am. J. Phys., 60, 142

[36] Weinberg, S. 1972, Gravitation and cosmology, (New York: Wiley)

[37] Wilson, O. C. 1939, ApJ, 90, 634

[38] Wright, E. L. 2001, Errors in tired light cosmology, www.astro.ucla.edu/∼wright/tiredlit.htm

References from "Scalar Speed Limits and Cosmology Acceleration from D-cceleration"

References

[1] D. Kabat and G. Lifschytz, "*Gauge theory origins of supergravity causal structure*" JHEP **9905**, 005 (1999) [arXiv:hep-th/9902073].

[2] S. Kachru, R. Kallosh, A. Linde, J. Maldacena, L. McAllister and S. P. Trivedi, "*Towards inflation in string theory*" arXiv:hep-th/0308055.

[3] C. Armendariz-Picon, T. Damour and V. Mukhanov, "*k-inflation*" Phys. Lett. B **458**, 209 (1999) [arXiv:hep-th/9904075].

[4] T. Chiba, T. Okabe and M. Yamaguchi, "*Kinetically driven quintessence*" Phys. Rev. D **62**, 023511 (2000) [arXiv:astro-ph/9912463].
C. Armendariz-Picon, V. Mukhanov and P. J. Steinhardt, "*A dynamical solution to the problem of a small cosmological constant and late-time cosmic acceleration*", Phys. Rev. Lett. **85**, 4438 (2000) [arXiv:astro-ph/0004134].
C. Armendariz-Picon, V. Mukhanov and P. J. Steinhardt, "*Essentials of k-essence*" Phys. Rev. D **63**, 103510 (2001) [arXiv:astro-ph/0006373].

[5] A. Sen, "*Rolling tachyon*" JHEP **0204**, 048 (2002) [arXiv:hep-th/0203211].
A. Sen, "*Tachyon matter*" JHEP **0207**, 065 (2002) [arXiv:hep-th/0203265].

[6] A. Maloney, A. Strominger and X. Yin, "*S-brane thermodynamics*" arXiv:hep-th /0302146.

[7] N. Lambert, H. Liu and J. Maldacena, "*Closed strings from decaying D-branes*", arXiv:hep-th/0303139.

[8] G. W. Gibbons, "*Cosmological evolution of the rolling tachyon*" Phys. Lett. B **537**, 1 (2002) [arXiv:hep-th/0204008].

[9] A. Feinstein, "*Power-law inflation from the rolling tachyon*", Phys. Rev. D **66**, 063511 (2002) [arXiv:hep-th/0204140].
T. Padmanabhan, "*Accelerated expansion of the universe driven by tachyonic matter*" Phys. Rev. D **66**, 021301 (2002) [arXiv:hep-th/0204150].
L. R. Abramo and F. Finelli, "*Cosmological dynamics of the tachyon with an inverse power-law potential*", arXiv:astro-ph/0307208.

[10] J. H. Horne and G. W. Moore, "*Chaotic coupling constants*" Nucl. Phys. B **432**, 109 (1994) [arXiv:hep-th/9403058].

[11] T. Banks, M. Berkooz, S. H. Shenker, G. W. Moore and P. J. Steinhardt, "*Modular cosmology*" Phys. Rev. D **52**, 3548 (1995) [arXiv:hep-th/9503114].

[12] T. Banks, W. Fischler and L. Motl, "*Dualities versus singularities*", JHEP **9901**, 019 (1999) [arXiv:hep-th/9811194].

[13] F. Quevedo, "*Lectures on string / brane cosmology*" Class. Quant. Grav. **19**, 5721 (2002) [arXiv:hep-th/0210292].

[14] C. Bachas, "*D-brane dynamics*", Phys. Lett. B **374**, 37 (1996) [arXiv:hep-th/9511043].

[15] M. R. Douglas, D. Kabat, P. Pouliot and S. H. Shenker, "*D-branes and short distances in string theory*", Nucl. Phys. B **485**, 85 (1997) [arXiv:hep-th/9608024].

[16] C. P. Burgess, P. Martineau, F. Quevedo and R. Rabadan, "*Branonium*", JHEP **0306**, 037 (2003) [arXiv:hep-th/0303170].

[17] M. Brandle and A. Lukas, "*Flop transitions in M-theory cosmology*", Phys. Rev. D **68**, 024030 (2003) [arXiv:hep-th/0212263].

[18] L. Jarv, T. Mohaupt and F. Saueressig, "*M-theory cosmologies from singular Calabi-Yau compactifications*", arXiv:hep-th/0310174.

[19] I. L. Buchbinder, A. Y. Petrov and A. A. Tseytlin, "*Two-loop N = 4 super Yang Mills effective action and interaction between D3-branes*" Nucl. Phys. B **621**, 179 (2002) [arXiv:hep-th/0110173].

[20] O. Aharony, S. S. Gubser, J. M. Maldacena, H. Ooguri and Y. Oz, "*Large N field theories, string theory and gravity*" Phys. Rept. **323**, 183 (2000) [arXiv:hep-th/9905111].

[21] L. Randall and R. Sundrum, "*A large mass hierarchy from a small extra dimension*" Phys. Rev. Lett. **83**, 3370 (1999) [arXiv:hep-ph/9905221].

[22] J. M. Maldacena, "*The large N limit of superconformal field theories and supergravity*" Adv. Theor. Math. Phys. **2**, 231 (1998) [Int. J. Theor. Phys. **38**, 1113 (1999)] [arXiv:hep-th/9711200].

[23] P. C. Aichelburg and R. U. Sexl, "*On The Gravitational Field Of A Massless Particle*" Gen. Rel. Grav. **2**, 303 (1971).

[24] A. Sen, "*Open-closed duality: Lessons from matrix model*" arXiv:hep-th/0308068.

[25] S. Kachru, J. Pearson and H. Verlinde, "*Brane/flux annihilation and the string dual of a non-supersymmetric field theory*", JHEP **0206**, 021 (2002) [arXiv:hep-th/0112197].

[26] H. Verlinde, "*Holography and compactification*" Nucl. Phys. B **580**, 264 (2000) [arXiv: hep-th/9906182].

[27] S. B. Giddings, S. Kachru and J. Polchinski, "*Hierarchies from fluxes in string compactifications*" Phys. Rev. D **66**, 106006 (2002) [arXiv:hep-th/0105097].

[28] N. Seiberg and E. Witten, "*The D1/D5 system and singular CFT*" JHEP **9904**, 017 (1999) [arXiv:hep-th/9903224].

[29] A. Fotopoulos and A. A. Tseytlin, "*On gravitational couplings in D-brane action*" JHEP **0212**, 001 (2002) [arXiv:hep-th/0211101].

[30] R. Bousso and J. Polchinski, "*Quantization of four-form fluxes and dynamical neutralization of the cosmological constant*", JHEP **0006**, 006 (2000) [arXiv:hep-th/0004134].

[31] I. R. Klebanov and M. J. Strassler, "*Supergravity and a confining gauge theory: Duality cascades and χSB-resolution of naked singularities*" JHEP **0008**, 052 (2000) [arXiv:hep-th/0007191].

[32] M. J. Strassler, "*Non-supersymmetric theories with light scalar fields and large hierarchies*" arXiv:hep-th/0309122.

[33] S. Dimopoulos, S. Kachru, N. Kaloper, A. E. Lawrence and E. Silverstein, "*Small numbers from tunneling between brane throats*" Phys. Rev. D **64**, 121702 (2001) [arXiv:hep-th/0104239].
S. Dimopoulos, S. Kachru, N. Kaloper, A. E. Lawrence and E. Silverstein, "*Generating small numbers by tunneling in multi-throat compactifications*", arXiv:hep-th/0106128.

[34] A. G. Muslimov, "*On The Scalar Field Dynamics In A Spatially Flat Friedman Universe*" Class. Quant. Grav. **7**, 231 (1990).
D. S. Salopek and J. R. Bond, "*Nonlinear Evolution Of Long Wavelength Metric Fluctuations In Inflationary Models*", Phys. Rev. D **42**, 3936 (1990).
W. H. Kinney, "*A Hamilton-Jacobi approach to non-slow-roll inflation*" Phys. Rev. D **56**, 2002 (1997) [arXiv:hep-ph/9702427].

[35] O. DeWolfe, D. Z. Freedman, S. S. Gubser and A. Karch, "*Modeling the fifth dimension with scalars and gravity*" Phys. Rev. D **62**, 046008 (2000) [arXiv:hep-th/9909134].

[36] M. R. Garousi, "*Tachyon couplings on non-BPS D-branes and Dirac-Born-Infeld action*" Nucl. Phys. B **584**, 284 (2000) [arXiv:hep-th/0003122].

[37] S. Kachru and L. McAllister, "*Bouncing brane cosmologies from warped string compactifications*" JHEP **0303**, 018 (2003) [arXiv:hep-th/0205209].

[38] A. Kehagias and E. Kiritsis, "*Mirage cosmology*" JHEP **9911**, 022 (1999) [arXiv:hep-th/9910174].

[39] M. Alishahiha, E. Silverstein and D. Tong, "DBI in the sky," arXiv:hep-th/0404084.

D.0.2 From Famous Reviews

From Challenges for LCDM - An update by Perivolaropoulos, Skara

References

[1] Aaronson, M., J. Huchra, and J. Mould (1979), ApJ **229**, 1.

[2] Aasi, J., *et al.* (LIGO Scientific) (2015), Class. Quant. Grav. **32**, 074001, arXiv:1411.4547 [gr-qc] .

[3] Abadi, T., and E. D. Kovetz (2021), Phys. Rev. D **103** (2), 023530, arXiv:2011.13853 [astro-ph.CO] .

[4] Abate, A., *et al.* (LSST Dark Energy Science) (2012), arXiv:1211.0310 [astro-ph.CO] .

[5] Abazajian, K. (2006), Phys. Rev. D **73**, 063513, arXiv:astro-ph/0512631 .

[6] Abazajian, K., *et al.* (2019), arXiv:1907.04473 [astro-ph.IM] .

[7] Abazajian, K., *et al.* (CMB-S4) (2022), Astrophys. J. **926** (1), 54, arXiv:2008.12619 [astro-ph.CO] .

[8] Abazajian, K. N., *et al.* (2012), arXiv:1204.5379 [hep-ph] .

[1] Abazajian, K. N., *et al.* (CMB-S4) (2016), arXiv:1610.02743 [astro-ph.CO] .

[10] Abbott, B., *et al.* (LIGO Scientific, Virgo, 1M2H, Dark Energy Camera GW-E, DES, DLT40, Las Cumbres Observatory, VINROUGE, MASTER) (2017), Nature **551** (7678), 85, arXiv:1710.05835 [astro-ph.CO] .

[11] Abbott, B., *et al.* (LIGO Scientific, Virgo) (2019), Phys. Rev. X **9** (3), 031040, arXiv:1811.12907 [astro-ph.HE] .

[12] Abbott, B. P., *et al.* (LIGO Scientific) (2009), Rept. Prog. Phys. **72**, 076901, arXiv:0711.3041 [gr-qc] .

[13] Abbott, B. P., *et al.* (LIGO Scientific, Virgo) (2016), Phys. Rev. Lett. **116** (6), 061102, arXiv:1602.03837 [gr-qc] .

[14] Abbott, B. P., *et al.* (LIGO Scientific) (2017), Class. Quant. Grav. **34** (4), 044001, arXiv:1607.08697 [astro-ph.IM] .

[15] Abbott, B. P., *et al.* (LIGO Scientific, Virgo) (2017), Phys. Rev. Lett. **119** (14), 141101, arXiv:1709.09660 [gr-qc] .

[16] Abbott, B. P., *et al.* (KAGRA, LIGO Scientific, VIRGO) (2018), Living Rev. Rel. **21** (1), 3, arXiv:1304.0670 [gr-qc] .

[17] Abbott, B. P., *et al.* (LIGO Scientific, Virgo) (2021), Astrophys. J. **909** (2), 218, arXiv:1908.06060 [astro-ph.CO] .

[18] Abbott, R., *et al.* (LIGO Scientific, Virgo) (2020), Astrophys. J. Lett. **896** (2), L44, arXiv:2006.12611 [astro-ph.HE] .

[19] Abbott, R., *et al.* (LIGO Scientific, Virgo) (2020), arXiv:2010.14527 [gr-qc] .

[20] Abbott, R., *et al.* (LIGO Scientific, Virgo) (2020), Astrophys. J. Lett. **900** (1), L13, arXiv:2009.01190 [astro-ph.HE] .

[21] Abbott, R., *et al.* (LIGO Scientific, VIRGO, KAGRA) (2021), arXiv:2111.03604 [astro-ph.CO] .

[22] Abbott, R., *et al.* (LIGO Scientific, VIRGO, KAGRA) (2021), arXiv:2111.03606 [gr-qc] .

[23] Abbott, T., *et al.* (DES) (2005), arXiv:astro-ph/0510346 .

[105] Abbott, T., *et al.* (DES) (2018), Mon. Not. Roy. Astron. Soc. **480** (3), 3879, arXiv:1711.00403 [astro-ph.CO] .

[25] Abbott, T., *et al.* (DES, NOAO Data Lab) (2018), Astrophys. J. Suppl. **239** (2), 18, arXiv:1801.03181 [astro-ph.IM] .

[26] Abbott, T., *et al.* (DES) (2020), Phys. Rev. D **102** (2), 023509, arXiv:2002.11124 [astro-ph.CO] .

[106] Abbott, T. M. C., *et al.* (DES) (2018), Phys. Rev. **D98** (4), 043526, arXiv:1708.01530 [astro-ph.CO] .

[107] Abbott, T. M. C., *et al.* (DES) (2019), Mon. Not. Roy. Astron. Soc. **483** (4), 4866, arXiv:1712.06209 [astro-ph.CO] .

[29] Abbott, T. M. C., *et al.* (DES) (2022), Phys. Rev. D **105** (2), 023520, arXiv:2105.13549 [astro-ph.CO] .

[30] Abdalla, E., *et al.* (2022), in *2022 Snowmass Summer Study*, arXiv:2203.06142 [astro-ph.CO] .

[31] Abdullah, M. H., A. Klypin, and G. Wilson (2020), Astrophys. J. **901** (2), 90, arXiv:2002.11907 [astro-ph.CO] .

[32] Abdullah, M. H., G. Wilson, A. Klypin, L. Old, E. Praton, and G. Ali (2020), Astrophys. J. Suppl. **246** (1), 2, arXiv:1907.05061 [astro-ph.CO] .

[33] Abel, T., G. L. Bryan, and M. L. Norman (2002), Science **295**, 93, arXiv:astro-ph/0112088 .

[34] Abell, P. A., *et al.* (LSST Science, LSST Project) (2009), arXiv:0912.0201 [astro-ph.IM] .

[35] Abellán, G. F., R. Murgia, and V. Poulin (2021), Phys. Rev. D **104** (12), 123533, arXiv:2102.12498 [astro-ph.CO] .

[36] Abellán, G. F., R. Murgia, V. Poulin, and J. Lavalle (2022), Phys. Rev. D **105** (6), 063525, arXiv:2008.09615 [astro-ph.CO] .

[37] Abitbol, M. H., *et al.* (Simons Observatory) (2019), Bull. Am. Astron. Soc. **51**, 147, arXiv:1907.08284 [astro-ph.IM] .

[38] Aboubrahim, A., M. Klasen, and P. Nath (2022), arXiv:2202.04453 [astro-ph.CO] .

[39] Abramovici, A., *et al.* (1992), Science **256**, 325.

[40] Acernese, F., *et al.* (2008), Class. Quant. Grav. **25**, 114045.

[41] Acernese, F., *et al.* (VIRGO) (2015), Class. Quant. Grav. **32** (2), 024001, arXiv:1408.3978 [gr-qc] .

[42] Ackerman, L., M. R. Buckley, S. M. Carroll, and M. Kamionkowski (2009), Phys. Rev. D **79**, 023519, arXiv:0810.5126 [hep-ph] .

[43] Acquaviva, G., O. Akarsu, N. Katirci, and J. A. Vazquez (2021), Phys. Rev. D **104** (2), 023505, arXiv:2104.02623 [astro-ph.CO] .

[44] Adam, R., *et al.* (Planck) (2016), Astron. Astrophys. **594**, A1, arXiv:1502.01582 [astro-ph.CO] .

[45] Adam, R., *et al.* (Euclid) (2019), Astron. Astrophys. **627** (627), A23, arXiv:1906.04707 [astro-ph.CO] .

[46] Addazi, A., *et al.* (2021), Prog. Part. Nucl. Phys. **103948**, 2022, arXiv:2111.05659 [hep-ph] .

[6] Addison, G., Y. Huang, D. Watts, C. Bennett, M. Halpern, G. Hinshaw, and J. Weiland (2016), Astrophys. J. **818** (2), 132, arXiv:1511.00055 [astro-ph.CO] .

[9] Addison, G., D. Watts, C. Bennett, M. Halpern, G. Hinshaw, and J. Weiland (2018), Astrophys. J. **853** (2), 119, arXiv:1707.06547 [astro-ph.CO] .

[49] Addison, G. E., G. Hinshaw, and M. Halpern (2013), Mon. Not. Roy. Astron. Soc. **436**, 1674, arXiv:1304.6984 [astro-ph.CO] .

[50] Ade, P., *et al.* (Planck) (2016), Astron.Astrophys. **594**, A13, arXiv:1502.01589 [astro-ph.CO] .

[51] Ade, P., *et al.* (Planck) (2016), Astron. Astrophys. **594**, A15, arXiv:1502.01591 [astro-ph.CO] .

[52] Ade, P., *et al.* (Planck) (2016), Astron. Astrophys. **594**, A16, arXiv:1506.07135 [astro-ph.CO] .

[53] Ade, P., *et al.* (Planck) (2016), Astron. Astrophys. **594**, A24, arXiv:1502.01597 [astro-ph.CO] .

[54] Ade, P., *et al.* (Simons Observatory) (2019), JCAP **02**, 056, arXiv:1808.07445 [astro-ph.CO] .

[55] Ade, P. A. R., *et al.* (Planck) (2014), Astron. Astrophys. **571**, A16, arXiv:1303.5076 [astro-ph.CO] .

[56] Ade, P. A. R., *et al.* (Planck) (2014), Astron. Astrophys. **571**, A23, arXiv:1303.5083 [astro-ph.CO] .

[57] Ade, P. A. R., *et al.* (Planck) (2016), Astron. Astrophys. **594**, A14, arXiv:1502.01590 [astro-ph.CO] .

[58] Adhikari, S., A.-S. Deutsch, and S. Shandera (2018), Phys. Rev. D **98** (2), 023520, arXiv:1805.00037 [astro-ph.CO] .

[59] Adhikari, S., and D. Huterer (2020), Phys. Dark Univ. **28**, 100539, arXiv:1905.02278 [astro-ph.CO] .

[60] Adhikary, P., S. Das, S. Basilakos, and E. N. Saridakis (2021), Phys. Rev. D **104** (12), 123519, arXiv:2104.13118 [gr-qc] .

[61] Adler, R. J., P. Chen, and D. I. Santiago (2001), Gen. Rel. Grav. **33**, 2101, arXiv:gr-qc/0106080 [gr-qc] .

[62] Adler, S. L. (2019), Phys. Rev. D **100** (12), 123503, arXiv:1905.08228 [astro-ph.CO] .

[63] Affeldt, C., *et al.* (2014), Class. Quant. Grav. **31** (22), 224002.

[64] Aghamousa, A., *et al.* (DESI) (2016), arXiv:1611.00036 [astro-ph.IM] .

[65] Aghamousa, A., *et al.* (DESI) (2016), arXiv:1611.00037 [astro-ph.IM] .

[66] Aghanim, N., *et al.* (Planck) (2016), Astron. Astrophys. **596**, A107, arXiv:1605.02985 [astro-ph.CO] .

[67] Aghanim, N., *et al.* (Planck) (2017), Astron. Astrophys. **607**, A95, arXiv:1608.02487 [astro-ph.CO] .

[68] Aghanim, N., *et al.* (Planck) (2020), Astron. Astrophys. **641**, A1, arXiv:1807.06205 [astro-ph.CO] .

[69] Aghanim, N., *et al.* (Planck) (2020), Astron. Astrophys. **641**, A1, arXiv:1807.06205 [astro-ph.CO] .

[70] Aghanim, N., *et al.* (Planck) (2020), Astron. Astrophys. **641**, A3, arXiv:1807.06207 [astro-ph.CO] .

[71] Aghanim, N., *et al.* (Planck) (2020), Astron. Astrophys. **641**, A5, arXiv:1907.12875 [astro-ph.CO] .

[72] Aghanim, N., *et al.* (Planck) (2020), Astron. Astrophys. **641**, A6, [Erratum: Astron.Astrophys. 652, C4 (2021)], arXiv:1807.06209 [astro-ph.CO] .

[73] Aghanim, N., *et al.* (Planck) (2020), Astron. Astrophys. **641**, A8, arXiv:1807.06210 [astro-ph.CO] .

[11] Agrawal, P., F.-Y. Cyr-Racine, D. Pinner, and L. Randall (2019), arXiv:1904.01016 [astro-ph.CO] .

[75] Agrawal, P., G. Obied, and C. Vafa (2021), Phys. Rev. D **103** (4), 043523, arXiv:1906.08261 [astro-ph.CO] .

[76] Aguiar, O. D. (2011), Res. Astron. Astrophys. **11**, 1, arXiv:1009.1138 [astro-ph.IM] .

[77] Aguiar, O. D., *et al.* (2002), Class. Quant. Grav. **19**, 1949.

[78] Ahumada, R., *et al.* (SDSS-IV) (2020), Astrophys. J. Suppl. **249** (1), 3, arXiv:1912.02905 [astro-ph.GA] .

[79] Aihara, H., *et al.* (2018), Publ. Astron. Soc. Jap. **70**, S8, arXiv:1702.08449 [astro-ph.IM] .

[80] Aiola, S., *et al.* (ACT) (2020), JCAP **12**, 047, arXiv:2007.07288 [astro-ph.CO] .

[81] Akaike, H. (1974), IEEE Transactions on Automatic Control **19** (6), 716.

[82] Akarsu, Ö., J. D. Barrow, L. A. Escamilla, and J. A. Vazquez (2020), Phys. Rev. D **101** (6), 063528, arXiv:1912.08751 [astro-ph.CO] .

[83] Akarsu, O., S. Kumar, E. Özülker, and J. A. Vazquez (2021), Phys. Rev. D **104** (12), 123512, arXiv:2108.09239 [astro-ph.CO] .

[84] Akarsu, Ö., S. Kumar, S. Sharma, and L. Tedesco (2019), Phys. Rev. D **100** (2), 023532, arXiv:1905.06949 [astro-ph.CO] .

[85] Akita, K., and M. Yamaguchi (2020), JCAP **08**, 012, arXiv:2005.07047 [hep-ph] .

[86] Akrami, Y., Y. Fantaye, A. Shafieloo, H. K. Eriksen, F. K. Hansen, A. J. Banday, and K. M. Górski (2014), Astrophys. J. Lett. **784**, L42, arXiv:1402.0870 [astro-ph.CO] .

[87] Akrami, Y., *et al.* (Planck) (2020), Astron. Astrophys. **641**, A4, arXiv:1807.06208 [astro-ph.CO] .

[88] Akrami, Y., *et al.* (Planck) (2020), Astron. Astrophys. **641**, A7, arXiv:1906.02552 [astro-ph.CO] .

[89] Akrami, Y., *et al.* (Planck) (2020), Astron. Astrophys. **641**, A10, arXiv:1807.06211 [astro-ph.CO] .

[90] Akrami, Y., *et al.* (Planck) (2020), Astron. Astrophys. **643**, A42, arXiv:2007.04997 [astro-ph.CO] .

[91] Akrami, Y., *et al.* (Planck) (2020), Astron. Astrophys. **643**, A42, arXiv:2007.04997 [astro-ph.CO] .

[92] Akutsu, T., *et al.* (KAGRA) (2020), arXiv:2008.02921 [gr-qc] .

[15] Alam, S., *et al.* (BOSS) (2017), Mon. Not. Roy. Astron. Soc. **470** (3), 2617, arXiv:1607.03155 [astro-ph.CO] .

[94] Alam, S., *et al.* (eBOSS) (2021), Phys. Rev. D **103** (8), 083533, arXiv:2007.08991 [astro-ph.CO] .

[95] Alam, S., *et al.* (eBOSS) (2021), Phys. Rev. D **103** (8), 083533, arXiv:2007.08991 [astro-ph.CO] .

[96] Alam, U., S. Bag, and V. Sahni (2017), Phys. Rev. D **95** (2), 023524, arXiv:1605.04707 [astro-ph.CO] .

[97] Albareti, F. D., *et al.* (SDSS) (2017), Astrophys. J. Suppl. **233** (2), 25, arXiv:1608.02013 [astro-ph.GA] .

[15] Albrecht, A., and P. J. Steinhardt (1987), Adv. Ser. Astrophys. Cosmol. **3**, 158.

[99] Alcaniz, J., N. Bernal, A. Masiero, and F. S. Queiroz (2021), Phys. Lett. B **812**, 136008, arXiv:1912.05563 [astro-ph.CO] .

[100] Alestas, G., I. Antoniou, and L. Perivolaropoulos (2021), Universe **7**, 366, arXiv:2104.14481 [astro-ph.CO] .

[101] Alestas, G., D. Camarena, E. Di Valentino, L. Kazantzidis, V. Marra, S. Nesseris, and L. Perivolaropoulos (2022), Phys. Rev. D **105** (6), 063538, arXiv:2110.04336 [astro-ph.CO] .

[102] Alestas, G., L. Kazantzidis, and L. Perivolaropoulos (2020), Phys. Rev. D **101** (12), 123516, arXiv:2004.08363 [astro-ph.CO] .

[103] Alestas, G., L. Kazantzidis, and L. Perivolaropoulos (2021), Phys. Rev. D **103** (8), 083517, arXiv:2012.13932 [astro-ph.CO] .

[104] Alestas, G., G. V. Kraniotis, and L. Perivolaropoulos (2020), Phys. Rev. D **102** (10), 104015, arXiv:2005.11702 [gr-qc] .

[105] Alestas, G., and L. Perivolaropoulos (2021), Mon. Not. Roy. Astron. Soc. **504**, 3956, arXiv:2103.04045 [astro-ph.CO] .

[106] Alestas, G., L. Perivolaropoulos, and K. Tanidis (2022), arXiv:2201.05846 [astro-ph.CO] .

[107] Alexander, S., and E. McDonough (2019), Phys. Lett. **B797**, 134830, arXiv:1904.08912 [astro-ph.CO] .

[108] Ali, A. F., M. Faizal, and M. M. Khalil (2015), JCAP **1509** (09), 025, arXiv:1505.06963 [gr-qc] .

[109] Aljaf, M., D. Gregoris, and M. Khurshudyan (2021), Eur. Phys. J. C **81** (6), 544, arXiv:2005.01891 [astro-ph.CO] .

[110] Allali, I. J., M. P. Hertzberg, and F. Rompineve (2021), Phys. Rev. D **104** (8), L081303, arXiv:2104.12798 [astro-ph.CO] .

[111] Allen, S. W., A. E. Evrard, and A. B. Mantz (2011), Ann. Rev. Astron. Astrophys. **49**, 409, arXiv:1103.4829 [astro-ph.CO] .

[112] Aloni, D., A. Berlin, M. Joseph, M. Schmaltz, and N. Weiner (2021), arXiv:2111.00014 [astro-ph.CO] .

[113] Alonso, D., J. Sanchez, and A. Slosar (LSST Dark Energy Science) (2019), Mon. Not. Roy. Astron. Soc. **484** (3), 4127, arXiv:1809.09603 [astro-ph.CO] .

[114] Aluri, P. K., and P. Jain (2012), Mon. Not. Roy. Astron. Soc. **419**, 3378, arXiv:1108.5894 [astro-ph.CO] .

[115] Alvarez, P. D., B. Koch, C. Laporte, and A. Rincón (2021), JCAP **06**, 019, arXiv:2009.02311 [gr-qc] .

[116] Amaldi, E., *et al.* (1990), Europhys. Lett. **12**, 5.

[117] Amann, F., *et al.* (2020), Rev. Sci. Instrum. **91** (9), 9, arXiv:2003.03434 [physics.ins-det] .

[118] Amaro-Seoane, P., *et al.* (2012), Class. Quant. Grav. **29**, 124016, arXiv:1202.0839 [gr-qc] .

[119] Amaro-Seoane, P., *et al.* (2013), GW Notes **6**, 4, arXiv:1201.3621 [astro-ph.CO] .

[120] Amaro-Seoane, P., *et al.* (LISA) (2017), arXiv:1702.00786 [astro-ph.IM] .

[121] Amati, L., R. D'Agostino, O. Luongo, M. Muccino, and M. Tantalo (2019), Mon. Not. Roy. Astron. Soc. **486** (1), L46, arXiv:1811.08934 [astro-ph.HE] .

[122] Amati, L., C. Guidorzi, F. Frontera, M. Della Valle, F. Finelli, R. Landi, and E. Montanari (2008), Mon. Not. Roy. Astron. Soc. **391**, 577, arXiv:0805.0377 [astro-ph] .

[123] Amati, L., *et al.* (2002), Astron. Astrophys. **390**, 81, arXiv:astro-ph/0205230 .

[21] Amendola, L. (2000), Phys. Rev. D **62**, 043511, arXiv:astro-ph/9908023 .

[125] Amendola, L., R. Gannouji, D. Polarski, and S. Tsujikawa (2007), Phys. Rev. D **75**, 083504, arXiv:gr-qc/0612180 .

[126] Amendola, L., and S. Tsujikawa (2015), *Dark Energy: Theory and Observations* (Cambridge University Press).

[127] Amendola, L., *et al.* (Euclid Theory Working Group) (2013), Living Rev. Rel. **16**, 6, arXiv:1206.1225 [astro-ph.CO] .

[128] Amirhashchi, H., and A. K. Yadav (2020), arXiv:2001.03775 [astro-ph.CO] .

[129] Amon, A., *et al.* (DES) (2022), Phys. Rev. D **105** (2), 023514, arXiv:2105.13543 [astro-ph.CO] .

[130] Amorisco, N. C., and N. W. Evans (2012), Mon. Not. Roy. Astron. Soc. **419**, 184, arXiv:1106.1062 [astro-ph.CO] .

[131] An, R., A. A. Costa, L. Xiao, J. Zhang, and B. Wang (2019), Mon. Not. Roy. Astron. Soc. **489** (1), 297, arXiv:1809.03224 [astro-ph.CO] .

[132] An, R., C. Feng, and B. Wang (2018), JCAP **02**, 038, arXiv:1711.06799 [astro-ph.CO] .

[133] Anand, G. S., R. B. Tully, L. Rizzi, A. G. Riess, and W. Yuan (2021), arXiv:2108.00007 [astro-ph.CO] .

[134] Anchordoqui, L. A. (2020), arXiv:2005.01217 [astro-ph.CO] .

[135] Anchordoqui, L. A. (2021), Phys. Rev. D **103** (3), 035025, arXiv:2010.09715 [hep-ph] .

[136] Anchordoqui, L. A., I. Antoniadis, D. Lüst, J. F. Soriano, and T. R. Taylor (2020), Phys. Rev. D **101**, 083532, arXiv:1912.00242 [hep-th] .

[137] Anchordoqui, L. A., V. Barger, H. Goldberg, X. Huang, D. Marfatia, L. H. M. da Silva, and T. J. Weiler (2015), Phys. Rev. D **92** (6), 061301, [Erratum: Phys.Rev.D 94, 069901 (2016)], arXiv:1506.08788 [hep-ph] .

[138] Anchordoqui, L. A., V. Barger, D. Marfatia, and J. F. Soriano (2022), arXiv:2203.04818 [astro-ph.CO] .

[139] Anchordoqui, L. A., E. Di Valentino, S. Pan, and W. Yang (2021), JHEAp **32**, 121, arXiv:2107.13932 [astro-ph.CO] .

[140] Anchordoqui, L. A., and S. E. Perez Bergliaffa (2019), Phys. Rev. D **100** (12), 123525, arXiv:1910.05860 [astro-ph.CO] .

[141] Andersson, N., and K. D. Kokkotas (1996), Phys. Rev. Lett. **77**, 4134, arXiv:gr-qc/9610035 .

[142] Ando, M., *et al.* (TAMA) (2001), Phys. Rev. Lett. **86**, 3950, arXiv:astro-ph/0105473 .

[143] Ando, S., and K. Ishiwata (2015), JCAP **05**, 024, arXiv:1502.02007 [astro-ph.CO] .

[144] Antoniadis, I., N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali (1998), Phys. Lett. B **436**, 257, arXiv:hep-ph/9804398 .

[145] Antoniou, I., and L. Perivolaropoulos (2010), JCAP **12**, 012, arXiv:1007.4347 [astro-ph.CO] .

[146] Antoniou, I., and L. Perivolaropoulos (2017), Phys. Rev. D **96** (10), 104002, arXiv:1708.02117 [gr-qc] .

[147] Applegate, D. E., A. von der Linden, P. L. Kelly, M. T. Allen, S. W. Allen, P. R. Burchat, D. L. Burke, H. Ebeling, A. Mantz, and R. G. Morris (2014), Mon. Not. Roy. Astron. Soc. **439** (1), 48, arXiv:1208.0605 [astro-ph.CO] .

[148] Arai, K. (TAMA) (2008), J. Phys. Conf. Ser. **120**, 032010.

[149] de Araujo, J. C. N., A. De Felice, S. Kumar, and R. C. Nunes (2021), Phys. Rev. D **104** (10), 104057, arXiv:2106.09595 [astro-ph.CO] .

[150] Archidiacono, M., S. Bohr, S. Hannestad, J. H. Jørgensen, and J. Lesgourgues (2017), JCAP **11**, 010, arXiv:1706.06870 [astro-ph.CO] .

[24] Archidiacono, M., S. Gariazzo, C. Giunti, S. Hannestad, R. Hansen, M. Laveder, and T. Tram (2016), JCAP **08**, 067, arXiv:1606.07673 [astro-ph.CO] .

[152] Archidiacono, M., S. Gariazzo, C. Giunti, S. Hannestad, and T. Tram (2020), JCAP **12**, 029, arXiv:2006.12885 [astro-ph.CO] .

[153] Archidiacono, M., E. Giusarma, S. Hannestad, and O. Mena (2013), Adv. High Energy Phys. **2013**, 191047, arXiv:1307.0637 [astro-ph.CO] .

[154] Archidiacono, M., S. Hannestad, R. S. Hansen, and T. Tram (2015), Phys. Rev. D **91** (6), 065021, arXiv:1404.5915 [astro-ph.CO] .

[155] Archidiacono, M., S. Hannestad, R. S. Hansen, and T. Tram (2016), Phys. Rev. D **93** (4), 045004, arXiv:1508.02504 [astro-ph.CO] .

[25] Archidiacono, M., D. C. Hooper, R. Murgia, S. Bohr, J. Lesgourgues, and M. Viel (2019), JCAP **10**, 055, arXiv:1907.01496 [astro-ph.CO] .

[157] Arendse, N., *et al.* (2020), Astron. Astrophys. **639**, A57, arXiv:1909.07986 [astro-ph.CO] .

[158] Aresté Saló, L., D. Benisty, E. I. Guendelman, and J. d. Haro (2021), JCAP **07**, 007, arXiv:2102.09514 [astro-ph.CO] .

[159] Aretakis, S. (2011), Commun. Math. Phys. **307**, 17, arXiv:1110.2007 [gr-qc] .

[160] Arevalo, F., A. Cid, and J. Moya (2017), Eur. Phys. J. C **77** (8), 565, arXiv:1610.09330 [astro-ph.CO] .

[161] Arias-Aragon, F., E. Fernandez-Martinez, M. Gonzalez-Lopez, and L. Merlo (2021), Eur. Phys. J. C **81** (1), 28, arXiv:2009.01848 [hep-ph] .

[162] Arjona, R., J. García-Bellido, and S. Nesseris (2020), Phys. Rev. D **102** (10), 103526, arXiv:2006.01762 [astro-ph.CO] .

[163] Arjona, R., and S. Nesseris (2020), Phys. Rev. D **101** (12), 123525, arXiv:1910.01529 [astro-ph.CO] .

[164] Arkani-Hamed, N., S. Dimopoulos, and G. R. Dvali (1998), Phys. Lett. B **429**, 263, arXiv:hep-ph/9803315 .

[165] Arkani-Hamed, N., S. Dimopoulos, and G. R. Dvali (1999), Phys. Rev. D **59**, 086004, arXiv:hep-ph/9807344 .

[166] Arkani-Hamed, N., H. Georgi, and M. D. Schwartz (2003), Annals Phys. **305**, 96, arXiv:hep-th/0210184 .

[167] Armano, M., *et al.* (2016), Phys. Rev. Lett. **116** (23), 231101.

[168] Armano, M., *et al.* (2018), Astropart. Phys. **98**, 28, arXiv:1711.07427 [astro-ph.IM] .

[169] Arnaboldi, M., M. J. Neeser, L. C. Parker, P. Rosati, M. Lombardi, J. P. Dietrich, and W. Hummel (2007), The Messenger **127**, 28.

[170] Artymowski, M., I. Ben-Dayan, and U. Kumar (2021), Phys. Rev. D **103** (12), L121303, arXiv:2010.02998 [hep-ph] .

[171] Asadi, P., *et al.* (2022), arXiv:2203.06680 [hep-ph] .

[172] Asencio, E., I. Banik, and P. Kroupa (2020), Mon. Not. Roy. Astron. Soc. **500** (4), 5249, arXiv:2012.03950 [astro-ph.CO] .

[173] Asgari, M., *et al.* (2020), Astron. Astrophys. **634**, A127, arXiv:1910.05336 [astro-ph.CO] .

[174] Asgari, M., *et al.* (KiDS) (2021), Astron. Astrophys. **645**, A104, arXiv:2007.15633 [astro-ph.CO] .

[175] Ashoorioon, A., J. L. Hovdebo, and R. B. Mann (2005), Nucl. Phys. **B727**, 63, arXiv:gr-qc/0504135 [gr-qc] .

[176] Ashoorioon, A., A. Kempf, and R. B. Mann (2005), Phys. Rev. **D71**, 023503, arXiv:astro-ph/0410139 [astro-ph] .

[177] Ashoorioon, A., and T. Koivisto (2016), Phys. Rev. D **94** (4), 043009, arXiv:1507.03514 [astro-ph.CO] .

[178] Ashoorioon, A., and R. B. Mann (2005), Phys. Rev. **D71**, 103509, arXiv:gr-qc/0410053 [gr-qc] .

[179] Ashoorioon, A., and R. B. Mann (2005), Nucl. Phys. **B716**, 261, arXiv:gr-qc/0411056 [gr-qc] .

[180] Aso, Y., Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto (KAGRA) (2013), Phys. Rev. D **88** (4), 043007, arXiv:1306.6747 [gr-qc] .

[181] Asplund, M., D. L. Lambert, P. E. Nissen, F. Primas, and V. V. Smith (2006), Astrophys. J. **644**, 229, arXiv:astro-ph/0510636 .

[182] Astier, P. (2001), Phys. Lett. B **500**, 8, arXiv:astro-ph/0008306 .

[183] Astone, P., *et al.* (1996), Phys. Lett. B **385**, 421.

[184] Astone, P., *et al.* (1997), Astropart. Phys. **7**, 231.

[185] Astone, P., *et al.* (2010), Phys. Rev. D **82**, 022003, arXiv:1002.3515 [gr-qc] .

[186] Asvesta, K., L. Kazantzidis, L. Perivolaropoulos, and C. G. Tsagas (2022), arXiv:2202.00962 [astro-ph.CO] .

[187] Atek, H., J. Richard, J.-P. Kneib, and D. Schaerer (2018), Mon. Not. Roy. Astron. Soc. **479** (4), 5184, arXiv:1803.09747 [astro-ph.GA] .

[29] Aubourg, E., *et al.* (2015), Phys. Rev. D **92** (12), 123516, arXiv:1411.1074 [astro-ph.CO] .

[189] Aurich, R., T. Buchert, M. J. France, and F. Steiner (2021), Class. Quant. Grav. **38** (22), 225005, arXiv:2106.13205 [astro-ph.CO] .

[190] Austermann, J. E., *et al.* (2012), in *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VI*, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8452, edited by W. S. Holland and J. Zmuidzinas, p. 84521E, arXiv:1210.4970 [astro-ph.IM] .

[191] Avila, F., A. Bernui, R. C. Nunes, E. de Carvalho, and C. P. Novaes (2021), Mon. Not. Roy. Astron. Soc. **509** (2), 2994, arXiv:2111.08541 [astro-ph.CO] .

[192] Avni, Y., and H. Tananbaum (1986), ApJ **305**, 83.

[193] Axelsson, M., Y. Fantaye, F. K. Hansen, A. J. Banday, H. K. Eriksen, and K. M. Gorski (2013), Astrophys. J. Lett. **773**, L3, arXiv:1303.5371 [astro-ph.CO] .

[194] Aylor, K., M. Joy, L. Knox, M. Millea, S. Raghunathan, and W. L. K. Wu (2019), Astrophys. J. **874** (1), 4, arXiv:1811.00537 [astro-ph.CO] .

[195] Bacon, D. J., A. R. Refregier, and R. S. Ellis (2000), Mon. Not. Roy. Astron. Soc. **318**, 625, arXiv:astro-ph/0003008 .

[196] Bacon, D. J., *et al.* (SKA) (2020), Publ. Astron. Soc. Austral. **37**, e007, arXiv:1811.02743 [astro-ph.CO] .

[197] Bag, S., V. Sahni, A. Shafieloo, and Y. Shtanov (2021), Astrophys. J. **923** (2), 212, arXiv:2107.03271 [astro-ph.CO] .

[198] Bahamonde, S., M. Caruana, K. F. Dialektopoulos, V. Gakis, M. Hohmann, J. Levi Said, E. N. Saridakis, and J. Sultana (2021), Phys. Rev. D **104** (8), 084082, arXiv:2105.13243 [gr-qc] .

[199] Bahamonde, S., K. F. Dialektopoulos, C. Escamilla-Rivera, G. Farrugia, V. Gakis, M. Hendry, M. Hohmann, J. L. Said, J. Mifsud, and E. Di Valentino (2021), arXiv:2106.13793 [gr-qc] .

[200] Bahamonde, S., K. F. Dialektopoulos, M. Hohmann, and J. Levi Said (2020), Class. Quant. Grav. **38** (2), 025006, arXiv:2003.11554 [gr-qc] .

[201] Bahamonde, S., K. F. Dialektopoulos, M. Hohmann, J. L. Said, C. Pfeifer, and E. N. Saridakis (2022), arXiv:2203.00619 [gr-qc] .

[202] Bahamonde, S., K. F. Dialektopoulos, and J. Levi Said (2019), Phys. Rev. **D100** (6), 064018, arXiv:1904.10791 [gr-qc] .

[203] Bainbridge, M. B., *et al.* (2017), Universe **3** (2), 32, arXiv:1702.01757 [astro-ph.CO] .

[204] Baker, T., *et al.* (2021), Rev. Mod. Phys. **93** (1), 015003, arXiv:1908.03430 [astro-ph.CO] .

[205] Balkenhol, L., *et al.* (SPT-3G) (2021), Phys. Rev. D **104** (8), 083509, arXiv:2103.13618 [astro-ph.CO] .

[206] Ballardini, M., M. Braglia, F. Finelli, D. Paoletti, A. A. Starobinsky, and C. Umiltà (2020), JCAP **10**, 044, arXiv:2004.14349 [astro-ph.CO] .

[207] Ballesteros, G., A. Notari, and F. Rompineve (2020), JCAP **11**, 024, arXiv:2004.05049 [astro-ph.CO] .

[208] Bamba, K., S. Capozziello, S. Nojiri, and S. D. Odintsov (2012), Astrophys. Space Sci. **342**, 155, arXiv:1205.3421 [gr-qc] .

[209] Banerjee, A., H. Cai, L. Heisenberg, E. O. Colgáin, M. M. Sheikh-Jabbari, and T. Yang (2021), Phys. Rev. D **103** (8), L081305, arXiv:2006.00244 [astro-ph.CO] .

[210] Banerjee, A., E. O. Colgáin, M. Sasaki, M. M. Sheikh-Jabbari, and T. Yang (2021), Phys. Lett. B **818**, 136366, arXiv:2009.04109 [astro-ph.CO] .

[211] Banerjee, N., S. Das, and K. Ganguly (2010), Pramana **74**, L481, arXiv:0801.1204 [gr-qc] .

[212] Banerjee, R., and K. Jedamzik (2004), Phys. Rev. D **70**, 123003, arXiv:astro-ph/0410032 .

[213] Banerjee, S., D. Benisty, and E. I. Guendelman (2021), Bulg. J. Phys. **48** (2), 117, arXiv:1910.03933 [gr-qc] .

[214] Banihashemi, A., N. Khosravi, and A. Shafieloo (2021), JCAP **06**, 003, arXiv:2012.01407 [astro-ph.CO] .

[215] Banihashemi, A., N. Khosravi, and A. H. Shirazi (2019), Phys. Rev. **D99** (8), 083509, arXiv:1810.11007 [astro-ph.CO] .

[216] Banihashemi, A., N. Khosravi, and A. H. Shirazi (2020), Phys. Rev. D **101** (12), 123521, arXiv:1808.02472 [astro-ph.CO] .

[217] Banik, N., and P. Sikivie (2013), Phys. Rev. D **88**, 123517, arXiv:1307.3547 [astro-ph.GA] .

[218] Barboza, E. M., Jr., and J. S. Alcaniz (2008), Phys. Lett. B **666**, 415, arXiv:0805.1713 [astro-ph] .

[219] Bardeen, J. M. (1980), Phys. Rev. D **22**, 1882.

[220] Barkana, R., N. J. Outmezguine, D. Redigolo, and T. Volansky (2018), Phys. Rev. D **98** (10), 103005, arXiv:1803.03091 [hep-ph] .

[221] Baron, E. A., P. E. Nugent, D. Branch, and P. H. Hauschildt (2004), Astrophys. J. Lett. **616**, L91, arXiv:astro-ph/0410153 .

[222] Barros, B. J., L. Amendola, T. Barreiro, and N. J. Nunes (2019), JCAP **01**, 007, arXiv:1802.09216 [astro-ph.CO] .

[223] Barrow, J. D., and A. A. H. Graham (2013), Phys. Rev. D **88**, 103513, arXiv:1307.6816 [gr-qc] .

[224] Barrow, J. D., and S. Z. W. Lip (2012), Phys. Rev. D **85**, 023514, arXiv:1110.3120 [gr-qc] .

[225] Barrow, J. D., H. B. Sandvik, and J. Magueijo (2002), Phys. Rev. D **65**, 063504, arXiv:astro-ph/0109414 .

[35] Bartelmann, M., and P. Schneider (2001), Phys. Rept. **340**, 291, arXiv:astro-ph/9912508 [astro-ph] .

[227] Bartolo, N., V. De Luca, G. Franciolini, A. Lewis, M. Peloso, and A. Riotto (2019), Phys. Rev. Lett. **122** (21), 211301, arXiv:1810.12218 [astro-ph.CO] .

[228] Bartolo, N., V. De Luca, G. Franciolini, M. Peloso, D. Racco, and A. Riotto (2019), Phys. Rev. D **99** (10), 103521, arXiv:1810.12224 [astro-ph.CO] .

[229] Basilakos, S., N. E. Mavromatos, and J. Solà (2016), Universe **2** (3), 14, arXiv:1505.04434 [gr-qc] .

[230] Basilakos, S., N. E. Mavromatos, and J. Solà Peracaula (2019), PoS **CORFU2018**, 044, arXiv:1905.05685 [hep-ph] .

[231] Basilakos, S., N. E. Mavromatos, and J. Solà Peracaula (2020), Phys. Rev. D **101** (4), 045001, arXiv:1907.04890 [hep-ph] .

[232] Basilakos, S., N. E. Mavromatos, and J. Solà Peracaula (2020), Phys. Lett. B **803**, 135342, arXiv:2001.03465 [gr-qc] .

[233] Basilakos, S., and S. Nesseris (2016), Phys. Rev. D **94** (12), 123525, arXiv:1610.00160 [astro-ph.CO] .

[234] Basilakos, S., and S. Nesseris (2017), Phys. Rev. D **96** (6), 063517, arXiv:1705.08797 [astro-ph.CO] .

[235] Basilakos, S., and L. Perivolaropoulos (2008), Mon. Not. Roy. Astron. Soc. **391**, 411, arXiv:0805.0875 [astro-ph] .

[236] Basilakos, S., D. Polarski, and J. Sola (2012), Phys. Rev. D **86**, 043010, arXiv:1204.4806 [gr-qc] .

[237] Battaglia, G., A. Helmi, E. Tolstoy, M. Irwin, V. Hill, and P. Jablonka (2008), Astrophys. J. Lett. **681**, L13, arXiv:0802.4220 [astro-ph] .

[238] Battye, R. A., and A. Moss (2014), Phys. Rev. Lett. **112** (5), 051303, arXiv:1308.5870 [astro-ph.CO] .

[239] Baumann, D., A. Nicolis, L. Senatore, and M. Zaldarriaga (2012), JCAP **07**, 051, arXiv:1004.2488 [astro-ph.CO] .

[240] Baume, F., and E. Palti (2016), JHEP **08**, 043, arXiv:1602.06517 [hep-th] .

[37] Bautista, J. E., *et al.* (2017), Astron. Astrophys. **603**, A12, arXiv:1702.00176 [astro-ph.CO] .

[242] Bautista, J. E., *et al.* (2020), Mon. Not. Roy. Astron. Soc. **500** (1), 736, arXiv:2007.08993 [astro-ph.CO] .

[243] Baxter, E. J., and B. D. Sherwin (2021), Mon. Not. Roy. Astron. Soc. **501** (2), 1823, arXiv:2007.04007 [astro-ph.CO] .

[244] Bayer, A. E., and U. Seljak (2020), JCAP **10**, 009, arXiv:2007.13821 [physics.data-an] .

[245] Beaton, R. L., *et al.* (2016), Astrophys. J. **832** (2), 210, arXiv:1604.01788 [astro-ph.CO] .

[246] Beck, R., I. Csabai, G. Rácz, and I. Szapudi (2018), Mon. Not. Roy. Astron. Soc. **479** (3), 3582, arXiv:1801.08566 [astro-ph.CO] .

[247] Becker, N., D. C. Hooper, F. Kahlhoefer, J. Lesgourgues, and N. Schöneberg (2021), JCAP **02**, 019, arXiv:2010.04074 [astro-ph.CO] .

[248] Bekenstein, J. D. (1982), Phys. Rev. D **25**, 1527.

[249] Belczynski, K., *et al.* (2016), Astron. Astrophys. **594**, A97, arXiv:1607.03116 [astro-ph.HE] .

[250] Belgacem, E., Y. Dirian, A. Finke, S. Foffa, and M. Maggiore (2020), JCAP **04**, 010, arXiv:2001.07619 [astro-ph.CO] .

[39] Belgacem, E., Y. Dirian, S. Foffa, and M. Maggiore (2018), JCAP **1803** (03), 002, arXiv:1712.07066 [hep-th] .

[252] Belgacem, E., S. Foffa, M. Maggiore, and T. Yang (2020), Phys. Rev. D **101** (6), 063505, arXiv:1911.11497 [astro-ph.CO] .

[253] Bellm, E. C., *et al.* (2018), Publications of the Astronomical Society of the Pacific **131** (995), 018002.

[254] Beltran Jimenez, J., D. Bettoni, and P. Brax (2020), Int. J. Mod. Phys. D **29** (14), 2043010, arXiv:2007.11029 [gr-qc] .

[255] Beltran Jimenez, J., D. Bettoni, and P. Brax (2021), Class. Quant. Grav. **38** (13), 134001, arXiv:2103.03627 [astro-ph.CO] .

[256] Beltrán Jiménez, J., D. Bettoni, D. Figueruelo, F. A. Teppa Pannia, and S. Tsujikawa (2021), Phys. Rev. D **104** (10), 103503, arXiv:2106.11222 [astro-ph.CO] .

[257] Benaoum, H. B., W. Yang, S. Pan, and E. Di Valentino (2022), Int. J. Mod. Phys. D **31** (03), 2250015, arXiv:2008.09098 [gr-qc] .

[258] Benedict, G. F., B. E. McArthur, M. W. Feast, T. G. Barnes, T. E. Harrison, R. J. Patterson, J. W. Menzies, J. L. Bean, and W. L. Freedman (2007), Astron. J. **133**, 1810, [Erratum: Astron.J. 133, 2980 (2007)], arXiv:astro-ph/0612465 .

[259] Benetti, M., H. Borges, C. Pigozzo, S. Carneiro, and J. Alcaniz (2021), JCAP **08**, 014, arXiv:2102.10123 [astro-ph.CO] .

[260] Benetti, M., W. Miranda, H. A. Borges, C. Pigozzo, S. Carneiro, and J. S. Alcaniz (2019), JCAP **12**, 023, arXiv:1908.07213 [astro-ph.CO] .

[261] Benevento, G., W. Hu, and M. Raveri (2020), Phys. Rev. D **101** (10), 103517, arXiv:2002.11707 [astro-ph.CO] .

[262] Bengaly, C. (2022), Phys. Dark Univ. **35**, 100966, arXiv:2111.06869 [gr-qc] .

[263] Bengaly, C., R. Maartens, and M. Santos (2018), JCAP **04**, 031, arXiv:1710.08804 [astro-ph.CO] .

[264] Bengaly, C. A. P. (2020), Mon. Not. Roy. Astron. Soc. **499** (1), L6, arXiv:1912.05528 [astro-ph.CO] .

[265] Bengaly, C. A. P., C. Clarkson, and R. Maartens (2020), JCAP **05**, 053, arXiv:1908.04619 [astro-ph.CO] .

[266] Bengaly, C. A. P., J. E. Gonzalez, and J. S. Alcaniz (2020), Eur. Phys. J. C **80** (10), 936, arXiv:2007.13789 [astro-ph.CO] .

[267] Benichou, R., and J. Estes (2012), Phys. Lett. B **712**, 456, arXiv:1112.0565 [hep-th] .

[268] Benisty, D. (2021), Phys. Dark Univ. **31**, 100766, arXiv:2005.03751 [astro-ph.CO] .

[269] Benisty, D., and A.-C. Davis (2022), Phys. Rev. D **105** (2), 024052, arXiv:2108.06286 [astro-ph.CO] .

[270] Benisty, D., E. I. Guendelman, A. van de Venn, D. Vasak, J. Struckmeier, and H. Stoecker (2022), Eur. Phys. J. C **82** (3), 264, arXiv:2109.01052 [astro-ph.CO] .

[271] Benisty, D., J. Mifsud, J. L. Said, and D. Staicova (2022), arXiv:2202.04677 [astro-ph.CO] .

[272] Benisty, D., and D. Staicova (2021), arXiv:2107.14129 [astro-ph.CO] .

[43] Bennett, C., *et al.* (WMAP) (2003), Astrophys. J. Suppl. **148**, 1, arXiv:astro-ph/0302207 .

[274] Bennett, C., *et al.* (WMAP) (2003), Astrophys. J. **583**, 1, arXiv:astro-ph/0301158 .

[275] Bennett, C. L., *et al.* (2011), Astrophys. J. Suppl. **192**, 17, arXiv:1001.4758 [astro-ph.CO] .

[42] Bennett, C. L., *et al.* (WMAP) (2013), Astrophys. J. Suppl. **208**, 20, arXiv:1212.5225 [astro-ph.CO] .

[277] Bennett, J. J., G. Buldgen, P. F. De Salas, M. Drewes, S. Gariazzo, S. Pastor, and Y. Y. Y. Wong (2021), JCAP **04**, 073, arXiv:2012.02726 [hep-ph] .

[278] Benson, B. A., *et al.* (SPT-3G) (2014), Proc. SPIE Int. Soc. Opt. Eng. **9153**, 91531P, arXiv:1407.2973 [astro-ph.IM] .

[279] Berbig, M., S. Jana, and A. Trautner (2020), Phys. Rev. D **102** (11), 115008, arXiv:2004.13039 [hep-ph] .

[280] Berechya, D., and U. Leonhardt (2021), Mon. Not. Roy. Astron. Soc. **507** (3), 3473, arXiv:2008.04789 [gr-qc] .

[281] Berengut, J. C., V. V. Flambaum, and V. F. Dmitriev (2010), Phys. Lett. B **683**, 114, arXiv:0907.2288 [nucl-th] .

[282] Berengut, J. C., V. V. Flambaum, A. Ong, J. K. Webb, J. D. Barrow, M. A. Barstow, S. P. Preval, and J. B. Holberg (2013), Phys. Rev. Lett. **111** (1), 010801, arXiv:1305.1337 [astro-ph.CO] .

[283] Berezhiani, Z., A. Dolgov, and I. Tkachev (2015), Phys. Rev. D **92** (6), 061303, arXiv:1505.03644 [astro-ph.CO] .

[284] Berger, M. S., and H. Shojaei (2006), Phys. Rev. D **73**, 083528, arXiv:gr-qc/0601086 .

[285] Berghaus, K. V., and T. Karwal (2020), Phys. Rev. D **101** (8), 083537, arXiv:1911.06281 [astro-ph.CO] .

[286] Bernal, J. L., L. Verde, R. Jimenez, M. Kamionkowski, D. Valcin, and B. D. Wandelt (2021), Phys. Rev. D **103** (10), 103533, arXiv:2102.05066 [astro-ph.CO] .

[44] Bernal, J. L., L. Verde, and A. G. Riess (2016), JCAP **1610** (10), 019, arXiv:1607.05617 [astro-ph.CO] .

[288] Bernardeau, F., S. Colombi, E. Gaztanaga, and R. Scoccimarro (2002), Phys. Rept. **367**, 1, arXiv:astro-ph/0112551 .

[289] Bernardo, R. C., J. L. Said, M. Caruana, and S. Appleby (2021), JCAP **10**, 078, arXiv:2107.08762 [gr-qc] .

[290] Bernui, A., C. P. Novaes, T. S. Pereira, and G. D. Starkman (2018), arXiv:1809.05924 [astro-ph.CO] .

[291] Bernui, A., A. F. Oliveira, and T. S. Pereira (2014), JCAP **10**, 041, arXiv:1404.2936 [astro-ph.CO] .

[292] Bertone, G., D. Hooper, and J. Silk (2005), Phys. Rept. **405**, 279, arXiv:hep-ph/0404175 [hep-ph] .

[293] Betoule, M., *et al.* (SDSS) (2014), Astron. Astrophys. **568**, A22, arXiv:1401.4064 [astro-ph.CO] .

[47] Beutler, F., C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson (2011), Mon. Not. Roy. Astron. Soc. **416**, 3017, arXiv:1106.3366 [astro-ph.CO] .

[48] Beutler, F., C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, G. B. Poole, L. Campbell, Q. Parker, W. Saunders, and F. Watson (2012), Mon. Not. Roy. Astron. Soc. **423**, 3430, arXiv:1204.4725 [astro-ph.CO] .

[296] Bianchini, F., *et al.* (SPT) (2020), Phys. Rev. D **102** (8), 083504, arXiv:2006.08061 [astro-ph.CO] .

[297] Binder, T., M. Gustafsson, A. Kamada, S. M. R. Sandner, and M. Wiesner (2018), Phys. Rev. **D97** (12), 123004, arXiv:1712.01246 [astro-ph.CO] .

[298] Birkinshaw, M. (1999), Phys. Rept. **310**, 97, arXiv:astro-ph/9808050 .

[299] Birrer, S., *et al.* (2020), Astron. Astrophys. **643**, A165, arXiv:2007.02941 [astro-ph.CO] .

[300] Biscardi, I., G. Raimondo, M. Cantiello, and E. Brocato (2008), Astrophys. J. **678**, 168, arXiv:0802.2509 [astro-ph] .

[301] Bisogni, S., E. Lusso, F. Civano, E. Nardini, G. Risaliti, M. Elvis, and G. Fabbiano (2021), Astron. Astrophys. **655**, A109, arXiv:2109.03252 [astro-ph.GA] .

[302] Biswas, R., K. Heitmann, S. Habib, A. Upadhye, A. Pope, and N. Frontiere (2019), arXiv:1901.10690 [astro-ph.CO] .

[303] Biswas, T., A. Conroy, A. S. Koshelev, and A. Mazumdar (2014), Class. Quant. Grav. **31**, 015022, [Erratum: Class.Quant.Grav. 31, 159501 (2014)], arXiv:1308.2319 [hep-th] .

[304] Bjaelde, O. E., S. Das, and A. Moss (2012), JCAP **10**, 017, arXiv:1205.0553 [astro-ph.CO] .

[305] Blackadder, G., and S. M. Koushiappas (2014), Phys. Rev. **D90** (10), 103527, arXiv:1410.0683 [astro-ph.CO] .

[306] Blair, D. G., E. N. Ivanov, M. E. Tobar, P. J. Turner, F. van Kann, and I. S. Heng (1995), Phys. Rev. Lett. **74**, 1908.

[307] Blake, C., and J. Wall (2002), Nature **416**, 150, arXiv:astro-ph/0203385 .

[308] Blake, C., *et al.* (2011), Mon. Not. Roy. Astron. Soc. **418**, 1707, arXiv:1108.2635 [astro-ph.CO] .

[309] Blake, C., *et al.* (2011), Mon. Not. Roy. Astron. Soc. **415**, 2876, arXiv:1104.2948 [astro-ph.CO] .

[53] Blake, C., *et al.* (2012), Mon. Not. Roy. Astron. Soc. **425**, 405, arXiv:1204.3674 [astro-ph.CO] .

[311] Blake, C., *et al.* (2016), Mon. Not. Roy. Astron. Soc. **462** (4), 4240, arXiv:1608.02668 [astro-ph.CO] .

[312] Blakeslee, J. (2012), Astrophys. Space Sci. **341**, 179, arXiv:1202.0581 [astro-ph.CO] .

[313] Blakeslee, J. P., E. A. Ajhar, and J. L. Tonry (1999), Astrophys. Space Sci. Libr. **237**, 181, arXiv:astro-ph/9807124 .

[314] Blakeslee, J. P., J. B. Jensen, C.-P. Ma, P. A. Milne, and J. E. Greene (2021), Astrophys. J. **911** (1), 65, arXiv:2101.02221 [astro-ph.CO] .

[315] Blakeslee, J. P., A. Jordan, S. Mei, P. Cote, L. Ferrarese, L. Infante, E. W. Peng, J. L. Tonry, and M. J. West (2009), Astrophys. J. **694**, 556, arXiv:0901.1138 [astro-ph.CO] .

[316] Blanchard, A., *et al.* (2021), Exper. Astron. **51** (3), 1623, arXiv:2102.03931 [astro-ph.CO] .

[317] Blanton, M. R., *et al.* (SDSS) (2017), Astron. J. **154** (1), 28, arXiv:1703.00052 [astro-ph.GA] .

[318] Bleem, L., *et al.* (SPT, DES) (2020), Astrophys. J. Suppl. **247** (1), 25, arXiv:1910.04121 [astro-ph.CO] .

[319] Bleem, L. E., *et al.* (SPT) (2015), Astrophys. J. Suppl. **216** (2), 27, arXiv:1409.0850 [astro-ph.CO] .

[320] Blennow, M., S. Clementz, and J. Herrero-Garcia (2017), JCAP **03**, 048, arXiv:1612.06681 [hep-ph] .

[321] Blennow, M., E. Fernandez-Martinez, O. Mena, J. Redondo, and P. Serra (2012), JCAP **07**, 022, arXiv:1203.5803 [hep-ph] .

[322] Blinov, N., C. Keith, and D. Hooper (2020), JCAP **06**, 005, arXiv:2004.06114 [astro-ph.CO] .

[323] Blinov, N., K. J. Kelly, G. Z. Krnjaic, and S. D. McDermott (2019), Phys. Rev. Lett. **123** (19), 191102, arXiv:1905.02727 [astro-ph.CO] .

[324] Blinov, N., and G. Marques-Tavares (2020), JCAP **09**, 029, arXiv:2003.08387 [astro-ph.CO] .

[325] de Blok, W. J. G. (2010), Adv. Astron. **2010**, 789293, arXiv:0910.3538 [astro-ph.CO] .

[57] Blomqvist, M., *et al.* (2019), Astron. Astrophys. **629**, A86, arXiv:1904.03430 [astro-ph.CO] .

[327] Blumenhagen, R., I. Valenzuela, and F. Wolf (2017), JHEP **07**, 145, arXiv:1703.05776 [hep-th] .

[328] Bocquet, S., *et al.* (SPT) (2019), Astrophys. J. **878** (1), 55, arXiv:1812.01679 [astro-ph.CO] .

[329] Bode, P., J. P. Ostriker, and N. Turok (2001), Astrophys. J. **556**, 93, arXiv:astro-ph/0010389 .

[330] Boisseau, B., G. Esposito-Farese, D. Polarski, and A. A. Starobinsky (2000), Phys. Rev. Lett. **85**, 2236, arXiv:gr-qc/0001066 .

[331] Bolejko, K. (2018), Phys. Rev. D **97** (10), 103529, arXiv:1712.02967 [astro-ph.CO] .

[332] Bolotin, Y. L., A. Kostenko, O. A. Lemets, and D. A. Yerokhin (2014), Int. J. Mod. Phys. D **24** (03), 1530007, arXiv:1310.0085 [astro-ph.CO] .

[333] Bolton, A. S., S. Burles, L. V. E. Koopmans, T. Treu, and L. A. Moustakas (2006), Astrophys. J. **638**, 703, arXiv:astro-ph/0511453 .

[334] Bond, H. E., E. P. Nelan, D. A. VandenBerg, G. H. Schaefer, and D. Harmer (2013), Astrophys. J. Lett. **765**, L12, arXiv:1302.3180 [astro-ph.SR] .

[335] Bond, J. R., G. Efstathiou, and J. Silk (1980), Phys. Rev. Lett. **45**, 1980.

[336] Bonetti, L., L. R. dos Santos Filho, J. A. Helayël-Neto, and A. D. A. M. Spallicci (2017), Phys. Lett. B **764**, 203, arXiv:1607.08786 [hep-ph] .

[337] Bonetti, L., L. R. dos Santos Filho, J. A. Helayël-Neto, and A. D. A. M. Spallicci (2018), Eur. Phys. J. C **78** (10), 811, arXiv:1709.04995 [hep-th] .

[338] Bonilla, A., S. Kumar, and R. C. Nunes (2021), Eur. Phys. J. C **81** (2), 127, arXiv:2011.07140 [astro-ph.CO] .

[339] Bonilla, A., S. Kumar, R. C. Nunes, and S. Pan (2021), arXiv:2102.06149 [astro-ph.CO] .

[340] Bonvin, V., *et al.* (2017), Mon. Not. Roy. Astron. Soc. **465** (4), 4914, arXiv:1607.01790 [astro-ph.CO] .

[341] van den Bosch, F. C., A. Burkert, and R. A. Swaters (2001), Mon. Not. Roy. Astron. Soc. **326**, 1205, arXiv:astro-ph/0105082 .

[342] Bose, B., and L. Lombriser (2021), Phys. Rev. D **103** (8), L081304, arXiv:2006.16149 [astro-ph.CO] .

[343] Bosma, A. (1981), Astron. J. **86**, 1825.

[344] Bottke, W. F., D. Vokrouhlický, and D. Nesvorný (2007), Nature **449**, 48.

[114] du Mas des Bourboux, H., *et al.* (2017), Astron. Astrophys. **608**, A130, arXiv:1708.02225 [astro-ph.CO] .

[346] du Mas des Bourboux, H., *et al.* (2020), Astrophys. J. **901** (2), 153, arXiv:2007.08995 [astro-ph.CO] .

[347] Bouwens, R. J., *et al.* (2015), Astrophys. J. **803** (1), 34, arXiv:1403.4295 [astro-ph.CO] .

[348] Bowen, R., S. H. Hansen, A. Melchiorri, J. Silk, and R. Trotta (2002), Mon. Not. Roy. Astron. Soc. **334**, 760, arXiv:astro-ph/0110636 .

[349] Bowman, J. D., A. E. E. Rogers, R. A. Monsalve, T. J. Mozdzen, and N. Mahesh (2018), Nature **555** (7694), 67, arXiv:1810.05912 [astro-ph.CO] .

[350] Boyarsky, A., D. Iakubovskyi, O. Ruchayskiy, A. Rudakovskyi, and W. Valkenburg (2019), Phys. Rev. D **100** (12), 123005, arXiv:1904.03097 [astro-ph.CO] .

[351] Boyarsky, A., M. Ovchynnikov, N. Sabti, and V. Syvolap (2021), Phys. Rev. D **104** (3), 035006, arXiv:2103.09831 [hep-ph] .

[352] Boyarsky, A., O. Ruchayskiy, and M. Shaposhnikov (2009), Ann. Rev. Nucl. Part. Sci. **59**, 191, arXiv:0901.0011 [hep-ph] .

[353] Boyd, R. N., C. R. Brune, G. M. Fuller, and C. J. Smith (2010), Phys. Rev. D **82**, 105005, arXiv:1008.0848 [astro-ph.CO] .

[354] Boylan-Kolchin, M., J. S. Bullock, and M. Kaplinghat (2011), Mon. Not. Roy. Astron. Soc. **415**, L40, arXiv:1103.0007 [astro-ph.CO] .

[355] Boylan-Kolchin, M., J. S. Bullock, and M. Kaplinghat (2012), Mon. Not. Roy. Astron. Soc. **422**, 1203, arXiv:1111.2048 [astro-ph.CO] .

[356] Bozek, B., *et al.* (2019), Mon. Not. Roy. Astron. Soc. **483** (3), 4086, arXiv:1803.05424 [astro-ph.GA] .

[357] Braglia, M., M. Ballardini, W. T. Emond, F. Finelli, A. E. Gumrukcuoglu, K. Koyama, and D. Paoletti (2020), Phys. Rev. D **102** (2), 023529, arXiv:2004.11161 [astro-ph.CO] .

[358] Braglia, M., M. Ballardini, F. Finelli, and K. Koyama (2021), Phys. Rev. D **103** (4), 043528, arXiv:2011.12934 [astro-ph.CO] .

[359] Braglia, M., W. T. Emond, F. Finelli, A. E. Gumrukcuoglu, and K. Koyama (2020), Phys. Rev. D **102** (8), 083513, arXiv:2005.14053 [astro-ph.CO] .

[360] Brainerd, T. G., R. D. Blandford, and I. Smail (1996), Astrophys. J. **466**, 623, arXiv:astro-ph/9503073 .

[361] Brax, P., C. van de Bruck, S. Clesse, A.-C. Davis, and G. Sculthorpe (2014), Phys. Rev. D **89** (12), 123507, arXiv:1312.3361 [astro-ph.CO] .

[362] Brax, P., C. van de Bruck, A. C. Davis, J. Khoury, and A. Weltman (2004), AIP Conf. Proc. **736** (1), 105, arXiv:astro-ph/0410103 .

[363] Brax, P., C. van de Bruck, A.-C. Davis, J. Khoury, and A. Weltman (2004), Phys. Rev. D **70**, 123518, arXiv:astro-ph/0408415 .

[364] Brax, P., C. van de Bruck, A.-C. Davis, D. F. Mota, and D. J. Shaw (2007), Phys. Rev. D **76**, 124034, arXiv:0709.2075 [hep-ph] .

[365] Brax, P., C. van de Bruck, A.-C. Davis, and D. J. Shaw (2008), Phys. Rev. D **78**, 104021, arXiv:0806.3415 [astro-ph] .

[366] Brax, P., C. van de Bruck, D. F. Mota, N. J. Nunes, and H. A. Winther (2010), Phys. Rev. D **82**, 083503, arXiv:1006.2796 [astro-ph.CO] .

[367] Brax, P., S. Casas, H. Desmond, and B. Elder (2021), Universe **8** (1), 11, arXiv:2201.10817 [gr-qc] .

[368] Brax, P., A.-C. Davis, B. Li, and H. A. Winther (2012), Phys. Rev. **D86**, 044015, arXiv:1203.4812 [astro-ph.CO] .

[369] Brdar, V., M. Lindner, S. Vogl, and X.-J. Xu (2020), Phys. Rev. D **101** (11), 115001, arXiv:2003.05339 [hep-ph] .

[370] Breuval, L., *et al.* (2020), Astron. Astrophys. **643**, A115, arXiv:2006.08763 [astro-ph.SR] .

[371] Briffa, R., S. Capozziello, J. Levi Said, J. Mifsud, and E. N. Saridakis (2020), Class. Quant. Grav. **38** (5), 055007, arXiv:2009.14582 [gr-qc] .

[372] Brinckmann, T., J. H. Chang, and M. LoVerde (2020), arXiv:2012.11830 [astro-ph.CO] .

[373] Bringmann, T., F. Kahlhoefer, K. Schmidt-Hoberg, and P. Walia (2018), Phys. Rev. D **98** (2), 023543, arXiv:1803.03644 [astro-ph.CO] .

[374] Broggini, C., L. Canton, G. Fiorentini, and F. L. Villante (2012), JCAP **06**, 030, arXiv:1202.5232 [astro-ph.CO] .

[375] Bromm, V. (2013), Rept. Prog. Phys. **76**, 112901, arXiv:1305.5178 [astro-ph.CO] .

[376] Bromm, V., P. S. Coppi, and R. B. Larson (2002), Astrophys. J. **564**, 23, arXiv:astro-ph/0102503 .

[377] Bromm, V., and R. B. Larson (2004), Ann. Rev. Astron. Astrophys. **42**, 79, arXiv:astro-ph/0311019 .

[378] Bronnikov, K. A., S. A. Kononogov, and V. N. Melnikov (2006), Gen. Rel. Grav. **38**, 1215, arXiv:gr-qc/0601114 .

[379] Brout, D., *et al.* (DES) (2019), Astrophys. J. **874** (1), 106, arXiv:1811.02378 [astro-ph.IM] .

[380] Brout, D., *et al.* (2022), arXiv:2202.04077 [astro-ph.CO] .

[381] Brown, A., *et al.* (Gaia) (2018), Astron. Astrophys. **616**, A1, arXiv:1804.09365 [astro-ph.GA] .

[382] Brown, A. G. A., *et al.* (Gaia) (2016), Astron. Astrophys. **595** (Gaia Data Release 1), A2, arXiv:1609.04172 [astro-ph.IM] .

[383] Brown, M. L., P. G. Castro, and A. N. Taylor (2005), Mon. Not. Roy. Astron. Soc. **360**, 1262, arXiv:astro-ph/0410394 .

[384] Brownsberger, S., D. Brout, D. Scolnic, C. W. Stubbs, and A. G. Riess (2021), arXiv:2110.03486 [astro-ph.CO] .

[385] Brune, T., and H. Päs (2019), Phys. Rev. D **99** (9), 096005, arXiv:1808.08158 [hep-ph] .

[386] Buchert, T., A. A. Coley, H. Kleinert, B. F. Roukema, and D. L. Wiltshire (2016), Int. J. Mod. Phys. D **25** (03), 1630007, arXiv:1512.03313 [astro-ph.CO] .

[387] Buen-Abad, M. A., R. Emami, and M. Schmaltz (2018), Phys. Rev. D **98** (8), 083517, arXiv:1803.08062 [hep-ph] .

[388] Buen-Abad, M. A., G. Marques-Tavares, and M. Schmaltz (2015), Phys. Rev. D **92** (2), 023531, arXiv:1505.03542 [hep-ph] .

[65] Buen-Abad, M. A., M. Schmaltz, J. Lesgourgues, and T. Brinckmann (2018), JCAP **01**, 008, arXiv:1708.09406 [astro-ph.CO] .

[390] Bueno Sanchez, J. C., and L. Perivolaropoulos (2011), Phys. Rev. D **84**, 123516, arXiv:1110.2587 [astro-ph.CO] .

[391] Bull, P., *et al.* (2016), Phys. Dark Univ. **12**, 56, arXiv:1512.05356 [astro-ph.CO] .

[392] Bullock, J. S. (2010), arXiv:1009.4505 [astro-ph.CO] .

[393] Bullock, J. S., and M. Boylan-Kolchin (2017), Ann. Rev. Astron. Astrophys. **55**, 343, arXiv:1707.04256 [astro-ph.CO] .

[394] Bullock, J. S., A. Dekel, T. S. Kolatt, A. V. Kravtsov, A. A. Klypin, C. Porciani, and J. R. Primack (2001), Astrophys. J. **555**, 240, arXiv:astro-ph/0011001 .

[395] Bunn, E. F., M. Zaldarriaga, M. Tegmark, and A. de Oliveira-Costa (2003), Phys. Rev. D **67**, 023501, arXiv:astro-ph/0207338 .

[396] Burgess, C. (2015), in *100e Ecole d'Ete de Physique: Post-Planck Cosmology*, pp. 149–197, arXiv:1309.4133 [hep-th] .

[397] Burns, C. R., *et al.* (CSP) (2018), Astrophys. J. **869** (1), 56, arXiv:1809.06381 [astro-ph.CO] .

[398] Burrage, C., and J. Sakstein (2016), JCAP **11**, 045, arXiv:1609.01192 [astro-ph.CO] .

[399] Burrage, C., and J. Sakstein (2018), Living Rev. Rel. **21** (1), 1, arXiv:1709.09071 [astro-ph.CO] .

[400] Busti, V. C., C. Clarkson, and M. Seikel (2014), Mon. Not. Roy. Astron. Soc. **441**, 11, arXiv:1402.5429 [astro-ph.CO] .

[401] Byrnes, C., G. Domènech, M. Sasaki, and T. Takahashi (2016), JCAP **12**, 020, arXiv:1610.02650 [astro-ph.CO] .

[402] Byrnes, C. T., D. Regan, D. Seery, and E. R. M. Tarrant (2016), Phys. Rev. D **93** (12), 123003, arXiv:1601.01970 [astro-ph.CO] .

[403] Byrnes, C. T., D. Regan, D. Seery, and E. R. M. Tarrant (2016), JCAP **06**, 025, arXiv:1511.03129 [astro-ph.CO] .

[404] Cabass, G., M. Gerbino, E. Giusarma, A. Melchiorri, L. Pagano, and L. Salvati (2015), Phys. Rev. D **92** (6), 063534, arXiv:1507.07586 [astro-ph.CO] .

[405] Cai, R.-G., Z. Cao, Z.-K. Guo, S.-J. Wang, and T. Yang (2017), Natl. Sci. Rev. **4** (5), 687, arXiv:1703.00187 [gr-qc] .

[406] Cai, R.-G., Z.-K. Guo, L. Li, S.-J. Wang, and W.-W. Yu (2021), Phys. Rev. D **103** (12), 121302, arXiv:2102.02020 [astro-ph.CO] .

[407] Cai, R.-G., Z.-K. Guo, S.-J. Wang, W.-W. Yu, and Y. Zhou (2022), arXiv:2202.12214 [astro-ph.CO] .

[408] Cai, R.-G., Z.-K. Guo, S.-J. Wang, W.-W. Yu, and Y. Zhou (2022), Phys. Rev. D **105** (2), L021301, arXiv:2107.13286 [astro-ph.CO] .

[409] Cai, R.-g., S. Pi, and M. Sasaki (2019), Phys. Rev. Lett. **122** (20), 201101, arXiv:1810.11000 [astro-ph.CO] .

[410] Cai, R.-G., N. Tamanini, and T. Yang (2017), JCAP **05**, 031, arXiv:1703.07323 [astro-ph.CO] .

[411] Cai, R.-G., and A. Wang (2005), JCAP **03**, 002, arXiv:hep-th/0411025 .

[412] Cai, Y.-C., M. Neyrinck, Q. Mao, J. A. Peacock, I. Szapudi, and A. A. Berlind (2017), Mon. Not. Roy. Astron. Soc. **466** (3), 3364, arXiv:1609.00301 [astro-ph.CO] .

[413] Cai, Y.-F., S. Capozziello, M. De Laurentis, and E. N. Saridakis (2016), Rept. Prog. Phys. **79** (10), 106901, arXiv:1511.07586 [gr-qc] .

[414] Cai, Y.-F., M. Khurshudyan, and E. N. Saridakis (2020), Astrophys. J. **888**, 62, arXiv:1907.10813 [astro-ph.CO] .

[66] Calabrese, E., A. Slosar, A. Melchiorri, G. F. Smoot, and O. Zahn (2008), Phys. Rev. **D77**, 123531, arXiv:0803.2309 [astro-ph] .

[416] Calderón, R., R. Gannouji, B. L'Huillier, and D. Polarski (2021), Phys. Rev. D **103** (2), 023526, arXiv:2008.10237 [astro-ph.CO] .

[417] Caldwell, R., R. Dave, and P. J. Steinhardt (1998), Phys. Rev. Lett. **80**, 1582, arXiv:astro-ph/9708069 .

[418] Caldwell, R. R., W. Komp, L. Parker, and D. A. T. Vanzella (2006), Phys. Rev. D **73**, 023513, arXiv:astro-ph/0507622 .

[419] Camarena, D., and V. Marra (2020), Mon. Not. Roy. Astron. Soc. **495** (3), 2630, arXiv:1910.14125 [astro-ph.CO] .

[420] Camarena, D., and V. Marra (2020), Phys. Rev. Res. **2** (1), 013028, arXiv:1906.11814 [astro-ph.CO] .

[421] Camarena, D., and V. Marra (2021), 10.1093/mnras/stab1200, arXiv:2101.08641 [astro-ph.CO] .

[422] Camera, S., M. Martinelli, and D. Bertacca (2019), Phys. Dark Univ. **23**, 100247, arXiv:1704.06277 [astro-ph.CO] .

[423] Cameron, E., and T. Pettitt (2012), arXiv:1207.6223 [astro-ph.CO] .

[424] del Campo, S., R. Herrera, G. Olivares, and D. Pavon (2006), Phys. Rev. D **74**, 023501, arXiv:astro-ph/0606520 .

[425] del Campo, S., R. Herrera, and D. Pavon (2009), JCAP **01**, 020, arXiv:0812.2210 [gr-qc] .

[426] Cantiello, M., *et al.* (2018), Astrophys. J. Lett. **854** (2), L31, arXiv:1801.06080 [astro-ph.GA] .

[427] Cao, S., M. Dainotti, and B. Ratra (2022), Mon. Not. Roy. Astron. Soc. **512**, 439, arXiv:2201.05245 [astro-ph.CO] .

[428] Cao, S., N. Khadka, and B. Ratra (2022), Mon. Not. Roy. Astron. Soc. **510** (2), 2928, arXiv:2110.14840 [astro-ph.CO] .

[429] Cao, S., and B. Ratra (2022), arXiv:2203.10825 [astro-ph.CO] .

[430] Cao, S., J. Ryan, N. Khadka, and B. Ratra (2021), Mon. Not. Roy. Astron. Soc. **501** (1), 1520, arXiv:2009.12953 [astro-ph.CO] .

[431] Cao, S., J. Ryan, and B. Ratra (2020), Mon. Not. Roy. Astron. Soc. **497** (3), 3191, arXiv:2005.12617 [astro-ph.CO] .

[432] Cao, S., J. Ryan, and B. Ratra (2021), Mon. Not. Roy. Astron. Soc. **504** (1), 300, arXiv:2101.08817 [astro-ph.CO] .

[433] Cao, S., J. Ryan, and B. Ratra (2022), Mon. Not. Roy. Astron. Soc. **509**, 4745, arXiv:2109.01987 [astro-ph.CO] .

[434] Capozzi, F., E. Di Valentino, E. Lisi, A. Marrone, A. Melchiorri, and A. Palazzo (2021), Phys. Rev. D **104** (8), 083031, arXiv:2107.00532 [hep-ph] .

[435] Capozziello, S. (2002), Int. J. Mod. Phys. D **11**, 483, arXiv:gr-qc/0201033 .

[436] Capozziello, S., M. Benetti, and A. D. A. M. Spallicci (2020), Found. Phys. **50** (9), 893, arXiv:2007.00462 [gr-qc] .

[437] Caprini, C., *et al.* (2016), JCAP **04**, 001, arXiv:1512.06239 [astro-ph.CO] .

[438] Cardona, W., R. Arjona, A. Estrada, and S. Nesseris (2021), JCAP **05**, 064, arXiv:2012.05282 [astro-ph.CO] .

[439] Cardone, V. F., M. G. Dainotti, S. Capozziello, and R. Willingale (2010), Mon. Not. Roy. Astron. Soc. **408**, 1181, arXiv:1005.0122 [astro-ph.CO] .

[440] Cardoso, J., M. Le Jeune, J. Delabrouille, M. Betoule, and G. Patanchon (2008), IEEE Journal of Selected Topics in Signal Processing **2** (5), 735.

[441] Carlson, E. D., M. E. Machacek, and L. J. Hall (1992), Astrophys. J. **398**, 43.

[442] Carlstrom, J. E., G. P. Holder, and E. D. Reese (2002), Ann. Rev. Astron. Astrophys. **40**, 643, arXiv:astro-ph/0208192 .

[443] Carneiro, S., P. C. de Holanda, C. Pigozzo, and F. Sobreira (2019), Phys. Rev. **D100** (2), 023505, arXiv:1812.06064 [astro-ph.CO] .

[444] Carrasco, J. J. M., M. P. Hertzberg, and L. Senatore (2012), JHEP **09**, 082, arXiv:1206.2926 [astro-ph.CO] .

[445] Carrilho, P., K. Carrion, B. Bose, A. Pourtsidou, J. C. Hidalgo, L. Lombriser, and M. Baldi (2021), 10.1093/mnras/stac641, arXiv:2111.13598 [astro-ph.CO] .

[446] Carrilho, P., C. Moretti, B. Bose, K. Markovi**v**c, and A. Pourtsidou (2021), JCAP **10**, 004, arXiv:2106.13163 [astro-ph.CO] .

[447] Carrillo González, M., Q. Liang, J. Sakstein, and M. Trodden (2021), JCAP **04**, 063, arXiv:2011.09895 [astro-ph.CO] .

[448] Carroll, S. M. (1998), Phys. Rev. Lett. **81**, 3067, arXiv:astro-ph/9806099 .

[449] Carroll, S. M. (2001), Living Rev. Rel. **4**, 1, arXiv:astro-ph/0004075 .

[450] Carroll, S. M., G. B. Field, and R. Jackiw (1990), Phys. Rev. D **41**, 1231.

[451] Carroll, S. M., W. H. Press, and E. L. Turner (1992), Ann. Rev. Astron. Astrophys. **30**, 499.

[452] Casertano, S., *et al.* (2016), Astrophys. J. **825** (1), 11, arXiv:1512.09371 [astro-ph.SR] .

[453] Cayuso, J. I., and M. C. Johnson (2020), Phys. Rev. D **101** (12), 123508, arXiv:1904.10981 [astro-ph.CO] .

[454] Cerdonio, M., *et al.* (1997), Class. Quant. Grav. **14**, 1491.

[455] Chacko, Z., Y. Cui, S. Hong, T. Okui, and Y. Tsai (2016), JHEP **12**, 108, arXiv:1609.03569 [astro-ph.CO] .

[456] Chakraborty, N., B. D. Fields, and K. A. Olive (2011), Phys. Rev. D **83**, 063006, arXiv:1011.0722 [astro-ph.CO] .

[457] Chamberlain, K., and N. Yunes (2017), Phys. Rev. D **96** (8), 084039, arXiv:1704.08268 [gr-qc] .

[458] Chang, C., *et al.* (DES, SPT) (2022), arXiv:2203.12440 [astro-ph.CO] .

[459] Chang, C. L., *et al.* (2022), in *2022 Snowmass Summer Study*, arXiv:2203.07638 [astro-ph.CO] .

[460] Chavanis, P.-H. (2015), Phys. Rev. D **92** (10), 103004, arXiv:1412.0743 [gr-qc] .

[461] Chávez, R., M. Plionis, S. Basilakos, R. Terlevich, E. Terlevich, J. Melnick, F. Bresolin, and A. L. González-Morán (2016), Mon. Not. Roy. Astron. Soc. **462** (3), 2431, arXiv:1607.06458 [astro-ph.CO] .

[462] Chávez, R., E. Terlevich, R. Terlevich, M. Plionis, F. Bresolin, S. Basilakos, and J. Melnick (2012), Mon. Not. Roy. Astron. Soc. **425**, 56, arXiv:1203.6222 [astro-ph.CO] .

[463] Chávez, R., R. Terlevich, E. Terlevich, F. Bresolin, J. Melnick, M. Plionis, and S. Basilakos (2014), Mon. Not. Roy. Astron. Soc. **442** (4), 3565, arXiv:1405.4010 [astro-ph.GA] .

[464] Chen, G., and B. Ratra (2011), Publ. Astron. Soc. Pac. **123**, 1127, arXiv:1105.5206 [astro-ph.CO] .

[465] Chen, G. C. F., *et al.* (2019), Mon. Not. Roy. Astron. Soc. **490** (2), 1743, arXiv:1907.02533 [astro-ph.CO] .

[466] Chen, H.-Y., M. Fishbach, and D. E. Holz (2018), Nature **562** (7728), 545, arXiv:1712.06531 [astro-ph.CO] .

[467] Chen, H.-Y., C.-J. Haster, S. Vitale, W. M. Farr, and M. Isi (2020), arXiv:2009.14057 [astro-ph.CO] .

[468] Chen, S.-F., Z. Vlah, and M. White (2022), JCAP **02** (02), 008, arXiv:2110.05530 [astro-ph.CO] .

[469] Chen, Y., S. Kumar, and B. Ratra (2017), Astrophys. J. **835** (1), 86, arXiv:1606.07316 [astro-ph.CO] .

[470] Cheoun, M.-K., T. Kajino, M. Kusakabe, and G. J. Mathews (2011), Phys. Rev. D **84**, 043001, arXiv:1104.5547 [astro-ph.CO] .

[471] Chevallier, M., and D. Polarski (2001), Int. J. Mod. Phys. **D10**, 213, arXiv:gr-qc/0009008 [gr-qc] .

[472] Chiang, C.-T., and A. Slosar (2018), arXiv:1811.03624 [astro-ph.CO] .

[473] Chiba, T., N. Sugiyama, and T. Nakamura (1997), Mon. Not. Roy. Astron. Soc. **289**, L5, arXiv:astro-ph/9704199 .

[474] Choi, G., M. Suzuki, and T. T. Yanagida (2020), Phys. Rev. D **101** (7), 075031, arXiv:2002.00036 [hep-ph] .

[475] Choi, G., M. Suzuki, and T. T. Yanagida (2020), Phys. Lett. B **805**, 135408, arXiv:1910.00459 [hep-ph] .

[476] Choi, G., T. T. Yanagida, and N. Yokozaki (2021), JHEP **01**, 127, arXiv:2010.06892 [hep-ph] .

[477] Choi, K.-Y., J. Kim, and C. Rott (2019), Phys. Rev. D **99** (8), 083018, arXiv:1903.03302 [astro-ph.CO] .

[78] Chu, X., B. Dasgupta, and J. Kopp (2015), JCAP **10**, 011, arXiv:1505.02795 [hep-ph] .

[479] Chuang, C.-H., *et al.* (2013), Mon. Not. Roy. Astron. Soc. **433**, 3559, arXiv:1303.4486 [astro-ph.CO] .

[480] Chudaykin, A., D. Gorbunov, and N. Nedelko (2020), JCAP **08**, 013, arXiv:2004.13046 [astro-ph.CO] .

[481] Chudaykin, A., D. Gorbunov, and N. Nedelko (2021), Phys. Rev. D **103** (4), 043529, arXiv:2011.04682 [astro-ph.CO] .

[482] Chudaykin, A., D. Gorbunov, and I. Tkachev (2016), Phys. Rev. D **94**, 023528, arXiv:1602.08121 [astro-ph.CO] .

[483] Chudaykin, A., D. Gorbunov, and I. Tkachev (2018), Phys. Rev. D **97** (8), 083508, arXiv:1711.06738 [astro-ph.CO] .

[484] Cimatti, A., R. Laureijs, B. Leibundgut, S. Lilly, R. Nichol, A. Refregier, P. Rosati, M. Steinmetz, N. Thatte, and E. Valentijn (2009), arXiv:0912.0914 [astro-ph.CO] .

[485] Ciufolini, I., and F. Fidecaro (1997), *Proceedings of the International Conference on Gravitational Waves: Sources and Detectors : Cascina (Pisa), Italy 19-23 March 1996*, Edoardo Amaldi Foundation series (World Scientific).

[486] Clara, M., and C. Martins (2020), Astron. Astrophys. **633**, L11, arXiv:2001.01787 [astro-ph.CO] .

[487] Clark, S. J., K. Vattis, J. Fan, and S. M. Koushiappas (2021), arXiv:2110.09562 [astro-ph.CO] .

[488] Clark, S. J., K. Vattis, and S. M. Koushiappas (2021), Phys. Rev. D **103** (4), 043014, arXiv:2006.03678 [astro-ph.CO] .

[489] Coc, A., S. Goriely, Y. Xu, M. Saimpert, and E. Vangioni (2012), Astrophys. J. **744**, 158, arXiv:1107.1117 [astro-ph.CO] .

[490] Coc, A., N. J. Nunes, K. A. Olive, J.-P. Uzan, and E. Vangioni (2007), Phys. Rev. D **76**, 023511, arXiv:astro-ph/0610733 .

[491] Coc, A., and E. Vangioni (2017), Int. J. Mod. Phys. E **26** (08), 1741002, arXiv:1707.01004 [astro-ph.CO] .

[492] Colas, T., G. D'amico, L. Senatore, P. Zhang, and F. Beutler (2020), JCAP **06**, 001, arXiv:1909.07951 [astro-ph.CO] .

[493] Cole, S., *et al.* (2dFGRS) (2005), Mon. Not. Roy. Astron. Soc. **362**, 505, arXiv:astro-ph/0501174 .

[494] Colgáin, E. O. (2019), JCAP **09**, 006, arXiv:1903.11743 [astro-ph.CO] .

[495] Colgáin, E. O., and M. M. Sheikh-Jabbari (2021), Class. Quant. Grav. **38** (17), 177001, arXiv:2102.09816 [gr-qc] .

[496] Colgáin, E. O., M. M. Sheikh-Jabbari, and L. Yin (2021), Phys. Rev. D **104** (2), 023510, arXiv:2104.01930 [astro-ph.CO] .

[497] Colgáin, E. O., and H. Yavartanoo (2019), Phys. Lett. B **797**, 134907, arXiv:1905.02555 [astro-ph.CO] .

[498] Colin, J., R. Mohayaee, M. Rameez, and S. Sarkar (2017), Mon. Not. Roy. Astron. Soc. **471** (1), 1045, arXiv:1703.09376 [astro-ph.CO] .

[499] Colin, J., R. Mohayaee, M. Rameez, and S. Sarkar (2019), Astron. Astrophys. **631**, L13, arXiv:1808.04597 [astro-ph.CO] .

[500] Colless, M., *et al.* (2DFGRS) (2001), Mon. Not. Roy. Astron. Soc. **328**, 1039, arXiv:astro-ph/0106498 .

[501] Comelli, D., M. Pietroni, and A. Riotto (2003), Phys. Lett. B **571**, 115, arXiv:hep-ph/0302080 .

[502] Condon, J. J., W. D. Cotton, E. W. Greisen, Q. F. Yin, R. A. Perley, G. B. Taylor, and J. J. Broderick (1998), Astron. J. **115**, 1693.

[503] Conlon, J. P., and M. C. D. Marsh (2013), JHEP **10**, 214, arXiv:1304.1804 [hep-ph] .

[504] Conn, A. R., *et al.* (2013), Astrophys. J. **766**, 120, arXiv:1301.7131 [astro-ph.CO] .

[505] Cooke, L. H., *et al.* (2022), 10.1093/mnras/stac588, arXiv:2203.00694 [astro-ph.GA] .

[506] Cooray, A. R., and D. Huterer (1999), Astrophys. J. Lett. **513**, L95, arXiv:astro-ph/9901097 .

[507] Copeland, E. J., A. R. Liddle, and D. Wands (1998), Phys. Rev. D **57**, 4686, arXiv:gr-qc/9711068 .

[508] Copeland, E. J., M. Sami, and S. Tsujikawa (2006), Int. J. Mod. Phys. **D15**, 1753, arXiv:hep-th/0603057 [hep-th] .

[509] Copi, C., D. Huterer, D. Schwarz, and G. Starkman (2007), Phys. Rev. D **75**, 023507, arXiv:astro-ph/0605135 .

[510] Copi, C. J., D. Huterer, D. Schwarz, and G. Starkman (2006), Mon. Not. Roy. Astron. Soc. **367**, 79, arXiv:astro-ph/0508047 .

[511] Copi, C. J., D. Huterer, D. J. Schwarz, and G. D. Starkman (2009), Mon. Not. Roy. Astron. Soc. **399**, 295, arXiv:0808.3767 [astro-ph] .

[512] Copi, C. J., D. Huterer, D. J. Schwarz, and G. D. Starkman (2010), Adv. Astron. **2010**, 847541, arXiv:1004.5602 [astro-ph.CO] .

[513] Copi, C. J., D. Huterer, D. J. Schwarz, and G. D. Starkman (2013), Mon. Not. Roy. Astron. Soc. **434**, 3590, arXiv:1303.4786 [astro-ph.CO] .

[514] Copi, C. J., D. Huterer, D. J. Schwarz, and G. D. Starkman (2015), Mon. Not. Roy. Astron. Soc. **451** (3), 2978, arXiv:1310.3831 [astro-ph.CO] .

[515] Copi, C. J., D. Huterer, D. J. Schwarz, and G. D. Starkman (2015), Mon. Not. Roy. Astron. Soc. **449** (4), 3458, arXiv:1311.4562 [astro-ph.CO] .

[516] Copi, C. J., D. Huterer, and G. D. Starkman (2004), Phys. Rev. D **70**, 043515, arXiv:astro-ph/0310511 .

[517] Copi, C. J., M. O'Dwyer, and G. D. Starkman (2016), Mon. Not. Roy. Astron. Soc. **463** (3), 3305, arXiv:1605.09732 [astro-ph.CO] .

[518] Cornish, N. J., and T. B. Littenberg (2015), Class. Quant. Grav. **32** (13), 135012, arXiv:1410.3835 [gr-qc] .

[519] Cornish, N. J., T. B. Littenberg, B. Bécsy, K. Chatziioannou, J. A. Clark, S. Ghonge, and M. Millhouse (2021), Phys. Rev. D **103** (4), 044006, arXiv:2011.09494 [gr-qc] .

[520] Corona, M. A., R. Murgia, M. Cadeddu, M. Archidiacono, S. Gariazzo, C. Giunti, and S. Hannestad (2021), arXiv:2112.00037 [astro-ph.CO] .

[521] Costa, A. A., X.-D. Xu, B. Wang, E. G. M. Ferreira, and E. Abdalla (2014), Phys. Rev. D **89** (10), 103531, arXiv:1311.7380 [astro-ph.CO] .

[522] Costanzi, M., B. Sartoris, M. Viel, and S. Borgani (2014), JCAP **10**, 081, arXiv:1407.8338 [astro-ph.CO] .

[523] Costanzi, M., *et al.* (DES) (2019), Mon. Not. Roy. Astron. Soc. **488** (4), 4779, arXiv:1810.09456 [astro-ph.CO] .

[524] Costanzi, M., *et al.* (DES, SPT) (2021), Phys. Rev. D **103** (4), 043522, arXiv:2010.13800 [astro-ph.CO] .

[525] Coulter, D., *et al.* (2017), Science **358**, 1556, arXiv:1710.05452 [astro-ph.HE] .

[526] Crowder, J., and N. J. Cornish (2005), Phys. Rev. D **72**, 083005, arXiv:gr-qc/0506015 .

[527] Cruz, M., L. Cayon, E. Martinez-Gonzalez, P. Vielva, and J. Jin (2007), Astrophys. J. **655**, 11, arXiv:astro-ph/0603859 .

[528] Cruz, M., E. Martinez-Gonzalez, and P. Vielva (2009), 10.1007/978-3-642-11250-8-37, arXiv:0901.1986 [astro-ph.CO] .

[529] Cruz, M., E. Martinez-Gonzalez, P. Vielva, and L. Cayon (2005), Mon. Not. Roy. Astron. Soc. **356**, 29, arXiv:astro-ph/0405341 .

[530] Cruz, M., M. Tucci, E. Martinez-Gonzalez, and P. Vielva (2006), Mon. Not. Roy. Astron. Soc. **369**, 57, arXiv:astro-ph/0601427 .

[531] Cucchiara, A., A. J. Levan, D. B. Fox, N. R. Tanvir, T. N. Ukwatta, E. Berger, T. Krühler, A. Küpcü Yoldaş, X. F. Wu, K. Toma, J. Greiner, F. E. Olivares, A. Rowlinson, L. Amati, T. Sakamoto, et. al. (2011), ApJ **736** (1), 7, arXiv:1105.4915 [astro-ph.CO] .

[532] Cuceu, A., J. Farr, P. Lemos, and A. Font-Ribera (2019), JCAP **1910** (10), 044, arXiv:1906.11628 [astro-ph.CO] .

[533] Cuesta, A. J., M. E. Gómez, J. I. Illana, and M. Masip (2021), arXiv:2109.07336 [hep-ph] .

[88] Cuesta, A. J., L. Verde, A. Riess, and R. Jimenez (2015), Mon. Not. Roy. Astron. Soc. **448** (4), 3463, arXiv:1411.1094 [astro-ph.CO] .

[535] Cunha, P. V. P., C. A. R. Herdeiro, and E. Radu (2017), Phys. Rev. D **96** (2), 024039, arXiv:1705.05461 [gr-qc] .

[536] Cyburt, R. H., B. D. Fields, and K. A. Olive (2003), Phys. Lett. B **567**, 227, arXiv:astro-ph/0302431 .

[537] Cyburt, R. H., B. D. Fields, and K. A. Olive (2004), Phys. Rev. D **69**, 123519, arXiv:astro-ph/0312629 .

[538] Cyburt, R. H., B. D. Fields, and K. A. Olive (2008), JCAP **11**, 012, arXiv:0808.2818 [astro-ph] .

[539] Cyburt, R. H., B. D. Fields, K. A. Olive, and T.-H. Yeh (2016), Rev. Mod. Phys. **88**, 015004, arXiv:1505.01076 [astro-ph.CO] .

[540] Cyr-Racine, F.-Y., R. de Putter, A. Raccanelli, and K. Sigurdson (2014), Phys. Rev. D **89** (6), 063517, arXiv:1310.3278 [astro-ph.CO] .

[541] Cyr-Racine, F.-Y., and K. Sigurdson (2014), Phys. Rev. D **90** (12), 123533, arXiv:1306.1536 [astro-ph.CO] .

[542] Cyr-Racine, F.-Y., K. Sigurdson, J. Zavala, T. Bringmann, M. Vogelsberger, and C. Pfrommer (2016), Phys. Rev. D **93** (12), 123527, arXiv:1512.05344 [astro-ph.CO] .

[543] Czerny, B., *et al.* (2018), Space Sci. Rev. **214** (1), 32, arXiv:1801.00598 [astro-ph.GA] .

[544] D'Agostino, R., and O. Luongo (2018), Phys. Rev. D **98** (12), 124013, arXiv:1807.10167 [gr-qc] .

[545] D'Agostino, R., and R. C. Nunes (2020), Phys. Rev. D **101** (10), 103505, arXiv:2002.06381 [astro-ph.CO] .

[546] Dai, W.-M., Y.-Z. Ma, and H.-J. He (2020), Phys. Rev. D **102**, 121302, arXiv:2003.03602 [astro-ph.CO] .

[547] Dai, Y., X.-G. Zheng, Z.-X. Li, H. Gao, and Z.-H. Zhu (2021), 10.1051/0004-6361/202140895, arXiv:2111.05544 [astro-ph.HE] .

[548] Daido, R., F. Takahashi, and W. Yin (2017), JCAP **05**, 044, arXiv:1702.03284 [hep-ph] .

[549] Daido, R., F. Takahashi, and W. Yin (2018), JHEP **02**, 104, arXiv:1710.11107 [hep-ph] .

[550] Dainotti, M., and R. Del Vecchio (2017), New Astron. Rev. **77**, 23, arXiv:1703.06876 [astro-ph.HE] .

[551] Dainotti, M. G., V. F. Cardone, E. Piedipalumbo, and S. Capozziello (2013), Mon. Not. Roy. Astron. Soc. **436**, 82, arXiv:1308.1918 [astro-ph.HE] .

[552] Dainotti, M. G., B. De Simone, T. Schiavone, G. Montani, E. Rinaldi, and G. Lambiase (2021), Astrophys. J. **912**, 150, arXiv:2103.02117 [astro-ph.CO] .

[553] Dainotti, M. G., B. De Simone, T. Schiavone, G. Montani, E. Rinaldi, G. Lambiase, M. Bogdan, and S. Ugale (2022), Galaxies **10** (1), 24, arXiv:2201.09848 [astro-ph.CO] .

[554] Dalal, N., D. E. Holz, S. A. Hughes, and B. Jain (2006), Phys. Rev. D **74**, 063006, arXiv:astro-ph/0601275 .

[555] Dálya, G., P. Raffai, and B. Bécsy (2021), Class. Quant. Grav. **38** (6), 065002, arXiv:2006.06256 [astro-ph.HE] .

[556] D'Amico, G., J. Gleyzes, N. Kokron, D. Markovic, L. Senatore, P. Zhang, F. Beutler, and H. Gil-Marín (2020), JCAP **05**, 005, arXiv:1909.05271 [astro-ph.CO] .

[557] D'Amico, G., Z. Huang, M. Mancarella, and F. Vernizzi (2017), JCAP **02**, 014, arXiv:1609.01272 [astro-ph.CO] .

[558] D'Amico, G., L. Senatore, and P. Zhang (2021), JCAP **01**, 006, arXiv:2003.07956 [astro-ph.CO] .

[559] D'Amico, G., L. Senatore, P. Zhang, and H. Zheng (2021), JCAP **05**, 072, arXiv:2006.12420 [astro-ph.CO] .

[560] Das, A., and S. Ghosh (2021), JCAP **07**, 038, arXiv:2011.12315 [astro-ph.CO] .

[561] Das, K. K., K. Sankharva, and P. Jain (2021), JCAP **07**, 035, arXiv:2101.11016 [astro-ph.CO] .

[562] Das, S., and N. Banerjee (2008), Phys. Rev. D **78**, 043512, arXiv:0803.3936 [gr-qc] .

[563] Das, S., A. Maharana, V. Poulin, and R. Kumar (2021), arXiv:2104.03329 [astro-ph.CO] .

[564] Davis, A.-C., E. A. Lim, J. Sakstein, and D. Shaw (2012), Phys. Rev. D **85**, 123006, arXiv:1102.5278 [astro-ph.CO] .

[565] Davis, M., G. Efstathiou, C. S. Frenk, and S. D. M. White (1985), Astrophys. J. **292**, 371.

[566] Dawson, K. S., *et al.* (BOSS) (2013), Astron. J. **145**, 10, arXiv:1208.0022 [astro-ph.CO] .

[567] Dawson, K. S., *et al.* (2016), Astron. J. **151**, 44, arXiv:1508.04473 [astro-ph.CO] .

[568] De Felice, A., S. Mukohyama, and M. C. Pookkillath (2021), Phys. Lett. B **816**, 136201, arXiv:2009.08718 [astro-ph.CO] .

[569] De Felice, A., and S. Tsujikawa (2010), Living Rev. Rel. **13**, 3, arXiv:1002.4928 [gr-qc] .

[570] de Vaucouleurs, G. (1986), in *Galaxy Distances and Deviations from Universal Expansion*, NATO Advanced Study Institute (ASI) Series C, Vol. 180, edited by B. F. Madore and R. B. Tully, pp. 1–6.

[571] de Vaucouleurs, G., and W. L. Peters (1985), ApJ **297**, 27.

[572] Deffayet, C., G. R. Dvali, G. Gabadadze, and A. I. Vainshtein (2002), Phys. Rev. D **65**, 044026, arXiv:hep-th/0106001 .

[573] Deffayet, C., X. Gao, D. A. Steer, and G. Zahariade (2011), Phys. Rev. D **84**, 064039, arXiv:1103.3260 [hep-th] .

[574] Dekker, H., S. D'Odorico, A. Kaufer, B. Delabre, and H. Kotzlowski (2000), in *Optical and IR Telescope Instrumentation and Detectors*, Vol. 4008, edited by M. Iye and A. F. M. Moorwood, International Society for Optics and Photonics (SPIE) pp. 534 – 545.

[575] Del Popolo, A., and M. Le Delliou (2017), Galaxies **5** (1), 17, arXiv:1606.07790 [astro-ph.CO] .

[576] Delubac, T., *et al.* (BOSS) (2015), Astron. Astrophys. **574**, A59, arXiv:1404.1801 [astro-ph.CO] .

[577] Demianski, M., E. Piedipalumbo, D. Sawant, and L. Amati (2017), Astron. Astrophys. **598**, A112, arXiv:1610.00854 [astro-ph.CO] .

[578] Demianski, M., E. Piedipalumbo, D. Sawant, and L. Amati (2021), Mon. Not. Roy. Astron. Soc. **506** (1), 903, arXiv:1911.08228 [astro-ph.CO] .

[579] Denzel, P., J. P. Coles, P. Saha, and L. L. R. Williams (2021), Mon. Not. Roy. Astron. Soc. **501** (1), 784, arXiv:2007.14398 [astro-ph.CO] .

[580] Deppisch, F. F., L. Graf, W. Rodejohann, and X.-J. Xu (2020), Phys. Rev. D **102** (5), 051701, arXiv:2004.11919 [hep-ph] .

[581] D'Eramo, F., R. Z. Ferreira, A. Notari, and J. L. Bernal (2018), JCAP **1811** (11), 014, arXiv:1808.07430 [hep-ph] .

[582] Desai, A., K. R. Dienes, and B. Thomas (2020), Phys. Rev. D **101** (3), 035031, arXiv:1909.07981 [astro-ph.CO] .

[583] Desmond, H., B. Jain, and J. Sakstein (2019), Phys. Rev. **D100** (4), 043537, arXiv:1907.03778 [astro-ph.CO] .

[584] Desmond, H., and J. Sakstein (2020), Phys. Rev. D **102** (2), 023007, arXiv:2003.12876 [astro-ph.CO] .

[585] Dessart, L., and D. J. Hillier (2005), Astron. Astrophys. **439**, 671, arXiv:astro-ph/0505465 .

[586] Dessart, L., *et al.* (2008), Astrophys. J. **675**, 644, arXiv:0711.1815 [astro-ph] .

[587] Desvignes, G., *et al.* (2016), Mon. Not. Roy. Astron. Soc. **458** (3), 3341, arXiv:1602.08511 [astro-ph.HE] .

[588] Detweiler, S. L. (1979), Astrophys. J. **234**, 1100.

[589] Dhawan, S., J. Alsing, and S. Vagnozzi (2021), Mon. Not. Roy. Astron. Soc. **506** (1), L1, arXiv:2104.02485 [astro-ph.CO] .

[590] Dhawan, S., D. Brout, D. Scolnic, A. Goobar, A. G. Riess, and V. Miranda (2020), Astrophys. J. **894** (1), 54, arXiv:2001.09260 [astro-ph.CO] .

[591] Dhawan, S., S. W. Jha, and B. Leibundgut (2018), Astron. Astrophys. **609**, A72, arXiv:1707.00715 [astro-ph.CO] .

[592] Dhawan, S., *et al.* (2022), arXiv:2203.04241 [astro-ph.CO] .

[593] Di, H., and Y. Gong (2018), JCAP **07**, 007, arXiv:1707.09578 [astro-ph.CO] .

[594] Di Bari, P., S. F. King, and A. Merle (2013), Phys. Lett. B **724**, 77, arXiv:1303.6267 [hep-ph] .

[595] Di Cintio, A., and F. Lelli (2016), Mon. Not. Roy. Astron. Soc. **456** (1), L127, arXiv:1511.06616 [astro-ph.GA] .

[596] Di Valentino, E. (2021), Mon. Not. Roy. Astron. Soc. **502** (2), 2065, arXiv:2011.00246 [astro-ph.CO] .

[108] Di Valentino, E., C. Bøehm, E. Hivon, and F. R. Bouchet (2018), Phys. Rev. **D97** (4), 043513, arXiv:1710.02559 [astro-ph.CO] .

[598] Di Valentino, E., and F. R. Bouchet (2016), JCAP **10**, 011, arXiv:1609.00328 [astro-ph.CO] .

[599] Di Valentino, E., R. Z. Ferreira, L. Visinelli, and U. Danielsson (2019), Phys. Dark Univ. **26**, 100385, arXiv:1906.11255 [astro-ph.CO] .

[600] Di Valentino, E., S. Gariazzo, C. Giunti, O. Mena, S. Pan, and W. Yang (2021), arXiv:2110.03990 [astro-ph.CO] .

[601] Di Valentino, E., S. Gariazzo, O. Mena, and S. Vagnozzi (2020), JCAP **07** (07), 045, arXiv:2005.02062 [astro-ph.CO] .

[602] Di Valentino, E., E. V. Linder, and A. Melchiorri (2018), Phys. Rev. **D97** (4), 043528, arXiv:1710.02153 [astro-ph.CO] .

[603] Di Valentino, E., E. V. Linder, and A. Melchiorri (2020), Phys. Dark Univ. **30**, 100733, arXiv:2006.16291 [astro-ph.CO] .

[604] Di Valentino, E., and A. Melchiorri (2021), arXiv:2112.02993 [astro-ph.CO] .

[605] Di Valentino, E., A. Melchiorri, Y. Fantaye, and A. Heavens (2018), Phys. Rev. D **98** (6), 063508, arXiv:1808.09201 [astro-ph.CO] .

[606] Di Valentino, E., A. Melchiorri, E. V. Linder, and J. Silk (2017), Phys. Rev. **D96** (2), 023523, arXiv:1704.00762 [astro-ph.CO] .

[607] Di Valentino, E., A. Melchiorri, and O. Mena (2017), Phys. Rev. **D96** (4), 043503, arXiv:1704.08342 [astro-ph.CO] .

[608] Di Valentino, E., A. Melchiorri, O. Mena, S. Pan, and W. Yang (2021), Mon. Not. Roy. Astron. Soc. **502** (1), L23, arXiv:2011.00283 [astro-ph.CO] .

[609] Di Valentino, E., A. Melchiorri, O. Mena, and S. Vagnozzi (2020), Phys. Dark Univ. **30**, 100666, arXiv:1908.04281 [astro-ph.CO] .

[610] Di Valentino, E., A. Melchiorri, O. Mena, and S. Vagnozzi (2020), Phys. Rev. D **101** (6), 063502, arXiv:1910.09853 [astro-ph.CO] .

[611] Di Valentino, E., A. Melchiorri, and J. Silk (2016), Phys. Rev. D **93** (2), 023513, arXiv:1509.07501 [astro-ph.CO] .

[612] Di Valentino, E., A. Melchiorri, and J. Silk (2016), Phys. Lett. **B761**, 242, arXiv:1606.00634 [astro-ph.CO] .

[613] Di Valentino, E., A. Melchiorri, and J. Silk (2019), Nature Astron. **4** (2), 196, arXiv:1911.02087 [astro-ph.CO] .

[614] Di Valentino, E., A. Melchiorri, and J. Silk (2020), JCAP **01**, 013, arXiv:1908.01391 [astro-ph.CO] .

[615] Di Valentino, E., O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri, D. F. Mota, A. G. Riess, and J. Silk (2021), Class. Quant. Grav. **38** (15), 153001, arXiv:2103.01183 [astro-ph.CO] .

[616] Di Valentino, E., and L. Mersini-Houghton (2017), JCAP **03**, 020, arXiv:1612.08334 [astro-ph.CO] .

[617] Di Valentino, E., A. Mukherjee, and A. A. Sen (2021), Entropy **23** (4), 404, arXiv:2005.12587 [astro-ph.CO] .

[618] Di Valentino, E., S. Pan, W. Yang, and L. A. Anchordoqui (2021), Phys. Rev. D **103** (12), 123527, arXiv:2102.05641 [astro-ph.CO] .

[619] Di Valentino, E., *et al.* (2021), Astropart. Phys. **131**, 102604, arXiv:2008.11285 [astro-ph.CO] .

[620] Di Valentino, E., *et al.* (2021), Astropart. Phys. **131**, 102605, arXiv:2008.11284 [astro-ph.CO] .

[621] Di Valentino, E., *et al.* (2021), Astropart. Phys. **131**, 102607, arXiv:2008.11286 [astro-ph.CO] .

[622] Dialektopoulos, K. F., J. L. Said, and Z. Oikonomopoulou (2021), 10.1140/epjc/s10052-022-10201-7, arXiv:2112.15045 [gr-qc] .

[623] Diaz Rivero, A., V. Miranda, and C. Dvorkin (2019), Phys. Rev. D **100** (6), 063504, arXiv:1903.03125 [astro-ph.CO] .

[624] Diego-Palazuelos, P., *et al.* (2022), Phys. Rev. Lett. **128** (9), 091302, arXiv:2201.07682 [astro-ph.CO] .

[625] Dienes, K. R., F. Huang, J. Kost, B. Thomas, and H.-B. Yu (2021), arXiv:2112.09105 [astro-ph.CO] .

[626] Ding, X., M. Biesiada, X. Zheng, K. Liao, Z. Li, and Z.-H. Zhu (2019), JCAP **04**, 033, arXiv:1801.05073 [astro-ph.CO] .

[627] Dirirsa, F. F., S. Razzaque, F. Piron, M. Arimoto, M. Axelsson, D. Kocevski, F. Longo, M. Ohno, and S. Zhu (2019), Astrophys. J. **887**, 13, arXiv:1910.07009 [astro-ph.HE] .

[628] Djorgovski, S., and M. Davis (1987), Astrophys. J. **313**, 59.

[629] Dodelson, S. (2003), *Modern Cosmology* (Academic Press, Amsterdam).

[630] Dodelson, S., K. Heitmann, C. Hirata, K. Honscheid, A. Roodman, U. Seljak, A. Slosar, and M. Trodden (2016), arXiv:1604.07626 [astro-ph.CO] .

[113] Dodelson, S., and L. M. Widrow (1994), Phys. Rev. Lett. **72**, 17, arXiv:hep-ph/9303287 .

[632] Domínguez, A., and F. Prada (2013), Astrophys. J. Lett. **771**, L34, arXiv:1305.2163 [astro-ph.CO] .

[633] Domínguez, A., R. Wojtak, J. Finke, M. Ajello, K. Helgason, F. Prada, A. Desai, V. Paliya, L. Marcotulli, and D. Hartmann (2019), 10.3847/1538-4357/ab4a0e, arXiv:1903.12097 [astro-ph.CO] .

[634] Dong, F., Y. Yu, J. Zhang, X. Yang, and P. Zhang (2020), Mon. Not. Roy. Astron. Soc. **500** (3), 3838, arXiv:2006.14202 [astro-ph.CO] .

[635] Donini, A., and S. G. Marimón (2016), Eur. Phys. J. C **76** (12), 696, arXiv:1609.05654 [hep-ph] .

[636] Drell, P. S., T. J. Loredo, and I. Wasserman (2000), Astrophys. J. **530**, 593, arXiv:astro-ph/9905027 .

[637] Drewes, M. (2013), Int. J. Mod. Phys. E **22**, 1330019, arXiv:1303.6912 [hep-ph] .

[638] Drinkwater, M. J., *et al.* (2010), Mon. Not. Roy. Astron. Soc. **401**, 1429, arXiv:0911.4246 [astro-ph.CO] .

[639] Du, M., W. Yang, L. Xu, S. Pan, and D. F. Mota (2019), Phys. Rev. D **100** (4), 043535, arXiv:1812.01440 [astro-ph.CO] .

[640] Dumont, V., and J. K. Webb (2017), Mon. Not. Roy. Astron. Soc. **468** (2), 1568, arXiv:1701.03176 [astro-ph.CO] .

[641] Dunkley, J., *et al.* (WMAP) (2009), Astrophys. J. Suppl. **180**, 306, arXiv:0803.0586 [astro-ph] .

[642] Durrer, R., and A. Neronov (2013), Astron. Astrophys. Rev. **21**, 62, arXiv:1303.7121 [astro-ph.CO] .

[643] Dutcher, D., *et al.* (SPT-3G) (2021), Phys. Rev. D **104** (2), 022003, arXiv:2101.01684 [astro-ph.CO] .

[644] Dutta, K., A. Roy, Ruchika, A. A. Sen, and M. Sheikh-Jabbari (2019), Phys. Rev. D **100** (10), 103501, arXiv:1908.07267 [astro-ph.CO] .

[645] Dutta, K., Ruchika, A. Roy, A. A. Sen, and M. M. Sheikh-Jabbari (2020), Gen. Rel. Grav. **52** (2), 15, arXiv:1808.06623 [astro-ph.CO] .

[646] Dutton, A. A. (2012), Mon. Not. Roy. Astron. Soc. **424**, 3123, arXiv:1206.1855 [astro-ph.CO] .

[647] Dvorkin, C., M. Wyman, D. H. Rudd, and W. Hu (2014), Phys. Rev. D **90** (8), 083503, arXiv:1403.8049 [astro-ph.CO] .

[648] Dzuba, V. A., V. V. Flambaum, and J. K. Webb (1999), Phys. Rev. Lett. **82**, 888, arXiv:physics/9802029 .

[649] Eastman, R. G., B. P. Schmidt, and R. Kirshner (1996), ApJ **466**, 911.

[650] Edholm, J., A. S. Koshelev, and A. Mazumdar (2016), Phys. Rev. D **94** (10), 104033, arXiv:1604.01989 [gr-qc] .

[651] Efstathiou, G. (1999), Mon. Not. Roy. Astron. Soc. **310**, 842, arXiv:astro-ph/9904356 .

[652] Efstathiou, G. (2021), Mon. Not. Roy. Astron. Soc. **505** (3), 3866, arXiv:2103.08723 [astro-ph.CO] .

[653] Efstathiou, G., and J. R. Bond (1999), Mon. Not. Roy. Astron. Soc. **304**, 75, arXiv:astro-ph/9807103 .

[654] Efstathiou, G., and S. Gratton (2019), arXiv:1910.00483 [astro-ph.CO] .

[655] Efstathiou, G., and S. Gratton (2020), 10.1093/mnrasl/slaa093, arXiv:2002.06892 [astro-ph.CO] .

[656] Efstathiou, G., Y.-Z. Ma, and D. Hanson (2010), Mon. Not. Roy. Astron. Soc. **407**, 2530, arXiv:0911.5399 [astro-ph.CO] .

[657] Eichler, D., M. Livio, T. Piran, and D. N. Schramm (1989), Nature **340**, 126.

[658] Eigenbrod, A., F. Courbin, S. Dye, G. Meylan, D. Sluse, P. Saha, C. Vuissoz, and P. Magain (2006), Astron. Astrophys. **451**, 747, arXiv:astro-ph/0510641 .

[659] Einstein, A. (1917), Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) **1917**, 142.

[660] Eisenhardt, P. R. M., F. Marocco, J. W. Fowler, A. M. Meisner, J. D. Kirkpatrick, N. Garcia, T. H. Jarrett, R. Koontz, E. J. Marchese, S. A. Stanford, D. Caselden, M. C. Cushing, R. M. Cutri, J. K. Faherty, C. R. Gelino, A. H. Gonzalez, A. Mainzer, B. Mobasher, D. J. Schlegel, D. Stern, H. I. Teplitz, and E. L. Wright (2020), The Astrophysical Journal Supplement Series **247** (2), 69.

[661] Eisenstein, D. J., and W. Hu (1998), Astrophys. J. **496**, 605, arXiv:astro-ph/9709112 .

[662] Eisenstein, D. J., H.-j. Seo, and M. J. White (2007), Astrophys. J. **664**, 660, arXiv:astro-ph/0604361 .

[663] Eisenstein, D. J., *et al.* (SDSS) (2005), Astrophys. J. **633**, 560, arXiv:astro-ph/0501171 [astro-ph] .

[664] El-Neaj, Y. A., *et al.* (AEDGE) (2020), EPJ Quant. Technol. **7**, 6, arXiv:1908.00802 [gr-qc] .

[665] Elizalde, E., J. Gluza, and M. Khurshudyan (2021), arXiv:2104.01077 [astro-ph.CO] .

[666] Elizalde, E., M. Khurshudyan, S. D. Odintsov, and R. Myrzakulov (2020), Phys. Rev. D **102** (12), 123501, arXiv:2006.01879 [gr-qc] .

[667] Ellis, G. F. R., and J. E. Baldwin (1984), Monthly Notices of the Royal Astronomical Society **206** (2), 377, https://academic.oup.com/mnras/article-pdf/206/2/377/18187025/mnras206-0377.pdf .

[668] Enqvist, K., S. Nadathur, T. Sekiguchi, and T. Takahashi (2015), JCAP **1509**, 067, arXiv:1505.05511 [astro-ph.CO] .

[121] Erben, T., *et al.* (2013), Mon. Not. Roy. Astron. Soc. **433**, 2545, arXiv:1210.8156 [astro-ph.CO] .

[670] Erickcek, A. L., S. M. Carroll, and M. Kamionkowski (2008), Phys. Rev. D **78**, 083012, arXiv:0808.1570 [astro-ph] .

[671] Erickcek, A. L., M. Kamionkowski, and S. M. Carroll (2008), Phys. Rev. D **78**, 123520, arXiv:0806.0377 [astro-ph] .

[672] Eriksen, H., F. Hansen, A. Banday, K. Gorski, and P. Lilje (2004), Astrophys. J. **605**, 14, [Erratum: Astrophys.J. 609, 1198 (2004)], arXiv:astro-ph/0307507 .

[673] Eriksen, H. K., A. J. Banday, K. M. Gorski, F. K. Hansen, and P. B. Lilje (2007), Astrophys. J. Lett. **660**, L81, arXiv:astro-ph/0701089 .

[674] Escamilla-Rivera, C., and J. Levi Said (2020), Class. Quant. Grav. **37** (16), 165002, arXiv:1909.10328 [gr-qc] .

[675] Escamilla-Rivera, C., J. Levi Said, and J. Mifsud (2021), JCAP **10**, 016, arXiv:2105.14332 [astro-ph.CO] .

[676] Escudero, M., and S. J. Witte (2020), Eur. Phys. J. C **80** (4), 294, arXiv:1909.04044 [astro-ph.CO] .

[677] Escudero, M., and S. J. Witte (2021), arXiv:2103.03249 [hep-ph] .

[678] Escudero Abenza, M. (2020), JCAP **05**, 048, arXiv:2001.04466 [hep-ph] .

[679] Esposito-Farese, G., and D. Polarski (2001), Phys. Rev. D **63**, 063504, arXiv:gr-qc/0009034 .

[680] Essick, R., S. Vitale, and M. Evans (2017), Phys. Rev. D **96** (8), 084004, arXiv:1708.06843 [gr-qc] .

[681] Evans, T. M., *et al.* (2014), Mon. Not. Roy. Astron. Soc. **445** (1), 128, arXiv:1409.1923 [astro-ph.CO] .

[682] Evrard, A. E. (1989), ApJ **341**, L71.

[683] Evslin, J. (2017), JCAP **04**, 024, arXiv:1604.02809 [astro-ph.CO] .

[684] Faber, S. M., and R. E. Jackson (1976), Astrophys. J. **204**, 668.

[685] Faizal, M. (2016), Phys. Lett. **B757**, 244, arXiv:1605.00925 [hep-th] .

[686] Faizal, M., M. M. Khalil, and S. Das (2016), Eur. Phys. J. **C76** (1), 30, arXiv:1501.03111 [physics.gen-ph] .

[687] Falco, E. E., M. V. Gorenstein, and I. I. Shapiro (1985), The Astrophysical Journal **289**, L1.

[688] Fanizza, G., B. Fiorini, and G. Marozzi (2021), Phys. Rev. D **104** (8), 083506, arXiv:2102.12419 [astro-ph.CO] .

[689] Fanizza, G., M. Gasperini, G. Marozzi, and G. Veneziano (2020), JCAP **02**, 017, arXiv:1911.09469 [gr-qc] .

[690] Farhang, M., and N. Khosravi (2021), Phys. Rev. D **103** (8), 083523, arXiv:2011.08050 [astro-ph.CO] .

[691] Farmer, A. J., and E. S. Phinney (2003), Mon. Not. Roy. Astron. Soc. **346**, 1197, arXiv:astro-ph/0304393 .

[692] Farooq, O., F. R. Madiyar, S. Crandall, and B. Ratra (2017), Astrophys. J. **835** (1), 26, arXiv:1607.03537 [astro-ph.CO] .

[693] Farr, W. M., M. Fishbach, J. Ye, and D. Holz (2019), Astrophys. J. Lett. **883** (2), L42, arXiv:1908.09084 [astro-ph.CO] .

[694] Farren, G. S., O. H. E. Philcox, and B. D. Sherwin (2021), 10.1103/PhysRevD.105.063503, arXiv:2112.10749 [astro-ph.CO] .

[695] Farrugia, C. R., J. Sultana, and J. Mifsud (2020), Phys. Rev. D **102** (2), 024013, arXiv:1812.02790 [astro-ph.CO] .

[696] Fassnacht, C. D., T. J. Pearson, A. C. S. Readhead, I. W. A. Browne, L. V. E. Koopmans, S. T. Myers, and P. N. Wilkinson (1999), Astrophys. J. **527**, 498, arXiv:astro-ph/9907257 .

[697] Fay, S., S. Nesseris, and L. Perivolaropoulos (2007), Phys. Rev. D **76**, 063504, arXiv:gr-qc/0703006 .

[698] Feeney, S. M., D. J. Mortlock, and N. Dalmasso (2018), Mon. Not. Roy. Astron. Soc. **476** (3), 3861, arXiv:1707.00007 [astro-ph.CO] .

[699] Feldman, H. A., R. Watkins, and M. J. Hudson (2010), Mon. Not. Roy. Astron. Soc. **407**, 2328, arXiv:0911.5516 [astro-ph.CO] .

[700] Feng, B., M. Li, Y.-S. Piao, and X. Zhang (2006), Phys. Lett. B **634**, 101, arXiv:astro-ph/0407432 .

[701] Feng, L., R.-Y. Guo, J.-F. Zhang, and X. Zhang (2022), Phys. Lett. B **827**, 136940, arXiv:2109.06111 [astro-ph.CO] .

[702] Feng, L., J.-F. Zhang, and X. Zhang (2017), Eur. Phys. J. C **77** (6), 418, arXiv:1703.04884 [astro-ph.CO] .

[703] Feng, W.-F., H.-T. Wang, X.-C. Hu, Y.-M. Hu, and Y. Wang (2019), Phys. Rev. D **99** (12), 123002, arXiv:1901.02159 [astro-ph.IM] .

[704] Ferlito, F., S. Vagnozzi, D. F. Mota, and M. Baldi (2022), 10.1093/mnras/stac649, arXiv:2201.04528 [astro-ph.CO] .

[705] Fernández-Cobos, R., P. Vielva, D. Pietrobon, A. Balbi, E. Martínez-González, and R. B. Barreiro (2014), Mon. Not. Roy. Astron. Soc. **441** (3), 2392, arXiv:1312.0275 [astro-ph.CO] .

[706] Fernandez-Martinez, E., M. Pierre, E. Pinsard, and S. Rosauro-Alcaraz (2021), Eur. Phys. J. C **81** (10), 954, arXiv:2106.05298 [hep-ph] .

[707] Fernández Arenas, D., E. Terlevich, R. Terlevich, J. Melnick, R. Chávez, F. Bresolin, E. Telles, M. Plionis, and S. Basilakos (2018), Mon. Not. Roy. Astron. Soc. **474** (1), 1250, arXiv:1710.05951 [astro-ph.CO] .

[708] Ferraro, R., and F. Fiorini (2007), Phys. Rev. **D75**, 084031, arXiv:gr-qc/0610067 [gr-qc] .

[709] Ferreira, P. G., and M. Joyce (1997), Phys. Rev. Lett. **79**, 4740, arXiv:astro-ph/9707286 .

[710] Ferreira, P. G., and M. Joyce (1998), Phys. Rev. D **58**, 023503, arXiv:astro-ph/9711102 .

[711] Ferrero, I., M. G. Abadi, J. F. Navarro, L. V. Sales, and S. Gurovich (2012), Mon. Not. Roy. Astron. Soc. **425**, 2817, arXiv:1111.6609 [astro-ph.CO] .

[712] Fields, B. D. (2011), Ann. Rev. Nucl. Part. Sci. **61**, 47, arXiv:1203.3551 [astro-ph.CO] .

[713] Filippenko, A. V. (1997), Ann. Rev. Astron. Astrophys. **35**, 309.

[714] Filippenko, A. V. (2000), AIP Conf. Proc. **522** (1), 123, arXiv:astro-ph/0002264 .

[715] Fishbach, M., *et al.* (LIGO Scientific, Virgo) (2019), Astrophys. J. Lett. **871** (1), L13, arXiv:1807.05667 [astro-ph.CO] .

[716] Fixsen, D. J., E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, and E. L. Wright (1996), Astrophys. J. **473**, 576, arXiv:astro-ph/9605054 .

[717] Flambaum, V. V., and I. B. Samsonov (2019), Phys. Rev. D **100** (6), 063541, arXiv:1908.09432 [astro-ph.CO] .

[718] Flender, S., S. Hotchkiss, and S. Nadathur (2013), JCAP **02**, 013, arXiv:1212.0776 [astro-ph.CO] .

[719] Floratos, E. G., and G. K. Leontaris (1999), Phys. Lett. B **465**, 95, arXiv:hep-ph/9906238 .

[720] Flores, M. M., and A. Kusenko (2021), Phys. Rev. Lett. **126** (4), 041101, arXiv:2008.12456 [astro-ph.CO] .

[721] Flores, R. A., and J. R. Primack (1994), Astrophys. J. Lett. **427**, L1, arXiv:astro-ph/9402004 .

[135] Follin, B., and L. Knox (2018), Mon. Not. Roy. Astron. Soc. **477** (4), 4534, arXiv:1707.01175 [astro-ph.CO] .

[723] Fondi, E., A. Melchiorri, and L. Pagano (2022), arXiv:2203.12930 [astro-ph.CO] .

[724] Font-Ribera, A., *et al.* (BOSS) (2014), JCAP **05**, 027, arXiv:1311.1767 [astro-ph.CO] .

[725] Foot, R., and S. Vagnozzi (2015), Phys. Rev. D **91**, 023512, arXiv:1409.7174 [hep-ph] .

[726] Foot, R., and S. Vagnozzi (2016), JCAP **07**, 013, arXiv:1602.02467 [astro-ph.CO] .

[727] Forastieri, F., M. Lattanzi, and P. Natoli (2019), Phys. Rev. D **100** (10), 103526, arXiv:1904.07810 [astro-ph.CO] .

[728] Ford, L. (1987), Phys. Rev. D **35**, 2339.

[729] Fosalba, P., and E. Gaztañaga (2020), arXiv:2011.00910 [astro-ph.CO] .

[730] Fosalba, P., and E. Gaztanaga (2004), Mon. Not. Roy. Astron. Soc. **350**, L37, arXiv:astro-ph/0305468 .

[731] Foster, R. S., and D. C. Backer (1990), ApJ **361**, 300.

[732] Franchino-Viñas, S. A., and M. E. Mosquera (2021), arXiv:2107.02243 [astro-ph.CO] .

[733] Fraser, S., *et al.* (2018), Phys. Lett. B **785**, 159, arXiv:1803.03245 [hep-ph] .

[734] Freedman, W. L. (2021), Astrophys. J. **919** (1), 16, arXiv:2106.15656 [astro-ph.CO] .

[735] Freedman, W. L., B. F. Madore, T. Hoyt, I. S. Jang, R. Beaton, M. G. Lee, A. Monson, J. Neeley, and J. Rich (2020), arXiv:2002.01550 [astro-ph.GA] .

[736] Freedman, W. L., B. F. Madore, V. Scowcroft, C. Burns, A. Monson, S. E. Persson, M. Seibert, and J. Rigby (2012), Astrophys. J. **758**, 24, arXiv:1208.3281 [astro-ph.CO] .

[737] Freedman, W. L., *et al.* (HST) (2001), Astrophys. J. **553**, 47, arXiv:astro-ph/0012376 .

[738] Freedman, W. L., *et al.* (2019), 10.3847/1538-4357/ab2f73, arXiv:1907.05922 [astro-ph.CO] .

[739] Freeman, K. (1970), Astrophys. J. **160**, 811.

[740] Freeman, K. C. (1970), ApJ **160**, 811.

[741] Freese, K., and M. W. Winkler (2021), Phys. Rev. D **104** (8), 083533, arXiv:2102.13655 [astro-ph.CO] .

[742] Frolov, V. P., and A. Zelnikov (2016), Phys. Rev. D **93** (6), 064048, arXiv:1509.03336 [hep-th] .

[743] Froustey, J., C. Pitrou, and M. C. Volpe (2020), JCAP **12**, 015, arXiv:2008.01074 [hep-ph] .

[744] Frusciante, N., S. Peirone, L. Atayde, and A. De Felice (2020), Phys. Rev. D **101** (6), 064001, arXiv:1912.07586 [astro-ph.CO] .

[745] Fu, X., A. Bressan, P. Molaro, and P. Marigo (2015), Monthly Notices of the Royal Astronomical Society **452** (3), 3256, https://academic.oup.com/mnras/article-pdf/452/3/3256/4901646/stv1384.pdf .

[746] Fujii, Y. (1982), Phys. Rev. D **26**, 2580.

[747] Fujita, T., Y. Minami, K. Murai, and H. Nakatsuka (2021), Phys. Rev. D **103** (6), 063508, arXiv:2008.02473 [astro-ph.CO] .

[748] Fujita, T., K. Murai, H. Nakatsuka, and S. Tsujikawa (2021), Phys. Rev. D **103**, 043509, arXiv:2011.11894 [astro-ph.CO] .

[749] Fung, L. W., L. Li, T. Liu, H. N. Luu, Y.-C. Qiu, and S. H. H. Tye (2021), arXiv:2105.01631 [astro-ph.CO] .

[750] Fung, L. W. H., L. Li, T. Liu, H. N. Luu, Y.-C. Qiu, and S. H. H. Tye (2021), JCAP **08**, 057, arXiv:2102.11257 [hep-ph] .

[751] Gaia Collaboration,, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. Babusiaux, and M. Biermann (2020), "Gaia early data release 3: Summary of the contents and survey properties," arXiv:2012.01533 [astro-ph.GA] .

[752] Galitzki, N., *et al.* (2018), Proc. SPIE Int. Soc. Opt. Eng. **10708**, 1070804, arXiv:1808.04493 [astro-ph.IM] .

[753] Galli, S., M. Martinelli, A. Melchiorri, L. Pagano, B. D. Sherwin, and D. N. Spergel (2010), Phys. Rev. D **82**, 123504, arXiv:1005.3808 [astro-ph.CO] .

[754] Gannouji, R., L. Kazantzidis, L. Perivolaropoulos, and D. Polarski (2018), Phys. Rev. D **98** (10), 104044, arXiv:1809.07034 [gr-qc] .

[755] Gannouji, R., L. Perivolaropoulos, D. Polarski, and F. Skara (2021), Phys. Rev. D **103** (6), 063509, arXiv:2011.01517 [gr-qc] .

[756] Gannouji, R., D. Polarski, A. Ranquet, and A. A. Starobinsky (2006), JCAP **09**, 016, arXiv:astro-ph/0606287 .

[757] Gao, F., J. A. Braatz, M. J. Reid, K. Y. Lo, J. J. Condon, C. Henkel, C. Y. Kuo, C. M. V. Impellizzeri, D. W. Pesce, and W. Zhao (2016), Astrophys. J. **817** (2), 128, arXiv:1511.08311 [astro-ph.GA] .

[758] Gao, L.-Y., Z.-W. Zhao, S.-S. Xue, and X. Zhang (2021), JCAP **07**, 005, arXiv:2101.10714 [astro-ph.CO] .

[759] Garcia-Berro, E., E. Gaztanaga, J. Isern, O. Benvenuto, and L. Althaus (1999), arXiv:astro-ph/9907440 .

[760] Garcia-Cely, C., and X. Chu (2017), in *52nd Rencontres de Moriond on EW Interactions and Unified Theories*, pp. 307–314, arXiv:1705.06221 [hep-ph] .

[761] García-García, C., J. R. Zapatero, D. Alonso, E. Bellini, P. G. Ferreira, E.-M. Mueller, A. Nicola, and P. Ruiz-Lapuente (2021), JCAP **10**, 030, arXiv:2105.12108 [astro-ph.CO] .

[762] Gardner, J. P., *et al.* (2006), Space Sci. Rev. **123**, 485, arXiv:astro-ph/0606175 .

[763] Gariazzo, S., E. Di Valentino, O. Mena, and R. C. Nunes (2021), arXiv:2111.03152 [astro-ph.CO] .

[764] Garrison-Kimmel, S., M. Boylan-Kolchin, J. S. Bullock, and E. N. Kirby (2014), Mon. Not. Roy. Astron. Soc. **444** (1), 222, arXiv:1404.5313 [astro-ph.GA] .

[765] Gasperini, M., G. Marozzi, F. Nugier, and G. Veneziano (2011), JCAP **07**, 008, arXiv:1104.1167 [astro-ph.CO] .

[766] Gavela, M., D. Hernandez, L. Lopez Honorez, O. Mena, and S. Rigolin (2009), JCAP **07**, 034, [Erratum: JCAP 05, E01 (2010)], arXiv:0901.1611 [astro-ph.CO] .

[767] Gayathri, V., J. Healy, J. Lange, B. O'Brien, M. Szczepanczyk, I. Bartos, M. Campanelli, S. Klimenko, C. O. Lousto, and R. O'Shaughnessy (2021), Astrophys. J. Lett. **908** (2), L34, arXiv:2009.14247 [astro-ph.HE] .

[768] Gaztanaga, E. (2020), Mon. Not. Roy. Astron. Soc. **494** (2), 2766, arXiv:2003.11544 [physics.gen-ph] .

[769] Gaztanaga, E. (2021), Mon. Not. Roy. Astron. Soc. **502** (1), 436, arXiv:2101.07368 [gr-qc] .

[770] Gaztanaga, E. (2022), Symmetry **14** (2), 300, arXiv:2202.00641 [astro-ph.CO] .

[771] Gaztanaga, E., E. Garcia-Berro, J. Isern, E. Bravo, and I. Dominguez (2002), Phys. Rev. D **65**, 023506, arXiv:astro-ph/0109299 .

[772] Gaztanaga, E. d. (2021), "The Black Hole Universe (BHU) from a FLRW cloud," Submitted to MNRAS, see darkcosmos.com.

[773] Geha, M., M. R. Blanton, M. Masjedi, and A. A. West (2006), Astrophys. J. **653**, 240, arXiv:astro-ph/0608295 .

[774] Gehrels, T. (1995), *Hazards Due to Comets and Asteroids*.

[775] Gelmini, G. B., M. Kawasaki, A. Kusenko, K. Murai, and V. Takhistov (2020), JCAP **09**, 051, arXiv:2005.06721 [hep-ph] .

[776] Gelmini, G. B., A. Kusenko, and V. Takhistov (2021), JCAP **06**, 002, arXiv:1906.10136 [astro-ph.CO] .

[777] Geng, C.-Q., C.-C. Lee, and E. N. Saridakis (2012), JCAP **1201**, 002, arXiv:1110.0913 [astro-ph.CO] .

[778] Geng, C.-Q., C.-C. Lee, E. N. Saridakis, and Y.-P. Wu (2011), Phys. Lett. **B704**, 384, arXiv:1109.1092 [hep-th] .

[779] Genina, A., A. Benítez-Llambay, C. S. Frenk, S. Cole, A. Fattahi, J. F. Navarro, K. A. Oman, T. Sawala, and T. Theuns (2017), Monthly Notices of the Royal Astronomical Society **474** (1), 1398–1411.

[780] Gerardi, F., S. M. Feeney, and J. Alsing (2021), arXiv:2104.02728 [astro-ph.CO] .

[781] Gertsenshtein, M. E., and V. I. Pustovoit (1962), Sov. Phys. JETP **16**, 433.

[782] Ghirlanda, G., G. Ghisellini, and D. Lazzati (2004), Astrophys. J. **616**, 331, arXiv:astro-ph/0405602 .

[783] Ghonge, S., K. Chatziioannou, J. A. Clark, T. Littenberg, M. Millhouse, L. Cadonati, and N. Cornish (2020), Phys. Rev. D **102** (6), 064056, arXiv:2003.09456 [gr-qc] .

[784] Ghosh, S. (2014), Phys. Rev. D **89**, 063518, arXiv:1309.6547 [astro-ph.CO] .

[785] Ghosh, S., R. Khatri, and T. S. Roy (2020), Phys. Rev. D **102** (12), 123544, arXiv:1908.09843 [hep-ph] .

[786] Ghosh, S., S. Kumar, and Y. Tsai (2021), arXiv:2107.09076 [astro-ph.CO] .

[787] Giannantonio, T., R. Crittenden, R. Nichol, and A. J. Ross (2012), Mon. Not. Roy. Astron. Soc. **426**, 2581, arXiv:1209.2125 [astro-ph.CO] .

[788] Giannantonio, T., R. Scranton, R. G. Crittenden, R. C. Nichol, S. P. Boughn, A. D. Myers, and G. T. Richards (2008), Phys. Rev. D **77**, 123520, arXiv:0801.4380 [astro-ph] .

[789] Giarè, W., E. Di Valentino, A. Melchiorri, and O. Mena (2021), Mon. Not. Roy. Astron. Soc. **505** (2), 2703, arXiv:2011.14704 [astro-ph.CO] .

[790] Giazotto, A. (1990), Nucl. Instrum. Meth. A **289**, 518.

[791] Gibelyou, C., and D. Huterer (2012), Mon. Not. Roy. Astron. Soc. **427**, 1994, arXiv:1205.6476 [astro-ph.CO] .

[792] Gil-Marin, H., *et al.* (2020), Mon. Not. Roy. Astron. Soc. **498** (2), 2492, arXiv:2007.08994 [astro-ph.CO] .

[793] Gil-Marín, H., W. J. Percival, L. Verde, J. R. Brownstein, C.-H. Chuang, F.-S. Kitaura, S. A. Rodríguez-Torres, and M. D. Olmstead (2017), Mon. Not. Roy. Astron. Soc. **465** (2), 1757, arXiv:1606.00439 [astro-ph.CO] .

[794] Gómez-Valent, A., V. Pettorino, and L. Amendola (2020), Phys. Rev. D **101** (12), 123513, arXiv:2004.00610 [astro-ph.CO] .

[795] Gómez-Valent, A., and J. Solà (2017), EPL **120** (3), 39001, arXiv:1711.00692 [astro-ph.CO] .

[796] Gómez-Valent, A., and J. Solà Peracaula (2018), Mon. Not. Roy. Astron. Soc. **478** (1), 126, arXiv:1801.08501 [astro-ph.CO] .

[797] Gogoi, A., R. K. Sharma, P. Chanda, and S. Das (2021), Astrophys. J. **915** (2), 132, arXiv:2005.11889 [astro-ph.CO] .

[798] Gómez-Valent, A., and L. Amendola (2018), JCAP **04**, 051, arXiv:1802.01505 [astro-ph.CO] .

[799] Gómez-Valent, A., J. Solà, and S. Basilakos (2015), JCAP **01**, 004, arXiv:1409.7048 [astro-ph.CO] .

[800] Gómez-Valent, A., Z. Zheng, L. Amendola, V. Pettorino, and C. Wetterich (2021), Phys. Rev. D **104** (8), 083536, arXiv:2107.11065 [astro-ph.CO] .

[801] Gonzalez, J. E., H. H. B. Silva, R. Silva, and J. S. Alcaniz (2018), Eur. Phys. J. C **78** (9), 730, arXiv:1809.00439 [astro-ph.CO] .

[802] Gonzalez, M., M. P. Hertzberg, and F. Rompineve (2020), JCAP **10**, 028, arXiv:2006.13959 [astro-ph.CO] .

[803] Gonzalez-Espinoza, M., R. Herrera, G. Otalora, and J. Saavedra (2021), Eur. Phys. J. C **81** (8), 731, arXiv:2106.06145 [gr-qc] .

[804] Gonzalez-Espinoza, M., G. Otalora, J. Saavedra, and N. Videla (2018), Eur. Phys. J. C **78** (10), 799, arXiv:1808.01941 [gr-qc] .

[805] González-Morán, A. L., R. Chávez, R. Terlevich, E. Terlevich, F. Bresolin, D. Fernández-Arenas, M. Plionis, S. Basilakos, J. Melnick, and E. Telles (2019), Mon. Not. Roy. Astron. Soc. **487** (4), 4669, arXiv:1906.02195 [astro-ph.GA] .

[806] Gordon, C. (2007), Astrophys. J. **656**, 636, arXiv:astro-ph/0607423 .

[807] Gordon, C., W. Hu, D. Huterer, and T. M. Crawford (2005), Phys. Rev. D **72**, 103002, arXiv:astro-ph/0509301 .

[808] Gorenstein, M. V., E. E. Falco, and I. I. Shapiro (1988), ApJ **327**, 693.

[809] Gottloeber, S., E. L. Lokas, A. Klypin, and Y. Hoffman (2003), Mon. Not. Roy. Astron. Soc. **344**, 715, arXiv:astro-ph/0305393 .

[810] Goudelis, A., M. Pospelov, and J. Pradler (2016), Phys. Rev. Lett. **116** (21), 211303, arXiv:1510.08858 [hep-ph] .

[811] Gould, R., and G. Schréder (1966), Phys. Rev. Lett. **16** (6), 252.

[812] Graham, M., *et al.* (2020), Phys. Rev. Lett. **124** (25), 251102, arXiv:2006.14122 [astro-ph.HE] .

[813] Graham, M. J., *et al.* (2019), Publ. Astron. Soc. Pac. **131** (1001), 078001, arXiv:1902.01945 [astro-ph.IM] .

[814] Grande, J., and L. Perivolaropoulos (2011), Phys. Rev. D **84**, 023514, arXiv:1103.4143 [astro-ph.CO] .

[815] Granett, B. R., M. C. Neyrinck, and I. Szapudi (2008), Astrophys. J. Lett. **683**, L99, arXiv:0805.3695 [astro-ph] .

[816] Granett, B. R., M. C. Neyrinck, and I. Szapudi (2008), arXiv:0805.2974 [astro-ph] .

[817] Graur, O., F. B. Bianco, M. Modjaz, I. Shivvers, A. V. Filippenko, W. Li, and N. Smith (2017), Astrophys. J. **837** (2), 121, arXiv:1609.02923 [astro-ph.HE] .

[818] Green, D., *et al.* (2019), Bull. Am. Astron. Soc. **51** (7), 159, arXiv:1903.04763 [astro-ph.CO] .

[819] Grier, J. A., A. S. McEwen, P. G. Lucey, M. Milazzo, and R. G. Strom (2001), Journal of Geophysical Research: planets **106** (E12), 32847.

[820] Gross, E., and O. Vitells (2010), Eur. Phys. J. C **70**, 525, arXiv:1005.1891 [physics.data-an] .

[821] Gruppuso, A. (2014), Mon. Not. Roy. Astron. Soc. **437** (3), 2076, arXiv:1310.2822 [astro-ph.CO] .

[822] Gruppuso, A., F. Finelli, P. Natoli, F. Paci, P. Cabella, A. De Rosa, and N. Mandolesi (2011), Mon. Not. Roy. Astron. Soc. **411**, 1445, arXiv:1006.1979 [astro-ph.CO] .

[823] Gruppuso, A., N. Kitazawa, M. Lattanzi, N. Mandolesi, P. Natoli, and A. Sagnotti (2018), Phys. Dark Univ. **20**, 49, arXiv:1712.03288 [astro-ph.CO] .

[824] Gu, Y., L. Wu, and B. Zhu (2021), arXiv:2105.07232 [hep-ph] .

[825] Gubser, S. S., and J. Khoury (2004), Phys. Rev. D **70**, 104001, arXiv:hep-ph/0405231 .

[826] Guo, R.-Y., L. Feng, T.-Y. Yao, and X.-Y. Chen (2021), JCAP **12** (12), 036, arXiv:2110.02536 [gr-qc] .

[827] Guo, R.-Y., J.-F. Zhang, and X. Zhang (2019), JCAP **1902**, 054, arXiv:1809.02340 [astro-ph.CO] .

[828] Guo, R.-Y., and X. Zhang (2017), Eur. Phys. J. C **77** (12), 882, arXiv:1704.04784 [astro-ph.CO] .

[829] Gurovich, S., S. S. McGaugh, K. C. Freeman, H. Jerjen, L. Staveley-Smith, and W. J. G. De Blok (2004), Publ. Astron. Soc. Austral. **21** (4), 412, arXiv:astro-ph/0411521 .

[830] Gurzadyan, V. G., and A. Stepanian (2019), Eur. Phys. J. C **79** (7), 568, arXiv:1905.03442 [astro-ph.CO] .

[831] Gurzadyan, V. G., and A. Stepanian (2021), Eur. Phys. J. Plus **136** (2), 235, arXiv:2102.10100 [gr-qc] .

[119] Guth, A. H. (1981), Phys. Rev. **D23**, 347.

[833] Guzzetti, M. C., N. Bartolo, M. Liguori, and S. Matarrese (2016), Riv. Nuovo Cim. **39** (9), 399, arXiv:1605.01615 [astro-ph.CO] .

[834] Guzzo, L., *et al.* (2008), Nature **451**, 541, arXiv:0802.1944 [astro-ph] .

[95] de Haan, T., *et al.* (SPT) (2016), Astrophys. J. **832** (1), 95, arXiv:1603.06522 [astro-ph.CO] .

[836] Haiman, Z., A. A. Thoul, and A. Loeb (1996), Astrophys. J. **464**, 523, arXiv:astro-ph/9507111 .

[837] Hajian, A. (2007), arXiv:astro-ph/0702723 .

[838] Hall, A. (2021), Mon. Not. Roy. Astron. Soc. **505** (4), 4935, arXiv:2104.12880 [astro-ph.CO] .

[158] Hamana, T., *et al.* (2020), Publ. Astron. Soc. Jap. **72** (1), Publications of the Astronomical Society of Japan, Volume 72, Issue 1, February 2020, 16, https://doi.org/10.1093/pasj/psz138, arXiv:1906.06041 [astro-ph.CO] .

[840] Hamilton, A. J. S. (1992), Astrophys. J. Lett. **385**, L5.

[841] Hamilton, A. J. S. (1997), in *Ringberg Workshop on Large Scale Structure*, arXiv:astro-ph/9708102 .

[842] Hammache, F., *et al.* (2013), Phys. Rev. C **88** (6), 062802, arXiv:1312.0894 [nucl-ex] .

[843] Hamuy, M., and P. A. Pinto (2002), Astrophys. J. Lett. **566**, L63, arXiv:astro-ph/0201279 .

[844] Handley, W. (2019), arXiv:1908.09139 [astro-ph.CO] .

[845] Hannestad, S., R. S. Hansen, and T. Tram (2014), Phys. Rev. Lett. **112** (3), 031802, arXiv:1310.5926 [astro-ph.CO] .

[846] Hansen, F. K., A. Banday, and K. Gorski (2004), Mon. Not. Roy. Astron. Soc. **354**, 641, arXiv:astro-ph/0404206 .

[847] Hansen, M., W. Zhao, A. M. Frejsel, P. D. Naselsky, J. Kim, and O. V. Verkhodanov (2012), Mon. Not. Roy. Astron. Soc. **426**, 57, arXiv:1202.1711 [astro-ph.CO] .

[848] Harari, D., and P. Sikivie (1992), Phys. Lett. B **289**, 67.

[849] Haridasu, B. S., V. V. Lukovi\'c, M. Moresco, and N. Vittorio (2018), JCAP **10**, 015, arXiv:1805.03595 [astro-ph.CO] .

[850] Haridasu, B. S., and M. Viel (2020), Mon. Not. Roy. Astron. Soc. **497** (2), 1757, arXiv:2004.07709 [astro-ph.CO] .

[851] Haridasu, B. S., M. Viel, and N. Vittorio (2021), Phys. Rev. D **103** (6), 063539, arXiv:2012.10324 [astro-ph.CO] .

[852] Harry, G. M. (LIGO Scientific) (2010), Class. Quant. Grav. **27**, 084006.

[853] Harry, G. M., P. Fritschel, D. A. Shaddock, W. Folkner, and E. S. Phinney (2006), Class. Quant. Grav. **23**, 4887, [Erratum: Class.Quant.Grav. 23, 7361 (2006)].

[854] Hart, L., and J. Chluba (2018), Mon. Not. Roy. Astron. Soc. **474** (2), 1850, arXiv:1705.03925 [astro-ph.CO] .

[855] Hart, L., and J. Chluba (2020), Mon. Not. Roy. Astron. Soc. **493** (3), 3255, arXiv:1912.03986 [astro-ph.CO] .

[856] Hart, L., and J. Chluba (2022), Mon. Not. Roy. Astron. Soc. **510** (2), 2206, arXiv:2107.12465 [astro-ph.CO] .

[857] Hartwig, T., V. Bromm, and A. Loeb (2018), Mon. Not. Roy. Astron. Soc. **479** (2), 2202, arXiv:1711.05742 [astro-ph.GA] .

[858] Hashim, M., W. El Hanafy, A. Golovnev, and A. A. El-Zant (2021), JCAP **07**, 052, arXiv:2010.14964 [astro-ph.CO] .

[859] Haslbauer, M., I. Banik, and P. Kroupa (2020), Mon. Not. Roy. Astron. Soc. **499** (2), 2845, arXiv:2009.11292 [astro-ph.CO] .

[860] Hasselfield, M., *et al.* (2013), JCAP **07**, 008, arXiv:1301.0816 [astro-ph.CO] .

[861] Hauser, M. G., and E. Dwek (2001), Ann. Rev. Astron. Astrophys. **39**, 249, arXiv:astro-ph/0105539 .

[862] Hawkins, E., *et al.* (2003), Mon. Not. Roy. Astron. Soc. **346**, 78, arXiv:astro-ph/0212375 .

[863] Hayakawa, S., *et al.* (2020), JPS Conf. Proc. **31**, 011036.

[864] Hazra, D. K., A. Shafieloo, and T. Souradeep (2019), JCAP **04**, 036, arXiv:1810.08101 [astro-ph.CO] .

[865] He, H.-J., Y.-Z. Ma, and J. Zheng (2020), JCAP **11**, 003, arXiv:2003.12057 [hep-ph] .

[866] He, J.-H., and B. Wang (2008), JCAP **06**, 010, arXiv:0801.4233 [astro-ph] .

[867] He, J.-H., B. Wang, and E. Abdalla (2009), Phys. Lett. B **671**, 139, arXiv:0807.3471 [gr-qc] .

[868] Heinesen, A., and T. Buchert (2020), Class. Quant. Grav. **37** (16), 164001, [Erratum: Class.Quant.Grav. 37, 229601 (2020)], arXiv:2002.10831 [gr-qc] .

[869] Heisenberg, L., and H. Villarrubia-Rojo (2021), JCAP **03**, 032, arXiv:2010.00513 [astro-ph.CO] .

[870] Heisenberg, L., H. Villarrubia-Rojo, and J. Zosso (2022), arXiv:2202.01202 [astro-ph.CO] .

[871] Heisenberg, L., H. Villarrubia-Rojo, and J. Zosso (2022), arXiv:2201.11623 [astro-ph.CO] .

[872] Heisenberg, W. (1927), Z. Phys. **43**, 172.

[873] Helayël-Neto, J. A., and A. D. A. M. Spallicci (2019), Eur. Phys. J. C **79** (7), 590, arXiv:1904.11035 [hep-ph] .

[874] Hellings, R. w., and G. s. Downs (1983), Astrophys. J. Lett. **265**, L39.

[875] Heng, I. S., D. G. Blair, E. N. Ivanov, and M. E. Tobar (1996), Phys. Lett. A **218**, 190.

[876] Henning, J. W., *et al.* (SPT) (2018), Astrophys. J. **852** (2), 97, arXiv:1707.09353 [astro-ph.CO] .

[877] Hernández-Almada, A., G. Leon, J. Magaña, M. A. García-Aspeitia, V. Motta, E. N. Saridakis, and K. Yesmakhanova (2021), 10.1093/mnras/stac255, arXiv:2111.00558 [astro-ph.CO] .

[878] Hernández-Almada, A., G. Leon, J. Magaña, M. A. García-Aspeitia, and V. Motta (2020), 10.1093/mnras/staa2052, arXiv:2002.12881 [astro-ph.CO] .

[879] Herold, L., E. G. M. Ferreira, and E. Komatsu (2021), arXiv:2112.12140 [astro-ph.CO] .

[880] Herrnstein, J. R., J. M. Moran, L. J. Greenhill, P. J. Diamond, M. Inoue, N. Nakai, M. Miyoshi, C. Henkel, and A. Riess (1999), Nature **400**, 539, arXiv:astro-ph/9907013 .

[171] Heymans, C., *et al.* (2012), Mon. Not. Roy. Astron. Soc. **427**, 146, arXiv:1210.0032 [astro-ph.CO] .

[882] Heymans, C., *et al.* (2013), Mon. Not. Roy. Astron. Soc. **432**, 2433, arXiv:1303.1808 [astro-ph.CO] .

[883] Heymans, C., *et al.* (2021), Astron. Astrophys. **646**, A140, arXiv:2007.15632 [astro-ph.CO] .

[172] Hikage, C., *et al.* (HSC) (2019), Publ. Astron. Soc. Jap. **71** (2), 43, arXiv:1809.09148 [astro-ph.CO] .

[885] Hild, S., S. Chelkowski, A. Freise, J. Franc, N. Morgado, R. Flaminio, and R. DeSalvo (2010), Class. Quant. Grav. **27**, 015003, arXiv:0906.2655 [gr-qc] .

[886] Hild, S., *et al.* (2011), Class. Quant. Grav. **28**, 094013, arXiv:1012.0908 [gr-qc] .

[887] Hildebrandt, H., *et al.* (2012), Mon. Not. Roy. Astron. Soc. **421**, 2355, arXiv:1111.4434 [astro-ph.CO] .

[174] Hildebrandt, H., *et al.* (2017), Mon. Not. Roy. Astron. Soc. **465**, 1454, arXiv:1606.05338 [astro-ph.CO] .

[175] Hildebrandt, H., *et al.* (2020), Astron. Astrophys. **633**, A69, arXiv:1812.06076 [astro-ph.CO] .

[890] Hill, J. C., and E. J. Baxter (2018), JCAP **08**, 037, arXiv:1803.07555 [astro-ph.CO] .

[891] Hill, J. C., E. McDonough, M. W. Toomey, and S. Alexander (2020), Phys. Rev. D **102** (4), 043507, arXiv:2003.07355 [astro-ph.CO] .

[892] Hill, J. C., *et al.* (2021), arXiv:2109.04451 [astro-ph.CO] .

[893] Hinrichsen, H., and A. Kempf (1996), J. Math. Phys. **37**, 2121, arXiv:hep-th/9510144 [hep-th] .

[894] Hinshaw, G., A. J. Banday, C. L. Bennett, K. M. Gorski, A. Kogut, C. H. Lineweaver, G. F. Smoot, and E. L. Wright (1996), Astrophys. J. Lett. **464**, L25, arXiv:astro-ph/9601061 .

[895] Hinshaw, G., D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright (2013), The Astrophysical Journal Supplement **208** (2), 19, arXiv:1212.5226 [astro-ph.CO] .

[896] Hinterbichler, K., and J. Khoury (2010), Phys. Rev. Lett. **104**, 231301, arXiv:1001.4525 [hep-th] .

[897] Hinterbichler, K., J. Khoury, A. Levy, and A. Matas (2011), Phys. Rev. D **84**, 103521, arXiv:1107.2112 [astro-ph.CO] .

[898] Hirano, S., T. Hosokawa, N. Yoshida, K. Omukai, and H. W. Yorke (2015), Mon. Not. Roy. Astron. Soc. **448** (1), 568, arXiv:1501.01630 [astro-ph.GA] .

[899] Hivon, E., K. M. Gorski, C. B. Netterfield, B. P. Crill, S. Prunet, and F. Hansen (2002), Astrophys. J. **567**, 2, arXiv:astro-ph/0105302 .

[900] Hobbs, G. (2013), Class. Quant. Grav. **30**, 224007, arXiv:1307.2629 [astro-ph.IM] .

[901] Hobson, M. P., G. P. Efstathiou, and A. N. Lasenby (2006), *General relativity: An introduction for physicists*.

[902] Hofmann, F., J. S. Sanders, N. Clerc, K. Nandra, J. Ridl, K. Dennerl, M. Ramos-Ceja, A. Finoguenov, and T. H. Reiprich (2017), Astron. Astrophys. **606**, A118, arXiv:1708.05205 [astro-ph.CO] .

[903] Hoftuft, J., H. Eriksen, A. Banday, K. Gorski, F. Hansen, and P. Lilje (2009), Astrophys. J. **699**, 985, arXiv:0903.1229 [astro-ph.CO] .

[904] Hogg, N. B., and M. Bruni (2021), 10.1093/mnras/stac324, arXiv:2109.08676 [astro-ph.CO] .

[905] Holsclaw, T., U. Alam, B. Sanso, H. Lee, K. Heitmann, S. Habib, and D. Higdon (2010), Phys. Rev. Lett. **105**, 241302, arXiv:1011.3079 [astro-ph.CO] .

[906] Holsclaw, T., U. Alam, B. Sanso, H. Lee, K. Heitmann, S. Habib, and D. Higdon (2010), Phys. Rev. D **82**, 103502, arXiv:1009.5443 [astro-ph.CO] .

[180] Holz, D. E., and S. A. Hughes (2005), Astrophys. J. **629**, 15, arXiv:astro-ph/0504616 .

[908] Hong, T., C. M. Springob, L. Staveley-Smith, M. I. Scrimgeour, K. L. Masters, L. M. Macri, B. S. Koribalski, D. H. Jones, and T. H. Jarrett (2014), Mon. Not. Roy. Astron. Soc. **445** (1), 402, arXiv:1409.0287 [astro-ph.CO] .

[909] Horndeski, G. W. (1974), Int. J. Theor. Phys. **10**, 363.

[910] Hotokezaka, K., E. Nakar, O. Gottlieb, S. Nissanke, K. Masuda, G. Hallinan, K. P. Mooley, and A. T. Deller (2019), Nature Astron. **3** (10), 940, arXiv:1806.10596 [astro-ph.CO] .

[911] Hou, J., *et al.* (2020), Mon. Not. Roy. Astron. Soc. **500** (1), 1201, arXiv:2007.08998 [astro-ph.CO] .

[912] Hou, S. Q., J. J. He, A. Parikh, D. Kahl, C. A. Bertulani, T. Kajino, G. J. Mathews, and G. Zhao (2017), Astrophys. J. **834** (2), 165, arXiv:1701.04149 [astro-ph.CO] .

[913] Hou, Z., R. Keisler, L. Knox, M. Millea, and C. Reichardt (2013), Phys. Rev. D **87**, 083008, arXiv:1104.2333 [astro-ph.CO] .

[182] Howlett, C., A. Ross, L. Samushia, W. Percival, and M. Manera (2015), Mon. Not. Roy. Astron. Soc. **449** (1), 848, arXiv:1409.3238 [astro-ph.CO] .

[915] Hoyt, T. J. (2021), arXiv:2106.13337 [astro-ph.GA] .

[916] Hryczuk, A., and K. Jodlowski (2020), Phys. Rev. D **102** (4), 043024, arXiv:2006.16139 [hep-ph] .

[917] Hu, J., *et al.* (2019), Mon. Not. Roy. Astron. Soc. **485** (4), 5050, arXiv:1812.11480 [astro-ph.SR] .

[918] Hu, J. P., F. Y. Wang, and Z. G. Dai (2021), Mon. Not. Roy. Astron. Soc. **507** (1), 730, arXiv:2107.12718 [astro-ph.CO] .

[919] Hu, W. (2001), Phys. Rev. D **64**, 083005, arXiv:astro-ph/0105117 .

[920] Hu, W., R. Barkana, and A. Gruzinov (2000), Phys. Rev. Lett. **85**, 1158, arXiv:astro-ph/0003365 .

[921] Hu, W., and S. Dodelson (2002), Ann. Rev. Astron. Astrophys. **40**, 171, arXiv:astro-ph/0110414 .

[922] Hu, W.-R., and Y.-L. Wu (2017), Natl. Sci. Rev. **4** (5), 685.

[923] Huang, C. D., A. G. Riess, W. Yuan, L. M. Macri, N. L. Zakamska, S. Casertano, P. A. Whitelock, S. L. Hoffmann, A. V. Filippenko, and D. Scolnic (2019), 10.3847/1538-4357/ab5dbd, arXiv:1908.10883 [astro-ph.CO] .

[924] Huang, C. D., *et al.* (2018), Astrophys. J. **857** (1), 67, arXiv:1801.02711 [astro-ph.CO] .

[925] Huang, G.-y., and W. Rodejohann (2021), Phys. Rev. D **103**, 123007, arXiv:2102.04280 [hep-ph] .

[926] Huang, Q.-G., and K. Wang (2016), Eur. Phys. J. C **76** (9), 506, arXiv:1606.05965 [astro-ph.CO] .

[927] Hudson, M. J., S. D. J. Gwyn, H. Dahle, and N. Kaiser (1998), Astrophys. J. **503**, 531, arXiv:astro-ph/9711341 .

[928] Huey, G., and B. D. Wandelt (2006), Phys. Rev. D **74**, 023519, arXiv:astro-ph/0407196 .

[929] Hui, L., A. Nicolis, and C. Stubbs (2009), Phys. Rev. D **80**, 104002, arXiv:0905.2966 [astro-ph.CO] .

[930] Hui, L., J. P. Ostriker, S. Tremaine, and E. Witten (2017), Phys. Rev. D **95** (4), 043541, arXiv:1610.08297 [astro-ph.CO] .

[931] Humphreys, E. M. L., M. J. Reid, J. M. Moran, L. J. Greenhill, and A. L. Argon (2013), Astrophys. J. **775**, 13, arXiv:1307.6031 [astro-ph.CO] .

[932] Hunt, P., and S. Sarkar (2010), Mon. Not. Roy. Astron. Soc. **401**, 547, arXiv:0807.4508 [astro-ph] .

[933] Huterer, D., and D. L. Shafer (2018), Rept. Prog. Phys. **81** (1), 016901, arXiv:1709.01091 [astro-ph.CO] .

[934] Ibata, N. G., R. A. Ibata, B. Famaey, and G. F. Lewis (2014), Nature **511**, 563, arXiv:1407.8178 [astro-ph.GA] .

[935] Ibata, R. A., *et al.* (2013), Nature **493**, 62, arXiv:1301.0446 [astro-ph.CO] .

[936] Iben, I., and A. Renzini (1983), Ann. Rev. Astron. Astrophys. **21**, 271.

[937] Ichiki, K., M. Oguri, and K. Takahashi (2004), Phys. Rev. Lett. **93**, 071302, arXiv:astro-ph/0403164 .

[938] Iliadis, C., and A. Coc (2020), Astrophys. J. **901** (2), 127, arXiv:2008.12200 [astro-ph.CO] .

[939] Ilic, S., M. Langer, and M. Douspis (2013), Astron. Astrophys. **556**, A51, arXiv:1301.5849 [astro-ph.CO] .

[940] Inoue, K. T., and J. Silk (2006), Astrophys. J. **648**, 23, arXiv:astro-ph/0602478 .

[941] Inoue, K. T., and J. Silk (2007), Astrophys. J. **664**, 650, arXiv:astro-ph/0612347 .

[942] Intema, H. T., P. Jagannathan, K. P. Mooley, and D. A. Frail (2017), Astron. Astrophys. **598**, A78, arXiv:1603.04368 [astro-ph.CO] .

[943] Iocco, F. (2012), Mem. Soc. Astron. Ital. Suppl. **22**, 19, arXiv:1206.2396 [astro-ph.GA] .

[944] Iocco, F., G. Mangano, G. Miele, O. Pisanti, and P. D. Serpico (2009), Phys. Rept. **472**, 1, arXiv:0809.0631 [astro-ph] .

[945] Iorio, G., F. Fraternali, C. Nipoti, E. Di Teodoro, J. I. Read, and G. Battaglia (2016), Monthly Notices of the Royal Astronomical Society , stw3285.

[946] Irvsivc, V., M. Viel, M. G. Haehnelt, J. S. Bolton, and G. D. Becker (2017), Phys. Rev. Lett. **119** (3), 031302, arXiv:1703.04683 [astro-ph.CO] .

[947] Ishak, M. (2019), Living Rev. Rel. **22** (1), 1, arXiv:1806.10122 [astro-ph.CO] .

[948] Ishikawa, S., *et al.* (2020), JPS Conf. Proc. **31**, 011037.

[949] Itoh, Y., K. Yahata, and M. Takada (2010), Phys. Rev. D **82**, 043530, arXiv:0912.1460 [astro-ph.CO] .

[950] Ivanov, M. M. (2021), Phys. Rev. D **104** (10), 103514, arXiv:2106.12580 [astro-ph.CO] .

[951] Ivanov, M. M., Y. Ali-Haïmoud, and J. Lesgourgues (2020), Phys. Rev. D **102** (6), 063515, arXiv:2005.10656 [astro-ph.CO] .

[952] Ivanov, M. M., E. McDonough, J. C. Hill, M. Simonović, M. W. Toomey, S. Alexander, and M. Zaldarriaga (2020), Phys. Rev. D **102** (10), 103502, arXiv:2006.11235 [astro-ph.CO] .

[953] Ivanov, M. M., M. Simonović, and M. Zaldarriaga (2020), JCAP **05**, 042, arXiv:1909.05277 [astro-ph.CO] .

[954] Ivezić, v. Z., *et al.* (LSST) (2019), Astrophys. J. **873** (2), 111, arXiv:0805.2366 [astro-ph] .

[955] Jackson, J. D. (1998), *Classical Electrodynamics* (Wiley).

[956] Jaeckel, J., and W. Yin (2021), JCAP **02**, 044, arXiv:2007.15006 [hep-ph] .

[957] Jaeckel, J., and W. Yin (2021), Phys. Rev. D **103** (11), 115019, arXiv:2102.00006 [hep-ph] .

[958] de Jaeger, T., L. Galbany, A. G. Riess, B. E. Stahl, B. J. Shappee, A. V. Filippenko, and W. Zheng (2022), arXiv:2203.08974 [astro-ph.CO] .

[959] de Jaeger, T., B. Stahl, W. Zheng, A. Filippenko, A. Riess, and L. Galbany (2020), 10.1093/mnras/staa1801, arXiv:2006.03412 [astro-ph.CO] .

[960] de Jaeger, T., *et al.* (2015), Astrophys. J. **815**, 121, arXiv:1511.05145 [astro-ph.HE] .

[961] de Jaeger, T., *et al.* (2017), Mon. Not. Roy. Astron. Soc. **472** (4), 4233, arXiv:1709.01513 [astro-ph.HE] .

[962] de Jaeger, T., *et al.* (DES) (2020), 10.1093/mnras/staa1402, arXiv:2005.09757 [astro-ph.HE] .

[963] Jaffe, T. R., A. J. Banday, H. K. Eriksen, K. M. Gorski, and F. K. Hansen (2005), Astrophys. J. Lett. **629**, L1, arXiv:astro-ph/0503213 .

[964] Jain, B., and J. Khoury (2010), Annals Phys. **325**, 1479, arXiv:1004.3294 [astro-ph.CO] .

[965] Janssen, G., *et al.* (2015), PoS **AASKA14**, 037, arXiv:1501.00127 [astro-ph.IM] .

[966] Jaranowski, P., A. Krolak, K. D. Kokkotas, and G. Tsegas (1996), Class. Quant. Grav. **13**, 1279.

[967] Jarvis, M., *et al.* (DES) (2016), Mon. Not. Roy. Astron. Soc. **460** (2), 2245, arXiv:1507.05603 [astro-ph.IM] .

[968] Jassal, H. K., J. S. Bagla, and T. Padmanabhan (2005), Phys. Rev. D **72**, 103503, arXiv:astro-ph/0506748 .

[969] Jazayeri, S., A. V. Sadr, and H. Firouzjahi (2017), Phys. Rev. D **96** (2), 023512, arXiv:1703.05714 [astro-ph.CO] .

[970] Jedamzik, K., and L. Pogosian (2020), Phys. Rev. Lett. **125** (18), 181302, arXiv:2004.09487 [astro-ph.CO] .

[971] Jedamzik, K., L. Pogosian, and G.-B. Zhao (2021), Commun. in Phys. **4**, 123, arXiv:2010.04158 [astro-ph.CO] .

[972] Jedamzik, K., and A. Saveliev (2019), Phys. Rev. Lett. **123** (2), 021301, arXiv:1804.06115 [astro-ph.CO] .

[973] Jenkins, A., C. Frenk, S. D. White, J. Colberg, S. Cole, A. E. Evrard, H. Couchman, and N. Yoshida (2001), Mon. Not. Roy. Astron. Soc. **321**, 372, arXiv:astro-ph/0005260 .

[974] Jensen, J. B., J. L. Tonry, R. I. Thompson, E. A. Ajhar, T. R. Lauer, M. J. Rieke, M. Postman, and M. C. Liu (2001), Astrophys. J. **550**, 503, arXiv:astro-ph/0011288 .

[975] Jiang, J.-Q., and Y.-S. Piao (2021), Phys. Rev. D **104** (10), 103524, arXiv:2107.07128 [astro-ph.CO] .

[976] Jiang, J.-Q., and Y.-S. Piao (2022), arXiv:2202.13379 [astro-ph.CO] .

[977] Jiménez, J. B., D. Bettoni, and P. Brax (2021), Phys. Rev. D **103** (10), 103505, arXiv:2004.13677 [astro-ph.CO] .

[978] Jiménez, J. B., D. Bettoni, D. Figueruelo, and F. A. Teppa Pannia (2020), JCAP **08**, 020, arXiv:2004.14661 [astro-ph.CO] .

[979] Jimenez, R., A. Cimatti, L. Verde, M. Moresco, and B. Wandelt (2019), JCAP **03**, 043, arXiv:1902.07081 [astro-ph.CO] .

[980] Jimenez, R., and A. Loeb (2002), Astrophys. J. **573**, 37, arXiv:astro-ph/0106145 .

[981] Jiménez Cruz, N. M., and C. Escamilla-Rivera (2021), Eur. Phys. J. Plus **136** (1), 51, arXiv:2011.09623 [gr-qc] .

[982] Johnson, A., *et al.* (2014), Mon. Not. Roy. Astron. Soc. **444** (4), 3926, arXiv:1404.3799 [astro-ph.CO] .

[983] Johnson, J. L. (2010), Mon. Not. Roy. Astron. Soc. **404**, 1425, arXiv:0911.1294 [astro-ph.CO] .

[984] Johnson, J. P., A. Sangwan, and S. Shankaranarayanan (2022), JCAP **01** (01), 024, arXiv:2102.12367 [astro-ph.CO] .

[985] Jones, D. H., *et al.* (2009), Mon. Not. Roy. Astron. Soc. **399**, 683, arXiv:0903.5451 [astro-ph.CO] .

[986] Jones, D. O., *et al.* (2017), Astrophys. J. **843** (1), 6, arXiv:1611.07042 [astro-ph.CO] .

[987] Jones, D. O., *et al.* (2018), Astrophys. J. **867** (2), 108, arXiv:1805.05911 [astro-ph.CO] .

[988] Jones, D. O., *et al.* (2022), arXiv:2201.07801 [astro-ph.CO] .

[989] de Jong, J. T. A., G. A. Verdoes Kleijn, K. H. Kuijken, and E. A. Valentijn (Astro-WISE, KiDS) (2013), Exper. Astron. **35**, 25, arXiv:1206.1254 [astro-ph.CO] .

[990] de Jong, J. T. A., *et al.* (2015), Astron. Astrophys. **582**, A62, arXiv:1507.00742 [astro-ph.CO] .

[991] de Jong, J. T. A., *et al.* (2017), Astron. Astrophys. **604**, A134, arXiv:1703.02991 [astro-ph.GA] .

[992] Joudaki, S., M. Kaplinghat, R. Keeley, and D. Kirkby (2018), Phys. Rev. D **97** (12), 123501, arXiv:1710.04236 [astro-ph.CO] .

[993] Joudaki, S., *et al.* (2017), Mon. Not. Roy. Astron. Soc. **465** (2), 2033, arXiv:1601.05786 [astro-ph.CO] .

[994] Joudaki, S., *et al.* (2017), Mon. Not. Roy. Astron. Soc. **471** (2), 1259, arXiv:1610.04606 [astro-ph.CO] .

[995] Joudaki, S., *et al.* (2018), Mon. Not. Roy. Astron. Soc. **474** (4), 4894, arXiv:1707.06627 [astro-ph.CO] .

[195] Joudaki, S., *et al.* (2020), Astron. Astrophys. **638**, L1, arXiv:1906.09262 [astro-ph.CO] .

[997] Joyce, A., B. Jain, J. Khoury, and M. Trodden (2015), Phys. Rept. **568**, 1, arXiv:1407.0059 [astro-ph.CO] .

[998] Kable, J. A., G. E. Addison, and C. L. Bennett (2020), Astrophys. J. **905** (2), 164, arXiv:2008.01785 [astro-ph.CO] .

[999] Kable, J. A., G. Benevento, N. Frusciante, A. De Felice, and S. Tsujikawa (2021), arXiv:2111.10432 [astro-ph.CO] .

[1000] Kaiser, N. (1987), Mon. Not. Roy. Astron. Soc. **227**, 1.

[1001] Kaiser, N., G. Wilson, and G. A. Luppino (2000), arXiv:astro-ph/0003338 .

[1002] Kaloper, N. (2019), Int. J. Mod. Phys. D **28** (14), 1944017, arXiv:1903.11676 [hep-th] .

[197] Kamionkowski, M., A. Kosowsky, and A. Stebbins (1997), Phys. Rev. D **55**, 7368, arXiv:astro-ph/9611125 .

[1004] Kamionkowski, M., J. Pradler, and D. G. E. Walker (2014), Phys. Rev. Lett. **113** (25), 251302, arXiv:1409.0549 [hep-ph] .

[1005] Kanekar, N., G. I. Langston, J. T. Stocke, C. L. Carilli, and K. M. Menten (2012), Astrophys. J. Lett. **746**, L16, arXiv:1201.3372 [astro-ph.CO] .

[1006] Kang, Y., Y.-W. Lee, Y.-L. Kim, C. Chung, and C. H. Ree (2020), Astrophys. J. **889** (1), 8, arXiv:1912.04903 [astro-ph.GA] .

[1007] Kaplinghat, M., M. Valli, and H.-B. Yu (2019), Mon. Not. Roy. Astron. Soc. **490** (1), 231, arXiv:1904.04939 [astro-ph.GA] .

[1008] Kapner, D. J., T. S. Cook, E. G. Adelberger, J. H. Gundlach, B. R. Heckel, C. D. Hoyle, and H. E. Swanson (2007), Phys. Rev. Lett. **98**, 021101, arXiv:hep-ph/0611184 .

[1009] Karachentsev, I. D., L. N. Makarova, R. B. Tully, L. Rizzi, and E. J. Shaya (2018), ApJ **858** (1), 62, arXiv:1804.00469 [astro-ph.GA] .

[1010] Karolyhazy, F. (1966), Nuovo Cim. **A42**, 390.

[1011] Karwal, T., and M. Kamionkowski (2016), Phys. Rev. D **94** (10), 103523, arXiv:1608.01309 [astro-ph.CO] .

[1012] Karwal, T., M. Raveri, B. Jain, J. Khoury, and M. Trodden (2022), Phys. Rev. D **105** (6), 063535, arXiv:2106.13290 [astro-ph.CO] .

[1013] Kasai, M., and T. Futamase (2019), PTEP **2019** (7), 073E01, arXiv:1904.09689 [gr-qc] .

[1014] Kase, R., and S. Tsujikawa (2019), Int. J. Mod. Phys. D **28** (05), 1942005, arXiv:1809.08735 [gr-qc] .

[1015] Kashlinsky, A., F. Atrio-Barandela, and H. Ebeling (2012), arXiv:1202.0717 [astro-ph.CO] .

[1016] Kashlinsky, A., F. Atrio-Barandela, H. Ebeling, A. Edge, and D. Kocevski (2010), Astrophys. J. Lett. **712**, L81, arXiv:0910.4958 [astro-ph.CO] .

[1017] Kashlinsky, A., F. Atrio-Barandela, D. Kocevski, and H. Ebeling (2009), Astrophys. J. Lett. **686**, L49, arXiv:0809.3734 [astro-ph] .

[1018] Kauffmann, G., S. D. M. White, and B. Guiderdoni (1993), Mon. Not. Roy. Astron. Soc. **264**, 201.

[1019] Kawamura, S., *et al.* (2006), Class. Quant. Grav. **23**, S125.

[1020] Kawamura, S., *et al.* (2008), J. Phys. Conf. Ser. **122**, 012006.

[1021] Kawamura, S., *et al.* (2008), J. Phys. Conf. Ser. **120**, 032004.

[1022] Kawamura, S., *et al.* (2011), Class. Quant. Grav. **28**, 094011.

[1023] Kawamura, S., *et al.* (2019), Int. J. Mod. Phys. D **28** (12), 1845001.

[1024] Kawamura, S., *et al.* (2021), PTEP **2021** (5), 05A105, arXiv:2006.13545 [gr-qc] .

[1025] Kawasaki, M., and M. Kusakabe (2011), Phys. Rev. D **83**, 055011, arXiv:1012.0435 [hep-ph] .

[1026] Kazantzidis, L., H. Koo, S. Nesseris, L. Perivolaropoulos, and A. Shafieloo (2021), Mon. Not. Roy. Astron. Soc. **501** (3), 3421, arXiv:2010.03491 [astro-ph.CO] .

[1027] Kazantzidis, L., and L. Perivolaropoulos (2018), Phys. Rev. D **97** (10), 103503, arXiv:1803.01337 [astro-ph.CO] .

[1028] Kazantzidis, L., and L. Perivolaropoulos (2019), arXiv:1907.03176 [astro-ph.CO] .

[1029] Kazantzidis, L., and L. Perivolaropoulos (2020), Phys. Rev. D **102** (2), 023520, arXiv:2004.02155 [astro-ph.CO] .

[1030] Kazantzidis, L., L. Perivolaropoulos, and F. Skara (2019), Phys. Rev. D **99** (6), 063537, arXiv:1812.05356 [astro-ph.CO] .

[1031] Keeley, R. E., S. Joudaki, M. Kaplinghat, and D. Kirkby (2019), JCAP **12**, 035, arXiv:1905.10198 [astro-ph.CO] .

[1032] Keeley, R. E., and A. Shafieloo (2021), arXiv:2111.04231 [astro-ph.CO] .

[1033] Keeley, R. E., A. Shafieloo, D. K. Hazra, and T. Souradeep (2020), JCAP **09**, 055, arXiv:2006.12710 [astro-ph.CO] .

[1034] Keeley, R. E., A. Shafieloo, B. L'Huillier, and E. V. Linder (2020), Mon. Not. Roy. Astron. Soc. **491** (3), 3983, arXiv:1905.10216 [astro-ph.CO] .

[1035] Keeley, R. E., A. Shafieloo, G.-B. Zhao, J. A. Vazquez, and H. Koo (2021), Astron. J. **161** (3), 151, arXiv:2010.03234 [astro-ph.CO] .

[1036] Keenan, R. C., A. J. Barger, and L. L. Cowie (2013), Astrophys. J. **775**, 62, arXiv:1304.2884 [astro-ph.CO] .

[1037] Keeton, C. R., and C. S. Kochanek (1997), Astrophys. J. **487**, 42, arXiv:astro-ph/9611216 .

[1038] Kehagias, A., and M. Maggiore (2014), JHEP **08**, 029, arXiv:1401.8289 [hep-th] .

[1039] Kehagias, A., and K. Sfetsos (2000), Phys. Lett. B **472**, 39, arXiv:hep-ph/9905417 .

[1040] Keisler, R., *et al.* (SPT) (2015), Astrophys. J. **807** (2), 151, arXiv:1503.02315 [astro-ph.CO] .

[1041] Kempf, A. (1996), in *21st International Colloquium on Group Theoretical Methods in Physics*, arXiv:hep-th/9612082 .

[1042] Kempf, A. (1997), J. Math. Phys. **38**, 1347, arXiv:hep-th/9602085 [hep-th] .

[1043] Kempf, A., G. Mangano, and R. B. Mann (1995), Phys. Rev. **D52**, 1108, arXiv:hep-th/9412167 [hep-th] .

[1044] Kennedy, J., L. Lombriser, and A. Taylor (2018), Phys. Rev. D **98** (4), 044051, arXiv:1804.04582 [astro-ph.CO] .

[1045] Kenworthy, W. D., D. Scolnic, and A. Riess (2019), Astrophys. J. **875** (2), 145, arXiv:1901.08681 [astro-ph.CO] .

[1046] Kerscher, M., and J. Weller (2019), SciPost Phys. Lect. Notes **9**, 1, arXiv:1901.07726 [astro-ph.CO] .

[1047] Khadka, N., O. Luongo, M. Muccino, and B. Ratra (2021), JCAP **09**, 042, arXiv:2105.12692 [astro-ph.CO] .

[1048] Khadka, N., and B. Ratra (2020), Mon. Not. Roy. Astron. Soc. **499** (1), 391, arXiv:2007.13907 [astro-ph.CO] .

[1049] Khadka, N., and B. Ratra (2020), Mon. Not. Roy. Astron. Soc. **492** (3), 4456, arXiv:1909.01400 [astro-ph.CO] .

[1050] Khadka, N., and B. Ratra (2021), Mon. Not. Roy. Astron. Soc. **502** (4), 6140, arXiv:2012.09291 [astro-ph.CO] .

[1051] Khadka, N., and B. Ratra (2022), Mon. Not. Roy. Astron. Soc. **510** (2), 2753, arXiv:2107.07600 [astro-ph.CO] .

[1052] Khetan, N., *et al.* (2021), Astron. Astrophys. **647**, A72, arXiv:2008.07754 [astro-ph.CO] .

[1053] Khosravi, N., S. Baghram, N. Afshordi, and N. Altamirano (2019), Phys. Rev. **D99** (10), 103526, arXiv:1710.09366 [astro-ph.CO] .

[1054] Khoury, J. (2010), arXiv:1011.5909 [astro-ph.CO] .

[1055] Khoury, J. (2013), Class. Quant. Grav. **30**, 214004, arXiv:1306.4326 [astro-ph.CO] .

[1056] Khoury, J., and A. Weltman (2004), Phys. Rev. D **69**, 044026, arXiv:astro-ph/0309411 .

[1057] Khoury, J., and A. Weltman (2004), Phys. Rev. Lett. **93**, 171104, arXiv:astro-ph/0309300 .

[1058] Kilbinger, M. (2015), Rept. Prog. Phys. **78**, 086901, arXiv:1411.0115 [astro-ph.CO] .

[1059] Kilbinger, M., *et al.* (2013), Mon. Not. Roy. Astron. Soc. **430**, 2200, arXiv:1212.3338 [astro-ph.CO] .

[1060] Kim, J., and P. Naselsky (2010), Astrophys. J. Lett. **714**, L265, arXiv:1001.4613 [astro-ph.CO] .

[1061] Kim, J., and P. Naselsky (2010), Phys. Rev. D **82**, 063002, arXiv:1002.0148 [astro-ph.CO] .

[1062] Kim, J., and P. Naselsky (2011), The Astrophysical Journal **739** (2), 79.

[1063] Kim, Y. J., J. Kang, M. G. Lee, and I. S. Jang (2020), Astrophys. J. **905** (2), 104, arXiv:2010.01364 [astro-ph.CO] .

[1064] Kim, Y.-L., M. Smith, M. Sullivan, and Y.-W. Lee (2018), ApJ **854** (1), 24, arXiv:1801.01192 [astro-ph.GA] .

[1065] King, J. A., J. K. Webb, M. T. Murphy, V. V. Flambaum, R. F. Carswell, M. B. Bainbridge, M. R. Wilczynska, and F. E. Koch (2012), Mon. Not. Roy. Astron. Soc. **422**, 3370, arXiv:1202.4758 [astro-ph.CO] .

[1066] Kirby, M., E. Rozo, R. G. Morris, S. W. Allen, M. Costanzi, T. E. Jeltema, A. B. Mantz, A. K. Romer, E. S. Rykoff, and A. von der Linden (2019), arXiv:1910.13548 [astro-ph.CO] .

[1067] Kirshner, R. P., and J. Kwan (1974), ApJ **193**, 27.

[1068] Kitazawa, N. (2020), arXiv:2010.12164 [astro-ph.CO] .

[1069] Kiuchi, K., M. Shibata, P. J. Montero, and J. A. Font (2011), Phys. Rev. Lett. **106**, 251102, arXiv:1105.5035 [astro-ph.HE] .

[1070] Klaewer, D., and E. Palti (2017), JHEP **01**, 088, arXiv:1610.00010 [hep-th] .

[1071] Klein, M., *et al.* (DES) (2018), Mon. Not. Roy. Astron. Soc. **474** (3), 3324, arXiv:1706.06577 [astro-ph.CO] .

[1072] Klypin, A., V. Poulin, F. Prada, J. Primack, M. Kamionkowski, V. Avila-Reese, A. Rodriguez-Puebla, P. Behroozi, D. Hellinger, and T. L. Smith (2021), Mon. Not. Roy. Astron. Soc. **504** (1), 769, arXiv:2006.14910 [astro-ph.CO] .

[1073] Klypin, A. A., A. V. Kravtsov, O. Valenzuela, and F. Prada (1999), Astrophys. J. **522**, 82, arXiv:astro-ph/9901240 .

[1074] Knox, L., and M. Millea (2020), Phys. Rev. D **101** (4), 043533, arXiv:1908.03663 [astro-ph.CO] .

[1075] Ko, P., N. Nagata, and Y. Tang (2017), Phys. Lett. B **773**, 513, arXiv:1706.05605 [hep-ph] .

[1076] Ko, P., and Y. Tang (2016), Phys. Lett. B **762**, 462, arXiv:1608.01083 [hep-ph] .

[1077] Ko, P., and Y. Tang (2017), Phys. Lett. B **768**, 12, arXiv:1609.02307 [hep-ph] .

[1078] Kobayashi, T. (2019), Rept. Prog. Phys. **82** (8), 086901, arXiv:1901.07183 [gr-qc] .

[1079] Kochanek, C. S. (2002), Astrophys. J. **578**, 25, arXiv:astro-ph/0205319 .

[1080] Kochanek, C. S. (2003), Astrophys. J. **583**, 49, arXiv:astro-ph/0206006 .

[1081] Kochanek, C. S. (2020), Mon. Not. Roy. Astron. Soc. **493** (2), 1725, arXiv:1911.05083 [astro-ph.CO] .

[1082] Kochanek, C. S. (2021), Mon. Not. Roy. Astron. Soc. **501** (4), 5021, arXiv:2003.08395 [astro-ph.CO] .

[1083] Kochanek, C. S., N. D. Morgan, E. E. Falco, B. A. McLeod, J. N. Winn, J. Dembicky, and B. Ketzeback (2006), Astrophys. J. **640**, 47, arXiv:astro-ph/0508070 .

[1084] Koda, J., C. Blake, T. Davis, C. Magoulas, C. M. Springob, M. Scrimgeour, A. Johnson, G. B. Poole, and L. Staveley-Smith (2014), Mon. Not. Roy. Astron. Soc. **445** (4), 4267, arXiv:1312.1022 [astro-ph.CO] .

[1085] Kofinas, G., E. Papantonopoulos, and E. N. Saridakis (2015), Phys. Rev. **D91** (10), 104034, arXiv:1501.00365 [gr-qc] .

[1086] Kofman, L., and A. A. Starobinsky (1985), Sov. Astron. Lett. **11**, 271.

[1087] Kogut, A., *et al.* (1993), Astrophys. J. **419**, 1, arXiv:astro-ph/9312056 .

[200] Köhlinger, F., *et al.* (2017), Mon. Not. Roy. Astron. Soc. **471** (4), 4412, arXiv:1706.02892 [astro-ph.CO] .

[1089] Kohri, K., and F. Takayama (2007), Phys. Rev. D **76**, 063507, arXiv:hep-ph/0605243 .

[1090] Kokkotas, K. D. (2008), Rev. Mod. Astron. **20**, 140, arXiv:0809.1602 [astro-ph] .

[1091] Kokkotas, K. D., and N. Stergioulas (2005), in *5th International Workshop on New Worlds in Astroparticle Physics*, arXiv:gr-qc/0506083 .

[157] Komatsu, E., *et al.* (WMAP) (2009), Astrophys. J. Suppl. **180**, 330, arXiv:0803.0547 [astro-ph] .

[1093] Komatsu, E., *et al.* (WMAP) (2011), Astrophys. J. Suppl. **192**, 18, arXiv:1001.4538 [astro-ph.CO] .

[1094] Koo, H., R. E. Keeley, A. Shafieloo, and B. L'Huillier (2022), JCAP **03** (03), 047, arXiv:2110.10977 [astro-ph.CO] .

[1095] Koo, H., A. Shafieloo, R. E. Keeley, and B. L'Huillier (2020), Astrophys. J. **899** (1), 9, arXiv:2001.10887 [astro-ph.CO] .

[1096] Korn, A. J., F. Grundahl, O. Richard, P. S. Barklem, L. Mashonkina, R. Collet, N. Piskunov, and B. Gustafsson (2006), Nature **442**, 657, arXiv:astro-ph/0608201 .

[1097] Kourkchi, E., R. B. Tully, G. S. Anand, H. M. Courtois, A. Dupuy, J. D. Neill, L. Rizzi, and M. Seibert (2020), Astrophys. J. **896** (1), 3, arXiv:2004.14499 [astro-ph.GA] .

[1098] Kovács, A. (2018), Mon. Not. Roy. Astron. Soc. **475** (2), 1777, arXiv:1701.08583 [astro-ph.CO] .

[1099] Kovács, A., *et al.* (DES) (2019), Mon. Not. Roy. Astron. Soc. **484**, 5267, arXiv:1811.07812 [astro-ph.CO] .

[1100] Kovács, A., *et al.* (2022), arXiv:2203.11306 [astro-ph.CO] .

[1101] Kovetz, E. D., V. Poulin, V. Gluscevic, K. K. Boddy, R. Barkana, and M. Kamionkowski (2018), Phys. Rev. D **98** (10), 103529, arXiv:1807.11482 [astro-ph.CO] .

[1102] Kraljic, D., and S. Sarkar (2015), JCAP **04**, 050, arXiv:1412.7719 [astro-ph.CO] .

[1103] Kraljic, D., and S. Sarkar (2016), JCAP **10**, 016, arXiv:1607.07377 [astro-ph.CO] .

[1104] Krall, R., F.-Y. Cyr-Racine, and C. Dvorkin (2017), JCAP **09**, 003, arXiv:1705.08894 [astro-ph.CO] .

[205] Krause, E., *et al.* (DES) (2017), arXiv:1706.09359 [astro-ph.CO] .

[1106] Kravtsov, A., and S. Borgani (2012), Ann. Rev. Astron. Astrophys. **50**, 353, arXiv:1205.5556 [astro-ph.CO] .

[206] Kreisch, C. D., F.-Y. Cyr-Racine, and O. Doré (2020), Phys. Rev. D **101** (12), 123505, arXiv:1902.00534 [astro-ph.CO] .

[1108] Krisciunas, K., *et al.* (2017), Astron. J. **154** (5), 211, arXiv:1709.05146 [astro-ph.IM] .

[1109] Krishnan, C., E. O. Colgáin, M. M. Sheikh-Jabbari, and T. Yang (2021), Phys. Rev. D **103** (10), 103509, arXiv:2011.02858 [astro-ph.CO] .

[1110] Krishnan, C., R. Mohayaee, E. O. Colgáin, M. M. Sheikh-Jabbari, and L. Yin (2021), Class. Quant. Grav. **38** (18), 184001, arXiv:2105.09790 [astro-ph.CO] .

[1111] Krolewski, A., S. Ferraro, and M. White (2021), JCAP **12** (12), 028, arXiv:2105.03421 [astro-ph.CO] .

[1112] Kroupa, P. (2012), Publ. Astron. Soc. Austral. **29**, 395, arXiv:1204.2546 [astro-ph.CO] .

[1113] Kroupa, P. (2015), Can. J. Phys. **93** (2), 169, arXiv:1406.4860 [astro-ph.GA] .

[1114] Kroupa, P., B. Famaey, K. S. de Boer, J. Dabringhausen, M. S. Pawlowski, C. M. Boily, H. Jerjen, D. Forbes, G. Hensler, and M. Metz (2010), Astron. Astrophys. **523**, A32, arXiv:1006.1647 [astro-ph.CO] .

[1115] Kroupa, P., C. Theis, and C. M. Boily (2005), Astron. Astrophys. **431**, 517, arXiv:astro-ph/0410421 .

[1116] Kuijken, K., *et al.* (2015), Mon. Not. Roy. Astron. Soc. **454** (4), 3500, arXiv:1507.00738 [astro-ph.CO] .

[1117] Kuijken, K., *et al.* (2019), Astron. Astrophys. **625**, A2, arXiv:1902.11265 [astro-ph.GA] .

[1118] Kumar, S. (2021), Phys. Dark Univ. **33**, 100862, arXiv:2102.12902 [astro-ph.CO] .

[1119] Kumar, S., and R. C. Nunes (2016), Phys. Rev. **D94** (12), 123511, arXiv:1608.02454 [astro-ph.CO] .

[1120] Kumar, S., and R. C. Nunes (2017), Phys. Rev. **D96** (10), 103511, arXiv:1702.02143 [astro-ph.CO] .

[1121] Kumar, S., R. C. Nunes, and S. K. Yadav (2018), Phys. Rev. D **98** (4), 043521, arXiv:1803.10229 [astro-ph.CO] .

[1122] Kumar, S., R. C. Nunes, and S. K. Yadav (2019), Eur. Phys. J. **C79** (7), 576, arXiv:1903.04865 [astro-ph.CO] .

[1123] Kuo, C., J. A. Braatz, M. J. Reid, F. K. Y. Lo, J. J. Condon, C. M. V. Impellizzeri, and C. Henkel (2013), Astrophys. J. **767**, 155, arXiv:1207.7273 [astro-ph.CO] .

[1124] Kuo, C. Y., J. A. Braatz, K. Y. Lo, M. J. Reid, S. H. Suyu, D. W. Pesce, J. J. Condon, C. Henkel, and C. M. V. Impellizzeri (2015), Astrophys. J. **800** (1), 26, arXiv:1411.5106 [astro-ph.GA] .

[1125] Kuroda, K. (LCGT) (2010), Class. Quant. Grav. **27**, 084004.

[1126] Kusakabe, M., K. S. Kim, M.-K. Cheoun, T. Kajino, Y. Kino, and G. J. Mathews (2014), Astrophys. J. Suppl. **214**, 5, arXiv:1403.4156 [astro-ph.CO] .

[1127] Kusenko, A. (2009), Phys. Rept. **481**, 1, arXiv:0906.2968 [hep-ph] .

[1128] La Posta, A., T. Louis, X. Garrido, and J. C. Hill (2021), arXiv:2112.10754 [astro-ph.CO] .

[1129] Lahav, O., P. B. Lilje, J. R. Primack, and M. J. Rees (1991), Mon. Not. Roy. Astron. Soc. **251**, 128.

[1130] Lake, M. J. (2019), Galaxies **7** (1), 11, arXiv:1712.00271 [gr-qc] .

[1131] Lake, M. J. (2020), arXiv:2008.13183 [gr-qc] .

[1132] Lamb, D. Q., and D. E. Reichart (2000), Astrophys. J. **536**, 1, arXiv:astro-ph/9909002 .

[1133] Lambiase, G., S. Mohanty, A. Narang, and P. Parashari (2019), Eur. Phys. J. C **79** (2), 141, arXiv:1804.07154 [astro-ph.CO] .

[209] Lancaster, L., F.-Y. Cyr-Racine, L. Knox, and Z. Pan (2017), JCAP **1707** (07), 033, arXiv:1704.06657 [astro-ph.CO] .

[1135] Land, K., and J. Magueijo (2005), Phys. Rev. D **72**, 101302, arXiv:astro-ph/0507289 .

[1136] Landy, S. D., and A. S. Szalay (1993), Astrophys. J. **412**, 64.

[1137] Larena, J., J.-M. Alimi, and A. Serna (2007), Astrophys. J. **658**, 1, arXiv:astro-ph/0511693 .

[1138] Laureijs, R., *et al.* (Euclid) (2011), arXiv:1110.3193 [astro-ph.CO] .

[1139] Leaf, K., and F. Melia (2018), Mon. Not. Roy. Astron. Soc. **474** (4), 4507, arXiv:1711.10793 [astro-ph.CO] .

[1140] Leandro, H., V. Marra, and R. Sturani (2022), Phys. Rev. D **105** (2), 023523, arXiv:2109.07537 [gr-qc] .

[1141] Leavitt, H. S. (1908), Harvard Obs. Annals **60**, 87.

[1142] Leavitt, H. S., and E. C. Pickering (1912), Harvard Obs. Circ. **173**, 1.

[1143] Lee, B.-H., W. Lee, E. O. Colgáin, M. M. Sheikh-Jabbari, and S. Thakur (2022), arXiv:2202.03906 [astro-ph.CO] .

[1144] Lee, C.-C., J. K. Webb, D. Milaković, and R. F. Carswell (2021), arXiv:2102.11648 [astro-ph.CO] .

[1145] Lee, T. D., and C.-N. Yang (1956), Phys. Rev. **104**, 254.

[1146] van Leeuwen, F., M. W. Feast, P. A. Whitelock, and C. D. Laney (2007), Mon. Not. Roy. Astron. Soc. **379**, 723, arXiv:0705.1592 [astro-ph] .

[1147] Leistedt, B., H. V. Peiris, and L. Verde (2014), Phys. Rev. Lett. **113**, 041301, arXiv:1404.5950 [astro-ph.CO] .

[1148] Lelli, F. (2022), Nature Astron. **6**, 35, arXiv:2201.11752 [astro-ph.GA] .

[1149] Lelli, F., S. S. McGaugh, and J. M. Schombert (2016), Astron. J. **152**, 157, arXiv:1606.09251 [astro-ph.GA] .

[1150] Lelli, F., S. S. McGaugh, and J. M. Schombert (2016), Astrophys. J. Lett. **816** (1), L14, arXiv:1512.04543 [astro-ph.GA] .

[1151] Lelli, F., S. S. McGaugh, J. M. Schombert, H. Desmond, and H. Katz (2019), Mon. Not. Roy. Astron. Soc. **484** (3), 3267, arXiv:1901.05966 [astro-ph.GA] .

[1152] Lemos, P., *et al.* (DES) (2021), Mon. Not. Roy. Astron. Soc. **505** (4), 6179, arXiv:2012.09554 [astro-ph.CO] .

[1153] Leon, G., A. Paliathanasis, E. N. Saridakis, and S. Basilakos (2022), arXiv:2203.14866 [gr-qc] .

[1154] Lesci, G. F., *et al.* (2022), Astron. Astrophys. **659**, A88, arXiv:2012.12273 [astro-ph.CO] .

[215] Lesgourgues, J., G. Marques-Tavares, and M. Schmaltz (2016), JCAP **02**, 037, arXiv:1507.04351 [astro-ph.CO] .

[216] Lesgourgues, J., and S. Pastor (2006), Phys. Rept. **429**, 307, arXiv:astro-ph/0603494 .

[1157] Lesgourgues, J., and S. Pastor (2012), Adv. High Energy Phys. **2012**, 608515, arXiv:1212.6154 [hep-ph] .

[1158] Lesgourgues, J., and S. Pastor (2014), New J. Phys. **16**, 065002, arXiv:1404.1740 [hep-ph] .

[1159] Levshakov, S. A., F. Combes, F. Boone, I. I. Agafonova, D. Reimers, and M. G. Kozlov (2012), Astron. Astrophys. **540**, L9, arXiv:1203.3649 [astro-ph.CO] .

[220] Lewis, A., and A. Challinor (2006), Phys. Rept. **429**, 1, arXiv:astro-ph/0601594 [astro-ph] .

[1161] Lewis, A., A. Challinor, and N. Turok (2002), Phys. Rev. D **65**, 023505, arXiv:astro-ph/0106536 .

[1162] L'Huillier, B., and A. Shafieloo (2017), JCAP **01**, 015, arXiv:1606.06832 [astro-ph.CO] .

[1163] L'Huillier, B., A. Shafieloo, and H. Kim (2018), Mon. Not. Roy. Astron. Soc. **476** (3), 3263, arXiv:1712.04865 [astro-ph.CO] .

[1164] L'Huillier, B., A. Shafieloo, D. Polarski, and A. A. Starobinsky (2020), Mon. Not. Roy. Astron. Soc. **494** (1), 819, arXiv:1906.05991 [astro-ph.CO] .

[1165] Li, B., Z. Chen, Y.-F. Cai, and Y. Mao (2019), Mon. Not. Roy. Astron. Soc. **487** (4), 5564, arXiv:1904.04683 [astro-ph.CO] .

[1166] Li, E.-K., M. Du, Z.-H. Zhou, H. Zhang, and L. Xu (2021), Mon. Not. Roy. Astron. Soc. **501** (3), 4452, arXiv:1911.12076 [astro-ph.CO] .

[1167] Li, H.-L., L. Feng, J.-F. Zhang, and X. Zhang (2019), Sci. China Phys. Mech. Astron. **62** (12), 120411, arXiv:1812.00319 [astro-ph.CO] .

[1168] Li, W., *et al.* (2011), Mon. Not. Roy. Astron. Soc. **412**, 1441, arXiv:1006.4612 [astro-ph.SR] .

[1169] Li, X., R. E. Keeley, A. Shafieloo, X. Zheng, S. Cao, M. Biesiada, and Z.-H. Zhu (2021), Mon. Not. Roy. Astron. Soc. **507** (1), 919, arXiv:2103.16032 [astro-ph.CO] .

[1170] Li, X., and A. Shafieloo (2019), Astrophys. J. Lett. **883** (1), L3, arXiv:1906.08275 [astro-ph.CO] .

[1171] Li, X., and A. Shafieloo (2020), Astrophys. J. **902** (1), 58, arXiv:2001.05103 [astro-ph.CO] .

[1172] Li, X., A. Shafieloo, V. Sahni, and A. A. Starobinsky (2019), Astrophys. J. **887**, 153, arXiv:1904.03790 [astro-ph.CO] .

[1173] Li, Z., Y. Jing, P. Zhang, and D. Cheng (2016), Astrophys. J. **833** (2), 287, arXiv:1609.03697 [astro-ph.CO] .

[1174] Liao, K., A. Shafieloo, R. E. Keeley, and E. V. Linder (2019), Astrophys. J. Lett. **886** (1), L23, arXiv:1908.04967 [astro-ph.CO] .

[1175] Liao, K., A. Shafieloo, R. E. Keeley, and E. V. Linder (2020), Astrophys. J. Lett. **895** (2), L29, arXiv:2002.10605 [astro-ph.CO] .

[1176] Libeskind, N. I., C. S. Frenk, S. Cole, J. C. Helly, A. Jenkins, J. F. Navarro, and C. Power (2005), Mon. Not. Roy. Astron. Soc. **363**, 146, arXiv:astro-ph/0503400 .

[1177] Liddle, A. R. (2004), Mon. Not. Roy. Astron. Soc. **351**, L49, arXiv:astro-ph/0401198 .

[1178] Liddle, A. R. (2007), Mon. Not. Roy. Astron. Soc. **377**, L74, arXiv:astro-ph/0701113 .

[1179] Liddle, A. R., and M. Cortês (2013), Phys. Rev. Lett. **111** (11), 111302, arXiv:1306.5698 [astro-ph.CO] .

[1180] Liddle, A. R., and R. J. Scherrer (1999), Phys. Rev. D **59**, 023509, arXiv:astro-ph/9809272 .

[1181] Lima, J. A. S., S. Basilakos, and J. Sola (2013), Mon. Not. Roy. Astron. Soc. **431**, 923, arXiv:1209.2802 [gr-qc] .

[1182] Lima, N. A., V. Smer-Barreto, and L. Lombriser (2016), Phys. Rev. D **94** (8), 083507, arXiv:1603.05239 [astro-ph.CO] .

[222] Lin, M.-X., G. Benevento, W. Hu, and M. Raveri (2019), Phys. Rev. D **100** (6), 063542, arXiv:1905.12618 [astro-ph.CO] .

[1184] Lin, M.-X., W. Hu, and M. Raveri (2020), Phys. Rev. D **102**, 123523, arXiv:2009.08974 [astro-ph.CO] .

[1185] Linares Cedeño, F. X., and U. Nucamendi (2021), Phys. Dark Univ. **32**, 100807, arXiv:2009.10268 [astro-ph.CO] .

[170] Linde, A. D. (1982), *Second Seminar on Quantum Gravity Moscow, USSR, October 13-15, 1981*, Phys. Lett. **108B**, 389.

[1187] Lindegren, L., U. Lammers, U. Bastian, and et al. (2016), Astron. Astrophys. **595**, A4, arXiv:1609.04303 [astro-ph.GA] .

[1188] Lindegren, L., *et al.* (2018), Astron. Astrophys. **616**, A2, arXiv:1804.09366 [astro-ph.IM] .

[1189] von der Linden, A., *et al.* (2014), Mon. Not. Roy. Astron. Soc. **439** (1), 2, arXiv:1208.0597 [astro-ph.CO] .

[1190] Linder, E. V. (2003), Phys. Rev. Lett. **90**, 091301, arXiv:astro-ph/0208512 [astro-ph] .

[1191] Linder, E. V. (2005), Phys. Rev. D **72**, 043529, arXiv:astro-ph/0507263 .

[1192] Linder, E. V. (2018), JCAP **03**, 005, arXiv:1801.01503 [astro-ph.CO] .

[1193] Linton, M. S., R. Crittenden, and A. Pourtsidou (2021), arXiv:2107.03235 [astro-ph.CO] .

[1194] Littenberg, T. B., and N. J. Cornish (2015), Phys. Rev. D **91** (8), 084034, arXiv:1410.3852 [gr-qc] .

[1195] Liu, J., and H. Wei (2015), Gen. Rel. Grav. **47** (11), 141, arXiv:1410.3960 [astro-ph.CO] .

[1196] Liu, M., and Z. Huang (2020), Astrophys. J. **897**, 166, arXiv:1910.05670 [astro-ph.CO] .

[1197] Liu, M., Z. Huang, X. Luo, H. Miao, N. K. Singh, and L. Huang (2020), Sci. China Phys. Mech. Astron. **63** (9), 290405, arXiv:1912.00190 [astro-ph.CO] .

[1198] Liu, S., Y.-M. Hu, J.-d. Zhang, and J. Mei (2020), Phys. Rev. D **101** (10), 103027, arXiv:2004.14242 [astro-ph.HE] .

[1199] Liu, Y., F. Chen, N. Liang, Z. Yuan, H. Yu, and P. Wu1 (2022), arXiv:2203.03178 [astro-ph.CO] .

[1200] Liu, Z., and H. Miao (2020), Int. J. Mod. Phys. D **29** (13), 2050088, arXiv:2002.05563 [astro-ph.CO] .

[1201] Lombriser, L. (2020), Phys. Lett. B **803**, 135303, arXiv:1906.12347 [astro-ph.CO] .

[1202] Loureiro, A., *et al.* (KiDS, Euclid) (2021), arXiv:2110.06947 [astro-ph.CO] .

[1203] Lovell, M. R., V. Gonzalez-Perez, S. Bose, A. Boyarsky, S. Cole, C. S. Frenk, and O. Ruchayskiy (2017), Mon. Not. Roy. Astron. Soc. **468** (3), 2836, arXiv:1611.00005 [astro-ph.GA] .

[1204] Lucca, M. (2020), Phys. Lett. B **810**, 135791, arXiv:2008.01115 [astro-ph.CO] .

[1205] Lucca, M. (2021), Phys. Rev. D **104** (8), 083510, arXiv:2106.15196 [astro-ph.CO] .

[1206] Lucca, M., and D. C. Hooper (2020), Phys. Rev. D **102** (12), 123502, arXiv:2002.06127 [astro-ph.CO] .

[1207] Lue, A., L.-M. Wang, and M. Kamionkowski (1999), Phys. Rev. Lett. **83**, 1506, arXiv:astro-ph/9812088 .

[1208] Luković, V. V., B. S. Haridasu, and N. Vittorio (2020), Mon. Not. Roy. Astron. Soc. **491** (2), 2075, arXiv:1907.11219 [astro-ph.CO] .

[1209] Luo, J., *et al.* (TianQin) (2016), Class. Quant. Grav. **33** (3), 035010, arXiv:1512.02076 [astro-ph.IM] .

[1210] Luo, Y., T. Kajino, M. Kusakabe, and G. J. Mathews (2019), Astrophys. J. **872** (2), 172, arXiv:1810.08803 [astro-ph.CO] .

[1211] Luongo, O., and M. Muccino (2021), Galaxies **9** (4), 77, arXiv:2110.14408 [astro-ph.HE] .

[1212] Luongo, O., M. Muccino, E. O. Colgáin, M. M. Sheikh-Jabbari, and L. Yin (2021), arXiv:2108.13228 [astro-ph.CO] .

[1213] Lusso, E., E. Piedipalumbo, G. Risaliti, M. Paolillo, S. Bisogni, E. Nardini, and L. Amati (2019), Astron. Astrophys. **628**, L4, arXiv:1907.07692 [astro-ph.CO] .

[1214] Lusso, E., *et al.* (2010), Astron. Astrophys. **512**, A34, arXiv:0912.4166 [astro-ph.CO] .

[1215] Lusso, E., *et al.* (2020), Astron. Astrophys. **642**, A150, arXiv:2008.08586 [astro-ph.GA] .

[1216] Lyth, D. H. (2015), JCAP **04**, 039, arXiv:1405.3562 [astro-ph.CO] .

[1217] Lyu, K.-F., E. Stamou, and L.-T. Wang (2021), Phys. Rev. D **103** (1), 015004, arXiv:2004.10868 [hep-ph] .

[1218] Ma, C.-P., and E. Bertschinger (1995), Astrophys. J. **455**, 7, arXiv:astro-ph/9506072 .

[1219] Ma, Y.-Z., and D. Scott (2013), Mon. Not. Roy. Astron. Soc. **428**, 2017, arXiv:1208.2028 [astro-ph.CO] .

[1220] Macaulay, E., I. K. Wehus, and H. K. Eriksen (2013), Phys. Rev. Lett. **111** (16), 161301, arXiv:1303.6583 [astro-ph.CO] .

[1221] Maggiore, M. (1993), Phys. Lett. **B304**, 65, arXiv:hep-th/9301067 [hep-th] .

[1222] Maggiore, M. (1993), Phys. Lett. **B319**, 83, arXiv:hep-th/9309034 [hep-th] .

[1223] Maggiore, M. (1994), Phys. Rev. **D49**, 5182, arXiv:hep-th/9305163 [hep-th] .

[1224] Maggiore, M. (2000), Phys. Rept. **331**, 283, arXiv:gr-qc/9909001 .

[1225] Maio, U., and M. Viel (2015), Mon. Not. Roy. Astron. Soc. **446**, 2760, arXiv:1409.6718 [astro-ph.CO] .

[1226] Mancini, A. S., and A. Pourtsidou (2021), 10.1093/mnrasl/slac019, arXiv:2110.07587 [astro-ph.CO] .

[1227] Mandal, S., D. Wang, and P. K. Sahoo (2020), Phys. Rev. D **102**, 124029, arXiv:2011.00420 [gr-qc] .

[1228] Mandelbaum, R. (2018), Ann. Rev. Astron. Astrophys. **56**, 393, arXiv:1710.03235 [astro-ph.CO] .

[1229] Mangano, G., A. Melchiorri, P. Serra, A. Cooray, and M. Kamionkowski (2006), Phys. Rev. D **74**, 043517, arXiv:astro-ph/0606190 .

[1230] Mangano, G., G. Miele, S. Pastor, T. Pinto, O. Pisanti, and P. D. Serpico (2005), Nucl. Phys. **B729**, 221, arXiv:hep-ph/0506164 [hep-ph] .

[1231] Mantz, A., S. W. Allen, D. Rapetti, and H. Ebeling (2010), Mon. Not. Roy. Astron. Soc. **406**, 1759, arXiv:0909.3098 [astro-ph.CO] .

[1232] Mantz, A. B., *et al.* (2015), Mon. Not. Roy. Astron. Soc. **446**, 2205, arXiv:1407.4516 [astro-ph.CO] .

[1233] Marcos-Caballero, A., R. Fernández-Cobos, E. Martínez-González, and P. Vielva (2016), JCAP **04**, 058, arXiv:1512.07412 [astro-ph.CO] .

[1234] Marcos-Caballero, A., and E. Martínez-González (2019), JCAP **10**, 053, arXiv:1909.06093 [astro-ph.CO] .

[1235] Mariano, A., and L. Perivolaropoulos (2012), Phys. Rev. D **86**, 083517, arXiv:1206.4055 [astro-ph.CO] .

[1236] Mariano, A., and L. Perivolaropoulos (2013), Phys. Rev. D **87** (4), 043511, arXiv:1211.5915 [astro-ph.CO] .

[1237] Marques, G. A., J. Liu, J. M. Z. Matilla, Z. Haiman, A. Bernui, and C. P. Novaes (2019), JCAP **06**, 019, arXiv:1812.08206 [astro-ph.CO] .

[1238] Marra, V., and L. Perivolaropoulos (2021), Phys. Rev. D **104** (2), L021303, arXiv:2102.06012 [astro-ph.CO] .

[1239] Marsh, D. J. E. (2011), Phys. Rev. D **83**, 123526, arXiv:1102.4851 [astro-ph.CO] .

[179] Marsh, D. J. E. (2016), Phys. Rept. **643**, 1, arXiv:1510.07633 [astro-ph.CO] .

[1241] Marshall, P., *et al.* (LSST) (2017), 10.5281/zenodo.842713, arXiv:1708.04058 [astro-ph.IM] .

[1242] Martin, J. (2012), Comptes Rendus Physique **13**, 566, arXiv:1205.3365 [astro-ph.CO] .

[1243] Martinelli, M., N. B. Hogg, S. Peirone, M. Bruni, and D. Wands (2019), Mon. Not. Roy. Astron. Soc. **488** (3), 3423, arXiv:1902.10694 [astro-ph.CO] .

[1244] Martinelli, M., and I. Tutusaus (2019), Symmetry **11** (8), 986, arXiv:1906.09189 [astro-ph.CO] .

[1245] Martinelli, M., *et al.* (EUCLID) (2020), Astron. Astrophys. **644**, A80, arXiv:2007.16153 [astro-ph.CO] .

[1246] Martinelli, M., *et al.* (Euclid) (2021), Astron. Astrophys. **654**, A148, arXiv:2105.09746 [astro-ph.CO] .

[1247] de Martino, I., S. S. Chakrabarty, V. Cesare, A. Gallo, L. Ostorero, and A. Diaferio (2020), Universe **6** (8), 107, arXiv:2007.15539 [astro-ph.CO] .

[1248] Martins, C. J. A. P. (2017), 10.1088/1361-6633/aa860e, arXiv:1709.02923 [astro-ph.CO] .

[1249] Martins, C. J. A. P., and A. M. M. Pinho (2017), Phys. Rev. D **95** (2), 023008, arXiv:1701.08724 [astro-ph.CO] .

[1250] Marulli, F., C. Carbone, M. Viel, L. Moscardini, and A. Cimatti (2011), Mon. Not. Roy. Astron. Soc. **418**, 346, arXiv:1103.0278 [astro-ph.CO] .

[1251] Mateo, M. (1998), Ann. Rev. Astron. Astrophys. **36**, 435, arXiv:astro-ph/9810070 .

[1252] Mather, J. C., *et al.* (1990), Astrophys. J. Lett. **354**, L37.

[1253] Mathews, G. J., A. Kedia, N. Sasankan, M. Kusakabe, Y. Luo, T. Kajino, D. Yamazaki, T. Makki, and M. E. Eid (2020), JPS Conf. Proc. **31**, 011033, arXiv:1909.01245 [astro-ph.CO] .

[1254] Matsubara, T. (2004), Astrophys. J. **615**, 573, arXiv:astro-ph/0408349 .

[1255] de Mattia, A., *et al.* (2021), Mon. Not. Roy. Astron. Soc. **501** (4), 5616, arXiv:2007.09008 [astro-ph.CO] .

[1256] Mauceli, E., Z. K. Geng, W. O. Hamilton, W. W. Johnson, S. Merkowitz, A. Morse, B. Price, and N. Solomonson (1996), Phys. Rev. D **54**, 1264, arXiv:gr-qc/9609058 .

[1257] Mavromatos, N. E. (2021), in *16th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories*, arXiv:2108.03998 [hep-th] .

[1258] Mavromatos, N. E. (2021), Universe **7** (12), 480, arXiv:2111.05675 [hep-th] .

[1259] Mavromatos, N. E. (2022), Phil. Trans. Roy. Soc. Lond. A **380**, 20210188, arXiv:2108.02152 [gr-qc] .

[1260] Mavromatos, N. E., and J. Solà Peracaula (2021), Eur. Phys. J. Plus **136** (11), 1152, arXiv:2105.02659 [hep-th] .

[1261] Mavromatos, N. E., and J. Solà Peracaula (2021), Eur. Phys. J. ST **230** (9), 9, arXiv:2012.07971 [hep-ph] .

[1262] Mazrouei, S., R. R. Ghent, W. F. Bottke, A. H. Parker, and T. M. Gernon (2019), Science **365** (6450), aaw9895.

[1263] Mazumdar, A., S. Mohanty, and P. Parashari (2020), arXiv:2011.13685 [hep-ph] .

[1264] McDonough, E., M.-X. Lin, J. C. Hill, W. Hu, and S. Zhou (2021), arXiv:2112.09128 [astro-ph.CO] .

[1265] McEwen, A. S., J. M. Moore, and E. M. Shoemaker (1997), Journal of Geophysical Research: planets **102** (E4), 9231.

[1266] McGaugh, S. (2012), Astron. J. **143**, 40, arXiv:1107.2934 [astro-ph.CO] .

[1267] McGaugh, S. S. (2005), Astrophys. J. **632**, 859, arXiv:astro-ph/0506750 .

[1268] McGaugh, S. S. (2011), Phys. Rev. Lett. **106**, 121303, [Erratum: Phys.Rev.Lett. 107, 229901 (2011)], arXiv:1102.3913 [astro-ph.CO] .

[1269] McGaugh, S. S., and W. J. G. de Blok (1998), Astrophys. J. **499**, 41, arXiv:astro-ph/9801123 .

[1270] McGaugh, S. S., J. M. Schombert, G. D. Bothun, and W. J. G. de Blok (2000), Astrophys. J. Lett. **533**, L99, arXiv:astro-ph/0003001 .

[1271] McKay, J. H., and D. L. Wiltshire (2016), Mon. Not. Roy. Astron. Soc. **457** (3), 3285, [Erratum: Mon.Not.Roy.Astron.Soc. 463, 3113 (2016)], arXiv:1503.04192 [astro-ph.CO] .

[1272] McKernan, B., K. Ford, I. Bartos, M. Graham, W. Lyra, S. Marka, Z. Marka, N. Ross, D. Stern, and Y. Yang (2019), Astrophys. J. Lett. **884** (2), L50, arXiv:1907.03746 [astro-ph.HE] .

[1273] McLaughlin, M. A. (2013), Class. Quant. Grav. **30**, 224008, arXiv:1310.0758 [astro-ph.IM] .

[1274] Mead, C. A. (1964), Phys. Rev. **135**, B849.

[1275] Mehrabi, A., and J. L. Said (2022), arXiv:2203.01817 [astro-ph.CO] .

[1276] Mei, J., *et al.* (TianQin) (2020), 10.1093/ptep/ptaa114, arXiv:2008.10332 [gr-qc] .

[1277] Mei, S., J. P. Blakeslee, J. L. Tonry, A. Jordan, E. W. Peng, P. Cote, L. Ferrarese, D. Merritt, M. Milosavljevic, and M. J. West (2005), Astrophys. J. Suppl. **156**, 113, arXiv:astro-ph/0501325 .

[1278] Meiksin, A., M. J. White, and J. A. Peacock (1999), Mon. Not. Roy. Astron. Soc. **304**, 851, arXiv:astro-ph/9812214 .

[1279] Melendez, J., L. Casagrande, I. Ramirez, M. Asplund, and W. Schuster (2010), Astron. Astrophys. **515**, L3, arXiv:1005.2944 [astro-ph.SR] .

[1280] Melia, F. (2017), Mon. Not. Roy. Astron. Soc. **464** (2), 1966, arXiv:1609.08576 [astro-ph.CO] .

[1281] Melia, F. (2019), Mon. Not. Roy. Astron. Soc. **489** (1), 517, arXiv:1907.13127 [astro-ph.CO] .

[1282] Melnick, J., R. Terlevich, and E. Terlevich (2000), Mon. Not. Roy. Astron. Soc. **311**, 629, arXiv:astro-ph/9908346 .

[1283] Menanteau, F., *et al.* (2012), Astrophys. J. **748**, 7, arXiv:1109.0953 [astro-ph.CO] .

[1284] Merloni, A., *et al.* (eROSITA) (2012), arXiv:1209.3114 [astro-ph.HE] .

[1285] Messenger, C., and J. Read (2012), Phys. Rev. Lett. **108**, 091101, arXiv:1107.5725 [gr-qc] .

[1286] Miao, H., and Z. Huang (2018), Astrophys. J. **868** (1), 20, arXiv:1803.07320 [astro-ph.CO] .

[1287] Milaković, D., C.-C. Lee, R. F. Carswell, J. K. Webb, P. Molaro, and L. Pasquini (2020), Mon. Not. Roy. Astron. Soc. **500** (1), 1, arXiv:2008.10619 [astro-ph.CO] .

[1288] Milgrom, M. (1983), Astrophys. J. **270**, 371.

[1289] Miller, L., *et al.* (2013), Mon. Not. Roy. Astron. Soc. **429**, 2858, arXiv:1210.8201 [astro-ph.CO] .

[1290] Millon, M., *et al.* (2020), Astron. Astrophys. **639**, A101, arXiv:1912.08027 [astro-ph.CO] .

[1291] Minami, Y. (2020), PTEP **2020** (6), 063E01, arXiv:2002.03572 [astro-ph.CO] .

[1292] Minami, Y., and E. Komatsu (2020), Phys. Rev. Lett. **125** (22), 221301, arXiv:2011.11254 [astro-ph.CO] .

[1293] Minami, Y., and E. Komatsu (2020), PTEP **2020** (10), 103E02, arXiv:2006.15982 [astro-ph.CO] .

[1294] Minami, Y., H. Ochi, K. Ichiki, N. Katayama, E. Komatsu, and T. Matsumura (2019), PTEP **2019** (8), 083E02, arXiv:1904.12440 [astro-ph.CO] .

[1295] Miyatake, H., *et al.* (2021), arXiv:2111.02419 [astro-ph.CO] .

[1296] Miyazaki, S., *et al.* (2015), Astrophys. J. **807** (1), 22, arXiv:1504.06974 [astro-ph.CO] .

[1297] Mohammad, F. G., *et al.* (2018), Astron. Astrophys. **619**, A17, arXiv:1807.05999 [astro-ph.CO] .

[1298] Mohammadi, A., A. F. Ali, T. Golanbari, A. Aghamohammadi, K. Saaidi, and M. Faizal (2017), Annals Phys. **385**, 214, arXiv:1505.04392 [gr-qc] .

[1299] Molaro, P., *et al.* (2013), Astron. Astrophys. **555**, A68, arXiv:1305.1884 [astro-ph.CO] .

[1300] Monaghan, J. J. (1992), Ann. Rev. Astron. Astrophys. **30**, 543.

[1301] Monteserin, C., R. Barreiro, P. Vielva, E. Martinez-Gonzalez, M. Hobson, and A. Lasenby (2008), Mon. Not. Roy. Astron. Soc. **387**, 209, arXiv:0706.4289 [astro-ph] .

[1302] Moore, B. (1994), Nature **370**, 629.

[1303] Moore, B., S. Ghigna, F. Governato, G. Lake, T. R. Quinn, J. Stadel, and P. Tozzi (1999), Astrophys. J. Lett. **524**, L19, arXiv:astro-ph/9907411 .

[1304] Moore, B., T. R. Quinn, F. Governato, J. Stadel, and G. Lake (1999), Mon. Not. Roy. Astron. Soc. **310**, 1147, arXiv:astro-ph/9903164 .

[1305] Moreno-Pulido, C., and J. S. Peracaula (2022), arXiv:2201.05827 [gr-qc] .

[1306] Moreno-Pulido, C., and J. Solà (2020), Eur. Phys. J. C **80** (8), 692, arXiv:2005.03164 [gr-qc] .

[1307] Moresco, M., L. Pozzetti, A. Cimatti, R. Jimenez, C. Maraston, L. Verde, D. Thomas, A. Citro, R. Tojeiro, and D. Wilkinson (2016), JCAP **05**, 014, arXiv:1601.01701 [astro-ph.CO] .

[1308] Moresco, M., *et al.* (2012), JCAP **08**, 006, arXiv:1201.3609 [astro-ph.CO] .

[1309] Moresco, M., *et al.* (2022), arXiv:2201.07241 [astro-ph.CO] .

[1310] Mori, K., and M. Kusakabe (2019), Phys. Rev. D **99** (8), 083013, arXiv:1901.03943 [astro-ph.CO] .

[1311] Mortonson, M. J., W. Hu, and D. Huterer (2009), Phys. Rev. D **80**, 067301, arXiv:0908.1408 [astro-ph.CO] .

[1312] Mortsell, E., A. Goobar, J. Johansson, and S. Dhawan (2021), arXiv:2105.11461 [astro-ph.CO] .

[1313] Mortsell, E., A. Goobar, J. Johansson, and S. Dhawan (2021), arXiv:2106.09400 [astro-ph.CO] .

[1314] Mosbech, M. R., C. Boehm, S. Hannestad, O. Mena, J. Stadler, and Y. Y. Y. Wong (2021), JCAP **03**, 066, arXiv:2011.04206 [astro-ph.CO] .

[1315] Moshafi, H., S. Baghram, and N. Khosravi (2021), Phys. Rev. D **104** (6), 063506, arXiv:2012.14377 [astro-ph.CO] .

[1316] Moss, A., E. Copeland, S. Bamford, and T. Clarke (2021), arXiv:2109.14848 [astro-ph.CO] .

[1317] Moss, G. E., L. R. Miller, and R. L. Forward (1971), Appl. Opt. **10**, 2495.

[1318] Mota, D. F., and D. J. Shaw (2006), Phys. Rev. Lett. **97**, 151102, arXiv:hep-ph/0606204 .

[1319] Mota, D. F., and D. J. Shaw (2007), Phys. Rev. D **75**, 063501, arXiv:hep-ph/0608078 .

[1320] Motloch, P., and W. Hu (2020), Phys. Rev. D **101** (8), 083515, arXiv:1912.06601 [astro-ph.CO] .

[1321] Mroczkowski, T., *et al.* (2019), Space Sci. Rev. **215** (1), 17, arXiv:1811.02310 [astro-ph.CO] .

[1322] Mörtsell, E., and S. Dhawan (2018), JCAP **1809** (09), 025, arXiv:1801.07260 [astro-ph.CO] .

[1323] Muñoz, J. B., and A. Loeb (2018), Nature **557** (7707), 684, arXiv:1802.10094 [astro-ph.CO] .

[1324] Muir, J., S. Adhikari, and D. Huterer (2018), Phys. Rev. D **98** (2), 023521, arXiv:1806.02354 [astro-ph.CO] .

[1325] Mukhanov, V. F., H. A. Feldman, and R. H. Brandenberger (1992), Phys. Rept. **215**, 203.

[1326] Mukherjee, P., and N. Banerjee (2021), Phys. Rev. D **103** (12), 123530, arXiv:2105.09995 [astro-ph.CO] .

[1327] Mukherjee, S., A. Ghosh, M. J. Graham, C. Karathanasis, M. M. Kasliwal, I. Magaña Hernandez, S. M. Nissanke, A. Silvestri, and B. D. Wandelt (2020), arXiv:2009.14199 [astro-ph.CO] .

[1328] Mukherjee, S., A. Krolewski, B. D. Wandelt, and J. Silk (2022), arXiv:2203.03643 [astro-ph.CO] .

[1329] Mukherjee, S., and T. Souradeep (2016), Phys. Rev. Lett. **116** (22), 221301, arXiv:1509.06736 [astro-ph.CO] .

[1330] Muller, J., J. G. Williams, and S. G. Turyshev (2008), Astrophys. Space Sci. Libr. **349**, 457, arXiv:gr-qc/0509114 .

[1331] Murgia, R., G. F. Abellán, and V. Poulin (2021), Phys. Rev. D **103** (6), 063502, arXiv:2009.10733 [astro-ph.CO] .

[1332] Murphy, M. T., J. K. Webb, and V. V. Flambaum (2003), Mon. Not. Roy. Astron. Soc. **345**, 609, arXiv:astro-ph/0306483 .

[1333] Murphy, M. T., J. K. Webb, V. V. Flambaum, V. A. Dzuba, C. W. Churchill, J. X. Prochaska, J. D. Barrow, and A. M. Wolfe (2001), Mon. Not. Roy. Astron. Soc. **327**, 1208, arXiv:astro-ph/0012419 .

[1334] Nadathur, S., S. Hotchkiss, and S. Sarkar (2012), JCAP **06**, 042, arXiv:1109.4126 [astro-ph.CO] .

[1335] Nadathur, S., M. Lavinto, S. Hotchkiss, and S. Räsänen (2014), Phys. Rev. D **90** (10), 103510, arXiv:1408.4720 [astro-ph.CO] .

[1336] Naidoo, K., A. Benoit-Lévy, and O. Lahav (2016), Mon. Not. Roy. Astron. Soc. **459** (1), L71, arXiv:1512.02694 [astro-ph.CO] .

[1337] Nakama, T., J. Chluba, and M. Kamionkowski (2017), Phys. Rev. D **95** (12), 121302, arXiv:1703.10559 [astro-ph.CO] .

[1338] Namikawa, T., *et al.* (2020), Phys. Rev. D **101** (8), 083527, arXiv:2001.10465 [astro-ph.CO] .

[1339] Navarro, J. F., V. R. Eke, and C. S. Frenk (1996), Mon. Not. Roy. Astron. Soc. **283**, L72, arXiv:astro-ph/9610187 .

[1340] Navarro, J. F., C. S. Frenk, and S. D. M. White (1996), Astrophys. J. **462**, 563, arXiv:astro-ph/9508025 .

[1341] Navarro, J. F., C. S. Frenk, and S. D. M. White (1997), Astrophys. J. **490**, 493, arXiv:astro-ph/9611107 .

[1342] Nesseris, S., S. Basilakos, E. N. Saridakis, and L. Perivolaropoulos (2013), Phys. Rev. **D88**, 103010, arXiv:1308.6142 [astro-ph.CO] .

[1343] Nesseris, S., and J. Garcia-Bellido (2013), JCAP **08**, 036, arXiv:1210.7652 [astro-ph.CO] .

[1344] Nesseris, S., G. Pantazis, and L. Perivolaropoulos (2017), Phys. Rev. D **96** (2), 023542, arXiv:1703.10538 [astro-ph.CO] .

[1345] Nesseris, S., and L. Perivolaropoulos (2005), Phys. Rev. D **72**, 123519, arXiv:astro-ph/0511040 .

[1346] Nesseris, S., and L. Perivolaropoulos (2007), Phys. Rev. D **75**, 023517, arXiv:astro-ph/0611238 .

[1347] Nesseris, S., and D. Sapone (2015), Phys. Rev. D **92** (2), 023013, arXiv:1505.06601 [astro-ph.CO] .

[1348] Nesseris, S., D. Sapone, and S. Sypsas (2020), Phys. Dark Univ. **27**, 100413, arXiv:1907.05608 [astro-ph.CO] .

[1349] Nesseris, S., *et al.* (Euclid) (2021), 10.1051/0004-6361/202142503, arXiv:2110.11421 [astro-ph.CO] .

[1350] Neveux, R., *et al.* (2020), Mon. Not. Roy. Astron. Soc. **499** (1), 210, arXiv:2007.08999 [astro-ph.CO] .

[1351] Ni, W.-T. (1977), Phys. Rev. Lett. **38**, 301.

[1352] Niedermann, F., and M. S. Sloth (2020), Phys. Rev. D **102** (6), 063527, arXiv:2006.06686 [astro-ph.CO] .

[1353] Niedermann, F., and M. S. Sloth (2021), arXiv:2112.00759 [hep-ph] .

[1354] Niedermann, F., and M. S. Sloth (2021), Phys. Rev. D **103** (4), L041303, arXiv:1910.10739 [astro-ph.CO] .

[1355] Niedermann, F., and M. S. Sloth (2021), Phys. Rev. D **103** (10), 103537, arXiv:2009.00006 [astro-ph.CO] .

[1356] Niedermann, F., and M. S. Sloth (2022), Phys. Rev. D **105** (6), 063509, arXiv:2112.00770 [hep-ph] .

[1357] Nissanke, S., D. E. Holz, N. Dalal, S. A. Hughes, J. L. Sievers, and C. M. Hirata (2013), arXiv:1307.2638 [astro-ph.CO] .

[1358] Nissanke, S., D. E. Holz, S. A. Hughes, N. Dalal, and J. L. Sievers (2010), Astrophys. J. **725**, 496, arXiv:0904.1017 [astro-ph.CO] .

[1359] Nojiri, S., and S. D. Odintsov (2002), Phys. Lett. B **548**, 215, arXiv:hep-th/0209066 .

[1360] Nojiri, S., and S. D. Odintsov (2006), eConf **C0602061**, 06, arXiv:hep-th/0601213 .

[1361] Nojiri, S., and S. D. Odintsov (2007), J. Phys. A **40**, 6725, arXiv:hep-th/0610164 .

[1362] Nojiri, S., and S. D. Odintsov (2011), Phys. Rept. **505**, 59, arXiv:1011.0544 [gr-qc] .

[1363] Nojiri, S., S. D. Odintsov, D. Saez-Chillon Gomez, and G. S. Sharov (2021), Phys. Dark Univ. **32**, 100837, arXiv:2103.05304 [gr-qc] .

[1364] Normann, B. D., and I. H. Brevik (2021), Mod. Phys. Lett. A **36** (27), 2150198, arXiv:2107.13533 [gr-qc] .

[1365] Nozari, K., P. Pedram, and M. Molkara (2012), Int. J. Theor. Phys. **51**, 1268, arXiv:1111.2204 [gr-qc] .

[1366] Ntampaka, M., K. Rines, and H. Trac (2019), Astrophys. J. **880**, 154, arXiv:1906.07729 [astro-ph.CO] .

[1367] Nunes, R. C. (2018), JCAP **05**, 052, arXiv:1802.02281 [gr-qc] .

[1368] Nunes, R. C., and E. Di Valentino (2021), Phys. Rev. D **104** (6), 063529, arXiv:2107.09151 [astro-ph.CO] .

[1369] Nunes, R. C., and S. Vagnozzi (2021), Mon. Not. Roy. Astron. Soc. **505** (4), 5427, arXiv:2106.01208 [astro-ph.CO] .

[1370] Nunes, R. C., S. Vagnozzi, S. Kumar, E. Di Valentino, and O. Mena (2022), arXiv:2203.08093 [astro-ph.CO] .

[1371] Nunes, R. C., S. K. Yadav, J. F. Jesus, and A. Bernui (2020), Mon. Not. Roy. Astron. Soc. **497** (2), 2133, arXiv:2002.09293 [astro-ph.CO] .

[1372] Nusser, A., and M. Davis (2011), Astrophys. J. **736**, 93, arXiv:1101.1650 [astro-ph.CO] .

[1373] Nygaard, A., T. Tram, and S. Hannestad (2021), JCAP **05**, 017, arXiv:2011.01632 [astro-ph.CO] .

[1374] Ó Colgáin, E., M. H. P. M. van Putten, and H. Yavartanoo (2019), Phys. Lett. B **793**, 126, arXiv:1807.07451 [hep-th] .

[1375] Ó Colgáin, E., and M. M. Sheikh-Jabbari (2021), Eur. Phys. J. C **81** (10), 892, arXiv:2101.08565 [astro-ph.CO] .

[1376] Obata, I. (2021), arXiv:2108.02150 [astro-ph.CO] .

[1377] Odintsov, S. D., and V. K. Oikonomou (2022), 10.1209/0295-5075/ac52dc, arXiv:2201.07647 [gr-qc] .

[1378] Odintsov, S. D., D. Sáez-Chillón Gómez, and G. S. Sharov (2021), Nucl. Phys. B **966**, 115377, arXiv:2011.03957 [gr-qc] .

[1379] O'Dwyer, M., C. J. Copi, J. M. Nagy, C. B. Netterfield, J. Ruhl, and G. D. Starkman (2019), 10.1093/mnras/staa3049, arXiv:1912.02376 [astro-ph.CO] .

[1380] Oguri, M. (2007), Astrophys. J. **660**, 1, arXiv:astro-ph/0609694 .

[1381] Oguri, M. (2019), Rept. Prog. Phys. **82** (12), 126901, arXiv:1907.06830 [astro-ph.CO] .

[1382] Okamatsu, F., T. Sekiguchi, and T. Takahashi (2021), Phys. Rev. D **104** (2), 023523, arXiv:2105.12312 [astro-ph.CO] .

[250] Okumura, T., *et al.* (2016), Publ. Astron. Soc. Jap. **68** (3), 38, arXiv:1511.08083 [astro-ph.CO] .

[251] Oldengott, I. M., T. Tram, C. Rampf, and Y. Y. Y. Wong (2017), JCAP **11**, 027, arXiv:1706.02123 [astro-ph.CO] .

[1385] Olivares E., F., *et al.* (2010), Astrophys. J. **715**, 833, arXiv:1004.2534 [astro-ph.CO] .

[1386] de Oliveira-Costa, A., M. Tegmark, M. Zaldarriaga, and A. Hamilton (2004), Phys. Rev. D **69**, 063516, arXiv:astro-ph/0307282 .

[1387] Olmez, S., V. Mandic, and X. Siemens (2010), Phys. Rev. D **81**, 104028, arXiv:1004.0890 [astro-ph.CO] .

[1388] Ooba, J., B. Ratra, and N. Sugiyama (2019), Astrophys. Space Sci. **364** (10), 176, arXiv:1802.05571 [astro-ph.CO] .

[1389] Ooguri, H., and C. Vafa (2007), Nucl. Phys. B **766**, 21, arXiv:hep-th/0605264 .

[1390] Ortiz, C. (2020), Int. J. Mod. Phys. D **29** (16), 2050115, arXiv:2011.02317 [gr-qc] .

[1391] Pacaud, F., *et al.* (XXL) (2018), Astron. Astrophys. **620**, A10, arXiv:1810.01624 [astro-ph.CO] .

[1392] Paci, F., A. Gruppuso, F. Finelli, P. Cabella, A. De Rosa, N. Mandolesi, and P. Natoli (2010), Mon. Not. Roy. Astron. Soc. **407**, 399, arXiv:1002.4745 [astro-ph.CO] .

[1393] Padmanabhan, T. (2003), Phys. Rept. **380**, 235, arXiv:hep-th/0212290 .

[1394] Padmanabhan, T. (2005), Curr. Sci. **88**, 1057, arXiv:astro-ph/0411044 .

[1395] Page, L., *et al.* (WMAP) (2003), Astrophys. J. Suppl. **148**, 233, arXiv:astro-ph/0302220 .

[1396] Paliathanasis, A. (2021), Universe **7** (7), 244, arXiv:2107.05880 [gr-qc] .

[1397] Palmese, A., C. R. Bom, S. Mucesh, and W. G. Hartley (2021), arXiv:2111.06445 [astro-ph.CO] .

[1398] Pan, S., E. N. Saridakis, and W. Yang (2018), Phys. Rev. D **98** (6), 063510, arXiv:1712.05746 [astro-ph.CO] .

[1399] Pan, S., G. S. Sharov, and W. Yang (2020), Phys. Rev. D **101** (10), 103533, arXiv:2001.03120 [astro-ph.CO] .

[1400] Pan, S., W. Yang, E. Di Valentino, E. N. Saridakis, and S. Chakraborty (2019), Phys. Rev. **D100** (10), 103520, arXiv:1907.07540 [astro-ph.CO] .

[1401] Pan, S., W. Yang, E. Di Valentino, A. Shafieloo, and S. Chakraborty (2020), JCAP **06** (06), 062, arXiv:1907.12551 [astro-ph.CO] .

[1402] Pan, S., W. Yang, and A. Paliathanasis (2020), 10.1093/mnras/staa213, arXiv:2002.03408 [astro-ph.CO] .

[1403] Pan, S., W. Yang, C. Singha, and E. N. Saridakis (2019), Phys. Rev. **D100** (8), 083539, arXiv:1903.10969 [astro-ph.CO] .

[1404] Pandey, K. L., T. Karwal, and S. Das (2020), JCAP **07**, 026, arXiv:1902.10636 [astro-ph.CO] .

[1405] Panpanich, S., P. Burikham, S. Ponglertsakul, and L. Tannukij (2021), Chin. Phys. C **45** (1), 015108, arXiv:1908.03324 [gr-qc] .

[1406] Papastergis, E., R. Giovanelli, M. P. Haynes, and F. Shankar (2015), Astron. Astrophys. **574**, A113, arXiv:1407.4665 [astro-ph.GA] .

[1407] Parker, L., and A. Raval (2000), Phys. Rev. D **62**, 083503, [Erratum: Phys.Rev.D 67, 029903 (2003)], arXiv:gr-qc/0003103 .

[1408] Parker, L., and D. A. T. Vanzella (2004), Phys. Rev. D **69**, 104009, arXiv:gr-qc/0312108 .

[1409] Paul, A., A. Chatterjee, A. Ghoshal, and S. Pal (2021), JCAP **10**, 017, arXiv:2104.04760 [hep-ph] .

[1410] Pavon, D., and W. Zimdahl (2005), Phys. Lett. B **628**, 206, arXiv:gr-qc/0505020 .

[1411] Pawlowski, M. S. (2018), Mod. Phys. Lett. A **33** (06), 1830004, arXiv:1802.02579 [astro-ph.GA] .

[1412] Pawlowski, M. S., J. Pflamm-Altenburg, and P. Kroupa (2012), Mon. Not. Roy. Astron. Soc. **423**, 1109, arXiv:1204.5176 [astro-ph.GA] .

[1413] Pawlowski, M. S., *et al.* (2014), Mon. Not. Roy. Astron. Soc. **442** (3), 2362, arXiv:1406.1799 [astro-ph.GA] .

[1414] Peacock, J. A., *et al.* (2001), Nature **410**, 169, arXiv:astro-ph/0103143 .

[1415] Peebles, P. (1968), Astrophys. J. **153**, 1.

[1416] Peebles, P. (2007), Nuovo Cim. B **122**, 1035, arXiv:0712.2757 [astro-ph] .

[1417] Peebles, P. J. E. (1980), *The large-scale structure of the universe*.

[1418] Peebles, P. J. E. (1984), Astrophys. J. **284**, 439.

[1419] Peebles, P. J. E. (2001), Astrophys. J. **557**, 495, arXiv:astro-ph/0101127 [astro-ph] .

[1420] Peebles, P. J. E. (2005), Nucl. Phys. B Proc. Suppl. **138**, 5, arXiv:astro-ph/0311435 .

[1421] Peebles, P. J. E., R. A. Daly, and R. Juszkiewicz (1989), ApJ **347**, 563.

[1422] Peebles, P. J. E., and B. Ratra (2003), Rev. Mod. Phys. **75**, 559, [,592(2002)], arXiv:astro-ph/0207347 [astro-ph] .

[1423] Peiris, H. V. (2014), IAU Symp. **306**, 124, arXiv:1410.3837 [astro-ph.CO] .

[1424] Peirone, S., G. Benevento, N. Frusciante, and S. Tsujikawa (2019), Phys. Rev. D **100** (6), 063540, arXiv:1905.05166 [astro-ph.CO] .

[1425] Peracaula, J. S. (2022), arXiv:2203.13757 [gr-qc] .

[1426] Perez, A., D. Sudarsky, and E. Wilson-Ewing (2021), Gen. Rel. Grav. **53** (1), 7, arXiv:2001.07536 [astro-ph.CO] .

[1427] Perico, E. L. D., J. A. S. Lima, S. Basilakos, and J. Sola (2013), Phys. Rev. D **88**, 063531, arXiv:1306.0591 [astro-ph.CO] .

[1428] Perivolaropoulos, L. (2003), Phys. Rev. D **67**, 123516, arXiv:hep-ph/0301237 .

[1429] Perivolaropoulos, L. (2005), JCAP **10**, 001, arXiv:astro-ph/0504582 .

[1430] Perivolaropoulos, L. (2006), AIP Conf. Proc. **848** (1), 698, arXiv:astro-ph/0601014 .

[1431] Perivolaropoulos, L. (2008), arXiv:0811.4684 [astro-ph] .

[1432] Perivolaropoulos, L. (2011), J. Cosmol. **15**, 6054, arXiv:1104.0539 [astro-ph.CO] .

[1433] Perivolaropoulos, L. (2014), Galaxies **2**, 22, arXiv:1401.5044 [astro-ph.CO] .

[1434] Perivolaropoulos, L. (2017), Phys. Rev. D **95** (10), 103523, arXiv:1704.05681 [gr-qc] .

[1435] Perivolaropoulos, L. (2017), Phys. Rev. D **95** (8), 084050, arXiv:1611.07293 [gr-qc] .

[1436] Perivolaropoulos, L. (2021), "H_0 Crisis: Systematics of distance calibrators or The End of ΛCDM?" https://www.youtube.com/watch?v=RQ0DU88A2ik&t=15s.

[1437] Perivolaropoulos, L. (2021), "The tensions of the ΛCDM and a gravitational transition," https://www.youtube.com/watch?v=GKubKmPXDM8.

[1438] Perivolaropoulos, L. (2022), arXiv:2201.08997 [astro-ph.EP] .

[1439] Perivolaropoulos, L., and L. Kazantzidis (2019), Int. J. Mod. Phys. D **28** (05), 1942001, arXiv:1904.09462 [gr-qc] .

[1440] Perivolaropoulos, L., and F. Skara (2020), Phys. Rev. D **102** (10), 104034, arXiv:2009.05640 [gr-qc] .

[1441] Perivolaropoulos, L., and F. Skara (2021), Phys. Rev. D **104** (12), 123511, arXiv:2109.04406 [astro-ph.CO] .

[1442] Perivolaropoulos, L., and F. Skara (2022), arXiv:2203.10374 [astro-ph.CO] .

[1443] Perivolaropoulos, L., and C. Sourdis (2002), Phys. Rev. D **66**, 084018, arXiv:hep-ph/0204155 .

[1444] Perko, A., L. Senatore, E. Jennings, and R. H. Wechsler (2016), arXiv:1610.09321 [astro-ph.CO] .

[1445] Perlmutter, S., *et al.* (Supernova Cosmology Project) (1999), Astrophys.J. **517**, 565, arXiv:astro-ph/9812133 [astro-ph] .

[1446] Perryman, M. A. C., K. S. de Boer, G. Gilmore, E. Hog, M. G. Lattanzi, L. Lindegren, X. Luri, F. Mignard, O. Pace, and P. T. de Zeeuw (2001), Astron. Astrophys. **369**, 339, arXiv:astro-ph/0101235 .

[1447] Pesce, D., *et al.* (2020), Astrophys. J. Lett. **891** (1), L1, arXiv:2001.09213 [astro-ph.CO] .

[1448] Peterson, E. R., *et al.* (2021), arXiv:2110.03487 [astro-ph.CO] .

[1449] Pettorino, V. (2013), Phys. Rev. D **88**, 063519, arXiv:1305.7457 [astro-ph.CO] .

[1450] Pettorino, V., and L. Amendola (2015), Phys. Lett. B **742**, 353, arXiv:1408.2224 [astro-ph.CO] .

[1451] Pettorino, V., L. Amendola, and C. Wetterich (2013), Phys. Rev. D **87**, 083009, arXiv:1301.5279 [astro-ph.CO] .

[269] Pezzotta, A., *et al.* (2017), Astron. Astrophys. **604**, A33, arXiv:1612.05645 [astro-ph.CO] .

[1453] Philcox, O. H. E., and M. M. Ivanov (2022), Phys. Rev. D **105** (4), 043517, arXiv:2112.04515 [astro-ph.CO] .

[1454] Philcox, O. H. E., B. D. Sherwin, G. S. Farren, and E. J. Baxter (2021), Phys. Rev. D **103** (2), 023538, arXiv:2008.08084 [astro-ph.CO] .

[1455] Pierre, M., *et al.* (2016), Astron. Astrophys. **592**, A1, arXiv:1512.04317 [astro-ph.CO] .

[1456] Pietrzyński, G., D. Graczyk, A. Gallenne, W. Gieren, I. B. Thompson, B. Pilecki, P. Karczmarek, M. Gorski, K. Suchomska, M. Taormina, B. Zgirski, P. Wielgorski, Z. Kolaczkowski, P. Konorski, S. Villanova, N. Nardetto, P. Kervella, F. Bresolin, R. P. Kudritzki, J. Storm, R. Smolec, and W. Narloch (2019), Nature **567** (7747), 200, arXiv:1903.08096 [astro-ph.GA] .

[1457] Pietrzyński, G., *et al.* (2013), Nature **495**, 76, arXiv:1303.2063 [astro-ph.GA] .

[1458] Pillepich, A., C. Porciani, and T. H. Reiprich (2012), Mon. Not. Roy. Astron. Soc. **422**, 44, arXiv:1111.6587 [astro-ph.CO] .

[1459] Pinho, A. M., S. Casas, and L. Amendola (2018), JCAP **11**, 027, arXiv:1805.00027 [astro-ph.CO] .

[1460] Pinsonneault, M. H., G. Steigman, T. P. Walker, and V. K. Narayanans (2002), Astrophys. J. **574**, 398, arXiv:astro-ph/0105439 .

[1461] Pitjeva, E. V., and N. P. Pitjev (2013), Mon. Not. Roy. Astron. Soc. **432**, 3431, arXiv:1306.3043 [astro-ph.EP] .

[272] Pitrou, C., A. Coc, J.-P. Uzan, and E. Vangioni (2018), Phys. Rept. **754**, 1, arXiv:1801.08023 [astro-ph.CO] .

[1463] Pizzella, G. (1997), Class. Quant. Grav. **14**, 1481.

[1464] Pizzone, R. G., R. Sparta, C. A. Bertulani, C. Spitaleri, M. La Cognata, J. Lalmansingh, L. Lamia, A. Mukhamedzhanov, and A. Tumino (2014), Astrophys. J. **786**, 112, arXiv:1403.4909 [nucl-ex] .

[1465] Pizzuti, L., I. D. Saltas, S. Casas, L. Amendola, and A. Biviano (2019), Mon. Not. Roy. Astron. Soc. **486** (1), 596, arXiv:1901.01961 [astro-ph.CO] .

[1466] Plionis, M., R. Terlevich, S. Basilakos, F. Bresolin, E. Terlevich, J. Melnick, and R. Chavez (2011), Mon. Not. Roy. Astron. Soc. **416**, 2981, arXiv:1106.4558 [astro-ph.CO] .

[1467] Pogosian, L., and A. Silvestri (2008), Phys. Rev. D **77**, 023503, [Erratum: Phys.Rev.D 81, 049901 (2010)], arXiv:0709.0296 [astro-ph] .

[1468] Pogosian, L., A. Silvestri, K. Koyama, and G.-B. Zhao (2010), Phys. Rev. D **81**, 104023, arXiv:1002.2382 [astro-ph.CO] .

[1469] Pogosian, L., G.-B. Zhao, and K. Jedamzik (2020), Astrophys. J. Lett. **904** (2), L17, arXiv:2009.08455 [astro-ph.CO] .

[1470] Polarski, D., and R. Gannouji (2008), Phys. Lett. B **660**, 439, arXiv:0710.1510 [astro-ph] .

[1471] Porto, R. A., L. Senatore, and M. Zaldarriaga (2014), JCAP **05**, 022, arXiv:1311.2168 [astro-ph.CO] .

[1472] Poulin, V., K. K. Boddy, S. Bird, and M. Kamionkowski (2018), Phys. Rev. D **97** (12), 123504, arXiv:1803.02474 [astro-ph.CO] .

[1473] Poulin, V., and P. D. Serpico (2015), Phys. Rev. Lett. **114** (9), 091101, arXiv:1502.01250 [astro-ph.CO] .

[1474] Poulin, V., P. D. Serpico, and J. Lesgourgues (2016), JCAP **1608**, 036, arXiv:1606.02073 [astro-ph.CO] .

[1475] Poulin, V., T. L. Smith, and A. Bartlett (2021), Phys. Rev. D **104** (12), 123550, arXiv:2109.06229 [astro-ph.CO] .

[1476] Poulin, V., T. L. Smith, D. Grin, T. Karwal, and M. Kamionkowski (2018), Phys. Rev. D **98** (8), 083525, arXiv:1806.10608 [astro-ph.CO] .

[301] Poulin, V., T. L. Smith, T. Karwal, and M. Kamionkowski (2019), Phys. Rev. Lett. **122** (22), 221301, arXiv:1811.04083 [astro-ph.CO] .

[1478] Pourtsidou, A., C. Skordis, and E. Copeland (2013), Phys. Rev. D **88** (8), 083505, arXiv:1307.0458 [astro-ph.CO] .

[1479] Pourtsidou, A., and T. Tram (2016), Phys. Rev. D **94** (4), 043518, arXiv:1604.04222 [astro-ph.CO] .

[1480] Pranav, P. (2021), arXiv:2101.02237 [astro-ph.CO] .

[1481] Pranav, P., R. J. Adler, T. Buchert, H. Edelsbrunner, B. J. T. Jones, A. Schwartzman, H. Wagner, and R. van de Weygaert (2019), Astron. Astrophys. **627**, A163, arXiv:1812.07678 [astro-ph.CO] .

[1482] Predehl, P., *et al.* (2010), Proc. SPIE Int. Soc. Opt. Eng. **7732**, 77320U, arXiv:1001.2502 [astro-ph.CO] .

[1483] Press, W. H., and P. Schechter (1974), Astrophys. J. **187**, 425.

[1484] Prusti, T., *et al.* (Gaia) (2016), Astron. Astrophys. **595** (Gaia Data Release 1), A1, arXiv:1609.04153 [astro-ph.IM] .

[1485] Punturo, M., *et al.* (2010), Class. Quant. Grav. **27**, 194002.

[1486] van Putten, M. H. P. M. (2017), Astrophys. J. **848** (1), 28, arXiv:1709.05944 [astro-ph.GA] .

[1487] van Putten, M. H. P. M. (2019), Monthly Notices of the Royal Astronomical Society: Letters **491** (1), L6.

[1488] Qi, J.-Z., Y. Cui, W.-H. Hu, J.-F. Zhang, J.-L. Cui, and X. Zhang (2022), arXiv:2202.01396 [astro-ph.CO] .

[1489] Qi, J.-Z., J.-W. Zhao, S. Cao, M. Biesiada, and Y. Liu (2021), Mon. Not. Roy. Astron. Soc. **503** (2), 2179, arXiv:2011.00713 [astro-ph.CO] .

[1490] Rácz, G., L. Dobos, R. Beck, I. Szapudi, and I. Csabai (2017), Mon. Not. Roy. Astron. Soc. **469** (1), L1, arXiv:1607.08797 [astro-ph.CO] .

[1491] Raichoor, A., *et al.* (2020), Mon. Not. Roy. Astron. Soc. **500** (3), 3254, arXiv:2007.09007 [astro-ph.CO] .

[1492] Rameez, M., and S. Sarkar (2021), Class. Quant. Grav. **38** (15), 154005, arXiv:1911.06456 [astro-ph.CO] .

[1493] Rapetti, D., S. W. Allen, A. Mantz, and H. Ebeling (2009), Mon. Not. Roy. Astron. Soc. **400**, 699, arXiv:0812.2259 [astro-ph] .

[1494] Rashkovetskyi, M., J. B. Muñoz, D. J. Eisenstein, and C. Dvorkin (2021), Phys. Rev. D **104** (10), 103517, arXiv:2108.02747 [astro-ph.CO] .

[1495] Rasmussen, C. E., and C. K. I. Williams (2005), *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)* (The MIT Press).

[1496] Rassat, A., J.-L. Starck, P. Paykari, F. Sureau, and J. Bobin (2014), JCAP **08**, 006, arXiv:1405.1844 [astro-ph.CO] .

[1497] Ratra, B., and P. Peebles (1988), Phys. Rev. D **37**, 3406.

[1498] Read, J. I., M. I. Wilkinson, N. W. Evans, G. Gilmore, and J. T. Kleyna (2006), Mon. Not. Roy. Astron. Soc. **367**, 387, arXiv:astro-ph/0511759 .

[1499] Refregier, A. (EUCLID) (2009), Exper. Astron. **23**, 17, arXiv:0802.2522 [astro-ph] .

[1500] Refsdal, S. (1964), Mon. Not. Roy. Astron. Soc. **128**, 307.

[1501] Reid, B. A., *et al.* (2012), Mon. Not. Roy. Astron. Soc. **426**, 2719, arXiv:1203.6641 [astro-ph.CO] .

[1502] Reid, M., D. Pesce, and A. Riess (2019), Astrophys. J. Lett. **886** (2), L27, arXiv:1908.05625 [astro-ph.GA] .

[1503] Reid, M. J., J. A. Braatz, J. J. Condon, L. J. Greenhill, C. Henkel, and K. Y. Lo (2009), Astrophys. J. **695**, 287, arXiv:0811.4345 [astro-ph] .

[1504] Reid, M. J., J. A. Braatz, J. J. Condon, K. Y. Lo, C. Y. Kuo, C. M. V. Impellizzeri, and C. Henkel (2013), Astrophys. J. **767**, 154, arXiv:1207.7292 [astro-ph.CO] .

[1505] Reitze, D., *et al.* (2019), Bull. Am. Astron. Soc. **51** (7), 035, arXiv:1907.04833 [astro-ph.IM] .

[1506] Reitze, D., *et al.* (2019), Bull. Am. Astron. Soc. **51**, 141, arXiv:1903.04615 [astro-ph.IM] .

[1507] Ren, X., T. H. T. Wong, Y.-F. Cai, and E. N. Saridakis (2021), Phys. Dark Univ. **32**, 100812, arXiv:2103.01260 [astro-ph.CO] .

[1508] Ren, X., S.-F. Yan, Y. Zhao, Y.-F. Cai, and E. N. Saridakis (2022), arXiv:2203.01926 [astro-ph.CO] .

[1509] Renevey, C., R. McManus, C. Dalang, and L. Lombriser (2021), arXiv:2106.05678 [gr-qc] .

[1510] Renk, J., M. Zumalacárregui, F. Montanari, and A. Barreira (2017), JCAP **1710** (10), 020, arXiv:1707.02263 [astro-ph.CO] .

[1511] Renzi, F., N. B. Hogg, and W. Giarè (2021), arXiv:2112.05701 [astro-ph.CO] .

[1512] Renzi, F., and A. Silvestri (2020), arXiv:2011.10559 [astro-ph.CO] .

[1513] Rezaei, M., M. Malekjani, S. Basilakos, A. Mehrabi, and D. F. Mota (2017), Astrophys. J. **843** (1), 65, arXiv:1706.02537 [astro-ph.CO] .

[1514] Rezaei, M., T. Naderi, M. Malekjani, and A. Mehrabi (2020), Eur. Phys. J. C **80** (5), 374, arXiv:2004.08168 [astro-ph.CO] .

[1515] Rezaei, M., J. Solà Peracaula, and M. Malekjani (2021), Mon. Not. Roy. Astron. Soc. **509** (2), 2593, arXiv:2108.06255 [astro-ph.CO] .

[1516] Richard, O., G. Michaud, and J. Richer (2005), Astrophys. J. **619**, 538, arXiv:astro-ph/0409672 .

[1517] Richardson, D., R. L. Jenkins, J. Wright, and L. Maddox (2014), Astron. J. **147**, 118, arXiv:1403.5755 [astro-ph.SR] .

[1518] Rider, A. D., D. C. Moore, C. P. Blakemore, M. Louis, M. Lu, and G. Gratta (2016), Phys. Rev. Lett. **117** (10), 101101, arXiv:1604.04908 [hep-ex] .

[1519] Riess, A. G. (2019), Nature Rev. Phys. **2** (1), 10, arXiv:2001.03624 [astro-ph.CO] .

[1520] Riess, A. G., S. Casertano, J. Anderson, J. Mackenty, and A. V. Filippenko (2014), Astrophys. J. **785**, 161, arXiv:1401.0484 [astro-ph.IM] .

[1521] Riess, A. G., S. Casertano, W. Yuan, J. B. Bowers, L. Macri, J. C. Zinn, and D. Scolnic (2021), Astrophys. J. Lett. **908** (1), L6, arXiv:2012.08534 [astro-ph.CO] .

[306] Riess, A. G., S. Casertano, W. Yuan, L. M. Macri, and D. Scolnic (2019), Astrophys. J. **876** (1), 85, arXiv:1903.07603 [astro-ph.CO] .

[308] Riess, A. G., L. Macri, S. Casertano, H. Lampeitl, H. C. Ferguson, A. V. Filippenko, S. W. Jha, W. Li, and R. Chornock (2011), Astrophys. J. **730**, 119, [Erratum: Astrophys.J. 732, 129 (2011)], arXiv:1103.2976 [astro-ph.CO] .

[1524] Riess, A. G., *et al.* (Supernova Search Team) (1998), Astron.J. **116**, 1009, arXiv:astro-ph/9805201 [astro-ph] .

[1525] Riess, A. G., *et al.* (2009), Astrophys. J. **699**, 539, arXiv:0905.0695 [astro-ph.CO] .

[1526] Riess, A. G., *et al.* (2016), Astrophys. J. **826** (1), 56, arXiv:1604.01424 [astro-ph.CO] .

[311] Riess, A. G., *et al.* (2018), Astrophys. J. **861** (2), 126, arXiv:1804.10655 [astro-ph.CO] .

[1528] Riess, A. G., *et al.* (2018), Astrophys. J. **855** (2), 136, arXiv:1801.01120 [astro-ph.SR] .

[1529] Riess, A. G., *et al.* (2021), arXiv:2112.04510 [astro-ph.CO] .

[1530] Rigault, M., *et al.* (Nearby Supernova Factory) (2020), Astron. Astrophys. **644**, A176, arXiv:1806.03849 [astro-ph.CO] .

[1531] Risaliti, G., and E. Lusso (2015), Astrophys. J. **815**, 33, arXiv:1505.07118 [astro-ph.CO] .

[1532] Risaliti, G., and E. Lusso (2019), Nature Astron. **3** (3), 272, arXiv:1811.02590 [astro-ph.CO] .

[1533] Robertson, H. P. (1929), Phys. Rev. **34**, 163.

[1534] Rodríguez, O., A. Clocchiatti, and M. Hamuy (2014), Astron. J. **148**, 107, arXiv:1409.3198 [astro-ph.CO] .

[1535] Rose, B. M., P. M. Garnavich, and M. A. Berg (2019), Astrophys. J. **874** (1), 32, arXiv:1902.01433 [astro-ph.CO] .

[313] Ross, A. J., L. Samushia, C. Howlett, W. J. Percival, A. Burden, and M. Manera (2015), Mon. Not. Roy. Astron. Soc. **449** (1), 835, arXiv:1409.3242 [astro-ph.CO] .

[1537] Rossi, M., M. Ballardini, M. Braglia, F. Finelli, D. Paoletti, A. A. Starobinsky, and C. Umiltà (2019), Phys. Rev. D **100** (10), 103524, arXiv:1906.10218 [astro-ph.CO] .

[1538] Roy, N., S. Goswami, and S. Das (2022), arXiv:2201.09306 [astro-ph.CO] .

[1539] Roy Choudhury, S., S. Hannestad, and T. Tram (2021), JCAP **03**, 084, arXiv:2012.07519 [astro-ph.CO] .

[1540] Rozo, E., *et al.* (DSDD) (2010), Astrophys. J. **708**, 645, arXiv:0902.3702 [astro-ph.CO] .

[1541] Ruan, C.-Z., F. Melia, Y. Chen, and T.-J. Zhang (2019), Astrophys. J. **881**, 137, arXiv:1901.06626 [astro-ph.CO] .

[1542] Ruan, W.-H., Z.-K. Guo, R.-G. Cai, and Y.-Z. Zhang (2020), Int. J. Mod. Phys. A **35** (17), 2050075, arXiv:1807.09495 [gr-qc] .

[1543] Ruan, W.-H., C. Liu, Z.-K. Guo, Y.-L. Wu, and R.-G. Cai (2019), 10.34133/2021/6014164, arXiv:1909.07104 [gr-qc] .

[1544] Ruan, W.-H., C. Liu, Z.-K. Guo, Y.-L. Wu, and R.-G. Cai (2020), Nature Astron. **4**, 108, arXiv:2002.03603 [gr-qc] .

[1545] Rubart, M., and D. J. Schwarz (2013), Astron. Astrophys. **555**, A117, arXiv:1301.5559 [astro-ph.CO] .

[1546] Rubin, V., N. Thonnard, and J. Ford, W.K. (1980), Astrophys. J. **238**, 471.

[1547] Rubin, V. C., and W. K. Ford, Jr. (1970), Astrophys. J. **159**, 379.

[1548] Rudnick, L., S. Brown, and L. R. Williams (2007), Astrophys. J. **671**, 40, arXiv:0704.0908 [astro-ph] .

[1549] Ruiz-Zapatero, J., C. García-García, D. Alonso, P. G. Ferreira, and R. D. P. Grumitt (2022), 10.1093/mnras/stac431, arXiv:2201.07025 [astro-ph.CO] .

[1550] Rydberg, C.-E., E. Zackrisson, P. Lundqvist, and P. Scott (2013), Mon. Not. Roy. Astron. Soc. **429**, 3658, arXiv:1206.0007 [astro-ph.CO] .

[1551] Sabelhaus, P. A., and J. E. Decker (2004), in *Optical, Infrared, and Millimeter Space Telescopes*, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 5487, edited by J. C. Mather, pp. 550–563.

[1552] Sabla, V. I., and R. R. Caldwell (2021), Phys. Rev. D **103** (10), 103506, arXiv:2103.04999 [astro-ph.CO] .

[1553] Sabla, V. I., and R. R. Caldwell (2022), arXiv:2202.08291 [astro-ph.CO] .

[1554] Sabti, N., J. B. Muñoz, and D. Blas (2021), arXiv:2110.13161 [astro-ph.CO] .

[1555] Sabti, N., J. B. Muñoz, and D. Blas (2022), Phys. Rev. D **105** (4), 043518, arXiv:2110.13168 [astro-ph.CO] .

[1556] Sachs, R. K., and A. M. Wolfe (1967), Astrophys. J. **147**, 73.

[1557] Sagredo, B., S. Nesseris, and D. Sapone (2018), Phys. Rev. D **98** (8), 083543, arXiv:1806.10822 [astro-ph.CO] .

[1558] Saha, P., and L. L. R. Williams (2006), Astrophys. J. **653**, 936, arXiv:astro-ph/0608496 .

[1559] Sahni, V., A. Shafieloo, and A. A. Starobinsky (2014), Astrophys. J. Lett. **793** (2), L40, arXiv:1406.2209 [astro-ph.CO] .

[1560] Saini, T. D., J. Weller, and S. L. Bridle (2004), Mon. Not. Roy. Astron. Soc. **348**, 603, arXiv:astro-ph/0305526 .

[96] de Sainte Agathe, V., *et al.* (2019), Astron. Astrophys. **629**, A85, arXiv:1904.03400 [astro-ph.CO] .

[1562] Sakai, S., *et al.* (2000), Astrophys. J. **529**, 698, arXiv:astro-ph/9909269 .

[1563] Sakharov, A. D. (1991), Usp. Fiz. Nauk **161** (5), 64.

[1564] Sakr, Z., and D. Sapone (2022), JCAP **03** (03), 034, arXiv:2112.14173 [astro-ph.CO] .

[1565] Sakstein, J. (2013), Phys. Rev. D **88** (12), 124013, arXiv:1309.0495 [astro-ph.CO] .

[1566] Sakstein, J. (2018), Int. J. Mod. Phys. D **27** (15), 1848008, arXiv:2002.04194 [astro-ph.CO] .

[1567] Sakstein, J., and M. Trodden (2020), Phys. Rev. Lett. **124** (16), 161301, arXiv:1911.11760 [astro-ph.CO] .

[100] de Salas, P. F., and S. Pastor (2016), JCAP **1607** (07), 051, arXiv:1606.06986 [hep-ph] .

[1569] Sales, L. V., *et al.* (2017), Mon. Not. Roy. Astron. Soc. **464** (2), 2419, arXiv:1602.02155 [astro-ph.GA] .

[1570] Salucci, P. (2019), Astron. Astrophys. Rev. **27** (1), 2, arXiv:1811.08843 [astro-ph.GA] .

[1571] Salvatelli, V., A. Marchini, L. Lopez-Honorez, and O. Mena (2013), Phys. Rev. D **88** (2), 023531, arXiv:1304.7119 [astro-ph.CO] .

[1572] Salvatelli, V., N. Said, M. Bruni, A. Melchiorri, and D. Wands (2014), Phys. Rev. Lett. **113** (18), 181301, arXiv:1406.7297 [astro-ph.CO] .

[1573] Salvaterra, R., *et al.* (2009), Nature **461**, 1258, arXiv:0906.1578 [astro-ph.CO] .

[321] Salvati, L., M. Douspis, and N. Aghanim (2018), Astron. Astrophys. **614**, A13, arXiv:1708.00697 [astro-ph.CO] .

[1575] Salvati, L., L. Pagano, M. Lattanzi, M. Gerbino, and A. Melchiorri (2016), JCAP **08**, 022, arXiv:1606.06968 [astro-ph.CO] .

[1576] Salvestrini, F., G. Risaliti, S. Bisogni, E. Lusso, and C. Vignali (2019), Astron. Astrophys. **631**, A120, arXiv:1909.12309 [astro-ph.GA] .

[1577] Samushia, L., W. J. Percival, and A. Raccanelli (2012), Mon. Not. Roy. Astron. Soc. **420**, 2102, arXiv:1102.1014 [astro-ph.CO] .

[1578] Samushia, L., and B. Ratra (2010), Astrophys. J. **714**, 1347, arXiv:0905.3836 [astro-ph.CO] .

[1579] Sandage, A., G. Tammann, A. Saha, B. Reindl, F. Macchetto, and N. Panagia (2006), Astrophys. J. **653**, 843, arXiv:astro-ph/0603647 .

[1580] Sandage, A., and G. A. Tammann (1975), ApJ **196**, 313.

[1581] Sandage, A., and G. A. Tammann (1984), Nature **307** (5949), 326.

[1582] Sandvik, H. B., J. D. Barrow, and J. Magueijo (2002), Phys. Rev. Lett. **88**, 031302, arXiv:astro-ph/0107512 .

[1583] Santos-Santos, I. M., C. B. Brook, G. Stinson, A. Di Cintio, J. Wadsley, R. Domínguez-Tenreiro, S. Gottlöber, and G. Yepes (2016), Mon. Not. Roy. Astron. Soc. **455** (1), 476, arXiv:1510.02474 [astro-ph.GA] .

[1584] Sapone, D., S. Nesseris, and C. A. P. Bengaly (2021), Phys. Dark Univ. **32**, 100814, arXiv:2006.05461 [astro-ph.CO] .

[1585] Saridakis, E. N., *et al.* (CANTATA) (2021), arXiv:2105.12582 [gr-qc] .

[1586] Sartoris, B., *et al.* (2016), Mon. Not. Roy. Astron. Soc. **459** (2), 1764, arXiv:1505.02165 [astro-ph.CO] .

[1587] Sathyaprakash, B. S., and B. F. Schutz (2009), Living Rev. Rel. **12**, 2, arXiv:0903.0338 [gr-qc] .

[1588] Sato, J., T. Shimomura, and M. Yamanaka (2017), Int. J. Mod. Phys. E **26** (08), 1741005, arXiv:1604.04769 [hep-ph] .

[1589] Sato, S., *et al.* (2017), J. Phys. Conf. Ser. **840** (1), 012010.

[1590] Sawala, T., *et al.* (2016), Mon. Not. Roy. Astron. Soc. **457** (2), 1931, arXiv:1511.01098 [astro-ph.GA] .

[1591] Sbordone, L., *et al.* (2010), Astron. Astrophys. **522**, A26, arXiv:1003.4510 [astro-ph.GA] .

[1592] Scalisi, M., and I. Valenzuela (2019), JHEP **08**, 160, arXiv:1812.07558 [hep-th] .

[1593] Schaefer, B. E. (2007), Astrophys. J. **660**, 16, arXiv:astro-ph/0612285 .

[326] Schellenberger, G., and T. H. Reiprich (2017), Mon. Not. Roy. Astron. Soc. **471** (2), 1370, arXiv:1705.05843 [astro-ph.CO] .

[1595] Scherrer, R. J. (2022), arXiv:2202.01132 [gr-qc] .

[1596] Schewtschenko, J. A., C. M. Baugh, R. J. Wilkinson, C. Bœhm, S. Pascoli, and T. Sawala (2016), Mon. Not. Roy. Astron. Soc. **461** (3), 2282, arXiv:1512.06774 [astro-ph.CO] .

[1597] Schive, H.-Y., T. Chiueh, and T. Broadhurst (2014), Nature Phys. **10**, 496, arXiv:1406.6586 [astro-ph.GA] .

[1598] Schlafly, E. F., A. M. Meisner, and G. M. Green (2019), The Astrophysical Journal Supplement **240** (2), 30, arXiv:1901.03337 [astro-ph.IM] .

[1599] Schmitz, K. (2022), arXiv:2203.04757 [physics.pop-ph] .

[1600] Schöneberg, N., J. Lesgourgues, and D. C. Hooper (2019), JCAP **1910** (10), 029, arXiv:1907.11594 [astro-ph.CO] .

[1601] Schneider, A., D. Anderhalden, A. Maccio, and J. Diemand (2014), Mon. Not. Roy. Astron. Soc. **441**, 6, arXiv:1309.5960 [astro-ph.CO] .

[1602] Schneider, P., and D. Sluse (2014), Astron. Astrophys. **564**, A103, arXiv:1306.4675 [astro-ph.CO] .

[1603] Schombert, J., S. McGaugh, and F. Lelli (2020), Astron. J. **160** (2), 71, arXiv:2006.08615 [astro-ph.CO] .

[1604] Schöneberg, N., G. Franco Abellán, A. Pérez Sánchez, S. J. Witte, V. Poulin, and J. Lesgourgues (2021), arXiv:2107.10291 [astro-ph.CO] .

[1605] Schramm, D. N., and M. S. Turner (1998), Rev. Mod. Phys. **70**, 303, arXiv:astro-ph/9706069 .

[1606] Schultz, C., J. Oñorbe, K. N. Abazajian, and J. S. Bullock (2014), Mon. Not. Roy. Astron. Soc. **442** (2), 1597, arXiv:1401.3769 [astro-ph.CO] .

[1607] Schutz, B. F. (1986), Nature **323**, 310.

[1608] Schwarz, D. J., C. J. Copi, D. Huterer, and G. D. Starkman (2016), Class. Quant. Grav. **33** (18), 184001, arXiv:1510.07929 [astro-ph.CO] .

[1609] Schwarz, D. J., G. D. Starkman, D. Huterer, and C. J. Copi (2004), Phys. Rev. Lett. **93**, 221301, arXiv:astro-ph/0403353 .

[1610] Schwarz, G. (1978), Annals Statist. **6**, 461.

[1611] Scolnic, D., *et al.* (2021), arXiv:2112.03863 [astro-ph.CO] .

[329] Scolnic, D. M., *et al.* (2018), Astrophys. J. **859** (2), 101, arXiv:1710.00845 [astro-ph.CO] .

[1613] Secco, L. F., *et al.* (DES) (2022), Phys. Rev. D **105** (2), 023515, arXiv:2105.13544 [astro-ph.CO] .

[1614] Secrest, N. J., S. von Hausegger, M. Rameez, R. Mohayaee, S. Sarkar, and J. Colin (2021), Astrophys. J. Lett. **908** (2), L51, arXiv:2009.14826 [astro-ph.CO] .

[1615] Seikel, M., and C. Clarkson (2013), arXiv:1311.6678 [astro-ph.CO] .

[1616] Seikel, M., C. Clarkson, and M. Smith (2012), JCAP **06**, 036, arXiv:1204.2832 [astro-ph.CO] .

[1617] Sekiguchi, T., and T. Takahashi (2021), Phys. Rev. D **103** (8), 083507, arXiv:2007.03381 [astro-ph.CO] .

[1618] Sen, A. A., S. A. Adil, and S. Sen (2021), arXiv:2112.10641 [astro-ph.CO] .

[1619] Seto, N., S. Kawamura, and T. Nakamura (2001), Phys. Rev. Lett. **87**, 221103, arXiv:astro-ph/0108011 .

[1620] Seto, O., and Y. Toda (2021), Phys. Rev. D **103** (12), 123501, arXiv:2101.03740 [astro-ph.CO] .

[1621] Seto, O., and Y. Toda (2021), arXiv:2104.04381 [astro-ph.CO] .

[1622] Shafieloo, A., A. G. Kim, and E. V. Linder (2012), Phys. Rev. D **85**, 123530, arXiv:1204.2272 [astro-ph.CO] .

[1623] Shafieloo, A., B. L'Huillier, and A. A. Starobinsky (2018), Phys. Rev. D **98** (8), 083526, arXiv:1804.04320 [astro-ph.CO] .

[1624] Shah, P., P. Lemos, and O. Lahav (2021), Astron. Astrophys. Rev. **29** (1), 9, arXiv:2109.01161 [astro-ph.CO] .

[1625] Shajib, A., *et al.* (DES) (2020), Mon. Not. Roy. Astron. Soc. **494** (4), 6072, arXiv:1910.06306 [astro-ph.CO] .

[1626] Shapiro, I. L., and J. Sola (2000), Phys. Lett. B **475**, 236, arXiv:hep-ph/9910462 .

[1627] Shapiro, I. L., and J. Sola (2002), JHEP **02**, 006, arXiv:hep-th/0012227 .

[1628] Shapiro, I. L., and J. Sola (2004), Nucl. Phys. B Proc. Suppl. **127**, 71, arXiv:hep-ph/0305279 .

[1629] Shapiro, I. L., and J. Sola (2008), arXiv:0808.0315 [hep-th] .

[1630] Shapiro, I. L., and J. Sola (2009), Phys. Lett. B **682**, 105, arXiv:0910.4925 [hep-th] .

[1631] Shapiro, I. L., J. Sola, and H. Stefancic (2005), JCAP **01**, 012, arXiv:hep-ph/0410095 .

[1632] Sharma, M. K., and S. Sur (2022), Int. J. Mod. Phys. D **31** (03), 2250017, arXiv:2112.08477 [astro-ph.CO] .

[1633] Sharma, R., A. Mukherjee, and H. K. Jassal (2022), Eur. Phys. J. Plus **137** (2), 219, arXiv:2004.01393 [astro-ph.CO] .

[1634] Sheth, R. K., and G. Tormen (2002), Mon. Not. Roy. Astron. Soc. **329**, 61, arXiv:astro-ph/0105113 .

[1635] Shiraishi, M., J. B. Muñoz, M. Kamionkowski, and A. Raccanelli (2016), Phys. Rev. D **93** (10), 103506, arXiv:1603.01206 [astro-ph.CO] .

[1636] Shoemaker, E. M. (1998), Journal of the Royal Astronomical Society of Canada **92**, 297.

[1637] Siegel, E. R., R. Guzman, J. P. Gallego, M. Orduna Lopez, and P. Rodriguez Hidalgo (2005), Mon. Not. Roy. Astron. Soc. **356**, 1117, arXiv:astro-ph/0410612 .

[1638] da Silva, W. J. C., and R. Silva (2021), Eur. Phys. J. Plus **136** (5), 543, arXiv:2011.09520 [astro-ph.CO] .

[1639] da Silva, W. J. C., and R. Silva (2021), Eur. Phys. J. C **81** (5), 403, arXiv:2011.09516 [astro-ph.CO] .

[1640] Simpson, F., C. Blake, J. A. Peacock, I. Baldry, J. Bland-Hawthorn, A. Heavens, C. Heymans, J. Loveday, and P. Norberg (2016), Phys. Rev. D **93** (2), 023525, arXiv:1505.03865 [astro-ph.CO] .

[1641] Singal, A. K. (2011), Astrophys. J. Lett. **742**, L23, arXiv:1110.6260 [astro-ph.CO] .

[1642] Singal, A. K. (2019), Phys. Rev. D **100** (6), 063501, arXiv:1904.11362 [physics.gen-ph] .

[1643] Singal, A. K. (2019), Mon. Not. Roy. Astron. Soc. **488** (1), L104, arXiv:1405.4796 [astro-ph.CO] .

[1644] Singer, L. P., and L. R. Price (2016), Phys. Rev. D **93** (2), 024013, arXiv:1508.03634 [gr-qc] .

[1645] Skara, F., and L. Perivolaropoulos (2019), Phys. Rev. D **100** (12), 123527, arXiv:1907.12594 [gr-qc] .

[1646] Skara, F., and L. Perivolaropoulos (2020), Phys. Rev. D **101** (6), 063521, arXiv:1911.10609 [astro-ph.CO] .

[1647] Skugoreva, M. A., E. N. Saridakis, and A. V. Toporensky (2015), Phys. Rev. **D91**, 044023, arXiv:1412.1502 [gr-qc] .

[1648] Smith, K. M., L. Senatore, and M. Zaldarriaga (2015), arXiv:1502.00635 [astro-ph.CO] .

[1649] Smith, T. L., M. Lucca, V. Poulin, G. F. Abellan, L. Balkenhol, K. Benabed, S. Galli, and R. Murgia (2022), arXiv:2202.09379 [astro-ph.CO] .

[1650] Smith, T. L., V. Poulin, and M. A. Amin (2020), Phys. Rev. D **101** (6), 063523, arXiv:1908.06995 [astro-ph.CO] .

[1651] Smith, T. L., V. Poulin, J. L. Bernal, K. K. Boddy, M. Kamionkowski, and R. Murgia (2021), Phys. Rev. D **103** (12), 123542, arXiv:2009.10740 [astro-ph.CO] .

[1652] Snyder, H. S. (1947), Phys. Rev. **71**, 38.

[1653] Soares-Santos, M., *et al.* (DES, Dark Energy Camera GW-EM) (2017), Astrophys. J. Lett. **848** (2), L16, arXiv:1710.05459 [astro-ph.HE] .

[1654] Soares-Santos, M., *et al.* (DES, LIGO Scientific, Virgo) (2019), Astrophys. J. Lett. **876** (1), L7, arXiv:1901.01540 [astro-ph.CO] .

[1655] Sohn, S. T., E. Patel, M. A. Fardal, G. Besla, R. P. van der Marel, M. Geha, and P. Guhathakurta (2020), The Astrophysical Journal **901** (1), 43.

[1656] Solà, J., A. Gómez-Valent, and J. de Cruz Pérez (2015), Astrophys. J. Lett. **811**, L14, arXiv:1506.05793 [gr-qc] .

[1657] Solà, J., A. Gómez-Valent, and J. de Cruz Pérez (2017), Astrophys. J. **836** (1), 43, arXiv:1602.02103 [astro-ph.CO] .

[1658] Solà, J., A. Gómez-Valent, and J. de Cruz Pérez (2017), Phys. Lett. **B774**, 317, arXiv:1705.06723 [astro-ph.CO] .

[1659] Solà Peracaula, J., J. de Cruz Pérez, and A. Gómez-Valent (2018), Mon. Not. Roy. Astron. Soc. **478** (4), 4357, arXiv:1703.08218 [astro-ph.CO] .

[1660] Sola, J. (2008), J. Phys. A **41**, 164066, arXiv:0710.4151 [hep-th] .

[1661] Sola, J. (2011), J. Phys. Conf. Ser. **283**, 012033, arXiv:1102.1815 [astro-ph.CO] .

[1662] Sola, J. (2013), J. Phys. Conf. Ser. **453**, 012015, arXiv:1306.1527 [gr-qc] .

[1663] Sola, J. (2015), AIP Conf. Proc. **1606** (1), 19, arXiv:1402.7049 [gr-qc] .

[1664] Solà, J. (2016), Int. J. Mod. Phys. A **31** (23), 1630035, arXiv:1612.02449 [astro-ph.CO] .

[1665] Solà, J., and A. Gómez-Valent (2015), Int. J. Mod. Phys. D **24**, 1541003, arXiv:1501.03832 [gr-qc] .

[1666] Sola, J., and H. Stefancic (2005), Phys. Lett. B **624**, 147, arXiv:astro-ph/0505133 .

[1667] Sola, J., and H. Stefancic (2006), Mod. Phys. Lett. A **21**, 479, arXiv:astro-ph/0507110 .

[1668] Solà Peracaula, J. (2021), arXiv:2109.12086 [gr-qc] .

[1669] Solà Peracaula, J., A. Gomez-Valent, J. de Cruz Pérez, and C. Moreno-Pulido (2019), Astrophys. J. Lett. **886** (1), L6, arXiv:1909.02554 [astro-ph.CO] .

[1670] Solà Peracaula, J., A. Gómez-Valent, J. de Cruz Pérez, and C. Moreno-Pulido (2020), Class. Quant. Grav. **37** (24), 245003, arXiv:2006.04273 [astro-ph.CO] .

[1671] Solà Peracaula, J., A. Gómez-Valent, J. de Cruz Perez, and C. Moreno-Pulido (2021), EPL **134** (1), 19001, arXiv:2102.12758 [astro-ph.CO] .

[1672] Soltis, J., S. Casertano, and A. G. Riess (2021), Astrophys. J. Lett. **908** (1), L5, arXiv:2012.09196 [astro-ph.GA] .

[1673] Somiya, K. (KAGRA) (2012), Class. Quant. Grav. **29**, 124007, arXiv:1111.7185 [gr-qc] .

[1674] Sotiriou, T. P., and V. Faraoni (2010), Rev. Mod. Phys. **82**, 451, arXiv:0805.1726 [gr-qc] .

[1675] Spallicci, A. D. A. M., J. A. Helayël-Neto, M. López-Corredoira, and S. Capozziello (2021), Eur. Phys. J. C **81** (1), 4, arXiv:2011.12608 [astro-ph.CO] .

[1676] Spergel, D. N., *et al.* (WMAP) (2003), Astrophys. J. Suppl. **148**, 175, arXiv:astro-ph/0302209 .

[1677] Spergel, D. N., *et al.* (WMAP) (2007), Astrophys. J. Suppl. **170**, 377, arXiv:astro-ph/0603449 .

[1678] Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002), J. Roy. Statist. Soc. B **64** (4), 583.

[1679] Spite, M., F. Spite, and P. Bonifacio (2012), Mem. Soc. Astron. Ital. Suppl. **22**, 9, arXiv:1208.1190 [astro-ph.CO] .

[1680] Stadler, J., C. Bœhm, and O. Mena (2019), JCAP **08**, 014, arXiv:1903.00540 [astro-ph.CO] .

[260] Starobinsky, A. A. (1987), Adv. Ser. Astrophys. Cosmol. **3**, 130.

[1682] Steffen, A. T., I. Strateva, W. N. Brandt, D. M. Alexander, A. M. Koekemoer, B. D. Lehmer, D. P. Schneider, and C. Vignali (2006), Astron. J. **131**, 2826, arXiv:astro-ph/0602407 .

[1683] Steigman, G. (2007), Ann. Rev. Nucl. Part. Sci. **57**, 463, arXiv:0712.1100 [astro-ph] .

[1684] Steinhardt, P. J. (1997), in *Critical Problems in Physics*, edited by V. L. Fitch, D. R. Marlow, and M. A. E. Dementi, p. 123.

[1685] Steinhardt, P. J., L.-M. Wang, and I. Zlatev (1999), Phys. Rev. D **59**, 123504, arXiv:astro-ph/9812313 .

[1686] Strigari, L. E., M. Kaplinghat, and J. S. Bullock (2007), Phys. Rev. D **75**, 061303, arXiv:astro-ph/0606281 .

[1687] Subramanian, K. (2016), Rept. Prog. Phys. **79** (7), 076901, arXiv:1504.02311 [astro-ph.CO] .

[1688] Sun, W., K. Jiao, and T.-J. Zhang (2021), Astrophys. J. **915** (2), 123, arXiv:2105.12618 [astro-ph.CO] .

[1689] Sunyaev, R., and Y. Zeldovich (1970), Astrophys. Space Sci. **7**, 20.

[1690] Sunyaev, R., and Y. Zeldovich (1972), Comments Astrophys. Space Phys. **4**, 173.

[1691] Sunyaev, R. A., and Y. B. Zeldovich (1980), Ann. Rev. Astron. Astrophys. **18**, 537.

[1692] Susskind, L. (2003), arXiv:hep-th/0302219 .

[1693] Suyu, S. H., T.-C. Chang, F. Courbin, and T. Okumura (2018), Space Sci. Rev. **214** (5), 91, arXiv:1801.07262 [astro-ph.CO] .

[1694] Suyu, S. H., P. J. Marshall, M. W. Auger, S. Hilbert, R. D. Blandford, L. V. E. Koopmans, C. D. Fassnacht, and T. Treu (2010), Astrophys. J. **711**, 201, arXiv:0910.2773 [astro-ph.CO] .

[1695] Suyu, S. H., *et al.* (2012), arXiv:1202.4459 [astro-ph.CO] .

[1696] Suyu, S. H., *et al.* (2017), Mon. Not. Roy. Astron. Soc. **468** (3), 2590, arXiv:1607.00017 [astro-ph.CO] .

[1697] Takahashi, F., and W. Yin (2021), arXiv:2112.06710 [astro-ph.CO] .

[1698] Takahashi, F., and W. Yin (2021), JCAP **04**, 007, arXiv:2012.11576 [hep-ph] .

[1699] Tamone, A., *et al.* (2020), Mon. Not. Roy. Astron. Soc. **499** (4), 5527, arXiv:2007.09009 [astro-ph.CO] .

[1700] Tanabashi, M., *et al.* (Particle Data Group) (2018), Phys. Rev. D **98** (3), 030001.

[1701] Tang, C.-H., Y.-F. Huang, J.-J. Geng, and Z.-B. Zhang (2019), Astrophys. J. Suppl. **245** (1), 1, arXiv:1905.07929 [astro-ph.HE] .

[1702] Tang, Y. (2016), Phys. Lett. B **757**, 387, arXiv:1603.00165 [astro-ph.CO] .

[1703] Tanin, E. H., and T. Tenkanen (2021), JCAP **01**, 053, arXiv:2004.10702 [astro-ph.CO] .

[1704] Tanvir, N. R., *et al.* (2009), Nature **461**, 1254.

[1705] Tawfik, A. N., and A. M. Diab (2015), Rept. Prog. Phys. **78**, 126001, arXiv:1509.02436 [physics.gen-ph] .

[1706] Taylor, S. R., and J. R. Gair (2012), Phys. Rev. D **86**, 023502, arXiv:1204.6739 [astro-ph.CO] .

[1707] Tegmark, M., *et al.* (SDSS) (2006), Phys. Rev. **D74**, 123507, arXiv:astro-ph/0608632 [astro-ph] .

[1708] Terlevich, R., E. Terlevich, J. Melnick, R. Chávez, M. Plionis, F. Bresolin, and S. Basilakos (2015), Mon. Not. Roy. Astron. Soc. **451** (3), 3001, arXiv:1505.04376 [astro-ph.CO] .

[1709] Theodoropoulos, A., and L. Perivolaropoulos (2021), Universe **7** (8), 300, arXiv:2109.06256 [astro-ph.CO] .

[1710] Thiele, L., Y. Guan, J. C. Hill, A. Kosowsky, and D. N. Spergel (2021), Phys. Rev. D **104** (6), 063535, arXiv:2105.03003 [astro-ph.CO] .

[1711] Tikhonov, A., and A. Klypin (2009), Mon. Not. Roy. Astron. Soc. **395**, 1915, arXiv:0807.0924 [astro-ph] .

[1712] Tinker, J. L., A. V. Kravtsov, A. Klypin, K. Abazajian, M. S. Warren, G. Yepes, S. Gottlober, and D. E. Holz (2008), Astrophys. J. **688**, 709, arXiv:0803.2706 [astro-ph] .

[1713] Tinsley, B. M. (1968), ApJ **151**, 547.

[1714] Tiwari, P., R. Kothari, A. Naskar, S. Nadkarni-Ghosh, and P. Jain (2014), Astropart. Phys. **61**, 1, arXiv:1307.1947 [astro-ph.CO] .

[1715] Tiwari, P., and A. Nusser (2016), JCAP **03**, 062, arXiv:1509.02532 [astro-ph.CO] .

[1716] Tojeiro, R., *et al.* (2012), Mon. Not. Roy. Astron. Soc. **424**, 2339, arXiv:1203.6565 [astro-ph.CO] .

[1717] Tollerud, E. J., M. Boylan-Kolchin, and J. S. Bullock (2014), Mon. Not. Roy. Astron. Soc. **440** (4), 3511, arXiv:1403.6469 [astro-ph.GA] .

[1718] Tomboulis, E. T. (1997), arXiv:hep-th/9702146 .

[1719] Tonry, J., and D. P. Schneider (1988), AJ **96**, 807.

[1720] Tonry, J. L., J. P. Blakeslee, E. A. Ajhar, and A. Dressler (1997), Astrophys. J. **475**, 399, arXiv:astro-ph/9609113 .

[1721] Tornatore, L., A. Ferrara, and R. Schneider (2007), Mon. Not. Roy. Astron. Soc. **382**, 945, arXiv:0707.1433 [astro-ph] .

[1722] de la Torre, S., *et al.* (2013), Astron. Astrophys. **557**, A54, arXiv:1303.2622 [astro-ph.CO] .

[1723] Tram, T., R. Vallance, and V. Vennin (2017), JCAP **01**, 046, arXiv:1606.09199 [astro-ph.CO] .

[1724] Treu, T., and P. J. Marshall (2016), Astron. Astrophys. Rev. **24** (1), 11, arXiv:1605.05333 [astro-ph.CO] .

[1725] Trott, E., and D. Huterer (2021), arXiv:2112.00241 [astro-ph.CO] .

[355] Troxel, M. A., *et al.* (DES) (2018), Phys. Rev. **D98** (4), 043528, arXiv:1708.01538 [astro-ph.CO] .

[1727] Tröster, T., *et al.* (2020), Astron. Astrophys. **633**, L10, arXiv:1909.11006 [astro-ph.CO] .

[1728] Truemper, J. (1993), Science **260** (5115), 1769.

[1729] Tsagas, C. G. (2011), Phys. Rev. D **84**, 063503, arXiv:1107.4045 [astro-ph.CO] .

[1730] Tsagas, C. G., and M. I. Kadiltzoglou (2015), Phys. Rev. D **92** (4), 043515, arXiv:1507.04266 [gr-qc] .

[1731] Tsagas, C. G., M. I. Kadiltzoglou, and K. Asvesta (2021), Astrophys. Space Sci. **366** (9), 90, arXiv:2105.09267 [gr-qc] .

[1732] Tsiapi, P., and S. Basilakos (2019), Mon. Not. Roy. Astron. Soc. **485** (2), 2505, arXiv:1810.12902 [astro-ph.CO] .

[1733] Tsujikawa, S. (2013), Class. Quant. Grav. **30**, 214003, arXiv:1304.1961 [gr-qc] .

[1734] Tsujikawa, S. (2015), Phys. Rev. D **92** (4), 044029, arXiv:1505.02459 [astro-ph.CO] .

[1735] Tulin, S., and H.-B. Yu (2018), Phys. Rept. **730**, 1, arXiv:1705.02358 [hep-ph] .

[1736] Tully, R. B. (1988), Nature **334** (6179), 209.

[1737] Tully, R. B., and J. R. Fisher (1977), Astron. Astrophys. **54**, 661.

[1738] Turnbull, S. J., M. J. Hudson, H. A. Feldman, M. Hicken, R. P. Kirshner, and R. Watkins (2012), Mon. Not. Roy. Astron. Soc. **420**, 447, arXiv:1111.0631 [astro-ph.CO] .

[1739] Turner, M. S. (1983), Phys. Rev. D **28**, 1243.

[1740] Turner, M. S. (1991), Phys. Rev. D **44**, 3737.

[1741] Turner, M. S. (2022), 10.1146/annurev-nucl-111119-041046, arXiv:2201.04741 [astro-ph.CO] .

[1742] Turner, M. S., and L. M. Widrow (1988), Phys. Rev. D **37**, 2743.

[1743] Tutusaus, I., B. Lamine, and A. Blanchard (2019), Astron. Astrophys. **625**, A15, arXiv:1803.06197 [astro-ph.CO] .

[1744] Tutusaus, I., B. Lamine, A. Dupays, and A. Blanchard (2017), Astron. Astrophys. **602**, A73, arXiv:1706.05036 [astro-ph.CO] .

[358] van Uitert, E., *et al.* (2018), Mon. Not. Roy. Astron. Soc. **476** (4), 4662, arXiv:1706.05004 [astro-ph.CO] .

[1746] Unnikrishnan, C. S. (2013), Int. J. Mod. Phys. D **22**, 1341010, arXiv:1510.06059 [physics.ins-det] .

[1747] Upadhye, A., S. S. Gubser, and J. Khoury (2006), Phys. Rev. D **74**, 104024, arXiv:hep-ph/0608186 .

[1748] Upadhye, A., W. Hu, and J. Khoury (2012), Phys. Rev. Lett. **109**, 041301, arXiv:1204.3906 [hep-ph] .

[1749] Vachaspati, T. (2021), Rept. Prog. Phys. **84** (7), 074901, arXiv:2010.10525 [astro-ph.CO] .

[1750] Vagenas, E. C., L. Alasfar, S. M. Alsaleh, and A. F. Ali (2018), Nucl. Phys. **B931**, 72, arXiv:1706.06502 [hep-th] .

[1751] Vagnetti, F., S. Turriziani, D. Trevese, and M. Antonucci (2010), Astron. Astrophys. **519**, A17, arXiv:1005.0144 [astro-ph.CO] .

[1752] Vagnozzi, S. (2020), Phys. Rev. D **102** (2), 023518, arXiv:1907.07569 [astro-ph.CO] .

[1753] Vagnozzi, S. (2021), Phys. Rev. D **104** (6), 063524, arXiv:2105.10425 [astro-ph.CO] .

[1754] Vagnozzi, S., S. Dhawan, M. Gerbino, K. Freese, A. Goobar, and O. Mena (2018), Phys. Rev. D **98** (8), 083501, arXiv:1801.08553 [astro-ph.CO] .

[1755] Vagnozzi, S., E. Di Valentino, S. Gariazzo, A. Melchiorri, O. Mena, and J. Silk (2021), Phys. Dark Univ. **33**, 100851, arXiv:2010.02230 [astro-ph.CO] .

[1756] Vagnozzi, S., A. Loeb, and M. Moresco (2021), Astrophys. J. **908** (1), 84, arXiv:2011.11645 [astro-ph.CO] .

[1757] Vagnozzi, S., F. Pacucci, and A. Loeb (2021), arXiv:2105.10421 [astro-ph.CO] .

[1758] Vagnozzi, S., L. Visinelli, P. Brax, A.-C. Davis, and J. Sakstein (2021), Phys. Rev. D **104** (6), 063023, arXiv:2103.15834 [hep-ph] .

[1759] Vagnozzi, S., L. Visinelli, O. Mena, and D. F. Mota (2020), Mon. Not. Roy. Astron. Soc. **493** (1), 1139, arXiv:1911.12374 [gr-qc] .

[1760] Vainshtein, A. I. (1972), Phys. Lett. B **39**, 393.

[1761] Valcin, D., J. L. Bernal, R. Jimenez, L. Verde, and B. D. Wandelt (2020), arXiv:2007.06594 [astro-ph.CO] .

[1762] Valkenburg, W. (2012), JCAP **01**, 047, arXiv:1106.6042 [astro-ph.CO] .

[1763] Van De Bruck, C., and J. Mifsud (2018), Phys. Rev. D **97** (2), 023506, arXiv:1709.04882 [astro-ph.CO] .

[1764] VandenBerg, D. A., H. E. Bond, E. P. Nelan, P. E. Nissen, G. H. Schaefer, and D. Harmer (2014), The Astrophysical Journal **792** (2), 110.

[1765] Varma, V., S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer (2019), Phys. Rev. Research. **1**, 033015, arXiv:1905.09300 [gr-qc] .

[1766] Vattis, K., S. M. Koushiappas, and A. Loeb (2019), Phys. Rev. **D99** (12), 121302, arXiv:1903.06220 [astro-ph.CO] .

[1767] Velten, H., I. Costa, and W. Zimdahl (2021), Phys. Rev. D **104** (6), 063507, arXiv:2104.05352 [astro-ph.CO] .

[1768] Velten, H., and S. Gomes (2020), Phys. Rev. D **101** (4), 043502, arXiv:1911.11848 [astro-ph.CO] .

[1769] Velten, H., R. vom Marttens, and W. Zimdahl (2014), Eur. Phys. J. C **74** (11), 3160, arXiv:1410.2509 [astro-ph.CO] .

[1770] Verbiest, J. P. W., *et al.* (2016), Mon. Not. Roy. Astron. Soc. **458** (2), 1267, arXiv:1602.03640 [astro-ph.IM] .

[1771] Verde, L., P. Protopapas, and R. Jimenez (2013), Phys. Dark Univ. **2**, 166, arXiv:1306.6766 [astro-ph.CO] .

[1772] Verde, L., T. Treu, and A. G. Riess (2019), Nature Astron. **3**, 891, arXiv:1907.10625 [astro-ph.CO] .

[1773] Verheijen, M. A. W. (2001), Astrophys. J. **563**, 694, arXiv:astro-ph/0108225 .

[1774] Viel, M., G. D. Becker, J. S. Bolton, and M. G. Haehnelt (2013), Phys. Rev. D **88**, 043502, arXiv:1306.2314 [astro-ph.CO] .

[1775] Viel, M., J. Lesgourgues, M. G. Haehnelt, S. Matarrese, and A. Riotto (2005), Phys. Rev. D **71**, 063534, arXiv:astro-ph/0501562 .

[1776] Vielva, P. (2010), Adv. Astron. **2010**, 592094, arXiv:1008.3051 [astro-ph.CO] .

[1777] Vielva, P., E. Martinez-Gonzalez, R. Barreiro, J. Sanz, and L. Cayon (2004), Astrophys. J. **609**, 22, arXiv:astro-ph/0310273 .

[1778] Vignali, C., W. N. Brandt, and D. P. Schneider (2003), Astron. J. **125**, 433, arXiv:astro-ph/0211125 .

[1779] Vishwakarma, R. G. (2020), Int. J. Mod. Phys. D **29** (14), 2043025, arXiv:2011.12146 [physics.gen-ph] .

[1780] Visinelli, L., S. Vagnozzi, and U. Danielsson (2019), Symmetry **11** (8), 1035, arXiv:1907.07953 [astro-ph.CO] .

[1781] Vogel, H., and J. Redondo (2014), JCAP **02**, 029, arXiv:1311.2600 [hep-ph] .

[1782] Vogelsberger, M., J. Zavala, F.-Y. Cyr-Racine, C. Pfrommer, T. Bringmann, and K. Sigurdson (2016), Mon. Not. Roy. Astron. Soc. **460** (2), 1399, arXiv:1512.05349 [astro-ph.CO] .

[1783] Vogelsberger, M., J. Zavala, and A. Loeb (2012), Mon. Not. Roy. Astron. Soc. **423**, 3740, arXiv:1201.5892 [astro-ph.CO] .

[1784] de Waard, A., L. Gottardi, J. van Houwelingen, A. Shumack, and G. Frossati (2003), Class. Quant. Grav. **20**, S143.

[1785] van Waerbeke, L., *et al.* (2000), Astron. Astrophys. **358**, 30, arXiv:astro-ph/0002500 .

[1786] Walker, M. G., and J. Penarrubia (2011), Astrophys. J. **742**, 20, arXiv:1108.2404 [astro-ph.CO] .

[1787] Wandelt, B. D., E. Hivon, and K. M. Gorski (2001), Phys. Rev. D **64**, 083003, arXiv:astro-ph/0008111 .

[1788] Wang, B., E. Abdalla, F. Atrio-Barandela, and D. Pavon (2016), Rept. Prog. Phys. **79** (9), 096901, arXiv:1603.08299 [astro-ph.CO] .

[1789] Wang, D. (2021), Eur. Phys. J. C **81** (5), 482, arXiv:2008.03966 [astro-ph.CO] .

[1790] Wang, D., and X.-H. Meng (2017), Astrophys. J. **843** (2), 100, arXiv:1612.09023 [astro-ph.CO] .

[1791] Wang, D., and X.-H. Meng (2017), Phys. Rev. D **95** (2), 023508, arXiv:1708.07750 [astro-ph.CO] .

[1792] Wang, D., and D. Mota (2020), Phys. Rev. D **102** (6), 063530, arXiv:2003.10095 [astro-ph.CO] .

[1793] Wang, D., and D. Mota (2021), Phys. Dark Univ. **32**, 100813, arXiv:2103.12358 [astro-ph.CO] .

[1794] Wang, D.-G., W. Zhao, Y. Zhang, and Y.-F. Cai (2016), JCAP **02**, 019, arXiv:1509.02541 [astro-ph.CO] .

[1795] Wang, F. Y., Z. G. Dai, and E. W. Liang (2015), New Astron. Rev. **67**, 1, arXiv:1504.00735 [astro-ph.HE] .

[1796] Wang, H., and Y.-S. Piao (2022), arXiv:2201.07079 [astro-ph.CO] .

[1797] Wang, H.-T., *et al.* (2019), Phys. Rev. D **100** (4), 043003, arXiv:1902.04423 [astro-ph.HE] .

[1798] Wang, J., L. Hui, and J. Khoury (2012), Phys. Rev. Lett. **109**, 241301, arXiv:1208.4612 [astro-ph.CO] .

[1799] Wang, J. S., F. Y. Wang, K. S. Cheng, and Z. G. Dai (2016), Astron. Astrophys. **585**, A68, arXiv:1509.08558 [astro-ph.HE] .

[1800] Wang, L.-F., D.-Z. He, J.-F. Zhang, and X. Zhang (2021), arXiv:2102.09331 [astro-ph.CO] .

[1801] Wang, L.-M., and P. J. Steinhardt (1998), Astrophys. J. **508**, 483, arXiv:astro-ph/9804015 .

[1802] Wang, M.-Y., A. H. G. Peter, L. E. Strigari, A. R. Zentner, B. Arant, S. Garrison-Kimmel, and M. Rocha (2014), Mon. Not. Roy. Astron. Soc. **445** (1), 614, arXiv:1406.0527 [astro-ph.CO] .

[1803] Wang, P., and X.-H. Meng (2005), Class. Quant. Grav. **22**, 283, arXiv:astro-ph/0408495 .

[1804] Wang, Y., L. Pogosian, G.-B. Zhao, and A. Zucca (2018), Astrophys. J. Lett. **869**, L8, arXiv:1807.03772 [astro-ph.CO] .

[1805] Wang, Y., L. Xu, and G.-B. Zhao (2017), Astrophys. J. **849** (2), 84, arXiv:1706.09149 [astro-ph.CO] .

[1806] Ward, S. N., and S. Day (2007).

[1807] Warren, M. S., K. Abazajian, D. E. Holz, and L. Teodoro (2006), Astrophys. J. **646**, 881, arXiv:astro-ph/0506395 .

[1808] Watkins, R., H. A. Feldman, and M. J. Hudson (2009), Mon. Not. Roy. Astron. Soc. **392**, 743, arXiv:0809.4041 [astro-ph] .

[1809] Webb, J., J. King, M. Murphy, V. Flambaum, R. Carswell, and M. Bainbridge (2011), Phys. Rev. Lett. **107**, 191101, arXiv:1008.3907 [astro-ph.CO] .

[1810] Webb, J. K., V. V. Flambaum, C. W. Churchill, M. J. Drinkwater, and J. D. Barrow (1999), Phys. Rev. Lett. **82**, 884, arXiv:astro-ph/9803165 .

[1811] Webb, J. K., M. T. Murphy, V. V. Flambaum, V. A. Dzuba, J. D. Barrow, C. W. Churchill, J. X. Prochaska, and A. M. Wolfe (2001), Phys. Rev. Lett. **87**, 091301, arXiv:astro-ph/0012539 .

[1812] Webb, J. K., A. Wright, F. E. Koch, and M. T. Murphy (2014), Mem. Soc. Ast. It. **85** (1), 57.

[1813] Weber, J. (1960), Phys. Rev. **117**, 306.

[1814] Wei, J.-J., and F. Melia (2022), arXiv:2202.07865 [astro-ph.CO] .

[1815] Wei, J.-J., X.-F. Wu, and F. Melia (2016), Mon. Not. Roy. Astron. Soc. **463** (2), 1144, arXiv:1608.02070 [astro-ph.CO] .

[1816] Weinberg, D. H., J. S. Bullock, F. Governato, R. Kuzio de Naray, and A. H. G. Peter (2015), Proc. Nat. Acad. Sci. **112**, 12249, arXiv:1306.0913 [astro-ph.CO] .

[1817] Weinberg, D. H., M. J. Mortonson, D. J. Eisenstein, C. Hirata, A. G. Riess, and E. Rozo (2013), Phys. Rept. **530**, 87, arXiv:1201.2434 [astro-ph.CO] .

[1818] Weinberg, S. (1987), Phys. Rev. Lett. **59**, 2607.

[1819] Weinberg, S. (1989), Rev. Mod. Phys. **61**, 1.

[1820] Weinberg, S. (2008), *Cosmology*.

[1821] Weller, J., and A. Albrecht (2002), Phys. Rev. D **65**, 103512, arXiv:astro-ph/0106079 .

[1822] Wetterich, C. (1988), Nucl. Phys. B **302**, 668, arXiv:1711.03844 [hep-th] .

[1823] Wetterich, C. (1995), Astron. Astrophys. **301**, 321, arXiv:hep-th/9408025 .

[1824] Whalen, D. J., C. L. Fryer, D. E. Holz, A. Heger, S. E. Woosley, M. Stiavelli, W. Even, and L. L. Frey (2013), Astrophys. J. Lett. **762**, L6, arXiv:1209.3457 [astro-ph.CO] .

[1825] White, M., B. Reid, C.-H. Chuang, J. L. Tinker, C. K. McBride, F. Prada, and L. Samushia (2015), Mon. Not. Roy. Astron. Soc. **447**, 234, arXiv:1408.5435 [astro-ph.CO] .

[1826] White, M., *et al.* (2022), JCAP **02** (02), 007, arXiv:2111.09898 [astro-ph.CO] .

[1827] White, M. J. (2002), Astrophys. J. Suppl. **143**, 241, arXiv:astro-ph/0207185 .

[1828] Wilczynska, M. R., *et al.* (2020), Sci. Adv. **6** (17), eaay9672, arXiv:2003.07627 [astro-ph.CO] .

[1829] Wilkinson, R. J., C. Boehm, and J. Lesgourgues (2014), JCAP **05**, 011, arXiv:1401.7597 [astro-ph.CO] .

[1830] Willke, B., *et al.* (2002), Class. Quant. Grav. **19**, 1377.

[1831] Wiltshire, D. L., P. R. Smale, T. Mattsson, and R. Watkins (2013), Phys. Rev. D **88**, 083529, arXiv:1201.5371 [astro-ph.CO] .

[1832] Wittman, D. M., J. A. Tyson, D. Kirkman, I. Dell'Antonio, and G. Bernstein (2000), Nature **405**, 143, arXiv:astro-ph/0003014 .

[1833] Wittner, M., G. Laverda, O. F. Piattella, and L. Amendola (2020), JCAP **07**, 019, arXiv:2003.08950 [gr-qc] .

[366] Wong, K. C., *et al.* (2020), Mon. Not. Roy. Astron. Soc. **498** (1), 1420, arXiv:1907.04869 [astro-ph.CO] .

[1835] Wong, Y. Y. Y. (2011), Ann. Rev. Nucl. Part. Sci. **61**, 69, arXiv:1111.1436 [astro-ph.CO] .

[1836] Wright, A. H., H. Hildebrandt, J. L. van den Busch, C. Heymans, B. Joachimi, A. Kannawadi, and K. Kuijken (2020), Astron. Astrophys. **640**, L14, arXiv:2005.04207 [astro-ph.CO] .

[1837] Wright, B. S., and B. Li (2018), Phys. Rev. D **97** (8), 083505, arXiv:1710.07018 [astro-ph.CO] .

[1838] Wright, E. L., *et al.* (2010), Astron. J. **140**, 1868, arXiv:1008.0031 [astro-ph.IM] .

[1839] Wu, C. S., E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson (1957), Phys. Rev. **105**, 1413.

[1840] Wu, H.-Y., and D. Huterer (2017), Mon. Not. Roy. Astron. Soc. **471** (4), 4946, arXiv:1706.09723 [astro-ph.CO] .

[1841] Wyman, M., D. H. Rudd, R. Vanderveld, and W. Hu (2014), Phys. Rev. Lett. **112** (5), 051302, arXiv:1307.7715 [astro-ph.CO] .

[1842] Xia, D.-M., and S. Wang (2016), Mon. Not. Roy. Astron. Soc. **463** (1), 952, arXiv:1608.04545 [astro-ph.CO] .

[1843] Xiao, L., L. Zhang, R. An, C. Feng, and B. Wang (2020), JCAP **01**, 045, arXiv:1908.02668 [astro-ph.CO] .

[1844] Xu, H., K. Ahn, M. L. Norman, J. H. Wise, and B. W. O'Shea (2016), Astrophys. J. Lett. **832** (1), L5, arXiv:1607.02664 [astro-ph.GA] .

[1845] Yadav, S. K. (2019), Mod. Phys. Lett. **A33**, 1950358, arXiv:1907.05886 [astro-ph.CO] .

[1846] Yahya, S., M. Seikel, C. Clarkson, R. Maartens, and M. Smith (2014), Phys. Rev. D **89** (2), 023503, arXiv:1308.4099 [astro-ph.CO] .

[1847] Yamazaki, D. G., M. Kusakabe, T. Kajino, G. J. Mathews, and M.-K. Cheoun (2014), Phys. Rev. D **90** (2), 023001, arXiv:1407.0021 [astro-ph.CO] .

[1848] Yamazaki, D. G., M. Kusakabe, T. Kajino, G. J. Mathews, and M.-K. Cheoun (2017), Int. J. Mod. Phys. E **26** (08), 1741006.

[1849] Yang, C. N. (1947), Phys. Rev. **72**, 874, [,7(1947)].

[1850] Yang, Q., Y. Liu, and H. Di (2017), Phys. Rev. D **96** (8), 083516, arXiv:1612.03708 [astro-ph.CO] .

[1851] Yang, T., A. Banerjee, and E. O. Colgáin (2020), Phys. Rev. D **102** (12), 123532, arXiv:1911.01681 [astro-ph.CO] .

[1852] Yang, T., S. Birrer, and B. Hu (2020), Mon. Not. Roy. Astron. Soc. **497** (1), L56, arXiv:2003.03277 [astro-ph.CO] .

[1853] Yang, W., E. Di Valentino, O. Mena, and S. Pan (2020), Phys. Rev. D **102** (2), 023535, arXiv:2003.12552 [astro-ph.CO] .

[1854] Yang, W., E. Di Valentino, O. Mena, S. Pan, and R. C. Nunes (2020), Phys. Rev. D **101** (8), 083509, arXiv:2001.10852 [astro-ph.CO] .

[1855] Yang, W., E. Di Valentino, S. Pan, S. Basilakos, and A. Paliathanasis (2020), Phys. Rev. D **102** (6), 063503, arXiv:2001.04307 [astro-ph.CO] .

[1856] Yang, W., E. Di Valentino, S. Pan, and O. Mena (2021), Phys. Dark Univ. **31**, 100762, arXiv:2007.02927 [astro-ph.CO] .

[1857] Yang, W., E. Di Valentino, S. Pan, and O. Mena (2021), Phys. Dark Univ. **31**, 100762, arXiv:2007.02927 [astro-ph.CO] .

[1858] Yang, W., E. Di Valentino, S. Pan, A. Shafieloo, and X. Li (2021), Phys. Rev. D **104** (6), 063521, arXiv:2103.03815 [astro-ph.CO] .

[1859] Yang, W., E. Di Valentino, S. Pan, Y. Wu, and J. Lu (2021), Mon. Not. Roy. Astron. Soc. **501** (4), 5845, arXiv:2101.02168 [astro-ph.CO] .

[1860] Yang, W., O. Mena, S. Pan, and E. Di Valentino (2019), Phys. Rev. **D100** (8), 083509, arXiv:1906.11697 [astro-ph.CO] .

[1861] Yang, W., A. Mukherjee, E. Di Valentino, and S. Pan (2018), Phys. Rev. **D98** (12), 123527, arXiv:1809.06883 [astro-ph.CO] .

[1862] Yang, W., R. C. Nunes, S. Pan, and D. F. Mota (2017), Phys. Rev. D **95** (10), 103522, arXiv:1703.02556 [astro-ph.CO] .

[1863] Yang, W., S. Pan, E. Di Valentino, R. C. Nunes, S. Vagnozzi, and D. F. Mota (2018), JCAP **1809** (09), 019, arXiv:1805.08252 [astro-ph.CO] .

[1864] Yang, W., S. Pan, E. Di Valentino, A. Paliathanasis, and J. Lu (2019), Phys. Rev. **D100** (10), 103518, arXiv:1906.04162 [astro-ph.CO] .

[1865] Yang, W., S. Pan, E. Di Valentino, and E. N. Saridakis (2019), Universe **5** (11), 219, arXiv:1811.06932 [astro-ph.CO] .

[1866] Yang, W., S. Pan, E. Di Valentino, E. N. Saridakis, and S. Chakraborty (2019), Phys. Rev. **D99** (4), 043543, arXiv:1810.05141 [astro-ph.CO] .

[1867] Yang, W., S. Pan, R. Herrera, and S. Chakraborty (2018), Phys. Rev. D **98** (4), 043517, arXiv:1808.01669 [gr-qc] .

[1868] Yang, W., S. Pan, and D. F. Mota (2017), Phys. Rev. D **96** (12), 123508, arXiv:1709.00006 [astro-ph.CO] .

[1869] Yang, W., S. Pan, R. C. Nunes, and D. F. Mota (2020), JCAP **04**, 008, arXiv:1910.08821 [astro-ph.CO] .

[1870] Yang, W., S. Pan, and A. Paliathanasis (2018), Mon. Not. Roy. Astron. Soc. **475** (2), 2605, arXiv:1708.01717 [gr-qc] .

[1871] Yang, W., S. Pan, A. Paliathanasis, S. Ghosh, and Y. Wu (2019), Mon. Not. Roy. Astron. Soc. **490** (2), 2071, arXiv:1904.10436 [gr-qc] .

[1872] Yang, W., S. Pan, S. Vagnozzi, E. Di Valentino, D. F. Mota, and S. Capozziello (2019), JCAP **1911**, 044, arXiv:1907.05344 [astro-ph.CO] .

[1873] Yang, W., S. Pan, L. Xu, and D. F. Mota (2019), Mon. Not. Roy. Astron. Soc. **482** (2), 1858, arXiv:1804.08455 [astro-ph.CO] .

[1874] Yao, Y., and X.-H. Meng (2021), Phys. Dark Univ. **33**, 100852, arXiv:2011.09160 [astro-ph.CO] .

[1875] Yao, Y.-H., and X.-H. Meng (2020), Phys. Dark Univ. **30**, 100729.

[1876] Ye, G., B. Hu, and Y.-S. Piao (2021), Phys. Rev. D **104** (6), 063510, arXiv:2103.09729 [astro-ph.CO] .

[1877] Ye, G., and Y.-S. Piao (2020), Phys. Rev. D **101** (8), 083507, arXiv:2001.02451 [astro-ph.CO] .

[1878] Ye, G., and Y.-S. Piao (2020), Phys. Rev. D **102** (8), 083523, arXiv:2008.10832 [astro-ph.CO] .

[1879] Ye, G., J. Zhang, and Y.-S. Piao (2021), arXiv:2107.13391 [astro-ph.CO] .

[1880] Yennapureddy, M. K., and F. Melia (2017), JCAP **11**, 029, arXiv:1711.03454 [astro-ph.CO] .

[1881] Yin, L. (2022), Eur. Phys. J. C **82** (1), 78, arXiv:2012.13917 [astro-ph.CO] .

[1882] York, D. G., *et al.* (SDSS) (2000), Astron. J. **120**, 1579, arXiv:astro-ph/0006396 .

[1883] Yoshida, N., T. Abel, L. Hernquist, and N. Sugiyama (2003), Astrophys. J. **592**, 645, arXiv:astro-ph/0301645 .

[1884] Yu, H., B. Ratra, and F.-Y. Wang (2018), Astrophys. J. **856** (1), 3, arXiv:1711.03437 [astro-ph.CO] .

[1885] Yuan, W., L. M. Macri, A. G. Riess, T. G. Brink, S. Casertano, A. V. Filippenko, S. L. Hoffmann, C. D. Huang, and D. Scolnic (2022), arXiv:2203.06681 [astro-ph.GA] .

[1886] Yuan, W., A. G. Riess, L. M. Macri, S. Casertano, and D. Scolnic (2019), Astrophys. J. **886**, 61, arXiv:1908.00993 [astro-ph.GA] .

[1887] Zackrisson, E., C. E. Rydberg, D. Schaerer, G. Ostlin, and M. Tuli (2011), Astrophys. J. **740**, 13, arXiv:1105.0921 [astro-ph.CO] .

[1888] Zaldarriaga, M., and U. Seljak (1997), Phys. Rev. D **55**, 1830, arXiv:astro-ph/9609170 .

[1889] Zaldarriaga, M., and U. Seljak (1998), Phys. Rev. D **58**, 023003, arXiv:astro-ph/9803150 .

[373] Zarrouk, P., *et al.* (2018), Mon. Not. Roy. Astron. Soc. **477** (2), 1639, arXiv:1801.03062 [astro-ph.CO] .

[1891] Zavala, J., M. Vogelsberger, and M. G. Walker (2013), Mon. Not. Roy. Astron. Soc. **431**, L20, arXiv:1211.6426 [astro-ph.CO] .

[1892] Zeng, H., and D. Yan (2019), 10.3847/1538-4357/ab35e3, arXiv:1907.10965 [astro-ph.HE] .

[1893] Zennaro, M., J. Bel, J. Dossett, C. Carbone, and L. Guzzo (2018), Mon. Not. Roy. Astron. Soc. **477** (1), 491, arXiv:1712.02886 [astro-ph.CO] .

[1894] Zentner, A. R., A. V. Kravtsov, O. Y. Gnedin, and A. A. Klypin (2005), Astrophys. J. **629**, 219, arXiv:astro-ph/0502496 .

[1895] Zhang, B. R., M. J. Childress, T. M. Davis, N. V. Karpenka, C. Lidman, B. P. Schmidt, and M. Smith (2017), Mon. Not. Roy. Astron. Soc. **471** (2), 2254, arXiv:1706.07573 [astro-ph.CO] .

[1896] Zhang, M.-J., and H. Li (2018), Eur. Phys. J. C **78** (6), 460, arXiv:1806.02981 [astro-ph.CO] .

[1897] Zhang, M.-J., and J.-Q. Xia (2016), JCAP **12**, 005, arXiv:1606.04398 [astro-ph.CO] .

[1898] Zhang, P., G. D'Amico, L. Senatore, C. Zhao, and Y. Cai (2022), JCAP **02** (02), 036, arXiv:2110.07539 [astro-ph.CO] .

[1899] Zhang, R., and D. Huterer (2010), Astropart. Phys. **33**, 69, arXiv:0908.3988 [astro-ph.CO] .

[1900] Zhang, X. (2005), Mod. Phys. Lett. A **20**, 2575, arXiv:astro-ph/0503072 .

[1901] Zhang, X., and Q.-G. Huang (2019), Commun. Theor. Phys. **71** (7), 826, arXiv:1812.01877 [astro-ph.CO] .

[1902] Zhao, C., *et al.* (2021), Mon. Not. Roy. Astron. Soc. **503** (1), 1149, arXiv:2007.08997 [astro-ph.CO] .

[1903] Zhao, G.-B., *et al.* (2017), Nature Astron. **1** (9), 627, arXiv:1701.08165 [astro-ph.CO] .

[1904] Zhao, G.-B., *et al.* (2019), Mon. Not. Roy. Astron. Soc. **482** (3), 3497, arXiv:1801.03043 [astro-ph.CO] .

[1905] Zhao, G.-B., *et al.* (2020), 10.1093/mnras/stab849, arXiv:2007.09011 [astro-ph.CO] .

[1906] Zhao, Q., M. Faizal, and Z. Zaz (2017), Phys. Lett. **B770**, 564, arXiv:1707.00636 [hep-th] .

[1907] Zhao, W. (2013), Mon. Not. Roy. Astron. Soc. **433**, 3498, arXiv:1209.1174 [astro-ph.CO] .

[1908] Zhou, Z., G. Liu, Y. Mu, and L. Xu (2022), Mon. Not. Roy. Astron. Soc. **511** (1), 595, arXiv:2105.04258 [astro-ph.CO] .

[1909] Zlatev, I., L.-M. Wang, and P. J. Steinhardt (1999), Phys. Rev. Lett. **82**, 896, arXiv:astro-ph/9807002 .

[378] Zubeldia, I. n., and A. Challinor (2019), Mon. Not. Roy. Astron. Soc. **489** (1), 401, arXiv:1904.07887 [astro-ph.CO] .

[1911] Zumalacarregui, M. (2020), Phys. Rev. D **102** (2), 023523, arXiv:2003.06396 [astro-ph.CO] .

[1912] Zwicky, F. (1933), Helv. Phys. Acta **6**, 110.

[1913] Zwicky, F. (1937), Astrophys. J. **86**, 217.

[1914] Zyla, P. A., *et al.* (Particle Data Group) (2020), PTEP **2020** (8), 083C01.

D.0.3 From Experimental Articles

From Plank 2018 (???)

References

[1] Abazajian, K. N. et al., CMB-S4 Science Book, First Edition. 2016, arXiv:1610.02743

[2] Abbott, B. P. et al., Upper Limits on the Stochastic Gravitational-Wave Background from Advanced LIGO's First Observing Run. 2017, Phys. Rev. Lett., 118, 121101, [Erratum: Phys. Rev. Lett.119 (2017) 029901], arXiv:1612.02029

[3] Achucarro, A., Atal, V., Hu, B., Ortiz, P., & Torrado, J., Inflation with moderately sharp features in the speed of sound: Generalized slow roll and in-in formalism for power spectrum and bispectrum. 2014, Phys. Rev., D90, 023511, arXiv:1404.7522

[4] Achúcarro, A., Atal, V., Ortiz, P., & Torrado, J., Localized correlated features in the CMB power spectrum and primordial bispectrum from a transient reduction in the speed of sound. 2014, Phys. Rev., D89, 103006, arXiv:1311.2552

[5] Achúcarro, A., Gong, J.-O., Hardeman, S., Palma, G. A., & Patil, S. P., Features of heavy physics in the CMB power spectrum. 2011, JCAP, 1101, 030, arXiv:1010.3693

[6] Ackerman, L., Carroll, S. M., & Wise, M. B., Imprints of a Primordial Preferred Direction on the Microwave Background. 2007, Phys. Rev., D75, 083502, arXiv:astro-ph/0701357

[7] Adams, F. C., Bond, J. R., Freese, K., Frieman, J. A., & Olinto, A. V., Natural inflation: Particle physics models, power law spectra for large scale structure, and constraints from COBE. 1993, Phys. Rev., D47, 426, arXiv:hep-ph/9207245

[8] Adams, J. A., Cresswell, B., & Easther, R., Inflationary perturbations from a potential with a step. 2001, Phys. Rev., D64, 123514, arXiv:astro-ph/0102236

[9] Addison, G. E., Watts, D. J., Bennett, C. L., et al., Elucidating ΛCDM: Impact of Baryon Acoustic Oscillation Measurements on the Hubble Constant Discrepancy. 2018, Astrophys. J., 853, 119, arXiv:1707.06547

[10] Ade, P. A. R. et al., Improved Constraints on Cosmology and Foregrounds from BICEP2 and Keck Array Cosmic Microwave Background Data with Inclusion of 95 GHz Band. 2016, Phys. Rev. Lett., 116, 031302, arXiv:1510.09217

[11] Ade, P. A. R. et al., BICEP2 / Keck Array X: Constraints on Primordial Gravitational Waves using Planck, WMAP, and New BICEP2/Keck Observations through the 2015 Season. 2018, Phys. Rev. Lett., 121, 221301, arXiv:1810.05216

[12] Adshead, P., Dvorkin, C., Hu, W., & Lim, E. A., Non-Gaussianity from Step Features in the Inflationary Potential. 2012, Phys. Rev., D85, 023531, arXiv:1110.3050

[13] Adshead, P. & Hu, W., Bounds on nonadiabatic evolution in single-field inflation. 2014, Phys. Rev., D89, 083531, arXiv:1402.1677

[15] Alam, S. et al., The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. 2016, Submitted to: Mon. Not. Roy. Astron. Soc., arXiv:1607.03155

[15] Albrecht, A. & Steinhardt, P. J., Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking. 1982, Phys. Rev. Lett., 48, 1220

[16] Amendola, L., Gordon, C., Wands, D., & Sasaki, M., Correlated perturbations from inflation and the cosmic microwave background. 2002, Phys. Rev. Lett., 88, 211302, arXiv:astro-ph/0107089

[30] Audren, B., Lesgourgues, J., Benabed, K., & Prunet, S., Conservative constraints on early cosmology: An illustration of the Monte Python cosmological parameter inference code. 2013, JCAP, 1302, 001, arXiv:1210.7183

[18] Bardeen, J. M., Steinhardt, P. J., & Turner, M. S., Spontaneous Creation of Almost Scale - Free Density Perturbations in an Inflationary Universe. 1983, Phys. Rev., D28, 679

[19] Bartolo, N., Cannone, D., & Matarrese, S., The Effective Field Theory of Inflation Models with Sharp Features. 2013, JCAP, 1310, 038, arXiv:1307.3483

[20] Bartolo, N. & Liddle, A. R., The Simplest curvaton model. 2002, Phys. Rev., D65, 121301, arXiv:astro-ph/0203076

[21] Bartolo, N., Matarrese, S., & Riotto, A., Adiabatic and isocurvature perturbations from inflation: Power spectra and consistency relations. 2001, Phys. Rev., D64, 123504, arXiv:astro-ph/0107502

[22] Bartolo, N., Matarrese, S., & Riotto, A., Gauge-invariant temperature anisotropies and primordial non-Gaussianity. 2004, Phys. Rev. Lett., 93, 231301, arXiv:astro-ph/0407505

[23] Bartolo, N., Matarrese, S., & Riotto, A., On non-Gaussianity in the curvaton scenario. 2004, Phys. Rev., D69, 043503, arXiv:hep-ph/0309033

[24] Bastero-Gil, M., Berera, A., Ramos, R. O., & Rosa, J. G., Warm Little Inflaton. 2016, Phys. Rev. Lett., 117, 151301, arXiv:1604.08838

[25] Bean, R., Dunkley, J., & Pierpaoli, E., Constraining Isocurvature Initial Conditions with WMAP 3-year data. 2006, Phys. Rev., D74, 063503, arXiv:astro-ph/0606685

[26] Behbahani, S. R., Dymarsky, A., Mirbabayi, M., & Senatore, L., (Small) Resonant non-Gaussianities: Signatures of a Discrete Shift Symmetry in the Effective Field Theory of Inflation. 2012, JCAP, 1212, 036, arXiv:1111.3373

[27] Behbahani, S. R. & Green, D., Collective Symmetry Breaking and Resonant Non-Gaussianity. 2012, JCAP, 1211, 056, arXiv:1207.2779

[28] Beltran, M., García-Bellido, J., Lesgourgues, J., & Riazuelo, A., Bounds on CDM and neutrino isocurvature perturbations from CMB and LSS data. 2004, Phys. Rev., D70, 103530, arXiv:astro-ph/0409326

[29] Beltrán, M., García-Bellido, J., Lesgourgues, J., & Viel, M., Squeezing the window on isocurvature modes with the Lyman-α forest. 2005, Phys. Rev., D72, 103515, arXiv:astro-ph/0509209

[30] Benetti, M., Updating constraints on inflationary features in the primordial power spectrum with the Planck data. 2013, Phys. Rev., D88, 087302, arXiv:1308.6406

[31] Bennett, C., Larson, D., Weiland, J., et al., Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. 2012, arXiv:1212.5225

[32] Bennett, C. L., Hill, R. S., Hinshaw, G., et al., Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Are There Cosmic Microwave Background Anomalies? 2011, ApJS, 192, 17, arXiv:1001.4758

[33] Bennett, C. L., Larson, D., Weiland, J. L., et al., Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. 2013, ApJS, 208, 20, arXiv:1212.5225

[34] Berera, A., Warm inflation. 1995, Phys. Rev. Lett., 75, 3218, arXiv:astro-ph/9509049

[47] Beutler, F., Blake, C., Colless, M., et al., The 6dF Galaxy Survey: Baryon Acoustic Oscillations and the Local Hubble Constant. 2011, MNRAS, 416, 3017, arXiv:1106.3366

[36] Bezrukov, F. & Shaposhnikov, M., The Standard Model Higgs boson as the inflaton. 2008, Phys. Lett., B659, 703

[37] Bezrukov, F. L. & Gorbunov, D. S., Distinguishing between R^2-inflation and Higgs-inflation. 2012, Phys. Lett., B713, 365, arXiv:1111.4397

[51] BICEP2/Keck Array and Planck Collaborations, Joint Analysis of BICEP2/Keck Array and Planck Data. 2015, Phys. Rev. Lett., 114, 101301, arXiv:1502.00612

[56] Blas, D., Lesgourgues, J., & Tram, T., The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes. 2011, JCAP, 1107, 034, arXiv:1104.2933

[40] Boubekeur, L. & Lyth, D., Hilltop inflation. 2005, JCAP, 0507, 010, arXiv:hep-ph/0502047

[41] Boyle, L. A., Steinhardt, P. J., & Turok, N., The Cosmic gravitational wave background in a cyclic universe. 2004, Phys. Rev., D69, 127302, arXiv:hep-th/0307170

[42] Bozza, V., Giovannini, M., & Veneziano, G., Cosmological perturbations from a new physics hypersurface. 2003, JCAP, 0305, 001, arXiv:hep-th/0302184

[43] Brandenberger, R. H., Nayeri, A., Patil, S. P., & Vafa, C., Tensor Modes from a Primordial Hagedorn Phase of String Cosmology. 2007, Phys. Rev. Lett., 98, 231302, arXiv:hep-th/0604126

[44] Brout, R., Englert, F., & Gunzig, E., The Creation of the Universe as a Quantum Phenomenon. 1978, Annals Phys., 115, 78

[45] Bucher, M., Dunkley, J., Ferreira, P., Moodley, K., & Skordis, C., The Initial conditions of the universe: How much isocurvature is allowed? 2004, Phys. Rev. Lett., 93, 081301, arXiv:astro-ph/0401417

[46] Bucher, M., Goldhaber, A. S., & Turok, N., An open universe from inflation. 1995, Phys. Rev., D52, 3314, arXiv:hep-ph/9411206

[47] Bucher, M., Moodley, K., & Turok, N., The General primordial cosmic perturbation. 2000, Phys. Rev., D62, 083508, arXiv:astro-ph/9904231

[48] Burgess, C., Martineau, P., Quevedo, F., Rajesh, G., & Zhang, R., Brane - anti-brane inflation in orbifold and orientifold models. 2002, JHEP, 0203, 052, arXiv:hep-th/0111025

[49] Burgess, C. P., Cicoli, M., & Quevedo, F., String Inflation After Planck 2013. 2013, JCAP, 1311, 003, arXiv:1306.3512

[50] Byrnes, C. T., Regan, D., Seery, D., & Tarrant, E. R. M., Implications of the cosmic microwave background power asymmetry for the early universe. 2016, Phys. Rev., D93, 123003, arXiv:1601.01970

[51] Byrnes, C. T., Regan, D., Seery, D., & Tarrant, E. R. M., The hemispherical asymmetry from a scale-dependent inflationary bispectrum. 2016, JCAP, 1606, 025, arXiv:1511.03129

[52] Byrnes, C. T. & Wands, D., Curvature and isocurvature perturbations from two-field inflation in a slow-roll expansion. 2006, Phys. Rev., D74, 043529, arXiv:astro-ph/0605679

[53] Cabass, G., Pagano, L., Salvati, L., et al., Updated Constraints and Forecasts on Primordial Tensor Modes. 2016, Phys. Rev., D93, 063508, arXiv:1511.05146

[66] Calabrese, E., Slosar, A., Melchiorri, A., Smoot, G. F., & Zahn, O., Cosmic Microwave Weak lensing data as a test for the dark universe. 2008, Phys. Rev. D, 77, 123531, arXiv:0803.2309

[55] Chen, X., Primordial Non-Gaussianities from Inflation Models. 2010, Advances in Astronomy, 2010, arXiv:1002.1416

[56] Chen, X., Primordial Features as Evidence for Inflation. 2012, JCAP, 1201, 038, arXiv:1104.1323

[57] Chen, X., Easther, R., & Lim, E. A., Large Non-Gaussianities in Single Field Inflation. 2007, JCAP, 0706, 023, arXiv:astro-ph/0611645

[58] Chen, X., Easther, R., & Lim, E. A., Generation and Characterization of Large Non-Gaussianities in Single Field Inflation. 2008, JCAP, 0804, 010, arXiv:0801.3295

[59] Chluba, J., Hamann, J., & Patil, S. P., Features and New Physical Scales in Primordial Observables: Theory and Observation. 2015, Int. J. Mod. Phys., D24, 1530023, arXiv:1505.01834

[60] Chung, D. J. H. & Upadhye, A., A Hint of a Blue Axion Isocurvature Spectrum? 2017, arXiv:1711.06736

[61] Cicoli, M., Burgess, C., & Quevedo, F., Fibre Inflation: Observable Gravity Waves from IIB String Compactifications. 2009, JCAP, 0903, 013, arXiv:0808.0691

[62] Cline, J. M., Crotty, P., & Lesgourgues, J., Does the small CMB quadrupole moment suggest new physics? 2003, JCAP, 0309, 010, arXiv:astro-ph/0304558

[63] Contaldi, C. R., Peloso, M., Kofman, L., & Linde, A. D., Suppressing the lower multipoles in the CMB anisotropies. 2003, JCAP, 0307, 002, arXiv:astro-ph/0303636

[64] Contreras, D., Hutchinson, J., Moss, A., Scott, D., & Zibin, J. P., Closing in on the large-scale CMB power asymmetry. 2018, Phys. Rev., D97, 063504, arXiv:1709.10134

[65] Contreras, D., Zibin, J. P., Scott, D., Banday, A. J., & Górski, K. M., Testing physical models for dipolar asymmetry with CMB polarization. 2017, Phys. Rev., D96, 123522, arXiv:1704.03143

[66] Danielsson, U. H., A Note on inflation and trans-Planckian physics. 2002, Phys. Rev., D66, 023511, arXiv:hep-th/0203198

[67] Dunkley, J., Bucher, M., Ferreira, P., Moodley, K., & Skordis, C., Measuring the geometry of the Universe in the presence of isocurvature modes. 2005, Phys. Rev. Lett., 95, 261303, arXiv:astro-ph/0507473

[68] Durakovic, A., Hunt, P., Mukherjee, S., Sarkar, S., & Souradeep, T., Reconstruction of a direction-dependent primordial power spectrum from Planck CMB data. 2018, JCAP, 1802, 012, arXiv:1711.08441

[69] Dvali, G., Shafi, Q., & Solganik, S., D-brane inflation. 2001, arXiv:hep-th/0105203

[70] Dvali, G. & Tye, S. H., Brane inflation. 1999, Phys. Lett., B450, 72, arXiv:hep-ph/9812483

[71] Dvali, G. R., Shafi, Q., & Schaefer, R. K., Large scale structure and supersymmetric inflation without fine tuning. 1994, Phys. Rev. Lett., 73, 1886, arXiv:hep-ph/9406319

[72] Easther, R. & Flauger, R., Planck Constraints on Monodromy Inflation. 2014, JCAP, 1402, 037, arXiv:1308.3736

[73] Easther, R., Kinney, W. H., & Peiris, H., Observing trans-Planckian signatures in the cosmic microwave background. 2005, JCAP, 0505, 009, arXiv:astro-ph/0412613

[74] Efstathiou, G. & Bond, J. R., Isocurvature cold dark matter fluctuations. 1986, MNRAS, 218, 103

[75] Efstathiou, G. & Bond, J. R., Microwave anisotropy constraints on isocurvature baryon models. 1987, MNRAS, 227, 33P

[76] Efstathiou, G. & Bond, J. R., Cosmic confusion: Degeneracies among cosmological parameters derived from measurements of microwave background anisotropies. 1999, Mon. Not. Roy. Astron. Soc., 304, 75, arXiv:astro-ph/9807103

[77] Enqvist, K., Kurki-Suonio, H., & Valiviita, J., Limits on isocurvature fluctuations from boomerang and MAXIMA. 2000, Phys. Rev., D62, 103003, arXiv:astro-ph/0006429

[78] Enqvist, K., Kurki-Suonio, H., & Valiviita, J., Open and closed CDM isocurvature models contrasted with the CMB data. 2002, Phys. Rev., D65, 043002, arXiv:astro-ph/0108422

[79] Enqvist, K. & Sloth, M. S., Adiabatic CMB perturbations in pre-big bang string cosmology. 2002, Nucl. Phys., B626, 395, arXiv:hep-ph/0109214

[80] Erickcek, A. L., Hirata, C. M., & Kamionkowski, M., A Scale-Dependent Power Asymmetry from Isocurvature Perturbations. 2009, Phys. Rev., D80, 083507, arXiv:0907.0705

[81] Eriksen, H. K., Hansen, F. K., Banday, A. J., Górski, K. M., & Lilje, P. B., Asymmetries in the Cosmic Microwave Background anisotropy field. 2004, Astrophys. J., 605, 14, [Erratum: Astrophys. J.609,1198(2004)], arXiv:astro-ph/0307507

[82] Fakir, R. & Unruh, W., Improvement on cosmological chaotic inflation through nonminimal coupling. 1990, Phys. Rev., D41, 1783

[83] Fergusson, J. R., Gruetjen, H. F., Shellard, E. P. S., & Liguori, M., Combining power spectrum and bispectrum measurements to detect oscillatory features. 2015, Phys. Rev., D91, 023502, arXiv:1410.5114

[84] Fergusson, J. R., Gruetjen, H. F., Shellard, E. P. S., & Wallisch, B., Polyspectra searches for sharp oscillatory features in cosmic microwave sky data. 2015, Phys. Rev. D, 91, 123506, arXiv:1412.6152

[85] Fergusson, J. R., Liguori, M., & Shellard, E. P. S., General CMB and primordial bispectrum estimation: Mode expansion, map making, and measures of F_{NL}. 2010, Phys. Rev. D, 82, 023502, arXiv:0912.5516

[86] Fergusson, J. R., Liguori, M., & Shellard, E. P. S., The CMB bispectrum. 2012, J. Cosmology Astropart. Phys., 12, 032, arXiv:1006.1642

[87] Fergusson, J. R. & Shellard, E. P. S., The shape of primordial non-Gaussianity and the CMB bispectrum. 2009, Phys. Rev., D80, 043510, arXiv:0812.3413

[88] Feroz, F., Hobson, M., & Bridges, M., MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. 2009, MNRAS, 398, 1601, arXiv:0809.3437

[89] Feroz, F., Hobson, M., Cameron, E., & Pettitt, A., Importance Nested Sampling and the MultiNest Algorithm. 2013, arXiv:1306.2144

[90] Ferrara, S. & Kallosh, R., Seven-disk manifold, α-attractors, and B modes. 2016, Phys. Rev., D94, 126015, arXiv:1610.04163

[91] Ferrara, S., Kallosh, R., Linde, A., & Porrati, M., Minimal Supergravity Models of Inflation. 2013, Phys. Rev., D88, 085038, arXiv:1307.7696

[92] Finelli, F., Hamann, J., Leach, S. M., & Lesgourgues, J., Single-field inflation constraints from CMB and SDSS data. 2010, JCAP, 1004, 011, arXiv:0912.0522

[93] Finelli, F. et al., Exploring cosmic origins with CORE: Inflation. 2018, JCAP, 1804, 016, arXiv:1612.08270

[94] Flauger, R., McAllister, L., Pajer, E., Westphal, A., & Xu, G., Oscillations in the CMB from Axion Monodromy Inflation. 2010, JCAP, 1006, 009, arXiv:0907.2916

[95] Flauger, R., McAllister, L., Pajer, E., Westphal, A., & Xu, G., Oscillations in the CMB from axion monodromy inflation. 2010, JCAP, 6, 9, arXiv:0907.2916

[96] Flauger, R., McAllister, L., Silverstein, E., & Westphal, A., Drifting Oscillations in Axion Monodromy. 2017, JCAP, 1710, 055, arXiv:1412.1814

[97] Flauger, R., Mirbabayi, M., Senatore, L., & Silverstein, E., Productive Interactions: Heavy particles and non-Gaussianity. 2017, JCAP, 1710, 058, arXiv:1606.00513

[98] Flauger, R. & Pajer, E., Resonant Non-Gaussianity. 2011, JCAP, 1101, 017, arXiv:1002.0833

[99] Freese, K., Frieman, J. A., & Olinto, A. V., Natural inflation with pseudo Nambu-Goldstone bosons. 1990, Phys. Rev. Lett., 65, 3233

[100] Garcia-Bellido, J., Figueroa, D. G., & Rubio, J., Preheating in the Standard Model with the Higgs-Inflaton coupled to gravity. 2009, Phys. Rev., D79, 063531, arXiv:0812.4624

[101] García-Bellido, J., Rabadan, R., & Zamora, F., Inflationary scenarios from branes at angles. 2002, JHEP, 0201, 036, arXiv:hep-th/0112147

[102] García-Bellido, J. & Wands, D., Metric perturbations in two-field inflation. 1996, Phys. Rev., D53, 5437, arXiv:astro-ph/9511029

[103] Garriga, J. & Mukhanov, V. F., Perturbations in k-inflation. 1999, Physics Letters B, 458, 219, arXiv:hep-th/9904176

[104] Gasperini, M. & Veneziano, G., Pre - big bang in string cosmology. 1993, Astropart. Phys., 1, 317, arXiv:hep-th/9211021

[105] Gauthier, C. & Bucher, M., Reconstructing the primordial power spectrum from the CMB. 2012, JCAP, 1210, 050, arXiv:astro-ph/1209.2147

[106] Goncharov, A. & Linde, A. D., Chaotic Inflation of the Universe in Supergravity. 1984, Sov. Phys. JETP, 59, 930

[107] Gong, J.-O. & Stewart, E. D., The density perturbation power spectrum to second order corrections in the slow roll expansion. 2001, Phys. Lett., B510, 1, arXiv:astro-ph/0101225

[108] Gong, J.-O. & Yamaguchi, M., Correlated primordial spectra in effective theory of inflation. 2017, Phys. Rev., D95, 083510, arXiv:1701.05875

[109] Gordon, C. & Lewis, A., Observational constraints on the curvaton model of inflation. 2003, Phys. Rev., D67, 123513, arXiv:astro-ph/0212248

[110] Gordon, C. & Pritchard, J. R., Forecasted 21 cm constraints on compensated isocurvature perturbations. 2009, Phys. Rev., D80, 063535, arXiv:0907.5400

[111] Gordon, C., Wands, D., Bassett, B. A., & Maartens, R., Adiabatic and entropy perturbations from inflation. 2001, Phys. Rev., D63, 023506, arXiv:astro-ph/0009131

[112] Gott, J., Creation of Open Universes from de Sitter Space. 1982, Nature, 295, 304

[113] Gott, J. & Statler, T., Constraints on the formation of bubble universes. 1984, Phys. Lett., B136, 157

[114] Graham, R. L., Knuth, D. E., & Patashnik, O. 1994, Concrete Mathematics: A Foundation for Computer Science, 2nd edn. (Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.)

[115] Grin, D., Dore, O., & Kamionkowski, M., Compensated Isocurvature Perturbations and the Cosmic Microwave Background. 2011, Phys. Rev., D84, 123003, arXiv:1107.5047

[116] Grin, D., Dore, O., & Kamionkowski, M., Do baryons trace dark matter in the early universe? 2011, Phys. Rev. Lett., 107, 261301, arXiv:1107.1716

[117] Grin, D., Hanson, D., Holder, G. P., Doré, O., & Kamionkowski, M., Baryons do trace dark matter 380,000 years after the big bang: Search for compensated isocurvature perturbations with WMAP 9-year data. 2014, Phys. Rev., D89, 023006, arXiv:1306.4319

[118] Gruppuso, A., Kitazawa, N., Mandolesi, N., Natoli, P., & Sagnotti, A., Pre-Inflationary Relics in the CMB? 2016, Phys. Dark Univ., 11, 68, arXiv:1508.00411

[119] Guth, A. H., The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems. 1981, Phys. Rev., D23, 347

[120] Guth, A. H., Kaiser, D. I., & Nomura, Y., Inflationary paradigm after Planck 2013. 2014, Phys. Lett., B733, 112, arXiv:1312.7619

[121] Guth, A. H. & Pi, S., Fluctuations in the New Inflationary Universe. 1982, Phys. Rev. Lett., 49, 1110

[122] Haga, T., Inomata, K., Ota, A., & Ravenni, A., Exploring compensated isocurvature perturbations with CMB spectral distortion anisotropies. 2018, arXiv:1805.08773

[123] Hamann, J., Shafieloo, A., & Souradeep, T., Features in the primordial power spectrum? A frequentist analysis. 2010, JCAP, 1004, 010, arXiv:0912.2728

[124] Handley, W., fgivenx: Functional Posterior Plotter. 2018, The Journal of Open Source Software, 3

[125] Handley, W., Lasenby, A., Peiris, H., & Hobson, M., Bayesian inflationary reconstructions from Planck 2018 data. 2019, Phys. Rev. D, in preparation

[126] Handley, W. J., Hobson, M. P., & Lasenby, A. N., POLYCHORD: nested sampling for cosmology. 2015, MNRAS, 450, L61, arXiv:1502.01856

[127] Handley, W. J., Hobson, M. P., & Lasenby, A. N., POLYCHORD: next-generation nested sampling. 2015, MNRAS, 453, 4384, arXiv:1506.00171

[128] Harrison, E. R., Fluctuations at the threshold of classical cosmology. 1970, Phys. Rev., D1, 2726

[129] Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E., A 408 MHz all-sky continuum survey. II - The atlas of contour maps. 1982, A&AS, 47, 1

[130] Hawking, S., The Development of Irregularities in a Single Bubble Inflationary Universe. 1982, Phys. Lett., B115, 295

[131] Hazra, D. K., Shafieloo, A., Smoot, G. F., & Starobinsky, A. A., Inflation with Whip-Shaped Suppressed Scalar Power Spectra. 2014, Phys. Rev. Lett., 113, 071301, arXiv:1404.0360

[132] Hazra, D. K., Shafieloo, A., Smoot, G. F., & Starobinsky, A. A., Primordial features and Planck polarization. 2016, JCAP, 1609, 009, arXiv:1605.02106

[133] Hazra, D. K., Shafieloo, A., & Souradeep, T., Primordial power spectrum from Planck. 2014, JCAP, 1411, 011, arXiv:1406.4827

[134] He, C., Grin, D., & Hu, W., Compensated isocurvature perturbations in the curvaton model. 2015, Phys. Rev., D92, 063018, arXiv:1505.00639

[135] Hee, S., Vázquez, J. A., Handley, W. J., Hobson, M. P., & Lasenby, A. N., Constraining the dark energy equation of state using Bayes theorem and the Kullback-Leibler divergence. 2016, ArXiv e-prints, arXiv:1607.00270

[136] Heinrich, C. H., Grin, D., & Hu, W., Lensing Bias to CMB Measurements of Compensated Isocurvature Perturbations. 2016, Phys. Rev., D94, 043534, arXiv:1605.08439

[137] Hivon, E., Mottet, S., & Ponthieu, N., QuickPol: Fast calculation of effective beam matrices for CMB polarization. 2017, Astron. Astrophys., 598, A25, arXiv:1608.08833

[138] Holder, G. P., Nollett, K. M., & van Engelen, A., On Possible Variation in the Cosmological Baryon Fraction. 2010, Astrophys. J., 716, 907, arXiv:0907.3919

[139] Hou, Z. et al., Constraints on Cosmology from the Cosmic Microwave Background Power Spectrum of the 2500 deg^2 SPT-SZ Survey. 2014, Astrophys. J., 782, 74, arXiv:1212.6267

[140] Hu, W., Generalized Slow Roll for Non-Canonical Kinetic Terms. 2011, Phys. Rev., D84, 027303, arXiv:1104.4500

[141] Hunt, P. & Sarkar, S., Reconstruction of the primordial power spectrum of curvature perturbations using multiple data sets. 2014, JCAP, 1401, 025, arXiv:1308.2317

[142] Hunt, P. & Sarkar, S., Search for features in the spectrum of primordial perturbations using Planck and other datasets. 2015, JCAP, 1512, 052, arXiv:1510.03338

[143] Ijjas, A. & Steinhardt, P. J., Implications of Planck 2015 for inflationary, ekpyrotic and anamorphic bouncing cosmologies. 2016, Class. Quant. Grav., 33, 044001, arXiv:1512.09010

[144] Ijjas, A., Steinhardt, P. J., & Loeb, A., Inflationary paradigm in trouble after Planck2013. 2013, Phys. Lett., B723, 261, arXiv:1304.2785

[145] Jackson, M. G. & Shiu, G., Study of the consistency relation for single-field inflation with power spectrum oscillations. 2013, Phys. Rev., D88, 123511, arXiv:1303.4973

[146] Jackson, M. G., Wandelt, B., & Bouchet, F., Angular Correlation Functions for Models with Logarithmic Oscillations. 2014, Phys. Rev., D89, 023510, arXiv:1303.3499

[147] Jeffreys, H. 1998, Theory of Probability, 3rd edn. (Oxford University Press)

[148] Kachru, S., Kallosh, R., Linde, A. D., et al., Towards inflation in string theory. 2003, JCAP, 0310, 013, arXiv:hep-th/0308055

[149] Kallosh, R., Linde, A., & Roest, D., Superconformal Inflationary α-Attractors. 2013, JHEP, 1311, 198, arXiv:1311.0472

[150] Kallosh, R., Linde, A., Wrase, T., & Yamada, Y., Maximal Supersymmetry and B-Mode Targets. 2017, JHEP, 04, 144, arXiv:1704.04829

[151] Kaloper, N., Lawrence, A., & Sorbo, L., An Ignoble Approach to Large Field Inflation. 2011, JCAP, 1103, 023, arXiv:1101.0026

[152] Kasuya, S. & Kawasaki, M., Axion isocurvature fluctuations with extremely blue spectrum. 2009, Phys. Rev., D80, 023516, arXiv:0904.3800

[153] Kazanas, D., Dynamics of the Universe and Spontaneous Symmetry Breaking. 1980, Astrophys. J., 241, L59

[154] Keskitalo, R., Kurki-Suonio, H., Muhonen, V., & Valiviita, J., Hints of Isocurvature Perturbations in the Cosmic Microwave Background. 2007, JCAP, 0709, 008, arXiv:astro-ph/0611917

[155] Kleban, M. & Schillo, M., Spatial Curvature Falsifies Eternal Inflation. 2012, JCAP, 1206, 029, arXiv:1202.5037

[156] Kobayashi, T., Yamaguchi, M., & Yokoyama, J., G-inflation: Inflation driven by the Galileon field. 2010, Phys. Rev. Lett., 105, 231302, arXiv:1008.0603

[157] Komatsu, E. et al., Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. 2009, Astrophys. J. Suppl., 180, 330, arXiv:0803.0547

[158] Kuhnel, F. & Schwarz, D. J., Large-Scale Suppression from Stochastic Inflation. 2010, Phys. Rev. Lett., 105, 211302, arXiv:1003.3014

[159] Kurki-Suonio, H., Muhonen, V., & Valiviita, J., Correlated Primordial Perturbations in Light of CMB and LSS Data. 2005, Phys. Rev., D71, 063005, arXiv:astro-ph/0412439

[160] Langlois, D., Correlated adiabatic and isocurvature perturbations from double inflation. 1999, Phys. Rev., D59, 123512, arXiv:astro-ph/9906080

[161] Langlois, D. & Riazuelo, A., Correlated mixtures of adiabatic and isocurvature cosmological perturbations. 2000, Phys. Rev., D62, 043504, arXiv:astro-ph/9912497

[162] Leach, S. M., Liddle, A. R., Martin, J., & Schwarz, D. J., Cosmological parameter estimation and the inflationary cosmology. 2002, Phys. Rev., D66, 023515, arXiv:astro-ph/0202094

[163] Lesgourgues, J., The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview. 2011, arXiv:1104.2932

[164] Lewis, A. & Bridle, S., Cosmological parameters from CMB and other data: A Monte Carlo approach. 2002, Phys. Rev. D, 66, 103511, arXiv:astro-ph/0205436

[220] Lewis, A. & Challinor, A., Weak gravitational lensing of the CMB. 2006, Phys. Rept., 429, 1, arXiv:astro-ph/0601594

[166] Lewis, A., Challinor, A., & Lasenby, A., Efficient Computation of CMB Anisotropies in Closed FRW Models. 2000, Astrophys. J., 538, 473, arXiv:astro-ph/9911177

[167] Liddle, A. R. & Leach, S. M., How long before the end of inflation were observable perturbations produced? 2003, Phys. Rev., D68, 103503, arXiv:astro-ph/0305263

[168] Linde, A. 2015, in Proceedings, 100th Les Houches Summer School: Post-Planck Cosmology: Les Houches, France, July 8 - August 2, 2013, 231–316

[169] Linde, A. 2017, in Black Holes, Gravitational Waves and Spacetime Singularities Rome, Italy, May 9-12, 2017

[170] Linde, A. D., A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems. 1982, Phys. Lett., B108, 389

[171] Linde, A. D., Chaotic Inflation. 1983, Phys. Lett., B129, 177

[172] Linde, A. D., Generation of Isothermal Density Perturbations in the Inflationary Universe. 1985, Phys. Lett., B158, 375

[173] Linde, A. D. & Mukhanov, V. F., Non-Gaussian isocurvature perturbations from inflation. 1997, Phys. Rev., D56, 535, arXiv:astro-ph/9610219

[174] Lucchin, F., Matarrese, S., & Pollock, M. D., Inflation with a Non-minimally Coupled Scalar Field. 1986, Phys. Lett., 167B, 163

[175] Lyth, D. H. & Stewart, E. D., Inflationary density perturbations with Omega $<$ 1. 1990, Phys. Lett., B252, 336

[176] Lyth, D. H., Ungarelli, C., & Wands, D., The Primordial Density Perturbation in the Curvaton Scenario. 2003, Phys. Rev., D67, 023503, arXiv:astro-ph/0208055

[177] Lyth, D. H. & Wands, D., Generating the curvature perturbation without an inflaton. 2002, Phys. Lett., B524, 5, arXiv:hep-ph/0110002

[178] Maleknejad, A., Sheikh-Jabbari, M. M., & Soda, J., Gauge Fields and Inflation. 2013, Phys. Rept., 528, 161, arXiv:1212.2921

[179] Marsh, D. J. E., Axion Cosmology. 2016, Phys. Rept., 643, 1, arXiv:1510.07633

[180] Martin, J. & Brandenberger, R., On the dependence of the spectra of fluctuations in inflationary cosmology on trans-Planckian physics. 2003, Phys. Rev., D68, 063513, arXiv:hep-th/0305161

[181] Martin, J. & Ringeval, C., First CMB Constraints on the Inflationary Reheating Temperature. 2010, Phys. Rev., D82, 023511, arXiv:1004.5525

[182] Martin, J., Ringeval, C., Trotta, R., & Vennin, V., The Best Inflationary Models After Planck. 2014, JCAP, 1403, 039, arXiv:1312.3529

[183] Martin, J., Ringeval, C., & Vennin, V., Encyclopædia Inflationaris. 2014, Phys. Dark Univ., arXiv:1303.3787

[184] McAllister, L., Silverstein, E., & Westphal, A., Gravity Waves and Linear Inflation from Axion Monodromy. 2010, Phys. Rev., D82, 046003, arXiv:0808.0706

[185] McAllister, L., Silverstein, E., Westphal, A., & Wrase, T., The Powers of Monodromy. 2014, JHEP, 09, 123, arXiv:1405.3652

[186] Meerburg, P. D., Alleviating the tension at low ℓ through axion monodromy. 2014, Phys. Rev., D90, 063529, arXiv:1406.3243

[187] Meerburg, P. D., Hloçek, R., Hadzhiyska, B., & Meyers, J., Multiwavelength constraints on the inflationary consistency relation. 2015, Phys. Rev., D91, 103505, arXiv:1502.00302

[188] Meerburg, P. D. & Münchmeyer, M., Optimal CMB estimators for bispectra from excited states. 2015, Phys. Rev., D92, 063527, arXiv:1505.05882

[189] Meerburg, P. D., Münchmeyer, M., & Wandelt, B., Joint resonant CMB power spectrum and bispectrum estimation. 2016, Phys. Rev., D93, 043536, arXiv:1510.01756

[190] Meerburg, P. D. & Pajer, E., Observational Constraints on Gauge Field Production in Axion Inflation. 2013, JCAP, 1302, 017, arXiv:1203.6076

[191] Meerburg, P. D., Spergel, D. N., & Wandelt, B. D. 2014, in Proceedings, 49th Rencontres de Moriond on Cosmology: La Thuile, Italy, March 15-22, 2014, 27–32

[192] Meerburg, P. D., Spergel, D. N., & Wandelt, B. D., Searching for oscillations in the primordial power spectrum. I. Perturbative approach. 2014, Phys. Rev., D89, 063536, arXiv:1308.3704

[193] Meerburg, P. D., Spergel, D. N., & Wandelt, B. D., Searching for oscillations in the primordial power spectrum. II. Constraints from Planck data. 2014, Phys. Rev., D89, 063537, arXiv:1308.3705

[194] Meerburg, P. D., van der Schaar, J. P., & Corasaniti, P. S., Signatures of Initial State Modifications on Bispectrum Statistics. 2009, JCAP, 0905, 018, arXiv:0901.4044

[195] Miranda, V. & Hu, W., Inflationary Steps in the Planck Data. 2014, Phys. Rev., D89, 083529, arXiv:1312.0946

[196] Miranda, V., Hu, W., & Adshead, P., Warp Features in DBI Inflation. 2012, Phys. Rev., D86, 063529, arXiv:1207.2186

[197] Miranda, V., Hu, W., & Dvorkin, C., Polarization Predictions for Inflationary CMB Power Spectrum Features. 2015, Phys. Rev., D91, 063514, arXiv:1411.5956

[198] Mollerach, S., Isocurvature baryon perturbations and inflation. 1990, Phys. Rev., D42, 313

[199] Moodley, K., Bucher, M., Dunkley, J., Ferreira, P., & Skordis, C., Constraints on isocurvature models from the WMAP first-year data. 2004, Phys. Rev., D70, 103520, arXiv:astro-ph/0407304

[200] Mooij, S., Palma, G. A., Panotopoulos, G., & Soto, A., Consistency relations for sharp inflationary non-Gaussian features. 2016, JCAP, 1609, 004, arXiv:1604.03533

[201] Moroi, T. & Takahashi, T., Effects of cosmological moduli fields on cosmic microwave background. 2001, Phys. Lett., B522, 215, arXiv:hep-ph/0110096

[202] Moss, A., Scott, D., Zibin, J. P., & Battye, R., Tilted Physics: A Cosmologically Dipole-Modulated Sky. 2011, Phys. Rev., D84, 023014, arXiv:1011.2990

[243] Muñoz, J. B., Grin, D., Dai, L., Kamionkowski, M., & Kovetz, E. D., Search for Compensated Isocurvature Perturbations with Planck Power Spectra. 2016, Phys. Rev., D93, 043008, arXiv:1511.04441

[204] Mukhanov, V. F., Gravitational Instability of the Universe Filled with a Scalar Field. 1985, JETP Lett., 41, 493

[205] Mukhanov, V. F. & Chibisov, G., Quantum Fluctuation and Nonsingular Universe. (In Russian). 1981, JETP Lett., 33, 532

[206] Mukhanov, V. F. & Chibisov, G., The Vacuum energy and large scale structure of the universe. 1982, Sov. Phys. JETP, 56, 258

[207] Münchmeyer, M., Bouchet, F., Jackson, M. G., & Wandelt, B., The Komatsu Spergel Wandelt estimator for oscillations in the cosmic microwave background bispectrum. 2014, Astron. Astrophys., 570, A94, arXiv:1405.2550

[208] Münchmeyer, M., Meerburg, P. D., & Wandelt, B. D., Optimal estimator for resonance bispectra in the CMB. 2015, Phys. Rev., D91, 043534, arXiv:1412.3461

[209] Obied, G., Dvorkin, C., Heinrich, C., Hu, W., & Miranda, V., Inflationary Features and Shifts in Cosmological Parameters from Planck 2015 Data. 2017, Phys. Rev., D96, 083526, arXiv:1706.09412

[210] Okamoto, T. & Hu, W., CMB lensing reconstruction on the full sky. 2003, Phys. Rev., D67, 083002, arXiv:astro-ph/0301031

[211] Ooba, J., Ratra, B., & Sugiyama, N., Planck 2015 constraints on the non-flat ΛCDM inflation model. 2017, arXiv:1707.03452

[212] Palma, G. A., Untangling features in the primordial spectra. 2015, JCAP, 1504, 035, arXiv:1412.5615

[213] Peebles, P. & Yu, J., Primeval adiabatic perturbation in an expanding universe. 1970, Astrophys. J., 162, 815

[214] Peebles, P. J. E. & Yu, J. T., Primeval Adiabatic Perturbation in an Expanding Universe. 1970, Astrophys. J., 162, 815

[215] Peiris, H., Easther, R., & Flauger, R., Constraining Monodromy Inflation. 2013, JCAP, 1309, 018, arXiv:1303.2616

[216] Peiris, H. et al., First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Implications for inflation. 2003, Astrophys. J. Suppl., 148, 213, arXiv:astro-ph/0302225

[217] Pierpaoli, E., Garcia-Bellido, J., & Borgani, S., Microwave background anisotropies and large scale structure constraints on isocurvature modes in a two field model of inflation. 1999, JHEP, 10, 015, arXiv:hep-ph/9909420

[275] Planck Collaboration XV, *Planck* 2013 results. XV. CMB power spectra and likelihood. 2014, A&A, 571, A15, arXiv:1303.5075

[279] Planck Collaboration XXII, *Planck* 2013 results. XXII. Constraints on inflation. 2014, A&A, 571, A22, arXiv:1303.5082

[220] Planck Collaboration XXIII, *Planck* 2013 results. XXIII. Isotropy and statistics of the CMB. 2014, A&A, 571, A23, arXiv:1303.5083

[280] Planck Collaboration XXIV, *Planck* 2013 results. XXIV. Constraints on primordial non-Gaussianity. 2014, A&A, 571, A24, arXiv:1303.5084

[282] Planck Collaboration I, *Planck* 2015 results. I. Overview of products and results. 2016, A&A, 594, A1, arXiv:1502.01582

[283] Planck Collaboration XI, *Planck* 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters. 2016, A&A, 594, A11, arXiv:1507.02704

[284] Planck Collaboration XIII, *Planck* 2015 results. XIII. Cosmological parameters. 2016, A&A, 594, A13, arXiv:1502.01589

[286] Planck Collaboration XV, *Planck* 2015 results. XV. Gravitational lensing. 2016, A&A, 594, A15, arXiv:1502.01591

[226] Planck Collaboration XVI, *Planck* 2015 results. XVI. Isotropy and statistics of the CMB. 2016, A&A, 594, A16, arXiv:1506.07135

[287] Planck Collaboration XVII, *Planck* 2015 results. XVII. Constraints on primordial non-Gaussianity. 2016, A&A, 594, A17, arXiv:1502.01592

[228] Planck Collaboration XVIII, *Planck* 2015 results. XVIII. Background geometry and topology of the Universe. 2016, A&A, 594, A18, arXiv:1502.01593

[288] Planck Collaboration XX, *Planck* 2015 results. XX. Constraints on inflation. 2016, A&A, 594, A20, arXiv:1502.02114

[230] Planck Collaboration I, *Planck* 2018 results. I. Overview, and the cosmological legacy of *Planck*. 2018, A&A, submitted, arXiv:1807.06205

[290] Planck Collaboration II, *Planck* 2018 results. II. Low Frequency Instrument data processing. 2018, A&A, in press, arXiv:1807.06206

[291] Planck Collaboration III, *Planck* 2018 results. III. High Frequency Instrument data processing. 2018, A&A, in press, arXiv:1807.06207

[292] Planck Collaboration IV, *Planck* 2018 results. IV. Diffuse component separation. 2018, A&A, in press, arXiv:1807.06208

[293] Planck Collaboration V, *Planck* 2018 results. V. Power spectra and likelihoods. 2018, A&A, submitted, arXiv:1907.12875

[235] Planck Collaboration VI, *Planck* 2018 results. VI. Cosmological parameters. 2018, A&A, submitted, arXiv:1807.06209

[236] Planck Collaboration VII, *Planck* 2018 results. VII. Isotropy and statistics. 2018, A&A, in press, arXiv:1906.02552

[294] Planck Collaboration VIII, *Planck* 2018 results. VIII. Gravitational lensing. 2018, A&A, in press, arXiv:1807.06210

[295] Planck Collaboration IX, *Planck* 2018 results. IX. Constraints on primordial non-Gaussianity. 2018, A&A, submitted, arXiv:1905.05697

[297] Planck Collaboration Int. XLVI, *Planck* intermediate results. XLVI. Reduction of large-scale systematic effects in HFI polarization maps and estimation of the reionization optical depth. 2016, A&A, 596, A107, arXiv:1605.02985

[298] Planck Collaboration Int. XLVII, *Planck* intermediate results. XLVII. Constraints on reionization history. 2016, A&A, 596, A108, arXiv:1605.03507

[241] Polarski, D. & Starobinsky, A. A., Isocurvature perturbations in multiple inflationary models. 1994, Phys. Rev., D50, 6123, arXiv:astro-ph/9404061

[242] Ratra, B. & Peebles, P., Inflation in an open universe. 1995, Phys. Rev., D52, 1837

[243] Raveri, M., Martinelli, M., Zhao, G., & Wang, Y., Information Gain in Cosmology: From the Discovery of Expansion to Future Surveys. 2016, ArXiv e-prints, arXiv:1606.06273

[313] Ross, A. J., Samushia, L., Howlett, C., et al., The clustering of the SDSS DR7 main Galaxy sample – I. A 4 per cent distance measure at $z = 0.15$. 2015, Mon. Not. Roy. Astron. Soc., 449, 835, arXiv:1409.3242

[245] Salopek, D., Bond, J., & Bardeen, J. M., Designing Density Fluctuation Spectra in Inflation. 1989, Phys. Rev., D40, 1753

[246] Sasaki, M., Valiviita, J., & Wands, D., Non-Gaussianity of the primordial perturbation in the curvaton model. 2006, Phys. Rev., D74, 103003, arXiv:astro-ph/0607627

[247] Sato, K., First-order phase transition of a vacuum and the expansion of the Universe. 1981, MNRAS, 195, 467

[248] Savelainen, M., Valiviita, J., Walia, P., Rusak, S., & Kurki-Suonio, H., Constraints on neutrino density and velocity isocurvature modes from WMAP-9 data. 2013, Phys. Rev., D88, 063010, arXiv:1307.4398

[249] Scott, D., Contreras, D., Narimani, A., & Ma, Y.-Z., The information content of cosmic microwave background anisotropies. 2016, JCAP, 1606, 046, arXiv:1603.03550

[332] Seljak, U., Gravitational lensing effect on cosmic microwave background anisotropies: A Power spectrum approach. 1996, ApJ, 463, 1, arXiv:astro-ph/9505109

[251] Silverstein, E. & Westphal, A., Monodromy in the CMB: Gravity Waves and String Inflation. 2008, Phys. Rev., D78, 106003, arXiv:0803.3085

[252] Sinha, R. & Souradeep, T., Post-WMAP assessment of infrared cutoff in the primordial spectrum from inflation. 2006, Phys. Rev., D74, 043518, arXiv:astro-ph/0511808

[253] Skilling, J. 2004, in American Institute of Physics Conference Series, Vol. 735, AIP Conference Series, ed. R. Fischer, R. Preuss, & U. V. Toussaint, 395–405

[345] Smith, T. L., Muñoz, J. B., Smith, R., Yee, K., & Grin, D., Baryons still trace dark matter: Probing CMB lensing maps for hidden isocurvature. 2017, Phys. Rev., D96, 083508, arXiv:1704.03461

[255] Soda, J., Statistical Anisotropy from Anisotropic Inflation. 2012, Class. Quant. Grav., 29, 083001, arXiv:1201.6434

[256] Soumagnac, M. T., Barkana, R., Sabiu, C. G., et al., Large-Scale Distribution of Total Mass versus Luminous Matter from Baryon Acoustic Oscillations: First Search in the Sloan Digital Sky Survey III Baryon Oscillation Spectroscopic Survey Data Release 10. 2016, Phys. Rev. Lett., 116, 201302, arXiv:1602.01839

[257] Soumagnac, M. T., Sabiu, C. G., Barkana, R., & Yoo, J., Large scale distribution of mass versus light from Baryon Acoustic Oscillations: Measurement in the final SDSS-III BOSS Data Release 12. 2018, arXiv:1802.10368

[258] Spokoiny, B., Inflation and Generation of Perturabtions in Broken Symmetric Theory of Gravity. 1984, Phys. Lett., B147, 39

[259] Starobinsky, A., The Perturbation Spectrum Evolving from a Nonsingular Initially de Sitter Cosmology and the Microwave Background Anisotropy. 1983, Sov. Astron. Lett., 9, 302

[260] Starobinsky, A. A., A New Type of Isotropic Cosmological Models Without Singularity. 1980, Phys. Lett., B91, 99

[261] Starobinsky, A. A., Dynamics of Phase Transition in the New Inflationary Universe Scenario and Generation of Perturbations. 1982, Phys. Lett., B117, 175

[262] Starobinsky, A. A., Spectrum of adiabatic perturbations in the universe when there are singularities in the inflation potential. 1992, JETP Lett., 55, 489, [Pisma Zh. Eksp. Teor. Fiz.55,477(1992)]

[263] Stewart, E. D., Inflation, supergravity and superstrings. 1995, Phys. Rev., D51, 6847, arXiv:hep-ph/9405389

[264] Stewart, E. D. & Lyth, D. H., A more accurate analytic calculation of the spectrum of cosmological perturbations produced during inflation. 1993, Phys. Lett. B, 302, 171, arXiv:gr-qc/9302019

[265] Stompor, R., Banday, A. J., & Gorski, Krzysztof, M., Flat dark matter dominated models with hybrid adiabatic plus isocurvature initial conditions. 1996, Astrophys. J., 463, 8, arXiv:astro-ph/9511087

[266] Takahashi, T., Yamaguchi, M., & Yokoyama, S., Primordial Non-Gaussianity in Models with Dark Matter Isocurvature Fluctuations. 2009, Phys. Rev., D80, 063524, arXiv:0907.3052

[267] Torrado, J., Hu, B., & Achucarro, A., Robust predictions for an oscillatory bispectrum in Planck 2015 data from transient reductions in the speed of sound of the inflaton. 2017, Phys. Rev., D96, 083515, arXiv:1611.10350

[268] Trotta, R., Applications of Bayesian model selection to cosmological parameters. 2007, Mon. Not. Roy. Astron. Soc., 378, 72, arXiv:astro-ph/0504022

[269] Trotta, R., The isocurvature fraction after WMAP 3-year data. 2007, MNRAS, 375, L26, arXiv:astro-ph/0608116

[270] Trotta, R., Bayes in the sky: Bayesian inference and model selection in cosmology. 2008, Contemp. Phys., 49, 71, arXiv:0803.4089

[271] Tsujikawa, S., Distinguishing between inflationary models from cosmic microwave background. 2014, PTEP, 2014, 06B104, arXiv:1401.4688

[357] Valiviita, J., Power Spectra Based Planck Constraints on Compensated Isocurvature, and Forecasts for LiteBIRD and CORE Space Missions. 2017, JCAP, 1704, 014, arXiv:1701.07039

[273] Valiviita, J. & Giannantonio, T., Constraints on primordial isocurvature perturbations and spatial curvature by Bayesian model selection. 2009, Phys. Rev., D80, 123516, arXiv:0909.5190

[274] Valiviita, J. & Muhonen, V., Correlated adiabatic and isocurvature cmb fluctuations in the wake of WMAP. 2003, Phys. Rev. Lett., 91, 131302, arXiv:astro-ph/0304175

[275] Valiviita, J., Savelainen, M., Talvitie, M., Kurki-Suonio, H., & Rusak, S., Constraints on scalar and tensor perturbations in phenomenological and two-field inflation models: Bayesian evidence for primordial isocurvature and tensor modes. 2012, Astrophys. J., 753, 151, arXiv:1202.2852

[276] van Tent, B., Multiple-field inflation and the CMB. 2004, Class. Quant. Grav., 21, 349, arXiv:astro-ph/0307048

[277] Verde, L., Jimenez, R., Alvarez-Gaume, L., Heavens, A. F., & Matarrese, S., Multi-variate joint PDF for non-Gaussianities: Exact formulation and generic approximations. 2013, JCAP, 1306, 023, arXiv:1301.6017

[278] Vilenkin, A. & Ford, L. H., Gravitational Effects upon Cosmological Phase Transitions. 1982, Phys. Rev., D26, 1231

[279] Wands, D., Bartolo, N., Matarrese, S., & Riotto, A., An observational test of two-field inflation. 2002, Phys. Rev., D66, 043520, arXiv:astro-ph/0205253

[280] Waterhouse, T. P. & Zibin, J. P., The cosmic variance of Ω. 2008, arXiv:0804.1771

[281] Yamamoto, K., Sasaki, M., & Tanaka, T., Large angle CMB anisotropy in an open universe in the one bubble inflationary scenario. 1995, Astrophys. J., 455, 412, arXiv:astro-ph/9501109

[282] Zeldovich, Y., A hypothesis unifying the structure and the entropy of the universe. 1972, MNRAS, 160, 1P

From Plank's VI 2018

References

[1] Abazajian, K. N., Canac, N., Horiuchi, S., & Kaplinghat, M., Astrophysical and Dark Matter Interpretations of Extended Gamma-Ray Emission from the Galactic Center. 2014, Phys. Rev. D, 90, 023526, arXiv:1402.4090

[2] Abbott, B. et al., GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. 2017, Phys. Rev. Lett., 119, 161101, arXiv:1710.05832

[3] Abbott, B. P., Abbott, R., Abbott, T. D., et al., A gravitational-wave standard siren measurement of the Hubble constant. 2017, Nature, 551, 85, arXiv:1710.05835

[4] Abe, K. et al., Atmospheric neutrino oscillation analysis with external constraints in Super-Kamiokande I-IV. 2018, Phys. Rev., D97, 072001, arXiv:1710.09126

[5] Adamson, P. et al., Constraints on Oscillation Parameters from ν_e Appearance and ν_μ Disappearance in NOvA. 2017, Phys. Rev. Lett., 118, 231801, arXiv:1703.03328

[6] Addison, G. E., Huang, Y., Watts, D. J., et al., Quantifying discordance in the 2015 Planck CMB spectrum. 2016, ApJ, 818, 132, arXiv:1511.00055

[7] Addison, G. E., Watts, D. J., Bennett, C. L., et al., Elucidating ΛCDM: Impact of Baryon Acoustic Oscillation Measurements on the Hubble Constant Discrepancy. 2018, ApJ, 853, 119, arXiv:1707.06547

[8] Ade, P. A. R. et al., BICEP2 / Keck Array x: Constraints on Primordial Gravitational Waves using Planck, WMAP, and New BICEP2/Keck Observations through the 2015 Season. 2018, Phys. Rev. Lett., 121, 221301, arXiv:1810.05216

[9] Adelberger, E. G. et al., Solar fusion cross sections II: the pp chain and CNO cycles. 2011, Rev. Mod. Phys., 83, 195, arXiv:1004.2318

[10] Agafonova, N. et al., Final results of the search for $\nu_\mu \rightarrow \nu_e$ oscillations with the OPERA detector in the CNGS beam. 2018, JHEP, 06, 151, arXiv:1803.11400

[11] Agrawal, P., Cyr-Racine, F.-Y., Pinner, D., & Randall, L., Rock 'n' Roll Solutions to the Hubble Tension. 2019, arXiv e-prints, arXiv:1904.01016

[12] Aguilar-Arevalo, A. et al., Evidence for neutrino oscillations from the observation of anti-neutrino(electron) appearance in a anti-neutrino(muon) beam. 2001, Phys. Rev., D64, 112007, arXiv:hep-ex/0104049

[13] Aguilar-Arevalo, A. A. et al., Significant Excess of ElectronLike Events in the MiniBooNE Short-Baseline Neutrino Experiment. 2018, Phys. Rev. Lett., 121, 221801, arXiv:1805.12028

[14] Akrami, Y., Kallosh, R., Linde, A., & Vardanyan, V., Dark energy, α-attractors, and large-scale structure surveys. 2018, J. Cosmology Astropart. Phys., 6, 041, arXiv:1712.09693

[15] Alam, S. et al., The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. 2017, MNRAS, 470, 2617, arXiv:1607.03155

[16] Alcock, C. & Paczynski, B., An evolution free test for non-zero cosmological constant. 1979, Nature, 281, 358

[17] Ali-Haïmoud, Y. & Hirata, C. M., Ultrafast effective multilevel atom method for primordial hydrogen recombination. 2010, Phys. Rev. D, 82, 063521, arXiv:1006.1355

[18] Ali-Haimoud, Y. & Hirata, C. M., HyRec: A fast and highly accurate primordial hydrogen and helium recombination code. 2011, Phys. Rev. D, 83, 043513, arXiv:1011.3758

[19] Allison, R., Caucal, P., Calabrese, E., Dunkley, J., & Louis, T., Towards a cosmological neutrino mass detection. 2015, Phys. Rev., D92, 123535, arXiv:1509.07471

[20] Almazán, H. et al., Sterile neutrino exclusion from the STEREO experiment with 66 days of reactor-on data. 2018, ArXiv e-prints, arXiv:1806.02096

[21] Amendola, L., Coupled quintessence. 2000, Phys. Rev., D62, 043511, arXiv:astro-ph/9908023

[22] Amendola, L., Kunz, M., & Sapone, D., Measuring the dark side (with weak lensing). 2008, J. Cosmology Astropart. Phys., 0804, 013, arXiv:0704.2421

[23] Anderson, L. et al., The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations in the Data Release 10 and 11 galaxy samples. 2014, MNRAS, 441, 24, arXiv:1312.4877

[24] Archidiacono, M., Gariazzo, S., Giunti, C., et al., Pseudoscalar—sterile neutrino interactions: reconciling the cosmos with neutrino oscillations. 2016, JCAP, 1608, 067, arXiv:1606.07673

[25] Archidiacono, M., Hooper, D. C., Murgia, R., et al., Constraining Dark Matter-Dark Radiation interactions with CMB, BAO, and Lyman-α. 2019, JCAP, 1910, 055, arXiv:1907.01496

[26] Arnaud, M., Pratt, G. W., Piffaretti, R., et al., The universal galaxy cluster pressure profile from a representative sample of nearby systems (REXCESS) and the Y_{SZ} - M_{500} relation. 2010, A&A, 517, A92, arXiv:0910.1234

[27] Ashenfelter, J. et al., First search for short-baseline neutrino oscillations at HFIR with PROSPECT. 2018, Phys. Rev. Lett., 121, 251802, arXiv:1806.02784

[28] Ata, M. et al., The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: first measurement of baryon acoustic oscillations between redshift 0.8 and 2.2. 2018, MNRAS, 473, 4773, arXiv:1705.06373

[29] Aubourg, É. et al., Cosmological implications of baryon acoustic oscillation measurements. 2015, Phys. Rev. D, 92, 123516, arXiv:1411.1074

[30] Audren, B., Lesgourgues, J., Benabed, K., & Prunet, S., Conservative Constraints on Early Cosmology: an illustration of the Monte Python cosmological parameter inference code. 2013, JCAP, 1302, 001, arXiv:1210.7183

[31] Aver, E., Olive, K. A., & Skillman, E. D., The effects of He I ?10830 on helium abundance determinations. 2015, J. Cosmology Astropart. Phys., 1507, 011, arXiv:1503.08146

[32] Aylor, K., Hou, Z., Knox, L., et al., A Comparison of Cosmological Parameters Determined from CMB Temperature Power Spectra from the South Pole Telescope and the Planck Satellite. 2017, ApJ, 850, 101, arXiv:1706.10286

[33] Baker, T., Bellini, E., Ferreira, P. G., et al., Strong constraints on cosmological gravity from GW170817 and GRB 170817A. 2017, Phys. Rev. Lett., 119, 251301, arXiv:1710.06394

[34] Banados, E. et al., An 800 million solar mass black hole in a significantly neutral universe at redshift 7.5. 2018, Nature, 553, 473, arXiv:1712.01860

[35] Bartelmann, M. & Schneider, P., Weak Gravitational Lensing. 2001, Phys. Rept., 340, 291, arXiv:astro-ph/9912508

[36] Baumann, D., Green, D., & Wallisch, B., Searching for light relics with large-scale structure. 2018, JCAP, 1808, 029, arXiv:1712.08067

[37] Bautista, J. E. et al., Measurement of baryon acoustic oscillation correlations at $z = 2.3$ with SDSS DR12 Lyα-Forests. 2017, A&A, 603, A12, arXiv:1702.00176

[38] Bautista, J. E. et al., The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations at redshift of 0.72 with the DR14 Luminous Red Galaxy Sample. 2018, ApJ, 863, 110, arXiv:1712.08064

[39] Belgacem, E., Dirian, Y., Foffa, S., & Maggiore, M., Nonlocal gravity. Conceptual aspects and cosmological predictions. 2018, JCAP, 1803, 002, arXiv:1712.07066

[40] Bellini, E. & Sawicki, I., Maximal freedom at minimum cost: linear large-scale structure in general modifications of gravity. 2014, J. Cosmology Astropart. Phys., 7, 050, arXiv:1404.3713

[41] Ben Achour, J., Crisostomi, M., Koyama, K., et al., Degenerate higher order scalar-tensor theories beyond Horndeski up to cubic order. 2016, JHEP, 12, 100, arXiv:1608.08135

[42] Bennett, C. et al., Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. 2013, ApJS, 208, 20, arXiv:1212.5225

[43] Bennett, C. L. et al., First Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results. 2003, ApJS, 148, 1, arXiv:astro-ph/0302207

[44] Bernal, J. L., Verde, L., & Riess, A. G., The trouble with H_0. 2016, JCAP, 1610, 019, arXiv:1607.05617

[45] Betoule, M., Kessler, R., Guy, J., et al., Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples. 2014, A&A, 568, A22, arXiv:1401.4064

[46] Betoule, M., Marriner, J., Regnault, N., et al., Improved photometric calibration of the SNLS and the SDSS supernova surveys. 2013, A&A, 552, A124, arXiv:1212.4864

[47] Beutler, F., Blake, C., Colless, M., et al., The 6dF Galaxy Survey: Baryon Acoustic Oscillations and the Local Hubble Constant. 2011, MNRAS, 416, 3017, arXiv:1106.3366

[48] Beutler, F., Blake, C., Colless, M., et al., The 6dF Galaxy Survey: $z \approx 0$ measurement of the growth rate and σ_8. 2012, MNRAS, 423, 3430, arXiv:1204.4725

[49] Beutler, F., Seo, H.-J., Saito, S., et al., The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: anisotropic galaxy clustering in Fourier space. 2017, MNRAS, 466, 2242, arXiv:1607.03150

[50] Beutler, F. et al., The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Fourier space. 2017, MNRAS, 464, 3409, arXiv:1607.03149

[51] BICEP2/Keck Array and Planck Collaborations, Joint Analysis of BICEP2/Keck Array and Planck Data. 2015, Phys. Rev. Lett., 114, 101301, arXiv:1502.00612

[52] Biffi, V., Borgani, S., Murante, G., et al., On the Nature of Hydrostatic Equilibrium in Galaxy Clusters. 2016, ApJ, 827, 112, arXiv:1606.02293

[53] Blake, C. et al., The WiggleZ Dark Energy Survey: Joint measurements of the expansion and growth history at $z < 1$. 2012, MNRAS, 425, 405, arXiv:1204.3674

[54] Blake, C. et al., Galaxy And Mass Assembly (GAMA): improved cosmic growth measurements using multiple tracers of large-scale structure. 2013, MNRAS, 436, 3089, arXiv:1309.5556

[55] Blanchard, A. & Schneider, J., Gravitational lensing effect on the fluctuations of the cosmic background radiation. 1987, A&A, 184, 1

[56] Blas, D., Lesgourgues, J., & Tram, T., The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes. 2011, J. Cosmology Astropart. Phys., 1107, 034, arXiv:1104.2933

[57] Blomqvist, M. et al., Baryon acoustic oscillations from the cross-correlation of Lyα absorption and quasars in eBOSS DR14. 2019, Astron. Astrophys., 629, A86, arXiv:1904.03430

[58] Böhringer, H., Chon, G., & Collins, C. A., The extended ROSAT-ESO Flux Limited X-ray Galaxy Cluster Survey (REFLEX II) IV. X-ray Luminosity Function and First Constraints on Cosmological Parameters. 2014, A&A, 570, A31, arXiv:1403.2927

[59] Böhringer, H., Chon, G., Retzlaff, J., et al., The extended Northern ROSAT Galaxy Cluster Survey (NORAS II) I. Survey Construction and First Results. 2017, AJ, 153, 220, arXiv:1704.06489

[60] Borsanyi, S. et al., Calculation of the axion mass based on high-temperature lattice quantum chromodynamics. 2016, Nature, 539, 69, arXiv:1606.07494

[61] Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al., Reionization after Planck: The Derived Growth of the Cosmic Ionizing Emissivity Now Matches the Growth of the Galaxy UV Luminosity Density. 2015, ApJ, 811, 140, arXiv:1503.08228

[62] Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J., & Mahesh, N., An Absorption Profile Centred at 78 Megahertz in the Sky-Averaged Spectrum. 2018, Nature, 555, 67

[63] Bridle, S., Elvin-Poole, J., Evans, J., et al., A combined view of sterile-neutrino constraints from CMB and neutrino oscillation measurements. 2017, Physics Letters B, 764, 322, arXiv:1607.00032

[64] Brinckmann, T. & Lesgourgues, J., MontePython 3: boosted MCMC sampler and other features. 2019, Phys. Dark Univ., 24, 100260, arXiv:1804.07261

[65] Buen-Abad, M. A., Schmaltz, M., Lesgourgues, J., & Brinckmann, T., Interacting Dark Sector and Precision Cosmology. 2018, JCAP, 1801, 008, arXiv:1708.09406

[66] Calabrese, E., Slosar, A., Melchiorri, A., Smoot, G. F., & Zahn, O., Cosmic Microwave Weak lensing data as a test for the dark universe. 2008, Phys. Rev. D, 77, 123531, arXiv:0803.2309

[67] Calore, F., Cholis, I., McCabe, C., & Weniger, C., A Tale of Tails: Dark Matter Interpretations of the Fermi GeV Excess in Light of Background Model Systematics. 2015, Phys. Rev. D, 91, 063003, arXiv:1411.4647

[68] Capozzi, F., Lisi, E., Marrone, A., & Palazzo, A., Current unknowns in the three neutrino framework. 2018, Prog. Part. Nucl. Phys., 102, 48, arXiv:1804.09678

[69] Cardona, W., Kunz, M., & Pettorino, V., Determining H_0 with Bayesian hyper-parameters. 2017, J. Cosmology Astropart. Phys., 3, 056, arXiv:1611.06088

[70] Carron, J., Lewis, A., & Challinor, A., Internal delensing of Planck CMB temperature and polarization. 2017, J. Cosmology Astropart. Phys., 1705, 035, arXiv:1701.01712

[71] Carter, P., Beutler, F., Percival, W. J., et al., Low Redshift Baryon Acoustic Oscillation Measurement from the Reconstructed 6-degree Field Galaxy Survey. 2018, MNRAS, 481, 2371, arXiv:1803.01746

[72] Casas, S., Kunz, M., Martinelli, M., & Pettorino, V., Linear and non-linear Modified Gravity forecasts with future surveys. 2017, Phys. Dark Univ., 18, 73, arXiv:1703.01271

[73] Charles, E. et al., Sensitivity Projections for Dark Matter Searches with the Fermi Large Area Telescope. 2016, Phys. Rept., 636, 1, arXiv:1605.02016

[74] Chen, H.-Y., Fishbach, M., & Holz, D. E., A two per cent Hubble constant measurement from standard sirens within five years. 2018, Nature, 562, 545, arXiv:1712.06531

[75] Cheung, C., Fitzpatrick, A. L., Kaplan, J., Senatore, L., & Creminelli, P., The effective field theory of inflation. 2008, Journal of High Energy Physics, 3, 014, arXiv:0709.0293

[76] Chluba, J. & Sunyaev, R. A., Induced two-photon decay of the 2s level and the rate of cosmological hydrogen recombination. 2006, A&A, 446, 39, arXiv:astro-ph/0508144

[77] Chluba, J. & Thomas, R. M., Towards a complete treatment of the cosmological recombination problem. 2011, MNRAS, 412, 748, arXiv:1010.3631

[78] Chu, X., Dasgupta, B., & Kopp, J., Sterile neutrinos with secret interactions—lasting friendship with cosmology. 2015, JCAP, 1510, 011, arXiv:1505.02795

[79] Cirelli, M., Kadastik, M., Raidal, M., & Strumia, A., Model-independent implications of the e+-, anti-proton cosmic ray spectra on properties of Dark Matter. 2009, Nucl. Phys., B813, 1, [Addendum: Nucl. Phys.B873,530(2013)], arXiv:0809.2409

[80] Cole, S., Percival, W. J., Peacock, J. A., et al., The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. 2005, MNRAS, 362, 505, arXiv:astro-ph/0501174

[81] Cooke, R., Pettini, M., Jorgenson, R. A., Murphy, M. T., & Steidel, C. C., Precision measures of the primordial abundance of deuterium. 2014, ApJ, 781, 31, arXiv:1308.3240

[82] Cooke, R. J., Pettini, M., & Steidel, C. C., One Percent Determination of the Primordial Deuterium Abundance. 2018, ApJ, 855, 102, arXiv:1710.11129

[83] Couchot, F., Henrot-Versillé, S., Perdereau, O., et al., Relieving tensions related to the lensing of the cosmic microwave background temperature power spectra. 2017, A&A, 597, A126, arXiv:1510.07600

[84] Creminelli, P., D'Amico, G., Norena, J., & Vernizzi, F., The Effective Theory of Quintessence: the w<-1 Side Unveiled. 2009, JCAP, 0902, 018, arXiv:0811.0827

[85] Creminelli, P. & Vernizzi, F., Dark Energy after GW170817 and GRB170817A. 2017, Phys. Rev. Lett., 119, 251302, arXiv:1710.05877

[86] Crisostomi, M. & Koyama, K., Self-accelerating universe in scalar-tensor theories after GW170817. 2018, Phys. Rev. D, 97, 084004, arXiv:1712.06556

[87] Crites, A. T. et al., Measurements of E-Mode Polarization and Temperature-E-Mode Correlation in the Cosmic Microwave Background from 100 Square Degrees of SPTpol Data. 2015, ApJ, 805, 36, arXiv:1411.1042

[88] Cuesta, A. J., Verde, L., Riess, A., & Jimenez, R., Calibrating the cosmic distance scale ladder: the role of the sound horizon scale and the local expansion rate as distance anchors. 2015, MNRAS, 448, 3463, arXiv:1411.1094

[89] Cuoco, A., Heisig, J., Korsmeier, M., & Krämer, M., Probing dark matter annihilation in the Galaxy with antiprotons and gamma rays. 2017, J. Cosmology Astropart. Phys., 1710, 053, arXiv:1704.08258

[90] Cuoco, A., Krämer, M., & Korsmeier, M., Novel Dark Matter Constraints from Antiprotons in Light of AMS-02. 2017, Phys. Rev. Lett., 118, 191102, arXiv:1610.03071

[91] Dalal, N., Dore, O., Huterer, D., & Shirokov, A., The imprints of primordial non-gaussianities on large- scale structure: scale dependent bias and abundance of virialized objects. 2008, Phys. Rev. D, 77, 123514, arXiv:0710.4560

[92] Das, S., Louis, T., Nolta, M. R., et al., The Atacama Cosmology Telescope: temperature and gravitational lensing power spectrum measurements from three seasons of data. 2014, J. Cosmology Astropart. Phys., 4, 14, arXiv:1301.1037

[93] Davies, F. B. et al., Quantitative Constraints on the Reionization History from the IGM Damping Wing Signature in Two Quasars at $z > 7$. 2018, ApJ, 864, 142, arXiv:1802.06066

[94] Daylan, T., Finkbeiner, D. P., Hooper, D., et al., The characterization of the gamma-ray signal from the central Milky Way: A case for annihilating dark matter. 2016, Phys. Dark Univ., 12, 1, arXiv:1402.6703

[95] de Haan, T., Benson, B. A., Bleem, L. E., et al., Cosmological Constraints from Galaxy Clusters in the 2500 Square-degree SPT-SZ Survey. 2016, ApJ, 832, 95, arXiv:1603.06522

[96] de Sainte Agathe, V. et al., Baryon acoustic oscillations at z = 2.34 from the correlations of Lyα absorption in eBOSS DR14. 2019, Astron. Astrophys., 629, A85, arXiv:1904.03400

[97] de Salas, P. F., Forero, D. V., Ternes, C. A., Tortola, M., & Valle, J. W. F., Status of neutrino oscillations 2018: 3σ hint for normal mass ordering and improved CP sensitivity. 2018, Phys. Lett., B782, 633, arXiv:1708.01186

[98] de Salas, P. F., Gariazzo, S., Mena, O., Ternes, C. A., & Tórtola, M., Neutrino Mass Ordering from Oscillations and Beyond: 2018 Status and Future Prospects. 2018, Front. Astron. Space Sci., 5, 36, arXiv:1806.11051

[99] de Salas, P. F., Lattanzi, M., Mangano, G., et al., Bounds on very low reheating scenarios after Planck. 2015, Phys. Rev., D92, 123534, arXiv:1511.00672

[100] de Salas, P. F. & Pastor, S., Relic neutrino decoupling with flavour oscillations revisited. 2016, JCAP, 1607, 051, arXiv:1606.06986

[101] Deffayet, C., Pujolas, O., Sawicki, I., & Vikman, A., Imperfect Dark Energy from Kinetic Gravity Braiding. 2010, JCAP, 1010, 026, arXiv:1008.0048

[102] Delubac, T. et al., Baryon acoustic oscillations in the Lyα forest of BOSS DR11 quasars. 2015, A&A, 574, A59, arXiv:1404.1801

[103] Dentler, M., Hernández-Cabezudo, Á., Kopp, J., et al., Updated Global Analysis of Neutrino Oscillations in the Presence of eV-Scale Sterile Neutrinos. 2018, JHEP, 08, 010, arXiv:1803.10661

[104] Dentler, M., Hernández-Cabezudo, Á., Kopp, J., Maltoni, M., & Schwetz, T., Sterile neutrinos or flux uncertainties? — Status of the reactor anti-neutrino anomaly. 2017, JHEP, 11, 099, arXiv:1709.04294

[105] DES Collaboration, Dark Energy Survey Year 1 Results: A Precise H_0 Measurement from DES Y1, BAO, and D/H Data. 2018, MNRAS, 480, 3879, arXiv:1711.00403

[106] DES Collaboration, Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing. 2018, Phys. Rev., D98, 043526, arXiv:1708.01530

[107] DES Collaboration, Dark Energy Survey Year 1 Results: Measurement of the Baryon Acoustic Oscillation scale in the distribution of galaxies to redshift 1. 2019, MNRAS, 483, 4866, arXiv:1712.06209

[108] Di Valentino, E., Bøehm, C., Hivon, E., & Bouchet, F. R., Reducing the H_0 and σ_8 tensions with Dark Matter-neutrino interactions. 2018, Phys. Rev., D97, 043513, arXiv:1710.02559

[109] Di Valentino, E., Gustavino, C., Lesgourgues, J., et al., Probing nuclear rates with Planck and BICEP2. 2014, Phys. Rev. D, 90, 023543, arXiv:1404.7848

[110] Dima, A. & Vernizzi, F., Vainshtein Screening in Scalar-Tensor Theories before and after GW170817: Constraints on Theories beyond Horndeski. 2018, Phys. Rev., D97, 101302, arXiv:1712.04731

[111] Dirian, Y., Changing the Bayesian prior: Absolute neutrino mass constraints in nonlocal gravity. 2017, Phys. Rev. D, 96, 083513, arXiv:1704.04075

[112] Dirian, Y., Foffa, S., Kunz, M., Maggiore, M., & Pettorino, V., Non-local gravity and comparison with observational datasets. II. Updated results and Bayesian model comparison with ΛCDM. 2016, JCAP, 1605, 068, arXiv:1602.03558

[113] Dodelson, S. & Widrow, L. M., Sterile-neutrinos as dark matter. 1994, Phys. Rev. Lett., 72, 17, arXiv:hep-ph/9303287

[114] du Mas des Bourboux, H. et al., Baryon acoustic oscillations from the complete SDSS-III Lyα-quasar cross-correlation function at $z = 2.4$. 2017, A&A, 608, A130, arXiv:1708.02225

[115] Efstathiou, G., H_0 revisited. 2014, MNRAS, 440, 1138, arXiv:1311.3461

[116] Efstathiou, G. & Bond, J. R., Cosmic confusion: degeneracies among cosmological parameters derived from measurements of microwave background anisotropies. 1999, MNRAS, 304, 75, arXiv:astro-ph/9807103

[117] Efstathiou, G. & Gratton, S., A Detailed Description of the CamSpec Likelihood Pipeline and a Reanalysis of the Planck High Frequency Maps. 2019, arXiv e-prints, arXiv:1910.00483, arXiv:1910.00483

[118] Einstein, A., Cosmological Considerations in the General Theory of Relativity. 1917, Sitzungsber.Preuss.Akad.Wiss.Berlin (Math.Phys.), 1917, 142

[119] Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al., Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. 2005, ApJ, 633, 560, arXiv:astro-ph/0501171

[120] Elvin-Poole, J. et al., Dark Energy Survey year 1 results: Galaxy clustering for combined probes. 2018, Phys. Rev., D98, 042006, arXiv:1708.01536

[121] Erben, T. et al., CFHTLenS: The Canada-France-Hawaii Telescope Lensing Survey - Imaging Data and Catalogue Products. 2013, MNRAS, 433, 2545, arXiv:1210.8156

[122] Ewall-Wice, A., Chang, T. C., Lazio, J., et al., Modeling the Radio Background from the First Black Holes at Cosmic Dawn: Implications for the 21 cm Absorption Amplitude. 2018, Astrophys. J., 868, 63, arXiv:1803.01815

[123] Ezquiaga, J. M. & Zumalacarregui, M., Dark Energy After GW170817: Dead Ends and the Road Ahead. 2017, Phys. Rev. Lett., 119, 251304, arXiv:1710.05901

[124] Fang, W., Hu, W., & Lewis, A., Crossing the Phantom Divide with Parameterized Post-Friedmann Dark Energy. 2008, Phys. Rev. D, 78, 087303, arXiv:0808.3125

[125] Farhang, M., Bond, J. R., & Chluba, J., Semi-blind Eigen Analyses of Recombination Histories Using Cosmic Microwave Background Data. 2012, ApJ, 752, 88, arXiv:1110.4608

[126] Farhang, M., Bond, J. R., Chluba, J., & Switzer, E. R., Constraints on Perturbations to the Recombination History from Measurements of the Cosmic Microwave Background Damping Tail. 2013, ApJ, 764, 137, arXiv:1211.4634

[127] Feeney, S. M., Mortlock, D. J., & Dalmasso, N., Clarifying the Hubble constant tension with a Bayesian hierarchical model of the local distance ladder. 2018, MNRAS, 476, 3861, arXiv:1707.00007

[128] Feeney, S. M., Peiris, H. V., Williamson, A. R., et al., Prospects for resolving the Hubble constant tension with standard sirens. 2019, Phys. Rev. Lett., 122, 061105, arXiv:1802.03404

[129] Ferrara, S., Kallosh, R., Linde, A., & Porrati, M., Minimal supergravity models of inflation. 2013, Phys. Rev. D, 88, 085038, arXiv:1307.7696

[130] Ferraro, S. & Smith, K. M., Characterizing the epoch of reionization with the small-scale CMB: Constraints on the optical depth and duration. 2018, Phys. Rev., D98, 123519, arXiv:1803.07036

[131] Fields, B. D., The Primordial Lithium Problem. 2011, Annual Review of Nuclear and Particle Science, 61, 47, arXiv:1203.3551

[132] Fields, B. D., Molaro, P., & Sarkar, S., Big-Bang Nucleosynthesis. 2014, Chin. Phys., C38, 339, arXiv:1412.1408

[133] Finkbeiner, D. P., Galli, S., Lin, T., & Slatyer, T. R., Searching for Dark Matter in the CMB: A Compact Parameterization of Energy Injection from New Physics. 2012, Phys. Rev. D, 85, 043522, arXiv:1109.6322

[134] Fixsen, D., The Temperature of the Cosmic Microwave Background. 2009, ApJ, 707, 916, arXiv:0911.1955

[135] Follin, B. & Knox, L., Insensitivity of the distance ladder Hubble constant determination to Cepheid calibration modelling choices. 2018, MNRAS, 477, 4534, arXiv:1707.01175

[136] Font-Ribera, A., Kirkby, D., Busca, N., et al., Quasar-Lyman α forest cross-correlation from BOSS DR11: Baryon Acoustic Oscillations. 2014, J. Cosmology Astropart. Phys., 5, 27, arXiv:1311.1767

[137] Freedman, W. L., Madore, B. F., Gibson, B. K., et al., Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant. 2001, ApJ, 553, 47, arXiv:astro-ph/0012376

[138] Freedman, W. L., Madore, B. F., Hatt, D., et al., The Carnegie-Chicago Hubble Program. VIII. An Independent Determination of the Hubble Constant Based on the Tip of the Red Giant Branch. 2019, ApJ, 882, 34, arXiv:1907.05922

[139] Freedman, W. L., Madore, B. F., Hoyt, T., et al., Calibration of the Tip of the Red Giant Branch. 2020, ApJ, 891, 57, arXiv:2002.01550

[140] Freedman, W. L., Madore, B. F., Scowcroft, V., et al., Carnegie Hubble Program: A Mid-infrared Calibration of the Hubble Constant. 2012, ApJ, 758, 24, arXiv:1208.3281

[141] Galante, M., Kallosh, R., Linde, A., & Roest, D., Unity of Cosmological Inflation Attractors. 2015, Phys. Rev. Lett., 114, 141302, arXiv:1412.3797

[142] Galli, S., Benabed, K., Bouchet, F., et al., CMB Polarization can constrain cosmology better than CMB temperature. 2014, Phys. Rev., D90, 063504, arXiv:1403.5271

[143] Galli, S., Slatyer, T. R., Valdes, M., & Iocco, F., Systematic Uncertainties In Constraining Dark Matter Annihilation From The Cosmic Microwave Background. 2013, Phys. Rev. D, 88, 063502, arXiv:1306.0563

[144] Gao, F., Braatz, J. A., Reid, M. J., et al., The Megamaser Cosmology Project. VIII. A Geometric Distance to NGC 5765b. 2016, ApJ, 817, 128, arXiv:1511.08311

[145] Gariazzo, S., Giunti, C., Laveder, M., & Li, Y. F., Model-independent $\bar{\nu}_e$ short-baseline oscillations from reactor spectral ratios. 2018, Phys. Lett., B782, 13, arXiv:1801.06467

[146] George, E. M. et al., A measurement of secondary cosmic microwave background anisotropies from the 2500-square-degree SPT-SZ survey. 2015, ApJ, 799, 177, arXiv:1408.3161

[147] Gil-Marín, H. et al., The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: structure growth rate measurement from the anisotropic quasar power spectrum in the redshift range $0.8 < z < 2.2$. 2018, MNRAS, 477, 1604, arXiv:1801.02689

[148] Gleyzes, J., Langlois, D., Piazza, F., & Vernizzi, F., Essential building blocks of dark energy. 2013, J. Cosmology Astropart. Phys., 8, 025, arXiv:1304.4840

[149] Gleyzes, J., Langlois, D., Piazza, F., & Vernizzi, F., Healthy theories beyond Horndeski. 2015, Phys. Rev. Lett., 114, 211101, arXiv:1404.6495

[150] Gnedin, N. Y. & Gnedin, O. Y., Cosmological Neutrino Background Revisited. 1998, ApJ, 509, 11, arXiv:astro-ph/9712199

[151] Gomez-Valent, A. & Amendola, L., H_0 from cosmic chronometers and Type Ia supernovae, with Gaussian Processes and the novel Weighted Polynomial Regression method. 2018, J. Cosmology Astropart. Phys., 2018, 051, arXiv:1802.01505

[152] Gordon, C. & Macias, O., Dark Matter and Pulsar Model Constraints from Galactic Center Fermi-LAT Gamma Ray Observations. 2013, Phys. Rev. D, 88, 083521, [Erratum: PRD89,no.4,049901(2014)], arXiv:1306.5725

[153] Grieb, J. N., Sánchez, A. G., Salazar-Albornoz, S., et al., The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological implications of the Fourier space wedges of the final sample. 2017, MNRAS, 467, 2085, arXiv:1607.03143

[154] Gruzinov, A. & Hu, W., Secondary CMB anisotropies in a universe reionized in patches. 1998, Astrophys. J., 508, 435, arXiv:astro-ph/9803188

[155] Gubitosi, G., Piazza, F., & Vernizzi, F., The effective field theory of dark energy. 2013, J. Cosmology Astropart. Phys., 2, 032, arXiv:1210.0201

[156] Gustavino, C. 2017, in RICAP16, 6th Roma International Conference on Astroparticle Physics, Roma, Italy, Edited by Morselli, A.; Capone, A.; Rodriguez Fernandez, G.; EPJ Web of Conferences, Volume 136, id.01009, Vol. 136

[157] Guth, A. H., Kaiser, D. I., & Nomura, Y., Inflationary paradigm after Planck 2013. 2014, Physics Letters B, 733, 112, arXiv:1312.7619

[158] Hamana, T. et al., Cosmological constraints from cosmic shear two-point correlation functions with HSC survey first-year data. 2020, Publ. Astron. Soc. Jap., 72, Publications of the Astronomical Society of Japan, Volume 72, Issue 1, February 2020, 16, https://doi.org/10.1093/pasj/psz138, arXiv:1906.06041

[159] Handley, W. & Millea, M., Maximum-Entropy Priors with Derived Parameters in a Specified Distribution. 2019, Entropy, 21, 272, arXiv:1804.08143

[160] Hannestad, S., Sloth Hansen, R., & Tram, T., Can active-sterile neutrino oscillations lead to chaotic behavior of the cosmological lepton asymmetry? 2013, J. Cosmology Astropart. Phys., 4, 032, arXiv:1302.7279

[161] Hanson, D., Challinor, A., & Lewis, A., Weak lensing of the CMB. 2010, General Relativity and Gravitation, 42, 2197, arXiv:0911.0612

[162] Hassan, S. F., Rosen, R. A., & Schmidt-May, A., Ghost-free Massive Gravity with a General Reference Metric. 2012, JHEP, 02, 026, arXiv:1109.3230

[163] Hazra, D. K. & Smoot, G. F., Witnessing the reionization history using Cosmic Microwave Background observation from Planck. 2017, JCAP, 1711, 028, arXiv:1708.04913

[164] Heavens, A., Jimenez, R., & Verde, L., Standard Rulers, Candles, and Clocks from the Low-Redshift Universe. 2014, Phys. Rev. Lett., 113, 241302, arXiv:1409.6217

[165] Heinrich, C. & Hu, W., Does *Planck* 2015 polarization data favor high redshift reionization? 2018, Phys. Rev., D98, 063514, arXiv:1802.00791

[166] Heinrich, C. & Hu, W., RELIKE: Reionization Effective Likelihood from Planck 2018 Data. 2021, arXiv:2104.13998 [astro-ph]

[167] Heinrich, C. H., Miranda, V., & Hu, W., Complete Reionization Constraints from Planck 2015 Polarization. 2017, Phys. Rev., D95, 023513, arXiv:1609.04788

[168] Hellings, R. W. & Nordtvedt, K., Vector-Metric Theory of Gravity. 1973, Phys. Rev., D7, 3593

[169] Henning, J. W. et al., Measurements of the Temperature and E-mode Polarization of the CMB from 500 Square Degrees of SPTpol Data. 2018, ApJ, 852, 97, arXiv:1707.09353

[170] Heymans, C., Van Waerbeke, L., Miller, L., et al., CFHTLenS: the Canada-France-Hawaii Telescope Lensing Survey. 2012, MNRAS, 427, 146, arXiv:1210.0032

[171] Heymans, C. et al., CFHTLenS: The Canada-France-Hawaii Telescope Lensing Survey. 2012, MNRAS, 427, 146, arXiv:1210.0032

[172] Hikage, C. et al., Cosmology from cosmic shear power spectra with Subaru Hyper Suprime-Cam first-year data. 2019, Publ. Astron. Soc. Jap., 71, arXiv:1809.09148

[173] Hildebrandt, H., Choi, A., Heymans, C., et al., RCSLenS: The Red Cluster Sequence Lensing Survey. 2016, MNRAS, 463, 635, arXiv:1603.07722

[174] Hildebrandt, H. et al., KiDS-450: Cosmological parameter constraints from tomographic weak gravitational lensing. 2017, MNRAS, 465, 1454, arXiv:1606.05338

[175] Hildebrandt, H. et al., KiDS+VIKING-450: Cosmic shear tomography with optical+infrared data. 2020, Astron. Astrophys., 633, A69, arXiv:1812.06076

[176] Hirata, C. M. & Seljak, U., Intrinsic alignment-lensing interference as a contaminant of cosmic shear. 2004, Phys. Rev. D, 70, 063526, [Erratum: PRD82,049901(2010)], arXiv:astro-ph/0406275

[177] Hivon, E., Mottet, S., & Ponthieu, N., QuickPol: Fast calculation of effective beam matrices for CMB polarization. 2017, A&A, 598, A25, arXiv:1608.08833

[178] Hoekstra, H., Herbonnet, R., Muzzin, A., et al., The Canadian Cluster Comparison Project: detailed study of systematics and updated weak lensing masses. 2015, MNRAS, 449, 685, arXiv:1502.01883

[179] Hojjati, A., Pogosian, L., & Zhao, G.-B., Testing gravity with CAMB and CosmoMC. 2011, J. Cosmology Astropart. Phys., 1108, 005, arXiv:1106.4543

[180] Holz, D. E. & Hughes, S. A., Using gravitational-wave standard sirens. 2005, ApJ, 629, 15, arXiv:astro-ph/0504616

[181] Hou, Z., Aylor, K., Benson, B. A., et al., A Comparison of Maps and Power Spectra Determined from South Pole Telescope and Planck Data. 2018, ApJ, 853, 3, arXiv:1704.00884

[182] Howlett, C., Ross, A., Samushia, L., Percival, W., & Manera, M., The clustering of the SDSS main galaxy sample – II. Mock galaxy catalogues and a measurement of the growth of structure from redshift space distortions at $z = 0.15$. 2015, MNRAS, 449, 848, arXiv:1409.3238

[183] Hu, B., Raveri, M., Frusciante, N., & Silvestri, A., Effective Field Theory of Cosmic Acceleration: an implementation in CAMB. 2014, Phys. Rev., D89, 103530, arXiv:1312.5742

[184] Hu, W. & Holder, G. P., Model - independent reionization observables in the CMB. 2003, Phys. Rev., D68, 023001, arXiv:astro-ph/0303400

[185] Hu, W., Scott, D., Sugiyama, N., & White, M., Effect of physical assumptions on the calculation of microwave background anisotropies. 1995, Phys. Rev. D, 52, 5498, arXiv:astro-ph/9505043

[186] Huang, Y., Addison, G. E., Weiland, J. L., & Bennett, C. L., Assessing Consistency Between WMAP 9-year and Planck 2015 Temperature Power Spectra. 2018, ApJ, 869, 38, arXiv:1804.05428

[187] Huterer, D., Shafer, D., Scolnic, D., & Schmidt, F., Testing ΛCDM at the lowest redshifts with SN Ia and galaxy velocities. 2017, J. Cosmology Astropart. Phys., 1705, 015, arXiv:1611.09862

[188] Ijjas, A. & Steinhardt, P. J., Implications of Planck2015 for inflationary, ekpyrotic and anamorphic bouncing cosmologies. 2016, Classical and Quantum Gravity, 33, 044001, arXiv:1512.09010

[189] Ijjas, A., Steinhardt, P. J., & Loeb, A., Inflationary paradigm in trouble after Planck2013. 2013, Physics Letters B, 723, 261, arXiv:1304.2785

[190] Iliadis, C., Anderson, K., Coc, A., Timmes, F., & Starrfield, S., Bayesian Estimation of Thermonuclear Reaction Rates. 2016, ApJ, 831, 107, arXiv:1608.05853

[191] Izotov, Y. I., Thuan, T. X., & Guseva, N. G., A new determination of the primordial He abundance using the He?i λ10830 **r**A emission line: cosmological implications. 2014, MNRAS, 445, 778, arXiv:1408.6953

[192] Jee, M. J., Tyson, J. A., Hilbert, S., et al., Cosmic Shear Results from the Deep Lens Survey - II: Full Cosmological Parameter Constraints from Tomography. 2016, ApJ, 824, 77, arXiv:1510.03962

[193] Jimenez, R. & Loeb, A., Constraining Cosmological Parameters Based on Relative Galaxy Ages. 2002, ApJ, 573, 37, arXiv:astro-ph/0106145

[194] Joudaki, S., Blake, C., Johnson, A., et al., KiDS-450 + 2dFLenS: Cosmological parameter constraints from weak gravitational lensing tomography and overlapping redshift-space galaxy clustering. 2018, MNRAS, 474, 4894, arXiv:1707.06627

[195] Joudaki, S. et al., KiDS+VIKING-450 and DES-Y1 combined: Cosmology with cosmic shear. 2019, arXiv e-prints, arXiv:1906.09262

[196] Kallosh, R., Linde, A., & Roest, D., Superconformal inflationary α-attractors. 2013, Journal of High Energy Physics, 11, 198, arXiv:1311.0472

[197] Kamionkowski, M., Kosowsky, A., & Stebbins, A., Statistics of Cosmic Microwave Background Polarization. 1997, Phys. Rev. D, 55, 7368, arXiv:astro-ph/9611125

[198] Kaplinghat, M. et al., Probing the Reionization History of the Universe using the Cosmic Microwave Background Polarization. 2003, ApJ, 583, 24, arXiv:astro-ph/0207591

[199] Kazin, E. A., Koda, J., Blake, C., et al., The WiggleZ Dark Energy Survey: improved distance measurements to $z = 1$ with reconstruction of the baryonic acoustic feature. 2014, MNRAS, 441, 3524, arXiv:1401.0358

[200] Köhlinger, F. et al., KiDS-450: The tomographic weak lensing power spectrum and constraints on cosmological parameters. 2017, MNRAS, 471, 4412, arXiv:1706.02892

[201] Kholupenko, E. E., Ivanchik, A. V., & Varshalovich, D. A., Rapid HeII-HeI recombination and radiation arising from this process. 2007, MNRAS, 378, L39, arXiv:astro-ph/0703438

[202] Knee, A. M., Contreras, D., & Scott, D., Cosmological constraints on sterile neutrino oscillations from Planck. 2018, J. Cosmology Astropart. Phys., in press, arXiv:1812.02102

[203] Knox, L., Scoccimarro, R., & Dodelson, S., The Impact of inhomogeneous reionization on cosmic microwave background anisotropy. 1998, Phys. Rev. Lett., 81, 2004, arXiv:astro-ph/9805012

[204] Komatsu, E. & Seljak, U., The Sunyaev-Zel'dovich angular power spectrum as a probe of cosmological parameters. 2002, MNRAS, 336, 1256, arXiv:astro-ph/0205468

[205] Krause, E. et al., Dark Energy Survey Year 1 Results: Multi-Probe Methodology and Simulated Likelihood Analyses. 2017, Submitted to: Phys. Rev. D, arXiv:1706.09359

[206] Kreisch, C. D., Cyr-Racine, F.-Y., & Doré, O., The Neutrino Puzzle: Anomalies, Interactions, and Cosmological Tensions. 2019, arXiv e-prints, arXiv:1902.00534

[207] Kunz, M. & Sapone, D., Crossing the Phantom Divide. 2006, Phys. Rev., D74, 123503, arXiv:astro-ph/0609040

[208] Kuo, C. Y., Braatz, J. A., Reid, M. J., et al., The Megamaser Cosmology Project. V. An Angular-diameter Distance to NGC 6264 at 140 Mpc. 2013, ApJ, 767, 155, arXiv:1207.7273

[209] Lancaster, L., Cyr-Racine, F.-Y., Knox, L., & Pan, Z., A tale of two modes: Neutrino free-streaming in the early universe. 2017, JCAP, 1707, 033, arXiv:1704.06657

[210] Langlois, D., Mancarella, M., Noui, K., & Vernizzi, F., Effective Description of Higher-Order Scalar-Tensor Theories. 2017, JCAP, 1705, 033, arXiv:1703.03797

[211] Langlois, D., Saito, R., Yamauchi, D., & Noui, K., Scalar-tensor theories and modified gravity in the wake of GW170817. 2018, Phys. Rev., D97, 061501, arXiv:1711.07403

[212] Larsen, P., Challinor, A., Sherwin, B. D., & Mak, D., Demonstration of cosmic microwave background delensing using the cosmic infrared background. 2016, Phys. Rev. Lett., 117, 151102, arXiv:1607.05733

[213] Lemos, P., Lee, E., Efstathiou, G., & Gratton, S., Model independent $H(z)$ reconstruction using the cosmic inverse distance ladder. 2019, MNRAS, 483, 4803, arXiv:1806.06781

[214] Lesgourgues, J., Mangano, G., Miele, G., & Pastor, S. 2013, Neutrino Cosmology (Cambridge)

[215] Lesgourgues, J., Marques-Tavares, G., & Schmaltz, M., Evidence for dark matter interactions in cosmological precision data? 2016, JCAP, 1602, 037, arXiv:1507.04351

[216] Lesgourgues, J. & Pastor, S., Massive neutrinos and cosmology. 2006, Phys. Rept., 429, 307, arXiv:astro-ph/0603494

[217] Lewis, A., Cosmological parameters from WMAP 5-year temperature maps. 2008, Phys. Rev. D, 78, 023002, arXiv:0804.3865

[218] Lewis, A., Efficient sampling of fast and slow cosmological parameters. 2013, Phys. Rev. D, 87, 103529, arXiv:1304.4473

[219] Lewis, A. & Bridle, S., Cosmological parameters from CMB and other data: A Monte Carlo approach. 2002, Phys. Rev. D, 66, 103511, arXiv:astro-ph/0205436

[220] Lewis, A. & Challinor, A., Weak gravitational lensing of the CMB. 2006, Phys. Rept., 429, 1, arXiv:astro-ph/0601594

[221] Lewis, A., Challinor, A., & Lasenby, A., Efficient computation of CMB anisotropies in closed FRW models. 2000, ApJ, 538, 473, arXiv:astro-ph/9911177

[222] Lin, M.-X., Benevento, G., Hu, W., & Raveri, M., Acoustic Dark Energy: Potential Conversion of the Hubble Tension. 2019, Phys. Rev., D100, 063542, arXiv:1905.12618

[223] Linde, A., On the problem of initial conditions for inflation. 2018, Found. Phys., 48, 1246, arXiv:1710.04278

[224] Lombriser, L. & Taylor, A., Breaking a Dark Degeneracy with Gravitational Waves. 2016, JCAP, 1603, 031, arXiv:1509.08458

[225] Louis, T., Addison, G. E., Hasselfield, M., et al., The Atacama Cosmology Telescope: cross correlation with Planck maps. 2014, J. Cosmology Astropart. Phys., 7, 016, arXiv:1403.0608

[226] Louis, T. et al., The Atacama Cosmology Telescope: Two-Season ACTPol Spectra and Parameters. 2017, J. Cosmology Astropart. Phys., 1706, 031, arXiv:1610.02360

[227] LoVerde, M. & Afshordi, N., Extended Limber Approximation. 2008, Phys. Rev. D, 78, 123506, arXiv:0809.5112

[228] Mak, D. S. Y., Challinor, A., Efstathiou, G., & Lagache, G., Measurement of CIB power spectra over large sky areas from Planck HFI maps. 2017, MNRAS, 466, 286, arXiv:1609.08942

[229] Maltoni, M., (eV) sterile neutrinos: the global picture. 2018, Neutrino 2018 Conference, Heidelberg

[230] Mangano, G., Miele, G., Pastor, S., et al., Relic neutrino decoupling including flavour oscillations. 2005, Nuclear Physics B, 729, 221, arXiv:hep-ph/0506164

[231] Mantz, A. B., von der Linden, A., Allen, S. W., et al., Weighing the giants - IV. Cosmology and neutrino mass. 2015, MNRAS, 446, 2205, arXiv:1407.4516

[232] Marcucci, L. E., Mangano, G., Kievsky, A., & Viviani, M., Implication of the proton-deuteron radiative capture for Big Bang Nucleosynthesis. 2016, Phys. Rev. Lett., 116, 102501, [Erratum: PRL117,no.4,049901(2016)], arXiv:1510.07877

[233] Marcucci, L. E., Viviani, M., Schiavilla, R., Kievsky, A., & Rosati, S., Electromagnetic structure of A=2 and 3 nuclei and the nuclear current operator. 2005, Phys. Rev. C, 72, 014001, arXiv:nucl-th/0502048

[234] Mason, C. A., Treu, T., Dijkstra, M., et al., The Universe Is Reionizing at $z \sim 7$: Bayesian Inference of the IGM Neutral Fraction Using Lyα Emission from Galaxies. 2018, ApJ, 856, 2, arXiv:1709.05356

[235] McManus, R., Lombriser, L., & Peñarrubia, J., Finding Horndeski theories with Einstein gravity limits. 2016, JCAP, 1611, 006, arXiv:1606.03282

[236] Mead, A., Heymans, C., Lombriser, L., et al., Accurate halo-model matter power spectra with dark energy, massive neutrinos and modified gravitational forces. 2016, MNRAS, 459, 1468, arXiv:1602.02154

[237] Mead, A., Peacock, J., Heymans, C., Joudaki, S., & Heavens, A., An accurate halo model for fitting non-linear cosmological power spectra and baryonic feedback models. 2015, MNRAS, 454, 1958, arXiv:1505.07833

[238] Millea, M. & Bouchet, F., Cosmic Microwave Background Constraints in Light of Priors Over Reionization Histories. 2018, Astron. Astrophys., 617, A96, arXiv:1804.08476

[239] Miranda, V., Lidz, A., Heinrich, C. H., & Hu, W., Signatures of Metal-Free Star Formation in Planck 2015 Polarization Data. 2017, MNRAS, 467, 4050, arXiv:1610.00691

[240] Mortonson, M. J. & Hu, W., Model-Independent Constraints on Reionization from Large-Scale CMB Polarization. 2008, ApJ, 672, 737, arXiv:0705.1132

[241] Mosher, J., Guy, J., Kessler, R., et al., Cosmological Parameter Uncertainties from SALT-II Type Ia Supernova Light Curve Models. 2014, ApJ, 793, 16, arXiv:1401.4065

[242] Motloch, P. & Hu, W., Tensions between direct measurements of the lens power spectrum from Planck data. 2018, Phys. Rev., D97, 103536, arXiv:1803.11526

[243] Muñoz, J. B., Grin, D., Dai, L., Kamionkowski, M., & Kovetz, E. D., Search for Compensated Isocurvature Perturbations with Planck Power Spectra. 2016, Phys. Rev. D, 93, 043008, arXiv:1511.04441

[244] Naess, S. et al., The Atacama Cosmology Telescope: CMB Polarization at $200 < \ell < 9000$. 2014, J. Cosmology Astropart. Phys., 1410, 007, arXiv:1405.5524

[245] Nagai, D., Kravtsov, A. V., & Vikhlinin, A., Effects of Galaxy Formation on Thermodynamics of the Intracluster Medium. 2007, ApJ, 668, 1, arXiv:astro-ph/0703661

[246] Nollett, K. M. & Burles, S., Estimating reaction rates and uncertainties for primordial nucleosynthesis. 2000, Phys. Rev. D, 61, 123505, arXiv:astro-ph/0001440

[247] Nollett, K. M. & Holder, G. P., An analysis of constraints on relativistic species from primordial nucleosynthesis and the cosmic microwave background. 2011, ArXiv e-prints, arXiv:1112.2683

[248] Obied, G., Dvorkin, C., Heinrich, C., Hu, W., & Miranda, V., Inflationary versus reionization features from *Planck* 2015 data. 2018, Phys. Rev., D98, 043518, arXiv:1803.01858

[249] Oka, A., Saito, S., Nishimichi, T., Taruya, A., & Yamamoto, K., Simultaneous constraints on the growth of structure and cosmic expansion from the multipole power spectra of the SDSS DR7 LRG sample. 2014, MNRAS, 439, 2515, arXiv:1310.2820

[250] Okumura, T. et al., The Subaru FMOS galaxy redshift survey (FastSound). IV. New constraint on gravity theory from redshift space distortions at $z \sim 1.4$. 2016, PASJ, 68, 24, arXiv:1511.08083

[251] Oldengott, I. M., Tram, T., Rampf, C., & Wong, Y. Y. Y., Interacting neutrinos in cosmology: exact description and constraints. 2017, JCAP, 1711, 027, arXiv:1706.02123

[252] Palanque-Delabrouille, N. et al., Neutrino masses and cosmology with Lyman-alpha forest power spectrum. 2015, J. Cosmology Astropart. Phys., 1511, 011, arXiv:1506.05976

[253] Parkinson, D., Riemer-Sorensen, S., Blake, C., et al., The WiggleZ Dark Energy Survey: Final data release and cosmological results. 2012, Phys. Rev. D, 86, 103518, arXiv:1210.2130

[254] Parsa, S., Dunlop, J. S., & McLure, R. J., No evidence for a significant AGN contribution to cosmic hydrogen reionization. 2018, MNRAS, 474, 2904, arXiv:1704.07750

[255] Patrignani, C. et al., Review of Particle Physics. 2016, Chin. Phys., C40, 100001

[256] Peebles, P. J. E., Recombination of the Primeval Plasma. 1968, ApJ, 153, 1

[257] Peebles, P. J. E. & Yu, J. T., Primeval Adiabatic Perturbation in an Expanding Universe. 1970, ApJ, 162, 815

[258] Peel, A., Pettorino, V., Giocoli, C., Starck, J.-L., & Baldi, M., Breaking degeneracies in modified gravity with higher (than 2nd) order weak-lensing statistics. 2018, A&A, 619, A38, arXiv:1805.05146

[259] Peimbert, A., Peimbert, M., & Luridiana, V., The primordial helium abundance and the number of neutrino families. 2016, Rev. Mex. Astron. Astrofis., 52, 419, arXiv:1608.02062

[260] Peiris, H. V., Komatsu, E., Verde, L., et al., First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications For Inflation. 2003, ApJS, 148, 213, arXiv:astro-ph/0302225

[261] Peloton, J., Schmittfull, M., Lewis, A., Carron, J., & Zahn, O., Full covariance of CMB and lensing reconstruction power spectra. 2017, Phys. Rev. D, 95, 043508, arXiv:1611.01446

[262] Penna-Lima, M., Bartlett, J. G., Rozo, E., et al., Calibrating the Planck Cluster Mass Scale with CLASH. 2017, A&A, 604, A89, arXiv:1608.05356

[263] Percival, W. J. & White, M., Testing cosmological structure formation using redshift-space distortions. 2009, MNRAS, 393, 297, arXiv:0808.0003

[264] Percival, W. J. et al., The 2dF Galaxy Redshift Survey: The power spectrum and the matter content of the universe. 2001, MNRAS, 327, 1297, arXiv:astro-ph/0105252

[265] Percival, W. J. et al., Parameter constraints for flat cosmologies from CMB and 2dFGRS power spectra. 2002, MNRAS, 337, 1068, arXiv:astro-ph/0206256

[266] Perlmutter, S., Aldering, G., Goldhaber, G., et al., Measurements of Ω and Λ from 42 High-Redshift Supernovae. 1999, ApJ, 517, 565, arXiv:astro-ph/9812133

[267] Pettorino, V., Testing modified gravity with Planck: The case of coupled dark energy. 2013, Phys. Rev. D, 88, 063519, arXiv:1305.7457

[268] Pettorino, V. & Baccigalupi, C., Coupled and Extended Quintessence: theoretical differences and structure formation. 2008, Phys. Rev., D77, 103003, arXiv:0802.1086

[269] Pezzotta, A. et al., The VIMOS Public Extragalactic Redshift Survey (VIPERS): The growth of structure at $0.5 < z < 1.2$ from redshift-space distortions in the clustering of the PDR-2 final sample. 2017, A&A, 604, A33, arXiv:1612.05645

[270] Pierpaoli, E., Scott, D., & White, M., Power-spectrum normalization from the local abundance of rich clusters of galaxies. 2001, MNRAS, 325, 77, arXiv:astro-ph/0010039

[271] Pisanti, O., Cirillo, A., Esposito, S., et al., PArthENoPE: Public Algorithm Evaluating the Nucleosynthesis of Primordial Elements. 2008, Comput. Phys. Commun., 178, 956, arXiv:0705.0290

[272] Pitrou, C., Coc, A., Uzan, J.-P., & Vangioni, E., Precision big bang nucleosynthesis with improved Helium-4 predictions. 2018, Phys. Rept., 04, 005, arXiv:1801.08023

[273] Planck Collaboration. 2018, Planck Legacy Archive, `https://pla.esac.esa.int`

[274] Planck Collaboration ES. 2018, The Legacy Explanatory Supplement, `https://www.cosmos.esa.int/web/planck/pla` (ESA)

[275] Planck Collaboration XV, *Planck* 2013 results. XV. CMB power spectra and likelihood. 2014, A&A, 571, A15, arXiv:1303.5075

[276] Planck Collaboration XVI, *Planck* 2013 results. XVI. Cosmological parameters. 2014, A&A, 571, A16, arXiv:1303.5076

[277] Planck Collaboration XVII, *Planck* 2013 results. XVII. Gravitational lensing by large-scale structure. 2014, A&A, 571, A17, arXiv:1303.5077

[278] Planck Collaboration XX, *Planck* 2013 results. XX. Cosmology from Sunyaev-Zeldovich cluster counts. 2014, A&A, 571, A20, arXiv:1303.5080

[279] Planck Collaboration XXII, *Planck* 2013 results. XXII. Constraints on inflation. 2014, A&A, 571, A22, arXiv:1303.5082

[280] Planck Collaboration XXIV, *Planck* 2013 results. XXIV. Constraints on primordial non-Gaussianity. 2014, A&A, 571, A24, arXiv:1303.5084

[281] Planck Collaboration XXX, *Planck* 2013 results. XXX. Cosmic infrared background measurements and implications for star formation. 2014, A&A, 571, A30, arXiv:1309.0382

[282] Planck Collaboration I, *Planck* 2015 results. I. Overview of products and results. 2016, A&A, 594, A1, arXiv:1502.01582

[283] Planck Collaboration XI, *Planck* 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters. 2016, A&A, 594, A11, arXiv:1507.02704

[284] Planck Collaboration XIII, *Planck* 2015 results. XIII. Cosmological parameters. 2016, A&A, 594, A13, arXiv:1502.01589

[285] Planck Collaboration XIV, *Planck* 2015 results. XIV. Dark energy and modified gravity. 2016, A&A, 594, A14, arXiv:1502.01590

[286] Planck Collaboration XV, *Planck* 2015 results. XV. Gravitational lensing. 2016, A&A, 594, A15, arXiv:1502.01591

[287] Planck Collaboration XVII, *Planck* 2015 results. XVII. Constraints on primordial non-Gaussianity. 2016, A&A, 594, A17, arXiv:1502.01592

[288] Planck Collaboration XX, *Planck* 2015 results. XX. Constraints on inflation. 2016, A&A, 594, A20, arXiv:1502.02114

[289] Planck Collaboration XXIV, *Planck* 2015 results. XXIV. Cosmology from Sunyaev-Zeldovich cluster counts. 2016, A&A, 594, A24, arXiv:1502.01597

[290] Planck Collaboration II, *Planck* 2018 results. II. Low Frequency Instrument data processing. 2020, A&A, 641, A2, arXiv:1807.06206

[291] Planck Collaboration III, *Planck* 2018 results. III. High Frequency Instrument data processing. 2020, A&A, 641, A3, arXiv:1807.06207

[292] Planck Collaboration IV, *Planck* 2018 results. IV. Diffuse component separation. 2020, A&A, 641, A4, arXiv:1807.06208

[293] Planck Collaboration V, *Planck* 2018 results. V. Power spectra and likelihoods. 2020, A&A, 641, A5, arXiv:1907.12875

[294] Planck Collaboration VIII, *Planck* 2018 results. VIII. Gravitational lensing. 2020, A&A, 641, A8, arXiv:1807.06210

[295] Planck Collaboration IX, *Planck* 2018 results. IX. Constraints on primordial non-Gaussianity. 2020, A&A, 641, A9, arXiv:1905.05697

[296] Planck Collaboration X, *Planck* 2018 results. X. Constraints on inflation. 2020, A&A, 641, A10, arXiv:1807.06211

[297] Planck Collaboration Int. XLVI, *Planck* intermediate results. XLVI. Reduction of large-scale systematic effects in HFI polarization maps and estimation of the reionization optical depth. 2016, A&A, 596, A107, arXiv:1605.02985

[298] Planck Collaboration Int. XLVII, *Planck* intermediate results. XLVII. Constraints on reionization history. 2016, A&A, 596, A108, arXiv:1605.03507

[299] Planck Collaboration Int. LI, *Planck* intermediate results. LI. Features in the cosmic microwave background temperature power spectrum and shifts in cosmological parameters. 2017, A&A, 607, A95, arXiv:1608.02487

[300] Polnarev, A. G., Polarization and Anisotropy Induced in the Microwave Background by Cosmological Gravitational Waves. 1985, Soviet Astronomy, 29, 607

[301] Poulin, V., Smith, T. L., Karwal, T., & Kamionkowski, M., Early Dark Energy Can Resolve The Hubble Tension. 2019, Phys. Rev. Lett., 122, 221301, arXiv:1811.04083

[302] Prat, J. et al., Dark Energy Survey year 1 results: Galaxy-galaxy lensing. 2018, Phys. Rev., D98, 042005, arXiv:1708.01537

[303] Raveri, M., Hu, B., Frusciante, N., & Silvestri, A., Effective Field Theory of Cosmic Acceleration: constraining dark energy with CMB data. 2014, Phys. Rev., D90, 043513, arXiv:1405.1022

[304] Raveri, M. & Hu, W., Concordance and Discordance in Cosmology. 2019, Phys. Rev., D99, 043506, arXiv:1806.04649

[305] Reid, M. J., Braatz, J. A., Condon, J. J., et al., The Megamaser Cosmology Project. IV. A Direct Measurement of the Hubble Constant from UGC 3789. 2013, ApJ, 767, 154, arXiv:1207.7292

[306] Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D., Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics Beyond LambdaCDM. 2019, ApJ, 876, 85, arXiv:1903.07603

[307] Riess, A. G., Filippenko, A. V., Challis, P., et al., Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. 1998, AJ, 116, 1009, arXiv:astro-ph/9805201

[308] Riess, A. G., Macri, L., Casertano, S., et al., A 3% Solution: Determination of the Hubble Constant with the Hubble Space Telescope and Wide Field Camera 3. 2011, ApJ, 730, 119, arXiv:1103.2976

[309] Riess, A. G., Macri, L. M., Hoffmann, S. L., et al., A 2.4% Determination of the Local Value of the Hubble Constant. 2016, ApJ, 826, 56, arXiv:1604.01424

[310] Riess, A. G. et al., New Parallaxes of Galactic Cepheids from Spatially Scanning the Hubble Space Telescope: Implications for the Hubble Constant. 2018, ApJ, 855, 136, arXiv:1801.01120

[311] Riess, A. G. et al., Milky Way Cepheid Standards for Measuring Cosmic Distances and Application to Gaia DR2: Implications for the Hubble Constant. 2018, ApJ, 861, 126, arXiv:1804.10655

[312] Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S., Cosmic Reionization and Early Star-forming Galaxies: A Joint Analysis of New Constraints from Planck and the Hubble Space Telescope. 2015, ApJ, 802, L19, arXiv:1502.02024

[313] Ross, A. J., Samushia, L., Howlett, C., et al., The clustering of the SDSS DR7 main Galaxy sample – I. A 4 per cent distance measure at $z = 0.15$. 2015, MNRAS, 449, 835, arXiv:1409.3242

[314] Ross, A. J. et al., The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Observational systematics and baryon acoustic oscillations in the correlation function. 2017, MNRAS, 464, 1168, arXiv:1607.03145

[315] Rosset, C., Tristram, M., Ponthieu, N., et al., *Planck* pre-launch status: High Frequency Instrument polarization calibration. 2010, A&A, 520, A13, arXiv:1004.2595

[316] Rubiño-Martín, J. A., Chluba, J., Fendt, W. A., & Wandelt, B. D., Estimating the impact of recombination uncertainties on the cosmological parameter constraints from cosmic microwave background experiments. 2010, MNRAS, 403, 439, arXiv:0910.4383

[317] Saikawa, K. & Shirai, S., Primordial gravitational waves, precisely: The role of thermodynamics in the Standard Model. 2018, JCAP, 1805, 035, arXiv:1803.01038

[318] Sakstein, J. & Jain, B., Implications of the Neutron Star Merger GW170817 for Cosmological Scalar-Tensor Theories. 2017, Phys. Rev. Lett., 119, 251303, arXiv:1710.05893

[319] Saltas, I. D., Sawicki, I., Amendola, L., & Kunz, M., Anisotropic Stress as a Signature of Nonstandard Propagation of Gravitational Waves. 2014, Physical Review Letters, 113, 191101, arXiv:1406.7139

[320] Saltas, I. D., Sawicki, I., & Lopes, I., White dwarfs and revelations. 2018, JCAP, 1805, 028, arXiv:1803.00541

[321] Salvati, L., Douspis, M., & Aghanim, N., Constraints from thermal Sunyaev-Zeldovich cluster counts and power spectrum combined with CMB. 2018, Astron. Astrophys., 614, A13, arXiv:1708.00697

[322] Salvati, L., Pagano, L., Consiglio, R., & Melchiorri, A., Cosmological constraints on the neutron lifetime. 2016, J. Cosmology Astropart. Phys., 1603, 055, arXiv:1507.07243

[323] Sánchez, A. G., Scoccimarro, R., Crocce, M., et al., The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological implications of the configuration-space clustering wedges. 2017, MNRAS, 464, 1640, arXiv:1607.03147

[324] Satpathy, S., Alam, S., Ho, S., et al., The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: on the measurement of growth rate using galaxy correlation functions. 2017, MNRAS, 469, 1369, arXiv:1607.03148

[325] Saviano, N., Mirizzi, A., Pisanti, O., et al., Multi-momentum and multi-flavour active-sterile neutrino oscillations in the early universe: role of neutrino asymmetries and effects on nucleosynthesis. 2013, Phys. Rev., D87, 073006, arXiv:1302.1200

[326] Schellenberger, G. & Reiprich, T. H., HICOSMO: cosmology with a complete sample of galaxy clusters? II. Cosmological results. 2017, MNRAS, 471, 1370, arXiv:1705.05843

[327] Schmittfull, M. M., Challinor, A., Hanson, D., & Lewis, A., On the joint analysis of CMB temperature and lensing-reconstruction power spectra. 2013, Phys. Rev. D, 88, 063012, arXiv:1308.0286

[328] Scolnic, D., Casertano, S., Riess, A., et al., Supercal: Cross-calibration of Multiple Photometric Systems to Improve Cosmological Measurements with Type Ia Supernovae. 2015, ApJ, 815, 117, arXiv:1508.05361

[329] Scolnic, D. M. et al., The Complete Light-curve Sample of Spectroscopically Confirmed Type Ia Supernovae from Pan-STARRS1 and Cosmological Constraints from The Combined Pantheon Sample. 2018, ApJ, 859, 101, arXiv:1710.00845

[330] Seager, S., Sasselov, D. D., & Scott, D., A New Calculation of the Recombination Epoch. 1999, ApJ, 523, L1, arXiv:astro-ph/9909275

[331] Seager, S., Sasselov, D. D., & Scott, D., How Exactly Did the Universe Become Neutral? 2000, ApJS, 128, 407, arXiv:astro-ph/9912182

[332] Seljak, U., Gravitational lensing effect on cosmic microwave background anisotropies: A Power spectrum approach. 1996, ApJ, 463, 1, arXiv:astro-ph/9505109

[333] Seljak, U., Sugiyama, N., White, M., & Zaldarriaga, M., Comparison of cosmological Boltzmann codes: Are we ready for high precision cosmology? 2003, Phys. Rev. D, 68, 083507, arXiv:astro-ph/0306052

[334] Seljak, U. & Zaldarriaga, M., Signature of gravity waves in polarization of the microwave background. 1997, Phys. Rev. DLett., 78, 2054, arXiv:astro-ph/9609169

[335] Serebrov, A. P. et al., Neutron lifetime measurements with a large gravitational trap for ultracold neutrons. 2018, Phys. Rev., C97, 055503, arXiv:1712.05663

[336] Serenelli, A. & Basu, S., Determining the initial helium abundance of the Sun. 2010, ApJ, 719, 865, arXiv:1006.0244

[337] Sereno, M., Covone, G., Izzo, L., et al., PSZ2LenS. Weak lensing analysis of the Planck clusters in the CFHTLenS and in the RCSLenS. 2017, MNRAS, 472, 1946, arXiv:1703.06886

[338] Shaw, J. R. & Chluba, J., Precise cosmological parameter estimation using COSMOREC. 2011, MNRAS, 415, 1343, arXiv:1102.3683

[339] Sievers, J. L. et al., The Atacama Cosmology Telescope: Cosmological parameters from three seasons of data. 2013, JCAP, 1310, 060, arXiv:1301.0824

[340] Sifon, C., Battaglia, N., Hasselfield, M., et al., The Atacama Cosmology Telescope: dynamical masses for 44 SZ-selected galaxy clusters over 755 square degrees. 2016, MNRAS, 461, 248, arXiv:1512.00910

[341] Slatyer, T. R., Indirect dark matter signatures in the cosmic dark ages. I. Generalizing the bound on s-wave dark matter annihilation from Planck results. 2016, Phys. Rev., D93, 023527, arXiv:1506.03811

[342] Slatyer, T. R., Indirect Dark Matter Signatures in the Cosmic Dark Ages II. Ionization, Heating and Photon Production from Arbitrary Energy Injections. 2016, Phys. Rev., D93, 023521, arXiv:1506.03812

[343] Smith, K. M. & Ferraro, S., Detecting Patchy Reionization in the Cosmic Microwave Background. 2017, Phys. Rev. Lett., 119, 021301, arXiv:1607.01769

[344] Smith, R. E., Peacock, J. A., Jenkins, A., et al., Stable clustering, the halo model and nonlinear cosmological power spectra. 2003, MNRAS, 341, 1311, arXiv:astro-ph/0207664

[345] Smith, T. L., Muñoz, J. B., Smith, R., Yee, K., & Grin, D., Baryons still trace dark matter: probing CMB lensing maps for hidden isocurvature. 2017, Phys. Rev., D96, 083508, arXiv:1704.03461

[346] Smoot, G. F., Bennett, C. L., Kogut, A., et al., Structure in the COBE differential microwave radiometer first-year maps. 1992, ApJ, 396, L1

[347] Spergel, D. N., Flauger, R., & Hlovzek, R., Planck Data Reconsidered. 2015, Phys. Rev. D, 91, 023518, arXiv:1312.3313

[348] Springob, C. M., Magoulas, C., Colless, M., et al., The 6dF Galaxy Survey: peculiar velocity field and cosmography. 2014, MNRAS, 445, 2677, arXiv:1409.6161

[349] Stöcker, P., Krämer, M., Lesgourgues, J., & Poulin, V., Exotic energy injection with ExoCLASS: Application to the Higgs portal model and evaporating black holes. 2018, JCAP, 1803, 018, arXiv:1801.01871

[350] Story, K. T., Reichardt, C. L., Hou, Z., et al., A Measurement of the Cosmic Microwave Background Damping Tail from the 2500-Square-Degree SPT-SZ Survey. 2013, ApJ, 779, 86, arXiv:1210.7231

[351] Sunyaev, R. A. & Zeldovich, Y. B., Small-Scale Fluctuations of Relic Radiation. 1970, Ap&SS, 7, 3

[352] Suyu, S. H., Auger, M. W., Hilbert, S., et al., Two Accurate Time-delay Distances from Strong Lensing: Implications for Cosmology. 2013, ApJ, 766, 70, arXiv:1208.6010

[353] Switzer, E. R. & Hirata, C. M., Primordial helium recombination. I. Feedback, line transfer, and continuum opacity. 2008, Phys. Rev. D, 77, 083006, arXiv:astro-ph/0702143

[354] Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M., Revising the Halofit Model for the Nonlinear Matter Power Spectrum. 2012, ApJ, 761, 152, arXiv:1208.2701

[355] Troxel, M. A. et al., Dark Energy Survey Year 1 results: Cosmological constraints from cosmic shear. 2018, Phys. Rev., D98, 043528, arXiv:1708.01538

[356] Troxel, M. A. et al., Survey geometry and the internal consistency of recent cosmic shear measurements. 2018, MNRAS, 479, 4998, arXiv:1804.10663

[357] Valiviita, J., Power Spectra Based Planck Constraints on Compensated Isocurvature, and Forecasts for LiteBIRD and CORE Space Missions. 2017, J. Cosmology Astropart. Phys., 1704, 014, arXiv:1701.07039

[358] van Uitert, E. et al., KiDS+GAMA: Cosmology constraints from a joint analysis of cosmic shear. 2018, MNRAS, 476, 4662, arXiv:1706.05004

[359] Vargas-Magaña, M. et al., The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: theoretical systematics and Baryon Acoustic Oscillations in the galaxy correlation function. 2018, MNRAS, 477, 1153, arXiv:1610.03506

[360] Vázquez, J. A., Bridges, M., Hobson, M. P., & Lasenby, A. N., Model Selection Applied to Reconstruction of the Primordial Power Spectrum. 2012, J. Cosmology Astropart. Phys., 06, 006, arXiv:1203.1252

[361] Viero, M. P., Wang, L., Zemcov, M., et al., HerMES: Cosmic Infrared Background Anisotropies and the Clustering of Dusty Star-forming Galaxies. 2013, ApJ, 772, 77, arXiv:1208.5049

[362] Vikman, A., Can dark energy evolve to the phantom? 2005, Phys. Rev., D71, 023515, arXiv:astro-ph/0407107

[363] Villanueva-Domingo, P., Gariazzo, S., Gnedin, N. Y., & Mena, O., Was there an early reionization component in our universe? 2018, JCAP, 1804, 024, arXiv:1712.02807

[364] von der Linden, A., Mantz, A., Allen, S. W., et al., Robust weak-lensing mass calibration of Planck galaxy clusters. 2014, MNRAS, 443, 1973, arXiv:1402.2670

[365] Weinberg, S., Anthropic bound on the cosmological constant. 1987, Phys. Rev. Lett., 59, 2607

[366] Wong, K. C. et al., H0LiCOW XIII. A 2.4% measurement of H_0 from lensed quasars: 5.3σ tension between early and late-Universe probes. 2019, arXiv e-prints, arXiv:1907.04869

[367] Wong, W. Y., Moss, A., & Scott, D., How well do we understand cosmological recombination? 2008, MNRAS, 386, 1023, arXiv:0711.1357

[368] Yèche, C., Palanque-Delabrouille, N., Baur, J., & du Mas des Bourboux, H., Constraints on neutrino masses from Lyman-alpha forest power spectrum with BOSS and XQ-100. 2017, J. Cosmology Astropart. Phys., 1706, 047, arXiv:1702.03314

[369] Yuan, W., Riess, A. G., Macri, L. M., Casertano, S., & Scolnic, D. M., Consistent Calibration of the Tip of the Red Giant Branch in the Large Magellanic Cloud on the Hubble Space Telescope Photometric System and a Redetermination of the Hubble Constant. 2019, ApJ, 886, 61, arXiv:1908.00993

[370] Yue, A. T., Dewey, M. S., Gilliam, D. M., et al., Improved Determination of the Neutron Lifetime. 2013, Phys. Rev. Lett., 111, 222501, arXiv:1309.2623

[371] Zahn, O., Reichardt, C. L., Shaw, L., et al., Cosmic Microwave Background Constraints on the Duration and Timing of Reionization from the South Pole Telescope. 2012, ApJ, 756, 65, arXiv:1111.6386

[372] Zaldarriaga, M., Spergel, D., & Seljak, U., Microwave Background Constraints on Cosmological Parameters. 1997, ApJ, 488, 1, arXiv:astro-ph/9702157

[373] Zarrouk, P. et al., The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: measurement of the growth rate of structure from the anisotropic correlation function between redshift 0.8 and 2.2. 2018, MNRAS, 477, 1639, arXiv:1801.03062

[374] Zeldovich, Y. B., Kurt, V. G., & Syunyaev, R. A., Recombination of Hydrogen in the Hot Model of the Universe. 1968, Zhurnal Eksperimental noi i Teoreticheskoi Fiziki, 55, 278

[375] Zhang, B. R., Childress, M. J., Davis, T. M., et al., A blinded determination of H_0 from low-redshift Type Ia supernovae, calibrated by Cepheid variables. 2017, MNRAS, 471, 2254, arXiv:1706.07573

[376] Zhang, P., Liguori, M., Bean, R., & Dodelson, S., Probing Gravity at Cosmological Scales by Measurements which Test the Relationship between Gravitational Lensing and Matter Overdensity. 2007, Phys. Rev. Lett., 99, 141302, arXiv:0704.1932

[377] Zhao, G.-B., Pogosian, L., Silvestri, A., & Zylberberg, J., Searching for modified growth patterns with tomographic surveys. 2009, Phys. Rev. D, 79, 083513, arXiv:0809.3791

[378] Zubeldia, i. & Challinor, A., Cosmological constraints from Planck galaxy clusters with CMB lensing mass bias calibration. 2019, MNRAS, 489, 401, arXiv:1904.07887

[379] Zumalacárregui, M. & García-Bellido, J., Transforming gravity: from derivative couplings to matter to second-order scalar-tensor theories beyond the Horndeski Lagrangian. 2014, Phys. Rev., D89, 064046, arXiv:1308.4685