

Gravitational Waves

Yury Holubeu *

October 29, 2024

This note isn't aimed for distribution.

Gravitational waves and applications are discussed in details. Links below show [problems](#) and good theory by Michele Maggiore.

The project is stopped, because I am not working with professionals in GW and there are too many complicated phenomena in GW physics, there is no sense to study them in advance. If I find a professional, I might continue to study them.

ADD COOL PROBLEMS AND ARTICLES IN THAT SECTION. THOMAS MIELING HAS GOOD MODELS! AFTER MAKING THIS GOOD SECTION WHOLE NOTE WILL BE MUCH BETTER. Now there is too much theory in comparison to the number of problems.

Contents

1	Preface and main motivation	8
I	Gravitational Waves on a Nutshell	9
2	Основное о гравитационных волнах	9
2.1	О свойствах типичных гравитационных волн	9
2.1.1	Основные свойства и формулы простой гравитационной волны	9
2.1.2	Simple estimations and examples (!!!!)	12
2.1.3	Об обосновании существования гравитационных волн	12
2.1.4	Некоторые типичные оценки свойств волн (??)	14
2.1.5	On other properties of GW	14
2.2	On radiation and sources of GW	14
2.2.1	О типичной теории излучения (?)	14
2.2.2	On other about radiation of GW (?)	16
2.3	On math and physics for GW (?)	16
2.3.1	О слабом пределе гравитации	16
2.3.2	On main math for GW	18
2.4	On detection of GW	18
2.4.1	Обзор экспериментальных результатов	18
2.4.2	О детектировании в интерферометре Майкельсона	19
2.4.3	О формализме необходимой математике для описания шумов (!!??)	20
2.4.4	Обзор других способов детектирования	21
2.5	On noises in interferometric GW detectors and their reduction (!!??)	21
2.5.1	Overview of noises	21
2.5.2	О материалах и покрытиях для детекторов (!!??)	28
3	On extra topics in GW	28
3.1	On information from GW	28
3.1.1	О волнах как способе проверить обобщения ОТО	28
3.1.2	Об информации о Вселенной на ранних этапах	28
3.2	О typical models with GW	28
3.2.1	On interaction with EM field	28
3.2.2	On interaction with light near star	28
3.2.3	On interaction with matter	28
3.2.4	On other possible interactions	28
3.2.5	Overview of models of T. Mieling	29
3.3	О других гравитационных волнах	29
3.3.1	О сильных волнах (?)	29

*yuri.holubev@gmail.com

3.3.2	О других волнах	30
3.4	О другом про гравитационные волны	30
3.5	О других способах детектирования волн	31
3.5.1	Об очень малоперспективных способах	32

II Properties of Gravitational Waves 33

4	Типичные гравитационные волны	33
4.1	Свойства гравитационных волн	33
4.1.1	Обоснование существования волн (???)	33
4.1.2	Поляризация	33
4.1.3	Основные свойства гравитационных волн	33
4.1.4	Second-order equations	33
4.1.5	Другие свойства гравитационных волн	33
4.1.6	Построение теории гравитационных волн с нуля	34
4.2	Излучение гравитационных волн	34
4.2.1	Типичное излучение гравитационных волн	34
4.2.2	Волны от типичных объектов	34
4.2.3	Другое про излучение гравитационных волн	35
5	Sathyaprakash's GW article	35
5.1	A New Window onto the Universe	35
5.1.1	Birth of gravitational astronomy	37
5.1.2	What this review is about	37
5.2	Gravitational Wave Observables	39
5.2.1	Gravitational field vs gravitational waves	39
5.2.2	Gravitational wave polarizations	40
5.2.3	Direction to a source	40
5.2.4	Amplitude of gravitational waves - the quadrupole approximation	41
5.2.5	Frequency of gravitational waves	43
5.2.6	Luminosity in gravitational waves	44
5.3	Sources of Gravitational Waves	45
5.3.1	Man-made sources	45
5.3.2	Gravitational wave bursts from gravitational collapse	46
5.3.3	Gravitational wave pulsars	47
5.3.4	Radiation from a binary star system	48
5.3.5	Quasi-normal modes of a black hole	53
5.3.6	Stochastic background	54
5.4	Physics with Gravitational Waves by Sathyaprakash	56
5.4.1	Speed of gravitational waves	56
5.4.2	Polarization of gravitational waves	57
5.4.3	Gravitational radiation reaction	57
5.4.4	Black hole spectroscopy	58
5.4.5	The two-body problem in general relativity	60
5.4.6	Tests of general relativity	72
5.5	Astrophysics with Gravitational Waves	77
5.5.1	Interacting compact binaries	77
5.5.2	Black hole astrophysics	79
5.5.3	Neutron star astrophysics	82
5.5.4	Multimessenger gravitational-wave astronomy	86

6	Typical models by Michele Maggiore	88
6.0.1	The geometric approach to GWs	88
6.0.2	1.1 Expansion around flat space	89
6.0.3	1.2 The transverse-traceless gauge	93
6.0.4	1.3 Interaction of GWs with test masses	98
6.0.5	1.3.1 Geodesic equation and geodesic deviation	98
6.0.6	1.3.2 Local inertial frames and freely falling frames	100
6.0.7	1.3.3 TT frame and proper detector frame	101
6.0.8	The TT frame	101
6.0.9	The proper detector frame	103
6.0.10	1.4 The energy of GWs	111
6.0.11	1.4.1 Separation of GWs from the background	111
6.0.12	1.4.2 How GWs curve the background	113
6.0.13	1.4.3 The energy-momentum tensor of GWs	119
6.0.14	The energy flux	121
6.0.15	1.5 Propagation in curved space-time	125
6.0.16	1.5.1 Geometric optics in curved space	127
6.0.17	Electromagnetic waves	127
6.0.18	Gravitational waves	129
6.0.19	1.5.2 Absorption and scattering of GWs	131
6.0.20	1.6 Solved problems	132
6.0.21	Problem 1.1. Linearization of the Riemann tensor in curved space	132
6.0.22	Problem 1.2. Gauge transformation of $h_{\mu\nu}$ and $R_{\mu\nu\rho\sigma}^{(1)}$	133
6.0.23	Further reading	135
6.0.24	2	136
6.0.25	The field-theoretical approach to GWs	136
6.0.26	2.1 Linearized gravity as a classical field theory	137
6.0.27	2.1.1 Noether's theorem	137
6.0.28	2.1.2 The energy-momentum tensor of GWs	142
6.0.29	2.1.3 The angular momentum of GWs	144
6.0.30	2.2 Gravitons	149
6.0.31	2.2.1 Why a spin-2 field?	149
6.0.32	2.2.2 The Pauli-Fierz action	154
6.0.33	Massless particles in field theory	154
6.0.34	The graviton and its action	155
6.0.35	The graviton propagator	157
6.0.36	2.2.3 From gravitons to gravity	158
6.0.37	2.2.4 Effective field theories and the Planck scale	162
6.0.38	2.3 Massive gravitons	165
6.0.39	2.3.1 Phenomenological bounds	165
6.0.40	2.3.2 Field theory of massive gravitons	167
6.0.41	A warm-up: massive photons	167
6.0.42	Massive gravitons	170
6.0.43	2.4 Solved problems	177
6.0.44	Problem 2.1. The helicity of gravitons	177
6.0.45	Problem 2.2. Angular momentum and parity of graviton states	180
6.0.46	Further reading	182
6.0.47	Generation of GWs in linearized theory	182
6.0.48	3.1 Weak-field sources with arbitrary velocity	184
6.0.49	3.2 Low-velocity expansion	187
6.0.50	3.3 Mass quadrupole radiation	191
6.0.51	3.3.1 Amplitude and angular distribution	191
6.0.52	Angular distribution of quadrupole radiation	192

6.0.53	Absence of monopole and dipole gravitational radiation	194
6.0.54	3.3.2 Radiated energy	195
6.0.55	3.3.3 Radiated angular momentum	197
6.0.56	3.3.4 Radiation reaction on non-relativistic sources	199
6.0.57	Gravitational radiation and the equivalence principle	202
6.0.58	3.3.5 Radiation from a closed system of point masses	203
6.0.59	3.4 Mass octupole and current quadrupole	207
6.0.60	Mass octupole	207
6.0.61	Current quadrupole	208
6.0.62	Linear momentum losses	212
6.0.63	3.5 Systematic multipole expansion	213
6.0.64	Multi-index notation	215
6.0.65	3.5.1 Symmetric-trace-free (STF) form	216
6.0.66	Scalar fields	217
6.0.67	Vector field	218
6.0.68	Gravitational field	219
6.0.69	3.5.2 Spherical tensor form	221
6.0.70	Spherical components of tensors	221
6.0.71	Vector and tensor spherical harmonics	226
6.0.72	Spin-2 tensor harmonics	229
6.0.73	Relation with the source moments	233
6.0.74	Radiated power	236
6.0.75	3.6 Solved problems	238
6.0.76	Problem 3.1. Quadrupole radiation from an oscillating mass	238
6.0.77	Further reading	251
6.0.78	Applications	251
6.0.79	4.1 Inspiral of compact binaries	252
6.0.80	4.1.1 Circular orbits. The chirp amplitude	254
6.0.81	4.1.2 Elliptic orbits. (I) Total power and frequency spectrum of the radiation emitted	261
6.0.82	Elliptic Keplerian orbits	261
6.0.83	Radiated power	264
6.0.84	Frequency spectrum	267
6.0.85	4.1.3 Elliptic orbits. (II) Evolution of the orbit under back-reaction	271
6.0.86	The evolution equations	271
6.0.87	Orbit circularization	274
6.0.88	The time to coalescence	275
6.0.89	4.1.4 Binaries at cosmological distances	277
6.0.90	A reminder of FRW kinematic	277
6.0.91	Propagation of GWs in a FRW Universe	281
6.0.92	4.2 Radiation from rotating rigid bodies	286
6.0.93	4.2.1 GWs from rotation around a principal axis	287
6.0.94	4.2.2 GWs from freely precessing rigid bodies	291
6.0.95	"Wobble" radiation from an axisymmetric rigid body	292
6.0.96	The back-reaction of GWs	296
6.0.97	Rotating and precessing triaxial bodies	299

III Detection of GW and Noises

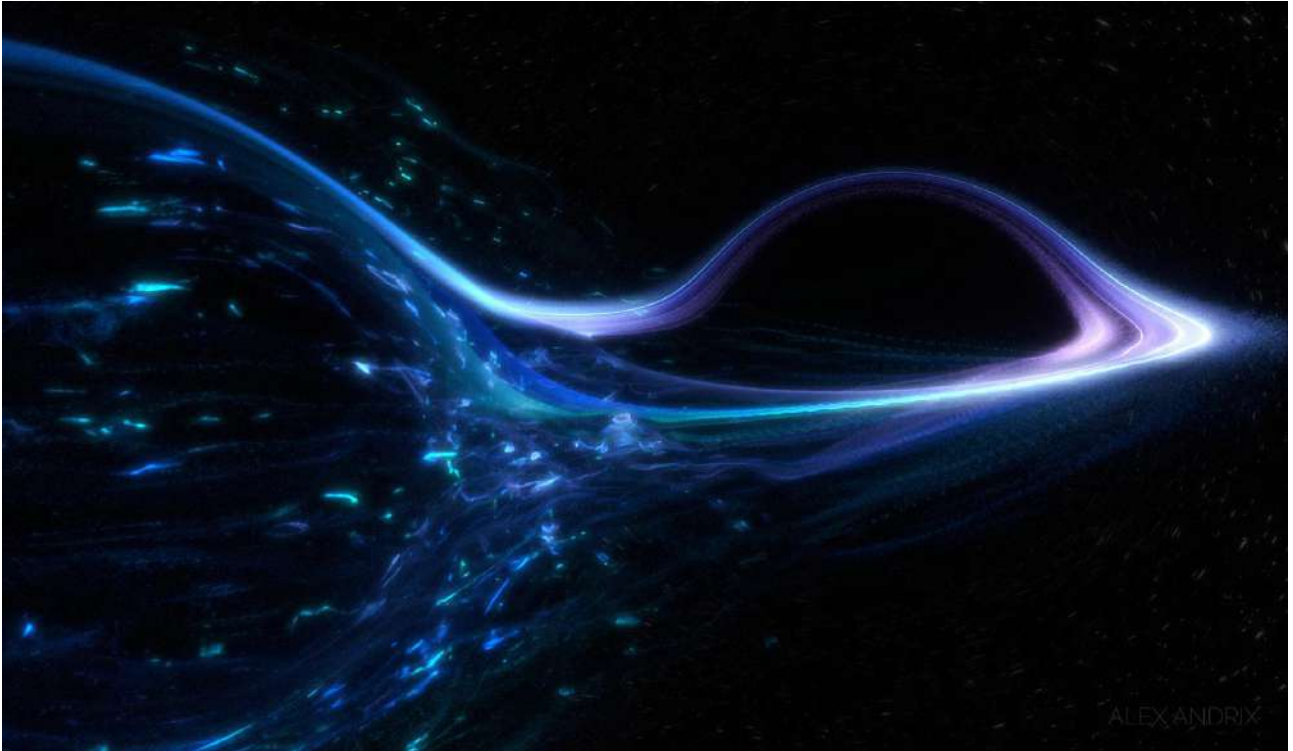
303

7 Известные методы	303
7.1 Gravitational Wave Detectors and Their Sensitivity by Sathyaprakash	303
7.1.1 Principles of the operation of resonant mass detectors	303
7.1.2 Principles of the operation of beam detectors	305
7.1.3 Practical issues of ground-based interferometers	310
7.1.4 Detection from space	314
7.1.5 Characterizing the sensitivity of a gravitational wave antenna	316
7.1.6 Source amplitudes vs sensitivity	319
7.1.7 Network detection	320
7.1.8 False alarms, detection threshold and coincident observation	323
7.2 Data Analysis by Sathyaprakash	324
7.2.1 Matched filtering and optimal signal-to-noise ratio	325
7.2.2 Suboptimal filtering methods	330
7.2.3 Measurement of parameters and source reconstruction	332
8 Шумы при детектировании интерферометром (!!!!!)	340
8.0.1 On thermal noise in interferometric due to dielectric optical coatings	340
8.0.2 On mechanical losses	343
8.0.3 On other typical noises (???)	343
8.0.4 Общие вопросы о шумах	343
9 Эксперименты	344
9.1 Метод LIGO	344
9.1.1 Общее описание метода	344
9.1.2 О способах обработки результатов	345
9.1.3 Результаты	345
9.2 VIRGO Experiment	345
9.2.1 Общее описание метода	345
9.2.2 Other about VIRGO	345
10 Другие методы детектирования волн	345
10.0.1 Detection by phase modulation of the light from a distant star	345
10.1 Другие методы детектирования волн	366
10.1.1 Методы на основе взаимодействия с электромагнитным полем	366
10.1.2 Методы на основе запаздывания времени вблизи пульсаров	366
10.2 Другие эксперименты	367
10.2.1 On ETpathfinder	367
10.2.2 Простейший эксперимент по детектированию волн	367
10.2.3 Исторически первый эксперимент	367
IV Problems	368
11 Основные задачи о гравитационных волнах	368
11.0.1 Задачи на идейное понимание гравитационных волн	368
11.0.2 Задачи на понимание формул гравитационных волн	368
11.0.3 Какие будут гравитационные волны от данной вращающейся системы?	368
12 Другие задачи о гравитационных волнах	368
12.1 Задачи на гравитационные волны в модифицированных теориях гравитации	368
V Other Models and Theories	369

13	Различные волны и явления	369
13.1	Волны от разных источников	369
13.1.1	Волны от взрыва	369
13.2	Разные объёмные волны	369
13.2.1	Объёмные волны Бонди - Пирани - Робинсона	369
13.2.2	Сильная гравитационная волна Бонди-Робинсона	369
13.2.3	Метрика Такено	373
13.2.4	Метрика Розена	373
13.2.5	Метрика Переса	374
13.2.6	Цилиндрические волны Эйнштейна - Розена	374
13.2.7	Другие гравитационные волны	374
13.2.8	Гравитационные волны в начале вселенной	374
13.3	Другие и редкие волны	374
13.3.1	Ударные гравитационные волны	374
13.3.2	Турбулентность в гравитационных волнах	375
13.3.3	Исследование коллапса гравитирующих систем	375
13.3.4	Гравитационные волны в космологии (?)	375
13.3.5	Многомерные гравитационные волны	375
14	Взаимодействия гравитационных волн	375
14.0.1	Взаимодействие гравитационных волн со сплошной средой	375
14.0.2	Взаимодействие гравитационных волн со сплошной средой по Лашкевичу	375
14.0.3	Взаимодействие гравитационных волн с электромагнитным полем	375
14.1	Взаимодействие гравитационной волны с электромагнитными полем	376
14.1.1	Гравитационные волны в электромагнитных полях	376
14.1.2	Колебания электромагнитного поля за счет гравитационной волны (?)	377
14.2	Прохождение света через гравитационную волну (!!!)	388
14.2.1	Линзирование гравитационной волной	388
14.2.2	Прохождение волн через сплошную среду	388
14.3	Прохождение волн через волновод	390
14.3.1	О разных факторах	390
14.3.2	The Response of Optical Fibres to Gravitational Waves by Thomas B. Mieling	391
15	Другие темы	408
15.1	Cosmology with Gravitational Wave Observations by Sathyaprakash	408
15.1.1	Detecting a stochastic gravitational wave background	409
15.1.2	Origin of a random background of gravitational waves	412
15.1.3	Cosmography: gravitational wave measurements of cosmological parameters	414
15.2	О других шумах и факторах	415
15.2.1	Шумы от вакуумных эффектов (????)	415
15.3	Более редкие случаи	415
15.3.1	Волны в искривленном 4-пространстве	415
15.4	Другие модели и вопросы	419
15.4.1	Другие вопросы для иллюстрации грав. волн	419
15.4.2	Другие идейные вопросы о грав. волнах	419
15.4.3	Гравитационные антенны	419
15.4.4	Рассеяние гравитационных волн	420
15.4.5	Дифракция гравитационной волной	420

A Введение	421
A.1 Еще мотивация	421
A.1.1 Множество приложений	421
A.1.2 О пользе в плане понимания природы и саморазвитии	421
A.1.3 Удивительные факты	421
A.2 Обзор гравитационных волн	421
A.2.1 Популярно о гравитационных волнах	421
A.3 Мышление профессионала в гравитационных волнах	422
A.3.1 Отношение к гравитационным волнам	422
A.3.2 Особенности изучения гравитационных волн	422
A.3.3 Особенности изучения детектирования волн	424
A.3.4 Как догадаться до всех главных идей самому?	424
A.4 Литература	424
A.4.1 Основная	424
A.4.2 Профессиональная и углубленная литература	424
A.4.3 Статьи о детектировании гравитационных волн и шумах	425
A.4.4 Статьи о разных эффектах и теоретических моделях	426
A.4.5 Дополнительная литература	426
A.4.6 Литература в помощь	427
A.5 Описание записи	427
A.5.1 Общие особенности записи	427
A.5.2 Обозначения и константы	427
A.6 Головоломки на разные случаи жизни	427
A.6.1 Бытовые головоломки	427
A.6.2 Принципиальные головоломки	427
A.6.3 Головоломки о деталях	427
A.7 Короткий исторический обзор	427
A.7.1 О первых открытиях	427
A.7.2 О результатах конца 20го и начала 21го веков	427
B Некоторые концепции математики	427
B.0.1 О методах диффгема для грав волн (!!!)	428
C Некоторые концепции физики	428
C.1 О теориях в помощь и других теоретических моделях	428
C.1.1 Об электродинамике в искривленном пространстве для грав волн	428
Bibliography	429

1 Preface and main motivation



(потом напишу предисловие, не до этого пока)

Part I

Gravitational Waves on a Nutshell

2 Основное о гравитационных волнах

2.1 О свойствах типичных гравитационных волн

2.1.1 Основные свойства и формулы простой гравитационной волны

Главные итоговые формулы

(тут абзацем текстом напишу, откуда они, в ниже параграфах уже с формулами вывод этого будет указан!)

В плоскости, поперечной к направлению движения волны имеем

$$dl_{\perp}^2 = (1 - h_{+}) (dx^2)^2 + (1 + h_{+}) (dx^3)^2 - 2h_{\times} dx^2 dx^3$$

Рассмотрим маленькую окружность радиуса dr в плоскости x^2x^3 (точнее, в соответствующем касательном пространстве), перпендикулярной волновому вектору. Будем задавать точку на окружности углом θ , отсчитываемым от оси x^2 . Тогда координаты этой точки отличаются от координат центра окружности на величины

$$dx^2 = dr \cos \theta, \quad dx^3 = dr \sin \theta.$$

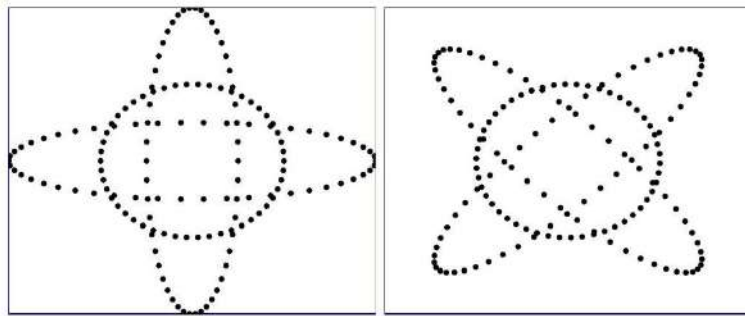
Отсюда имеем

$$dl_{\perp} = dr \left(1 - \frac{h_{+}}{2} \cos 2\theta - \frac{h_{\times}}{2} \sin 2\theta \right) = dr \left(1 - \frac{h_{*}}{2} \cos 2(\theta - \varphi) \right),$$

где

$$h_{+} = h_{*} \cos 2\varphi, \quad h_{\times} = h_{*} \sin 2\varphi.$$

(важные слова про поляризацию!)



О порядках малости в грав волнах (???)

(это мб ключевым, так что отдельно сразу пишу. до какого порядка нам нужны формулы???? пока не уверен, все просто берут формулы и все, не особо акцентируя на этом.)

Основные свойства

Волновое уравнение и условие калибровки:

$$\square\psi_{\mu\nu} = 0, \quad \psi_{\nu,\mu}^{\bar{\mu}} = 0.$$

Подставим:

$$\begin{aligned}\psi_{\mu\nu}(x) &= A_{\mu\nu}e^{-ikx}, \quad A_{\nu\mu} = A_{\mu\nu}, \\ \psi'_{\mu\nu} &= \psi_{\mu\nu} + \xi_{\bar{\mu},\nu} + \xi_{\bar{\nu},\mu} - \eta_{\mu\nu}\xi_{,\lambda}^{\lambda}, \\ \square\xi^{\mu} &= 0\end{aligned}$$

У нас просто плоские волны с разными амплитудами и соотношениями на них: Общее решение имеет вид линейной комбинации всех возможных плоских волн

$$\begin{aligned}\psi_{\mu\nu}(x) &= \frac{1}{2} \sum_{\varepsilon=\pm} \int \frac{d^{d-1}k}{(2\pi)^{d-1}} A_{\mu\nu}^{\varepsilon}(\mathbf{k}) e^{-i\varepsilon(|\mathbf{k}|x^0 - \mathbf{k}\mathbf{r})} \\ |\mathbf{k}|A_{0\nu}^{\pm}(\mathbf{k}) + k^{\bar{i}}A_{i\nu}^{\pm}(\mathbf{k}) &= 0, \quad A_{\nu\mu}^{\pm}(\mathbf{k}) = A_{\mu\nu}^{\pm}(\mathbf{k}), \quad A_{\mu\nu}^{-}(\mathbf{k}) = (A_{\mu\nu}^{+}(\mathbf{k}))^{*}\end{aligned}$$

Просто какие-то коэффициенты тут будут нулем.

Тут (!!)

$$\alpha, \beta, \dots = 0, 1; \quad a, b, \dots = 2, \dots, d-1.$$

Вообще, условием

$$\psi'_{\mu\nu} = \psi_{\mu\nu} + \xi_{\bar{\mu},\nu} + \xi_{\bar{\nu},\mu} - \eta_{\mu\nu}\xi_{,\lambda}^{\lambda}.$$

где $\square\xi^{\mu} = 0$ можно убить многие компоненты. Возьмем ξ^{μ} :

$$\xi^{\mu} = \text{Re } X^{\mu} e^{-i\omega(x^0 - x^1)}.$$

получим новые амплитуды

$$\begin{aligned}A'_{00} &= A_{00} - i\omega(X^0 + X^1), \\ A'_{0a} &= A_{0a} + i\omega X^a, \\ A'_{ab} &= A_{ab} - i\omega(X^0 - X^1)\delta_{ab}.\end{aligned}$$

Устраняются еще d компонент, если например, положить

$$X^0 + X^1 = -i\omega^{-1}A_{00}, \quad X^a = i\omega^{-1}A_{0a}, \quad X^0 - X^1 = -\frac{i\omega^{-1}}{d-2} \sum_{a=1}^{d-2} A_{aa}$$

Гравитационные волны через световой конус по Лашкевичу

Есть уравнение

$$\square\psi_{\mu\nu} = 0, \quad \psi_{\nu,\mu}^{\bar{\mu}} = 0$$

(вот это слишком уж затратно по энергии, это я не знаю вообще.)

Для постоянных по x^a функций уравнений принимают вид

$$\partial_+\partial_-\psi_{\mu\nu} = 0, \quad \partial_+\psi_{-\mu} + \partial_-\psi_{+\mu} = 0$$

$$\psi_{-\mu}^L = \psi_{+\mu}^R = 0.$$

$$dl^2 = (dx^1)^2 + (\delta_{ab} - h_{ab}) dx^a dx^b$$

Если мы возьмем $(d - 2)$ -мерную сферу в плоскости, перпендикулярной волне, то функции h_{ab} описывают небольшие центрально-симметричные деформации такой сферы, причем величина h_{aa} дает нам уменьшение диаметра сферы в направлении оси x^a , а $h_{ab} (a \neq b)$ дает уменьшение диаметра сферы в диагональном направлении $x^a = x^b$ в плоскости $x^a x^b$ и увеличение диаметра в перпендикулярном ему диагональном направлении $x^a = -x^b$. В силу бесследовости объем сферы в первом порядке по h_{ab} сохраняется: $\det(\gamma_{ab}) = \det(\delta_{ab} - h_{ab}) \simeq 1 - \text{tr}(h_{ab}) = 1$.

(!! да, выше это известное свойство, напишу четче)

Важно, что эта $(d - 2)$ -мерная сфера может быть физически реализована как система частиц. Именно, свободные пробные частицы, неподвижные в системе координат, в которой реализуется поперечная калибровка, остаются неподвижными в этой системе координат в течение всего времени прохождения волны. Действительно, в силу калибровки $h_{0\mu} = 0$ имеем $\Gamma_{00}^\mu = \frac{1}{2}g^{\mu\nu}(2h_{0\nu,0} - h_{00,\nu}) = 0$. Это значит, что уравнение движения свободной частицы имеет вид

$$\ddot{x}^\mu = -2\Gamma_{i0}^\mu \dot{x}^i \dot{x}^0 - \Gamma_{ij}^\mu \dot{x}^i \dot{x}^j,$$

То есть, если частица имеет нулевую скорость ($\dot{x}^i = 0$) в какой-то момент времени, то она имеет и нулевое ускорение ($\ddot{x}^i = 0$). Значит, она будет оставаться неподвижной в поле волны сколь угодно долго.

Энергия от гравитационных волн

(хз, спишу у Лашкевича)

Основное о связи величины волн и энергии источника, а также уносимой энергии

Radiation is the quadrupole formula, and it depends only on the density (ρ) and velocity fields of the Newtonian system. If we define the spatial tensor Q_{jk} , the second moment of the mass distribution, by the equation

$$Q_{jk} = \int \rho x_j x_k d^3x$$

then the amplitude of the emitted gravitational wave is, at lowest order, the three-tensor

$$h_{jk} = \frac{2}{r} \frac{d^2 Q_{jk}}{dt^2}.$$

$$T_{\alpha\beta} = \frac{1}{32\pi} \left\langle h_{jk,\alpha}^{\text{TT}} h_{,\beta}^{\text{TT}jk} \right\rangle,$$

$$L_{gw} = \frac{1}{5} \left(\sum_{j,k} \ddot{Q}_{jk} \ddot{Q}_{jk} - \frac{1}{3} \ddot{Q}^2 \right),$$

$$L_0 = c^5/G = 3.6 \times 10^{52} \text{ W}.$$

$$-\frac{d\mathcal{E}}{dt} = I = \frac{G}{45} \ddot{D}_{ij}^2.$$

On estimated values

On frequency of GWs

In most cases the frequency will be related to the natural frequency for a self-gravitating body, defined as

$$\omega_0 = \sqrt{\pi G \bar{\rho}}, \text{ or } f_0 = \omega_0/2\pi = \sqrt{G \bar{\rho}/4\pi},$$

where $\bar{\rho}$ is the mean density of mass-energy in the source.

The characteristic frequency of the radiation of a compact object of mass M and radius R is

$$f_0 = \frac{1}{4\pi} \left(\frac{3M}{R^3} \right)^{1/2} \simeq \text{kHz} \left(\frac{10M_\odot}{M} \right).$$

2.1.2 Simple estimations and examples (!!!!)

Волны от двух вращающихся тел

Волны от типичных объектов

(человек махнул рукой, машина проехала, т.п. оценки на эти случаи соберу)

2.1.3 Об обосновании существования гравитационных волн

(wiki)

Main ideas why they should exist and why they can't exist

(??? так и не знаю, что это за статьи, которые заставили Эйнштейна сомневаться? но очень интересно посмотреть было бы, когда-то подумаю про это, если встречу того, кто разбирается в этом.)

Historical overview of their existence

(wiki, i'll later rewrite)

The possibility of gravitational waves and that those might travel at the speed of light was discussed in 1893 by Oliver Heaviside, using the analogy between the inverse-square law of gravitation and the electrostatic force.[22] In 1905, Henri Poincaré proposed gravitational waves, emanating from a body and propagating at the speed of light, as being required by the Lorentz transformations[23] and suggested that, in analogy to an accelerating electrical charge producing electromagnetic waves, accelerated masses in a relativistic field theory of gravity should produce gravitational waves.[24][25]

In 1915 Einstein published his general theory of relativity, a complete relativistic theory of gravitation. He conjectured, like Poincaré, that the equation would produce gravitational waves, but, as he mentions in a letter to Schwarzschild in February 1916,[25] these could not be similar to electromagnetic waves. Electromagnetic waves are produced by dipole motion, requiring both a positive and a negative charge; gravitation has no equivalent to negative charge. Einstein continued to work through the complexity of the equations of general relativity to find an alternative wave model. The result was published in June 1916,[4] and there he came to the conclusion that the gravitational wave must propagate with the speed of light, and there must, in fact, be three types of gravitational waves dubbed longitudinal-longitudinal, transverse-longitudinal, and transverse-transverse by Hermann Weyl.[25]

However, the nature of Einstein's approximations led many (including Einstein himself) to doubt the result. In 1922, Arthur Eddington showed that two of Einstein's types of waves were

artifacts of the coordinate system he used, and could be made to propagate at any speed by choosing appropriate coordinates, leading Eddington to jest that they “propagate at the speed of thought”.

This also cast doubt on the physicality of the third (transverse–transverse) type that Eddington showed always propagate at the speed of light regardless of coordinate system. In 1936, Einstein and Nathan Rosen submitted a paper to *Physical Review* in which they claimed gravitational waves could not exist in the full general theory of relativity because any such solution of the field equations would have a singularity. The journal sent their manuscript to be reviewed by Howard P. Robertson, who anonymously reported that the singularities in question were simply the harmless coordinate singularities of the employed cylindrical coordinates. Einstein, who was unfamiliar with the concept of peer review, angrily withdrew the manuscript, never to publish in *Physical Review* again. Nonetheless, his assistant Leopold Infeld, who had been in contact with Robertson, convinced Einstein that the criticism was correct, and the paper was rewritten with the opposite conclusion and published elsewhere. In 1956, Felix Pirani remedied the confusion caused by the use of various coordinate systems by rephrasing the gravitational waves in terms of the manifestly observable Riemann curvature tensor.[27]

At the time, Pirani’s work was overshadowed by the community’s focus on a different question: whether gravitational waves could transmit energy. This matter was settled by a thought experiment proposed by Richard Feynman during the first “GR” conference at Chapel Hill in 1957. In short, his argument known as the “sticky bead argument” notes that if one takes a rod with beads then the effect of a passing gravitational wave would be to move the beads along the rod; friction would then produce heat, implying that the passing wave had done work. Shortly after, Hermann Bondi published a detailed version of the “sticky bead argument”.[25] This later led to a series of articles (1959 to 1989) by Bondi and Pirani that established the existence of plane wave solutions for gravitational waves.[28]

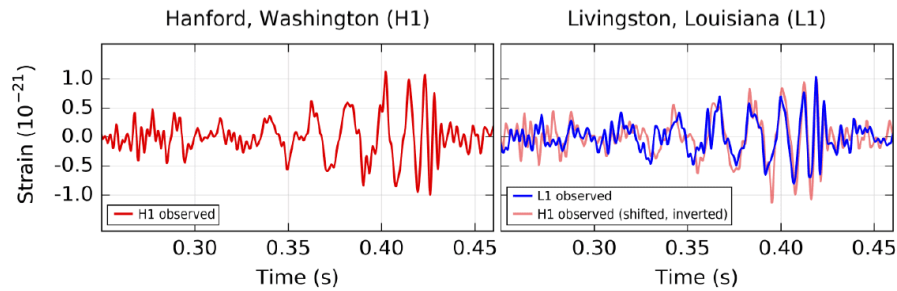
Paul Dirac further postulated the existence of gravitational waves, declaring them to have “physical significance” in his 1959 lecture at the Lindau Meetings.[29] Further, it was Dirac who predicted gravitational waves with a well defined energy density in 1964.[30]

After the Chapel Hill conference, Joseph Weber started designing and building the first gravitational wave detectors now known as Weber bars. In 1969, Weber claimed to have detected the first gravitational waves, and by 1970 he was “detecting” signals regularly from the Galactic Center; however, the frequency of detection soon raised doubts on the validity of his observations as the implied rate of energy loss of the Milky Way would drain our galaxy of energy on a timescale much shorter than its inferred age. These doubts were strengthened when, by the mid-1970s, repeated experiments from other groups building their own Weber bars across the globe failed to find any signals, and by the late 1970s consensus was that Weber’s results were spurious.[25]

In the same period, the first indirect evidence of gravitational waves was discovered. In 1974, Russell Alan Hulse and Joseph Hooton Taylor, Jr. discovered the first binary pulsar, which earned them the 1993 Nobel Prize in Physics.[31] Pulsar timing observations over the next decade showed a gradual decay of the orbital period of the Hulse–Taylor pulsar that matched the loss of energy and angular momentum in gravitational radiation predicted by general relativity.[32][33][25]

Об экспериментальном доказательстве существования гравитационных волн

(укажу на эксперименты, тоже очень полезно.)



О теоретическом обосновании существования (???)

(мб пару абзацев будет, посмотрим.)

2.1.4 Некоторые типичные оценки свойств волн (??)

(в духе - быстро оценить, не влияет ли что-то такое особенное на волны? потом напишу это, просто хотелось бы это понимать.)

2.1.5 On other properties of GW

(все очень редко нужное)

On degrees of freedom (???)

(четко напишу про это)

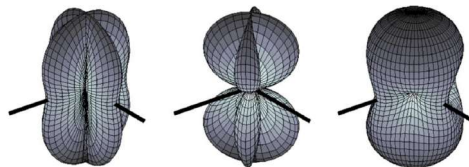
О влиянии искривленности пространства (????)

2.2 On radiation and sources of GW

(тут много всего может быть)

2.2.1 О типичной теории излучения (?)

Об излучении по Тонгу (??)



(мб да, все-таки пойму, как дойти до диаграммы излученной энергии в зависимости от направления, тут ее приведу тогда. пока не знаю.)

Об излучении по Лашкевичу для произвольной размерности (??)

Основные формулы

$$\begin{aligned}\psi_{\mu\nu}(t, \mathbf{r}) &= -16\pi G \int_{-\infty}^t dt' \int d^{d-1}x' G^R(t-t', |\mathbf{r}-\mathbf{r}'|) T_{\mu\nu}(t', \mathbf{r}') = \\ &= -4G \int d^3x' \frac{T_{\mu\nu}(t-|\mathbf{r}-\mathbf{r}'|, \mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|},\end{aligned}$$

$G^R(t, \mathbf{r})$: $\square G^R(t, \mathbf{r}) = \delta(t)\delta(\mathbf{r})$, и переходя в Фурье и обратно, получаем для $d = 4$:
 $G^R(t, \mathbf{r}) = \frac{\delta(t-r)}{4\pi r}$.

Переходим в волновую зону ($R \gg r_{\text{ист}}, r_{\text{ист}}^2/\lambda$):

$$\psi_{\mu\nu}(t, \mathbf{R}) = -\frac{4GM_{\mu\nu}(t-R, \mathbf{n})}{R}, \quad M_{\mu\nu}(t, \mathbf{n}) = \int d^3x T_{\mu\nu}(t + \mathbf{n}\mathbf{r}, \mathbf{r})$$

(тут дальше особые обозначения, которые потом укажу, пока очень коротко пишу итог. разбирать их отдельно нужно.)

$$h_+(t, \mathbf{R}) = -\frac{4GM_+(t-R, \mathbf{n})}{R}, \quad h_\times(t, \mathbf{R}) = -\frac{4GM_\times(t-R, \mathbf{n})}{R},$$

где, если выбрать декартовы координаты так, чтобы было $\mathbf{n} = \partial_1$, имеем

$$\begin{aligned}M_+(t) &= \partial_0^2 \int d^3x T_{++}(t+x^1, \mathbf{r}) \left((x^2)^2 - (x^3)^2 \right), \\ M_\times(t) &= \partial_0^2 \int d^3x T_{++}(t+x^1, \mathbf{r}) \cdot 2x^2x^3.\end{aligned}$$

(??? про T напишу, что это??)

в случае нерелятивистского источника имеем

$$\begin{aligned}M_+(t) &= \frac{1}{4}\partial_0^2 \int d^3x \rho(t, \mathbf{r}) \left((x^2)^2 - (x^3)^2 \right), \\ M_\times(t) &= \frac{1}{2}\partial_0^2 \int d^3x \rho(t, \mathbf{r}) x^2x^3.\end{aligned}$$

Псевдотензор энергии-импульса и поток энергии:

$$\begin{aligned}t^{01}(t+R, R\partial_3) &= \frac{G}{36\pi R^2} \left(\left(\frac{\ddot{D}_{22}(t) - \ddot{D}_{33}(t)}{2} \right)^2 + \ddot{D}_{23}^2(t) \right) \\ dI(t, \partial_1) &= t^{01}(t+R, R\partial_1) R^2 d\phi = \frac{G}{36\pi} \left(\left(\frac{\ddot{D}_{22}(t) - \ddot{D}_{33}(t)}{2} \right)^2 + \ddot{D}_{23}^2(t) \right) d\phi\end{aligned}$$

О выводе свойств гравитационных волн

Это следует из следующих соображений:

определим $T_\nu^- := 2T_{+\nu} = T_{0\nu} + T_{1\nu}$

Уравнение непрерывности:

$$\begin{aligned}\dot{T}_\nu^- (t+x^1, \mathbf{r}) &= \partial_k T_{k\nu} (t+x^1, \mathbf{r}). \\ M_\nu^-(t) &= M_{0\nu}(t) + M_{1\nu}(t) = \text{const}\end{aligned}$$

Умножив его для $\nu = j$ на x^i и проинтегрировав по всему пространству, получим

$$\partial_0 \int d^3x T_j^- (t + x^1, \mathbf{r}) x^i = \int d^3x (\partial_k T_{kj} (t + x^1, \mathbf{r})) x^i = \int d^3x \partial_k (T_{kj} (t + x^1, \mathbf{r}) x^i) - M_{ij}.$$

Первый член в правой части представляет собой интеграл по удаленной границе, где $T_{ki} = 0$, и мы его опустим. Симметризуя по индексам i, j , получаем (????)

$$M_{ij} = -\partial_0 \int d^3x (T_{+i} (t + x^1, \mathbf{r}) x^j + T_{+j} (t + x^1, \mathbf{r}) x^i).$$

Теперь умножим уравнение непрерывности для $\nu = +$ на $x^i x^j$ и проинтегрируем по всему пространству (????)

$$\partial_0 \int d^3x 2T_{++} (t + x^1, \mathbf{r}) x^i x^j = \int d^3x \partial_k (T_{+k} (t + x^1, \mathbf{r}) x^i x^j) - \int d^3x (T_{+i} (t + x^1, \mathbf{r}) x^j + T_{+j} (t + x^1, \mathbf{r}) x^i).$$

Опять первый интеграл обращается в нуль, а второй интеграл совпадает интегралом в правой части (9.20). Отсюда получаем

$$M_{ij}(t) = \partial_0^2 \int d^3x 2T_{++} (t + x^1, \mathbf{r}) x^i x^j$$

2.2.2 On other about radiation of GW (?)

2.3 On math and physics for GW (?)

(тоже всяких указаний полезных наверняка можно сделать. после разбора в теории буду выписывать это!)

2.3.1 О слабом пределе гравитации

(по сути тут все основы гравитации)

Main Formulas

$$G |T^{\mu\nu}| l^2 \ll 1$$

где l — характерные масштабы, на которых меняется метрика.

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\bar{\mu}\bar{\nu}}, \quad |g| = 1 + h, \quad h = h_{\bar{\mu}}^{\bar{\mu}}.$$

$$\eta^{\lambda\kappa} h_{\mu\nu, \lambda\kappa} - 2h_{(\mu, \nu)\lambda}^{\bar{\lambda}} + h_{, \mu\nu} = -16\pi G \left(T_{\mu\nu} - \frac{\eta_{\mu\nu}}{d-2} T \right)$$

$$\psi_{\nu, \mu}^{\bar{\mu}} = 0, \quad \psi_{\mu\nu} = h_{\mu\nu} - \frac{\eta_{\mu\nu}}{2} h.$$

$$h_{\mu\nu} = \psi_{\mu\nu} - \frac{\eta_{\mu\nu}}{d-2} \psi, \quad \psi = \psi_{\mu}^{\bar{\mu}} = -\frac{d-2}{2} h.$$

$$R_{\mu\nu}^{(1)} = -\frac{1}{2} \square h_{\mu\nu}.$$

$$\square \psi_{\mu\nu} = -16\pi G T_{\mu\nu}.$$

$$0 = \square \psi_{\nu, \mu}^{\bar{\mu}} = -16\pi G T_{\nu, \mu}^{\bar{\mu}}$$

Следует понять, что нерелятивистский предел имеет Ньютонов вид:

$$\phi(t, \mathbf{r}) = -\frac{8\pi G}{(d-2)S_{d-1}} \sum_a \frac{m_a}{|\mathbf{r} - \mathbf{r}_a(t)|^{d-3}}.$$

Это есть решение уравнений

$$\begin{aligned}\Delta\psi_{00}(t, \mathbf{r}) &= 16\pi G\rho(t, \mathbf{r}), \\ \Delta\psi_{0i}(t, \mathbf{r}) &= \Delta\psi_{ij}(t, \mathbf{r}) = 0,\end{aligned}$$

где $\Delta = \sum_{i=1}^{d-1} \partial_i^2$ - пространственный лапласиан. Время t в этом уравнении играет роль параметра. Решение:

$$\psi_{00}(t, \mathbf{r}) = -\frac{16\pi G}{(d-3)S_{d-1}} \int d^{d-1}x' \frac{\rho(t, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^{d-3}}, \quad \psi_{0i}(t, \mathbf{r}) = \psi_{ij}(t, \mathbf{r}) = 0,$$

где $S_n = n\pi^{n/2}/\Gamma(\frac{n}{2} + 1)$ - площадь поверхности единичной гиперболы в n -мерном евклидовом пространстве. Ну и остается выразить метрику $h_{00} = \frac{d-3}{d-2}\psi_{00}$ и $h_{ij} = \frac{\psi_{00}}{d-2}\delta_{ij}$ и посмотреть на

$$S = \int dt \left(-m + \left(1 - \frac{2\phi(t, \mathbf{r})}{d-3} \right) \frac{m\mathbf{v}^2}{2} - m\phi(t, \mathbf{r}) \right)$$

On other formulas

Об уравнениях в d -мерном пространстве по Лашевичу

$$h'_{\mu\nu} = h_{\mu\nu} + \xi_{\bar{\mu},\nu} + \xi_{\bar{\nu},\mu}$$

(?? ну калибровка, и что??)

Наложим d условий:

$$\psi_{\bar{\nu},\mu} = 0, \quad \psi_{\mu\nu} := h_{\mu\nu} - \frac{\eta_{\mu\nu}}{2}h.$$

Обратно,

$$h_{\mu\nu}(\psi) = \psi_{\mu\nu} - \frac{\eta_{\mu\nu}}{d-2}\psi, \quad \psi = \psi_{\bar{\mu}}^{\bar{\mu}} = -\frac{d-2}{2}h.$$

On second order equations

Условимся, что буквы h, ψ без индексов теперь будут обозначать матрицы, отвечающие соответствующим тензорам, а следы будут обозначаться значком tr : $\text{tr } a = a_{\bar{\mu}}^{\bar{\mu}} = \eta^{\mu\nu}a_{\mu\nu}$, $(ab)_{\mu\nu} = a_{\bar{\mu}}\bar{\lambda}_{\nu}$ и т.п. Отсюда легко находим:

$$\begin{aligned}g^{\mu\nu} &= \eta^{\mu\nu} - h^{(1)\bar{\mu}\bar{\nu}} - h^{(2)\bar{\mu}\bar{\nu}} + (h^{(1)2})^{\bar{\mu}\bar{\nu}} + \dots \\ |g| &= 1 + \text{tr } h^{(1)} + \text{tr } h^{(2)} + \frac{1}{2} \left((\text{tr } h^{(1)})^2 - \text{tr } h^{(1)2} \right) + \dots\end{aligned}$$

В нашей задаче в силу калибровочных условий $h_{\alpha\nu}^{(1)} = 0$, $\text{tr } h^{(1)} = 0$. Отсюда получаем во втором порядке

$$\begin{aligned}g^{\alpha\beta} &= \eta^{\alpha\beta} - h^{(2)\bar{\alpha}\bar{\beta}} \\ g^{\alpha\bar{b}} &= -h^{(2)\bar{\alpha}\bar{b}} \\ g^{ab} &= \eta^{ab} - h^{(1)\bar{a}\bar{b}} - h^{(2)\bar{a}\bar{b}} + (h^{(1)2})^{\bar{a}\bar{b}} \\ |g| &= 1 + \text{tr } h^{(2)} - \frac{1}{2} \text{tr } h^{(1)2}.\end{aligned}$$

На поправки второго порядка наложим такое же калибровочное условие, как и на поправки первого порядка:

$$\psi_{,\nu}^{(2)\bar{\mu}\nu} = 0, \quad \psi_{\mu\nu}^{(2)} = h_{\mu\nu}^{(2)} - \frac{1}{2}\eta_{\mu\nu} \operatorname{tr} h^{(2)}$$

Единственные ненулевые кристоффеля равны:

$$\tilde{\Gamma}_{\alpha b}^c = \frac{1}{2} \left(\eta^{cd} - h^{(1)\bar{c}d} \right) h_{bd,\alpha}^{(1)}, \quad \tilde{\Gamma}_{ab}^\gamma = -\frac{1}{2} h_{ab}^{(1),\gamma}.$$

Тензор Риччи:

$$R_{\mu\nu} = -\frac{1}{2} \square h_{\mu\nu}^{(2)} + \tilde{R}_{\mu\nu},$$

где

$$\begin{aligned} \tilde{R}_{\alpha\beta} &= \frac{1}{2} \operatorname{tr} \left(h^{(1)} h_{,\alpha\beta}^{(1)} \right) + \frac{1}{4} \operatorname{tr} \left(h_{,\alpha}^{(1)} h_{,\beta}^{(1)} \right), \quad \tilde{R}_{\alpha b} = 0, \quad \tilde{R}_{ab} = \frac{1}{2} \left(h^{(1),\bar{\alpha}} h^{(1),\alpha} \right)_{ab}, \\ \tilde{R} &= \frac{3}{4} \operatorname{tr} \left(h^{(1),\bar{\alpha}} h^{(1),\alpha} \right). \end{aligned}$$

$$\begin{aligned} t^{\alpha\beta} &= \frac{1}{32\pi G} \left(\operatorname{tr} h^{,\alpha} h^{,\beta} - \frac{1}{4} \eta^{\alpha\beta} \square \operatorname{tr} h^2 \right), \\ t^{\alpha b} &= 0, \\ t^{ab} &= \frac{1}{32\pi G} \left(\square (h^2)^{ab} - \frac{1}{4} \eta^{ab} \square \operatorname{tr} h^2 \right). \end{aligned}$$

Для распространяющейся вправо волны в лоренц-инвариантном виде можно написать

$$t^{\mu\nu} = \frac{1}{32\pi G} \operatorname{tr} h^{,\mu} h^{,\nu} \equiv \frac{1}{32\pi G} h_{\kappa}^{\lambda,\mu} h_{\lambda}^{\kappa,\nu}.$$

2.3.2 On main math for GW

Main definitions and notation

$$a_{(\mu_1 \dots \mu_n)} := \frac{1}{n!} \sum_{\sigma \in S^n} a_{\mu_{\sigma_1} \dots \mu_{\sigma_n}}$$

2.4 On detection of GW

2.4.1 Обзор экспериментальных результатов

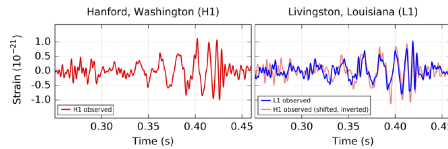
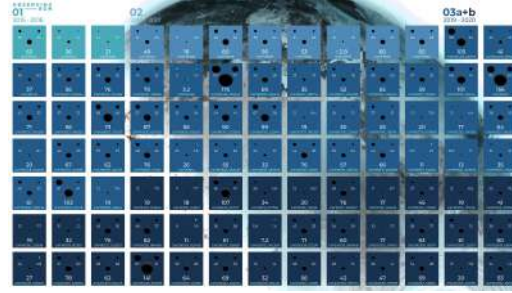
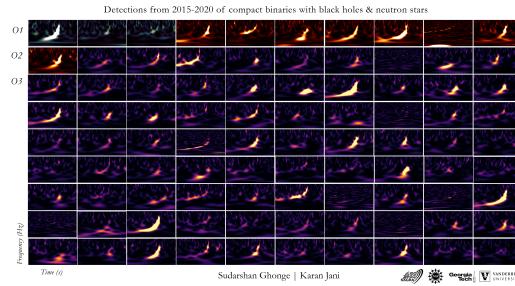
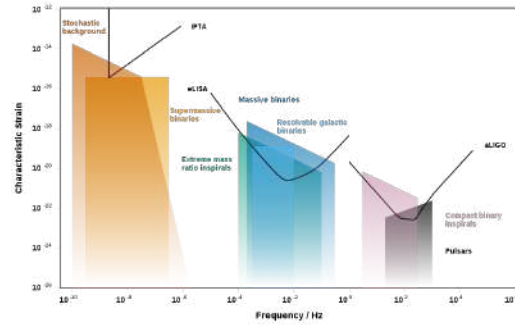


Figure 1



2.4.2 О детектировании в интерферометре Майкельсона

Основные свойства



Основные формулы

Типично:

$$h \sim 10^{-23}$$

Смещение 1 км за одно прохождение гравитационной волны:

$$\sim 10^{-21} \text{ m.}$$

Проблема, что размеры атома в огромное число раз меньше, атом $\sim 10^{-10} \text{ m.}$

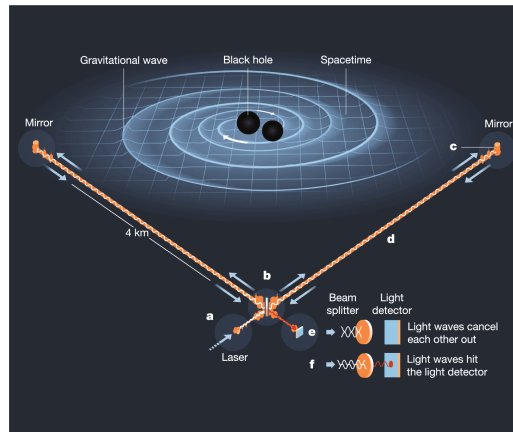
$$\int_0^{\tau_{\text{out}}} dt = \frac{1}{c} \int_0^L \sqrt{1 + h_{11}} dx \approx \frac{1}{c} \int_0^L \left(1 + \frac{1}{2} h_{11} (2\pi f t - \vec{k} \cdot \vec{x}) \right) dx$$

travel time difference between the two interferometer arms is

$$\Delta\tau(t) = h(t) \frac{2L}{c} = h(t) \tau_{\text{rt}0}$$

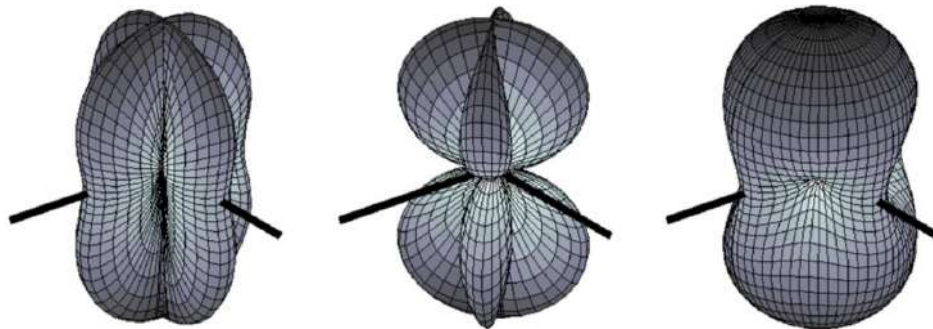
or that the optical phase difference between the two arms is

$$\Delta\phi(t) = h(t) \tau_{\text{rt}0} \frac{2\pi c}{\lambda}$$



$$\Delta\phi(t) = h(t)\tau_{\text{rt0}} \frac{2\pi c}{\lambda} \left(\frac{1}{2} (1 + \cos^2 \Theta) \cos 2\Phi \cos 2\Psi - \cos \Theta \sin 2\Phi \sin 2\Psi \right)$$

where Θ and Φ are the polar and azimuthal angles, and Ψ is the polarization angle of the wave. Pattern with $\times$ polarisation, $+$ polarisation and averaged:



This pattern is remarkably isotropic, with a very large range of angles within a factor of two of the maximum sensitivity.

(?? еще не полностью понял эту формулу, тут укажу, почему это так? в теории - большое объяснение)

О материалах для детекторов

(немного скучная тема, но многие про это прямо парятся, так что когда-то тоже напишу это. Слушал про это.)

Другие формулы и основные принципиальные вопросы

Сколько раз фотон пройдет, пока гравитационная волна движется перпендикулярно интерферометру?

(???? так и не понял, подумаю - напишу. оценю для типичных волн! там известная из графиков типичная частота грав волн. как бы так ни было, что он лишь пару раз пройдет, хотя экспериментаторы говорят, что проходит сотни раз. все еще вопрос)

2.4.3 О формализме необходимой математике для описания шумов (!!???)

(ответ про то, как понимать формулы ниже?)

Parameters, FDT, (???)

(выписка из статфиза, очень важная!)

2.4.4 Обзор других способов детектирования

(другие способы мало кто рассматривает, так что я не ставлю запись этого раздела в большой приоритет)

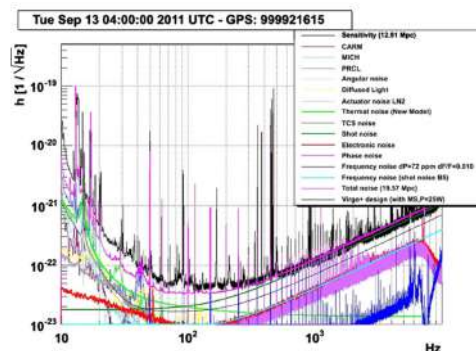
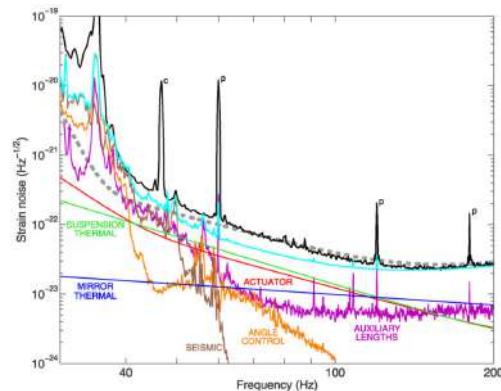
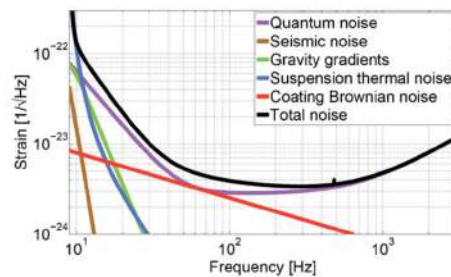
2.5 On noises in interferometric GW detectors and their reduction (!!??)

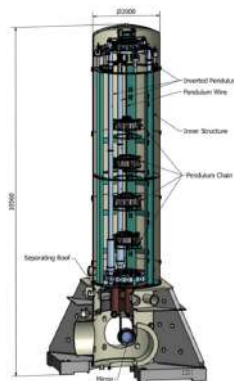
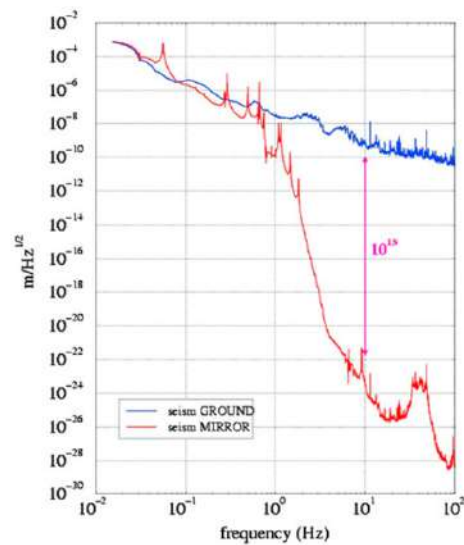
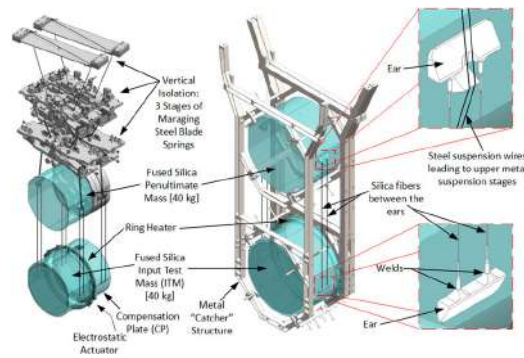
(постараюсь выписать всю суть нормально!)

(!!! очень важно разобраться, откуда по сути шумы берутся!?!?!?!?)

2.5.1 Overview of noises

Main properties of noises





Quantitative explanation of noises in an interferometer

(!! here is a very important general explanation, why noises are as so???)

(???? how this curved has been measured????? a big question!!! I'll ask one who knows.)

Thermal noise main concepts and formulas (???)

(here - by Saulson)

$$F_{\text{ext}} = Zv$$

$$v = YF_{\text{ext}}$$

$$F_{\text{therm}}^2(f) = 4k_B T \operatorname{Re}(Z(f))$$

Equivalently, the fluctuating motion of the system is characterized by a power spectrum of

$$x_{\text{therm}}^2(f) = \frac{k_B T}{\pi^2 f^2} \operatorname{Re}(Y(f))$$

In terms of this model of dissipation, we can rewrite the power spectrum of thermal noise of a test mass as

$$x^2(f) = \frac{4k_B T k \phi}{2\pi f [(k - m(2\pi f)^2)^2 + k^2 \phi^2]}$$

This discussion of thermal noise from internal friction is necessary for understanding why a pendulum suspension has been the universal choice for all gravitational wave detectors built to date. A pendulum has a very special property: the “spring” that provides most of the restoring force on the mass is the gravitational force (through the tension in the wire).

Only a very small fraction come from elasticity in flexing pendulum wire. The gravitational/tensional spring constant is dissipationless; only the small extra spring constant from the flexing wire contributes its dissipation. The combination of the two springs is one effective spring whose loss fraction ϕ_{pend} is given by

$$\phi_{\text{pend}} = \phi_{\text{wire}} \frac{k_{\text{el}}}{k_{\text{el}} + k_{\text{grav}}}$$

Типичная работа про это Levin-а. (абзац про суть ее)

И далее работа Lui and Thorn (абзац про суть ее)

What is quantum noise and what are main its formulas

(из самого популярного на ютбуе видео идеи, пропишу потом их лучше, тут совсем не очевидные идеи)

foreign so the way that gravitational wave detectors detect gravitational waves is by measuring the light travel time between a laser and a mirror and the mirror is usually some kilometers away so there's two things we have to do to make this gravitational wave detector work at the level of sensitivity that's needed and let me say what is that level of sensitivity what we're trying to do is we're measuring the displacements of mirrors that are kilometers apart and those displacements are at the level of 10 to the minus 18 meters so this is not terribly easy to do and so one piece of the of the enormous Technologies involved in ligo are making the mirror very still so you know shielding it from all the all external forces like seismic noise and thermal noise and another piece of the work that we must do is to make a measurement of that of the position of the mirror at that level of sensitivity the measurement tool in our case is the laser light itself it is the laser light and the wavelength of light that acts as the meter stick by which we measure these small displacement so the way that the ligo measurement works is we shine laser light in an into an interferometer and so the the interferometer basically works by splitting the laser light on two orthogonal parts two paths that are 90 degrees to each other the light travels for four kilometers and it returns to the beam splitter and it interferes there and we measure that interference pattern and we can arrange the path lengths to such that the interference pattern can be completely constructive or completely destructive so that's the principle of the of the measurement now one of the big limits to how well you can do comes to us from quantum mechanics it you know we know that light is quantized and the quantization of light puts a limit on how precisely we can measure the mirror position so one easy way this is

called the shot noise limit and what it one easy way to to think about it is it comes about because of the Heisenberg uncertainty principle so we know from the Heisenberg uncertainty principle that we can't measure two complementary variables with with infinite precision and so the way that the quantization of light interferes with our measurement is simply that our measure our signal is is proportional to the number of photons to the amount of light that we have hitting the mirror so the more photons we we can count the more uh signal we have now the noise the quantum noise on the light it's a poisson distributed noise source and so the noise goes as the square root of the number of photons or the square root of the power so you can immediately see that our signal increases with power our noise increases with square root of power so the signal to noise ratio improves with increasing the laser power and that's a trick that is commonly played in in Optical measurements where you try to use more laser power to make a better measurement now turns out that there is another price to be paid when you do that in a gravitational wave detector now for the the mirror to to displace under the influence of a gravitational wave we suspend it as a pendulum so it's a quasi-free particle so now you can think about about it this way we have a mirror that's hanging from a pendulum and its natural frequency is about one Hertz so if you just pull the mirror and let it swing it would swing at one Hertz so at frequencies well above one Hertz this is a free quasi-free particle it will respond freely to the forces uh uh that that push on it now it turns out of course that these these very same photons that have their fluctuating Quantum noise that gives us the square root of of power relationship these very same photons also carry momentum and now that momentum when the photon hits the mirror it exchanges momentum with the mirror and it kicks the mirror so in ligo we have this very incredible and rather fundamental Quantum noise limit which is that we would like to use more laser power to make a better measurement to get to do better than the shot noise limit but when we use more laser power that are the the momentum of the photons kicks the mirror so the very Act of making the measurement is disturbing the measurement and this is sometimes called the quantum measurement problem or the quantum noise problem and it turns out that in our field and it we also have something we call the standard Quantum limit which is where these two competing effects equate so it's a rather fundamental question actually you could ask about more or

less any measurement which is if I use light to measure the position of a particle how well can I do and the answer is it's limited by the Heisenberg uncertainty principle for this exact reason if you use more light to make a stronger measurement you get a stronger back action which is that your measurement kicks the particle and so we in our Laboratories also do research on that question of of how do we make detections that can do better than this standard Quantum limit or or exceed it and to do that there you need techniques that always always honor the Heisenberg uncertainty principle because that is is is a you know it's a fundamental law of nature but we can manipulate it so we like to think of of this as as we we get to to play with the Heisenberg uncertainty principle and our research allows us to do that so let me now talk about one piece of this Quantum noise problem which was the shot noise problem where does the short noise come from and how can we do better so the way let's think of a very simple thought experiment the thought experiment is that I have a simple Michaelson in a parameter that some of you will be familiar from with the Michaelson Morley experiment it's a device where you start with a laser beam you split the laser beam in exactly a half on a beam splitter half the light gets reflected by the bean splitter half the light gets transmitted by the beam splitter and then you have two mirrors that reflect the light back to the beam splitter and by arranging the path length between the mirrors and the beam splitter of course you can make perfect destructive or constructive interference at the output now suppose for a moment that I arrange these path lengths so I make perfect destructive interference at the output right that means in the ideal case no light goes to the output and by conservation of energy all the light has to go back to the laser so now let me let me frame the the the the thought experiment

question where does the noise that comes with the laser go now of course if the noise is simply attached to the laser light then it should all go right back to the laser and if I now put a detector at the output Port that is the other side of the beam splitter not the the laser side I should measure a perfect nothing right the truth is we don't it turns out that in the quantum mechanical formulation of this problem the interferometer is surrounded by all the modes of the electromagnetic vacuum and where do these come from these come simply from from uh you know spontaneous creation and annihilation of virtual particles so the way that I like to think of the quantum mechanical world is no matter where we are we are surrounded by this popcorn of vacuum states which is just part particle virtual particles being being created and annihilated so this output Port of our beam splitter is a natural place where these vacuum fluctuations enter the the interferometer and by the symmetry of the beam splitter the vacuum fluctuations come right back

out so now if I make this measurement even this idealized measurement where all my laser noise went back to the laser when I measure any signal coming out due to the relative displacements of the mirrors that signal has to beat with this vacuum fluctuations that have also traveled in the interferometer and that's where this shot noise comes from the shot noise comes from the vacuum fluctuation superposing on the light the signal carry light coming out of the interferometer this was an idea that was introduced in the very early 1980s by Carl Caves and collaborators and it's a very important distinction to understand if we want to now do better because if you believe all the noise came with the laser light then you'd spend all your time making a better and better laser but that noise all goes back to the laser and doesn't really help your measurement if you're truly limited by this Quantum noise so what we have done in our Laboratories and we're certainly not the first to to to to do this as we have created uh what are called squeezed states of light so let's talk about squeezed states what are these so let's start with an N with an a classical state of light the classical state of light simply says that you have some light wave it's it has an amplitude and a phase as any wave does now in quantum mechanics we also have to impose the Heisenberg uncertainty principle and what the Heisenberg uncertainty principle says about light is that we can't know the amplitude and phase of the light perfectly well at any given time in fact they are uncertain the product of their uncertainties is is given by the Heisenberg limit and so what we have done in our Laboratories is to create a squeezed state of light where we take the uncertainty in the amplitude or the phase and we make it smaller than the shot noise limit but of course that would violate Heisenberg if that was all we did we have to at the same time as we make the the uncertainty in the amplitude smaller we must make the uncertainty in the phase bigger so that the product is still the same as what Heisenberg would require the product still gives us $\hbar$ and that's what we do we have experiments in our Laboratories where we can create these squeezed states of light using non-linear interactions in optical materials and we have actually shown that when you inject these squeezed states of light into the LIGO interferometers you can make the you can improve the signal to noise ratio to better than the shot noise limit and that's very exciting because

now what we can do is we can use these specially Quantum engineered states of light to make a measurement that is that does better than the shot noise limit squeezed states of light were proposed in the early 1980s by Caves and collaborators the very first squeezed states were made in 1985 by Slusher at all and it has taken us almost 20 years more to make them work in gravitational wave detectors in a useful way and the reason for this is that squeezed states are actually very fragile so when you create a squeezed state first of all to create a squeezed state you need you need to have these non-linear interactions which are usually not very strong so you don't create a very strongly squeezed state to begin with but then the squeezed state very easily wants to degrade back to the vacuum state so every time in your optical system you encounter an optical loss your squeezed state mixes with the ordinary vacuum state and wants

to degrade back to the ordinary vacuum state so once you generate the squeeze State delivering it with very little as little loss as possible to the interferometer is one of the big challenges in in these experiments is how do you create a squeeze tape that's very very squeezed most squeezing we would like and then how do you maintain that degree of squeezing as the squeeze State travels in your Optical system to the interferometer and back to the detection Port so that's one of the big challenges in in squeeze State generation is making enough squeezing and then preserving that that squeeze state in our experiments another thing that's very difficult to do and something that's actually a topic of active research in our group right now is how to make a squeeze state where at different frequencies you're squeezing either the amplitude or the phase degrees of freedom now why would we like to do that well it turns out that in the in gravitational wave detectors we are at high frequencies we're limited by the shot noise on the phase fluctuations but at low frequencies are where our mirror responds more easily to external forces we're limited by that that that momentum transfer the radiation pressure noise and so we would like to be able to make a squeeze state where at high frequencies we're squeezing along one axis and at low frequencies we're squeezing along the opposite axis and at intermediate frequencies we have you know so we'd like to have a continuously rotating squeeze State and that's another thing that we are working on in our Laboratories right now and to do that we are using these devices that are called Optical cavities and they are used to continuously rotate the squeeze state foreign

On Thermal noise due to dielectric optical coatings

У нас меняется просто поверхность линзы. (абзац про то, что физически происходит???)

$$S_x(f) = \frac{4k_B T}{4\pi f^2 \Re\{Z(f)\}} \sim T$$

(?? крч, такую формулу пишут!)

$$S_x(f) = \frac{2k_B T}{\pi^{3/2} f} \frac{1 - \sigma^2}{wY} \left\{ \phi_{\text{substrate}} + \frac{2}{\sqrt{\pi}} \frac{(1 - 2\sigma)}{(1 - \sigma)} \frac{d}{w} \phi_{\parallel} \right\}$$

(абзац про вывод словами в целом, что делать и почему? пока не до этого.)

How to reduce it?

On Substrate noises main concepts and formulas

$$S^{\delta z}(f) = \frac{a\beta^2 k_B T^2}{2\pi\kappa} \times (e^{it} E_1(it) + e^{-it} E_1(-it)).$$

(абзац про то, что физически происходит???)

Coating thermal noise amplitude spectral density—from now on referred to as coating thermal noise (CTN)—is given by [30]

$$x(f) = \sqrt{\frac{2k_B T}{\pi^2 f} \frac{1}{w^2} \frac{1 - \sigma_{\text{sub}} - 2\sigma_{\text{sub}}^2}{Y_{\text{sub}}} \sum_j b_j d_j \phi_j}.$$

(article by Jessica Steinlechner and Iain W. Martin, they got it from W. Yam, S. Gras, and M. Evans, Multimaterial coatings with reduced thermal noise 2015)

The lower coating layers contribute more to the total CTN than the upper layers

$$S_x(f) = \frac{2k_B T}{\pi^2 f Y_{\text{sub}}} \frac{d}{w^2} \left(\frac{Y_{\text{coat}}}{Y_{\text{sub}}} \phi_{\parallel} + \frac{Y_{\text{sub}}}{Y_{\text{coat}}} \phi_{\perp} \right)$$

(from J. Steinlechner)

(?? а какие выводы - я хз примерно, там предлагается много использовать линий. хз. типа проблема в том, что у нас расстояние меняется, так что не угадать, где будет этот конец волны.)

How to reduce it?

Seismic noise main concepts and formulas

(скорее всего в теории про это большое обсуждение будет.)

$$x(f) = 10^{-9} \text{ m}/\sqrt{\text{Hz}}$$

at frequencies above 10 Hz.

(абзац про то, что физически происходит????)

How to reduce it?

Suspension thermal noise main concepts and formulas

$$Y(f) = Z(f)^{-1} = \frac{v(f)}{F_{\text{thermal}}} = \frac{2\pi i f x(f)}{F_{\text{thermal}}}$$

(абзац про то, что физически происходит????)

How to reduce it?

Mechanical Dissipation main concepts and formulas

The mechanical dissipation factor can be defined as:

$$\phi(f_0) = \frac{E_{\text{lost per cycle}}}{2\pi E_{\text{total}}} \equiv \frac{\Delta f}{f_0} = \frac{1}{Q}$$

(абзац про то, что физически происходит????) Думаю, это от колебания зеркала, от его трения. ХЗ пока.

How to reduce it?

Residual gas noises

$$S_L(f) = \frac{4\rho(2\pi\alpha)^2}{v_0} \int_0^L \frac{\exp[-2\pi f w(z)/v_0]}{w(z)} dz$$

Here, ρ is the gas number density for each molecular species, α is the polarizability, $v_0 = \sqrt{\frac{2k_b T}{m}}$ is the molecule's most probable velocity, L is the arm-length and $w(z)$ is the Gaussian beam radius.

(абзац про то, что физически происходит????)

How to reduce it?

2.5.2 О материалах и покрытиях для детекторов (!!??)

(просто список результатов. очень коротко, потому что это совсем мало к грав волнам относится, но тем не менее, многие люди над этим работают.)

3 On extra topics in GW

3.1 On information from GW

(все это потом напишу)

3.1.1 О волнах как способе проверить обобщения ОТО

3.1.2 Об информации о Вселенной на ранних этапах

(выпишу когда-то, уже создан раздел)

3.2 О typical models with GW

(про все это большие статьи, тут только самую суть пишу их, как в итоге это работает и какие оценки этих эффектов? пока не актуально, не пишу об этом!!)

3.2.1 On interaction with EM field

Main formulas

3.2.2 On interaction with light near star

Main formulas

3.2.3 On interaction with matter

Main formulas

3.2.4 On other possible interactions

(тут скорее всего многое будет, но большинство - отсылками, оценками и словами.)

3.2.5 Overview of models of T. Mieling

(у него очень хорошие статьи, которые стоят отдельного указания. потом в каталоге задач многие задачи будут - как раз части его статей. действительно, очень интересно изучать это, когда такого качества наработки сделаны!!! это не актуально, но очень интересно!!)

3.3 О других гравитационных волнах

3.3.1 О сильных волнах (?)

О типичной сильной волне Бонди-Роберсона

(тут много вопросов пока что, многое не понимаю, но они в теории, тут укажу совсем мало информации. не такая актуальная тема.)

Рассмотрим решение уравнений гравитационного поля в вакууме, которое зависит лишь от одной координаты, которую обозначим x^- :

$$g_{\mu\nu} = g_{\mu\nu}(x^-) \Leftrightarrow \partial_i g_{\mu\nu} = 0 \quad (i \neq -).$$

Пусть $\det(g_{ij}) = 0$ (??? так и не понял, почему нам именно такое условие нужно?). (??? из каких-то калибровок, про которые в теории подумаю) Метрику можно записать в виде

$$ds^2 = dx^- dx^+ + g_{ab}(dx^a + g^a dx^+)(dx^b + g^b dx^+).$$

Тогда $(-)$ -компонента уравнения Эйнштейна имеет вид

$$R_{--} = -\frac{1}{2}\dot{\kappa}_a^a - \frac{1}{4}\kappa_a^b \kappa_b^a = -\frac{1}{2}g^{ab}\ddot{g}_{ab} + \frac{1}{4}g^{ab}\dot{g}_{bc}g^{cd}\dot{g}_{da} = 0.$$

Это - единственное уравнение на компоненты g_{ab} . Поэтому $d(d-3)/2$ из них можно выбрать произвольно, и только на оставшуюся компоненту писать уравнение. Удобно это сделать в виде:

$$g_{ab} = -\chi^2 \gamma_{ab}, \quad \det(\gamma_{ab}) = 1$$

$$\ddot{\chi} = \frac{1}{4(d-2)}\dot{\gamma}^{ab}\dot{\gamma}_{ab}\chi.$$

$$x^0 := \frac{x^+ + x^-}{2}, \quad x^1 := \frac{x^+ - x^-}{2}.$$

Рассмотрим опять слабую волну $g_{ab} = \eta_{ab} + h_{ab}$ тогда $\chi(x^-) \simeq 1$ и

$$\dot{\chi} \simeq -\frac{1}{4(d-2)} \int_0^t dt (\dot{\gamma}_{ab})^2.$$

Величина под интегралом, усредненная по времени, есть, в сущности, интенсивность волны. Если эта интенсивность не меняется, χ по мере движения волны буде убывать по закону $\chi \simeq 1 - \text{const} \cdot t^2$. Хотя этот закон будет меняться, наступит момент t_0 , когда $\chi(t_0) = 0$. В окрестности этой области имеем

$$ds^2 = dx^- dx^+ - A^2 (x^- - t_0)^2 \delta_{ab} dx^a dx^b.$$

Казалось бы, мы имеем особенность (я остановился на том, что это именно для этой волны, а не вообще для каждой!). Однако замена переменных

$$x^- = y^- + t_0, \quad x^a = \frac{y^a}{Ay^-}, \quad x^+ = y^+ - \frac{\delta_{ab}y^a y^b}{y^-}$$

разрешает эту особенность:

$$ds^2 = dy^- dy^+ - \delta_{ab} dy^a dy^b.$$

Поэтому на самом деле распространение гравитационной волны не приводит к особенности метрики пространства-времени.

О других сильных волнах (??)

(там многое есть, вот и напишу.)

3.3.2 О других волнах

(пока все остальное просто здесь и все.)

О многомерных волнах

(вот и напишу их суть)

О волнах в модифицированных теориях гравитации (??)

(вот и напишу их суть)

3.4 О другом про гравитационные волны

(указание на разное совсем не часто нужное, просто очень много тут всего всякого)

Conclusions and Future Directions by Sathyaprakash

The development of gravitational wave detectors to their present capability has required patience, ingenuity, and dedication by an entire generation of experimental physicists. No less dedication and vision have been required by scientific funding organizations of a half-dozen nations and two major space agencies. The initial data runs of the LIGO and VIRGO detectors at their first sensitivity goals (bursts with amplitudes of 10^{-21}) have not so far yielded any detections, but this is certainly not surprising. The operation of these detectors at this sensitivity level has demonstrated that the technology is understood, and the analysis of the data has provided important early experience and the opportunity to organize the efforts into the LSC and VIRGO collaborations. As the detectors are upgraded during the period 2008-2014, the first detection could occur at any time; if the advanced detectors do not make early detections, then there will inevitably be serious questions about general relativity. The field of gravitational wave detection has never before been at the point where it could test the fundamental theory.

Once the first detection is made, there will be increasing emphasis on the fundamental physics and astrophysics that will follow from further detections. As we have discussed in this review, one can look forward soon thereafter to a detailed comparison of black hole mergers with theory, to exploring the relationship between compact-object mergers and gamma-ray bursts, to using this association to make a precise and calibration-free measurement of the

Hubble constant, and to population studies of neutron stars and black holes. In this early phase of gravitational wave astronomy there are very exciting (but less certain) potential observations: an unexpectedly strong cosmological background, which would revolutionize early-universe physics; the detection of mass asymmetry or normal-mode oscillations of rotating neutron stars, either of which would for the first time probe the interior physics of these complex objects and would help unravel the mystery of the pulsar phenomenon; the first studies of the interior core dynamics of a supernova, if one happens to occur nearby; the detection of populations of compact dark objects, like cosmic strings or small black holes; the discovery of exceptionally-massive black holes, around $100M_{\odot}$; or the association of gravitational wave events with transient phenomena other than gamma-ray bursts, such as transient radio bursts.

When LISA is launched, the physics and astrophysics consequences become even richer. LISA will study black hole mergers during the early phases of galaxy formation, exploring the mysterious link between the two. It will map in detail black hole spacetimes and verify the black-hole uniqueness and area theorems of general relativity. It is likely to map the history of the expansion of the universe through measuring the distances to massive black hole mergers, and from that look for evidence that the dark energy has been evolving with time. It will discover every short-period binary system in our galaxy, calibrating white-dwarf masses, mapping their mass distribution, determining the population of neutron stars in binaries. As with ground-based detectors, LISA might make other discoveries that are harder to predict, such as a cosmological background, cosmic strings, intermediate-mass black holes, even g -mode oscillations of the sun. LISA has enough sensitivity to be able to make discoveries even of sources for which there are no signal models to aid data analysis. And if LISA does not see its verification binary sources, that will be fatal for general relativity.

Gravitational wave detections may also come from other technologies, such as pulsar timing searches or observations of the cosmic microwave background. The spectrum of gravitational waves is enormous, and present technologies can explore only a tiny fraction of it. Beyond the LISA timeframe, say after 2020, new technologies may come into the field and make possible detectors that extend the ground-based detection band to lower frequencies (such as the Einstein Telescope project), observing in space in the 0.1 Hz band, going up to megahertz frequencies.

The present review has attempted to give a good overview of the science that can be done with gravitational waves, but it is certainly not complete. Future revisions are planned to add more on LISA, more on data analysis issues, and considerably more on detectors that might go beyond Advanced LIGO and VIRGO. This is a field that is developing rapidly. For example, the launch of LISA is 10 years away (at the time of writing, 2008), but already the scientific literature contains many hundreds of refereed papers on LISA science and technology, and every second year there is a major international symposium on the subject. This is probably unprecedented among space missions. Living Reviews in Relativity is planning to release a suite of articles in the near future on LISA, which will cover cosmology, tests of general relativity, galactic astrophysics, black hole astrophysics, and observations of low-frequency gravitational wave sources with LISA. Until the next revision, readers interested in keeping up with the field should also consult the proceedings of the regular conferences on gravitational waves: the Amaldi meetings, GWDAW (Gravitational Wave Data Analysis Workshops), GWADW (Gravitational Wave Advanced Detectors Workshops), and the LISA Symposium.

3.5 О других способах детектирования волн

(потом напишу что-то сюда, пока не знаю, что еще там есть за методы, да и не нужно это знать.)

3.5.1 Об очень малоперспективных способах

О детектировании оптоволоконном

Part II

Properties of Gravitational Waves

(позже опишу их, пока в записи про ОТО про них.)

4 Типичные гравитационные волны

4.1 Свойства гравитационных волн

4.1.1 Обоснование существования волн (???)

(с этого все и начинается)

4.1.2 Поляризация

Обсудим подробно поляризацию гравитационных волн.

4.1.3 Основные свойства гравитационных волн

Дипольная компонента в излучении всегда отсутствует

(еще раз про это, по идее должно это быть очевидно, но нужно еще додумать это, тут четко напишу, почему так.)

4.1.4 Second-order equations

Here I'll discuss in details second order equations for GW.

Теория по Лашкевичу

4.1.5 Другие свойства гравитационных волн

(есть много очень интересных свойств, которые не так актуальны или просто я не готов про это писать, так что пока они тут побудут.)

Длина гравитационной волны

хз, тоже интересно.

Энергия гравитационной волны

(она вообще малая или большая?)

Конечность размеров гравитационной волны

я вот хз, по идее прямо всё пространство-время не может быть же ей занято?
так а где там учтено, что вне какого-то скажем шара волны нет?
имеется в виду в плоскости, в которой метрика деформируется.

Почему калибровками мы все-таки можем остаться в нашей обычной системе отсчета?

Мы говорим, что делаем калибровки, так что в какой-то системе будет гравитационная волна такого вида. Круто, замечательно, однако что по поводу нашей системе отсчета, где наш детектор, разве в ней тоже будет такой вид иметь гравитационная волна?

(оставил пока вопрос, потом подумаю. мб да, ничего тут нет особенного.)

Особенности гравитационных волн

(разные вопросы про волны пока тут соберу)

Временная компонента волны

Может ли волна иметь временную компоненту?

а то что-то только пространственные не так интересно.

я не могу объяснить, почему не может?

Устраняемая сингулярность за волной (?????)

(Ландау писал об этом в главе про сильные волны, спишу, интересный вопрос. потом разберусь и дам наглядную интерпретацию, пока что даже наглядно не представляю и не актуально это.)

Определение времени и места источника излучения волн

как понять, что данная волна, зафиксированная на Земле была излучена столько-то лет назад такими-то черными дырами?

(не знаю пока)

4.1.6 Построение теории гравитационных волн с нуля

(потом раскрою, еще я не профессионал, а вопрос этот самый профессиональный)

4.2 Излучение гравитационных волн

Обсудим подробно излучение волн.

(подробно и разное об этом. странно, что еще не написано, скоро напишу)

4.2.1 Типичное излучение гравитационных волн

Теория

(проходил, напишу, просто не до этого пока)

4.2.2 Волны от типичных объектов

(давно идея прорешать эти задачи, когда-то займусь.)

Волны от взмаха рукой

Волны от проезжающей машины

Волны от кружения пары людей

4.2.3 Другое про излучение гравитационных волн

(тоже многое указать можно, пока просто не особо думал.)

5 Sathyaprakash's GW article

(крутая статья, но не читал, и не до этого, вряд ли польза есть в чтении ее так вот просто. можно забить и начать исследовать и по потребности брать нужные темы!)

5.1 A New Window onto the Universe

The last six decades have witnessed a great revolution in astronomy, driven by improvements in observing capabilities across the electromagnetic spectrum: very large optical telescopes, radio antennas and arrays, a host of satellites to explore the infrared, X-ray, and gamma-ray parts of the spectrum, and the development of key new technologies (CCDs, adaptive optics). Each new window of observation has brought new surprises that have dramatically changed our understanding of the universe. These serendipitous discoveries have included:

- the relic cosmic microwave background radiation (Penzias and Wilson [287]), which has become our primary tool for exploring the Big Bang;
- the fact that quasi-stellar objects are at cosmological distances (Maarten Schmidt [323]), which has developed into the understanding that they are powered by supermassive black holes;
- pulsars (Hewish and Bell [189]), which opened up the study of neutron stars and illuminated one endpoint for stellar evolution;
- X-ray binary systems (Giacconi and collaborators [326]), which now enable us to make detailed studies of black holes and neutron stars;
- gamma-ray bursts coming from immense distances (Klebesadel et al. [216]), which are not fully explained even today;
- the fact that the expansion of the universe is accelerating (two teams [313, 288]), which has led to the hunt for the nature of dark energy.

None of these discoveries was anticipated by the observing team, and in many cases the instruments were built to observe completely different phenomena.

Within a few years another new window on the universe will open up, with the first direct detection of gravitational waves. There is keen interest in observing gravitational waves directly, in order to test Einstein's theory of general relativity and to observe some of the most exotic objects in nature, particularly black holes. But, in addition, the potential of gravitational wave observations to produce more surprises is very high

The gravitational wave spectrum is completely distinct from, and complementary to, the electromagnetic spectrum. The primary emitters of electromagnetic radiation are charged elementary particles, mainly electrons; because of overall charge neutrality, electromagnetic radiation is typically emitted in small regions, with short wavelengths, and conveys direct information

about the physical conditions of small portions of the astronomical sources. By contrast, gravitational waves are emitted by the cumulative mass and momentum of entire systems, so they have long wavelengths and convey direct information about large-scale regions. Electromagnetic waves couple strongly to charges and so are easy to detect but are also easily scattered or absorbed by material between us and the source; gravitational waves couple extremely weakly to matter, making them very hard to detect but also allowing them to travel to us substantially unaffected by intervening matter, even from the earliest moments of the Big Bang.

These contrasts, and the history of serendipitous discovery in astronomy, all suggest that electromagnetic observations may be poor predictors of the phenomena that gravitational wave detectors will eventually discover. Given that 96% of the mass-energy of the universe carries no charge, gravitational waves provide us with our first opportunity to observe directly a major part of the universe. It might turn out to be as complex and interesting as the charged minor component, the part that we call "normal" matter.

Several long-baseline interferometric gravitational-wave detectors planned over a decade ago [Laser Interferometer Gravitational-Wave Observatory (LIGO) [18], GEO [244], VIRGO [109] and TAMA [363]] have begun initial operations [3, 245, 19] with unprecedented sensitivity levels and wide bandwidths at acoustic frequencies (10 Hz – 10kHz) [197]. These large interferometers are superseding a world-wide network of narrow-band resonant bar antennas that operated for several decades at frequencies near 1kHz. Before 2020 the space-based LISA [71] gravitational wave detector may begin observations in the low-frequency band from 0.1mHz to 0.1 Hz. This suite of detectors can be expected to open up the gravitational wave window for astronomical exploration, and at the same time perform stringent tests of general relativity in its strong-field dynamic sector.

Gravitational wave antennas are essentially omni-directional, with linearly polarized quadrupolar antenna patterns that typically have a response better than 50% of its average over 75% of the sky [197]. Their nearly all-sky sensitivity is an important difference from pointed astronomical antennas and telescopes. Gravitational wave antennas operate as a network, with the aim of taking data continuously. Ground-based interferometers can at present (2008) survey a volume of order 10^4Mpc^3 for inspiraling compact star binaries - among the most promising sources for these detectors - and plan to enhance their range more than tenfold with two major upgrades (to enhanced and then advanced detectors) during the period 2009-2014. For the advanced detectors, there is great confidence that the resulting thousandfold volume increase will produce regular detections. It is this second phase of operation that will be more interesting from the astrophysical point of view, bringing us physical and astrophysical insights into populations of neutron star and black hole binaries, supernovae and formation of compact objects, populations of isolated compact objects in our galaxy, and potentially even completely unexpected systems. Following that, LISA's ability to survey the entire universe for black hole coalescences at milliHertz frequencies will extend gravitational wave astronomy into the cosmological arena.

However, the present initial phase of observation, or observations after the first enhancements, may very well produce the first detections. Potential sources include coalescences of binaries consisting of black holes at a distance of 100 – 200Mpc and spinning neutron stars in our galaxy with ellipticities greater than about 10^{-6} . Observations even at this initial level may, of course, also reveal new sources not observable in any other way. These initial detections, though not expected to be frequent, would be important from the fundamental physics point of view and could enable us to directly test general relativity in the strongly nonlinear regime.

Gravitational wave detectors register gravitational waves coherently by following the phase of the wave and not just measuring its intensity. Since the phase is determined by large-scale motions of matter inside the sources, much of the astrophysical information is extracted from the phase. This leads to different kinds of data analysis methods than one normally encounters

in astronomy, based on matched filtering and searches over large parameter spaces of potential signals. This style of data analysis requires the input of pre-calculated template signals, which means that gravitational wave detection depends more strongly than most other branches of astronomy on theoretical input. The better the input, the greater the range of the detectors.

The fact that detectors are omni-directional and detect coherently the phase of the incoming wave makes them in many ways more like microphones for sound than like conventional telescopes. The analogy with sound can be helpful, since microphones can be used to monitor environments for disturbances in any location, and since when we listen to sounds our brains do a form of matched filtering to allow us to interpret the sounds we want to understand against a background of noise. In a very real sense, gravitational wave detectors will be listening to the sounds of a restless universe. The gravitational wave "window" will actually be a listening post, a monitor for the most dramatic events that occur in the universe.

5.1.1 Birth of gravitational astronomy

Gravity is the dominant interaction in most astronomical systems. The big surprise of the last three decades of the 20th century was that relativistic gravitation is relevant in so many of these systems. Strong gravitational fields are Nature's most efficient converters of mass into energy. Examples where strong-field relativistic gravity is important include the following:

- neutron stars, the residue of supernova explosions, represent up to 0.1% (by number) of the entire stellar population of any galaxy;
- stellar-mass black holes power many binary X-ray sources and tend to concentrate near the centers of globular clusters;
- massive black holes in the range $10^6 - 10^9 M_\odot$ seem almost ubiquitous in galaxies that have central bulges, and power active galaxies, quasars, and giant radio jets;
- and, of course, the Big Bang is the only naked singularity we expect to be able to see.

Most of these systems are either dynamical or were formed in catastrophic events; many are or were, therefore, strong sources of gravitational radiation. As the 21st century opens, we are on the threshold of using this radiation to gain a new perspective on the observable universe.

The theory of gravitational radiation already makes an important contribution to the understanding of a number of astronomical systems, such as neutron star binaries, cataclysmic variables, young neutron stars, low-mass X-ray binaries, and even the anisotropy of the microwave background radiation. As the understanding of relativistic phenomena improves, it can be expected that gravitational radiation will play a crucial role as a theoretical tool in modeling relativistic astrophysical systems.

5.1.2 What this review is about

The first three-quarters of the 20th century were required to place the mathematical theory of gravitational radiation on a sound footing. Many of the most fundamental constructs in general relativity, such as null infinity and the theory of conserved quantities, were developed at least in part to help solve the technical problems of gravitational radiation. We will not cover this history here, for which there are excellent reviews [259, 132]. There are still many open questions, since it is impossible to construct exact solutions for most interesting situations. For example, we still lack a full understanding of the two-body problem, and we will review the theoretical work on this problem below. But the fundamentals of the theory of gravitational radiation are no longer in doubt. Indeed, the observation of the orbital decay in the binary

pulsar PSR B1913+16 [388] has lent irrefutable support to the correctness of the theoretical foundations aimed at computing gravitational wave emission, in particular to the energy and angular momentum carried away by the radiation.

It is, therefore, to be expected that the evolution of astrophysical systems under the influence of strong tidal gravitational fields will be associated with the emission of gravitational waves. Consequently, these systems are of interest both to a physicist, whose aim is to understand fundamental interactions in nature, their inter-relationships and theories describing them, and to an astrophysicist, who wants to dig deeper into the environs of dense or nonlinearly gravitating systems in solving the mysteries associated with relativistic phenomena listed in Sections 6, 7 and 8. Indeed, some of the gravitational wave antennas that are being built are capable of observing systems to cosmological distances, and even to the edge of the universe. The new window, therefore, is also of interest to cosmologists.

This is a living review of the prospects that lie ahead for gravitational antennas to test the predictions of general relativity as a fundamental theory, for using relativistic gravitation as a means to understand highly energetic sources, for interpreting gravitational waves to uncover the (electromagnetically) dark universe, and ultimately for employing networks of gravitational wave detectors to observe the first fraction of a second of the evolution of the universe.

We begin in Section 2 with a brief review of the physical nature of gravitational waves, giving a heuristic derivation of the formulas involved in the calculation of the gravitational wave observables such as the amplitude, frequency and luminosity of gravitational waves. This is followed in Section 3 by a discussion of the astronomical sources of gravitational waves, their expected event rates, amplitudes, waveforms and spectra. In Section 4 we then give a detailed description of the existing and upcoming gravitational wave antennas and their sensitivity. Included in Section 4 are bar and interferometric antennas covering both ground and space-based experiments. Section 4 also compares the sensitivity of the antennas with the strengths of astronomical sources and expected signal-to-noise ratios (SNRs). We then turn in Section 5 to data analysis, which is a central component of gravitational wave astronomy, focusing on those aspects of analysis that are crucial in gleaning physical, astrophysical and cosmological information from gravity wave observations.

Sections 7-9 treat in some detail how gravitational wave observations will aid in a better understanding of nonlinear gravity and test some of its fundamental predictions. In Section 6 we review the physics implications of gravitational wave observations, including new tests of general relativity that can be performed via gravitational wave observations, how these observations may help in formulating and gaining insight into the two-body problem in general relativity, and how gravitational wave observations may help to probe the structure of the universe and the nature of dark energy. In Section 7 we look at the astronomical information returned by gravitational wave observations, and how these observations will affect our understanding of black holes, neutron stars, supernovae, and other relativistic phenomena. Section 8 is devoted to the cosmological implications of gravitational wave observations, including placing constraints on inflation, early phase transitions associated with spontaneous symmetry breaking, and the large-scale structure of the universe.

This review is by no means exhaustive. We plan to expand it to include other key topics in gravitational wave astronomy with subsequent revisions.

Unless otherwise specified we shall use a system of units in which $c = G = 1$, which means $1M_{\odot} \simeq 5 \times 10^{-6} \text{ s} \simeq 1.5 \text{ km}$, $1\text{Mpc} \simeq 10^{14} \text{ s}$. We shall assume a universe with cold dark-matter density of $\Omega_M = 0.3$, dark energy of $\Omega_{\Lambda} = 0.7$, and a Hubble constant of $H_0 = 70 \text{ km s}^{-1}\text{Mpc}^{-1}$.

5.2 Gravitational Wave Observables

To benefit from gravitational wave observations we must first understand what are the attributes of gravitational waves that we can observe. This section is devoted to a short discussion of the nature of gravitational radiation.

5.2.1 Gravitational field vs gravitational waves

Gravitational waves are propagating oscillations of the gravitational field, just as light and radio waves are propagating oscillations of the electromagnetic field. Whereas light and radio waves are emitted by accelerated electrically-charged particles, gravitational waves are emitted by accelerated masses. However, since there is only one sign of mass, gravitational waves never exist on their own: they are never more than a small part of the overall external gravitational field of the emitter. One may wonder, therefore, how it is possible to infer the presence of an astronomical body by the gravitational waves that it emits, when it is clearly not possible to sense its much larger stationary (essentially Newtonian) gravitational potential. There are, in fact, two reasons:

- In general relativity, the effects of both the stationary field and gravitational radiation are described by the tidal forces they produce on free test masses. In other words, single geodesics alone cannot detect gravity or gravitational radiation; we need at least a pair of geodesics. While the stationary tidal force due to the Newtonian potential ϕ of a self-gravitating source at a distance r falls off as $\nabla\nabla\phi \sim r^{-3}$, the tidal force due to the gravitational wave amplitude h that it emits at wavelength λ decreases as $\nabla\nabla h \sim r^{-1}\lambda^{-2}$. Therefore, the stationary coulomb gravitational potential is the dominant tidal force close to the gravitating body (in the near zone, where $r \leq \lambda$). However, in the far zone ($r \gg \lambda$) the tidal effect of the waves is much stronger.
- The stationary part of the tidal field is a DC effect, and simply adds to the stationary tidal forces of all other objects in the universe. It is not possible to discriminate one source from another. Gravitational waves carry time-dependent tidal forces, and so they can be discriminated from the stationary field if one knows what kind of time dependence to look for. Interferometers are ideal detectors in this respect because they sense only changes in the position of an interference fringe, which makes them insensitive to the DC part of the tidal field.

Because gravitational waves couple so weakly to our detectors, those astronomical sources that we can detect must be extremely luminous in gravitational radiation. Even at the distance of the Virgo cluster of galaxies, a detectable source could be as luminous as the full Moon, if only for a millisecond! Indeed, while radio astronomers deal with flux levels of Jy, mJy and even μ Jy, in the case of gravitational wave sources we encounter fluxes that are typically 10^{20} Jy or larger. Gravitational wave astronomy therefore is biased toward looking for highly energetic, even catastrophic, events.

Extracting useful physical, astrophysical and cosmological information from gravitational wave observations is made possible by measuring a number of gravitational wave attributes that are related to the properties of the source. In the rest of this section we discuss those attributes of gravitational radiation that can be measured via gravitational wave observations. In the process we will review the basic formulas used in computing the gravitational wave amplitude and luminosity of a source. These will then be used in Section 3 to make an order-of-magnitude estimate of the strength of astronomical sources of gravitational waves.

5.2.2 Gravitational wave polarizations

Because of the equivalence principle, single isolated particles cannot be used to measure gravitational waves: they fall freely in any gravitational field and experience no effects from the passage of the wave. Instead, one must look for inhomogeneities in the gravitational field, which are the tidal forces carried by the waves, and which can be measured only by comparing the positions or interactions of two or more particles.

In general relativity, gravitational radiation is represented by a second rank, symmetric tracefree tensor. In a general coordinate system, and in an arbitrary gauge (coordinate choice), this tensor has ten independent components. However, as in the electromagnetic case, gravitational radiation has only two independent states of polarization in Einstein's theory: the plus polarization and the cross polarization (the names being derived from the shape of the equivalent force fields that they produce). In contrast to electromagnetic waves, the angle between the two polarization states is $\pi/4$ rather than $\pi/2$. This is illustrated in Figure 1, where the response of a ring of free particles in the (x, y) plane to plus-polarized and cross-polarized gravitational waves traveling in the z -direction is shown. The effect of the waves is to cause a tidal deformation of the circular ring into an elliptical ring with the same area. This tidal deformation caused by passing gravitational waves is the basic principle behind the construction of gravitational wave antennas.

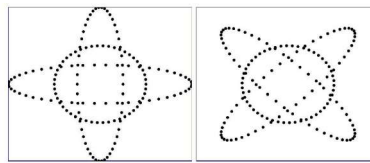


Figure 1: In Einstein's theory, gravitational waves have two independent polarizations. The effect on proper separations of particles in a circular ring in the (x, y) -plane due to a plus-polarized wave traveling in the z -direction is shown in (a) and due to a cross-polarized wave is shown in (b). The ring continuously gets deformed into one of the ellipses and back during the first half of a gravitational wave period and gets deformed into the other ellipse and back during the next half.

The two independent polarizations of gravitational waves are denoted h_+ and $h_\times$. These are the two primary time-dependent observables of a gravitational wave. The polarization of gravitational waves from a source, such as a binary system, depends on the orientation of the dynamics inside the source relative to the observer. Therefore, measuring the polarization provides information about, for example, the orientation of the binary system.

5.2.3 Direction to a source

Gravitational wave antennas are linearly-polarized quadrupolar detectors and do not have good directional sensitivity. As a result we cannot deduce the direction to a source using a single antenna. One normally needs simultaneous observation using three or more detectors so that the source can be triangulated in the sky by measuring the time differences in signal arrival times at various detectors in a network. Ground-based detectors have typical separation baselines of $L \sim 3 \times 10^6$ m, so that at a wavelength of $\lambda = 3 \times 10^5$ m = 1 ms (a frequency of 1kHz) the network has a resolution of $\delta\theta = \lambda/L = 0.1$ rad. If the amplitude SNR is high, then one can localize the source by a factor of $1/\text{SNR}$ better than this.

For long-lived sources, however, a single antenna synthesizes many antennas by observing the source at different points along its orbit around the sun. The baseline for such observations is 2AU, so that, for a source emitting radiation at 1kHz, the resolution is as good as $\Delta\theta = 10^{-6}$ rad, which is smaller than an arcsecond.

For space-based detectors orbiting the sun, like LISA, the baseline is again 2 AU, but the observing frequency is some five or six orders of magnitude lower, so the basic resolution is only of order 1 radian. However, as we shall see later, some of the sources that a space-based detector will observe have huge amplitude SNRs in the range of $\text{SNR} \sim 10^3 - 10^4$, which improves the resolution to arcminute accuracies in the best cases.

5.2.4 Amplitude of gravitational waves - the quadrupole approximation

The Einstein equations are too difficult to solve analytically in the generic case of a strongly gravitating source to compute the luminosity and amplitude of gravitational waves from an astronomical source. We will discuss numerical solutions later; the most powerful available analytic approach is called the post-Newtonian approximation scheme. This approximation is suited to gravitationally bound systems, which constitute the majority of expected sources. In this scheme [79, 169], solutions are expanded in the small parameter $(v/c)^2$, where v is the typical dynamical speed inside the system. Because of the virial theorem, the dimensionless Newtonian gravitational potential ϕ/c^2 is of the same order, so that the expansion scheme links orders in the expanded metric with those in the expanded source terms. The lowest-order post-Newtonian approximation for the emitted radiation is the quadrupole formula, and it depends only on the density (ρ) and velocity fields of the Newtonian system. If we define the spatial tensor Q_{jk} , the second moment of the mass distribution, by the equation

$$Q_{jk} = \int \rho x_j x_k \, d^3x$$

then the amplitude of the emitted gravitational wave is, at lowest order, the three-tensor

$$h_{jk} = \frac{2}{r} \frac{d^2 Q_{jk}}{dt^2}$$

This is to be interpreted as a linearized gravitational wave in the distant almost-flat geometry far from the source, in a coordinate system (gauge) called the Lorentz gauge.

Wave amplitudes and polarization in TT-gauge

A useful specialization of the Lorentz gauge is the TT-gauge, which is a comoving coordinate system: free particles remain at constant coordinate locations, even as their proper separations change. To get the TT-amplitude of a wave traveling outwards from its source, project the tensor in Equation (2) perpendicular to its direction of travel and remove the trace of the projected tensor. The result of doing this to a symmetric tensor is to produce, in the transverse plane, a two-dimensional matrix with only two independent elements:

$$h_{ab} = \begin{pmatrix} h_+ & h_\times \\ h_\times & -h_+ \end{pmatrix}$$

This is the definition of the wave amplitudes h_+ and $h_\times$ that are illustrated in Figure 1. These amplitudes are referred to as the coordinates chosen for that plane. If the coordinate unit basis vectors in this plane are $\hat{e}_x$ and $\hat{e}_y$, then we can define the basis tensors

$$\begin{aligned} \mathbf{e}_+ &= \hat{e}_x \otimes \hat{e}_x - \hat{e}_y \otimes \hat{e}_y \\ \mathbf{e}_\times &= \hat{e}_x \otimes \hat{e}_y + \hat{e}_y \otimes \hat{e}_x \end{aligned}$$

In terms of these, the TT-gravitational wave tensor can be written as

$$\mathbf{h} = h_+ \mathbf{e}_+ + h_\times \mathbf{e}_\times$$

If the coordinates in the transverse plane are rotated by an angle ψ , then one obtains new amplitudes h'_+ and $h'_\times$ given by

$$\begin{aligned} h'_+ &= \cos 2\psi h_+ + \sin 2\psi h_\times, \\ h'_\times &= -\sin 2\psi h_+ + \cos 2\psi h_\times. \end{aligned}$$

This shows the quadrupolar nature of the polarizations, and is consistent with our remark in association with Figure 1 that a rotation of $\pi/4$ changes one polarization into the other.

It should be clear from the TT projection operation that the emitted radiation is not isotropic: it will be stronger in some directions than in others¹. It should also be clear from this that spherically-symmetric motions do not emit any gravitational radiation: when the trace is removed, nothing remains.

Simple estimates

A typical component of d^2Q_{jk}/dt^2 will (from Equation (1)) have magnitude $(Mv^2)_{\text{nonsph}}$, where $(Mv^2)_{\text{nonsph}}$ is twice the nonspherical part of the kinetic energy inside the source. So a bound on any component of Equation (2) is

$$h \lesssim \frac{2(Mv^2)_{\text{nonsph}}}{r}$$

It is interesting to observe that the ratio ϵ of the wave amplitude to the Newtonian potential ϕ_{ext} of its source at the observer's distance r is simply bounded by

$$h/\phi_{\text{ext}} < 2v_{\text{nonsph}}^2$$

and this bound is attained if the entire mass of the source is involved in the nonspherical motions, so that $(Mv^2)_{\text{nonsph}} \sim Mv_{\text{nonsph}}^2$. By the virial theorem for self-gravitating bodies

$$v_{\text{nonsph}}^2 \leq \phi_{\text{int}},$$

where ϕ_{int} is the maximum value of the Newtonian gravitational potential inside the system. This provides a convenient bound in practice [328]:

$$h \lesssim 2\phi_{\text{int}} \phi_{\text{ext}}.$$

The bound is attained if the system is highly nonspherical. An equal-mass star binary system is a good example of a system that attains this bound.

For a neutron star source, one has $\phi_{\text{int}} \sim 0.2$. If the star is in the Virgo cluster ($r \sim 18\text{Mpc}$) and has a mass of $1.4M_\odot$, and if it is formed in a highly-nonspherical gravitational collapse, then the upper limit on the amplitude of the radiation from such an event is 1.5×10^{-21} . This is a simple way to get the number that has been the goal of detector development for decades, to make detectors that can observe waves at or below an amplitude of about 10^{-21}

¹ In the case of an inspiraling binary, the root mean square of the two polarization amplitudes in a direction orthogonal to the orbital plane will be a factor $2\sqrt{2}$ larger than in the plane.

5.2.5 Frequency of gravitational waves

The signals for which the best waveform predictions are available have well-defined frequencies. In some cases the frequency is dominated by an existing motion, such as the spin of a pulsar. But in most cases the frequency will be related to the natural frequency for a self-gravitating body, defined as

$$\omega_0 = \sqrt{\pi G \bar{\rho}}, \quad \text{or} \quad f_0 = \omega_0/2\pi = \sqrt{G \bar{\rho}/4\pi},$$

where $\bar{\rho}$ is the mean density of mass-energy in the source. This is of the same order as the binary orbital frequency and the fundamental pulsation frequency of the body. Even though this is a Newtonian formula, it provides a remarkably good order-of-magnitude prediction of natural frequencies, even for highly relativistic systems such as black holes.

The frequency of the emitted gravitational waves need not be the natural frequency, of course, even if the mechanism is an oscillation with that frequency. In many cases, such as binary systems, the radiation comes out at twice the oscillation frequency. But since, at this point, we are not trying to be more accurate than a few factors, we will ignore this distinction here. In later sections, with specific source models, we will get the factors right.

The mean density and hence the frequency are determined by the size R and mass M of the source, taking $\bar{\rho} = 3M/4\pi R^3$. For a neutron star of mass $1.4M_\odot$ and radius 10 km, the natural frequency is $f_0 = 1.9\text{kHz}$. For a black hole of mass $10M_\odot$ and radius $2M = 30\text{ km}$, it is $f_0 = 1\text{kHz}$. And for a large black hole of mass $2.5 \times 10^6 M_\odot$, such as the one at the center of our galaxy, this goes down in inverse proportion to the mass to $f_0 = 4\text{mHz}$. In general, the characteristic frequency of the radiation of a compact object of mass M and radius R is

$$f_0 = \frac{1}{4\pi} \left(\frac{3M}{R^3} \right)^{1/2} \simeq 1\text{kHz} \left(\frac{10M_\odot}{M} \right)$$

Figure 2 shows the mass-radius diagram for likely sources of gravitational waves. Three lines of constant natural frequency are plotted: $f_0 = 10^4\text{ Hz}$, $f_0 = 1\text{ Hz}$, and $f_0 = 10^{-4}\text{ Hz}$. These are interesting frequencies from the point of view of observing techniques: gravitational waves between 1 and 10^4 Hz are in principle accessible to ground-based detectors, while lower frequencies are observable only from space. Also shown is the line marking the black-hole boundary. This has the equation $R = 2M$. There are no objects below this line, because they would be smaller than the horizon size for their mass. This line cuts through the ground-based frequency band in such a way as to restrict ground-based instruments to looking at stellar-mass objects. No system with a mass above about $10^4 M_\odot$ can produce quadrupole radiation in the ground-based frequency band.

A number of typical relativistic objects are placed in the diagram: a neutron star, a pair of neutron stars that spiral together as they orbit, some black holes. Two other interesting lines are drawn. The lower (dashed) line is the 1-year coalescence line, where the orbital shrinking timescale due to gravitational radiation backreaction (cf. Equation (28)) is less than one year. The upper (solid) line is the 1-year chirp line: if a binary lies below this line, then its orbit will shrink enough to make its orbital frequency increase by a measurable amount in one year. (In a one-year observation one can, in principle, measure changes in frequency of 1yr^{-1} , or $3 \times 10^{-8}\text{ Hz}$.)

It is clear from the Figure that any binary system that is observed from the ground will coalesce within an observing time of one year. Since pulsar binary statistics suggest that neutronstar-binary coalescences happen less often than once every 10^5 years in our galaxy, ground-based detectors must be able to register these events in a volume of space containing at least 10^6 galaxies in order to have a hope of seeing occasional coalescences. That corresponds to

a volume of radius roughly 100Mpc. For comparison, first-generation ground-based interferometric detectors have a reach of around 20Mpc for such binaries, while advanced interferometers should extend that to about 200Mpc.

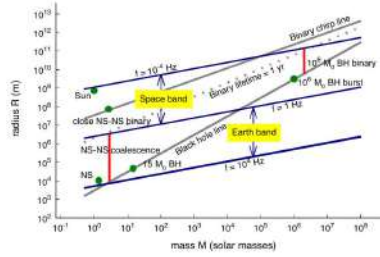


Figure 2: Mass-radius plot for gravitational wave sources. The horizontal axis is the total mass of a radiating system, and the vertical axis is its size. Typical values from various sources for ground-based and space-based detectors are shown. Lines give order-of-magnitude constraints and relations. Characteristic frequencies are estimated from $f \sim (G\rho/4\pi)^{1/2}$. The black-hole and binary lines are described in the text.

5.2.6 Luminosity in gravitational waves

The general formula for the local stress-energy of a gravitational wave field propagating through flat spacetime, using the TT-gauge, is given by the Isaacson expression [259, 332]

$$T_{\alpha\beta} = \frac{1}{32\pi} \langle h_{jk,\alpha}^{\text{TT}} h_{,\beta}^{\text{TT}jk} \rangle$$

where the angle brackets denote averages over regions of the size of a wavelength and times of the length of a period of the wave. The energy flux of a wave in the x^i direction is the T^{0i} component.

The gravitational wave luminosity in the quadrupole approximation is obtained by integrating the energy flux from Equation (14) over a distant sphere. When one correctly takes into account the projection factors mentioned after Equation (2), one obtains [259]

$$L_{gw} = \frac{1}{5} \left(\sum_{j,k} \ddot{Q}_{jk} \ddot{Q}_{jk} - \frac{1}{3} \ddot{Q}^2 \right)$$

where Q is the trace of the matrix Q_{jk} . This equation can be used to estimate the backreaction effect on a system that emits gravitational radiation.

Notice that the expression for L_{gw} is dimensionless when $c = G = 1$. It can be converted to normal luminosity units by multiplying by the scale factor

$$L_0 = c^5/G = 3.6 \times 10^{52} \text{ W}.$$

This is an enormous luminosity. By comparison, the luminosity of the sun is only 3.8×10^{26} W, and that of a typical galaxy would be 10^{37} W. All the galaxies in the visible universe emit, in visible light, on the order of 10^{49} W. We will see that gravitational wave systems always emit at a fraction of L_0 , but that the gravitational wave luminosity can come close to L_0 and can greatly exceed typical electromagnetic luminosities. Close binary systems normally radiate much more energy in gravitational waves than in light. Black hole mergers can, during their peak few cycles, compete in luminosity with the steady luminosity of the entire universe!

Combining Equations (2) and (15) one can derive a simple expression for the apparent luminosity of radiation $\mathcal{F}$, at great distances from the source, in terms of the gravitational wave amplitude [332]:

$$\mathcal{F} \sim \frac{|\dot{h}|^2}{16\pi}$$

The above relation can be used to make an order-of-magnitude estimate of the gravitational wave amplitude from a knowledge of the rate at which energy is emitted by a source in the form of gravitational waves. If a source at a distance r radiates away energy E in a time T , predominantly at a frequency f , then writing $\dot{h} = 2\pi f h$ and noting that $\mathcal{F} \sim E/(4\pi r^2 T)$, the amplitude of gravitational waves is

$$h \sim \frac{1}{\pi f r} \sqrt{\frac{E}{T}}$$

When the time development of a signal is known, one can filter the detector output through a copy of the expected signal (see Section 5 on matched filtering). This leads to an enhancement in the SNR, as compared to its narrow-band value, by roughly the square root of the number of cycles the signal spends in the detector band. It is useful, therefore, to define an effective amplitude of a signal, which is a better measure of its detectability than its raw amplitude:

$$h_{\text{eff}} \equiv \sqrt{n} h.$$

Now, a signal lasting for a time T around a frequency f would produce $n \simeq fT$ cycles. Using this we can eliminate T from Equation (18) and get the effective amplitude of the signal in terms of the energy, the emitted frequency and the distance to the source:

$$h_{\text{eff}} \sim \frac{1}{\pi r} \sqrt{\frac{E}{f}}$$

Notice that this depends on the energy only through the total fluence, or time-integrated flux $E/4\pi r^2$ of the wave. As in many other branches of astronomy, the detectability of a source is ultimately a function of its apparent luminosity and the observing time. However, one should not ignore the dependence on frequency in this formula. Two sources with the same fluence are not equally easy to detect if they are at different frequencies: higher frequency signals have smaller amplitudes.

5.3 Sources of Gravitational Waves

5.3.1 Man-made sources

One source can unfortunately be ruled out as undetectable: man-made gravitational radiation. Imagine creating a wave generator with the following extreme properties. It consists of two masses of 10^3 kg each (a small car) at opposite ends of a beam 10 m long. At its center the beam pivots about an axis. This centrifuge rotates 10 times per second. All the velocity is nonspherical, so v_{nonsph}^2 in Equation (9) is about $10^5 \text{ m}^2 \text{ s}^{-2}$. The frequency of the waves will actually be 20 Hz, since the mass distribution of the system is periodic in time with a period of half the rotation period. The wavelength of the waves will, therefore, be 1.5×10^7 m, similar to the diameter of the earth. In order to detect gravitational waves, not near-zone Newtonian gravity, the detector must be at least one wavelength from the source, say diametrically opposite the centrifuge on the Earth. Then the amplitude h can be deduced from Equation (9): $h \sim 5 \times 10^{-43}$. This is far too small to contemplate detecting! The story changes, fortunately, when we consider astrophysical sources of gravitational waves, where nature arranges for masses that are 10^{27} times larger than our centrifuge to move at speeds close to the speed of light!

Until observations of gravitational waves are successfully made, one can only make intelligent guesses about most of the sources that will be seen. There are many that could be strong enough to be seen by the early detectors: star binaries, supernova explosions, neutron stars, the early universe. In this section, we make rough luminosity estimates using the quadrupole formula and other approximations, which are usually accurate to within factors of order two, and, very importantly, they show how key observables scale with the properties of the systems. Where appropriate we also make use of predictions from the much more accurate modelling that is available for some sources, such as binary systems and black hole mergers. The detectability depends, of course, not only on the intrinsic luminosity of the source, but on how far away it is. Often the biggest uncertainties in making predictions are the spatial density and event rate of any particular class of sources. This is not surprising, since our information at present comes from electromagnetic observations, and as our earlier remarks about the differences between the mechanisms of emission of gravitational and electromagnetic radiation make clear, electromagnetic observations may not strongly constrain the source population.

5.3.2 Gravitational wave bursts from gravitational collapse

Neutron stars and black holes are formed from the gravitational collapse of a highly evolved star or the core collapse of an accreting white dwarf. In either case, if the collapse is nonspherical, perhaps induced by strong rotation, then gravitational waves could carry away some of the binding energy and angular momentum depending on the geometry of the collapse. Collapse events are thought to produce supernovae of various types, and increasingly there is evidence that they also produce most of the observed gamma-ray bursts [191] in hypernovae and collapsars [397, 249]. Supernovae of Type II are believed to occur at a rate of between 0.1 and 0.01 per year in a milkyway equivalent galaxy (MWEG); thus, within the Virgo supercluster, we might expect an event rate of about 30 per year. Hypernova events are considerably rarer and might only contribute observable gravitational-wave events in current and near-future detectors if they involve so much rotation that strong non-axisymmetric instabilities are triggered.

Simulating gravitational collapse is a very active area of numerical astrophysics, and most simulations also predict the energy and spectral characteristics of the emitted gravitational waves [167]. However, it is still beyond the capabilities of computers to simulate a gravitational collapse event with all the physics that might be necessary to give reliable predictions: three-dimensional hydrodynamics, neutrino transport, realistic nuclear physics, magnetic fields, rotation. In fact, it is still by no means clear why Type II supernovae explode at all: simulations typically have great difficulty reversing the inflow and producing an explosion with the observed light-curves and energetics. It may be that the answer lies in some of the physics that has to be oversimplified in order to be used in current simulations, or in some neutrino physics that we do not yet know, or in some unexplored hydrodynamic mechanism [276]. In a typical supernova, simulations suggest that gravitational waves might extract between about 10^{-7} and 10^{-5} of the total available massenergy [264, 147, 148], and the waves could come off in a burst whose frequency might lie in the range of $\sim 200 - 1000$ Hz.

We can use Equation (18) to make a rough estimate of the amplitude, if the emitted energy and timescale are known. Using representative values for a supernova in our galaxy, lying at 10kpc, emitting the energy equivalent of $10^{-7}M_{\odot}$ at a frequency of 1kHz, and lasting for 1 ms, the received amplitude would be

$$h \sim 6 \times 10^{-21} \left(\frac{E}{10^{-7}M_{\odot}} \right)^{1/2} \left(\frac{1 \text{ ms}}{T} \right)^{1/2} \left(\frac{1\text{kHz}}{f} \right) \left(\frac{10\text{kpc}}{r} \right)$$

The upper bound in Equation (11) would give the same amplitude for a source 60 times further away, which reflects the fact that simulations find it difficult to put significant energy

into gravitational waves. This amplitude is large enough for current ground-based detectors to observe with a reasonably high confidence, but of course the event rate within 10kpc is expected to be far too small to make an early detection likely.

5.3.3 Gravitational wave pulsars

Some likely gravitational wave sources behave like the centrifuge example we used in the first paragraph of this section, only on a grander scale. Suppose a neutron star of radius R and mass M spins with a frequency f and has an irregularity, a deformation of its otherwise axially symmetric shape. We idealize this as a "bump" of mass m on its surface, although of course it will really be a distribution of mass leading to an asymmetrical quadrupole tensor. The moment of inertia of the bump will be mR^2 , and it is conventional to parameterize the bump in terms of the fractional asymmetry it creates in the moment of inertia tensor itself. If we idealize the star as having uniform density, then the spherical moment of inertia is $2MR^2/5$, and so the bump has fractional asymmetry

$$\epsilon = \frac{5}{2} \frac{m}{M}, \quad m = 0.4\epsilon M$$

The bump will emit gravitational radiation at frequency $2f$ because the star spins about its net center of mass, so it effectively has mass excesses on both sides of the star. The nonspherical velocity will be just $v_{\text{nonsph}} = 2\pi Rf$. The radiation amplitude will be, from Equation (9),

$$h_{\text{bump}} \sim (4/5)(2\pi Rf)^2 \epsilon M/r$$

and the luminosity, from Equation (15) (assuming that roughly four comparable components of Q_{jk} contribute to the sum),

$$L_{\text{bump}} \sim (16/125)(2\pi f)^6 \epsilon^2 M^2 R^4$$

The radiated energy would presumably come from the rotational energy of the star $Mv^2/5$. This would lead to a spindown of the star on a timescale

$$t_{\text{spindown}} \sim \frac{1}{5} Mv^2 / L_{\text{bump}} \sim \frac{25}{32\pi} \epsilon^{-2} f^{-1} \left(\frac{M}{R} \right)^{-1} v^{-3}$$

It is believed that neutron star crusts are not strong enough to support fractional asymmetries larger than about $\epsilon \sim 10^{-6}$ [370], and realistic asymmetries may be much smaller.

From these considerations one can estimate the likelihood that the observed spindown timescales of pulsars are due to gravitational radiation. In most cases, it seems that gravitational wave losses could account for a substantial amount of the spindown: the required asymmetries are much smaller than 10^{-4} , often smaller than 10^{-7} . But an interesting exception is the Crab pulsar, PSR J0534+2200, whose young age and consequently short spindown time (measured to be 8.0×10^{10} s, about 2500yr) would require an exceptionally large asymmetry. If we take the neutron star's radius to be 10 km, so that $M/R \sim 0.21$ and the speed of any irregularity is $v/c \sim 6.2 \times 10^{-3}$, then Equation (24) would require an asymmetry of $\epsilon \sim 1.4 \times 10^{-3}$. Of course, we have made a lot of approximations to get here, only keeping our estimates of amplitudes and energies correct to within factors of two, but a more careful calculation reduces this only by a factor of two to $\epsilon \sim 7 \times 10^{-4}$ [12]. What makes this interesting is the fact that an asymmetry this large would produce radiation detectable by first-generation interferometers. Conversely, an upper limit from first-generation interferometers would provide direct observational limits on the asymmetry and on the fraction of energy lost by the Crab pulsar to gravitational waves.

From Equation (23) the Crab pulsar would, if its spindown is dominated by gravitational wave losses, produce an amplitude at the Earth of $h \sim 1.5 \times 10^{-24}$, if its distance is 2kpc. Is this detectable when present instruments are only capable of seeing millisecond bursts of radiation at levels of 10^{-21} ? The answer is yes, if the observation time is long enough. Indeed, the latest LIGO observations have not detected any gravitational waves from the Crab pulsar, which has been used to set an upper limit on the asymmetry in its mass distribution [12]. The limit depends on the model assumed for the pulsar. If one assumes that gravitational waves are produced at exactly twice the pulsar spin frequency and uses the inferred values of the pulsar orientation and polarization angle, then for a canonical value of the moment-of-inertia $I = 10^{38} \text{ kg m}^2$, one gets an upper limit on the ellipticity of $\epsilon \leq 1.8 \times 10^{-4}$, assuming the pulsar is at 2kpc. This is a factor of 4.2 below the spindown limit [12]. If, however, one assumes that gravitational waves are emitted at a frequency close, but not exactly equal, to twice the spin frequency and one uses a uniform prior for the orientation and polarization angle, then one gets $\epsilon \leq 9 \times 10^{-4}$, which is 0.8 of the limit derived from the spin-down rate.

Indeed, even signals weaker than the amplitude determined by the Crab spindown rate will be observable by present detectors, and these may be coming from a larger variety of neutron stars, in particular low-mass X-ray binary systems (LMXBs). The neutron stars in them are accreting mass and angular momentum, so they should be spinning up. Observations suggest that most neutron stars are spinning at speeds between about 300 and 600 Hz, far below their maximum, which is greater than 1000 Hz. The absence of faster stars suggests that something stops them from spinning up beyond this range. Bildsten suggested [77] that the limiting mechanism may be the re-radiation of the accreted angular momentum in gravitational waves, possibly due to a quadrupole moment created by asymmetrical heating induced by the accreted matter. Another possible mechanism [285] is that a "bump" of the kind we have treated is formed by accreting matter channeled onto the surface by the star's magnetic field. It is also possible that accretion drives an instability in the star that leads to steady emission [308, 270]. In either case, the stars could turn out to be long-lived sources of gravitational waves. This idea, which is a variant of one proposed long ago by Wagoner [383], is still speculative, but the numbers make a plausible case. We discuss it in more detail in Section 7.3.5.

5.3.4 Radiation from a binary star system

Radiation from a binary system and its backreaction

A binary star system can also be treated as a "centrifuge". Two stars of the same mass m in a circular orbit of radius R have all their mass in nonspherical motion, so that

$$(Mv^2)_{\text{nonsph}} = M(\Omega R)^2 = \frac{M^2}{R}$$

where Ω is the orbital angular velocity. The gravitational wave amplitude can then be written

$$h_{\text{binary}} \sim 2 \frac{M}{r} \frac{M}{R}$$

Since the internal radius R of the orbit is not an observable, it is sometimes convenient to replace R by the orbital angular frequency Ω using the above orbit equation, giving

$$h_{\text{binary}} \sim \frac{2}{r} M^{5/3} \Omega^{2/3}$$

The gravitational wave luminosity of such a system is, by a calculation analogous to that for bumps on neutron stars (assuming that four components of Q_{ij} to be significant),

$$L_{\text{binary}} \sim \frac{4}{5} \left(\frac{M}{R} \right)^5$$

in units given by the fundamental luminosity L_0 in Equation (16). This shows that self-gravitating systems always emit at a fraction of L_0 , since M/R is always smaller than 1, but it can approach L_0 for highly-relativistic systems where $M/R \sim 1$.

The radiation of energy by the orbital motion causes the orbit to shrink. The shrinking will make any observed gravitational waves increase in frequency with time. This is called a chirp. The timescale² for this in a binary system with equal masses is

$$t_{\text{chirp}} = \frac{Mv^2}{2}/L_{\text{binary}} \sim \frac{5M}{8} \left(\frac{M}{R} \right)^{-4}$$

As the binary evolves, the frequency and amplitude of the wave grow and this drives the binary to evolve even more rapidly. The signal's frequency, however, will not increase indefinitely; the slow inspiral phase ends either when the stars begin to interact and merge or (if they are very compact) when the distance between the stars is roughly at the last stable orbit (LSO) $R = 6M$, which corresponds to a gravitational wave frequency of

$$f_{\text{LSO}} \sim 220 \left(\frac{20M_{\odot}}{M} \right) \text{ Hz}$$

where we have normalized this to a binary with $M = 20M_{\odot}$. This is the last stable orbit (LSO) frequency.

A compact-object binary that coalesces after passing through the last stable orbit is a powerful source of gravitational waves, with a luminosity that approaches the limiting luminosity L_0 . This is called a coalescing binary in gravitational wave searches. Since a typical search might last on the order of one year, a coalescing binary can be defined as a system that has a chirp time smaller than one year. In Figure 2 the coalescence line is shown as a straight line with slope 3/4 (set t_{chirp} to a constant in Equation (28)). Binary systems below this line have a chirp time smaller than one year. It is evident from the figure that all binary systems observable by ground-based detectors will coalesce in less than a year.

As mentioned for gravitational wave pulsars, the raw amplitude of the radiation from a longlived system is not by itself a good guide to its detectability, if the waveform can be predicted. Data analysis techniques like matched filtering are able to eliminate most of the detector noise and allow the recognition of weaker signals. The improvement in amplitude sensitivity is roughly proportional to the square root of the number of cycles of the waveform that one observes. For neutron stars that are observed from a frequency of 10 Hz until they coalesce, there could be on the order of 10^4 cycles, meaning that the sensitivity of a second-generation interferometric detector would effectively be 100 times better than its broadband (prefiltering) sensitivity. Such detectors could see typical coalescences at $\sim 200\text{Mpc}$. The event rate for second-generation detectors is estimated at around 40 events per year, with rather large error bars [101, 211, 242].

Chirping binaries as standard sirens

When we consider real binaries we must do the calculation for systems that have unequal masses. Still assuming for the moment that the binary orbit is circular, the quadrupole amplitude turns out to be

² In Sections 5.1 we will use parameters called chirp times, instead of the masses, to characterize a binary. The timescale defined here is closely related to the chirp times.

$$h_{\text{binary}} \sim \frac{1}{r} \mathcal{M}^{5/3} \Omega^{2/3}$$

where we define the chirp mass $\mathcal{M}$ as

$$\mathcal{M} = \mu^{3/5} M^{2/5} = \nu^{3/5} M, \quad \nu = \frac{\mu}{M}$$

with μ the reduced mass, M the total mass and ν the symmetric mass ratio. We have left out of Equation (30) a factor of order one that depends on the angle from which the binary system is viewed. The two polarization amplitudes can be found in Equation (132).

Remarkably, the other observable, namely the shrinking of the orbit as measured by the rate of change of the orbital frequency P_b also depends on the masses just through $\mathcal{M}$ [290]:

$$\dot{P}_b = -\frac{192\pi}{5} \left(\frac{2\pi\mathcal{M}}{P_b} \right)^{5/3}$$

In this case, the chirp time is

$$t_{\text{chirp}} = \frac{5M}{96} \frac{1}{\nu} \left(\frac{M}{R} \right)^{-4}$$

This is just the equal-mass chirp time of Equation (28) scaled inversely with the symmetric mass ratio $\nu = m_1 m_2 / M^2$. From this equation it is clear that systems with large mass ratios between the components can spend a long time in highly relativistic orbits, whereas equal-mass binaries can be expected to merge after only a few orbits in the highly relativistic regime.

If one observes P_b and $\dot{P}_b$, one can infer $\mathcal{M}$ from Equation (32). Then, from the observed amplitude in Equation (30), the only remaining unknown is the distance r to the source. Gravitational wave observations of orbits that shrink because of gravitational energy losses can therefore directly determine the distance to the source [329]. By analogy with the "standard candles" of electromagnetic astronomy, these systems are now being called "standard sirens". Although our calculation here assumed an equal-mass circular system, the conclusion is robust: any binary, even with ellipticity and extreme mass ratio, encodes its distance in its gravitational wave signal.

This is another way in which gravitational wave observations are complementary to electromagnetic ones, providing information that is hard to obtain electromagnetically. One consequence is the possibility that observations of coalescing compact object binaries could allow one to measure the Hubble constant [329] or other cosmological parameters. This will be particularly interesting for the LISA project, whose observations of black hole binaries could contribute an independent measurement of the acceleration of the universe [195, 131, 48].

Because chirping systems are so interesting we have also drawn, in Figure 2, a line where the chirp time can be measured in one year. This means that the change in frequency due to the chirp must be larger than the frequency resolution 1yr^{-1} . A little algebra shows that the condition for the chirp to be resolved in an observation time T in a binary with period P_b is

$$P_b t_{\text{chirp}} = T^2.$$

Since $P_b \propto R^{3/2} M^{-1/2}$, this condition leads to a line of slope 7/11 in the logarithmic plot in Figure 2. The line drawn there corresponds to a resolution time T of one year. All binaries below this line will chirp in a short enough time for their distances to be measured.

Binary pulsar tests of gravitational radiation theory

The most famous example of the effects of gravitational radiation on an orbiting system is the Hulse-Taylor Binary Pulsar, PSR B1913+16. In this system, two neutron stars orbit in a close eccentric orbit. The pulsar provides a regular clock that allows one to deduce, from post-Newtonian effects, all the relevant orbital parameters and the masses of the stars. The key to the importance of this binary system is that all of the important parameters of the system can be measured before one takes account of the orbital shrinking due to gravitational radiation reaction. This is because a number of post-Newtonian effects on the arrival time of pulses at the Earth, such as the precession of the position of the periastron and the time-dependent gravitational redshift of the pulsar period as it approaches and recedes from its companion, can be measured accurately, and they fully determine the masses, the semi-major axis and the eccentricity of their orbit [394, 344].

Equation (28) for the chirp time predicts that this system would change its orbital period $P_b = 7.75\text{hrs}$ on the timescale (taking $M = 1.4M_\odot$ and $R = 10^6\text{ km}$)

$$t_{\text{chirp}} = P_b/\dot{P}_b \sim 1.9 \times 10^{18} \text{ s}$$

From this one can infer that $\dot{P}_b \sim 1.5 \times 10^{-14}$. But this has to be corrected for our oversimplification of the orbit as circular: an eccentric orbit evolves much faster because, during the phase of closest approach, the velocities are much higher, and the emitted luminosity is a very strong function of the velocity. Using equations first computed by Peters and Mathews [290], for the actual eccentricity of 0.62, one finds (see Equation (109) below) $\dot{P}_T = -(2.40242 \pm 0.00002) \times 10^{-12}$. Observations [394, 388] currently give $\dot{P}_O = -(2.4184 \pm 0.0009) \times 10^{-12}$. There is a significant discrepancy between these, but it can be removed by realizing that the binary system is accelerating toward the center of our galaxy, which produces a small period change. Taking this into account gives a corrected prediction of $-(2.4056 \pm 0.0051) \times 10^{-12}$, and this agrees with the observation within the uncertainties [394, 355]. This is the most sensitive test that we have of the correctness of Einstein's equations with respect to gravitational radiation, and it leaves little room for doubt in the validity of the quadrupole formula for other systems that may generate detectable radiation.

A number of other binary systems are now known in which such tests are possible [344]. The most important of the other systems is the "double pulsar" in which both neutron stars are seen as pulsars [246]. This system will soon overtake the Hulse-Taylor binary as the most accurate test of gravitational radiation.

White-dwarf binaries

Binary systems at lower frequencies are much more abundant than coalescing binaries, and they have much longer lifetimes. LISA will look for close white-dwarf binaries in our galaxy, and will probably see thousands of them. White dwarfs are not as compact as black holes or neutron stars. Although their masses can be similar to that of a neutron star their sizes are much larger, typically 3,000 km in radius. As a result, white-dwarf binaries never reach the last stable orbit, which would occur at roughly 1.5kHz for these masses. We will discuss the implications of multi-messenger astronomy for white-dwarf binaries in Section 7.4.

The maximum amplitude of the radiation from a white-dwarf binary will be several orders of magnitude smaller than that of a neutron star or black hole binary at the same distance but close to coalescence. However, a binary system with a short period is long lived, so the effective amplitude (after matched filtering) improves as the square root of the observing time. Besides that, these sources are nearer: there are many thousands of such systems in our galaxy radiating in the LISA frequency window above about 1mHz, and LISA should be able to see all of them. Below 1mHz there are even more sources, so many that LISA will not resolve

them individually, but will see them blended together in a stochastic background of radiation, as shown in Figure 5.

Supermassive black hole binaries

Observations indicate that the center of every galaxy probably hosts a black hole whose mass is in the range of $10^6 - 10^9 M_\odot$ [305], with the black holes mass correlating well with the mass of the galactic bulge. A black hole whose mass is in the above range is called a supermassive black hole (SMBH). There is now abundant observational evidence that galaxies often collide and merge, and there are good reasons to believe that when that happens, friction between the SMBHs and the stars and gas of the irregular merged galaxy will lead the SMBHs to spiral into a common nucleus and (on a timescale of some 10^8yr) even get close enough to be driven into complete orbital decay by gravitational radiation reaction. In many systems this should lead to a black hole merger within a Hubble time [221]. For a binary with two nonspinning $M = 10^6 M_\odot$ black holes, the frequency of emitted gravitational waves at the last stable orbit is, from Equation (29), $f_{\text{LSO}} = 4 \text{mHz}$; during and after the merger the frequency rises from 4mHz to the quasi-normal-mode frequency of 24mHz (if the spin of the final black hole is negligible). (Naturally, all these frequencies simply scale inversely with the mass for other mass ranges.) This is in the frequency range of LISA, and observing these mergers is one of the central purposes of the mission.

SMBH mergers are so spectacularly strong that they will be visible in LISA's data stream even before applying any matched filter, although good models of the inspiral and particularly the merger radiation will be needed to extract source parameters. Because the masses of such black holes are so large, LISA can see essentially any merger that happens in its frequency band anywhere in the universe, even out to extremely high redshifts. It can thereby address astrophysical questions about the origin, growth and population of SMBHs. The recent discovery of an SMBH binary [221] and the association of X-shaped radio lobes with the merger of SMBH binaries [254] has further raised the optimism concerning SMBH merger rates, as has the suggestion that an SMBH has been observed to have been expelled from the center of its galaxy, an event that could only have happened as a result of a merger between two SMBHs [222]. The rate at which galaxies merge is about 1yr^{-1} out to a red-shift of $z = 5$ [185], and LISA's detection rate for SMBH mergers might be roughly the same.

Modelling of the merger of two black holes requires numerical relativity, and the accuracy and reliability of numerical simulations is now becoming good enough that they will soon become an integral part of gravitational wave searches.

Extreme and intermediate mass-ratio inspiral sources

The SMBH environment of our own galaxy is known to contain a large number of compact objects and white dwarfs. Near the central SMBH there is a disproportionately large number of stellar-mass black holes, which have sunk there through random gravitational encounters with the normal stellar population (dynamical friction). Three body interaction will occasionally drive one of these compact objects into a capture orbit of the central SMBH. The compact object will sometimes be captured [305, 338, 337] into a highly eccentric trajectory ($e > 0.99$) with the periastron close to the last stable orbit of the SMBH. Since the mass of the captured object will be about $1 - 100 M_\odot$, while the SMBH will have a far greater mass, we essentially have a "test mass" falling in the geometry of a Kerr black hole. By Equation (33) we would expect that the small body would spend many orbits in the relativistic regime near the horizon of the large black hole: a $10 M_\odot$ black hole falling into a $10^6 M_\odot$ black hole would require on the order of 10^5 orbits. The emitted gravitational radiation [317, 179, 178, 67, 171, 57] would consist of a very long wave train that carries information about the nearly geodesic trajectory of the

test body, thereby providing a very clean probe to survey the spacetime geometry of the central object (which could be a Kerr black hole or some other compact object) and test whether or not this geometry is as predicted by general relativity [318, 198, 177, 176, 68].

This kind of event happens occasionally to every SMBH that lives in the center of a galaxy. Indeed, since the SNR from matched filtering builds up in proportion to the square root of the observation time $t_{\text{chirp}} \propto \nu^{-1} = (\mu/M)^{-1}$ [cf. Equation (33)] and the inherent amplitude of the radiation is linear in ν [cf. Equation (30)], the SNR varies with the symmetric mass ratio as $\sqrt{\nu}$ and typical SNR will be about ten to a thousand times smaller than an SMBH binary at the same distance. LISA will, therefore, be able to see such sources only to much smaller distances, say between 1 to 10Gpc depending on the mass ratio. For events at such distances LISA's SNR after matched filtering could be in the range 10 – 100, but matched filtering will be very difficult because of the complexity of the orbit, especially of its evolution due to radiation effects. However, this volume of space contains a large number of galaxies, and the event rate is expected to be several tens to hundreds per year [67]. There will be a background from more distant sources that might in the end set the limit on how much sensitivity LISA has to these events.

Due to relativistic frame dragging, for each passage of the apastron the test body could execute several nearly circular orbits at its periastron. Therefore, long periods of low-frequency, small-amplitude radiation will be followed by several cycles of high-frequency, large-amplitude radiation [317, 179, 178, 67, 171, 57]. The apastron slowly shrinks, while the periastron remains more or less at the same location, until the final plunge of the compact object before merger. Moreover, if the central black hole has a large spin then spin-orbit coupling leads to precession of the orbital plane thereby changing the polarization of the wave as seen by LISA.

Thus, there is a lot of structure in the waveforms owing to a number of different physical effects: contribution from higher-order multipoles due to an eccentric orbit, precession of the orbital plane, precession of the periastron, etc., and gravitational radiation backreaction plays a pivotal role in the dynamics of these systems. If one looks at the time-frequency map of such a signal one notices that the signal power is greatly smeared across the map [320], as compared to that of a sharp chirp from a nonspinning black-hole binary. For this reason, this spin modulated chirp is sometimes referred to as a smirch [322]. More commonly, such sources are called extreme mass ratio inspirals (EMRIs) and represent systems whose mass ratio is in the range of $\sim 10^{-3} - 10^{-6}$. Inspirals of systems with their mass ratio in the range $\sim 10^{-2} - 10^{-3}$ are termed intermediate mass ratio inspirals or IMRIs. These latter systems correspond to the inspiral of intermediate mass black holes of mass $\sim 10^3 - 10^4 M_{\odot}$ and might constitute a prominent source in LISA provided the central SMBH grew in mass as a result of a number of mergers of small black holes [30, 31, 32].

While black hole perturbation theory with a careful treatment of radiation reaction is necessary for the description of EMRIs, IMRIs may be amenable to a description using a hybrid scheme of post-Newtonian approximations and perturbation theory. This is an area that requires more study.

5.3.5 Quasi-normal modes of a black hole

In 1970, Vishveshwara [381] discussed a gedanken experiment, similar in philosophy to Rutherford's (real) experiment with the atom. In Vishveshwara's experiment, he scattered gravitational radiation off a black hole to explore its properties. With the aid of such a gedanken experiment, he demonstrated for the first time that gravitational waves scattered off a black hole will have a characteristic waveform, when the incident wave has frequencies beyond a certain value, depending on the size of the black hole. It was soon realized that perturbed black holes have quasi-normal modes (QNMs) of vibration and in the process emit gravitational

radiation, whose amplitude, frequency and damping time are characteristic of its mass and angular momentum [296, 220]. We will discuss in Section 6.4 how observations of QNMs could be used in testing strong field predictions of general relativity.

We can easily estimate the amplitude of gravitational waves emitted when a black hole forms at a distance r from Earth as a result of the coalescence of compact objects in a binary. The effective amplitude is given by Equation (20), which involves the energy E put into gravitational waves and the frequency f at which the waves come off. By dimensional arguments E is proportional to the total mass M of the resulting black hole. The efficiency at which the energy is converted into radiation depends on the symmetric mass ratio ν of the merging objects. One does not know the fraction of the total mass emitted nor the exact dependence on ν . Flanagan and Hughes [164] argue that $E \sim 0.03(4\nu)^2 M$. The frequency f is inversely proportional to M ; indeed, for Schwarzschild black holes $f = (2\pi M)^{-1}$. Thus, the formula for the effective amplitude takes the form

$$h_{\text{eff}} \sim \frac{4\alpha\nu M}{\pi r}$$

where α is a number that depends on the (dimensionless) angular momentum a of the black hole and has a value between 0.7 (for $a = 0$, Schwarzschild black hole) and 0.4 (for $a = 1$, maximally spinning Kerr black hole). For stellar mass black holes at a distance of 200 Mpc the amplitude is:

$$h_{\text{eff}} \simeq 10^{-21} \left(\frac{\nu}{0.25} \right) \left(\frac{M}{20M_{\odot}} \right) \left(\frac{r}{200\text{Mpc}} \right)^{-1}$$

For SMBHs, even at cosmological distances, the amplitude of quasinormal mode signals is pretty large:

$$h_{\text{eff}} \simeq 3 \times 10^{-17} \left(\frac{\nu}{0.25} \right) \left(\frac{M}{2 \times 10^6 M_{\odot}} \right) \left(\frac{r}{6.5\text{Gpc}} \right)^{-1}$$

In the first case we have a pair of $10M_{\odot}$ black holes inspiraling and merging to form a single black hole. In this case the waves come off at a frequency of around 500 Hz [cf. Equation (13)]. The initial ground-based network of detectors might be able to pick these waves up by matched filtering, especially when an inspiral event precedes the ringdown signal. A $100M_{\odot}$ black hole plunging into a $10^6 M_{\odot}$ black hole at a distance of 6.5Gpc ($z \simeq 1$) gives out radiation at a frequency of about 15mHz. Note that in the latter case the radiation is redshifted from 30mHz to 15mHz. Such an event produces an amplitude just large enough to be detected by LISA. At the same distance, a pair of $10^6 M_{\odot}$ SMBHs spiral in and merge to produce a fantastic amplitude of 3×10^{-17} , way above the LISA background noise. In this case, the signals would be given off at about 7.5mHz and will be loud and clear to LISA. It will not only be possible to detect these events, but also to accurately measure the masses and spins of the objects before and after merger and thereby test the black hole no-hair theorem and confirm whether the result of the merger is indeed a black hole or some other exotic object (e.g., a boson star or a naked singularity).

5.3.6 Stochastic background

In addition to radiation from discrete sources, the universe should have a random gravitational wave field that results from a superposition of countless discrete systems and also from fundamental processes, such as the Big Bang. Observing any of these backgrounds would bring useful information, but the ultimate goal of detector development is the observation of the background radiation from the Big Bang. It is expected to be very weak, but it will come to us

unhindered from as early as 10^{-30} s, and it could illuminate the nature of the laws of physics at energies far higher than we can hope to reach in the laboratory.

It is usual to characterize the intensity of a random field of gravitational waves by its energy density as a function of frequency. Since the energy density of a plane wave is the same as its flux (when $c = 1$), we have from Equation (17)

$$\rho_{\text{gw}} = \frac{\pi}{4} f^2 h^2$$

But the wave field in this case is a random variable, so we must replace h^2 by a statistical mean square amplitude per unit frequency (Fourier transform power per unit frequency) called $S_{\text{gw}}(f)$, so that the energy density per unit frequency is proportional to $f^2 S_{\text{gw}}(f)$. It is then conventional to talk about the energy density per unit logarithm of the frequency, which means multiplying by f . The result, after being careful about averaging over all directions of the waves and all independent polarization components, is [27, 359]

$$\frac{d\rho_{\text{gw}}}{d\ln f} = 4\pi^2 f^3 S_{\text{gw}}(f)$$

Finally, what is of most interest is the energy density as a fraction of the closure or critical cosmological density, given by the Hubble constant H_0 as $\rho_c = 3H_0^2/8\pi$. The resulting ratio is called $\Omega_{\text{gw}}(f)$:

$$\Omega_{\text{gw}}(f) = \frac{10\pi^2}{3H_0^2} f^3 S_{\text{gw}}(f)$$

The only tight constraint on Ω_{gw} from non-gravitational-wave astronomy is that it must be smaller than 10^{-5} , in order not to disturb the agreement between the standard Big Bang model of nucleosynthesis (of helium and other light elements) and observation. If the universe contains this much gravitational radiation today, then at the time of nucleosynthesis the (blue-shifted) energy density of this radiation would have been comparable to that of the photons and the three neutrino species. Although the radiation would not have participated in the nuclear reactions, its extra energy density would have required that the expansion rate of the universe at that time be significantly faster, in order to evolve into the universe we see today. In turn, this faster expansion would have provided less time for the nuclear reactions to "freeze out", altering the abundances from the values that are observed today [281, 346]. First-generation interferometers should be able to set direct limits on the cosmological background at around this level. Radiation in the lower-frequency LISA band, from galactic and extra-galactic binaries, is expected to be much smaller than this bound.

Random radiation is indistinguishable from instrumental noise in a single detector, at least for short observing times. If the random field is produced by an anisotropically-distributed set of astrophysical sources (the binaries in our galaxy, for example) then over a year, as the detector changes its orientation, the noise from this background should rise and fall in a systematic way, allowing it to be identified. But this is a rather crude way of detecting the radiation, and a better way is to perform a cross-correlation between two detectors, if available.

In cross-correlation, which amounts to multiplying the outputs and integrating, the random signal in one detector essentially acts as a template for the signal in the other detector. If they match, then there will be a stronger-than-expected correlation. Notice that they can only match well if the wavelength of the gravitational waves is longer than the separation between the detectors: otherwise time delays for waves reaching one detector before the other degrade the match. The outcome is not like standard matched filtering, however, since the "filter" of the first detector has as much noise superimposed on its template as the other detector. As a result, the amplitude SNR of the correlated field grows only with observing time T as $T^{1/4}$, rather than the square root growth that characterizes matched filtering [359].

5.4 Physics with Gravitational Waves by Sathyaprakash

Classical general relativity has passed all possible experimental and observational tests so far. The theory is elegant, self-consistent and mathematically complete (i.e., its equations are, in principle, solvable). However, theorists are uncomfortable with general relativity because it has so far eluded all efforts of quantization, making it a unique modern theory, whose quantum mechanical analogue is unknown. Although general relativity arises as a by-product in certain string theories, the physical relevance of such theories is unclear. Therefore, it has been proposed that general relativity is a low-energy limit of a more general theory, which in itself is amenable to both quantization and unification. There are also other theoretical motivations to look for modifications of general relativity or new theories of gravity. While there are some alternative candidates (including the Brans-Dicke theory), none has predictions that contradict general relativistic predictions in linear and mildly nonlinear gravitational fields. More precisely, the extra parameters of these other theories of gravity are constrained by the present experimental and astronomical observations, however, they are expected to significantly deviate from general relativistic predictions under conditions of strong gravitational fields.

Gravitational wave observations provide a unique opportunity to test strongly nonlinear and highly relativistic gravity and hence provide an unprecedented testbed for confronting different theories of gravity. Every nonlinear gravitational effect in general relativity will have a counterpart in alternative theories and therefore a measurement of such an effect would provide an opportunity to compare the performance of general relativity with its competitors. Indeed, a single measurement of the full polarization of an incident gravitational wave can potentially rule out general relativity. This is a field that would benefit from an in-depth study. What we are lacking is a systematic study of higher-order post-Newtonian effects in alternative theories of gravity. For instance, we do not know how tails of gravitational waves or tails of tails would appear in any theory other than general relativity.

In what follows we present strong field tests of general relativity afforded by future gravitational wave observations. We will begin with observations of single black holes followed by black hole binaries (more generally, coalescing binaries of compact objects).

5.4.1 Speed of gravitational waves

Association of a gravitational wave event with an electromagnetic event, such as the observation of a gamma or X-ray burst coincidentally with a gravitational wave event, would help to deduce the speed of gravitational waves to a phenomenal accuracy. The best candidate sources for the simultaneous observations of both are the well-known extra-galactic gamma-ray bursts (GRBs). Depending on the model that produces the GRB, the delay between the emission of a GRB and gravitational waves might be either a fraction of a second (as in GRBs generated by internal shocks in a fireball [306]) or 100's of seconds (as in GRBs generated when the fireball is incident on an external medium [256]). It is unlikely that high-redshift gamma-ray observations will be visible in the gravitational wave band, since the amplitude of gravitational waves might be rather low. However, advanced detectors might see occasional low-redshift events, especially if the GRB is caused by black-hole-neutron-star mergers. Third generation detectors would be sensitive to such events up to $z = 2$. A single unambiguous association can verify the speed of gravitational waves relative to light to a fantastic precision.

For instance, even a day's delay in the arrival times of gravitational and electromagnetic radiation from a source at a distance of one giga light year (distance to a low-redshift GRB detectable by advanced detectors) would determine the relative speeds to better than one part in 10^{11}) $1 \text{ day} / 10^9 \text{ yr} \sim 3 \times 10^{-12}$). Coincident detection of GRBs and gravitational waves would require good timing accuracy to determine the direction of the source so that astronomical observations of associated gamma rays (and afterglows in other spectral bands) can be made.

Consequently, gravitational wave antennas around the globe will have to make a coincident detection of the event.

If the speed of gravitational waves is less than that of light, then this could indicate that the graviton has an effective nonzero mass.

This would have other observable effects, in particular dispersion; different frequencies should move at different speeds. Will [393] pointed out that LISA's observations of coalescences of SMBHs at high redshifts will place extremely tight constraints on dispersion, and may, therefore, indirectly set the best available limits on the speed of gravitational waves. This and other bounds on the graviton mass are discussed in Section 6.6.1.

5.4.2 Polarization of gravitational waves

As noted in Section 2, in Einstein's theory gravitational waves have two independent polarizations, usually denoted h_+ and $h_\times$ [259]. A general wave will be a linear combination of both. Rotating sources typically emit both polarizations with a phase delay between them, leading to elliptical polarization patterns. Depending on the nature of the source such polarizations can be detected either with a single detector (in the case of continuous wave sources) or with a network of detectors (in the case of burst sources).

While Einstein's general relativity predicts only two independent polarizations, there are other theories of gravitation in which there are additional states of polarization. For instance, in Fierz-Jordan-Brans-Dicke theory [394] there are four polarization degrees of freedom more than in Einstein's theory. Therefore, an unambiguous determination of the polarization of the waves will be of fundamental importance.

In the case of a burst source, to determine two polarization states, source direction and amplitude requires three detectors, observing other polarizations would require the use of more than three detectors (see, for example, Will [394]). The scalar polarization mode of Brans-Dicke, for example, expands a transverse ring of test particles without changing its shape. This is the breathing mode, or monopole polarization. If such a wave is incident from above on an interferometer, it will not register at all. But if it comes in along one of the arms, then, since it acts transversely, it will affect only the other arm and leave a signal. If the wave is seen with enough detectors, then it is possible to determine that it has scalar polarization. Note that a measurement such as this can make a qualitative change in physics: a single measurement could put general relativity in jeopardy.

Polarization measurements have an important application in astronomy. The polarization of the waves contains orientation information. For example, a binary system emits purely circular polarization along the angular momentum axis, but purely linear polarization in its equatorial plane. By measuring the polarization of waves from a binary (or from a spinning neutron star) one can determine the orientation and inclination of its spin axis. This is a piece of information that is usually very hard to extract from optical observations. We will return to this discussion in Section 7.1.1.

5.4.3 Gravitational radiation reaction

In 1974, Hulse and Taylor discovered the first double neutron star binary PSR B1913+16, a system in which the emission of gravitational radiation has an observable effect [200, 356]. General relativity predicts that the loss of energy and angular momentum due to the emission of gravitational waves should cause the period of the system to decrease and, by carefully monitoring the orbital period of the binary, that it would be possible to measure the rate at which the period changes. The rate at which the period decays can be computed using the quadrupole formula for the luminosity of the emitted radiation combined with the energy-

balance equation; namely that the energy carried away by the waves comes at the expense of the binding energy of the system.

For a binary consisting of stars of masses m_1 and m_2 , in an orbit of eccentricity e and period P_b , the period decay is given by the generalization of Equation (32) [290]:

$$\dot{P}_b = -\frac{192\pi}{5} \left(\frac{2\pi\mathcal{M}}{P_b} \right)^{5/3} \left(1 + \frac{73}{24}e^2 + \frac{37}{96}e^4 \right) (1 - e^2)^{-7/2}$$

where we recall that

$$\mathcal{M} = (m_1 m_2)^{3/5} (m_1 + m_2)^{-1/5} = \mu^{3/5} M^{2/5}$$

is the chirpmass of the binary that we defined in Equation (31). (In the third expression here, μ is the reduced mass of the binary and M its total mass.) Since the masses of the binary and the eccentricity of the orbit can be measured by other means, one can use these parameters in the above equation to infer the rate at which the period is predicted to decrease according to general relativity. For the Hulse-Taylor binary the relevant values are: $m_1 = 1.4414M_\odot$, $m_2 = 1.3867M_\odot$, $e = 0.6171338$, $P_b = 2.790698 \times 10^4$ s. The predicted value $\dot{P}_b^{\text{GR}} = -(2.40242 \pm 0.00002) \times 10^{-12}$, while the observed period decay (after subtracting the apparent decay due to the acceleration of the pulsar in the gravitational field of our galaxy, as described in Section 3.4.3) is $\dot{P}_b^{\text{Obs}} = -(2.4056 \pm 0.0051) \times 10^{-12}$ and the two are in agreement to better than a tenth of a percent [394].

Observation of the decay of the orbital period in PSR B1913+16 is an unambiguous direct observation of the effect of gravitational radiation backreaction on the dynamics of the system. PSR B1913+16 was the first system in which the effect of gravitational radiation reaction force was measured. In 2004, a new binary pulsar PSR J0737-3039 was discovered [101, 247]. J0737 is in a tighter orbit than PSR B1913+16; with an orbital period of only 2.4hrs, the orbit is shrinking by about 7 mm each day in good agreement with the general relativistic prediction. Several other systems are also known [242]. In Sections 6.5, 6.5.2 and 6.5.3 we will discuss in some detail the dynamics of relativistic binaries and the radiation reaction as predicted by post-Newtonian theory and numerical relativity simulations.

5.4.4 Black hole spectroscopy

An important question relating to the structure of a black hole is its stability. Studies that began in the 1970s [307, 381, 382, 398, 296, 357, 358, 297] showed that a black hole is stable under external perturbation. A formalism was developed to study how a black hole responds to generic external perturbations, which has come to be known as black hole perturbation theory [112]. What we now know is that a distorted Kerr black hole relaxes to its axisymmetric state by partially emitting the energy in the distortion as gravitational radiation. The radiation consists of a superposition of QNMs, whose frequency and damping time depend uniquely on the mass M and spin angular momentum J of the parent black hole and not on the nature of the external perturbation. The amplitudes and damping times of different modes, however, are determined by the details of the perturbation and are not easy to calculate, except in some simple cases.

The uniqueness of the QNMs is related to the "no-hair" theorem of general relativity according to which a black hole is completely specified by its mass and spin⁹. Thus, observing QNMs would not only confirm the source to be a black hole, but would be an unambiguous proof of the uniqueness theorem of general relativity.

The end state of a black hole binary will lead to the formation of a single black hole, which is initially highly distorted. Therefore, one can expect coalescing black holes to end their

lives with the emission of QNM radiation, often called ringdown radiation. It was realized quite early on [164] that the energy emitted during the ringdown phase of a black-hole-binary coalescence could be pretty large. Although, the initial quantitative estimates [164] have proven to be rather high, the qualitative nature of the prediction has proven to be correct. Indeed, numerical relativity simulations show that about 1-2% of a binary's total mass would be emitted in QNMs [298]. The effective one-body (EOB) model [96, 97], the only analytical treatment of the merger dynamics, gives the energy in the ringdown radiation to be about 0.7% of the total mass, consistent with numerical results. Thus, it is safe to expect that the ringdown will be as luminous an event as the inspiral and the merger phases. The fact that QNMs can be used to test the no-hair theorem puts a great emphasis on understanding their properties, especially the frequencies, damping times and relative amplitudes of the different modes that will be excited during the merger of a black hole binary and how accurately they can be measured.

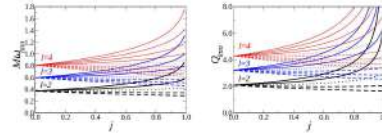


Figure 8: Normal mode frequencies (left) and corresponding quality factors (right) of fundamental modes with $l = 2, 3, 4$, as a function of the dimensionless black hole spin j , for different values of $m = l, \dots, 0, \dots, -l$ (for each l , different line styles from top to bottom correspond to decreasing values of m). Figure reprinted with permission from [76]. (c) The American Physical Society.

QNMs are characterized by a complex frequency ω that is determined by three "quantum" numbers, (l, m, n) (see, e.g., [76]). Here (l, m) are indices that are similar to those for standard spherical harmonics. For each pair of (l, m) there are an infinitely large number of resonant modes characterized by another integer n . The time dependence of the oscillations is given by $\exp(i\omega t)$, where ω is a complex frequency, its real part determining the mode frequency and the imaginary part (which is always positive) giving the damping time: $\omega = \omega_{lmn} + i/\tau_{lmn}$, $\omega_{lmn} = 2\pi f_{lmn}$ defining the angular frequency and τ_{lmn} the damping time. The ringdown wave will appear in a detector as the linear combination $h(t)$ of the two polarizations h_+ and $h_\times$, that is $h(t) = F_+ h_+ + F_\times h_\times$, F_+ and $F_\times$ being the antenna pattern functions as defined in Equation (57). The polarization amplitudes for a given mode are given by

$$h_+ = \frac{A(f_{lmn}, Q_{lmn}, \epsilon_{rd})}{r} (1 + \cos^2 \iota) \exp\left(\frac{-\pi f_{lmn} t}{Q_{lmn}}\right) \cos(2\pi f_{lmn} t + \varphi_{lmn}),$$

$$h_\times = \frac{A(f_{lmn}, Q_{lmn}, \epsilon_{rd})}{r} 2 \cos \iota \exp\left(\frac{-\pi f_{lmn} t}{Q_{lmn}}\right) \sin(2\pi f_{lmn} t + \varphi_{lmn}),$$

where ι is the angle between the black hole's spin axis and the observer's line of sight and φ_{lmn} is an unknown constant phase. The quality factor Q_{lmn} of a mode is defined as $Q_{lmn} = \omega_{lmn} \tau_{lmn} / 2$ and gives roughly the number of oscillations that are observable before the mode dies out. Figure 8 [76] plots frequencies and quality factors for the first few QNMs as a function of the dimensionless spin parameter $j = J/M^2$. The mode of a Schwarzschild black hole corresponding to $l = 2, m = n = 0$, is given by

$$f_{200} = \pm 1.207 \times 10^3 \frac{10 M_\odot}{M} \text{ Hz}, \quad \tau_{200} = 5.537 \times 10^{-4} \frac{M}{10 M_\odot} \text{ s}$$

For stellar-mass-black-hole coalescences expected to be observed in ground-based detectors the ringdown signal is a transient that lasts for a very short time. However, for space-based LISA the signal would last several minutes for a black hole of $M = 10^7 M_\odot$. In the latter case,

⁹ A black hole can, in principle, carry an electric charge in addition to mass and spin angular momentum. However, astrophysical black holes are believed to be electrically neutral

the ringdown waves could carry the energy equivalent of $10^5 M_\odot$ converted to gravitational waves a phenomenal amount of energy compared even to the brightest quasars and gamma ray bursts. Thus, LISA should be able to see QNMs from black hole coalescences anywhere in the universe, provided the final (redshifted) mass of the black hole is larger than about $10^6 M_\odot$, as otherwise the signal lasts for far too short a time for the detector to accumulate the SNR.

Berti et al. [76] have carried out an exhaustive study, in which they find that the LISA observations of SMBH binary mergers could be an excellent testbed for the no-hair theorem. Figure 9 (left panel) plots the fractional energy ϵ_{rd} that must be deposited in the ringdown mode so that the event is observable at a distance of 3Gpc. Black holes at 3Gpc with mass M in the range of $10^6 - 10^8 M_\odot$ would be observable (i.e., will have an SNR of 10 or more) even if a fraction $\epsilon_{\text{rd}} \simeq 10^{-7} M$ of energy is in the ringdown phase. Numerical relativity predicts that as much as 1% of the energy could be emitted as QNMs, when two black holes merge, implying that the ringdown phase could be observed with an SNR of 100 or greater all the way up to $z \sim 10$, provided their mass lies in the appropriate range¹⁰. Furthermore, they find that at this redshift it should be possible to resolve the fundamental $l = 2, m = 2$ mode. Since black holes forming from primordial gas clouds at $z = 10 - 15$ could well be the seeds of galaxy formation and large-scale structure, LISA could indeed witness their formation through out the cosmic history of the universe.

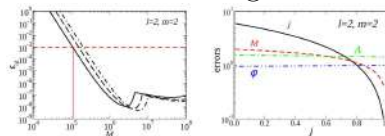


Figure 9: The smallest fraction of black hole mass in ringdown waveforms that is needed to observe the fundamental mode at a distance of 3Gpc (left) for three values of the black hole spin, $j = 0$ (solid line) $j = 0.80$ (dashed line) and $j = 0.98$ (dot-dashed line) and the error in the measurement of the various parameters as a function of the black hole spin for the same mode (right). Figure reprinted with permission from [76]. (c) The American Physical Society.

Figure 9 (right panel) shows SNR-normalized errors (i.e., one-sigma deviations multiplied by the SNR) in the measurement of the various QNM parameters (the mass of the hole M , its spin j , the QNM amplitude A and phase φ) for the fundamental $l = m = 2$ mode. We see that, for expected ringdown efficiencies of $\epsilon_{\text{rd}} \simeq 10^{-2} M$ into the fundamental mode of an a-million-solarmass black hole with spin $j = 0.8$ at 3Gpc ($\rho \sim 2000$), the mass and spin of the black hole can be measured to an accuracy of a tenth of a percent.

By observing a mode's frequency and damping time, one can deduce the (redshifted) mass and spin of the black hole. However, this is not enough to test the no-hair theorem. It would be necessary, although by no means sufficient, to observe at least one other mode (whose damping time and frequency can again be used to find the black hole's mass and spin) to see if the two are consistent with each other. Berti et al [76] find that such a measurement should be possible if the event occurs within a redshift of $z \sim 0.5$.

5.4.5 The two-body problem in general relativity

The largest effort in gravitational radiation theory in recent years has been to study the two-body problem using various approximations. The reason is that gravitationally bound binary systems are likely to be important gravitational wave sources, and until the evolution of such a system is thoroughly understood, it will not be possible to extract the maximum possible information from the observations.

¹⁰ Note that a black hole of physical mass M at a redshift of z will appear as a black hole of mass $M_z = (1+z)M$. This shifts the frequency of the QNM to the lower end of the spectrum. Assuming a frequency cutoff of 10^{-4} Hz for LISA, this means that only black holes of intrinsic mass $M < 1.2 \times 10^8 M_\odot / (1+z)$ can be observed at a redshift z .

From Figure 2, we see that ground-based detectors will be sensitive to compact binaries with mass in the range of $[1, 10^4] M_\odot$ while LISA will be sensitive to the mass range $[10^4, 10^8] M_\odot$. As we have seen in Section 3, most classes of binary sources will follow orbits that evolve strongly due to gravitational radiation reaction. In the case of ground-based detectors, they will all merge within a year of entering the observation band. In the case of LISA, we might observe sources (both stellar mass binaries as well as SMBH binaries), whose frequency hardly changes.

In contrast to Newtonian gravity, modeling a bound binary in general relativity is complicated by the existence of gravitational radiation and the nonlinearity of Einstein's equations. It must therefore be done approximately. The three most important approximation methods for solving gravitational wave problems are:

- The post-Newtonian scheme. This is a combination of a low-velocity expansion (v/c small) and a weak-field expansion (M/R small), in which the two small parameters are linked because a gravitationally-bound binary satisfies the virial relation $v^2 \sim M/R$, even in relativity. The zero-order solution is the Newtonian binary system. The post-Newtonian (PN) approximation has now been developed to a very high order in v/c because the velocities in late-stage binaries, just before coalescence, are very high.
- Perturbation theory. This is an expansion in which the small parameter is the massratio of the binary components. The zero-order solution is the field of the more massive component, and linear field corrections due to the second component determine the binary's orbital motion and the emitted radiation. This approximation is fully relativistic at all orders. It is being used to study the signals emitted by compact stars and stellar-mass black holes as they fall into SMBHs, an important source for LISA.
- Numerical approaches. With numerical relativity one can in principle simulate any desired relativistic system, no matter how strong the fields or high the velocities. It is being used to study the final stage of the evolution of binaries, including their coalescence, after the PN approximation breaks down. Although it deals with fully relativistic and nonlinear general relativity, the method needs to be regarded as an approximate one, since spacetime is not resolved to infinite precision. The accuracy of a numerical simulation is normally judged by performing convergence tests, that is by doing the simulation at a variety of resolutions and showing that there are no unexpected differences between them.

We will review the physics that can be learned from models using each of these approximation schemes. But first we treat a subject that is common to all binaries that evolve due to radiation reaction, which is that one can estimate their distance from a gravitational wave observation.

Binaries as standard candles: distance estimation

Astronomers refer to systems as standard candles if their intrinsic luminosity is known, so that when the apparent luminosity of a particular system is measured, then its distance can be deduced. As mentioned in Section 3.4.2, radiating binaries have this property, if one can measure the effects of radiation reaction on their orbits [329]. Because of the one-dimensional nature of gravitational wave data, some scientists have begun calling these standard sirens [195]. Over cosmological distances, the distance measured from the observation is the luminosity distance. We discuss in Section 8 below how this can be used to determine the Hubble constant and even the acceleration of the universe in methods independent of any cosmic distance ladder.

Numerical approaches to the two-body problem

From the point of view of relativity, the simplest two-body problem is that of two black holes. There are no matter fields and no point particles, just pure gravity. Therefore, the physics is entirely governed by Einstein's equations, which are highly nonlinear and rather difficult to solve. A number of teams have worked for over three decades towards developing accurate numerical solutions for the coalescence of two black holes, using fully three-dimensional numerical simulations.

A breakthrough came in early 2005 with Pretorius [298] announcing the results from the first stable simulation ever, followed by further breakthroughs by two other groups [104, 61] with successful simulations. The main results from numerical simulations of nonspinning black holes are rather simple. Indeed, just as the EOB had predicted, and probably contrary to what many people had expected, the final merger is just a continuation of the adiabatic inspiral, leading on smoothly to merger and ringdown. In Figure 10 we show the results from one of the numerical simulations (right panel) and that of the EOB (left panel), both for the same initial conditions. There is also good agreement in the prediction of the total energy emitted by the system, being $5.0\%(\pm 0.4\%)$ (for a review see [299]) and 3.1% [97], by numerical simulations and EOB, respectively, as well as the spin of the final black hole (respectively, 0.69 and 0.8) that results from the merger.

The total energy emitted and the spin angular momentum of the black hole both depend on the spin angular momenta of the parent black holes and how they are aligned with respect to the orbital angular momentum. In the test-mass limit, it is well known that the last stable orbit of a test particle in prograde orbit will be closer to, and that of a retrograde orbit will be farther from, the black hole as compared to the Schwarzschild case. Thus, prograde orbits last longer and radiate more compared to retrograde orbits. The same is true even in the case of spinning black holes of comparable masses; the emitted energy will be greater when the spins are aligned with the orbital angular momentum and least when they are anti-aligned. For instance, for two equal mass black holes, each with its spin angular momentum equal to 0.76, the total energy radiated in the aligned (anti-aligned) case is $6.7\%(2.2\%)$ and the spin of the final black hole is $0.89(0.44)$ [105, 294]. Heuristically, in the aligned case the black holes experience a repulsive force, deferring the merger of the two bodies to a much later time than in the anti-aligned case, where they experience an attractive force, accelerating the merger.

Detailed comparisons [137, 284, 88] show that we should be able to deploy the analytical templates from EOB [100, 135, 138, 139] (and other approximants [22]) that better fit the numerical data in our searches. With the availability of merger waveforms from numerical simulations and analytical templates, it will now be possible to search for compact binary coalescences with a greater sensitivity. The visibility of the signal improves significantly for binaries with their component masses in the range $[10, 100]M_\odot$. Currently, an effort is underway to evaluate how to make use of numerical relativity simulations in gravitational wave searches [352], which should help to increase the distance reach of interferometric detectors by a factor of two and correspondingly nearly an order-of-magnitude increase in event rate.

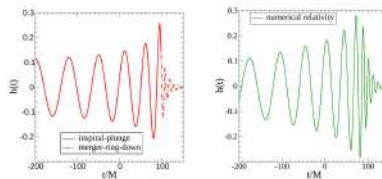


Figure 10: Comparison of waveforms from the analytical EOB approach (left) and numerical relativity simulations (right) for the same initial conditions. The two approaches predict very similar values for the total energy emitted in gravitational waves and the final spin of the black hole. Figure from [94].

Numerical relativity simulations have now greatly matured, allowing a variety of different

studies. Some are studying the effect of the spin orientations of the component black holes on the linear momentum carried away by the final black hole, fancifully called kicks [188, 92, 62, 105, 180, 294]; some have focused on the dependence of the emitted waveform phase and energy on the mass ratio; and yet others have strived to evolve the system with high accuracy and for a greater number of cycles so as to push the techniques of numerical relativity to the limit [89, 88].

Of particular interest are the numerical values of black hole kicks that have been obtained for certain special configurations of the component spins. Velocities as large as 4000 km s^{-1} have been reported by several groups, but such velocities are only achieved when both black holes have large ¹¹ spins. Such velocities are in excess of escape velocities typical of normal galaxies and are, therefore, of great astronomical significance. These high velocities, however, are not seen for generic geometries of the initial spin orientations; therefore, their astronomical significance is not yet clear.

What is the physics behind kicks? Beamed emission of radiation from a binary could result in imparting a net linear momentum to the final black hole. The radiation could be beamed either because the masses of the two black holes are not the same (resulting in asymmetric emission in the orbital plane) or because of the precession of the orbital plane arising from spin-orbit and spin-spin interactions, or both. In the case of black holes with unequal masses, the largest kick one can get is around 170 km s^{-1} , corresponding to a mass ratio of about 3 : 1. It was really with the advent of numerical simulations that superkicks begin to be realized, but only when black holes had large spins. The spin-orbit configurations that produce large kicks are rather unusual and at first sight unexpected. When the component black holes are both of the same mass and have equal but opposite spin angular momenta that lie in the orbital plane, frame dragging can lead to tilting and oscillation of the orbital plane, which, in the final phases of the evolution, could result in a rather large kick [299]. SMBHs are suspected to have large spins and, therefore, the effect of spin on the evolution of a binary and the final spin and kick velocity could be of astrophysical interest too.

Curiously, a recent optical observation of a distant quasar, SDSS J0927 12.65+294344.0, could well be the first identification of a superkick, causing the SMBH to escape from the parent galaxy [222]. From a fundamental physics point of view, kicks offer a new way of testing frame dragging in the vicinity of black holes, but much work is needed in this direction.

More recently, there has been an effort to understand and predict [98, 309, 136] the spin of the final black hole, which should help in further exploring interesting regions of the spin parameter space. In the relatively simple case of two black holes with equal and aligned spins of magnitude a , but unequal masses, with the symmetric mass ratio being $\nu = m_1 m_2 / (m_1 + m_2)^2$, Rezzolla et al. [309] have obtained an excellent fit for the final spin a_{fin} of the black hole by enforcing basic constraints from the test-mass limit:

$$a_{\text{fin}} = a + \left(2\sqrt{3} + t_0 a + s_4 a^2\right) \nu + (s_5 a + t_2) \nu^2 + t_3 \nu^3,$$

where $t_0 = -2.686 \pm 0.065$, $t_2 = -3.454 \pm 0.132$, $t_3 = 2.353 \pm 0.548$, $s_4 = -0.129 \pm 0.012$, and $s_5 = -0.384 \pm 0.261$. The top and middle panels of Figure 11 compare as functions of black hole spin and the symmetric mass ratio the goodness of their fit (blue short-dashed line, top panels) with the predictions of numerical simulations (circles and stars) from different groups (AEI [311], FAU-Jena [251], Jena [75] and Goddard [100]). Their residuals (red dotted lines, bottom panels) are less than a percent over the entire parameter space observed. These figures also show the fits obtained for the equal-mass but variable-spin case (green long-dashed line, left panel) [98] and for the nonspinning but unequal-mass case (green long-dashed line, middle

¹¹ By large spins we mean values that are close to the maximum value allowed by general relativity. If $\mathbf{J}$ is the magnitude of the spin angular momentum then general relativity requires that $|\mathbf{J}| \leq M^2$

panel) [136].

For the simple case of two equal mass black holes with aligned spins, the above analytical formula predicts that minimal and maximal final spin values of $a_{\text{fin}} = 0.35 \pm 0.03$ and $a_{\text{fin}} = 0.96 \pm 0.03$, respectively [309]. More interestingly, one can now ask what initial configurations of the mass ratios and spins would lead to the formation of a Schwarzschild black hole (i.e., $a_{\text{fin}}(a, \nu) = 0$) [199], which defines the boundary of the region on one side of which lie systems for which the spin of the final black hole flips relative to the initial total angular momentum (bottom panel in Figure 11).

Finally, the evolution of binaries composed of nonspinning bodies is characterized by a single parameter, namely the ratio of the masses of the two black holes. The study of systems with different mass ratios has allowed relativists to fit numerical waveforms with phenomenological waveforms [22]. The advantage of the latter waveforms is that one is able to more readily carry out data analysis in any part of the parameter space without needing the numerical data over the entire signal manifold.

Numerical relativity is still in its infancy and the parameter space is quite large. In the coming years more accurate simulations should become available, allowing the computation of waveforms with more cycles and less systematic errors. However, the challenge remains to systematically explore the effect of different spin orientations, mass ratios and eccentricity. One area that has not been explored using perturbative methods or post-Newtonian theory is that of intermediatemass-ratio inspirals. These are systems with moderate mass ratios of order 100:1, where neither black-hole perturbation theory nor post-Newtonian approximation might be adequate. Yet, the prospect for detecting such systems in ground and space-based detectors is rather high. Numerical relativity simulations might be the only way to set up effectual search templates for such systems.

Post-Newtonian approximation to the two-body problem

For the interpretation of observations of neutron-star-binary coalescences, which might be detected within five years by upgraded detectors that are now taking data, it is necessary to understand their orbital evolution to a high order in the PN expansion. The first effects of radiation reaction

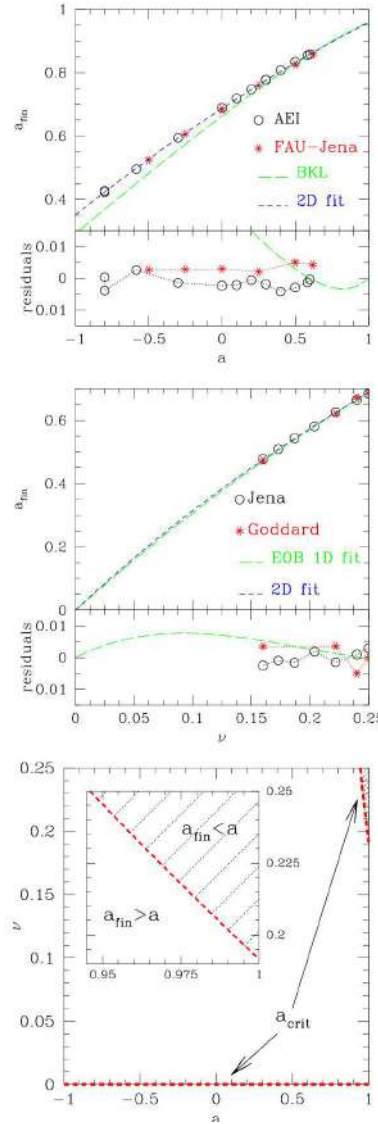


Figure 11: The final spin of a black hole that results from the merger of two equal mass black holes of aligned spins (top panel) and nonspinning unequal mass black holes (middle panel). The bottom panel shows the region in the parameter space that results in an overall flip in the spin-orbit orientation of the system. Figure reprinted with permission from [310]. See text for details. (c) The American Astronomical Society. are seen at 2.5 PN order (i.e., at order $(v/c)^5$ beyond Newtonian gravity), but we probably have to have control in the equations of motion over the expansion at least to 3.5PN order beyond the first radiation reaction (i.e., to order $(v/c)^{12}$ beyond Newtonian dynamics). There are many approaches to this, and we can not do justice here to the enormous effort that has gone into this field in recent years and refer the reader to the Living Reviews by Blanchet [79] and by Futamase & Itoh [169].

Most work on this problem so far has treated a binary system as if it were composed of two point masses. This is, strictly speaking, inconsistent in general relativity, since the masses should form black holes of finite size. Blanchet, Damour, Iyer, and collaborators [78] have avoided this problem by a method that involves generalized functions. They first expand in the nonlinearity parameter, and, when they have reached sufficiently high order, they obtain the velocity expansion of each order. By ordering terms in the post-Newtonian manner they have developed step-by-step the approximations up to 3.5PN order.

A different team, led by Will, works with a different method of regularizing the point-particle singularity and compares its results with those of Blanchet et al. at each order [82]. There is no guarantee that either method can be continued successfully to any particular order, but so far they have worked well and are in agreement. Their results form the basis of the templates

that are being designed to search for binary coalescences.

An interesting way of extending the validity of the expansion that is known to any order is to use Padé approximants [133, 134] (rational polynomials) of the fundamental quantities in the theory, namely the orbital energy and the gravitational wave luminosity. This has worked rather well in improving the convergence of PN theory. Buonanno and Damour [96, 97] have proposed an EOB approach to two-body dynamics, which makes it possible to compute the orbit of the binary and hence the phasing of the gravitational waves emitted beyond the last stable orbit into the merger and ringdown phases in the evolution of the black hole binary. This analytical approach has been remarkably successful and gained a lot of ground after the recent success in numerical relativity (see Section 6.5.2).

Other methods have been applied to this problem. Futamase [168] introduced a limit that combines the nonlinearity and velocity expansions in different ways in different regions of space, so that the orbiting bodies themselves have a regular (finite relativistic self-gravity) limit, while their orbital motion is treated in a Newtonian limit. This should not fail at any order [169], and has demonstrated its robustness by arriving at the same results as the other approaches, at least through 3 PN order. But it has a degree of arbitrariness in choosing initial data (see [327]) that could cause problems for gravitational wave search templates that integrate orbits for a long period of time.

Linear calculations of point particles around black holes are of interest in themselves and also for checking results of the full two-body calculations. These are well-developed for certain situations, e.g., [354, 258]. But the general equation of motion for such a body, taking into account all nongeodesic effects, has not yet been cast into a form suitable for practical calculations [106, 301]. This field is reviewed by two separate Living Reviews [292, 319].

Matched filtering, discussed in Section 5.1, is a plausible method of testing the validity of different approaches to computing the inspiral and merger waveforms from binary systems. Though a single observation is not likely to settle the question as to which methods are correct, a catalogue of events will help to evaluate the accuracy of different approaches by studying the statistics of the SNRs they measure.

6.5.3.1 Post-Newtonian expansions of energy and luminosity. Post-Newtonian calculations yield the expansion of the gravitational binding energy E and the gravitational wave luminosity $\mathcal{F}$ as a function of the post-Newtonian expansion parameter ^{12}v . This is related to the frequency f_{gw} of the dominant component of gravitational waves emitted by the binary system by

$$v^3 = \pi M f_{\text{gw}}$$

where M is the total mass of the system. The expansions for a circular binary are [80, 81, 79]

$$E = -\frac{\nu M v^2}{2} \left\{ 1 + \left(-\frac{9+\nu}{12} \right) v^2 + \left(\frac{-81+57\nu-\nu^2}{24} \right) v^4 \right. \\ \left. + \left(-\frac{675}{64} + \left[\frac{34445}{576} - \frac{205\pi^2}{96} \right] \nu - \frac{155}{96} \nu^2 - \frac{35}{5184} \nu^3 \right) v^6 + \mathcal{O}(v^8) \right\}$$

and

$$\begin{aligned} \mathcal{F} = & \frac{32\nu^2 v^{10}}{5} \left\{ 1 - \left(\frac{1247}{336} + \frac{35}{12}\nu \right) v^2 + 4\pi v^3 + \left(-\frac{44711}{9072} + \frac{9271}{504}\nu + \frac{65}{18}\nu^2 \right) v^5 \right. \\ & - \left(\frac{8191}{672} + \frac{583}{24} \right) \pi v^5 + \left[\frac{6643739519}{69854400} + \frac{16}{3}\pi^2 - \frac{1712}{105}(\gamma + \ln(4v)) \right. \\ & + \left(-\frac{4709005}{272160} + \frac{41}{48}\pi^2 \right) \nu - \frac{94403}{3024}\nu^2 - \frac{775}{324}\nu^3 \left. \right] v^6 \\ & + \left(-\frac{16285}{504} + \frac{214745}{1728}\nu + \frac{193385}{3024}\nu^2 \right) \pi v^7 + \mathcal{O}(v^8) \left. \right\} \end{aligned}$$

where $\gamma = 0.577\dots$ is Euler's constant.

6.5.3.2 Evolution equation for the orbital phase. Starting from these expressions, one requires that gravitational radiation comes at the expense of the binding energy of the system (see, e.g., [134]):

$$\mathcal{F} = -\frac{dE}{dt}$$

the energy balance equation. This can then be used to compute the (adiabatic) evolution of the various quantities as a function of time. For instance, the rate of change of the orbital velocity $\omega(t) = v^3/M$ (M being the total mass) can be computed using:

$$\frac{d\omega(t)}{dt} = \frac{d\omega}{dv} \frac{dv}{dE} \frac{dE}{dt} = \frac{3v^2}{M} \frac{\mathcal{F}(v)}{E'(v)}, \quad \frac{dv}{dt} = \frac{dv}{dE} \frac{dE}{dt} = \frac{-\mathcal{F}(v)}{E'(v)}$$

where $E'(v) = dE/dv$. Supplemented with a differential equation for t ,

$$dt = \frac{dt}{dE} \frac{dE}{dv} = -\frac{E'(v)}{\mathcal{F}}$$

one can solve for the evolution of the system's orbital velocity. Similarly, the evolution of the orbital phase $\varphi(t)$ can be computed using

$$\frac{d\varphi(t)}{dt} = \frac{v^3}{M}, \quad \frac{dv}{dt} = \frac{-\mathcal{F}(v)}{E'(v)}$$

6.5.3.3 Phasing formulas. The foregoing evolution equations for the orbital phase can be solved in several equivalent ways [134], each correct to the required post-Newtonian order, but numerically different from one another. For instance, one can retain the rational polynomial $\mathcal{F}(v)/E(v)$ in Equation (118) and solve the two differential equations numerically, thereby obtaining the time evolution of $\varphi(t)$. Alternatively, one might re-expand the rational function $\mathcal{F}(v)/E(v)$ as a polynomial in v , truncate it to order v^n (where n is the order to which the luminosity is given), thereby obtaining a parametric representation of the phasing formula in terms of polynomial expressions in v :

$$\varphi(v) = \varphi_{\text{ref}} + \sum_{k=0}^n \varphi_k v^k, \quad t(v) = t_{\text{ref}} + \sum_{k=0}^n t_k v^k$$

where φ_{ref} and t_{ref} are a reference phase and time, respectively. The standard post-Newtonian phasing formula goes one step further and inverts the second of the relations above

¹² In Newton's theory a two-body problem can be reduced to a one-body problem, in which a body of reduced mass μ moves in an effective potential. The parameter v is the velocity of the reduced mass, if the orbit is circular. In the extreme mass ratio limit $\nu \rightarrow 0$, v is the velocity of the smaller mass.

to express v as a polynomial in t (again truncated to appropriate order), which is then substituted in the first of the expressions above to obtain a phasing formula as an explicit function of time:

$$\begin{aligned} \varphi(t) = & \frac{-1}{\nu\tau^5} \left\{ 1 + \left(\frac{3715}{8064} + \frac{55}{96}\nu \right) \tau^2 - \frac{3\pi}{4}\tau^3 + \left(\frac{9275495}{14450688} + \frac{284875}{258048}\nu + \frac{1855}{2048}\nu^2 \right) \tau^4 \right. \\ & + \left(-\frac{38645}{172032} + \frac{65}{2048}\nu \right) \pi\tau^5 \ln \tau + \left[\frac{831032450749357}{57682522275840} - \frac{53}{40}\pi^2 - \frac{107}{56}(\gamma + \ln(2\tau)) \right. \\ & + \left(-\frac{126510089885}{4161798144} + \frac{2255}{2048}\pi^2 \right) \nu + \frac{154565}{1835008}\nu^2 - \frac{1179625}{1769472}\nu^3 \left. \right] \tau^6 \\ & + \left. \left(\frac{188516689}{173408256} + \frac{488825}{516096}\nu - \frac{141769}{516096}\nu^2 \right) \pi\tau^7 \right\} \\ v^2 = & \frac{\tau^2}{4} \left\{ 1 + \left(\frac{743}{4032} + \frac{11}{48}\nu \right) \tau^2 - \frac{\pi}{5}\tau^3 + \left(\frac{19583}{254016} + \frac{24401}{193536}\nu + \frac{31}{288}\nu^2 \right) \tau^4 \right. \\ & + \left(-\frac{11891}{53760} + \frac{109}{1920}\nu \right) \pi\tau^5 + \left[-\frac{10052469856691}{6008596070400} + \frac{\pi^2}{6} + \frac{107}{420}(\gamma + \ln 2\tau) \right. \\ & + \left(\frac{3147553127}{780337152} - \frac{451}{3072}\pi^2 \right) \nu - \frac{15211}{442368}\nu^2 + \frac{25565}{331776}\nu^3 \left. \right] \tau^6 \\ & + \left. \left(-\frac{113868647}{433520640} - \frac{31821}{143360}\nu + \frac{294941}{3870720}\nu^2 \right) \pi\tau^7 \right\}. \end{aligned}$$

In the above formulas $v = \pi M f_{\text{gw}}$ and $\tau = [\nu(t_C - t)/(5M)]^{-1/8}$, t_C being the time at which the two stars merge together and the gravitational wave frequency f_{gw} formally diverges.

6.5.3.4 Waveform polarizations. The post-Newtonian formalism also gives the two polarizations h_+ and $h_\times$ as multipole expansions in powers of the parameter v . To lowest order, the two polarizations of the radiation from a binary with a circular orbit, located at a distance D , with total mass M and symmetric mass ratio $\nu = m_1 m_2 / M^2$, are given by

$$h_+ = \frac{2\nu M}{D} v^2 (1 + \cos^2 \iota) \cos[2\varphi(t)], \quad h_\times = \frac{4\nu M}{D} v^2 \cos \iota \sin[2\varphi(t)]$$

where ι is the inclination of the orbital plane with the line of sight and v is the velocity parameter introduced earlier.

An interferometer will record a certain combination of the two polarizations given by

$$h(t) = F_+ h_+ + F_\times h_\times,$$

where the beam pattern functions F_+ and $F_\times$ are those discussed in Section 4.2.1. In the case of ground-based instruments, the signal duration is pretty small, at most 15 min for neutron star binaries and smaller for heavier systems. Consequently, one can assume the source direction to be unchanging during the course of observation and the above combination produces essentially the same functional form of the waveforms as in Equation (122). Indeed, it is quite straightforward to show that

$$h(t) = 4\nu M \frac{\mathcal{C}}{D} v^2 \cos[2\varphi(t) + 2\varphi_0]$$

where

$$\mathcal{C} = \sqrt{A^2 + B^2}, \quad A = \frac{1}{2} (1 + \cos^2 \iota) F_+, \quad B = \cos \iota F_\times, \quad \tan 2\varphi_0 = \frac{B}{A}.$$

The factor $\mathcal{C}$ is a function of the various angles and lies in the range $[0,1]$ with an RMS value of $2/5$ (see Section 4.2.1, especially the discussion following Equation (62)).

These waveforms form the basis for evaluating the science that can be extracted from future observations of neutron star and black hole binaries. We will discuss the astrophysical and cosmological measurements that are made possible with such high precision waveforms in several sections that follow (6.5.5 and 8.3). It is clear from the expressions for the waveform polarizations that, at the lowest order, the radiation from a binary is predominantly emitted at twice the orbital frequency. However, even in the case of quasi-circular orbits the waves come off at other harmonics of the orbital frequency. As we shall see below, these harmonics are very important for estimating the parameters of a binary, although they do not seem to contribute much to the SNR of the system.

Measuring the parameters of an inspiraling binary

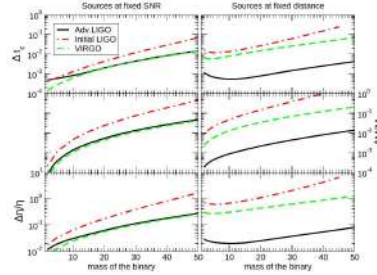


Figure 12: One-sigma errors in the time of coalescence, chirp mass and symmetric mass ratio for sources with a fixed SNR (left panels) and at a fixed distance (right panels). The errors in the time of coalescence are given in ms, while in the case of chirp mass and symmetric mass ratio they are fractional errors. These plots are for nonspinning black hole binaries; the errors reduce greatly when dynamical evolution of spins are included in the computation of the covariance matrix. Slightly modified figure from [49].

The issue of parameter estimation in the context of black hole binaries has received a lot of attention [114, 128, 165, 64, 293, 65, 44]. Most authors have used the covariance matrix for this purpose, although Markov Chain Monte Carlo (MCMC) techniques have also been used occasionally [116, 316, 315, 120], especially in the context of LISA [364, 121, 122, 123, 125]. Covariance matrix is often the preferred method, as one can explore a large parameter space without having to do expensive Monte Carlo simulations. However, when the parameter space is large, covariance matrix is not a reliable method for estimating parameter accuracies, especially at low SNRs [63, 64, 372]; but at high SNRs, as in the case of SMBH binaries in LISA, the problem might be that our waveforms are not accurate enough to facilitate a reliable extraction of the source parameters [129]. Although MCMC methods can give more reliable estimates, they suffer from being computationally extremely expensive. However, they are important in ascertaining the validity of results based on the covariance matrix, at least in a small subset of the parameter space, and should probably be employed in assessing parameter accuracies of candidate gravitational wave events.

In what follows we shall summarize the most recent work on parameter estimation in ground and space-based detectors for binaries with and without spin and the improvements brought about by including higher harmonics.

6.5.4.1 Ground-based detectors - nonspinning components. In Figure 12 we have plotted the one-sigma uncertainty in the measurement of the time of coalescence, chirp mass and symmetric mass ratio for initial and advanced LIGO and VIRGO [44]. The plots show errors for sources all producing a fixed SNR of 10 (left panels) or all at a fixed distance of 300 Mpc (right panels). The fractional error in chirp mass, even at a modest SNR of 10, can be as low as a few parts in

ten thousand for stellar mass binaries, but the error stays around 1%, even for heavier systems that have only a few cycles in a detector's sensitivity band. Error in the mass ratio is not as small, increasing to 100% at the higher end of the mass range explored. Thus, although the chirpmass can be measured to a good accuracy, poor estimation of the mass ratio means that the individual masses of the binary cannot be measured very well. Note also that the time of coalescence of the signal is determined pretty well, which means that we would be able to measure the location of the system in the sky quite well.

At a given SNR the accuracy is better in the case of low-mass binaries, since they spend a longer duration and a greater number of cycles in the detector band and the chirpmass can be determined better than the mass ratio, since to first order the frequency evolution of a binary is determined only by the chirpmass.

6.5.4.2 Measuring the parameters of supermassive black hole binaries in LISA. In the case of LISA, the merger of SMBHs produces events with extremely large SNRs, even at a redshift of $z = 1$ (100s to several thousands depending on the chirpmass of the source). Therefore, one expects to measure the parameters of a merger event in LISA to a phenomenal accuracy. Figure 13 depicts the distribution of the errors for a binary consisting of two SMBHs of masses $(10^6, 3 \times 10^5) M_\odot$ at a redshift of $z = 1$ [232]. The distribution was obtained for ten thousand samples of the system corresponding to random orientations of the binary at random sky locations with the starting frequency greater than 3×10^{-5} Hz and the ending frequency corresponding to the last stable orbit.

Each plot in Figure 13 shows the results of computations for binaries consisting of black holes with and without spins. Even in the absence of spin-induced modulations in the waveform, the parameter accuracies are pretty good. Note that spin-induced modulations in the waveform enable a far better estimation of parameters, chirpmass accuracy improving by more than an order of magnitude and reduced mass accuracy by two orders of magnitude. It is because of such accurate

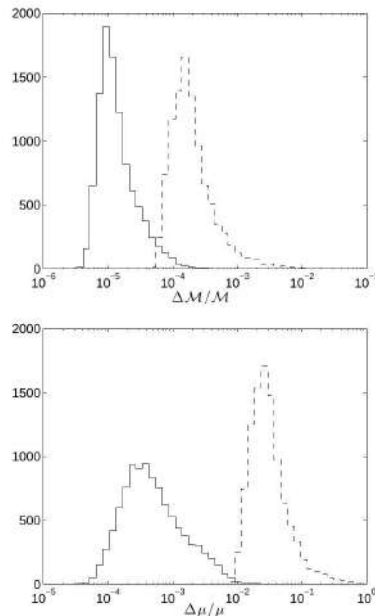


Figure 13: Distribution of measurement accuracy for a binary merger consisting of two black holes of masses $m_1 = 10^6 M_\odot$ and $m_2 = 3 \times 10^5 M_\odot$, based on 10,000 samples of the system in which the sky location and orientation of the binary are chosen randomly. Dashed lines are for nonspinning systems and solid lines are for systems with spin. Figure reprinted with permission from [231]. (c) The American Physical Society. measurements that it will be possible to use SMBH mergers to test general relativity in the strong field regime of the theory (see below).

Although Figure 13 corresponds to a binary with specific masses, the trends shown are found to be true more generically for other systems too, the actual parameter accuracies and

improvements due to spin both depending on the specific system studied.

Improvement from higher harmonics

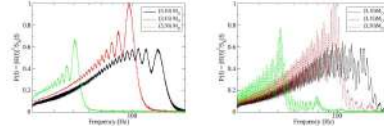


Figure 14: The SNR integrand of a restricted (left panel) and full waveform (right panel) as seen in initial LIGO. We have shown three systems, in which the smaller body’s mass is the same, to illustrate the effect of the mass ratio. In all cases the system is at 100Mpc and the binary’s orbit is oriented at 45° with respect to the line of sight.

The results discussed so far use the restricted post-Newtonian approximation in which the waveform polarizations contain only twice the orbital frequency, neglecting all higher-order corrections (including those to the second harmonic). The full waveform is a post-Newtonian expansion of the two polarizations as a power-series in v/c and consists of terms that have not only the dominant harmonic at twice the orbital frequency, but also other harmonics of the waveform. Schematically, the full waveform can be written as [79, 374]

$$h(t) = \frac{4M\eta}{D_L} \sum_{k=1}^7 \sum_{n=0}^5 A_{(k,n/2)} v^{n+2}(t) \cos [k\varphi(t) + \varphi_{(k,n/2)}]$$

where $\nu = m_1 m_2 / M^2$ is the symmetric mass ratio, the first sum (index k) is over the different harmonics of the waveform and the second sum (index n) is over the different post-Newtonian orders. Note that post-Newtonian order weighs down the importance of higher-order amplitude corrections by an appropriate factor of the small parameter v . In the restricted post-Newtonian approximation one keeps only the lowest-order term. Since $A_{1,0}$ happens to be zero, the dominant term corresponds to $k = 2$ and $n = 0$, containing twice the orbital frequency.

The various signal harmonics, and the associated additional structure in the waveform, can potentially enhance our ability to measure the parameters of a binary to a greater accuracy. The reason we can expect to do so can be seen by looking at the spectra of gravitational waves with and without these harmonics. For a binary that is oriented face on with respect to a detector only the second harmonic is seen, while for any other orientation the radiation is emitted at all other harmonics, the influence of the harmonics becoming more pronounced as the inclination angle changes from 0 to $\pi/2$. Figure 14 compares, in the frequency band of ground-based detectors, the spectrum of a source using the restricted post-Newtonian approximation (left panel) to the full waveform. In both cases the source is inclined to the line of sight at 45 degrees.

Following is a list of improvements brought about by higher harmonics. In the case of groundbased detectors Van Den Broeck and Sengupta [374, 375] found that, when harmonics are included, the SNR hardly changes, but is always smaller, relative to a restricted waveform. However, the presence of frequencies higher than twice the orbital frequency means that it will be possible to observe heavier systems, increasing the mass reach of ground-based detectors by a factor of 2 to 3 in advanced LIGO and third generation detectors [374,375]. The same effect was found in the case of LISA too, allowing LISA to observe SMBH masses up to a few $\times 10^8 M_\odot$ [47]. More than the increased mass reach, the harmonics reduce the error in the estimation of the chirpmass, symmetric mass ratio and the time of arrival by more than an order of magnitude for stellar-mass black hole binaries. The same is true to a greater extent in the case of SMBH binaries, allowing as well a far greater accuracy in the measurement of the luminosity distance and sky resolution in LISA’s observation of these sources [48, 361]. For instance, Figure 15

[361] shows the gain in LISA's angular resolution for two massive black-hole-binary mergers as a consequence of using higher harmonics for a specific orientation of the binary. Improvements of order 10 to 100 can be seen over large regions of the sky. This improved performance of LISA makes it a good probe of dark energy [48] (see Section 8.3).

A word of caution is in order with regard to the improvements brought about by higher harmonics. If the sensitivity of a detector has an abrupt lower frequency cutoff, or falls off rapidly below a certain frequency, then the harmonics bring about a more dramatic improvement than when the sensitivity falls off gently. Higher harmonics, nevertheless, always help in reducing the random errors associated with the measurement of parameters of a coalescing black-hole binary.

5.4.6 Tests of general relativity

Gravitational wave measurements of black holes automatically test general relativity in its strongfield regime. Observations of the mergers of comparable-mass black holes will be rich in details of their strong-field interactions. If measurements can determine the masses and spins of the initial black holes, as well as the eccentricity and orientation of their inspiral orbit, then one would hope to compare the actual observed waveform with the output of a numerical simulation of the same system. If measurements can also determine the final mass and spin (say from the ringdown radiation) then one can test the Hawking area theorem (the final area must exceed the sum of the areas of the initial holes) and the Penrose cosmic censorship conjecture (the final black hole should have $J/M^2 < 1$).

Observations of stellar mass black holes inspiraling into SMBHs, the extreme mass ratio inspirals (EMRIs), have an even greater potential for testing general relativity. The stellar mass black hole spends thousands of precessing (both of periastron and the orbital plane) orbits along highlyeccentric trajectories and slowly inspirals into the larger black hole. The emitted gravitational radiation literally carries the signature of the spacetime geometry around the central object. So fitting the orbit to theoretical templates could reveal small deviations of this geometry from that of Kerr. For example, if we know (from fitting the waveform) the mass and spin of the central black hole, then all its higher multipole moments are determined. If we can measure some of these and they deviate from Kerr, then that would indicate that either the central object is not a black hole or that general relativity needs to be corrected [177, 68].

Testing the post-Newtonian approximation

Current tests of general relativity rely on experiments in the solar system (using the sun's gravitational field) and observations of binary pulsars. In dimensionless units, the gravitational potential on the surface of the sun is about one part in a million and even in a binary pulsar the potential that each neutron star experiences due to its companion is no more than one part in ten thou-

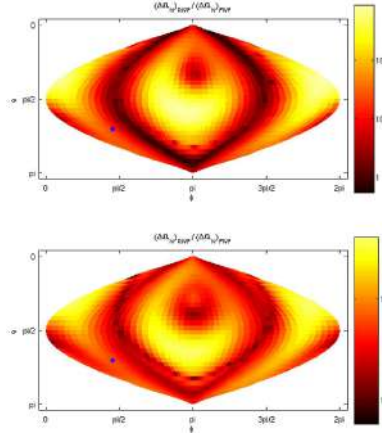


Figure 15: Sky map of the gain in angular resolution for LISA observations of the final year of inspirals using full waveforms with harmonics versus restricted post-Newtonian waveforms with only the dominant harmonic, corresponding to the equal mass case ($m_1 = m_2 = 10^7 M_\odot$, top) and a system with mass ratio of 10 ($m_1 = 10^7 M_\odot, m_2 = 10^6 M_\odot$, bottom). The sources are all at $z = 1$, have the same orientation ($\cos \theta_L = 0.2, \phi_L = 3$) and zero spins $\beta = \sigma = 0$. Figure reprinted with permission from [361]. (C) The American Physical Society. sand. These are mildly relativistic fields, with the corresponding escape velocity being as large as a thousandth and a hundredth that of light, respectively.

Thus, gravitational fields in the solar system or in a binary pulsar are still weak by comparison to the largest possible values. Indeed, close to the event horizon of a black hole, gravitational fields can get as strong as they can ever get, with the dimensionless potential being of order unity and the escape velocity equal to that of the speed of light. Although general relativity has been found to be consistent with experiments in the solar system and observations of binary pulsars, phenomena close to the event horizons of black holes would be a great challenge to the theory. It would be very exciting to test Einstein's gravity under such circumstances.

The large SNR that is expected from SMBH binaries makes it possible to test Einstein's theory under extreme conditions of gravity [45, 46]. To see how one might test the post-Newtonian structure of Einstein's theory, let us consider the waveform from a binary in the frequency domain. Since an inspiral wave's frequency changes rather slowly (adiabatic evolution) it is possible to apply a stationary phase approximation to compute the Fourier transform $H(f)$ of the waveform given in Equation (124):

$$H(f) = \mathcal{A} f^{-7/6} \exp \left[i\Psi(f) + i\frac{\pi}{4} \right]$$

with the Fourier amplitude $\mathcal{A}$ and phase $\Psi(f)$ given by

$$\mathcal{A} = \frac{\mathcal{C}}{D\pi^{2/3}} \sqrt{\frac{5\nu}{24}} M^{5/6}, \quad \Psi(f) = 2\pi f t_C + \Phi_C + \frac{3}{128\nu} \sum_k \alpha_k (\pi M f)^{(k-5)/3}$$

Here ν is the symmetric mass ratio defined before (see Equation 31), $\mathcal{C}$ is a function of the various angles, as in Equation (124), and t_C and Φ_C are the fiducial epoch of merger and the phase of the signal at that epoch, respectively. The coefficients in the PN expansion of the Fourier phase are given by

$$\begin{aligned}
\alpha_0 &= 1, \quad \alpha_1 = 0, \quad \alpha_2 = \frac{3715}{756} + \frac{55}{9}\nu, \quad \alpha_3 = -16\pi, \\
\alpha_4 &= \frac{15293365}{508032} + \frac{27145}{504}\nu + \frac{3085}{72}\nu^2, \quad \alpha_5 = \pi \left(\frac{38645}{756} - \frac{65}{9}\nu \right) [1 + \ln(6^{3/2}\pi Mf)], \\
\alpha_6 &= \frac{11583231236531}{4694215680} - \frac{640}{3}\pi^2 - \frac{6848}{21}\gamma + \left(-\frac{15737765635}{3048192} + \frac{2255}{12}\pi^2 \right) \nu \\
&\quad + \frac{76055}{1728}\nu^2 - \frac{127825}{1296}\nu^3 - \frac{6848}{63}\ln(64\pi Mf), \\
\alpha_7 &= \pi \left(\frac{77096675}{254016} + \frac{378515}{1512}\nu - \frac{74045}{756}\nu^2 \right).
\end{aligned}$$

These are the PN coefficients in Einstein's theory; in an alternative theory of gravity they will be different. In Einstein's theory the coefficients depend only on the two mass parameters, the total mass M and symmetric mass ratio ν . One of the tests we will discuss below concerns the consistency of the various coefficients. Note, in particular, that in Einstein's gravity the 0.5 PN term is absent, i.e., the coefficient of the term v is zero. Even with the very first observations of inspiral events, it should be possible to test if this is really so.

Figure 16 shows one such test that is possible with SMBH binaries [45, 46]. The observation of these systems in LISA makes it possible to measure the parameters associated with different physical effects. For example, the rate at which a signal chirps (i.e., the rate at which its frequency changes) depends on the binary's chirpmass. Given the chirpmass, the length of the signal depends on the system's symmetric mass ratio (the ratio of reduced mass to total mass). Another example would be the scattering of gravitational waves off the curved spacetime geometry of the binary, producing the tail effect in the emitted signal, which is determined principally by the system's total mass [85, 84]. Similarly, spin-orbit interaction, spin-spin coupling, etc. depend on other combinations of the masses.

The binary will be seen with a high SNR, which means that we can measure the mass parameters associated with many of these physical effects. If each parameter is known precisely, we can draw a curve corresponding to it in the space of masses. However, our observations are inevitably subject to statistical (and possibly systematic) errors. Therefore, each parameter corresponds to a region in the parameter space (shown in Figure 16 for the statistical errors only). If Einstein's theory of gravitation is correct, the regions corresponding to the different parameters must all have at least one common region, a region that contains the true parameters of the binary. This is because Einstein's theory, or an alternative, has to be used to project the observed data onto the space of masses. If the region corresponding to one or more of these parameters does not overlap with the common region of the rest of the parameters, then Einstein's theory, or its alternative, is in trouble.

In Brans-Dicke theory the system is expected to emit dipole radiation and the PN series would begin an order v^{-2} earlier than in Einstein's theory. In the notation introduced above we would have an α_{-2} term, which has the form [73, 74]

$$\alpha_{-2} = -\frac{5\mathcal{S}^2}{84\omega_{\text{BD}}}$$

Here $\mathcal{S}$ is the the difference in the scalar charges of the two bodies and ω_{BD} is the Brans-Dicke parameter. Although this term is formally two orders lower than the lowest-order quadrupole term of Einstein's gravity (i.e., it is $\mathcal{O}(v^{-2})$ order smaller), numerically its effect will be far smaller than the quadrupole term because of the rather large bound on $\omega_{\text{BD}} \gg 1$. Nevertheless, its importance lies in the fact that there is now a new parameter on which the phase depends.

Berti, Buonanno and Will conclude that LISA observations of massive black-hole binaries will enable scientists to set limits on $\omega_{\text{BD}} \sim 10^4 - 10^5$.

A massive graviton theory would also alter the phase. The dominant effect is at 1PN order and would change the coefficient α_2 to

$$\alpha_2 \rightarrow \alpha_2 - \frac{128\nu}{3} \frac{\pi^2 DM}{\lambda_g^2(1+z)}$$

where ν is the symmetric mass ratio. This term alters the time of arrival of waves of different frequencies, causing a dispersion, and a corresponding modulation, in the wave's phase, depending on the Compton wavelength λ_g and the distance D to the binary. Hence, by tracking the phase of the inspiral waves, one can bound the graviton's mass. Will [393] finds that one can bound the mass to 1.7×10^{13} km using ground-based detectors and 1.7×10^{17} km using space-based detectors, as also confirmed by more recent and exhaustive calculations [73]. These limits might improve if one takes into account the modulation of the waveform due to spin-orbit and spin-spin coupling, but the latter authors [73] looked at spinning, but nonprecessing, systems only.

Uniqueness of Kerr geometry

In Section 3 we pointed out that LISA should be able to see many hundreds of signals emitted by compact objects - black holes, neutron stars, even white dwarfs - orbiting around and being captured by massive black holes in the centers of galaxies. But for LISA to reach its full potential, we must model the orbits and their emitted radiation accurately. Since the wave trains may be many hundreds or thousands of cycles long in the LISA band, the challenge of constructing template waveforms that remain accurate to within about one radian over the whole evolution is significant.

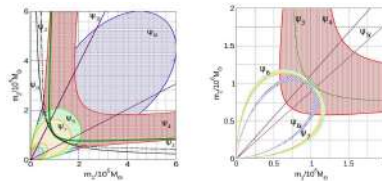


Figure 16: By fitting the Fourier transform of an observed signal to a post-Newtonian expansion, one can measure the various post-Newtonian coefficients $\psi_k(m_1, m_2)$, $k = 0, 2, 3, 4, 6, 7$ and coefficients of log-terms $\psi_{5l}(m_1, m_2)$ and $\psi_{6l}(m_1, m_2)$. In Einstein's theory, all the coefficients depend only on the two masses of the component black holes. By treating them as independent parameters one affords a test of the post-Newtonian theory. Given a measured value of a coefficient, one can draw a curve in the $m_1 - m_2$ plane. If Einstein's theory is correct, then the different curves must all intersect at one point within the allowed errors. These plots show what might be possible with LISA's observation of the merger of a binary consisting of a pair of $10^6 M_\odot$ black holes. In the right-hand plot all known post-Newtonian parameters are treated as independent, while in the left-hand plot only three parameters ψ_0, ψ_2 and one of the remaining post-Newtonian parameter are treated as independent and the procedure is repeated for each of the remaining parameters. The large SNR in LISA for SMBH binaries makes it possible to test various post-Newtonian effects, such as the tails of gravitational waves, tails of tails, the presence of log-terms, etc., associated with these parameters. Left-hand figure adapted from [46], right-hand figure reprinted with permission from [45]. (c) The American Physical Society.

The range of mass ratios is also wide. LISA's central black holes might have masses between 10^3 and $10^7 M_\odot$. Inspiring neutron stars and white dwarfs might have masses between 0.5 and $2 M_\odot$. Inspiring stellar-population black holes might be in the range of $7 - 50 M_\odot$, while intermediate mass black holes formed by the first generation of stars (Population III stars)

might have masses around $300M_{\odot}$ or even $1000M_{\odot}$. So the mass ratios might be anything in the range 10^{-7} to 1.

The techniques that must be used to compute these signals depend on the mass ratio. Ratios near one are treated by post-Newtonian methods until the objects are so close that only numerical relativity can follow their subsequent evolution. For ratios below 10^{-4} (a dividing line that is rather very uncertain and that depends on the bandwidth being used to observe the system, i.e., on how long the approximation must be valid for), systems are treated by fully-relativistic perturbation theory, expanding in the mass ratio. Intermediate mass ratios have not been studied in much detail yet; they can probably be treated by post-Newtonian methods up to a certain point, but it is not yet clear whether their final stages can be computed accurately by either numerical relativity or perturbation theory.

Post-Newtonian methods have been extensively discussed above. The basics of perturbation theory underlying this problem are treated in two Living Reviews [292, 319]. Once one has sufficiently good waveform templates, there remain the challenge of finding weak signals in LISA's noise. This depends on a number of factors, including the rates of sources. A recent study by a number of specialists [170] has concluded that the event rate is high enough and the detection methods robust enough for us to be very optimistic that LISA will detect hundreds of these sources. In fact, the opposite problem might materialize: LISA might find it has a confusion problem for the detection of these sources, as for the galactic binaries. Recent estimates of the magnitude of the problem [66] suggest that LISA's noise may at worst be raised effectively by a factor of two, but in return one gets a large number of sources of this kind.

Quantum gravity

It seems inevitable that general relativity's description of nature will one day yield to a quantumbased description, involving uncertainties in geometry and probabilities in the outcome of gravitational observations. This is one of the most active areas of research in fundamental physics today, and there are many speculations about how quantum effects might show up in gravitational wave observations.

The simplest idea might be to try to find evidence for "gravitons" directly in gravitational waves, by analogy with the way that astronomers count individual photons from astronomical sources. But this seems doomed to failure. The waves that we can observe have very low frequency, so the energy of each graviton is extremely small. And the total energy flux of the waves is, as we have seen, enormous. So the number of gravitons in a detectable gravitational wave is far more than the number of photons in the light from a distant quasar.

Quantum gravity might involve new gravity-like fields, whose effects might be seen indirectly in the inspiral signals of black holes or neutron stars, as we have noted above. String theory might lead to the production of cosmic strings, which might be observed through their gravitational wave emission [142]. If our universe is just a 4-dimensional subspace of a large-scale 10 or 11-dimensional space, then dynamics in the larger space might produce gravitational effects in our space, and in particular gravitational waves [304].

It might be possible to observe the quantum indeterminacy of geometry directly using gravitational wave detectors, if Hogan's principle of holographic indeterminacy is valid [193]. Hogan speculates that quantum geometry might be manifested by an uncertainty in the position of a beam splitter, and that this could be the explanation for an unexpectedly large amount of noise at low frequencies in the GEO600 detector. In this connection it is interesting to construct, from fundamental constants alone, a quantity with the dimensions of amplitude spectral noise density $(S_h)^{1/2}$. This has units of $s^{1/2}$, so one can define the "Planck noise power" $S_{Pl} = t_{Pl} = (G\hbar/c^5)^{1/2}$. Then the amplitude noise is $S_{Pl}^{1/2} = (G\hbar/c^5)^{1/4} = 2.3 \times 10^{-22} \text{ Hz}^{-1/2}$. This is

comparable to or larger than the instrumental noise in current interferometric gravitational wave detectors, as shown in Figure 5. This in itself does not mean that Planckian noise will show up in gravitational wave detectors, but Hogan argues that the particular design of GEO600 might indeed make it subject to this noise more strongly than other large interferometers.

5.5 Astrophysics with Gravitational Waves

Gravitational radiation plays an observable role in the dynamics of many known astronomical systems. In some, such as cataclysmic variables [157] and neutron-star-binary systems [355], the role of gravitational radiation has been understood for years. In others, such as young neutron stars [34] and low-mass X-ray binaries [77], the potential importance of gravitational radiation has been understood only recently. As further observations, particularly at X-ray wavelengths, become available, the usefulness of gravitational radiation as a tool for modelling astronomical systems should increase [385].

At this point in the progress of gravitational wave detection, the greatest emphasis in calculations of sources is on prediction: trying to anticipate what might be seen. Not only is this important in motivating the construction of detectors, but it also guides details of their design and, very importantly, the design of data analysis methods. Historically, many predictions of emission strengths and the capability of detectors to extract information from signals have relied on estimates using the quadrupole formula. This was justifiable because, given the uncertainties in our astrophysical understanding of potential sources, more accurate calculations would be unjustified in most cases.

But these rough estimates are now being replaced by more and more detailed source models where possible. This applies particularly in two cases. One is binary orbits, where the pointmass approximation is good over a large range of observable frequencies, so that fully relativistic calculations (using the post-Newtonian methods described above) are not only possible, but are necessary for the construction of sensitive search templates in the data analysis. The second exception is the numerical simulations of the merger of black holes and neutron stars, where the dynamics is so complex that none of our analytic approximations offers us reliable guidance. In fact, these two methods are currently being joined to produce uniform models of signal evolution over as long an observation time as the signal allows. From these models we not only improve detection algorithms, but we also learn much more about the kinds of information that detections will extract from the signals.

Once gravitational waves have been observed, there will of course be a welcome shift of emphasis to include interpretation. The emphasis will be on extracting observable parameters (waveforms, polarizations, source location, etc.) from noisy data or data where (in the case of LISA) there is source confusion. These issues need considerably more attention than they have so far received.

5.5.1 Interacting compact binaries

The first example of the use of gravitational radiation in modelling an observed astronomical system was the explanation by Faulkner [157] of how the activity of cataclysmic binary systems is regulated. Such systems, which include many novae, involve accretion by a white dwarf from a companion star. Unlike accretion onto neutron stars, where the accreted hydrogen is normally processed quickly into heavier elements, on a white dwarf the unprocessed material can build up until there is a nuclear chain reaction, which results in an outburst of visible radiation from the system.

Now, in a circular binary system that conserves total mass and angular momentum, a transfer of mass from a more massive to a less massive star will make the orbit shrink, while a

transfer in the opposite direction makes the orbit grow. If accretion onto a white dwarf begins with the dwarf as the less massive star, then the stars will draw together, and the accretion will get stronger. This runaway process stops when the stars are of equal mass, and then accretion begins to drive them apart again. Astronomers observed that in this phase accretion in certain very close binaries continued at a more or less steady rate, instead of shutting off as the stars separated more and more. Faulkner pointed out that gravitational radiation from the orbital motion would carry away angular momentum and drive the stars together. The two effects together result in steady accretion at a rate that can be predicted from the quadrupole formula and simple Newtonian orbital dynamics, and which is in good accord with observations in a number of systems.

Binaries consisting of two white dwarfs in very tight orbits will be direct LISA sources: we won't have to infer their radiation indirectly, but will actually be able to detect it. Some of them will also be close enough to tidally interact with one another, leading in some cases to mass transfer. Others will be relatively clean systems in which the dominant effect will be gravitational radiation reaction.

Observations during the last decade have identified a number of such systems with enough confidence to predict that LISA should see their gravitational waves. These are called verification binaries: if LISA does not see them then either the instrument is not working properly or general relativity is wrong! For a review of verification binaries, see [349].

Resolving the mass-inclination degeneracy

Gravitational-wave-polarization measurements can be very helpful in resolving the degeneracy that occurs in the measurement of the mass and inclination of a binary system. As is well known, astronomical observations of binaries cannot yield the total mass but only the combination $m \sin \iota$, where ι is the inclination of the binary's orbit to the line of sight. However, measurement of polarization can determine the angle ι since the polarization state depends on the binary's inclination with the line of sight.

For instance, consider a circular binary system with total mass M at a distance D . Suppose its orbital angular momentum vector makes an angle ι with the line of sight (the standard definition of the inclination of a binary orbit). The two observed polarizations are given in the quadrupole approximation by Equation (122). We can eliminate the distance R between the stars that is implicit in the velocity $v = R\omega$ (where ω is the instantaneous angular velocity of the orbit, the derivative of the orbital phase function $\varphi(t)$) by using the Newtonian orbital dynamics equation $\omega^2 = M/R^3$. Then we find

$$h_+ = \frac{2\nu M}{D} [\pi M f(t)]^{2/3} (1 + \cos^2 \iota) \cos[2\varphi(t)], \quad h_\times = \frac{4\nu M}{D} [\pi M f(t)]^{2/3} \cos \iota \sin[2\varphi(t)]$$

where M is the total mass of the binary and, as before, ν is the symmetric mass ratio $m_1 m_2 / M^2$. The frequency $f = \omega / \pi$ is the gravitational wave frequency, twice the orbital frequency. Notice that, consistent with Equation (30), the masses of the stars enter these equations only in the combination $\mathcal{M} = \nu^{3/5} M$.

It is clear that the ratio of the two polarization amplitudes determines the angle ι . In this connection it is interesting to relate the polarization to the orientation. When the binary is viewed from a point in its orbital plane, so that $\iota = \pi/2$, then $h_\times = 0$; the radiation has pure $+$ polarization. From the observer's point of view, the motion of the binary stars projected onto the sky is purely linear; the two stars simply go back and forth across the line of sight. This linear projected motion results in linearly polarized waves. At the other extreme, consider viewing the system down its orbital rotation axis, where $\iota = 0$. The stars execute a circular motion in the sky, and the polarization components h_+ and $h_\times$ have equal amplitude and are out

of phase by $\pi/2$. This is circularly polarized gravitational radiation. So, when the radiation is produced in the quadrupole approximation, the polarization has a very direct relationship to the motions of the masses when projected on the observer's sky plane. If the radiation is produced by higher multipoles it becomes more complex to make these relations, but it can be done. For example, see [333] for the case of current quadrupole radiation, which is emitted by the r-mode instability discussed in Section 7.3.4.2 below.

While a single detector is linearly polarized, it can still measure the two polarizations if the signal has a long enough duration for the detector to turn (due to the motion of the Earth) and change the polarization component it measures. Alternatively, a network of three detectors can determine the polarization and location of the source even over short observation times.

Such a measurement would lead to a potentially very interesting interplay between gravitational and electromagnetic observations, with applications in the observations of isolated neutrons stars, binary systems, etc. And would lead to synergies, for example, between the LISA and Gaia [289] missions.

5.5.2 Black hole astrophysics

Black holes are the most relativistic systems possible. Observing gravitational waves from them, individually or in binaries, helps to test some of the predictions of general relativity in the strongly nonlinear regime, such as the tails of gravitational waves, spin-orbit coupling induced precession, nonlinear amplitude terms, hereditary effects, etc [360, 84, 83, 322]. They are also good test beds to constrain other theories of gravity. Gravitational waves - emitted either during the inspiral and merger of rotating SMBHs or when a stellar-mass compact object falls into a SMBH - can be used to map the structure of spacetime and test uniqueness theorems on rotating black holes [360]. LISA will be able to see the formation of massive black holes at cosmological distances by detecting the waves emitted in the process [360]. We give below a brief discussion of the physics that will follow from the observation of gravitational waves from black holes.

Gravitational waves from stellar-mass black holes

Astronomers now recognize that there is an abundance of black holes in the universe. Observations across the electromagnetic spectrum have located black holes in X-ray binary systems in our galaxy in the centers of star clusters, and in the centers of galaxies.

These three classes of black holes have very different masses. Stellar black holes typically have masses of around $10M_{\odot}$, and are thought to have been formed by the gravitational collapse of the center of a large, evolved red giant star, perhaps in a supernova explosion. Black holes in clusters have been found in the range of $10^4M_{\odot}$, and are called intermediate-mass black holes. Black holes in galactic centers have masses between 10^6 and $10^{10}M_{\odot}$, and are called SMBHs. The higher masses are found in the centers of active galaxies and quasars. The history and method of formation of intermediate-mass and supermassive black holes are not yet well understood.

All three kinds of black hole can radiate gravitational waves. According to Figure 2, stellar black-hole radiation will be in the ground-based frequency range, while galactic holes are detectable only from space. Intermediate-mass black holes may lie at the upper end of the LISA band or between LISA and ground-based detectors. The radiation from an excited black hole itself is strongly damped, lasting only a few cycles at its natural frequency [see Equation (12) with $R = 2M$]:

$$f_{BH} \sim 1000 \left(\frac{M}{10M_{\odot}} \right)^{-1} \text{ Hz}$$

Stellar-mass black-hole binaries

Radiation from stellar-mass black holes is expected mainly from coalescing binary systems, when one or both of the components is a black hole. Although black holes are formed more rarely than neutron stars, the spatial abundance of binary systems consisting of neutron stars with black holes, or of two black holes, is amplified relative to neutron-star binaries because binary systems are much more easily broken up when a neutron star forms than when a black hole forms. When a neutron star forms, most of the progenitor star's mass ($6M_{\odot}$ or more) must be expelled from the system rapidly. This typically unbinds the binary: the companion star has the same speed as before but is held to the neutron star by only a fraction of the original gravitational attraction. Observed neutron-star binaries are thought to have survived because the neutron star was coincidentally given a kick against its orbital velocity when it formed. When a black hole forms, most of the original mass may simply go down into the hole, and the binary will have a higher survival probability. However, this argument may not lead to observable black hole binaries; there is a possibility that systems that would form black holes close enough to coalesce in a Hubble time do not become binaries, but rather the two progenitor stars are so close that they merge before forming black holes.

On the other hand, double black-hole binaries may in fact be formed abundantly by capture processes in globular clusters, which appear to be efficient factories for black-hole binaries [295]. Being more massive than the average star in a globular cluster, black holes sink towards the center, where three-body interactions can lead to the formation of binaries. The key point is that these binaries are not strongly bound to the cluster, so they can easily be expelled by later encounters. From that point on they evolve in isolation, and typically have a lifetime shorter than 10^{10} yrs.

The larger mass of stellar black-hole systems makes them visible from a greater distance than neutron-star binaries. If the abundance of binaries with black holes is comparable to that of neutron-star binaries, black hole events will be detected much more frequently than those involving neutron stars. They may even be seen by first-generation detectors in the S5 science run of the LSC (see Section 4.3.1), although that is still not very probable, even with optimistic estimates of the black-hole binary population. It seems very possible, however, that the first observations of binaries by interferometers will eventually be of black holes.

More speculatively, black hole binaries may even be part of the dark matter of the universe. Observations of Massive Compact Halo Objects (MACHOs) - microlensing of distant stars by compact objects in the halo of our galaxy - have indicated that up to half of the galactic halo could be made up of dark compact objects of $0.5M_{\odot}$ [25, 351]. This is difficult to understand in terms of stellar evolution, as we understand it today: neutron stars and black holes should be more massive than this, and white dwarfs of this mass should be bright enough to have been identified as the lensing objects. One speculative possibility is that the objects were formed primordially, when conditions may have allowed black holes of this mass to form. If so, there should also be a population of binaries among them, and occasional coalescences should, therefore, be expected. In fact, the abundance would be so high that the coalescence rate might be as large as one every 20 years in each galaxy, which is higher than the supernova rate. Since binaries are maximally non-axisymmetric, these systems could be easily detected by first-generation interferometers out to the distance of the Virgo Cluster [265].

The estimates used here of detectability of black hole systems depend mainly on the radiation emitted as the orbit decays, during which the point-particle post-Newtonian approximation should be adequate. But the inspiral phase will, of course, be followed by a burst of gravitational radiation from the merger of the black holes that will depend in detail on the masses and spins of the objects. Numerical simulations of such events will be used to interpret this signal and to provide templates for the detection of black holes too massive for their inspiral signals to be seen. There is an abundance of information in these signals: population studies of the masses

and spins of black holes, studies of typical kick velocities for realistic mergers, tests of general relativity.

Intermediate-mass black holes

Intermediate-mass black holes, with masses between $100M_{\odot}$ and $10^4M_{\odot}$, are expected on general evolutionary grounds, but have proved hard to identify because of their weaker effect on surrounding stellar motions. Very recently [275] strong evidence has been found for such a black hole in the star cluster Omega Centauri. If such black holes are reasonably abundant, then they may be LISA sources when they capture a stellar-mass black hole or a neutron star from the surrounding cluster.

For these merger events the mass ratio is not as extreme as for EMRIs, and so these are accordingly called IMRIs: Intermediate Mass-Ratio Inspirals.

The problem of modelling the signals from these systems has not yet been fully studied. If these signals can be detected, they will tell us how important black holes were in the early stellar population, and whether these black holes have anything to do with the central black holes in the same galaxies.

Supermassive black holes

Gravitational radiation is expected from SMBHs in two ways. In one scenario, two massive black holes spiral together in a much more powerful version of the coalescence we have just discussed. The frequency is much lower, in inverse proportion to their masses, and the amplitude is higher. Equation (128) implies that the effective signal amplitude (which is what appears in the expression for the SNR) is almost linear in the masses of the black holes, so that a signal from two $10^6M_{\odot}$ black holes will have an amplitude 10^5 times bigger than the signal from two $10M_{\odot}$ holes at the same distance. Even allowing for differences in technology, this indicates why space-based detectors will be able to study such events with a very high SNR, no matter where in the universe they occur. Observations of coalescing massive black-hole binaries will therefore provide unique insight into the behavior of strong gravitational fields in general relativity.

The event rate for such coalescences is not easy to predict, but is likely to be large. It seems that the central core of most galaxies may contain a black hole of at least $10^6M_{\odot}$. This is known to be true for our galaxy [151] and for a very large proportion of other galaxies that are near enough to be studied in sufficient detail [312]. SMBHs (up to a few times $10^9M_{\odot}$) are believed to power quasars and active galaxies, and there is a good correlation between the mass of the central black hole and the velocity dispersion of stars in the core of the host galaxy [174].

If black holes are formed with their galaxies, in a single spherical gravitational collapse event, and if nothing happens to them after that, then coalescences will never be seen. But this is unlikely for two reasons. First, it is believed that galaxies may have formed through the merger of smaller units, sub-galaxies of masses upwards of $10^6M_{\odot}$. If these units had their own black holes, then the mergers would have resulted in the coalescences of many of the black holes on a timescale shorter than the present age of the universe. This would give an event rate of several mergers per year in the universe, most of which would be observable by LISA, if the more massive black hole is not larger than about $10^7M_{\odot}$. If the $10^6M_{\odot}$ black holes were formed from smaller black holes in a hierarchical merger scenario, then the event rate could be hundreds or thousands per year. The second reason is that we see large galaxies merging frequently. Interacting galaxies are common, and if galaxies come together in such a way that their central black holes both remain in the central core, then dynamical friction with other stars will bring them close enough together to allow gravitational radiation to bring about a merger on a timescale of less than 10^{10} yrs. There is considerable evidence for black hole binaries in a number of external galaxies [255]. There is even a recent report of an SMBH having been

ejected from a galaxy, possibly by the kick following a merger [222] and of an SMBH binary that will coalesce in about 10,000 yrs [373]!

Besides mergers of holes with comparable masses, the capture of a small compact object by a massive black hole can also result in observable radiation. The tidal disruption of main-sequence or giant stars that stray too close to the black hole is thought to provide the gas that powers the quasar phenomenon. These disruptions are not expected to produce observable radiation. But the clusters will also contain a good number of neutron stars and stellar-mass black holes. They are too compact to be disrupted by the black hole, even if they fall directly into it.

Such captures, therefore, emit a gravitational wave signal that will be well approximated as that from a point mass near the black hole. This will again be a chirp of radiation, but in this case the orbit may be highly eccentric. The details of the waveform encode information about the geometry of spacetime near the black hole. In particular, it may be possible to measure the mass and spin of the black hole and thereby to test the uniqueness theorem for black holes. The event rate is not very dependent on the details of galaxy formation, and is probably high enough for many detections per year from a space-based detector [32], provided that theoretical calculations give data analysts accurate predictions of the motion of these point particles over many hundreds of thousands of orbits. These Extreme Mass-Ratio Inspiral sources (EMRIs) are a primary goal of the LISA detector. By observing them, LISA will provide information about the stellar population near central black holes. When combined with modelling and spectroscopic observations, this will facilitate a deep view of the centers of galaxies and their evolution.

5.5.3 Neutron star astrophysics

Gravitational collapse and the formation of neutron stars

The event that forms most neutron stars is the gravitational collapse that results in a supernova. It is difficult to predict the waveform or amplitude expected from this event. Although detecting this radiation has been a goal of detector development for decades, little more is known about what to expect than 30 years ago. The burst might be at any frequency between 100 Hz and 1kHz, and it might be a regular chirp (from a rotating deformed core) or a more chaotic signal (from convective motions in the core). Considerable energy is released by a collapse, and on simple energetic grounds this source could produce strong radiation: if the emitted energy is more than about $0.01M_{\odot}$, then second-generation detectors would have no trouble seeing events that occur in the Virgo Cluster. This energetic consideration drove the early development of bar detectors.

But numerical simulations tell a different story, and it seems very likely that radiation amplitudes will be much smaller, as described in Section 3. Such signals might be detectable by second-generation detectors from a supernova in our galaxy, but not from much greater distances. When they are finally detected, the gravitational waves will be extremely interesting, providing our only information about the dynamics inside the collapse, and helping to determine the equation of state of hot nuclear matter.

If gravitational collapse forms a neutron star spinning very rapidly, then it may be followed by a relatively long period (perhaps a year) of emission of nearly monochromatic gravitational radiation, as the r-mode instability (Section 7.3.4) forces the star to spin down to speeds of about 100 – 200 Hz [278]. If as few as 10% of all the neutron stars formed since star formation began (at a redshift of perhaps four) went through such a spindown, then they may have produced a detectable random background of gravitational radiation at frequencies down to 20 Hz [324].

Neutron-star-binary mergers

When two neutron stars merge, they will almost certainly have too much mass to remain as a star, and will eventually collapse to a black hole, unless they can somehow expel a significant amount of mass. The collision heats up the nuclear matter to a point where, at least initially, thermal pressure becomes significant. Numerical simulations can use theoretical equations of state (such as that of Lattimer and Swesty [233]) to predict the merger radiation, and observations will then test the nuclear physics assumptions that go into the equation of state. Simulations show that the choice of equation of state makes a big difference to the emitted waveform, as do the masses of the stars: there is no mass scaling as there is for black holes [60].

When a neutron star encounters a black hole in a stellar compact binary merger, the star may not be heated very much by the tidal forces, and the dynamics may be governed by the cold nuclearmatter equation of state, about which there is great uncertainty. Again, comparing observed with predicted waveforms may provide some insight into this equation of state. Simulations suggest that these systems may give rise to many of the observed short, hard gamma-ray bursts [155, 336].

Simultaneous gravitational wave and gamma ray detections would settle the issue and open the way to more detailed modeling of these systems.

Neutron-star normal mode oscillations

Gravitational wave observations at high frequencies of neutron-star vibrations may also constrain the cold-matter equation of state. In Figure 2 there is a dot for the typical neutron star. The corresponding frequency is the fundamental vibrational frequency of such an object. In fact, neutron stars have a rich spectrum of nonradial normal modes, which fall into several families: f , g , p , w , and r -modes have all been studied. These have been reviewed by Andersson and Comer [35]. If their gravitational wave emissions can be detected, then the details of their spectra would be a sensitive probe of their structure and of the equation of state of neutron stars, in much the same way that helioseismology probes the interior of the sun. Even knowing accurately the frequency and decay time of just the fundamental $\ell = 2$ -mode would be enough to eliminate most current equations of state [36].

This is a challenge to ground-based interferometers, which have so far focussed their efforts on frequencies below 1kHz. But Advanced LIGO and the upgraded GEO-HF detector (Section 4.3.1) may have the capability to perform narrow-banding and enhance their sensitivity considerably at frequencies up to perhaps 2kHz, which could put the f -modes of neutron stars into range.

The f -modes of neutron stars, which could be excited by glitches or by the nuclear explosions on accreting neutron stars that are thought to produce X-ray flares and soft gamma-ray repeater events. The rise-time of X-ray emission can be as short as a few milliseconds [173], which might be impulsive enough to excite acoustic vibrations. If the rise time of the explosion matches the period of the mode well enough, then a substantial fraction of the energy released could go into mechanical vibration, and almost all of this fraction would be carried away by gravitational waves, since other mode-damping mechanisms inside neutron stars are much less efficient.

Radio-pulsar glitches seem to release energies of order 10^{35} J, and X-ray and gamma ray events can be much more energetic. Using Equation (20), we can estimate that the release of that much energy into gravitational waves at 2kHz at a distance of 1kpc would create a wave of effective amplitude around 3×10^{-22} . (The effective amplitude assumes we can do matched filtering, which in this case is not very difficult.) This kind of amplitude should be within the reach of Advanced LIGO (Figure 5) and perhaps GEO-HF, provided they implement narrowbanding. This will not be easy, either scientifically or operationally, but the payoff in terms of our understanding of neutron star physics could be very substantial.

Observations of these modes would immediately constrain the cold-matter nuclear equation of state in significant ways [36, 35].

In fact, modes of neutron stars may have already been observed in X-rays [386]. But these are likely to be crustal modes, whose restoring force is the shear strength of the crust. While the physics of the crust is interesting in itself, such observations provide only weak constraints on the interior physics of the neutron star.

Stellar instabilities

7.3.4.1 The CFS instability. In 1971 Chandrasekhar [111] applied the quadrupole formula to calculate the corrections to the eigenfrequencies of the normal mode vibrations of rotating stars, and he found to his surprise that some modes were made unstable, i.e., that coupling to gravitational radiation could destabilize a rotating star. Subsequent work by Friedman and Schutz [166] showed that there was a key signature for the mode of a Newtonian star that would be unstable in general relativity. This was the pattern speed of the mode, i.e., the angular velocity at which the crests of the pattern rotated about the rotation axis of the star. If this speed was in the same sense as the rotation of the star, but slower than the star, then the mode would be unstable in a perfect-fluid star. This instability has come to be known as the CFS instability, after the three authors who explained it.

The basic theory was developed for perfect-fluid stars. However, Lindblom and Detweiler [238] showed that the effect of viscosity ran counter to that of radiation reaction, so that the instability was strongest in modes with the longest wavelengths, i.e., in the quadrupolar modes. Full numerical calculations on Newtonian stellar models with realistic viscosity models showed [239] that the standard fundamental and acoustic modes of rotating neutron stars were not vulnerable to this instability. Subsequent work on fully relativistic models [347] has hinted that the instability may be stronger than the Newtonian models indicate, but it is still at the margins of astrophysical interest.

7.3.4.2 The r-mode instability. The situation changed in 1997 when Andersson [34] pointed out that there is another class of modes of Newtonian stars that should be unstable in the same way, but which had not been studied in this context before, the Rossby or r-modes. These are momentum-dominated modes, where the gravitational radiation comes from the current quadrupole terms, rather than from the mass quadrupole. Investigations by a number of authors [241, 37, 278] have shown that this instability could be very strong in hot, rapidly-rotating stars. This is particularly relevant to young neutron stars, which may well be formed with rapid spin and which will certainly be hot. For their first year, stars spinning faster than about 100 Hz could spin down to about 100 Hz by losing angular momentum to gravitational radiation. The instability might also operate in old accreting neutron stars, such as those in LMXB X-ray binaries (see the next section).

However, the instability is, like other CFS instabilities, sensitive to viscosity, and there is great uncertainty about the amount of viscosity inside neutron stars [240, 228, 35].

Low-mass X-ray binaries

Observations by the Rossi satellite (RXTE) have given evidence that the class of X-ray sources called Low-Mass X-ray Binaries (LMXB's) contains neutron stars with a remarkably narrow range of spins, between perhaps 250 Hz and 320 Hz [376]. These are systems in which it is believed that neutron stars are spun up from the low angular velocities they have after their lifetime as normal pulsars to the high spins that millisecond pulsars have. One would expect, therefore, that the spins of neutron stars in such systems would be spread over a wide range. The fact that they are not requires an explanation.

The most viable explanation offered so far is the suggestion of Bildsten [77] that gravitational

radiation limits the rotation rate. The proposed mechanism is that anisotropic accretion onto the star creates a temperature gradient in the crust of the neutron star, which in turn creates a gradient in the mass of the nucleus that is in local equilibrium, and this in turn creates a density gradient that leads, via the rotation of the star, to the emission of gravitational radiation. This radiation carries away angular momentum, balancing that which is accreted, so that the star remains at an approximately constant speed.

According to the model, the gravitational wave luminosity of the star is proportional to the measured flux of X-rays, since the X-ray flux is itself proportional to the accreted angular momentum that has to be carried away by the gravitational waves. If this model is correct, then the X-ray source Sco X-1 might be marginally detectable by advanced interferometers, and other similar systems could also be candidates [385].

Galactic population of neutron stars

Neutron stars are known to astronomy through the pulsar phenomenon. As radio surveys improve, the number of known pulsars is pushing up toward 2000. There is a public catalogue on the web [55]. But the galactic population of neutron stars is orders of magnitude larger, perhaps as many as 10^8 . Most are much older than typical pulsars, which seem to stop emitting after a few million years. X-ray surveys reveal a number of unidentified point sources, which might be hot neutron stars, but older neutron stars are probably not even hot enough to show up in such surveys.

Gravitational wave observations have the potential to discover more neutron stars, but in the foreseeable future the numbers will not be large. Spinning neutron stars can be found in searches for continuous-wave signals, but there is no a priori reason to expect significant deformations that would lead to large gravitational wave amplitudes. One mechanism, proposed by Cutler [127], is that a large buried toroidal magnetic field could, by pulling in the waist of a spinning star, turn it into a prolate spheroid. This is classically unstable and would tip over and spin about a short axis, emitting gravitational waves. Millisecond pulsars could, in principle, be spinning down through the emission of gravitational waves in this way. Only deep observations by Advanced LIGO could begin to probe this possibility.

In fact, strong emission of gravitational waves is in some sense counterproductive, since it causes a neutron star to spin down and move out of the observing band quickly. This places important limits on the likely distribution of observable continuous-wave amplitudes from neutron stars [219]. This is important input into the blind searches for such signals being conducted by the LSC.

Radio observations of pulsars have, of course, revealed a fascinating population of binary systems containing neutron stars, including the original Hulse-Taylor pulsar [201] and the double pulsar PSR J0737-3039 [246]. But radio surveys only cover a small fraction of our galaxy, so there may be many more interesting and exotic systems waiting to be discovered, including neutron stars orbiting black holes. In fact, not all neutron stars are pulsars, so there are likely to be nearby binary systems containing neutron stars that are not known as pulsars at all.

LISA has enough sensitivity to detect all such binaries in the galaxy whose gravitational wave emission is above 1mHz, i.e., with orbital periods shorter than half an hour. Below that frequency, systems may just blend into the confusion noise of the white-dwarf background, unless they are particularly close. The Hulse-Taylor system is a bit below the LISA band, and even its higher harmonics are likely to be masked by the dense confusion noise of white-dwarf galaxies at low frequencies. Double pulsars should be detectable by LISA with low SNR (around five in five years) above the confusion background at a frequency of 0.2mHz [210]. In all, LISA might detect several tens or even hundreds of double neutron-star systems, and potentially even a handful of double black hole binaries.

Neutron stars are the fossils of massive stars, and so a population census of binaries can help

normalize our galaxy's star-formation rate in the past. The mass distribution of such systems will also be of interest: do all neutron-star binaries have stars whose masses are near $1.4M_{\odot}$, or is this only true of systems that become pulsars? LISA observations are likely to illuminate many puzzles of stellar evolution.

Finally, it is possible to search for gravitational waves from individual spinning neutron stars in binary systems. Although more rare than isolated neutron stars, these systems might have a different history and a different distribution of amplitudes. Searches are planned by the LSC, but they are difficult to do, since the parameter space is even larger than for isolated pulsars.

5.5.4 Multimessenger gravitational-wave astronomy

Multimessenger gravitational-wave astronomy refers to coordinated observations using different kinds of radiation and information carriers: electromagnetic, neutrino, cosmic ray, and gravitational wave. Joint coordinated observing has much to offer gravitational wave detection, by allowing it to target known interesting sources or locations, thereby reducing the parameter space that must be searched and improving the confidence of a detection. Even more importantly, the information obtained from gravitational wave observations is typically complementary to that which one can get from electromagnetic astronomy, and so there are big science gains to be realized from coordinated observations.

One can distinguish three broad classes of coordinated observations: triggered gravitational-wave searches, follow-up electromagnetic observations, and parameter refinement.

- Triggered searches use transient electromagnetic events, such as gamma-ray bursts, to narrow down the window of time for a search in the gravitational-wave data stream, and possibly also to restrict the ranges of various parameters. Since gravitational-wave-signal data is recorded, it is no problem to go back to data at the time of the triggering event and search it. This helps to lower the detection threshold, since gravitational wave events need, in this case, to be significant over a time scale of a few minutes rather than, say, an entire year.
- Follow-up searches use a, possibly tentative, gravitational-wave detection to mark an area in the sky and a timeframe for an electromagnetic search. A very interesting example of this will occur with LISA, which will be able to predict the location and time of the coalescence of two SMBHs with reasonable accuracy at least a week in advance. This will allow sufficient time for telescopes in all the electromagnetic wavelengths to prepare to observe the event.
- Parameter refinement refers to the use of electromagnetic observations of potential gravitational wave sources to improve the values of the parameters that must be used in the gravitational wave search. This has already been used in LSC searches, for example in trying to detect radiation from known radio pulsars: radio observations during the gravitational-wave observation period were used to track the changing frequency of the pulsar [11].

Finding electromagnetic counterparts to gravitational wave observations is important, of course, for learning about the nature of the events. But it has a more subtle benefit: it generally improves significantly the accuracy with which parameters can be estimated from the gravitational wave observation. The reason is that one of the biggest sources of parameter uncertainty is the sky location of a gravitational wave source. Interferometers have broad antenna patterns, which is helpful in that they can monitor essentially the entire sky continuously, but

which means that directional information for transient events can come only from time delay information among different detectors. The simple Rayleigh limit λ_{gw}/D for ground-based interferometers gives angular accuracies on the order of several degrees, divided by the amplitude SNR (never smaller than 5 for any reasonable detection). The covariance of angular errors with uncertainties in other parameters (distance, polarization, stellar masses, etc) is usually significant. Therefore, if a follow-up electromagnetic observation can provide a more accurate position, this can also improve the determination of all the other parameters measured gravitationally.

Triggered searches are already being performed by the LSC for gravitational waves associated with gamma-ray bursts [14]. The nondetection of any gravitational waves associated with the gamma-ray burst GRB 070201 showed that it was not created by the merger of neutron stars in the nearby galaxy M31, despite its positional coincidence on the sky [16]. In addition, the gravitational wave detectors are monitoring the triggers provided by both high-energy and lowenergy neutrino detectors in order to get instant warning of a supernova in our galaxy or of some more exotic event further away. As we have noted above, X-ray flares from neutron stars may signal normal-mode radiation from acoustic vibrations.

Triggers may also allow the first detection of gravitational waves from the normal modes of neutron stars, which as mentioned in Section 7.3.3, would provide our first "view" inside these exotic objects. These triggers could be radio-pulsar glitches, X-ray flares, or even the formation and subsequent ringdown of a neutron star.

Follow-up observations of neutron-star-binary coalescence events are likely to be particularly informative. It is possible that these events are associated with short gamma-ray bursts, in which case most events are missed because of the narrow beaming of the gamma rays. Gravitational waves, by contrast, are emitted nearly isotropically, so that they will pick out essentially all such events within the range of the detectors, and astronomers can subsequently search for afterglows and prompt X- and gamma-ray emission. The ability to study such events from all aspect angles will help model them reliably. Even if coalescences are not associated with gamma-ray bursts, it is difficult to imagine that they will not produce visible afterglows or other transient electromagnetic events that would presumably not have been recognized before. The same considerations apply to coalescences of neutron stars and black holes.

Gravitational wave events may also provide our first notice of a gravitational collapse event, if the event is a strong radiator and is too far away for neutrino detectors to see it. While supernova simulations generally suggest that the amplitude of emitted gravitational waves is small [147], numerical simulations of the aftermath of neutron-star coalescence suggest the possibility of very powerful gravitational-wave emission [60]. While this event seems to lead inevitably to a black hole, because the total mass is too large for a single neutron star, neutron stars might occasionally be formed in this way by mergers of white dwarfs, again with strong rotation and the possibility of the emission of strong gravitational radiation. In this connection the suggestion of Arons [43] that at least some magnetars are formed in events of some kind that involve strong magnetic field braking but also strong gravitational wave emission, and that these events are the source of the ultra-high-energy cosmic rays whose source, is so far unexplained [384].

LISA offers particularly interesting opportunities for follow-up observations with electromagnetic waves, beyond the direct monitoring of the merger events for SMBHs mentioned above. Because SMBHs often carry accretion disks, the merger event may be followed by the turning on of accretion after a delay of, perhaps, a year or so [257]. The merger may also cause a prompt shock in surrounding gas, due to the essentially instantaneous loss of several percent of the gravity of the central mass. These or similar effects may make it possible to identify the galaxies that host LISA mergers, which in turn will allow one to associate a redshift with the luminosity distance that the gravitational wave event provides. This will be important for

LISA's cosmographic capabilities (next section).

LISA will look for close white-dwarf binaries in our galaxy and will probably see thousands of them. White-dwarf binaries never reach the last stable orbit, which would occur at roughly 1.5kHz for these masses. Instead they undergo a tidal interaction and can either disrupt at much lower frequency or end up as AM CVns (see, for instance, [341, 271]). In the latter case, we have a close white-dwarf binary with orbital periods of minutes or hours, wherein the smaller of the two stars transfers mass to the more massive one. This mass loss leads to an increase in the orbital period as a result of redistribution of the angular momentum. So far only a handful of AM CVn systems are known. LISA could potentially discover a lot more of these as their orbital periods are right in the heart of LISA's sensitivity band and simultaneous observation of these systems in the gravitational and electromagnetic window has huge impacts on the science we can learn about these end products of stellar evolution and their eventual fate.

For each resolved white-dwarf binary LISA can determine the orbital period and the spatial orientation of the orbit, and it can give a relatively crude position. If the orbit is seen to decay during the observation, LISA can determine the distance to the binary. If the binary is known from optical or X-ray observations, then this can be very valuable additional information about the system, again complementary to that which is normally available from the electromagnetic observation. Even for systems that have not been identified, LISA's census of white-dwarf binaries will provide important statistics (on the mass function, distribution of separations, etc) that should lead to a better understanding of white-dwarf and binary evolution.

In the near term, one of the most practical applications of multimessenger astronomy is to use electromagnetic observations to refine the values of key search parameters for the gravitational wave data analysis. This has been extensively discussed for possible observations of low-mass

X-ray binaries, as described in Section 7.3.5. Watts et al. [385] surveyed the known ranges of parameters, such as spin rates and orbital parameters, and concluded that they need to be narrowed considerably if a practical search were to be possible, not just because of the computer power required, but more importantly because of the loss of significance if too large a parameter space has to be searched.

6 Typical models by Michele Maggiore

6.0.1 The geometric approach to GWs

In this chapter we discuss how gravitational waves (GWs) emerge from general relativity, and what their properties are. The most straightforward approach, pursued in Sections 1.1-1.3, is "linearized theory", and consists of expanding the Einstein equations around the flat Minkowski metric $\eta_{\mu\nu}$. This allows us to see immediately how a wave equation emerges (Section 1.1) and how the solutions can be put in an especially simple form by an appropriate gauge choice (Section 1.2); then, using standard tools of general relativity such as the geodesic equation and the equation of the geodesic deviation, we can study how these waves interact with a detector, idealized for the moment as a set of test masses (Section 1.3).

We next turn to the issue of what is the energy and momentum carried by GWs. Historically, this is a subject that has been surrounded by much confusion, to the extent that for a long time even the existence of physical effects associated with GWs was considered dubious. The heart of the problem is that general relativity has a huge local gauge invariance, the invariance under arbitrary coordinate transformations, and one can easily fall into the mistake of believing that the effect of GWs can be "gauged away", i.e. set to zero with an appropriate coordinate transformation. We will discuss these issues in details, paying special attention to

the conceptual aspects that are hidden behind the derivations. In this chapter we will approach the problem from a geometric point of view, identifying the energy-momentum tensor of GWs from their effect on the background curvature. This approach requires that we depart from linearized theory (i.e. from the expansion over a flat background) and take a broader point of view, where GWs are introduced as perturbations over a slowly varying, but otherwise generic, curved background, as discussed in Section 1.4. In Section 2.1 we will re-examine these conceptual problems from the point of view of field theory. The combination of the geometrical and field-theoretical perspectives gives in general a deeper understanding of the subject. As a byproduct of the study of the interaction between GWs and the background performed in Section 1.4, we will also find the equation governing the propagation of GWs in curves space, which is examined in Section 1.5. Finally, at the end of the chapter we collect in a Solved Problems section some detailed calculations and some more technical issues.

1

1.1 Expansion around flat space 4

1.2 The TT gauge 7

1.3 Interaction of GWs with test masses 13

1.4 The energy of GWs 26

1.5 Propagation in curved space-time

1.6 Solved problems 48

¹ The factor $1/c$ in eq. (1.2) compensates the fact that $d^4x = c dt d^3x$, see the Notation.

6.0.2 1.1 Expansion around flat space

The gravitational action is $S = S_E + S_M$, where

$$S_E = \frac{c^3}{16\pi G} \int d^4x \sqrt{-g} R \quad (1.1)$$

is the Einstein action and S_M is the matter action. The Ricci scalar R , as well as the Ricci tensor $R_{\mu\nu}$ and the Riemann tensor $R_{\mu\nu\rho\sigma}$ are defined in the Notation section. The energy-momentum tensor of matter, $T^{\mu\nu}$, is defined from the variation of the matter action S_M under a change of the metric $g_{\mu\nu} \rightarrow g_{\mu\nu} + \delta g_{\mu\nu}$, according to ¹

$$\delta S_M = \frac{1}{2c} \int d^4x \sqrt{-g} T^{\mu\nu} \delta g_{\mu\nu} \quad (1.2)$$

Taking the variation of the total action with respect to $g_{\mu\nu}$, one finds the Einstein equations,

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu} \quad (1.3)$$

General relativity is invariant under a huge symmetry group, the group of all possible coordinate transformations,

$$x^\mu \rightarrow x'^\mu(x) \quad (1.4)$$

where x'^μ is an arbitrary smooth function of x^μ . More precisely, we require that $x'^\mu(x)$ be invertible, differentiable, and with a differentiable inverse, i.e. $x'^\mu(x)$ is an arbitrary diffeomorphism. Under eq. (1.4), the metric transforms as

$$g_{\mu\nu}(x) \rightarrow g'_{\mu\nu}(x') = \frac{\partial x^\rho}{\partial x'^\mu} \frac{\partial x^\sigma}{\partial x'^\nu} g_{\rho\sigma}(x) \quad (1.5)$$

We will refer to this symmetry as the gauge symmetry of general relativity.

As a first step toward the understanding of GWs, we wish to study the expansion of the Einstein equations around the flat-space metric. Therefore we write

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad |h_{\mu\nu}| \ll 1 \quad (1.6)$$

and we expand the equations of motion to linear order in $h_{\mu\nu}$. The resulting theory is called linearized theory. Since the numerical values of the components of a tensor depend on the reference frame, what we really mean is that, in the physical situation in which we are interested, there exists a reference frame where eq. (1.6) holds, on a sufficiently large region of space. Choosing a reference frame breaks the invariance of general relativity under coordinate transformations. Indeed, breaking

a local invariance is in general the best way to get rid of spurious degrees of freedom and exposing the actual physical content of a field theory.

However, after choosing a frame where eq. (1.6) holds, a residual gauge symmetry remains. Consider in fact a transformation of coordinates

$$x^\mu \rightarrow x'^\mu = x^\mu + \xi^\mu(x) \quad (1.7)$$

where the derivatives $|\partial_\mu \xi_\nu|$ are at most of the same order of smallness as $|h_{\mu\nu}|$. Using the transformation law of the metric, eq. (1.5), we find that the transformation of $h_{\mu\nu}$, to lowest order, is

$$h_{\mu\nu}(x) \rightarrow h'_{\mu\nu}(x') = h_{\mu\nu}(x) - (\partial_\mu \xi_\nu + \partial_\nu \xi_\mu) \quad (1.8)$$

If $|\partial_\mu \xi_\nu|$ are at most of the same order of smallness as $|h_{\mu\nu}|$, the condition $|h_{\mu\nu}| \ll 1$ is preserved, and therefore these slowly varying diffeomorphisms are a symmetry of linearized theory.²

We can also perform finite, global (i.e. x -independent) Lorentz transformations

$$x^\mu \rightarrow \Lambda^\mu_\nu x^\nu \quad (1.9)$$

By definition of Lorentz transformation, the matrix Λ^μ_ν satisfies

$$\Lambda_\mu^\rho \Lambda_\nu^\sigma \eta_{\rho\sigma} = \eta_{\mu\nu} \quad (1.10)$$

Under a Lorentz transformation,

$$\begin{aligned} g_{\mu\nu}(x) \rightarrow g'_{\mu\nu}(x') &= \Lambda_\mu^\rho \Lambda_\nu^\sigma g_{\rho\sigma}(x) \\ &= \Lambda_\mu^\rho \Lambda_\nu^\sigma [\eta_{\rho\sigma} + h_{\rho\sigma}(x)] \\ &= \eta_{\mu\nu} + \Lambda_\mu^\rho \Lambda_\nu^\sigma h_{\rho\sigma}(x) \end{aligned} \quad (1.11)$$

where we used eq. (1.10). Therefore in the Lorentz-transformed frame we have $g'_{\mu\nu}(x') = \eta_{\mu\nu} + h'_{\mu\nu}(x')$, with

$$h'_{\mu\nu}(x') = \Lambda_\mu^\rho \Lambda_\nu^\sigma h_{\rho\sigma}(x) \quad (1.12)$$

This shows that $h_{\mu\nu}$ is a tensor under Lorentz transformations. Rotations never spoil the condition $|h_{\mu\nu}| \ll 1$, while for boosts we must limit ourselves to those that do not spoil this condition.

Besides, we see from eq. (1.5) that $h_{\mu\nu}$ is invariant under constant translations, i.e. transformations $x^\mu \rightarrow x'^\mu = x^\mu + a^\mu$, where a^μ is not restricted to be infinitesimal, but can be finite. Therefore linearized theory is invariant under finite Poincaré transformations (that is, the group formed by translations and Lorentz transformations), as well as under the infinitesimal local transformation (1.8). In contrast, full general relativity does not have Poincaré symmetry, since the flat space metric plays no special role, but has the full invariance under coordinate transformations, rather than the infinitesimal version (1.8).

To linear order in $h_{\mu\nu}$, the Riemann tensor becomes

$$R_{\mu\nu\rho\sigma} = \frac{1}{2} (\partial_\nu \partial_\rho h_{\mu\sigma} + \partial_\mu \partial_\sigma h_{\nu\rho} - \partial_\mu \partial_\rho h_{\nu\sigma} - \partial_\nu \partial_\sigma h_{\mu\rho}) \quad (1.13)$$

Symmetries of linearized theory

² The first corrections to the righthand side of eq. (1.8) are $O(\partial\xi\partial\xi)$ and $O(h\partial\xi)$. Note that corrections $O(\xi\partial^2\xi)$, which appear in the intermediate steps of the derivation of eq. (1.8), finally cancel, so it is not necessary to require that $|\xi^\mu|$ themselves are small but only that $|\partial_\mu \xi_\nu|$ are small. The condition $|\partial_\mu \xi_\nu| \ll 1$ is also all we need to invert iteratively the relation $x'^\mu = x^\mu + \xi^\mu(x)$, writing $x^\mu = x'^\mu - \xi^\mu(x) = x'^\mu - \xi^\mu(x' - \xi) \simeq x'^\mu - \xi^\mu(x') + O(\xi\partial\xi)$. When the background metric is not $\eta_{\mu\nu}$ we will also find a condition on $|\xi^\mu|$, see Problem 1.2.

Equations of motion of linearized theory ³ More generally, the harmonic (or De Donder gauge) is defined, in a curved background, by the condition $\partial_\mu (g^{\mu\nu} \sqrt{-g}) = 0$. In this form, we will use it extensively in Chapter 5. Writing $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ and expanding to linear order, the harmonic gauge reduces to the Lorentz gauge (1.18).

The denomination "Lorentz gauge" derives its name by the analogy with the Lorentz gauge of electromagnetism, $\partial_\mu A^\mu = 0$. It is amusing to observe that this is in fact a misnomer. In electromagnetism, this gauge was first used by L. V. Lorenz (without "t", the person who also invented retarded potentials) in 1867, when the better known H. A. Lorentz was just 14 years old! (See J. D. Jackson and L. B. Okun, 2001.) However, this "misprint" has by now entered universally into use, and we will conform to it.

To make things worse, none of the above denominations is historically correct. In general relativity, this gauge choice was in fact first suggested to Einstein by De Sitter, see Chapter 3 of Kennefick (2007).

(We will prove this in Problem 1.1, where we perform explicitly the linearization of the Riemann tensor over an arbitrary curved background). Plugging eq. (1.8) into eq. (1.13) we see that, under the residual gauge transformation (1.8), the linearized Riemann tensor is invariant (while, under arbitrary diffeomorphisms in the full non-linearized theory, it is rather covariant).

The linearized equations of motion are written more compactly defining

$$h = \eta^{\mu\nu} h_{\mu\nu} \quad (1.14)$$

and

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} h \quad (1.15)$$

Observe that $\bar{h} \equiv \eta^{\mu\nu} \bar{h}_{\mu\nu} = h - 2h = -h$, so eq. (1.15) can be inverted to give

$$h_{\mu\nu} = \bar{h}_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} \bar{h} \quad (1.16)$$

In linearized theory we use the convention that indices are raised and lowered with the flat metric $\eta_{\mu\nu}$. Using eq. (1.13) we can compute with straightforward algebra the linearization of

the Einstein tensor $G_{\mu\nu} = R_{\mu\nu} - (1/2)g_{\mu\nu}R$, and we find that the linearization of the Einstein equations (1.3) is

$$\square \bar{h}_{\mu\nu} + \eta_{\mu\nu} \partial^\rho \partial^\sigma \bar{h}_{\rho\sigma} - \partial^\rho \partial_\nu \bar{h}_{\mu\rho} - \partial^\rho \partial_\mu \bar{h}_{\nu\rho} = -\frac{16\pi G}{c^4} T_{\mu\nu} \quad (1.17)$$

We can now use the gauge freedom (1.8) to choose the Lorentz gauge (also called the Hilbert gauge, or the harmonic gauge, or the De Donder gauge) ³

$$\partial^\nu \bar{h}_{\mu\nu} = 0 \quad (1.18)$$

To prove that, using the symmetry transformation (1.8), we can impose the condition (1.18), we observe that, in terms of $\bar{h}_{\mu\nu}$, eq. (1.8) becomes

$$\bar{h}_{\mu\nu} \rightarrow \bar{h}'_{\mu\nu} = \bar{h}_{\mu\nu} - (\partial_\mu \xi_\nu + \partial_\nu \xi_\mu - \eta_{\mu\nu} \partial_\rho \xi^\rho) \quad (1.19)$$

and therefore

$$\partial^\nu \bar{h}_{\mu\nu} \rightarrow (\partial^\nu \bar{h}_{\mu\nu})' = \partial^\nu \bar{h}_{\mu\nu} - \square \xi_\mu \quad (1.20)$$

where, in the context of linearized theory, $\square$ is defined as the flat space d'Alembertian, $\square = \eta_{\mu\nu} \partial^\mu \partial^\nu = \partial_\mu \partial^\mu$. (Recall also that in linearized theory indices are raised and lowered with the flat metric $\eta_{\mu\nu}$.) Therefore, if the initial field configuration $h_{\mu\nu}$ is such that $\partial^\nu \bar{h}_{\mu\nu} = f_\mu(x)$, with $f_\mu(x)$ some function, to obtain $(\partial^\nu \bar{h}_{\mu\nu})' = 0$ we must choose $\xi_\mu(x)$ so that

$$\square \xi_\mu = f_\mu(x) \quad (1.21)$$

This equation always admits solutions, because the d'Alembertian operator is invertible. If we denote by $G(x)$ a Green's function of the d'Alembertian operator, so that

$$\square_x G(x - y) = \delta^4(x - y) \quad (1.22)$$

then the corresponding solution is

$$\xi_\mu(x) = \int d^4y G(x - y) f_\mu(y) \quad (1.23)$$

In this gauge the last three terms on the left-hand side of eq. (1.17) vanish, and we get a simple wave equation,

$$\square \bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4} T_{\mu\nu} \quad (1.24)$$

Observe that eq. (1.18) gives four conditions, that reduce the 10 independent components of the symmetric 4×4 matrix $h_{\mu\nu}$ to six independent components. Equations (1.18) and (1.24) together imply for consistency

$$\partial^\nu T_{\mu\nu} = 0 \quad (1.25)$$

which is the conservation of energy-momentum in the linearized theory. This should be contrasted with the conservation in the full theory, $D^\nu T_{\mu\nu} = 0$, where D^ν is the covariant derivative.

Physically, the approximations implicit in the linearized theory can be summarized as follows: the bodies that act as sources of GWs are taken to move in flat space-time, along the trajectories determined by their mutual influence. In particular, for a self-gravitating system such as a binary star, the fact that the background space-time metric is $\eta_{\mu\nu}$ means that we are describing its dynamics using Newtonian gravity, rather than full general relativity. The

response of test masses to the GW $h_{\mu\nu}$ generated by these bodies is rather computed using $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, and neglecting terms $O(h^2)$ when evaluating the Christoffel symbols or the Riemann tensor.

6.0.3 1.2 The transverse-traceless gauge

Equation (1.24) is the basic result for computing the generation of GWs within linearized theory. To study the propagation of GWs as well as the interaction with test masses (and therefore with a GW detector), we are rather interested in this equation outside the source, i.e. where $T_{\mu\nu} = 0$

$$\square \bar{h}_{\mu\nu} = 0 \quad (\text{outside the source}). \quad (1.26)$$

Since $\square = -(1/c^2) \partial_t^2 + \nabla^2$, eq. (1.26) implies that GWs travel at the speed of light. Outside the source we can greatly simplify the form of the metric, observing that eq. (1.18) does not fix the gauge completely; in fact, we saw in eq. (1.20) that, under the transformation (1.7), $\partial^\nu \bar{h}_{\mu\nu}$ ⁴ This closely parallels the situation in electrodynamics. The classical equation of motion obtained from the variation of the Maxwell Lagrangian with an external current is $\partial_\mu F^{\mu\nu} = j^\nu$, i.e. $\partial_\mu (\partial^\mu A^\nu - \partial^\nu A^\mu) = j^\nu$, and it becomes $\square A^\nu = j^\nu$ when we impose the Lorentz gauge $\partial_\mu A^\mu = 0$. The Lorentz gauge still leaves the residual gauge freedom $A_\mu \rightarrow A_\mu - \partial_\mu \theta$ with $\square \theta = 0$. Outside the source we have $j^\mu = 0$, and therefore $\square A^\mu = 0$, so the residual gauge freedom, i.e. the function θ which satisfies $\square \theta = 0$, can be used to set $A^0 = 0$. When $A^0 = 0$, the Lorentz gauge $\partial_\mu A^\mu = 0$ becomes a transversality condition on A^i , $\partial_i A^i = 0$. If instead $j^0 \neq 0$, we have $\square A^0 \neq 0$ and we cannot remove A^0 using a function θ which satisfies $\square \theta = 0$. transforms as in eq. (1.20). Then, the condition $\partial^\nu \bar{h}_{\mu\nu} = 0$ is not spoiled by a further coordinate transformation $x^\mu \rightarrow x^\mu + \xi^\mu$ with

$$\square \xi_\mu = 0 \quad (1.27)$$

If $\square \xi_\mu$ is zero, then also $\square \xi_{\mu\nu} = 0$, where

$$\xi_{\mu\nu} \equiv \partial_\mu \xi_\nu + \partial_\nu \xi_\mu - \eta_{\mu\nu} \partial_\rho \xi^\rho \quad (1.28)$$

since the flat space d'Alembertian $\square$ commutes with ∂_μ . Then eq. (1.19) tells us that, from the six independent components of $\bar{h}_{\mu\nu}$, which satisfy $\square \bar{h}_{\mu\nu} = 0$, we can subtract the functions $\xi_{\mu\nu}$, which depend on four independent arbitrary functions ξ_μ , and which satisfy the same equation, $\square \xi_{\mu\nu} = 0$. This means that we can choose the functions ξ_μ so as to impose four conditions on $h_{\mu\nu}$. In particular, we can choose ξ^0 such that the trace $\bar{h} = 0$. Note that if $\bar{h} = 0$, then $\bar{h}_{\mu\nu} = h_{\mu\nu}$. The three functions $\xi^i(x)$ are now chosen so that $h^{0i}(x) = 0$. Since $\bar{h}_{\mu\nu} = h_{\mu\nu}$, the Lorentz condition (1.18) with $\mu = 0$ reads

$$\partial^0 h_{00} + \partial^i h_{0i} = 0 \quad (1.29)$$

Having fixed $h_{0i} = 0$, this simplifies to

$$\partial^0 h_{00} = 0 \quad (1.30)$$

so h_{00} becomes automatically constant in time. A time-independent term h_{00} corresponds to the static part of the gravitational interaction, i.e. to the Newtonian potential of the source which generated the gravitational wave. The gravitational wave itself is the time-dependent part and therefore, as far as the GW is concerned, $\partial^0 h_{00} = 0$ means that $h_{00} = 0$. So, we have set all four components $h_{0\mu} = 0$ and we are left only with the spatial components h_{ij} , for which

the Lorentz gauge condition now reads $\partial^j h_{ij} = 0$, and the condition of vanishing trace becomes $h^i_i = 0$. In conclusion, we have set

$$h^{0\mu} = 0, \quad h^i_i = 0, \quad \partial^j h_{ij} = 0 \quad (1.31)$$

This defines the transverse-traceless gauge, or TT gauge. By imposing the Lorentz gauge, we have reduced the 10 degrees of freedom of the symmetric matrix $h_{\mu\nu}$ to six degrees of freedom, and the residual gauge freedom, associated to the four functions ξ^μ that satisfy eq. (1.27), has further reduced these to just two degrees of freedom. We will denote the metric in the TT gauge by h_{ij}^{TT} .

Observe that the TT gauge cannot be chosen inside the source, since in this case $\square \bar{h}_{\mu\nu} \neq 0$. Inside the source, once we have chosen the Lorentz gauge, we still have the freedom to perform a transformation with $\square \xi_\mu = 0$, and therefore $\square \xi_{\mu\nu} = 0$. However, now we cannot set to zero any further component of $\bar{h}_{\mu\nu}$, which satisfies $\square \bar{h}_{\mu\nu} \neq 0$, subtracting from it a function $\xi_{\mu\nu}$ which satisfies $\square \xi_{\mu\nu} = 0$.⁴

Equation (1.26) has plane wave solutions, $h_{ij}^{\text{TT}}(x) = e_{ij}(\mathbf{k})e^{ikx}$, with $k^\mu = (\omega/c, \mathbf{k})$ and $\omega/c = |\mathbf{k}|$ (and the usual convention that the real part is taken at the end of the computation). The tensor $e_{ij}(\mathbf{k})$ is called the polarization tensor. For a single plane wave with a given wave-vector $\mathbf{k}$ (or for a superposition of plane waves with different frequencies but all with the same direction of propagation $\hat{\mathbf{n}} = \mathbf{k}/|\mathbf{k}|$), we see from eq. (1.31) that the non-zero components of h_{ij}^{TT} are in the plane transverse to $\hat{\mathbf{n}}$ since, on a plane wave, the condition $\partial^j h_{ij} = 0$ becomes $n^i h_{ij} = 0$. Choosing for definiteness $\hat{\mathbf{n}}$ along the z axis, and, imposing that h_{ij} be symmetric and traceless, we have

$$h_{ij}^{\text{TT}}(t, z) = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}_{ij} \cos[\omega(t - z/c)] \quad (1.32)$$

or, more simply,

$$h_{ab}^{\text{TT}}(t, z) = \begin{pmatrix} h_+ & h_\times \\ h_\times & -h_+ \end{pmatrix}_{ab} \cos[\omega(t - z/c)] \quad (1.33)$$

where $a, b = 1, 2$ are indices in the transverse (x, y) plane; h_+ and $h_\times$ are called the amplitudes of the "plus" and "cross" polarization of the wave. In terms of the interval ds^2 ,

$$ds^2 = -c^2 dt^2 + dz^2 + \{1 + h_+ \cos[\omega(t - z/c)]\} dx^2 + \{1 - h_+ \cos[\omega(t - z/c)]\} dy^2 + 2h_\times \cos[\omega(t - z/c)] dx dy \quad (1.34)$$

Given a plane wave solution $h_{\mu\nu}(x)$ propagating in the direction $\hat{\mathbf{n}}$, outside the sources, already in the Lorentz gauge but not yet in the TT gauge, we can find the form of the wave in the TT gauge as follows. First we introduce the tensor

$$P_{ij}(\hat{\mathbf{n}}) = \delta_{ij} - n_i n_j \quad (1.35)$$

This tensor is symmetric, is transverse (i.e. $n^i P_{ij}(\hat{\mathbf{n}}) = 0$), is a projector (i.e. $P_{ik} P_{kj} = P_{ij}$), and its trace is $P_{ii} = 2$. With the help of P_{ij} we construct

$$\Lambda_{ij,kl}(\hat{\mathbf{n}}) = P_{ik} P_{jl} - \frac{1}{2} P_{ij} P_{kl} \quad (1.36)$$

This is still a projector, in the sense that

$$\Lambda_{ij,kl} \Lambda_{kl,mn} = \Lambda_{ij,mn} \quad (1.37)$$

Furthermore it is transverse on all indices, $n^i \Lambda_{ij,kl} = 0$, $n^j \Lambda_{ij,kl} = 0$, etc., it is traceless with respect to the (i, j) and (k, l) indices,

$$\Lambda_{ii,kl} = \Lambda_{ij,kk} = 0 \quad (1.38)$$

Projection onto the TT gauge; the Lambda tensor

⁵ This is the same tensor $\Lambda_{ij,kl}$ defined in Weinberg (1972), eq. (10.4.14). However, Weinberg only uses it in an equation where it is contracted with the tensor $T^{ij}T^{kl}$, which is symmetric under $i \leftrightarrow j$ and $k \leftrightarrow l$ and then he replaces the term $n_j n_l \delta_{ik} + n_i n_k \delta_{jl}$ in eq. (1.39) by $2n_j n_l \delta_{ik}$. and it is symmetric under the simultaneous exchange $(i, j) \leftrightarrow (k, l)$. In terms of $\hat{\mathbf{n}}$, its explicit form is

$$\begin{aligned} \Lambda_{ij,kl}(\hat{\mathbf{n}}) = & \delta_{ik} \delta_{jl} - \frac{1}{2} \delta_{ij} \delta_{kl} - n_j n_l \delta_{ik} - n_i n_k \delta_{jl} \\ & + \frac{1}{2} n_k n_l \delta_{ij} + \frac{1}{2} n_i n_j \delta_{kl} + \frac{1}{2} n_i n_j n_k n_l \end{aligned} \quad (1.39)$$

We shall meet $\Lambda_{ij,kl}$ often, and we call it the Lambda tensor. ⁵ We can now show that, given a plane wave $h_{\mu\nu}$ in the Lorentz gauge, but not yet in the TT gauge, the GW in the TT gauge is given in terms of the spatial components h_{ij} of $h_{\mu\nu}$ by

$$h_{ij}^{\text{TT}} = \Lambda_{ij,kl} h_{kl} \quad (1.40)$$

In fact by construction the right-hand side is transverse and traceless in (i, j) while, from the fact that $h_{\mu\nu}$ was a solution of the wave equation in the vacuum and that it was in the Lorentz gauge, it follows that $\square h_{ij}^{\text{TT}} = 0$. (Observe that it is important that $h_{\mu\nu}$ be in the Lorentz gauge already, otherwise the equation of motion that it satisfies would not simply be $\square h_{\mu\nu} = 0$.)

In general, given any symmetric tensor S_{ij} , we define its transversetraceless part as

$$S_{ij}^{\text{TT}} = \Lambda_{ij,kl} S_{kl} \quad (1.41)$$

Observe that S_{ij}^{TT} is still symmetric.

For later calculations, it is useful to spell out clearly our conventions and definitions for the plane wave expansion. In the TT gauge the equation of motion is $\square h_{ij}^{\text{TT}} = 0$ and therefore h_{ij}^{TT} can be expanded as

$$h_{ij}^{\text{TT}}(x) = \int \frac{d^3 k}{(2\pi)^3} (\mathcal{A}_{ij}(\mathbf{k}) e^{ikx} + \mathcal{A}_{ij}^*(\mathbf{k}) e^{-ikx}) \quad (1.42)$$

The four-vector k^μ , with dimensions of inverse length, is related to the frequency ω and to the wave-vector $\mathbf{k}$ by $k^\mu = (\omega/c, \mathbf{k})$, with $|\mathbf{k}| = \omega/c = (2\pi f)/c$ and $\mathbf{k}/|\mathbf{k}| = \hat{\mathbf{n}}$. Therefore $d^3 k = |\mathbf{k}|^2 d|\mathbf{k}| d\Omega = (2\pi/c)^3 f^2 df d\Omega$, with $f > 0$. We denote by $d^2 \hat{\mathbf{n}} = d\cos\theta d\phi$ the integration over the solid angle, so the above equation reads

$$h_{ij}^{\text{TT}}(x) = \frac{1}{c^3} \int_0^\infty df f^2 \int d^2 \hat{\mathbf{n}} (\mathcal{A}_{ij}(f, \hat{\mathbf{n}}) e^{-2\pi i f(t - \hat{\mathbf{n}} \cdot \mathbf{x}/c)} + c.c.) \quad (1.43)$$

Observe that, inside the parentheses, both the contribution written explicitly and its complex conjugate refer to a wave traveling in the direction $+\hat{\mathbf{n}}$, since they depend on t and $\mathbf{x}$ only through the combination $(t - \hat{\mathbf{n}} \cdot \mathbf{x}/c)$. Observe also that, in this form, only "physical" frequencies $f > 0$ enter in the expansion.

The TT gauge conditions (1.31) give $\mathcal{A}_i^i(\mathbf{k}) = 0$ and $k^i \mathcal{A}_{ij}(\mathbf{k}) = 0$. Of course, in a superposition of waves with different propagation directions, $h_{ij}(x)$ does not reduce to a 2×2 matrix; for instance, h_{12} gets contributions from the waves with $k^3 \neq 0$, h_{13} gets contributions from the waves with $k^2 \neq 0$, h_{23} from the waves with $k^1 \neq 0$, etc. This will be important when we consider stochastic backgrounds of GWs. However, when we observe on Earth a GW emitted by a single astrophysical source, the direction of propagation of the wave, $\hat{\mathbf{n}}_0$, is very well defined and we can write

$$\mathcal{A}_{ij}(\mathbf{k}) = A_{ij}(f) \delta^{(2)}(\hat{\mathbf{n}} - \hat{\mathbf{n}}_0) \quad (1.44)$$

The transversality condition now states that the only non-vanishing components are those in the plane transverse to the propagation direction $\hat{\mathbf{n}}_0$. We label by $a, b = 1, 2$ the indices in the transverse plane and we omit, for notational simplicity, the superscript TT, since the fact that we are in the TT gauge is already implicit in the use of the indices $a, b = 1, 2$ instead of $i, j = 1, 2, 3$. Then

$$h_{ab}(t, \mathbf{x}) = \int_0^\infty df \left(\tilde{h}_{ab}(f, \mathbf{x}) e^{-2\pi i f t} + \tilde{h}_{ab}^*(f, \mathbf{x}) e^{2\pi i f t} \right) \quad (1.45)$$

where

$$\begin{aligned} \tilde{h}_{ab}(f, \mathbf{x}) &= \frac{f^2}{c^3} \int d^2 \hat{\mathbf{n}} \mathcal{A}_{ab}(f, \hat{\mathbf{n}}) e^{2\pi i f \hat{\mathbf{n}} \cdot \mathbf{x} / c} \\ &= \frac{f^2}{c^3} A_{ab}(f) e^{2\pi i f \hat{\mathbf{n}}_0 \cdot \mathbf{x} / c} \end{aligned} \quad (1.46)$$

As we will see when we discuss the detectors, for resonant bars and ground based interferometers the linear dimensions of the detector are much smaller than the reduced wavelength $\lambda = \lambda/(2\pi)$ of the GWs to which they are sensitive. In this case, choosing the origin of the coordinate system centered on the detector, we have $\exp\{2\pi i f \hat{\mathbf{n}} \cdot \mathbf{x} / c\} = \exp\{i \hat{\mathbf{n}} \cdot \mathbf{x} / \lambda\} \simeq 1$ all over the detector. If we are interested in the GW at the detector location, we can therefore neglect all $\mathbf{x}$ -dependences and write simply

$$h_{ab}(t) = \int_0^\infty df \left(\tilde{h}_{ab}(f) e^{-2\pi i f t} + \tilde{h}_{ab}^*(f) e^{2\pi i f t} \right) \quad (1.47)$$

with $\tilde{h}_{ab}(f) = \tilde{h}_{ab}(f, \mathbf{x} = 0)$. Of course, the dependence on $\mathbf{x}$ must be kept when we compare the GW signal at two different detectors (e.g. when we consider the overlap reduction function in a two-detector correlation, see Section 7.8.3) or when we need spatial derivatives of $h_{ab}(t, \mathbf{x})$ (e.g. when we compute spatial components of the energymomentum tensor).

From eq. (1.33) it follows that

$$\tilde{h}_{ab}(f) = \begin{pmatrix} \tilde{h}_+(f) & \tilde{h}_\times(f) \\ \tilde{h}_\times(f) & -\tilde{h}_+(f) \end{pmatrix}_{ab} \quad (1.48)$$

The $+$ and $\times$ polarizations are defined with respect to a given choice of axes in the transverse plane. If we rotate by an angle ψ the system of

⁶ Of course eq. (1.52) could have been written directly setting $\mathbf{x} = 0$ from the beginning and writing a temporal Fourier transform. However, this derivation makes it clear that negative frequencies are not "unphysical", but are simply a way of rewriting the complex conjugate term indicated by c.c. in eq. (1.43), and that this contribution refers to a wave traveling in the $+\hat{\mathbf{n}}$

direction, exactly as the contribution with $f > 0$. axes used for their definition, we show in Problem 2.1 that h_+ and $h_\times$ transform as

$$h_+ \rightarrow h_+ \cos 2\psi - h_\times \sin 2\psi \quad (1.49)$$

$$h_\times \rightarrow h_+ \sin 2\psi + h_\times \cos 2\psi \quad (1.50)$$

Observe that, until now, only physical frequencies $f > 0$ entered our equations. However, eq. (1.45) can be rewritten in a slightly more compact form, extending the definition of $\tilde{h}_{ab}(f, \mathbf{x})$ to negative frequencies, by defining

$$\tilde{h}_{ab}(-f, \mathbf{x}) = \tilde{h}_{ab}^*(f, \mathbf{x}) \quad (1.51)$$

so that eq. (1.45) becomes ⁶ (we do not explicitly write the $\mathbf{x}$ dependence)

$$h_{ab}(t) = \int_{-\infty}^{\infty} df \tilde{h}_{ab}(f) e^{-2\pi i f t} \quad (1.52)$$

The inversion of eq. (1.52) is

$$\tilde{h}_{ab}(f) = \int_{-\infty}^{\infty} dt h_{ab}(t) e^{2\pi i f t} \quad (1.53)$$

Another useful form for the plane wave expansion is obtained by introducing the polarization tensors $e_{ij}^A(\hat{\mathbf{n}})$ (with $A = +, \times$ labeling the polarizations) defined as

$$e_{ij}^+(\hat{\mathbf{n}}) = \hat{\mathbf{u}}_i \hat{\mathbf{u}}_j - \hat{\mathbf{v}}_i \hat{\mathbf{v}}_j, \quad e_{ij}^\times(\hat{\mathbf{n}}) = \hat{\mathbf{u}}_i \hat{\mathbf{v}}_j + \hat{\mathbf{v}}_i \hat{\mathbf{u}}_j \quad (1.54)$$

with $\hat{\mathbf{u}}, \hat{\mathbf{v}}$ unit vectors orthogonal to the propagation direction $\hat{\mathbf{n}}$ and to each other. With this definition, the polarization tensors are normalized as

$$e_{ij}^A(\hat{\mathbf{n}}) e^{A', ij}(\hat{\mathbf{n}}) = 2\delta^{AA'} \quad (1.55)$$

In the frame where $\hat{\mathbf{n}}$ is along the $\hat{\mathbf{z}}$ direction, we can choose $\hat{\mathbf{u}} = \hat{\mathbf{x}}$ and $\hat{\mathbf{v}} = \hat{\mathbf{y}}$, so

$$e_{ab}^+ = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}_{ab}, \quad e_{ab}^\times = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}_{ab} \quad (1.56)$$

with $a, b = 1, 2$ spanning the (x, y) plane. In a generic frame, we can define the amplitudes $\tilde{h}_A(f, \hat{\mathbf{n}})$ from

$$\frac{f^2}{c^3} \mathcal{A}_{ij}(f, \hat{\mathbf{n}}) = \sum_{A=+, \times} \tilde{h}_A(f, \hat{\mathbf{n}}) e_{ij}^A(\hat{\mathbf{n}}) \quad (1.57)$$

Equation (1.43) then becomes

$$h_{ab}(t, \mathbf{x}) = \sum_{A=+, \times} \int_{-\infty}^{\infty} df \int d^2 \hat{\mathbf{n}} \tilde{h}_A(f, \hat{\mathbf{n}}) e_{ab}^A(\hat{\mathbf{n}}) e^{-2\pi i f(t - \hat{\mathbf{n}} \cdot \mathbf{x}/c)}$$

where again we have defined $\tilde{h}_A(-f, \hat{\mathbf{n}}) = \tilde{h}_A^*(f, \hat{\mathbf{n}})$.

6.0.4 1.3 Interaction of GWs with test masses

In the previous section we have seen how to describe a GW. In this section we discuss the interaction of GWs with a detector, idealized for the moment as a set of test masses. This is an issue that hides some subtleties because, even if the physics must finally be invariant under coordinate transformations, the language that we use to describe the GWs and the detector, as well as the intermediate steps of the computations, do depend on the reference frame that we choose.

In general relativity, the mathematical procedure of choosing a gauge corresponds, physically, to selecting a specific observer. We have seen that GWs have an especially simple form in the TT gauge, so we want to understand which reference frame corresponds to the TT gauge. We will also see that the description of the detector is more intuitive in another frame, the detector proper frame. It is therefore important, when we discuss the interaction of GWs with the detector, to be aware of which reference frame we are using, and to understand which is the appropriate language in that frame.

Two important tools for understanding the physical meaning of a given gauge choice are the geodesic equation and the equation of the geodesic deviation. We briefly recall these basic concepts of general relativity in the next subsection. We will then explore the interaction of GWs with test masses in different frames, in particular in the TT frame and in the detector proper frame.

6.0.5 1.3.1 Geodesic equation and geodesic deviation

In this subsection and in the next we recall some elementary notions of general relativity, referring the reader, e.g. to Misner, Thorne and Wheeler (1973) or to Hartle (2003) for more details and proofs. Consider, in some reference frame, a curve $x^\mu(\lambda)$, parametrized by a parameter λ . The interval ds between two points separated by $d\lambda$ is given by

$$\begin{aligned} ds^2 &= g_{\mu\nu} dx^\mu dx^\nu \\ &= g_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda} d\lambda^2 \end{aligned} \quad (1.59)$$

All along a space-like curve we have, by definition, $ds^2 > 0$, and we can use

$$ds = (g_{\mu\nu} dx^\mu dx^\nu)^{1/2} \quad (1.60)$$

to measure proper distances along the curve. A time-like curve is rather defined by the condition that all along it $ds^2 < 0$, and in this case we can define the proper time τ , from

$$c^2 d\tau^2 = -ds^2 = -g_{\mu\nu} dx^\mu dx^\nu \quad (1.61)$$

The proper time τ is the time measured by a clock carried along this trajectory. It is therefore natural to use τ itself as the parameter λ

⁷ More precisely, each geodesic is parametrized by its own proper time τ , and $\xi^\mu(\tau)$ connects point with the same value of τ on the two geodesics. which parametrizes the trajectory, so that $x^\mu = x^\mu(\tau)$. Observe, from eq. (1.61), that

$$g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = -c^2 \quad (1.62)$$

The four-velocity u^μ is defined as

$$u^\mu = \frac{dx^\mu}{d\tau} \quad (1.63)$$

so eq. (1.62) reads

$$g_{\mu\nu}u^\mu u^\nu = -c^2 \quad (1.64)$$

Among all possible time-like curves that satisfy the fixed boundary conditions $x^\mu(\tau_A) = x_A^\mu$ and $x^\mu(\tau_B) = x_B^\mu$, the classical trajectory of a point-like test mass m is obtained by extremizing the action

$$S = -m \int_{\tau_A}^{\tau_B} d\tau \quad (1.65)$$

This gives the geodesic equation,

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma_{\nu\rho}^\mu(x) \frac{dx^\nu}{d\tau} \frac{dx^\rho}{d\tau} = 0 \quad (1.66)$$

which is the classical equation of motion of a test mass in the curved background described by the metric $g_{\mu\nu}$, in the absence of external nongravitational forces. In terms of the four-velocity u^μ , the geodesic equation reads

$$\frac{du^\mu}{d\tau} + \Gamma_{\nu\rho}^\mu u^\nu u^\rho = 0 \quad (1.67)$$

Consider now two nearby geodesics, one parametrized by $x^\mu(\tau)$ and the other by $x^\mu(\tau) + \xi^\mu(\tau)$.⁷ Then $x^\mu(\tau)$ satisfies eq. (1.66), while $x^\mu(\tau) + \xi^\mu(\tau)$ satisfies

$$\frac{d^2 (x^\mu + \xi^\mu)}{d\tau^2} + \Gamma_{\nu\rho}^\mu(x + \xi) \frac{d(x^\nu + \xi^\nu)}{d\tau} \frac{d(x^\rho + \xi^\rho)}{d\tau} = 0 \quad (1.68)$$

If $|\xi^\mu|$ is much smaller than the typical scale of variation of the gravitational field, taking the difference between eqs. (1.68) and (1.66), and expanding to first order in ξ , we get

$$\frac{d^2 \xi^\mu}{d\tau^2} + 2\Gamma_{\nu\rho}^\mu(x) \frac{dx^\nu}{d\tau} \frac{d\xi^\rho}{d\tau} + \xi^\sigma \partial_\sigma \Gamma_{\nu\rho}^\mu(x) \frac{dx^\nu}{d\tau} \frac{dx^\rho}{d\tau} = 0 \quad (1.69)$$

This is the equation of the geodesic deviation. We can rewrite it in a more elegant way by introducing the covariant derivative of a vector field $V^\mu(x)$ along the curve $x^\mu(\tau)$,

$$\frac{DV^\mu}{D\tau} \equiv \frac{dV^\mu}{d\tau} + \Gamma_{\nu\rho}^\mu V^\nu \frac{dx^\rho}{d\tau} \quad (1.70)$$

Then, eq. (1.69) can be written as

$$\frac{D^2 \xi^\mu}{D\tau^2} = -R_{\nu\rho\sigma}^\mu \xi^\rho \frac{dx^\nu}{d\tau} \frac{dx^\sigma}{d\tau} \quad (1.71)$$

or, in terms of the four-velocity u^μ ,

$$\frac{D^2 \xi^\mu}{D\tau^2} = -R_{\nu\rho\sigma}^\mu u^\rho u^\sigma \quad (1.72)$$

This equation shows that two nearby time-like geodesics experience a tidal gravitational force, which is determined by the Riemann tensor. Writing explicitly the geodesic equation or the equation of the geodesic deviation in the reference frame of interest we can understand how test masses behave for the corresponding observer. We consider the most relevant examples in the next subsections.

6.0.6 1.3.2 Local inertial frames and freely falling frames

Before discussing the TT frame and the detector frame, it may be useful to recall some basic facts about the construction of local inertial frames and of freely falling frames.

It is a standard exercise in general relativity to show that it is always possible to perform a change of coordinates such that, at a given space-time point P , all the components of the Christoffel symbol vanish, $\Gamma_{\nu\rho}^{\mu}(P) = 0$. In such a frame, at P the geodesic equation (1.66) becomes

$$\left. \frac{d^2 x^{\mu}}{d\tau^2} \right|_P = 0 \quad (1.73)$$

so in this frame a test mass is free falling, although only at one point in space and at one moment in time. Such a frame is called a local inertial frame (sometimes abbreviated as LIF), and gives a realization of the equivalence principle.

An explicit construction of the corresponding system of coordinates can be done as follows (see, e.g. Hartle 2003, Section 8.4). At the point P we choose a basis of four orthonormal four-vectors, e_{α} , where $\alpha = 0, \dots, 3$ labels the four-vector. We choose them orthogonal to each other with respect to the flat-space metric $\eta_{\mu\nu}$, so $\eta_{\mu\nu} e_{\alpha}^{\mu} e_{\beta}^{\nu} = \eta_{\alpha\beta}$. Consider now the spatial geodesic that starts at P , in the direction of a space-like unit fourvector n . We parametrize the geodesic using the proper distance. Let Q be the point reached from P , moving along this geodesic, after a proper distance s , and let (n^0, n^1, n^2, n^3) be the components of n in the basis $\{e_{\alpha}\}$. Then we assign to Q the coordinates $x_Q = (sn^0, sn^1, sn^2, sn^3)$. Thus, for example, if we send out a geodesic along the direction e_3 , and we meet a point Q after a proper distance s , the coordinates of Q are $x_Q = (0, 0, 0, s)$. Similarly, we send out a time-like geodesic in the direction of a time-like unit four-vector n , we parametrize this geodesic using proper time, and we assign the coordinates $(\tau n^0, \tau n^1, \tau n^2, \tau n^3)$ to the point that we reach after a proper time τ .

We fill all of space-time with time-like or space-like geodesics (null geodesics can be approximated to arbitrary accuracy with time-like or space-like geodesics and therefore can be obtained by continuity), so all points are reached by at least one geodesic.

In a sufficiently small region of space, geodesics do not intersect (which of course is no longer true on large regions of space, as vividly shown for instance by the phenomenon of gravitational lensing), so each point in this small region is reached by one and only one geodesic. Thus, the above method allows us to assign coordinates unambiguously to all points of a sufficiently small space-time region around P . This coordinate system is known as Riemann normal coordinates. We can now check that it indeed gives an explicit realization of a local inertial frame.

First of all, the fact that $g_{\mu\nu}(P) = \eta_{\mu\nu}$ follows simply from the fact that the coordinates are referred to the basis $\{e_{\alpha}\}$, built at P , which by definition is orthonormal with respect to the flat space-time metric, $\eta_{\mu\nu} e_{\alpha}^{\mu} e_{\beta}^{\nu} = \eta_{\alpha\beta}$. To show that we also have $\Gamma_{\nu\rho}^{\mu}(P) = 0$ in this frame, consider the geodesic equation (1.66). Since, by definition, the coordinates are linear in proper time (if they are reached by a time-like geodesic; or linear in proper distance if they are reached by space-like geodesics, in which case in the geodesic equation (1.66) τ must be replaced by s), the term $d^2 x^{\mu}/d\tau^2$ (or $d^2 x^{\mu}/ds^2$ for space-like geodesics) vanishes, while $dx^{\mu}/d\tau = n^{\mu}$. Then the geodesic equation becomes

$$\Gamma_{\nu\rho}^{\mu}(P) n^{\nu} n^{\rho} = 0 \quad (1.74)$$

and, since this holds for all n^{μ} , we conclude that

$$\Gamma_{\nu\rho}^{\mu}(P) = 0 \quad (1.75)$$

Riemann normal coordinates therefore provide an explicit example of a local inertial frame.

In a local inertial frame a test mass moves freely only at one point in space and at one moment in time. We can however do much better than this, building a reference frame where a test mass is in free fall all along the geodesic. Such a frame can be built observing that a freely spinning object (like a gyroscope) that moves along a time-like geodesic $x^\mu(\tau)$ obeys the equation

$$\frac{ds^\mu}{d\tau} + \Gamma_{\nu\rho}^\mu s^\nu \frac{dx^\rho}{d\tau} = 0 \quad (1.76)$$

where s^μ is the spin four-vector, i.e. the four-vector that in the rest frame reduces to $s^\mu = (0, \mathbf{s})$. This equation is the covariant generalization of the equation $ds^\mu/d\tau = 0$ that expresses the conservation of angular momentum in flat space-time. We start by constructing a local inertial frame at P , as before, but using three gyroscopes to mark the direction of the spatial axes. We then propagate this reference frame along the geodesic, always orienting the spatial axes in the direction pointed out by the gyroscopes (while the time axis is in the direction of the four-velocity along the geodesic). By definition, then, in this frame the gyroscopes do not rotate with respect to the axes since they define the orientation of the axes. Then, $ds^\mu/d\tau = 0$ along the entire time-like geodesic and, from eq. (1.76), we see that $\Gamma_{\nu\rho}^\mu$ vanishes along the entire time-like geodesic, and not just at a single point P . Such a reference frame is called a freely

falling frame, and its coordinates (Riemann normal coordinates with axes marked by gyroscopes) are known as Fermi normal coordinates. A freely falling frame is therefore a local inertial frame along an entire geodesic.

Such a frame is practically given by drag-free satellites, in which an experimental apparatus is freely floating inside a satellite, which screens it from external disturbances (e.g. solar wind, micrometeorites, etc.). The satellite senses precisely the position of the experimental apparatus and adjusts its position, using thrusters, to remain centered about it. As we will see, these drag-free techniques are crucial for a space interferometer.

6.0.7 1.3.3 TT frame and proper detector frame

6.0.8 The TT frame

We have seen that there exists a gauge where GWs have an especially simple form, the TT gauge. We denote the corresponding reference frame as the TT frame and we now ask what it means, physically, to be in the TT frame.

Again, the answer can be found by looking at the geodesic equation, eq. (1.66). If a test mass is at rest at $\tau = 0$, we find from eq. (1.66) that

$$\begin{aligned} \left. \frac{d^2 x^i}{d\tau^2} \right|_{\tau=0} &= - \left[\Gamma_{\nu\rho}^i(x) \frac{dx^\nu}{d\tau} \frac{dx^\rho}{d\tau} \right]_{\tau=0} \\ &= - \left[\Gamma_{00}^i \left(\frac{dx^0}{d\tau} \right)^2 \right]_{\tau=0} \end{aligned} \quad (1.77)$$

where in the second line we used the fact that, by assumption, at $\tau = 0$ we have $dx^i/d\tau = 0$, since we took the mass initially at rest. Writing $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ and expanding to first order in $h_{\mu\nu}$, the Christoffel symbol $\Gamma_{\nu\rho}^\mu$ becomes

$$\Gamma_{\nu\rho}^\mu = \frac{1}{2} \eta^{\mu\sigma} (\partial_\nu h_{\rho\sigma} + \partial_\rho h_{\nu\sigma} - \partial_\sigma h_{\nu\rho}) \quad (1.78)$$

and therefore

$$\Gamma_{00}^i = \frac{1}{2} (2\partial_0 h_{0i} - \partial_i h_{00}) \quad (1.79)$$

However, in the TT gauge this quantity vanishes, because both h_{00} and h_{0i} are set to zero by the gauge condition. Therefore, if at time $\tau = 0$ $dx^i/d\tau$ is zero, in the TT gauge also its derivative $d^2x^i/d\tau^2$ vanishes, and therefore $dx^i/d\tau$ remains zero at all times. This shows that in the TT frame, particles which were at rest before the arrival of the wave remain at rest even after the arrival of the wave.⁸

In other words, the coordinates of the TT frame stretch themselves, in response to the arrival of the wave, in such a way that the position of free test masses initially at rest do not change. A physical implementation of the TT gauge can be obtained using the free test masses themselves to mark the coordinates. We can use a test mass to define the origin.⁸ Strictly speaking, this is true only to linear order in $h_{\mu\nu}$ since, if we also include the terms $O(h^2)$ in eq. (1.78), Γ_{00}^i no longer vanishes. However, given that on Earth one typically expects GWs with at most $h = O(10^{-21})$, going beyond the linear order is here of no interest.

⁹ To avoid misunderstandings, observe that ξ^i is a coordinate distance, since it is the difference between the coordinates of two test masses. It is not a proper distance. We will see below what happens if we consider proper distances.

¹⁰ Actually, this is true up to irrelevant correction $O(h^4)$ since $dx^i(\tau)/d\tau = O(h^2)$. Compare with Note 8. of the coordinates, a second one to define, e.g. the point with spatial coordinates $(x = 1, y = 0, z = 0)$, and so on; then, we state that, by definition, these masses still mark the origin, the point $(x = 1, y = 0, z = 0)$, etc. even when the GW is passing.

If the coordinates of test masses initially at rest remain constant, also their coordinate separation must remain constant, for arbitrary finite separation and therefore, of course, also when the separation is small with respect to the typical length-scale of variation of the GW, which is its reduced wavelength. In this limiting case, the equation of the geodesic deviation applies, and it is instructive to check explicitly, from the equation of the geodesic deviation in the TT frame, that the separation ξ^i between the coordinates of two test masses initially at rest does not change. To this purpose, we use the spatial component ($\mu = i$) of eq. (1.69). Since, at $\tau = 0$, $dx^i/d\tau = 0$ by assumption, while $dx^0/d\tau = c$, we get

$$\left. \frac{d^2\xi^i}{d\tau^2} \right|_{\tau=0} = - \left[2c\Gamma_{0\rho}^i \frac{d\xi^\rho}{d\tau} + c^2\xi^\sigma \partial_\sigma \Gamma_{00}^i \right]_{\tau=0} \quad (1.80)$$

However, we already saw that in the TT gauge Γ_{00}^i vanishes identically (at all values of space and time, since in the TT gauge h_{0i} and h_{00} vanish everywhere), and therefore in the first term in bracket, $\Gamma_{0\rho}^i$ is nonvanishing only if ρ is a spatial index, while in the second term $\partial_\sigma \Gamma_{00}^i = 0$. From eq. (1.78), in the TT gauge $\Gamma_{0j}^i = (1/2)\partial_0 h_{ij}$. Therefore, in the TT gauge, the equation of the geodesic deviation gives

$$\left. \frac{d^2\xi^i}{d\tau^2} \right|_{\tau=0} = - \left[\dot{h}_{ij} \frac{d\xi^j}{d\tau} \right]_{\tau=0} \quad (1.81)$$

and therefore, if at $\tau = 0$ we have $d\xi^i/d\tau = 0$, then also $d^2\xi^i/d\tau^2 = 0$, and the separation ξ^i remains constant at all times.⁹

Observe also that, since in the TT gauge we have $h_{00} = h_{0i} = 0$, the proper time on a time-like trajectory $x^\mu(\tau) = (x^0(\tau), x^i(\tau))$ is obtained from

$$\begin{aligned} c^2 d\tau^2 &= c^2 dt^2(\tau) - (\delta_{ij} + h_{ij}^{\text{TT}}) dx^i(\tau) dx^j(\tau) \\ &= c^2 dt^2(\tau) - (\delta_{ij} + h_{ij}^{\text{TT}}) \frac{dx^i}{d\tau} \frac{dx^j}{d\tau} d\tau^2 \end{aligned} \quad (1.82)$$

where we write $x^0(\tau) = ct(\tau)$. However, we have seen for a test mass initially at rest that $dx^i(\tau)/d\tau = 0$ at all times. Then, in the TT gauge the proper time τ measured by a clock sitting on a test mass initially at rest is the same as coordinate time t .¹⁰

The TT gauge illustrates in a particularly neat way the fact that, in general relativity, the physical effects are not expressed by what happens to the coordinates since the theory is invariant under coordinate transformations. At first sight one might be surprised that in the TT gauge the position of test masses does not change as a GW passes by. Of course, this does not mean that the GW had no physical effect, but only that we used the freedom in choosing the coordinate system to define the coordinates in such a way that they do not change. Physical effects can

instead be found monitoring proper distances, or proper times. Consider for instance two events at $(t, x_1, 0, 0)$ and at $(t, x_2, 0, 0)$, respectively. In the TT gauge, the coordinate distance $x_2 - x_1 = L$ remains constant, even if there is a GW propagating along the z axis. However, we see from eq. (1.34) that the proper distance s between these two events is

$$\begin{aligned} s &= (x_2 - x_1) [1 + h_+ \cos(\omega t)]^{1/2} \\ &\simeq L \left[1 + \frac{1}{2} h_+ \cos(\omega t) \right] \end{aligned} \quad (1.83)$$

where in the second line we only retained the term linear in h_+ . Therefore, the proper distance changes periodically in time because of the GW. More generally, if the spatial separation between the two events is given by a vector $\mathbf{L}$, the proper distance is given by $s^2 = L^2 + h_{ij}(t) L_i L_j$ and, to linear order in h , we have $s \simeq L + h_{ij} (L_i L_j / 2L)$, implying

$$\ddot{s} \simeq \frac{1}{2} \ddot{h}_{ij} \frac{L_i}{L} L_j \quad (1.84)$$

Writing $L_i/L = n_i$ and defining s_i from $s = n_i s_i$, we get

$$\begin{aligned} \ddot{s}_i &\simeq \frac{1}{2} \ddot{h}_{ij} L_j \\ &\simeq \frac{1}{2} \ddot{h}_{ij} s_j \end{aligned} \quad (1.85)$$

where in the second line we used the fact that, to lowest order in h , we have $L_j = s_j$. This is the geodesic equation in terms of proper distances, rather than coordinate distances.

If these two test masses are mirrors between which a light beam travels back and forth, it is the proper distance that determines the time taken by the light to make a round trip, so the fact that GWs affect the proper distance means that they can be detected measuring the round-trip time. We will see in detail in Chapter 9 how to analyze an interferometric GW detector in this way.

6.0.9 The proper detector frame

The TT frame has the advantage that GWs have a very simple form in it. However, it is not the frame normally used by an experimentalist to describe its apparatus. In a laboratory, positions are not marked by freely falling particles; rather, after choosing an origin, one ideally uses a rigid ruler to define the coordinates.¹¹ In this frame we expect that a test mass which is free to move (at least along some direction) will be displaced by the passage of the GWs, with respect to the position defined by the rigid ruler and by the test mass which defines the origin.

This is different from what happens in the TT frame, where the positions of the test masses are, by definition, unchanged by an incoming GW.

Conceptually, the simplest laboratory to analyze is one inside a dragfree satellite, so the apparatus is indeed in free fall in the total gravitational field, both of the Earth and of the GWs which might be present.¹¹ A rigid ruler is of course an idealization. When we study resonant bars, in Chapter 8, we will see that a bar (and hence also a ruler) is stretched by an incoming GW. We denote by $\xi_0(t)$ the oscillation amplitude of the fundamental elastic mode of a bar (or of a ruler) of length L , and by ω_0 and γ_0 the frequency and the dissipation coefficient of this elastic mode (with $\gamma_0 \ll \omega_0$), respectively. We will find in eq. (8.32) that, when the bar (or the ruler) is driven by a monochromatic GW with frequency ω and amplitude h_0 ,

$$\xi_0(t) = (2Lh_0\omega^2/\pi^2) \times \frac{(\omega^2 - \omega_0^2) \cos \omega t - \gamma_0 \omega \sin \omega t}{(\omega^2 - \omega_0^2)^2 + \gamma_0^2 \omega^2}$$

For a resonant bar, one chooses the frequency ω_0 of the fundamental elastic mode as close as possible to the frequency ω of the GW that one is searching, so the denominator becomes very small (typically, in a resonant bar $\gamma_0 = \omega_0/Q$ with $Q \sim 10^6$) and $\xi_0(t)$ is enhanced. A ruler is instead rigid, with respect to a GW with a frequency ω , if it has $\omega_0 \gg \omega$; then the above equation becomes $\xi_0(t) \simeq -(\Delta L) \cos \omega t$, with

$$\Delta L/L = (2/\pi^2) h_0 (\omega/\omega_0)^2$$

Then $\omega_0 \gg \omega$ implies that $\Delta L/L \ll h$. We will see in eq. (8.14) that $\omega_0 = \pi v_s/L$, where v_s is the speed of sound in the material. Thus, a rigid ruler can be obtained taking L small. Observe also that all experiments do not measure the absolute length of the apparatus, but rather the length variation induced by a GW. Therefore a very small rigid ruler is all that is needed.

¹² See Ni and Zimmermann (1978), or eq. (4.1) of Thorne (1983).

This means that, if we restrict our attention to a sufficiently small region of space, we can choose coordinates $(t, \mathbf{x})$ so that even in the presence of GWs, the metric is flat,

$$ds^2 \simeq -c^2 dt^2 + \delta_{ij} dx^i dx^j \quad (1.86)$$

We have seen in Section 1.3.2 how to explicitly construct such a freely falling frame along an entire geodesic, using Fermi normal coordinates.

To linear order in $|x^i|$ there are no corrections to this metric, since in a freely falling frame the derivatives of $g_{\mu\nu}$ vanish at the point P around which we expand. Pursuing the expansion to second order, and expressing the second derivatives of $g_{\mu\nu}$ in terms of the Riemann tensor (using again the fact that the Christoffel symbol vanishes at the point P around which we are expanding), the result is

$$ds^2 \simeq -c^2 dt^2 [1 + R_{0i0j} x^i x^j] - 2cdtdx^i \left(\frac{2}{3} R_{0jik} x^j x^k \right) + dx^i dx^j \left[\delta_{ij} - \frac{1}{3} R_{ikjl} x^k x^l \right] \quad (1.87)$$

where the Riemann tensor is evaluated at the point P . We see that, if L_B is the typical variation scale of the metric, so that $R_{\mu\nu\rho\sigma} = O(1/L_B^2)$, the corrections to the flat metric are $O(r^2/L_B^2)$, where $r^2 = x^i x^i$.

For an Earthbound detector, the situation seems more complicated since it is not in free fall with respect to the Earth's gravity (that is, it has an acceleration $\mathbf{a} = -\mathbf{g}$ with respect to

a local inertial frame), and furthermore it rotates relative to local gyroscopes (as illustrated for instance by a Foucault pendulum). The metric in this laboratory frame can be found by explicitly writing the coordinate transformation from the inertial frame to the frame which is accelerating and rotating, and transforming the metric accordingly. The result, up to $O(r^2)$, is

$$\begin{aligned}
 ds^2 \simeq & -c^2 dt^2 \left[1 + \frac{2}{c^2} \mathbf{a} \cdot \mathbf{x} + \frac{1}{c^4} (\mathbf{a} \cdot \mathbf{x})^2 - \frac{1}{c^2} (\boldsymbol{\Omega} \times \mathbf{x})^2 + R_{0i0j} x^i x^j \right] \\
 & + 2cdtdx^i \left[\frac{1}{c} \epsilon_{ijk} \Omega^j x^k - \frac{2}{3} R_{0jik} x^j x^k \right] \\
 & + dx^i dx^j \left[\delta_{ij} - \frac{1}{3} R_{ikjl} x^k x^l \right]
 \end{aligned} \tag{1.88}$$

where a^i is the acceleration of the laboratory with respect to a local free falling frame (i.e. a^i is minus the local "acceleration of gravity") and Ω^i is the angular velocity of the laboratory with respect to local gyroscopes. The term $2\mathbf{a} \cdot \mathbf{x}/c^2$ in eq. (1.88) gives the inertial acceleration. The term $(\mathbf{a} \cdot \mathbf{x}/c^2)^2$ is a gravitational redshift. The term $(\boldsymbol{\Omega} \times \mathbf{x}/c)^2$ gives a Lorentz time dilatation due to the angular velocity of the laboratory. The term $(1/c)\epsilon_{ijk}\Omega^j x^k$ is known as the Sagnac effect. Finally, the terms proportional to the Riemann tensor contain both the effect of slowly varying gravitational backgrounds and the effect of GWs.

The frame where the metric has the form (1.88) is called the proper detector frame, and is implicitly used by experimentalists in a laboratory on Earth: first of all, at zeroth order in r/L_B , this metric reduces to eq. (1.86), i.e. as long as we focus on regions of space smaller than the typical variation scale of the background, we live in the flat space-time of Newtonian physics. This should be contrasted with the TT gauge, where the GW is always present in the background space-time (and no simplification appears when performing an expansion in r/L_B ; the metric in the TT gauge is not an expansion in r/L_B). Next, in the proper frame there are corrections linear in r/L_B . Their effects can be described in terms of Newtonian forces (Newtonian gravity, Coriolis forces, centrifugal forces, etc.). In fact, writing the geodesic equation corresponding to the metric (1.88), and neglecting the terms $O(r^2)$ in the metric, we get

$$\frac{d^2 x^i}{d\tau^2} = -a^i - 2(\boldsymbol{\Omega} \times \mathbf{v})^i + \frac{f^i}{m} + O(x^i) \tag{1.89}$$

The term $-a^i$ is the acceleration of gravity while $-2(\boldsymbol{\Omega} \times \mathbf{v})^i$ is the Coriolis acceleration, and we also added an external force f^i which represents, for instance, the suspension mechanism that compensates the acceleration of gravity. Including the terms $O(x^i x^j)$ in the metric, we get corresponding terms $O(x^i)$ on the right-hand side of eq. (1.89), such as a term $-\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r})^i$ which gives the centrifugal acceleration, etc.¹³ Details aside, the point is that in this frame the evolution of the coordinate $x^i(\tau)$ of a test mass is described by the equations of motion of Newtonian physics, i.e. in terms of forces.

At quadratic order there are also the terms proportional to the Riemann tensor, to which both the slowly varying gravitational field of the Earth, and the GWs contribute. The effect of the GWs is therefore entirely in the term $O(r^2)$. In principle, GWs must therefore compete with a number of other effects, like static gravitational forces, Coriolis forces, etc., that practically are many order of magnitudes larger. What can save the situation is the fact that GWs can have high frequencies, compared to the typical variation time-scales of all other effects. In practice, as we will see when we discuss the various Earthbound detectors, GWs with frequencies lower than a few Hz are hopelessly lost into a sea of much higher Newtonian noises. At higher

frequencies, however, it is possible to have a frequency window ¹⁴ where sufficient isolation from external noises is possible, and an interesting sensitivity to GWs can be obtained.

To isolate the effect of GWs, we can therefore focus on the response of the detector in this frequency window. The acceleration a^i is compensated by the suspension mechanism, and all other effects produce slowly varying changes. We can then neglect all terms in eq. (1.88), and we only retain the part proportional to the Riemann tensor. This means that we can use eq. (1.87), i.e. the metric in the freely falling frame, and we deduce from it the geodesic equation. It is understood that we restrict to the components of $x^i(\tau)$ in the direction in which the test masses are left free to move by the suspension mechanism, and that we consider only the Fourier components of the motion in a frequency window where the ¹³ See eq. (20) of Ni and Zimmermann (1978), for the full expression including relativistic corrections.

¹⁴ Sufficiently high frequency is necessary to overcome the slowly varying Newtonian noises, as well as the seismic noise. However, we will see that above a certain frequency, other types of instrumental noise begin to dominate, and therefore only a frequency window is available for GW detection.

detector is sensitive to GWs. In this frequency window, we will assume that time-varying Newtonian gravitational forces are sufficiently small, so that only GWs contribute to the Riemann tensor.

Rather than using the geodesic equation, it is actually simpler to use the equation for the geodesic deviation, in the form (1.69). We use the fact that $\Gamma_{\nu\rho}^\mu$ vanishes at the expansion point P . Furthermore, since the detector moves non-relativistically, $dx^i/d\tau$ can be neglected with respect to $dx^0/d\tau$, and eq. (1.69) gives

$$\frac{d^2\xi^i}{d\tau^2} + \xi^\sigma \partial_\sigma \Gamma_{00}^i \left(\frac{dx^0}{d\tau} \right)^2 = 0 \quad (1.90)$$

The metric (1.87) depends quadratically on the distance x^i from the point P around which we are expanding, while it depends on t only through the Riemann tensor. In eq. (1.90), $\partial_\sigma \Gamma_{00}^i$ is evaluated at the point P , i.e. at $x^i = 0$. Since $g_{\mu\nu} = \eta_{\mu\nu} + O(x^i x^j)$, a non-zero contribution comes only from the terms in which the two derivatives on the metric present in $\partial_\sigma \Gamma_{00}^i$ are both spatial derivatives, and act on $x^i x^j$. In particular $\partial_0 \Gamma_{00}^i$ evaluated at P gives zero, so $\xi^\sigma \partial_\sigma \Gamma_{00}^i = \xi^j \partial_j \Gamma_{00}^i$. Furthermore, using the fact that, at the point P , both $\Gamma_{\nu\rho}^\mu = 0$ and $\partial_0 \Gamma_{0j}^i = 0$, we have $R^i_{0j0} = \partial_j \Gamma_{00}^i - \partial_0 \Gamma_{0j}^i = \partial_j \Gamma_{00}^i$, so eq. (1.90) becomes

$$\frac{d^2\xi^i}{d\tau^2} = -R^i_{0j0} \xi^j \left(\frac{dx^0}{d\tau} \right)^2 \quad (1.91)$$

If a test mass is initially at rest, it acquires a velocity $dx^i/d\tau = cO(h)$ after the passage of the GW. Therefore

$$\begin{aligned} dt^2 &= d\tau^2 \left[1 + \frac{1}{c^2} \frac{dx^i}{d\tau} \frac{dx^i}{d\tau} \right] \\ &= d\tau^2 [1 + O(h^2)] \end{aligned} \quad (1.92)$$

On the other hand, in eq. (1.91) the Riemann tensor R^i_{0j0} is already $O(h)$, since we are neglecting all effects of the background and we are considering only the GWs. Therefore, if in eq. (1.91) we limit ourselves to linear order in h , we can write $t = \tau$, so $dx^0/d\tau = c$, and eq. (1.91) becomes

$$\ddot{\xi}^i = -c^2 R^i_{0j0} \xi^j, \quad (1.93)$$

where the dot denotes the derivative with respect to the coordinate time t of the proper detector frame.

Next, we should compute the Riemann tensor R^i_{0j0} due to the GWs in the proper detector frame, where eq. (1.93) holds. However, as we discussed below eq. (1.13), in linearized theory the Riemann tensor is invariant, rather than just covariant as in full general relativity, and we can compute it in the frame that we prefer. Clearly, the best choice is to compute it in the TT frame, since in this frame GWs have the simplest form. Then, from eq. (1.13) we immediately find

$$R^i_{0j0} = R_{i0j0} = -\frac{1}{2c^2} \ddot{h}_{ij}^{\text{TT}} \quad (1.94)$$

In conclusion, the equation of the geodesic deviation in the proper detector frame is

$$\ddot{\xi}^i = \frac{1}{2} \ddot{h}_{ij}^{\text{TT}} \xi^j \quad (1.95)$$

This equation is remarkable in its simplicity, since it states that, in the proper detector frame, the effect of GWs on a point particle of mass m can be described in terms of a Newtonian force

$$F_i = \frac{m}{2} \ddot{h}_{ij}^{\text{TT}} \xi^j \quad (1.96)$$

and therefore the response of the detector to GWs can be analyzed in a purely Newtonian language, without any further reference to general relativity.

This means that, in practice, an experimenter in a laboratory on Earth can describe the situation as follows:

- He/she lives in flat space-time, where Newtonian intuition applies.
- There are a number of static or slowly varying Newtonian forces. The acceleration of gravity is compensated by a suspension mechanism, so an Earthbound detector is not free to move in the z direction. Still, it is free to move in the (x, y) plane, or at least in one direction in this plane, depending on the suspension mechanism and detector geometry, and the effect of GWs can show up in the motion of the detector in this free direction. Other slowly varying Newtonian forces perturb the experiment, and one takes great care to minimize their influence. This can be possible at most in a frequency window $[f_{\min}, f_{\max}]$, where $f_{\min}$ is sufficiently large that slowly varying Newtonian noises are under control (as well as the seismic noise, which from a practical point of view is more important at low frequencies), and $f_{\max}$ is not too large, otherwise other noises (e.g., as we will see, the shot noise in an interferometer) begin to dominate.
- Even the effect of GWs on test masses is described in terms of a Newtonian force, given by eq. (1.96).

Before exploring the consequences of eq. (1.95), it is worthwhile to add a few more remarks on its meaning and to stress its limit of validity.

- We have defined ξ^i to be a coordinate separation (rather than a proper distance), since it was introduced as the difference between the coordinates of two nearby geodesics, see eqs. (1.66) and (1.68). With this definition, we have found that eq. (1.95) holds in the proper detector frame. It does not hold in the TT frame, where the geodesic equation is eq. (1.81), and $d^2\xi^i/d\tau^2$ is proportional to $d\xi^i/d\tau$, rather than to ξ^i (consistent with the fact that, in the TT gauge, if $d\xi^i/d\tau$ initially vanishes, the coordinate separation ξ^i does not change).

Nevertheless, in eq. (1.95) the GW in the TT gauge, h_{ij}^{TT} , enters. This comes out because the Riemann tensor is invariant, and we can compute it in any frame, and in particular in the TT gauge, where the GW has the simplest form. Thus, eq. (1.95) combines two nice features: it holds in a frame where our Newtonian intuition is valid, and therefore the description of the detector is more intuitive; at the same time, the Newtonian force due to GWs is expressed in terms of h_{ij} in the TT gauge, where the form of the GW is simpler.

- In the proper detector frame, to a first approximation coordinate distances are the same as proper distances, since the metric is flat, up to $O(r^2/L_B^2)$. Since proper distances are an invariant concept, eq. (1.95) also describes the evolution of proper distances in any other frame (as long as $\tau = t$ to lowest order in h , and the spatial velocities are non-relativistic, otherwise we must use the most general form of the equation of the geodesic deviation, eq. (1.71)). In particular, if we substitute ξ^i with the proper distances s^i , eq. (1.95) holds also in the TT gauge, as we indeed found in eq. (1.85).
- In deriving the equation for the geodesic deviation we have expanded the Christoffel symbols to first order in ξ , neglecting all higher orders, see eqs. (1.68) and (1.69). This is valid as long as $|\xi^i|$ is much smaller than the typical scale over which the gravitational field changes substantially. For a GW, this length-scale is the reduced wavelength λ . Thus, if a detector has a characteristic linear size L , we can discuss its interaction with GWs using the equation of the geodesic deviation, if and only if

$$L \ll \lambda \quad (1.97)$$

As we will see in the chapters on experiments, this condition is satisfied by resonant bar detectors and (in a first approximation) by ground based interferometers. It is not satisfied by proposed spaceborne interferometers such as LISA, nor by the Doppler tracking of spacecraft.

In the former case it is possible to study the detector and its interaction with GWs using a simple and intuitive description in terms of Newtonian forces, supplemented by the "GW force" (1.96). In the latter case, a full general relativistic description, usually performed in the TT frame, is necessary.

- One should observe that, in general, analysis in the TT gauge can be more subtle, since in this frame our intuition can be misleading. For instance, in certain GW detectors (e.g. microwave cavities) one can have objects which are natural to treat as having rigid walls. However, this description is correct only in the proper detector frame. In fact, when a GW passes, the relative position of a freely falling mass and of an object which is not free to move (like the endpoint of a rigid ruler) changes; since, in the TT frame, the coordinates are marked by freely falling masses, an object which is described as rigid in the proper detector frame, gets a deformation $\Delta L/L = O(h)$ in the TT frame when a GW passes.

We can now use eq. (1.95) to study the effect of a GW on test masses. We use the language of the proper detector frame, which is more intuitive, so we consider a ring of test masses initially at rest in the proper detector frame and fix the origin in the center of the ring. Then ξ^i describe the distance of a test mass, with respect to this origin (coordinate distance or proper distance, since we have seen that in the proper detector frame, on a sufficiently small region of space, they are the same). We then use eq. (1.95) to see how these positions change under the effect of a GW. (Alternatively, as discussed above, if we wish to take the point of view of an observer in the TT frame, then eq. (1.95) describes the evolution of proper distances.)

We consider a GW propagating along the z direction, and a ring of test masses located in the (x, y) plane. First of all, for a wave propagating in the z direction, the components of h_{ij}^{TT} with $i = 3$ or $j = 3$ are zero and therefore we see from eq. (1.95) that, if a test particle is initially at $z = 0$, it will remain at $z = 0$, and the displacement will be confined to the (x, y) plane. Therefore, GWs are transverse not only from a mathematical point of view (i.e. they satisfy $\partial_i h_{ij} = 0$), but also in their physical effect: they displace the test masses transversally, with respect to their direction of propagation.¹⁵

To study the motion of the test particles in the (x, y) plane, we first consider the $+$ polarization. Then, at $z = 0$ (choosing the origin of time so that $h_{ij}^{\text{TT}} = 0$ at $t = 0$),

$$h_{ab}^{\text{TT}} = h_+ \sin \omega t \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (1.98)$$

and, as usual, $a, b = 1, 2$ are the indices in the transverse plane. We write $\xi_a(t) = (x_0 + \delta x(t), y_0 + \delta y(t))$, where (x_0, y_0) are the unperturbed positions and $\delta x(t), \delta y(t)$ are the displacements induced by the GW. Then eq. (1.95) becomes

$$\delta \ddot{x} = -\frac{h_+}{2} (x_0 + \delta x) \omega^2 \sin \omega t \quad (1.99)$$

$$\delta \ddot{y} = +\frac{h_+}{2} (y_0 + \delta y) \omega^2 \sin \omega t \quad (1.100)$$

Since δx is $O(h_+)$, on the right-hand side, to linear order in h , the terms $\delta x, \delta y$ can be neglected with respect to the constant parts x_0, y_0 , and the equations are immediately integrated, to give

$$\delta x(t) = \frac{h_+}{2} x_0 \sin \omega t \quad (1.101)$$

$$\delta y(t) = -\frac{h_+}{2} y_0 \sin \omega t \quad (1.102)$$

Similarly, for the cross polarization, we get

$$\delta x(t) = \frac{h_\times}{2} y_0 \sin \omega t \quad (1.103)$$

¹⁵ These two properties are however logically distinct. The condition $\partial_i h_{ij} = 0$ can be imposed exactly, as a gauge condition, and (as we will see in detail in Section 2.2) is basically a consequence of the fact that the graviton is described by a massless spin-2 field. The fact that the Newtonian force is transverse is valid only because we took a test mass at rest, i.e. $u^\mu = (c, 0, 0, 0)$. If we consider test masses with non-zero velocity, the geodesic deviation has a longitudinal term, although suppressed by a factor v^2/c^2 with respect to the transverse term. For example, if we have two test masses with initial spatial separation $\boldsymbol{\xi} = (0, 0, \xi_z)$, both moving with velocity v along the x axis, so that $u^\mu = (1/\gamma)(c, v, 0, 0)$, and the GW propagates along the z direction, the equation of geodesic deviation gives

$$\ddot{\xi}_z = -v^2 R_{zxzx} \xi_z$$

where the dot is the derivative with respect to coordinate time (recall that in this case $d\tau = \gamma dt$). For a GW propagating along z , $R_{zxzx} = -(1/2)\partial_z^2 h_{xx} = -(1/2c^2)\ddot{h}_+$, and therefore

$$\ddot{\xi}_z = \frac{v^2}{2c^2} \ddot{h}_+ \xi_z$$

so the relative displacement of two test masses in the direction of the GWs changes. This is easily understood by performing a boost with velocity $-v$; in the transformed frame the particles are at rest, but the propagation direction of the GW now has both a component along z and a component along x .

ωt	h_+	$h_\times$
0	$\vdots$	
$\pi/2$	$\vdots$	
π	$\vdots$	
$3\pi/2$	$\vdots$	

Fig. 1.1 The deformation of a ring of test masses due to the $+$ and $\times$ polarization.

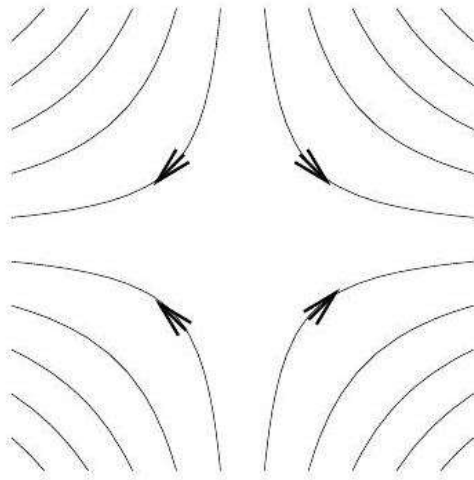


Fig. 1.2 The lines of force corresponding to the $+$ polarization. The arrows show the direction of the force when $\sin \omega t$ is positive. The force reverses when $\sin \omega t$ is negative.

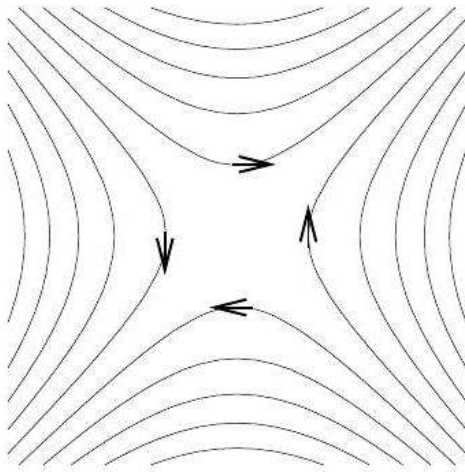


Fig. 1.3 The lines of force corresponding to the $\times$ polarization.

$$\delta y(t) = \frac{h_\times}{2} x_0 \sin \omega t \quad (1.104)$$

The resulting deformation of a ring of test masses located in the (x, y) plane is shown in Fig. 1.1. From eq. (1.96) we see that

$$\partial_i F_i = \frac{m}{2} \ddot{h}_{ij}^{\text{TT}} \delta_{ij} \quad (1.105)$$

and this vanishes, because h_{ij}^{TT} is traceless. Thus, the Newtonian force (1.96) has vanishing divergence, $\nabla \cdot \mathbf{F} = 0$. We can have a pictorial representation of $\mathbf{F}$ drawing its lines of force in the (x, y) plane (defined so that at each point (x, y) they go in the direction of the force, and their density is proportional to the modulus $|\mathbf{F}|$ of the force). The condition $\nabla \cdot \mathbf{F} = 0$ then implies that there are no sources nor sinks for the field lines, just as for the magnetic field in classical electrodynamics. The lines of force in the (x, y) plane obtained from eq. (1.96), for the h_+ and for the $h_\times$ polarization, are shown in Figs. 1.2 and 1.3. The symmetry axes of these lines of force have a typical quadrupolar pattern, with the shape of a + and of a $\times$ sign, respectively, and this is the origin of the denominations "plus" and "cross" polarizations. Observe that Fig. 1.3 is obtained from Fig. 1.2 by performing a rotation of 45 degrees, in agreement with eqs. (1.49) and (1.50).

6.0.10 1.4 The energy of GWs

Our next task is to understand the energy and momentum carried by gravitational waves. The fact that GWs do indeed carry energy and momentum is already clear from the discussion of the interaction of GWs with test masses presented above. We have seen that, in the proper detector frame, an incoming GW sets in motion a ring of test masses initially at rest (and, in fact, the action of the waves on nearby test masses can even be described in terms of a Newtonian force, see eq. (1.96)), so GWs impart kinetic energy to these masses. If, for instance, we connect these masses together with a loose spring with friction, this kinetic energy will be dissipated into heat. Thus, GWs can do work, and conservation of energy requires that the kinetic energy acquired by the test masses must necessarily come from the energy of the GWs. To get the explicit expression of the energy-momentum tensor of GWs we can follow two different routes, one more geometrical and the other more field-theoretical:

(1) Since, according to general relativity, any form of energy contributes to the curvature of space-time, we can ask whether GWs are themselves a source of space-time curvature.

(2) We can treat linearized gravity as any other classical field theory, and apply Noether's theorem, the standard field-theoretical tool that answers this question.

In this section we pursue the former approach, while in Section 2.1 we discuss the latter, and we will see that they both lead to the same answer.

6.0.11 1.4.1 Separation of GWs from the background

To discuss whether GWs curve the background space-time we must broaden our setting. Until now, we have linearized the Einstein equations expanding around the flat metric $\eta_{\mu\nu}$. In this setting the definition of GWs is relatively clear: the background space-time is flat, and the small fluctuations around it have been called "gravitational waves". The term "waves" is justified by the fact that, in a suitable gauge, $h_{\mu\nu}$ indeed satisfies a wave equation. However, to study whether GWs generate a curvature, we cannot define them as perturbation over the flat metric $\eta_{\mu\nu}$, otherwise we exclude from the beginning the possibility that GWs curve the background space-time. Rather, we must allow the background space-time to be dynamical, which means that we would like to define GWs as perturbations over some curved, dynamical, background metric $\bar{g}_{\mu\nu}(x)$, and write ¹⁶

$$g_{\mu\nu}(x) = \bar{g}_{\mu\nu}(x) + h_{\mu\nu}(x), \quad |h_{\mu\nu}| \ll 1 \quad (1.106)$$

However, a problem arises immediately. How do we decide which part of $g_{\mu\nu}$ is the background and which is the fluctuations? In principle, in eq. (1.106) we can move x -dependent

terms from $h_{\mu\nu}$ to $\bar{g}_{\mu\nu}$ or viceversa. The problem did not arise in linearized theory, where the background metric was chosen once and for all to be the constant flat-space metric $\eta_{\mu\nu}$.

As we will see in this section, this problem is not just an abstract issue of principle. On the contrary, the answer to this question allows us to understand properties of GWs such as their energy-momentum tensor, and to get rid of ambiguities concerning whether GWs can be "gauged away" or not.

In the most general setting, there is no unambiguous way to perform a separation of the type (1.106). The total metric $g_{\mu\nu}(x)$ can receive contributions which change, in space and in time, on all possible scales due, for example, to the time-varying Newtonian gravitational fields of nearby masses in movement. The situation is quite similar to that of waves in the sea. In principle, there is no unambiguous way to state which part of the vertical movement of the surface of the water belongs to a given wave, and which part belongs to a "background" originated by the incoherent superposition of perturbations of varied origin. Nevertheless, there are obviously situations where a description of the perturbation of the sea surface in terms of waves is useful, at least at the level of an effective description.

In particular, a natural splitting between the space-time background and gravitational waves arises when there is a clear separation of scales. For example, a natural distinction occurs if, in some coordinate system, we can write the metric as in eq. (1.106), where $\bar{g}_{\mu\nu}$ has a typical scale ¹⁷ For a function $f(x)$ oscillating as e^{ikx} with $k = 2\pi/\lambda = 1/\lambda$, the typical length-scale is λ rather than λ , in the sense that $|df/dx| = (1/\lambda)|f|$.

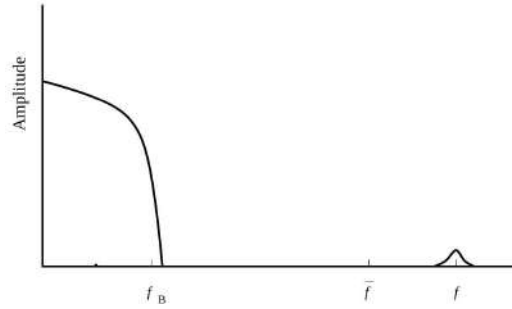


Fig. 1.4 A situation that allows us to separate the metric into a low-frequency background and a small high-frequency perturbation. The background is defined as the part with frequencies $f \ll \bar{f}$ and the GW as the part with $f \gg \bar{f}$. This definition is largely independent of the precise value of $\bar{f}$.

of spatial variation L_B , on top of which small amplitude perturbations are superimposed, characterized by a wavelength λ such that

$$\lambda \ll L_B \quad (1.107)$$

where $\lambda = \lambda/(2\pi)$ is the reduced wavelength. ¹⁷ In this case $h_{\mu\nu}$ has the physical meaning of small ripples on a smooth background. Alternatively, a natural distinction can be made in frequency space, if $\bar{g}_{\mu\nu}$ has frequencies up to a maximum value f_B , while $h_{\mu\nu}$ is peaked around a frequency f such that

$$f \gg f_B \quad (1.108)$$

In this case $h_{\mu\nu}$ is a high-frequency perturbation of a static or slowly varying background. The situation (1.108) is illustrated in Fig. 1.4. We will see below that in this case $h_{\mu\nu}$, in a suitable gauge, obeys a wave equation, and as a consequence its characteristic wavelength and

¹⁶ The condition $|h_{\mu\nu}| \ll 1$ assumes that we are using a coordinate system where the diagonal elements of $\bar{g}_{\mu\nu}$ are $O(1)$, on the region of space-time in which we are interested.

frequency, λ and f , are related by $\lambda = c/f$. However, the scales L_B and f_B that characterize the background are a priori unrelated, so the conditions (1.107) and (1.108) are independent, and it suffices that one of them be satisfied.

We can now ask two questions:

- How this high-frequency (or short wavelength) perturbation propagates in the background space-time with metric $\bar{g}_{\mu\nu}$. The answer to this question will justify the fact that the perturbation $h_{\mu\nu}$ is called a gravitational "wave".
- How this perturbation affects the background metric itself. The answer to this question will allow us to assign an energy-momentum tensor to GWs.

In the next subsections we will address these two questions. First we remark that, traditionally, the separation of the metric into a smooth background plus fluctuations is discussed using the condition (1.107), and the method is called the short-wave expansion. It should be observed, however, that from the point of view of GW detectors, the condition (1.108) is fulfilled instead. Consider for instance a GW with a frequency $f \sim 10^2 - 10^3$ Hz, corresponding to a reduced wavelength $\lambda \simeq 500-50$ km, which are typical GWs that can be searched by groundbased detectors. The Earth's gravitational potential is not spatially smooth over a scale of tens of kms, compared to the GW perturbation. On the contrary, fluctuations in the metric due to local density variations, mountains, etc. are many orders of magnitude bigger than the expected GWs: the Newtonian gravitational potential at the surface of the Earth is in fact $|h_{00}| = 2GM_\oplus/(R_\oplus c^2) \sim 10^{-9}$, while, as we will see, GWs arriving on Earth are expected to have at most $h \sim 10^{-21}$ so even a spatial variation of just one part in 10^{12} due to local inhomogeneities, is large compared to the expected GWs.

On the other hand, these Newtonian gravitational fields are essentially static, and it is much more difficult to find important temporal variations at large frequency scales, e.g. at $f \sim 1$ kHz, since it requires relatively large masses moving at these frequencies. A distinction between background and gravitational waves based on the condition $f \gg f_B$ becomes therefore possible.

Indeed, ground-based GW detectors have a size which is much smaller than the wavelength of the GWs that they are searching. A GW with frequency $f \sim 10^2 - 10^3$ Hz has a reduced wavelength $\sim 500 - 50$ km, which is much bigger than the size of the detector. Therefore, GW detectors do not monitor spatial variations of the gravitational field on lengthscales $L \gg \lambda$. Rather, their output is a time series, which is analyzed in Fourier space looking for temporal variations in their output induced by a passing GW. As we will see in Chapter 9, after suitable isolation, the residual noise due to seismic motion and Newtonian gravitational fields is important only at lower frequencies, say below $O(10)$ Hz. Therefore, we are actually searching for fast temporal variations in the detector output due to GWs, over a background which is slowly varying in time.

6.0.12 1.4.2 How GWs curve the background

We therefore consider the situation in which, in some reference frame, we can separate the metric into a background plus fluctuations, as in eq. (1.106), and this separation is based on the fact that there is a clear distinction of scales either in space, in which case eq. (1.107) applies, or in time, in which case eq. (1.108) applies.

As discussed above, our aim is to understand how the perturbation $h_{\mu\nu}$ propagates, and how it affects the background space-time. To address these questions, we begin by expanding the Einstein equations around the background metric $\bar{g}_{\mu\nu}$. In the expansion we have two small parameters: one is the typical amplitude $h \equiv O(|h_{\mu\nu}|)$, and the second

¹⁸ To be more precise, we should take into account that $\bar{R}_{\mu\nu}$ is non-linear in $\bar{g}_{\mu\nu}$. If $\bar{k}$ separates the low frequency from the high frequency modes, then $\bar{g}_{\mu\nu}$ has only modes up to a typical wave-vector $k_B \simeq 2\pi/L_B$ with $k_B \ll \bar{k}$. The Christoffel symbols of the background are quadratic in the background metric and therefore have modes up to $2k_B$. Terms quadratic in the Christoffel symbols, such as those which appear in the definition of the Ricci tensor, therefore have modes up to $\sim 4k_B$. In any case, if the separation of scales between the background and the GW is clearcut, we still have $4k_B \ll \bar{k}$. In this sense $\bar{R}_{\mu\nu}$ contains only low-frequency modes. is either λ/L_B or f_B/f , depending on whether eq. (1.107) or eq. (1.108) applies. The situation in which $\lambda/L_B \ll 1$ and the situation in which $f_B/f \ll 1$ can be treated in parallel, with the appropriate change of notation, and we will refer generically to both cases as the short-wave expansion.

As a first step, we expand to quadratic order in $h_{\mu\nu}$. It is convenient to cast the Einstein equations in the form

$$R_{\mu\nu} = \frac{8\pi G}{c^4} \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T \right) \quad (1.109)$$

where $T_{\mu\nu}$ is the energy-momentum tensor of matter and T its trace, and then we expand the Ricci tensor to $O(h^2)$,

$$R_{\mu\nu} = \bar{R}_{\mu\nu} + R_{\mu\nu}^{(1)} + R_{\mu\nu}^{(2)} + \dots \quad (1.110)$$

where $\bar{R}_{\mu\nu}$ is constructed with $\bar{g}_{\mu\nu}$ only, $R_{\mu\nu}^{(1)}$ is linear in $h_{\mu\nu}$ and $R_{\mu\nu}^{(2)}$ is quadratic in $h_{\mu\nu}$. The crucial observation now is the following. The quantity $\bar{R}_{\mu\nu}$ is constructed from $\bar{g}_{\mu\nu}$ and therefore contains only low-frequency modes. ¹⁸ $R_{\mu\nu}^{(1)}$ by definition is linear in $h_{\mu\nu}$ and therefore contains only high-frequency modes. $R_{\mu\nu}^{(2)}$ is quadratic in $h_{\mu\nu}$ and therefore contains both high and low frequencies: for instance, in a quadratic term $\sim h_{\mu\nu} h_{\rho\sigma}$ a mode with a high wave-vector $\mathbf{k}_1$ from $h_{\mu\nu}$ can combine with a mode with a high wave-vector $\mathbf{k}_2 \simeq -\mathbf{k}_1$ from $h_{\rho\sigma}$ to give a low wave-vector mode. Therefore the Einstein equations can be split into two separate equations for the low- and high-frequency parts,

$$\bar{R}_{\mu\nu} = - [R_{\mu\nu}^{(2)}]^{\text{Low}} + \frac{8\pi G}{c^4} \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T \right)^{\text{Low}} \quad (1.111)$$

and

$$R_{\mu\nu}^{(1)} = - [R_{\mu\nu}^{(2)}]^{\text{High}} + \frac{8\pi G}{c^4} \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T \right)^{\text{High}} \quad (1.112)$$

The superscript "Low" denotes the projection on the low momenta (i.e. long wavelengths) or on the low frequencies, depending on whether eq. (1.107) or eq. (1.108) applies, and similarly for the superscript "High". The explicit expression for $R_{\mu\nu}^{(1)}$ is computed in Problem 1.1, and is

$$R_{\mu\nu}^{(1)} = \frac{1}{2} (\bar{D}^\alpha \bar{D}_\mu h_{\nu\alpha} + \bar{D}^\alpha \bar{D}_\nu h_{\mu\alpha} - \bar{D}^\alpha \bar{D}_\alpha h_{\mu\nu} - \bar{D}_\nu \bar{D}_\mu h) \quad (1.113)$$

where $\bar{D}_\mu$ is the covariant derivative with respect to the background metric. At quadratic order one finds, after some long algebra,

$$\begin{aligned}
R_{\mu\nu}^{(2)} = & \frac{1}{2} \bar{g}^{\rho\sigma} \bar{g}^{\alpha\beta} \left[\frac{1}{2} \bar{D}_\mu h_{\rho\alpha} \bar{D}_\nu h_{\sigma\beta} + (\bar{D}_\rho h_{\nu\alpha}) (\bar{D}_\sigma h_{\mu\beta} - \bar{D}_\beta h_{\mu\sigma}) \right. \\
& + h_{\rho\alpha} (\bar{D}_\nu \bar{D}_\mu h_{\sigma\beta} + \bar{D}_\beta \bar{D}_\sigma h_{\mu\nu} - \bar{D}_\beta \bar{D}_\nu h_{\mu\sigma} - \bar{D}_\beta \bar{D}_\mu h_{\nu\sigma}) \\
& \left. + \left(\frac{1}{2} \bar{D}_\alpha h_{\rho\sigma} - \bar{D}_\rho h_{\alpha\sigma} \right) (\bar{D}_\nu h_{\mu\beta} + \bar{D}_\mu h_{\nu\beta} - \bar{D}_\beta h_{\mu\nu}) \right]. \quad (1.114)
\end{aligned}$$

In the next subsections we will closely examine eqs. (1.111) and (1.112). We will see that, from eq. (1.111), we can understand what the energymomentum tensor of GWs is, while eq. (1.112) is a wave equation that describes the propagation of $h_{\mu\nu}$ on the background space-time.

First we discuss, from the vantage point of the expansion over a generic curved background, why the expansion over flat space-time presented in Section 1.1 cannot be promoted to a systematic expansion. Consider first the situation in which there is no external matter, $T_{\mu\nu} = 0$. In eqs. (1.111) and (1.112) we have equated terms of different orders in the h -expansion. The reason is that we have a second small expansion parameter, which is λ/L_B (or f_B/f ; as usual, the two cases can be treated in parallel with just a change of notation, and for definiteness we use $\lambda(L_B)$, which can compensate for the smallness in h . The relative strength of these parameters is therefore fixed by the Einstein equations themselves. We use the notation $h = O(|h_{\mu\nu}|)$, while we take $\bar{g}_{\mu\nu} = O(1)$ (in a limited region of space, we can always set $\bar{g}_{\mu\nu} = O(1)$ with a suitable rescaling of the coordinates).

Since we set $T_{\mu\nu} = 0$, we see from eq. (1.111) that $\bar{R}_{\mu\nu}$ is determined only by $[R_{\mu\nu}^{(2)}]^{\text{Low}}$. From the explicit expression (1.114) we see that $R_{\mu\nu}^{(2)}$ is a sum of terms of order $(\partial h)^2$ and of terms of order $h\partial^2 h$. Let us anticipate that, when we compute the projection onto the low modes, these two terms give contributions which are of the same order of magnitude, and one finds that $[R_{\mu\nu}^{(2)}]^{\text{Low}}$ is of order $(\partial h)^2$ (compare with eqs. (1.125) and (1.133) below). Then, in order of magnitude, eq. (1.111) in the absence of matter fields reads

$$\bar{R}_{\mu\nu} \sim (\partial h)^2 \quad (1.115)$$

and expresses the fact that the derivatives of the perturbation $h_{\mu\nu}$ affect the curvature of the background metric $\bar{g}_{\mu\nu}$. The scale of variation of $\bar{g}_{\mu\nu}$ is L_B , while that of h is λ ; therefore, in order of magnitude,

$$\partial \bar{g}_{\mu\nu} \sim \frac{1}{L_B} \quad (1.116)$$

(recall that we took $\bar{g}_{\mu\nu} = O(1)$), while

$$\partial h \sim \frac{h}{\lambda} \quad (1.117)$$

Since the background curvature $\bar{R}_{\mu\nu}$ is constructed from the second derivatives of the background metric, eq. (1.116) implies that

$$\bar{R}_{\mu\nu} \sim \partial^2 \bar{g}_{\mu\nu} \sim \frac{1}{L_B^2} \quad (1.118)$$

while eq. (1.117) gives $(\partial h)^2 \sim (h/\lambda)^2$. Therefore eq. (1.115) gives the relation

$$\frac{1}{L_B^2} \sim \left(\frac{h}{\lambda} \right)^2 \quad (1.119)$$

¹⁹ In special cases one can find exact wave-like solutions of the full nonlinear Einstein equations, see, e.g. Misner, Thorne and Wheeler (1973), Section 35.9, and then there is no need to perform a separation between the background and the waves. However, it would be hopeless to look for exact solutions for the gravitational waves emitted by realistic astrophysical sources. that is,

$$h \sim \frac{\lambda}{L_B}, \quad (\text{curvature determined by GWs}). \quad (1.120)$$

Consider now the opposite limit where $T_{\mu\nu}$ is non-vanishing, and the contribution of GWs to the background curvature is negligible compared to the contribution of matter sources. In this case the total background curvature will be much bigger than the contribution of GWs, $1/L_B^2 \sim h^2/\star^2 + (\text{matter contribution}) \gg h^2/\star^2$, i.e.

$$h \ll \frac{\lambda}{L_B}, \quad (\text{curvature determined by matter}). \quad (1.121)$$

At this point we can understand why the linearized approximation of Section 1.1 cannot be extended beyond linear order. If we force the background metric to be $\eta_{\mu\nu}$, we are actually forcing $1/L_B$ to be strictly equal to zero, and therefore any arbitrarily small, but finite, value of h necessarily violates the condition $h \lesssim \lambda/L_B$, and the expansion in powers of h has no domain of validity. This means that the linearized expansion of the classical theory cannot be promoted to a systematic expansion, and if we want to compute higher-order corrections we cannot insist on a flat background metric.

We can also understand from eqs. (1.120) and (1.121) that the notion of GW is well defined only for small amplitudes, $h \ll 1$. If h becomes of order one, eqs. (1.120) and (1.121) tell us that λ/L_B also becomes at least of order one. Since the separation between λ and L_B is at the basis of the definition of GWs, when h becomes of order one the distinction between GWs and background vanishes. In a general context, there is nothing like "a GW of arbitrary amplitude". ¹⁹

We consider now eq. (1.111). When there is a clear-cut separation between the length-scale λ of the GWs and the length-scale L_B of the background, there is a simple way to perform the projection on the longwavelength modes: we introduce a scale $\bar{l}$ such that $\lambda \ll \bar{l} \ll L_B$, and we average over a spatial volume with side $\bar{l}$. In this way, modes with a wavelength of order L_B remain unaffected, because they are basically constant over the volume used for the averaging, while modes with a reduced wavelength of order λ are oscillating very fast and average to zero. Similarly, if $h_{\mu\nu}$ is a high-frequency perturbation of a quasi-static background, we can introduce a time-scale $\bar{t}$ which is much larger than the period $1/f$ of the GW and much smaller than the typical time-scale $1/f_B$ of the background, and average over this time $\bar{t}$, i.e. over several periods of the GW. We can therefore write eq. (1.111) as

$$\bar{R}_{\mu\nu} = -\langle R_{\mu\nu}^{(2)} \rangle + \frac{8\pi G}{c^4} \left\langle T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T \right\rangle \quad (1.122)$$

where $\langle \dots \rangle$ denotes a spatial average over many reduced wavelengths λ , if eq. (1.107) applies, and a temporal average over several periods $1/f$ of the GW, if rather eq. (1.108) applies.

In the context of general relativity and gravitational-wave physics, the usefulness of introducing some averaging procedure was understood in the 1960s. To put it into a broader theoretical framework, it is useful to realize that what we have done is basically a special case of a general technique, which is known as a renormalization group transformation, and which is nowadays one of the most important tools both in quantum field theory and in statistical

physics. The basic idea is to start from the fundamental equations of a theory and to "integrate out" the fluctuations that take place on a length-scale smaller than l , in order to obtain an effective theory that describes the physics at the length-scale l . These renormalization group transformations can be performed in coordinate space, which is the language that we used above; in momentum space, integrating out the high-momentum modes, in order to get the corresponding low-energy effective action; or in frequency space, in order to eliminate the fast temporal variations and to obtain the effective dynamics of the slowly varying degrees of freedom.

We now define an effective energy-momentum tensor of matter, that we denote by $\bar{T}^{\mu\nu}$, from

$$\left\langle T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T \right\rangle = \bar{T}^{\mu\nu} - \frac{1}{2}\bar{g}_{\mu\nu}\bar{T} \quad (1.123)$$

where $\bar{T} = \bar{g}_{\mu\nu}\bar{T}^{\mu\nu}$ is the trace. By definition, $\bar{T}^{\mu\nu}$ is a purely lowfrequency (or low-momentum) quantity, and is a smoothed form of the matter energy-momentum tensor $T_{\mu\nu}$; for instance, when the separation has been done on the basis of the condition $\lambda \ll L_B$, we can visualize it as a "macroscopic" (with respect to the scale λ) version of the energy-momentum tensor, while $T_{\mu\nu}$ is the fundamental ("microscopic") quantity.²⁰

We also define the quantity $t_{\mu\nu}$ as

$$t_{\mu\nu} = -\frac{c^4}{8\pi G} \left\langle R_{\mu\nu}^{(2)} - \frac{1}{2}\bar{g}_{\mu\nu}R^{(2)} \right\rangle \quad (1.125)$$

where

$$R^{(2)} = \bar{g}^{\mu\nu}R_{\mu\nu}^{(2)} \quad (1.126)$$

and we define its trace as²¹

$$\begin{aligned} t &= \bar{g}^{\mu\nu}t_{\mu\nu} \\ &= +\frac{c^4}{8\pi G} \langle R^{(2)} \rangle \end{aligned} \quad (1.127)$$

To go from the first to the second line, in eq. (1.127), we used the fact that $\bar{g}^{\mu\nu} \langle R_{\mu\nu}^{(2)} \rangle = \langle \bar{g}^{\mu\nu}R_{\mu\nu}^{(2)} \rangle$ since $\bar{g}^{\mu\nu}$ by definition is a purely lowfrequency quantity, as well as the obvious identity $\bar{g}^{\mu\nu}\bar{g}_{\mu\nu} = 4$. Inserting eq. (1.127) into eq. (1.125) (and using again the fact that $\bar{g}_{\mu\nu}$ is constant under the averaging procedure, so $\langle \bar{g}_{\mu\nu}R^{(2)} \rangle = \bar{g}_{\mu\nu} \langle R^{(2)} \rangle$) we see that

$$-\langle R_{\mu\nu}^{(2)} \rangle = \frac{8\pi G}{c^4} \left(t_{\mu\nu} - \frac{1}{2}\bar{g}_{\mu\nu}t \right) \quad (1.128)$$

²⁰ In a typical situation, the fundamental energy-momentum tensor $T^{\mu\nu}$ generated by a macroscopic matter distribution will already be quite smooth, so it will be approximately constant on the scale used for averaging. In this case

$$\begin{aligned} \left\langle T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T \right\rangle &\simeq T_{\mu\nu} - \frac{1}{2}\langle g_{\mu\nu} \rangle T \\ &= T_{\mu\nu} - \frac{1}{2}\bar{g}_{\mu\nu}T \end{aligned} \quad (1.124)$$

and therefore $\bar{T}_{\mu\nu} \simeq T_{\mu\nu}$. However, the definition (1.123) copes with the most general situation.

²¹ Observe that, since $R_{\mu\nu}^{(2)}$ is already quadratic in $h_{\mu\nu}$, we have

$$\begin{aligned} g^{\mu\nu} R_{\mu\nu}^{(2)} &= (\bar{g}^{\mu\nu} + h^{\mu\nu}) R_{\mu\nu}^{(2)} \\ &= \bar{g}^{\mu\nu} R_{\mu\nu}^{(2)} + O(h^3) \end{aligned}$$

and, since we are working up to $O(h^2)$, it is irrelevant whether we define the traces of $R_{\mu\nu}^{(2)}$ and of $t_{\mu\nu}$ contracting with $g^{\mu\nu}$ or with $\bar{g}^{\mu\nu}$.

²² Recall however that we limited ourselves to an expansion of $R_{\mu\nu}$ up to quadratic order in $h_{\mu\nu}$, so all higherorder non-linearities in $h_{\mu\nu}$ have been neglected. We will come back to these non-linear terms in Section 2.2.3 and especially in Chapter 5.

So, we can rewrite eq. (1.122) as

$$\bar{R}_{\mu\nu} = \frac{8\pi G}{c^4} \left(t_{\mu\nu} - \frac{1}{2} \bar{g}_{\mu\nu} t \right) + \frac{8\pi G}{c^4} \left(\bar{T}_{\mu\nu} - \frac{1}{2} \bar{g}_{\mu\nu} \bar{T} \right) \quad (1.129)$$

or, in an equivalent way,

$$\bar{R}_{\mu\nu} - \frac{1}{2} \bar{g}_{\mu\nu} \bar{R} = \frac{8\pi G}{c^4} (\bar{T}_{\mu\nu} + t_{\mu\nu}) \quad (1.130)$$

This can be appropriately called the "coarse-grained" form of the Einstein equations. These equations determine the dynamics of $\bar{g}_{\mu\nu}$, which is the long-wavelength (or low-frequency) part of the metric, in terms of the long-wavelength (or, respectively, low-frequency) part of the matter energy-momentum tensor, $\bar{T}_{\mu\nu}$, and of a tensor $t_{\mu\nu}$ which does not depend on the external matter but only on the gravitational field itself, and is quadratic in $h_{\mu\nu}$.²²

We can then summarize the results of this analysis as follows.

- At a "microscopic" level, there is no fundamental distinction between a background metric and fluctuations over it. The gravitational field is described by all its modes, and its dynamics is fully accounted for by the Einstein equations (1.3).
- If some fluctuations $h_{\mu\nu}$ are clearly distinguishable from the background because their typical length-scale λ is much smaller than the typical length-scale L_B that characterizes the spatial variations of the background, it becomes useful to introduce a "macroscopic" level of description, i.e. an approximate description which is valid at a length-scale $\bar{l}$, such that $\lambda \ll \bar{l}$ (but still $\bar{l} \ll L_B$). This is obtained "integrating out" the short-wavelength degrees of freedom, which, in practice, can be obtained by performing a spatial average of the Einstein equations over a box of size $\bar{l}$, i.e. over several wavelengths λ .

If the separation between fluctuations and background is based on the condition $f_B \ll f$ instead, we integrate out the fast-varying degrees of freedom, performing a temporal average over several periods $1/f$, and we are left with an effective dynamics for the slowly varying degrees of freedom.

- The result of this procedure (which, basically, is a renormalization group transformation) is summarized by eq. (1.130), together with the definitions of $t_{\mu\nu}$ and $\bar{T}_{\mu\nu}$ given in eqs. (1.123) and (1.125). The left-hand side of eq. (1.130) is the Einstein tensor for the slowly varying metric $\bar{g}_{\mu\nu}$. On the right-hand side we find, not surprisingly, a smoothed version of the matter energy-momentum tensor, $\bar{T}_{\mu\nu}$.

The most interesting aspect of eq. (1.130), however, is that it shows that the effect of GWs on the background curvature is formally identical to that of matter with energy-momentum tensor $t^{\mu\nu}$. We are therefore able to assign an energy-momentum tensor to GWs.

- It is useful to observe that $t_{\mu\nu}$ comes out automatically in an averaged form. This averaging procedure is not something that is imposed by hand afterwards. It comes out this way because, to derive the effect of GWs on the background, one is passing from a fundamental, "microscopic", description, to a coarse-grained, "macroscopic" description.

6.0.13 1.4.3 The energy-momentum tensor of GWs

We now compute explicitly $t_{\mu\nu}$, using eq. (1.125) with $R_{\mu\nu}^{(2)}$ given in eq. (1.114). We are interested in the energy and momentum carried by the GWs at large distances from the source (e.g. at the position of the detector), where we can approximate the background space-time as flat. In this case we can simply replace $\bar{D}^\mu \rightarrow \partial^\mu$ in eq. (1.114), so we get

$$\begin{aligned}
 R_{\mu\nu}^{(2)} = \frac{1}{2} \left[\frac{1}{2} \partial_\mu h_{\alpha\beta} \partial_\nu h^{\alpha\beta} + h^{\alpha\beta} \partial_\mu \partial_\nu h_{\alpha\beta} - h^{\alpha\beta} \partial_\nu \partial_\beta h_{\alpha\mu} - h^{\alpha\beta} \partial_\mu \partial_\beta h_{\alpha\nu} \right. \\
 + h^{\alpha\beta} \partial_\alpha \partial_\beta h_{\mu\nu} + \partial^\beta h_\nu^\alpha \partial_\beta h_{\alpha\mu} - \partial^\beta h_\nu^\alpha \partial_\alpha h_{\beta\mu} - \partial_\beta h^{\alpha\beta} \partial_\nu h_{\alpha\mu} \\
 + \partial_\beta h^{\alpha\beta} \partial_\alpha h_{\mu\nu} - \partial_\beta h^{\alpha\beta} \partial_\mu h_{\alpha\nu} - \frac{1}{2} \partial^\alpha h \partial_\alpha h_{\mu\nu} + \frac{1}{2} \partial^\alpha h \partial_\nu h_{\alpha\mu} \\
 \left. + \frac{1}{2} \partial^\alpha h \partial_\mu h_{\alpha\nu} \right] \quad (1.131)
 \end{aligned}$$

As we saw in Section 1.2, the 4×4 symmetric matrix $h_{\mu\nu}$ has 10 degrees of freedom, out of which eight are gauge modes and two are physical modes. Correspondingly, in $t_{\mu\nu}$ one can have in principle contributions both from the physical modes and from the gauge modes. The left-hand side of eq. (1.130), i.e. the Einstein tensor of the background metric $\bar{g}_{\mu\nu}$, is of course a quantity that depends on the coordinate system, since it is a tensor. Thus, in principle there is nothing wrong if, on the right-hand side, we have both physical contributions and coordinatedependent contributions, i.e contributions from gauge modes. The issue is how to distinguish the contribution to $t_{\mu\nu}$ due to the physical modes from the contribution of the gauge modes. The former will give the energy-momentum tensor of the GWs, and describe physical effects that cannot be gauged away, while the latter will be associated with ripples in space-times that are due to the choice of the coordinate system, and that can be made to vanish with an appropriate gauge choice.

The most straightforward way to get the contribution of the physical modes is to make use of the Lorentz gauge condition (1.18). This immediately eliminates four spurious degrees of freedom, leaving us with the two physical degrees of freedom contained in h_{ij}^{TT} and the four gauge modes ξ_μ which satisfy $\square \xi_\mu = 0$, as discussed in Section 1.2. We also choose the ξ_μ so that $h = 0$ (so that only three independent gauge modes remain). Then $\bar{h}_{\mu\nu} = h_{\mu\nu}$ and the Lorentz gauge condition becomes $\partial^\mu h_{\mu\nu} = 0$.

We can now drastically simplify $R_{\mu\nu}^{(2)}$ in eq. (1.131) observing that, inside the spatial or temporal average, the space-time derivative ∂_μ can be integrated by parts, neglecting the boundary term.²³ Performing in-²³ On generic functions, an integration by parts of ∂_t is possible only if we have performed an integral over time, while an integration by parts of ∂_i requires a spatial integral. Recall however that in the Lorentz gauge, outside the source, the equation of motion is a simple wave equation $\square h_{\mu\nu} = 0$. So, for a solution propagating for instance in the z direction, all quantities are functions of the combination $x^0 - z$, where $x^0 = ct$. In expressions such as $\int dz g(x^0 - z) \partial_0 f(x^0 - z)$ we can replace $\partial_0 f$ with $-\partial_z f$, integrate ∂_z by parts and then replace again $\partial_z g$ with $-\partial_0 g$. Therefore, for solutions of the wave equation, a spatial average allows us to integrate by parts not only the spatial derivative but even the time derivative, and similarly for a time average.

Observe also that, in the integration by parts, the boundary terms vanish only when the size of the box used for the integration is infinitely larger than λ . A more precise statement is that the non-zero terms are of higher order in λ/L_B . However, we will only need the result to leading order.

²⁴ The fact that this result holds only to leading order in λ/L_B (see Note 23) is not surprising, since it is only in this limit that the notion of GW is well defined. If λ/L_B approaches one, the distinction between the background and the perturbation fades away, and correspondingly one can no longer assign a gauge-invariant energy-momentum tensor to the perturbations. Integrations by parts and making use of the gauge conditions $\partial^\mu h_{\mu\nu} = 0$ and $h = 0$ and of the equation of motion $\square h_{\alpha\beta} = 0$, it is immediate to see that all terms in eq. (1.131) collapse to zero except the first two, which are related to each other by an integration by parts, giving

$$\langle R_{\mu\nu}^{(2)} \rangle = -\frac{1}{4} \langle \partial_\mu h_{\alpha\beta} \partial_\nu h^{\alpha\beta} \rangle \quad (1.132)$$

while $\langle R^{(2)} \rangle$ vanishes upon integration by parts and using the equation of motion $\square h_{\alpha\beta} = 0$. Recalling the factor $-c^4/(8\pi G)$ from eq. (1.125), we finally find

$$t_{\mu\nu} = \frac{c^4}{32\pi G} \langle \partial_\mu h_{\alpha\beta} \partial_\nu h^{\alpha\beta} \rangle \quad (1.133)$$

We can now verify that the residual gauge modes ξ_μ do not contribute to this expression. In fact, under the gauge transformation (1.8), the variation of $t_{\mu\nu}$ is

$$\begin{aligned} \delta t_{\mu\nu} &= \frac{c^4}{32\pi G} [\langle \partial_\mu h_{\alpha\beta} \partial_\nu (\delta h^{\alpha\beta}) \rangle + (\mu \leftrightarrow \nu)] \\ &= \frac{c^4}{32\pi G} [\langle \partial_\mu h_{\alpha\beta} \partial_\nu (\partial^\alpha \xi^\beta + \partial^\beta \xi^\alpha) \rangle + (\mu \leftrightarrow \nu)] \\ &= \frac{c^4}{16\pi G} [\langle \partial_\mu h_{\alpha\beta} \partial_\nu \partial^\alpha \xi^\beta \rangle + (\mu \leftrightarrow \nu)] \end{aligned} \quad (1.134)$$

and this vanishes since, inside $\langle \dots \rangle$, we can integrate by parts ∂^α , and then we can use the Lorentz condition $\partial^\alpha h_{\alpha\beta} = 0$.²⁴ Therefore $t_{\mu\nu}$ depends only on the physical modes h_{ij}^{TT} , and we can simply replace $h_{\mu\nu}$ in eq. (1.133) with the metric in the TT gauge. In particular, the gaugeinvariant energy density is

$$t^{00} = \frac{c^2}{32\pi G} \langle \dot{h}_{ij}^{\text{TT}} \dot{h}_{ij}^{\text{TT}} \rangle \quad (1.135)$$

(where the dot denotes $\partial_t = c\partial_0$) or, in terms of the amplitudes h_+ and $h_\times$,

$$t^{00} = \frac{c^2}{16\pi G} \langle \dot{h}_+^2 + \dot{h}_\times^2 \rangle \quad (1.136)$$

For a plane wave traveling along the z direction, h_{ij}^{TT} is a function of $t - z/c$, and therefore $t^{01} = t^{02} = 0$ while $\partial_z h_{ij}^{\text{TT}} = -\partial_0 h_{ij}^{\text{TT}} = +\partial^0 h_{ij}^{\text{TT}}$ and therefore

$$t^{03} = t^{00} \quad (1.137)$$

An alternative way of extracting the gauge-invariant part from $t_{\mu\nu}$ is to start from the full expression (1.131) without performing any prior gauge fixing, and consider its variation under a linearized gauge transformation (1.8), where now ξ_μ are generic, rather than being constrained to satisfy $\square \xi_\mu = 0$. Then, with straightforward algebra one finds that

$$t_{\mu\nu} \rightarrow t_{\mu\nu} + \partial_\rho U_{\mu\nu}^\rho \quad (1.138)$$

with $U_{\mu\nu}^\rho$ some tensor. The additional term is a total divergence, and we would like to throw it away inside the average, as we have done above. Here however we must be careful because, since we have not fixed the Lorentz gauge, the metric now does not satisfy a simple wave equation such as $\square h_{\mu\nu} = 0$. Thus, the argument discussed in Note 23, which allowed us to integrate by parts ∂_ρ inside a temporal average, or inside a spatial average, no longer goes through. However, we can integrate by parts ∂_μ inside a space-time average, that we denote as $\langle\langle \dots \rangle\rangle$. Then $\langle\langle \partial_\rho U_{\mu\nu}^\rho \rangle\rangle$ vanishes and $\langle\langle t_{\mu\nu} \rangle\rangle$ is gauge invariant (again to leading order in λ/L_B). Thus, an equivalent way to single out the gauge-invariant part of $t_{\mu\nu}$ is to average it over space-time, and the result gives again eq. (1.133).²⁵

Finally, observe that in eq. (1.130) the left-hand side is covariantly conserved with respect to $\bar{D}^\mu$, i.e. $\bar{D}^\mu (\bar{R}_{\mu\nu} - \frac{1}{2}\bar{g}_{\mu\nu}\bar{R}) = 0$, because of the Bianchi identity. Therefore, we have

$$\bar{D}^\mu (\bar{T}_{\mu\nu} + t_{\mu\nu}) = 0 \quad (1.140)$$

The fact that the covariantly conserved quantity is the sum of $\bar{T}_{\mu\nu}$ and $t_{\mu\nu}$, rather than each one separately, reflects the fact that there is in general exchange of energy and momentum between the matter sources and GWs. At large distances from the source the metric approaches the flat-space metric, so $\bar{D}^\mu$ approaches ∂^μ , while outside the source $\bar{T}_{\mu\nu} = 0$. Then, far from the sources, eq. (1.140) reduces to

$$\partial^\mu t_{\mu\nu} = 0 \quad (1.141)$$

6.0.14 The energy flux

Having obtained the energy-momentum tensor carried by the GWs, it is now straightforward to compute the corresponding energy flux, i.e. the energy of GWs flowing per unit time through a unit surface at a large distance from the source. We start from the conservation of the energy-momentum tensor, $\partial_\mu t^{\mu\nu} = 0$, which implies that

$$\int_V d^3x (\partial_0 t^{00} + \partial_i t^{i0}) = 0 \quad (1.142)$$

where V is a spatial volume in the far region, bounded by a surface S . The GW energy inside the volume V is

$$E_V = \int_V d^3x t^{00} \quad (1.143)$$

so eq. (1.142) can be written as

$$\begin{aligned} \frac{1}{c} \frac{dE_V}{dt} &= - \int_V d^3x \partial_i t^{0i} \\ &= - \int_S dA n_i t^{0i} \end{aligned} \quad (1.144)$$

²⁵ To make the proof simpler we worked at large distance from the source, where the background metric can be taken as flat. The argument can however be repeated in a generic background $\bar{g}_{\mu\nu}$, although it becomes technically more involved. In this case one makes use of the fact that, inside a spacetime average, to lowest order in λ/L_B : (i) Covariant divergences can be integrated by parts discarding the boundary term. In particular, expressions such as

$\langle\langle\bar{D}_\rho U_{\mu\nu}^\rho\rangle\rangle$ vanish. (ii) Covariant derivative commutes. Then one finds, from the full expression (1.114), that

$$t_{\mu\nu} \rightarrow t_{\mu\nu} + \bar{D}_\rho U_{\mu\nu}^\rho \quad (1.139)$$

and therefore $\langle\langle t_{\mu\nu}\rangle\rangle$ is gauge invariant, to leading order in λ/L_B . A further technical subtlety is that, in curved space, the sum of tensors at different points in space-time is not a tensor, so the result of integrating a tensor is also not a tensor. Thus, before integrating over d^4x , one must carry the tensors $t_{\mu\nu}(x)$ back to a single common point using parallel transport along geodesics. Details can be found in the Appendix of Isaacson (1968b), and references therein.

In principle, the same parallel transport procedure should be applied to the spatial and to the temporal averages that we introduced in Section 1.4.2. However, we will always end up computing these averages very far from the sources, where the background spacetime can be taken as flat.

²⁶ More precisely, we take as volume V a spherical shell centered on the source but far away from it, so that both its inner boundary S_1 and its outer boundary S_2 are in the far region, where the gravitational field is given simply by gravitational waves. Then in eq. (1.143) we can limit ourselves to the energy-momentum tensor t^{00} of GWs, neglecting the energy-momentum tensor of the quasi-static gravitational fields, as well as the energy-momentum tensor of matter. Therefore the time derivative of E_V is given by two terms: the energy flowing in through S_1 minus the energy flowing out from S_2 . We are interested in the energy flux through a unit surface at a given distance from the source (say, in the energy flowing through a unit surface of our detector), which for definiteness we choose to be on the outer surface S_2 , so in the following we simply take $S = S_2$. where n^i is the outer normal to the surface and dA is the surface element. ²⁶ Furthermore, outside the source, we can impose the TT gauge. Let S be a spherical surface at a large distance r from the source. Its surface element is $dA = r^2 d\Omega$, and its normal $\hat{n} = \hat{r}$ is the unit vector in the radial direction. Then eq. (1.144) gives

$$\frac{dE_V}{dt} = -c \int dA t^{0r} \quad (1.145)$$

where

$$t^{0r} = \frac{c^4}{32\pi G} \left\langle \partial^0 h_{ij}^{\text{TT}} \frac{\partial}{\partial r} h_{ij}^{\text{TT}} \right\rangle \quad (1.146)$$

A GW propagating radially outward, at sufficiently large distances r , has the general form

$$h_{ij}^{\text{TT}}(t, r) = \frac{1}{r} f_{ij}(t - r/c) \quad (1.147)$$

where $f_{ij}(t - r/c)$ is some function of retarded time $t_{\text{ret}} = t - r/c$. We will prove this result in Section 3.1, but it is in fact completely analogous to the result for electromagnetic waves. Therefore

$$\frac{\partial}{\partial r} h_{ij}^{\text{TT}}(t, r) = -\frac{1}{r^2} f_{ij}(t - r/c) + \frac{1}{r} \frac{\partial}{\partial r} f_{ij}(t - r/c) \quad (1.148)$$

On a function of the combination $t - r/c$ we have

$$\frac{\partial}{\partial r} f_{ij}(t - r/c) = -\frac{1}{c} \frac{\partial}{\partial t} f_{ij}(t - r/c) \quad (1.149)$$

and therefore

$$\begin{aligned}\frac{\partial}{\partial r} h_{ij}^{\text{TT}}(t, r) &= -\partial_0 h_{ij}^{\text{TT}}(t, r) + O(1/r^2) \\ &= +\partial^0 h_{ij}^{\text{TT}}(t, r) + O(1/r^2)\end{aligned}\quad (1.150)$$

Then, from eq. (1.146), we see that at large distances, $t^{0r} = +t^{00}$ (which could also have been derived more simply from eq. (1.137), observing that an observer sitting at large distances from the source sees a plane wavefront), and the energy inside the volume V satisfies

$$\frac{dE_V}{dt} = -c \int dA t^{00} \quad (1.151)$$

The fact that E_V decreases means that the outward-propagating GW carries away an energy flux

$$\begin{aligned}\frac{dE}{dA dt} &= +c t^{00} \\ &= \frac{c^3}{32\pi G} \langle \dot{h}_{ij}^{\text{TT}} \dot{h}_{ij}^{\text{TT}} \rangle\end{aligned}\quad (1.152)$$

or, writing the surface element $dA = r^2 d\Omega$,

$$\frac{dE}{dt} = \frac{c^3 r^2}{32\pi G} \int d\Omega \langle \dot{h}_{ij}^{\text{TT}} \dot{h}_{ij}^{\text{TT}} \rangle \quad (1.153)$$

In terms of h_+ and $h_\times$, we can rewrite the result as

$$\frac{dE}{dA dt} = \frac{c^3}{16\pi G} \langle \dot{h}_+^2 + \dot{h}_\times^2 \rangle \quad (1.154)$$

The total energy flowing through dA between $t = -\infty$ and $t = +\infty$ is therefore ²⁷

$$\frac{dE}{dA} = \frac{c^3}{16\pi G} \int_{-\infty}^{\infty} dt \langle \dot{h}_+^2 + \dot{h}_\times^2 \rangle \quad (1.155)$$

As discussed in the previous section, in the situations relevant for GW detectors the average $\langle \dots \rangle$ in eq. (1.155) is a purely temporal average, over a few periods. Then in eq. (1.155) we can first perform the integral over dt from $-\infty$ to $+\infty$, eliminating therefore any time dependence, and the subsequent temporal average is just the average of a constant. Therefore the average in eq. (1.155) can be omitted, and

$$\frac{dE}{dA} = \frac{c^3}{16\pi G} \int_{-\infty}^{\infty} dt \left(\dot{h}_+^2 + \dot{h}_\times^2 \right) \quad (1.156)$$

Inserting the plane wave expansion of $h_{+,\times}(t)$, given in eqs. (1.52) and (1.48), we get

$$\begin{aligned}\frac{dE}{dA} &= \frac{c^3}{16\pi G} \int_{-\infty}^{\infty} df (2\pi f)^2 \left(\left| \tilde{h}_+(f) \right|^2 + \left| \tilde{h}_\times(f) \right|^2 \right) \\ &= \frac{\pi c^3}{4G} \int_{-\infty}^{\infty} df f^2 \left(\left| \tilde{h}_+(f) \right|^2 + \left| \tilde{h}_\times(f) \right|^2 \right)\end{aligned}\quad (1.157)$$

Since the integrand is even under $f \rightarrow -f$, we can restrict it to physical frequencies, $f > 0$, writing

$$\frac{dE}{dA} = \frac{\pi c^3}{2G} \int_0^\infty df f^2 \left(\left| \tilde{h}_+(f) \right|^2 + \left| \tilde{h}_\times(f) \right|^2 \right) \quad (1.158)$$

Therefore

$$\frac{dE}{dA df} = \frac{\pi c^3}{2G} f^2 \left(\left| \tilde{h}_+(f) \right|^2 + \left| \tilde{h}_\times(f) \right|^2 \right) \quad (1.159)$$

We will always use the convention that the energy spectrum dE/df is the quantity that gives the total energy when it is integrated over the positive frequencies, rather than between $-\infty$ and $+\infty$. Writing $dA = r^2 d\Omega$, and integrating over a sphere surrounding the source, we find the energy spectrum

$$\frac{dE}{df} = \frac{\pi c^3}{2G} f^2 r^2 \int d\Omega \left(\left| \tilde{h}_+(f) \right|^2 + \left| \tilde{h}_\times(f) \right|^2 \right) \quad (1.160)$$

In the same way we can compute the flux of momentum. The momentum of the GWs inside a spherical shell V at large distances from the source is given by

$$P_V^k = \frac{1}{c} \int_V d^3x t^{0k} \quad (1.161)$$

²⁷ The integration over t from $-\infty$ to $+\infty$ is necessary if we want to resolve all possible frequencies. In an experiment one will integrate a signal only over a certain time interval Δt and one has a corresponding resolution in frequency $\Delta f \simeq 1/\Delta t$.

Considering again a GW propagating radially outward, and repeating the same steps leading from eq. (1.142) to eq. (1.154), we get

$$\begin{aligned} c \partial_0 P_V^k &= \int_V d^3x \partial_0 t^{0k} \\ &= - \int_S dA t^{0k} \end{aligned} \quad (1.162)$$

and therefore the momentum flux carried away by the outward-propagating GW is

$$\frac{dP^k}{dA dt} = +t^{0k} \quad (1.163)$$

Inserting the expression (1.133) for t^{0k} , we get

$$\frac{dP^k}{dt} = -\frac{c^3}{32\pi G} r^2 \int d\Omega \left\langle \dot{h}_{ij}^{\text{TT}} \partial^k h_{ij}^{\text{TT}} \right\rangle \quad (1.164)$$

Observe that, if t^{0k} is odd under a parity transformation $\mathbf{x} \rightarrow -\mathbf{x}$, then the angular integral vanishes.

6.0.15 1.5 Propagation in curved space-time

In the last section we have examined the consequences of the low-modes equation, eq. (1.111). We have seen that it determines the dynamics of the background metric $\bar{g}_{\mu\nu}$, and that it allows us to identify the energymomentum tensor of $h_{\mu\nu}$.

We now turn our attention to the high-mode equation (1.112). First of all, we examine it in the limiting case of no external matter, $T_{\mu\nu} = 0$, so eq. (1.112) becomes

$$R_{\mu\nu}^{(1)} = - [R_{\mu\nu}^{(2)}]^{\text{High}} \quad (1.165)$$

We are interested in the leading term in λ/L_B (or in f_B/f ; for definiteness we use λ/L_B). We therefore first perform an order of magnitude estimates of $R_{\mu\nu}^{(1)}$ and of $[R_{\mu\nu}^{(2)}]^{\text{High}}$. In principle, in the short-wave expansion we have two small parameters, $h \equiv O(h_{\mu\nu})$ and $\star/L_B$. Recall however, from eq. (1.120), that when $T_{\mu\nu} = 0$ the Einstein equations fix these two scales to the same order of magnitude. Therefore, in this case we have a single small parameter, that we denote by ϵ ,

$$\epsilon = O(h) = O(\lambda/L_B) \quad (1.166)$$

To simplify notation, we use units $L_B = 1$ when we estimate the order of magnitude of the various terms, so that $h \sim \lambda \sim \epsilon$. From eq. (1.113), the leading term of $R_{\mu\nu}^{(1)}$ is

$$R_{\mu\nu}^{(1)} \sim \partial^2 h \sim \frac{h}{\star^2} \sim \frac{1}{\epsilon} \quad (1.167)$$

while

$$R_{\mu\nu}^{(2)} \sim \partial^2 h^2 \sim \frac{h^2}{\lambda^2} \sim 1 \quad (1.168)$$

So $[R_{\mu\nu}^{(2)}]^{\text{High}}$ is at most $O(1)$ and can be neglected in eq. (1.165), compared to the leading term of $R_{\mu\nu}^{(1)}$, which is $O(1/\epsilon)$. Thus, if we limit ourselves to the leading term, eq. (1.112) simply becomes

$$[R_{\mu\nu}^{(1)}]_{1/\epsilon} = 0 \quad (1.169)$$

where $[\dots]_{1/\epsilon}$ means that we must extract the $O(1/\epsilon)$ part. Equation (1.169) can be written explicitly as

$$\eta^{\rho\sigma} (\partial_\rho \partial_\nu h_{\mu\sigma} + \partial_\rho \partial_\mu h_{\nu\sigma} - \partial_\nu \partial_\mu h_{\rho\sigma} - \partial_\rho \partial_\sigma h_{\mu\nu}) \simeq 0 \quad (1.170)$$

since the $O(1/\epsilon)$ part is obtained substituting the covariant derivatives with ordinary derivatives, and at the same time $\bar{g}^{\rho\sigma}$ in front of the parenthesis can be substituted with $\eta^{\rho\sigma}$, again to leading order in ϵ . This is just a propagation equation for the field $h_{\mu\nu}$ in a flat background, and it is the same equation that governs the propagation in the linearized theory discussed in Section 1.1, so we can again introduce $\bar{h}_{\mu\nu} = h_{\mu\nu} - (1/2)\eta_{\mu\nu}h$, impose the Lorentz gauge, and eq. (1.170) is nothing but

$$\square \bar{h}_{\mu\nu} \simeq 0 \quad (1.171)$$

where $\square = \partial_\mu \partial^\mu$ is the flat space d'Alembertian. So, this is the same as eq. (1.24) with the matter energy-momentum tensor $T_{\mu\nu} = 0$. We therefore discover that the high-frequency equation (1.112) is a wave equation for $h_{\mu\nu}$. We find that the propagation of GWs at $O(h)$ is the same as in the linearized theory because we considered the limit in which GWs are the

only source of curvature. We now turn to the more interesting case in which external matter is present and dominates the curvature, so the low-frequency equation (1.111) becomes

$$\bar{R}_{\mu\nu} - \frac{1}{2}\bar{g}_{\mu\nu}\bar{R} \simeq \frac{8\pi G}{c^4}\bar{T}_{\mu\nu} \quad (1.172)$$

To expand the high-frequency equation (1.112), we recall from eq. (1.121) that in this case $h \ll \lambda/L_B \ll 1$, so the expansions in h and in λ/L_B are different. We keep only the terms linear in h (terms quadratic in h , in a typical situation involving GWs, are utterly negligible), and we expand the result in powers of λ/L_B . If we limit only to the leading and next-to-leading order in λ/L_B , eq. (1.112) becomes simply²⁸

$$R_{\mu\nu}^{(1)} = 0 \quad (1.174)$$

Now $\bar{g}_{\mu\nu}$ is determined by $T_{\mu\nu}$ and is not close to flat, so $R_{\mu\nu}^{(1)}$ is a fully covariant quantity with respect to a non-trivial background metric. As we show in Problem 1.1, in a curved background the equation $R_{\mu\nu}^{(1)} = 0$, written explicitly, reads

$$\bar{g}^{\rho\sigma} (\bar{D}_\rho \bar{D}_\nu h_{\mu\sigma} + \bar{D}_\rho \bar{D}_\mu h_{\nu\sigma} - \bar{D}_\nu \bar{D}_\mu h_{\rho\sigma} - \bar{D}_\rho \bar{D}_\sigma h_{\mu\nu}) = 0 \quad (1.175)$$

The discussion of this equation parallels exactly that of Section 1.1,²⁸ In fact, in eq. (1.112), $[R_{\mu\nu}^{(2)}]^{\text{High}}$ is negligible with respect to $R_{\mu\nu}^{(1)}$, because it has one more power of h . To estimate the order of magnitude of $(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T)^{\text{High}}$, we observe that, if $T_{\mu\nu}$ is smooth, as we expect for macroscopic matter, its high-frequency part will come from the fact that the energy-momentum tensor $T_{\mu\nu}$ depends in general on the metric $g_{\mu\nu}$, and therefore will have a high-frequency component $O(h)$. Besides, also $g_{\mu\nu}T = (\bar{g}_{\mu\nu} + h_{\mu\nu})T$ has a high-frequency part $O(h)$ which comes from multiplying $\bar{g}_{\mu\nu}$ with the $O(h)$ high-frequency part of T , and another high-frequency part $O(h)$ which comes from multiplying $h_{\mu\nu}$ with the low-frequency part of T . So,

$$\left(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\right)^{\text{High}} = O(h/L_B^2) \quad (1.173)$$

Instead, $R_{\mu\nu}^{(1)} \sim \partial^2 h \sim h/\lambda^2$. Then $(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T)^{\text{High}}$ is smaller than $R_{\mu\nu}^{(1)}$ by a factor $O(\lambda^2/L_B^2)$ and, to leading and next-to-leading order in λ/L_B , it does not contribute.

²⁹ There is a slight notational clash here. The bar over $g_{\mu\nu}$ denotes the background metric, while over $h_{\mu\nu}$ it denotes the combination (1.176). with $\eta_{\mu\nu}$ replaced by $\bar{g}_{\mu\nu}$. The equation becomes simpler introducing $h = \bar{g}^{\mu\nu}h_{\mu\nu}$ and²⁹

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\bar{g}_{\mu\nu}h \quad (1.176)$$

We now impose the gauge condition

$$\bar{D}^\nu \bar{h}_{\mu\nu} = 0 \quad (1.177)$$

that we still call the Lorentz gauge. In this gauge the equation $R_{\mu\nu}^{(1)} = 0$ becomes

$$\bar{D}^\rho \bar{D}_\rho \bar{h}_{\mu\nu} + 2\bar{R}_{\mu\rho\nu\sigma}\bar{h}^{\rho\sigma} - \bar{R}_{\mu\rho}\bar{h}_\nu^\rho - \bar{R}_{\nu\rho}\bar{h}_\mu^\rho = 0 \quad (1.178)$$

Outside the matter sources, where $\bar{T}_{\mu\nu} = 0$, the Einstein equation for the background, eq. (1.172), tells us that $\bar{R}_{\mu\nu} = 0$. More precisely, using eq. (1.111), $\bar{R}_{\mu\nu}$ gets a contribution only from the term $[R_{\mu\nu}^{(2)}]^{\text{Low}}$, so $\bar{R}_{\mu\nu} = O(h^2/\lambda^2)$. Then, to linear order in h we can drop the terms $\bar{R}_{\mu\rho}\bar{h}_\nu^\rho$ and $\bar{R}_{\nu\rho}\bar{h}_\mu^\rho$ in eq. (1.178). Furthermore, $\bar{R}_{\mu\rho\nu\sigma}\bar{h}^{\rho\sigma} = O(h/L_B^2)$ while $\bar{D}^\rho \bar{D}_\rho \bar{h}_{\mu\nu} =$

$O(\hbar/\star^2)$. Thus, since we have already restricted ourselves to the leading term and next-to-leading term in λ/L_B (see Note 28) we simply have

$$\bar{D}^\rho \bar{D}_\rho \bar{h}_{\mu\nu} = 0 \quad (1.179)$$

Equations (1.177) and (1.179) determine the propagation of GWs in the curved background, in the limit $\lambda \ll L_B$. In conclusion we find that, after separating the Einstein equations into a low-frequency part and a high-frequency part, the low-frequency part describes the effect of GWs and of external matter on the background space-time, while the high-frequency part gives a wave equation in curved space, which describes the propagation of $h_{\mu\nu}$. This curved-space equation can be solved using the eikonal approximation of geometric optics, as we now discuss.

6.0.16 1.5.1 Geometric optics in curved space

6.0.17 Electromagnetic waves

We first recall how geometric optics works for electromagnetic waves, in a curved space with metric $\bar{g}_{\mu\nu}$.³⁰ The action of the electromagnetic field in this curved space is

$$S = -\frac{1}{4} \int d^4x \sqrt{-\bar{g}} \bar{g}_{\mu\alpha} \bar{g}_{\nu\beta} F^{\mu\nu} F^{\alpha\beta} \quad (1.180)$$

and its variation gives the equations of motion

$$\bar{D}_\mu (\bar{D}^\mu A^\nu - \bar{D}^\nu A^\mu) = 0 \quad (1.181)$$

(we raise and lower the indices with $\bar{g}_{\mu\nu}$), which generalize the flat-space pair of Maxwell equations $\partial_\mu F^{\mu\nu} = 0$. One imposes on the four-vector potential A^μ the curved-space generalization of the Lorentz gauge,

$$\bar{D}_\mu A^\mu = 0 \quad (1.182)$$

From the definition of covariant derivative, $\bar{D}_\mu \bar{D}^\nu A^\mu = \bar{D}^\nu \bar{D}_\mu A^\mu + R^\nu{}_\mu A^\mu$, where $R^\nu{}_\mu$ is the Ricci tensor. The term $\bar{D}^\nu \bar{D}_\mu A^\mu$ vanishes because of the gauge condition, so eq. (1.181) becomes

$$\bar{D}^\rho \bar{D}_\rho A^\mu - R^\mu{}_\rho A^\rho = 0 \quad (1.183)$$

Geometric optics is valid when λ is much smaller than the other lengthscales in the problem. So we must have $\lambda \ll L_B$, with L_B the typical scale of variation of the background metric, and also $\lambda \ll L_c$, where L_c is the characteristic length-scale over which the amplitude, polarization or wavelength of the electromagnetic field change substantially. In particular, λ must be much smaller than the curvature radius of the wavefront.

Under these conditions, we can use the eikonal approximation, which consists in looking for a solution with a phase θ rapidly varying, i.e. θ changes on the scale λ , while the amplitude changes only on the scale L_B or L_c (whichever is smaller), so it is slowly varying. To perform the expansion systematically it is convenient to write

$$A^\mu(x) = [a^\mu(x) + \varepsilon b^\mu(x) + \varepsilon^2 c^\mu(x) + \dots] e^{i\theta(x)/\varepsilon} \quad (1.184)$$

where ε is a fictitious parameter, to be finally set equal to unity, that reminds us that a term to which a factor ε^n is attached, is of order $(\lambda/L)^n$, where L is the smallest between L_B and L_c . An expansion of the form (1.184) is just an ansatz, and its validity is verified by substituting it in the equations.

Since $R^\mu{}_\rho A^\rho = O(A/L_B^2)$, where A is the typical amplitude of A^μ , while $\bar{D}^\rho \bar{D}_\rho A^\mu = O(A/\star^2)$, to leading and next-to-leading order in λ/L_B we can neglect $R^\mu{}_\rho A^\rho$, and the equation of motion is simply

$$\bar{D}^\rho \bar{D}_\rho A^\mu = 0 \quad (1.185)$$

Defining the wave-vector $k_\mu \equiv \partial_\mu \theta$ and plugging the ansatz (1.184) into eq. (1.182) we get, to lowest order,

$$\bar{g}_{\mu\nu} k^\mu a^\nu = 0 \quad (1.186)$$

From eq. (1.185) we get instead, to lowest order,

$$\bar{g}_{\mu\nu} k^\mu k^\nu = 0 \quad (1.187)$$

This is known as the eikonal equation. From this it also follows that $0 = \bar{D}_\nu (k_\mu k^\mu) = 2k^\mu \bar{D}_\nu k_\mu$ (recall again that indices are raised and lowered with $\bar{g}_{\mu\nu}$). Since θ is a scalar, and on a scalar the covariant derivatives commute, we have $\bar{D}_\nu \partial_\mu \theta = \bar{D}_\nu \bar{D}_\mu \theta = \bar{D}_\mu \bar{D}_\nu \theta = \bar{D}_\mu \partial_\nu \theta$,

³¹ Writing $k^\mu = dx^\mu/d\lambda$, where λ is the affine parameter along the geodesic, eq. (1.188) gives the geodesic equation, in the form

$$\frac{d^2 x^\mu}{d\lambda^2} + \Gamma^\mu_{\nu\rho} \frac{dx^\nu}{d\lambda} \frac{dx^\rho}{d\lambda} = 0$$

so we can interchange the indices, $\bar{D}_\nu k_\mu = \bar{D}_\mu k_\nu$, and the condition $k^\mu \bar{D}_\nu k_\mu = 0$ becomes

$$k^\mu \bar{D}_\mu k_\nu = 0 \quad (1.188)$$

This is the geodesic equation in the space-time of the background metric $\bar{g}_{\mu\nu}$,³¹ so eq. (1.188) states that the curves orthogonal to the surfaces of constant phase (the "rays" of the geometric optics approximation) travel along the null geodesics of $\bar{g}_{\mu\nu}$.

To next-to-leading order in ε , eq. (1.185) gives

$$2k_\rho \bar{D}^\rho a^\mu + (\bar{D}^\rho k_\rho) a^\mu = 0 \quad (1.189)$$

It is convenient to introduce the real scalar amplitude $a = (a^\mu a_\mu^*)^{1/2}$ and the polarization vector e^μ defined from $a^\mu = a e^\mu$, so $e^\mu e_\mu^* = 1$. An equation for the scalar amplitude is obtained observing that, on the one hand, one has the trivial identity $k^\mu \partial_\mu (a^2) = 2a k^\mu \partial_\mu a$. On the other hand, on a scalar such as a^2 , ∂_μ can be replaced by $\bar{D}_\mu$, so $k^\mu \partial_\mu (a^2) = k^\mu \bar{D}_\mu (a^\rho a_\rho^*) = -(\bar{D}^\rho k_\rho) a^2$, where we used eq. (1.189). Comparing these two results, we get an equation for the scalar amplitude,

$$k^\mu \partial_\mu a = -\frac{1}{2} (\bar{D}_\mu k^\mu) a \quad (1.190)$$

Finally, to obtain an equation for e^μ , we substitute $a^\mu = a e^\mu$ into eq. (1.189) and we use eq. (1.190). This gives

$$k^\rho \bar{D}_\rho e^\mu = 0 \quad (1.191)$$

Expanding the equations to still higher orders we could determine the corrections $b_\mu, c_\mu, \dots$ to the amplitude in terms of a_μ . Equations (1.186), (1.187), (1.188), (1.190) and (1.191) are the fundamental results of the geometric optics of electromagnetic waves in curved space. Equations (1.187) and (1.188) states that light rays (or photons, in a quantum language) travel along the null geodesics of $\bar{g}_{\mu\nu}$. Equation (1.186) states that the polarization vector e^μ is orthogonal to

the propagation direction, $k_\mu e^\mu = 0$, and eq. (1.191) states that it is parallel-transported along the null geodesics. Finally, eq. (1.190) expresses (in the quantum language) the conservation of the number of photons in the limit of geometric optics. This can be seen rewriting it in the form

$$\bar{D}^\mu (a^2 k^\mu) = 0 \quad (1.192)$$

This shows that the current $j^\mu = a^2 k^\mu$ is covariantly conserved. Its associated conserved charge, according to the Noether theorem that we will recall in Section 2.1.1, is the integral of $a^2 k^0$ over a spatial surface at constant time. In a plane wave, the energy density is proportional to $|\mathbf{E}|^2 + |\mathbf{B}|^2 = 2|\mathbf{E}|^2$. In the gauge $A_0 = 0$, the electric field is $\mathbf{E} = \partial_0 \mathbf{A}$, so its amplitude is $k^0 a$, and the energy density is proportional to $(k^0 a)^2$. Since each photon carries an energy k^0 , we see that $k^0 a^2$ is proportional to the number density of photons, so eq. (1.192) expresses the fact that, in the limit of geometric optics, the number of photons is conserved.

6.0.18 Gravitational waves

We can now discuss the eikonal approximation for GWs. We make the ansatz

$$\bar{h}_{\mu\nu}(x) = [A_{\mu\nu}(x) + \varepsilon B_{\mu\nu}(x) + \dots] e^{i\theta(x)/\varepsilon} \quad (1.193)$$

Again we define $k_\mu = \partial_\mu \theta$, and we write $A_{\mu\nu} = A e_{\mu\nu}$ where the polarization tensor $e_{\mu\nu}$ is normalized as $e^{\mu\nu} e_{\mu\nu}^* = 1$, and A is the scalar amplitude. Substituting this ansatz into eqs. (1.177) and (1.179) and repeating basically the same steps as for the electromagnetic case, we find that k_μ still obeys eqs. (1.187) and (1.188), so gravitons travel along the null geodesic of $\bar{g}_{\mu\nu}$. Just as for photons, the scalar amplitude satisfies

$$k^\mu \partial_\mu A = -\frac{1}{2} (\bar{D}_\mu k^\mu) A \quad (1.194)$$

which can be written as $\bar{D}^\mu (A^2 k^\mu) = 0$, and gives the conservation of the number of gravitons. Finally, the polarization tensor satisfies

$$k^\nu e_{\mu\nu} = 0 \quad (1.195)$$

$$k^\rho \bar{D}_\rho e_{\mu\nu} = 0 \quad (1.196)$$

so it is transverse and is parallel-propagated along the null geodesics.

Since gravitons propagate along null geodesics, just as photons, their propagation through curved space-time is the same as the propagation of photons, as long as geometric optics applies. For instance, they suffer gravitational deflection when passing near a massive body, with the same deflection angle as photons, and they undergo the same redshift in a gravitational potential.³²

One practical difference concerning the lensing of gravitational and electromagnetic waves is however worth observing. Both type of waves can in principle be lensed by a large mass situated between the source and the observer. When the different images of the source cannot be resolved we are in the regime of microlensing, where we have a single image which is magnified. The amplification factor $\mathcal{A}$ in the energy density, computed within geometric optics, is³³

$$\mathcal{A} = \frac{u^2 + 2}{u\sqrt{u^2 + 4}} \simeq \frac{1}{u} \quad (1.197)$$

where $u = \beta/\theta_E$, β is the angle of the source with respect to the observerlens axis (see Fig. 1.5), and θ_E is the Einstein angle, given by

$$\theta_E^2 = \frac{2R_S d_{SL}}{d_{OL}(d_{SL} + d_{OL})} \quad (1.198)$$

where R_S is the Schwarzschild radius of the lens and d_{SL} and d_{OL} are the source-lens and observer-lens distances, respectively. The second equality in eq. (1.197) holds when $u \ll 1$, i.e. when the source, lens and observer are well aligned, and in this case the amplification factor is large. In fact, it even becomes formally infinite if $u = 0$, i.e. when the source, lens and observer are perfectly aligned. However, the geometric ³² The redshift of gravitons in a FRW cosmological model will be discussed explicitly in Section 4.1.4.

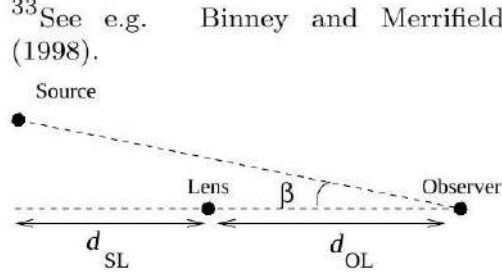


Fig. 1.5 The geometry for the lensing of gravitational or electromagnetic waves.

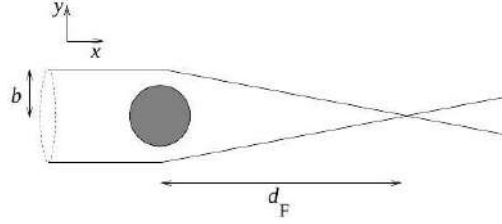


Fig. 1.6 The focusing of GWs from a source. ³⁴ A more accurate estimate can be obtained taking into account the detailed internal structure of the lens, see Bontz and Haugan (1981). optics approximation breaks down near the caustics, i.e. when the light rays coming from the source cross each other, since there the scale of variation of the wavefront is no longer small compared to λ .

The actual behavior near the caustics can be obtained taking into account diffraction (or, in a quantum language, the uncertainty principle). In order of magnitude the effect can be estimated as follows. Consider a circular ring of rays, which is part of a plane wavefront, arriving with impact parameter b on a star of mass M which acts as a gravitational lens, as in Fig. 1.6. The deflection angle due to the gravitational field of the star is given by the classical Einstein result, $\theta = 2R_S/b$, where $R_S = 2GM/c^2$ is the Schwarzschild radius of the star. The whole circular ring of rays with impact parameter b is then focused on a single point, at a focal distance d_F given by $b/d_F = \tan \theta \simeq \theta$, i.e.

$$d_F \simeq \frac{b^2}{2R_S} \quad (1.199)$$

The fact that a one-dimensional surface is focused onto a point is responsible for the formal infinite enhancement of the luminosity. In practice, however, diffraction forbids such a perfect focusing. Indeed, when we state that a photon has impact parameter b , we are implicitly assuming that the error Δy on its transverse position is smaller than b . Then, by the Heisenberg principle, it has an uncertainty on the transverse momentum $\Delta k_y \gtrsim \hbar/b$, and an angular spreading

$$\Delta\theta_s \simeq \frac{\Delta k_y}{k_x} \gtrsim \frac{\lambda}{b} \quad (1.200)$$

Propagating to a distance d_F this induces a transverse spread

$$\Delta y_s \simeq d_F \Delta\theta_s \gtrsim \frac{\star b}{2R_S} \quad (1.201)$$

The focusing is substantial only if $\Delta y_s \ll b$, which gives

$$\star \ll 2R_S \quad (1.202)$$

For a lens with a mass of order $M_\odot$, this means that a substantial focusing is possible only for waves with $\lambda \ll O(6)\text{km}$, i.e. a frequency $f \gg O(10)\text{kHz}$.³⁴ For electromagnetic waves in the visible spectrum, this condition is very well satisfied, and microlensing is indeed commonly observed. For GWs, however, we will see that no astrophysical source is expected to produce waves with frequencies much larger than $O(10)\text{kHz}$, and no significant amplification can be obtained for these waves from typical stellar-mass lenses.

6.0.19 1.5.2 Absorption and scattering of GWs

Finally, GWs are insensitive to absorption and scattering, during their propagation from astrophysical sources to the observer, because of the smallness of the gravitational cross-section. To make a quantitative

estimate, recall that the mean free path l of a particle scattering off an ensemble of target with number density n and cross-section σ , is given by

$$l = \frac{1}{n\sigma} \quad (1.203)$$

For a graviton of energy E , the scattering cross-section off a target of mass m is (using units $\hbar = c = 1$) $\sigma \sim G_N^2 s$, where s is the square of the center-of-mass energy.³⁵ Consider for instance the scattering of a graviton with four-momentum $k^\mu = (E, 0, 0, E)$ off a nucleon at rest, with four-momentum $p^\mu = (m_n, 0, 0, 0)$, where the mass $m_n \sim 1\text{GeV}$. Then the square of the center-of-mass energy is $s = -(p + k)^2 = m_n^2 + 2m_n E$. For all astrophysically plausible values of the GW frequency f_{gw} , $E = \hbar\omega_{\text{gw}}$ is totally negligible with respect to m_n . For instance, if $f_{\text{gw}} = 1\text{kHz}$, we have $\hbar\omega_{\text{gw}} = O(10^{-21})\text{GeV}$; so $s \simeq m_n^2$ and

$$\sigma \sim G_N^2 m_n^2 \quad (1.204)$$

We can now compare the absorption of electromagnetic and of gravitational waves, for instance from the Sun. For photons in a neutral plasma, such as the Sun, the most important process is the Thompson scattering on electrons, which has a cross-section $\sigma = 8\pi\alpha^2/(3m_e^2)$, where $\alpha \simeq 1/137$ is the fine-structure constant, and we use units $\hbar = c = 1$. Inserting the numerical values, we see that the Thomson cross-section for scattering of photons on electrons is larger than the gravitational cross-section for scattering of gravitons on nucleons, eq. (1.204), by a huge factor $O(10^{80})$. The number density n_e of electrons in the Sun (which is relevant for computing the electromagnetic mean free path due to electron-photon scattering) is about the same as the number density of protons, relevant to compute the gravitational mean free path due to proton-graviton scattering, so the mean-free path for gravitons in the Sun is larger by a factor $O(10^{80})$ compared to that of photons! Using the value of n_e of the Sun, one finds that the photon mean free path inside the Sun is $O(1)\text{cm}$. Using this value of l , one can show that a photon produced by thermonuclear reactions in the Sun core takes about $3 \times 10^4\text{yr}$ to reach the surface of the Sun and finally escape.³⁶ For a graviton the mean free path inside the Sun is

$O(10^{80})$ cm, which is huge even compared to the observable size of the Universe (consider that $1\text{Gpc} \simeq 3 \times 10^{27}$ cm). Therefore, for a GW the Sun is completely transparent.

Significant absorption of GWs can take place if the wave impinges on a black hole. In this case, we can use the result for the capture cross section of a relativistic particle by a black hole with Schwarzschild radius R_S , $\sigma = (27/4)\pi R_S^2$.³⁷ Another possibility is that the GW impinges on a neutron star, just with the right frequency to excite one of its normal modes. In this case, the wave interacts coherently with the neutron star (while, in the above estimate of scattering in the Sun, we computed the incoherent scattering off the single protons). However, as we see in microlensing experiments, the probability that a compact object lies on the path from an astrophysical source to the Earth is very small.³⁵ This can be shown most easily using the field-theoretical methods of Chapter 2, observing that the graviton-matter-matter vertex is proportional to $G_N^{1/2}$. In the Feynman diagram for the graviton-matter scattering there are two such vertices, so the amplitude is proportional to G_N and the crosssection to G_N^2 . The dependence on s is then fixed by dimensional arguments observing that, in units $\hbar = c = 1$, G_N is an inverse mass squared, as well as from Lorentz invariance, that dictates that the energy dependence is through the Lorentz-invariant quantity s . This is the same result that holds (at energies $E \ll M_W$) for neutrinos, with the Fermi constant G_F replacing Newton's constant G_N .

³⁶ See e.g. Exercise 1.2 of Maggiore (2005).

6.0.20 1.6 Solved problems

6.0.21 Problem 1.1. Linearization of the Riemann tensor in curved space

In this problem we compute the Riemann and Ricci tensors, linearized to first order in $h_{\mu\nu}$ over a generic curved background $\bar{g}_{\mu\nu}$. The inversion of $g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}$ is $g^{\mu\nu} = \bar{g}^{\mu\nu} - h^{\mu\nu} + O(h^2)$. Then it is straightforward to find that, to $O(h)$, the linearization of the Christoffel symbol gives

$$\Gamma_{\nu\rho}^\mu = \bar{\Gamma}_{\nu\rho}^\mu + \frac{1}{2}\bar{g}^{\mu\sigma}(\bar{D}_\nu h_{\rho\sigma} + \bar{D}_\rho h_{\nu\sigma} - \bar{D}_\sigma h_{\nu\rho}) \quad (1.205)$$

The calculation of the Riemann tensor at a given point x is enormously simplified if we perform it in a coordinate system where $\bar{\Gamma}_{\nu\rho}^\mu(x) = 0$ (paying attention, of course, to the fact that the derivatives of $\bar{\Gamma}_{\nu\rho}^\mu$ are non-zero!) We see from eq. (1.205) that in this frame $\Gamma_{\nu\rho}^\mu = O(h)$ and therefore the terms $\sim \Gamma\Gamma$ in the Riemann tensor are $O(h^2)$, so they do not contribute to $O(h)$, and we simply have $R_{\nu\rho\sigma}^\mu = \partial_\rho \Gamma_{\nu\sigma}^\mu - \partial_\sigma \Gamma_{\nu\rho}^\mu + O(h^2)$. Furthermore, substituting eq. (1.205) for $\Gamma_{\nu\rho}^\mu$, we only need to keep the terms where the derivative acts on the background Christoffel symbols $\bar{\Gamma}_{\nu\rho}^\mu$, since background Christoffel symbols on which no derivative act give zero. Since $\bar{\Gamma}_{\nu\rho}^\mu(x) = 0$, in the final expression (i.e. after having performed all derivatives) we are free to write derivatives as covariant derivatives with respect to the background, and we obtain an expression valid in all coordinate systems.

Then, in the frame where $\bar{\Gamma}_{\nu\rho}^\mu(x) = 0$, we get

$$\begin{aligned} R_{\mu\nu\rho\sigma} = \bar{R}_{\mu\nu\rho\sigma} + \frac{1}{2}(\bar{D}_\rho \bar{D}_\nu h_{\mu\sigma} + \bar{D}_\sigma \bar{D}_\mu h_{\nu\rho} - \bar{D}_\rho \bar{D}_\mu h_{\nu\sigma} - \bar{D}_\sigma \bar{D}_\nu h_{\mu\rho} \\ + h_\mu{}^\tau \bar{R}_{\tau\nu\rho\sigma} - h_\nu{}^\tau \bar{R}_{\tau\mu\rho\sigma}). \end{aligned} \quad (1.206)$$

We have performed the computation in a special frame. Since, however, the final result is expressed in terms of covariant quantities, this expression holds in any frame. The linearization

³⁷ See e.g. Landau and Lifshitz, Vol. II (1979), Section 102.

of the Ricci tensor is then obtained using $R_{\mu\nu} = g^{\alpha\beta} R_{\alpha\mu\beta\nu} = (\bar{g}^{\alpha\beta} - h^{\alpha\beta}) (\bar{R}_{\alpha\mu\beta\nu} + R_{\alpha\mu\beta\nu}^{(1)})$, so that the part linear in $h_{\mu\nu}$ is $R_{\mu\nu}^{(1)} = \bar{g}^{\alpha\beta} R_{\alpha\mu\beta\nu}^{(1)} - h^{\alpha\beta} \bar{R}_{\alpha\mu\beta\nu}$. Then we get ³⁸

$$R_{\mu\nu}^{(1)} = \frac{1}{2} (\bar{D}^\alpha \bar{D}_\mu h_{\nu\alpha} + \bar{D}^\alpha \bar{D}_\nu h_{\mu\alpha} - \bar{D}^\alpha \bar{D}_\alpha h_{\mu\nu} - \bar{D}_\nu \bar{D}_\mu h) \quad (1.207)$$

variant derivative, see e.g. Weinberg (1972), eq. (6.5.3). Pay attention to the fact that Weinberg has the opposite sign convention for the Riemann tensor. where $h = \bar{g}^{\alpha\beta} h_{\alpha\beta}$. Observe that $\bar{D}_\nu \bar{D}_\mu h = \bar{D}_\nu \partial_\mu h = \partial_\nu \partial_\mu h - \bar{\Gamma}_{\nu\mu}^\rho \partial_\rho h$ is symmetric under the exchange $\mu \leftrightarrow \nu$, and therefore $R_{\mu\nu}^{(1)}$ is also symmetric, as it should be.

If the background is flat, $\bar{g}_{\mu\nu} = \eta_{\mu\nu}$, the covariant derivatives become ordinary derivatives and therefore commute. If we impose the Lorentz gauge condition $\partial^\mu \bar{h}_{\mu\nu} = 0$, i.e. $\partial^\mu h_{\mu\nu} = (1/2) \partial_\nu h$, the linearized Ricci tensor takes a very simple form,

$$R_{\mu\nu}^{(1)} = -\frac{1}{2} \square h_{\mu\nu}, \quad (\text{flat background}) \quad (1.208)$$

6.0.22 Problem 1.2. Gauge transformation of $h_{\mu\nu}$ and $R_{\mu\nu\rho\sigma}^{(1)}$

In the text we showed that, when the background space-time is $\eta_{\mu\nu}$, the resulting linearized theory has a gauge symmetry given by eq. (1.8), and that the linearized Riemann tensor $R_{\mu\nu\rho\sigma}^{(1)}$ is gauge-invariant. It is interesting to see how these results are modified when the background space-time is curved.

Under the coordinate transformation $x^\mu \rightarrow x'^\mu = x^\mu + \xi^\mu(x)$, the usual transformation law of the metric is

$$g_{\mu\nu}(x) \rightarrow g'_{\mu\nu}(x') = \frac{\partial x^\rho}{\partial x'^\mu} \frac{\partial x^\sigma}{\partial x'^\nu} g_{\rho\sigma}(x) \quad (1.209)$$

Writing $g'_{\mu\nu}(x) = g'_{\mu\nu}(x' - \xi)$ and expanding to first order in ξ , we have

$$g'_{\mu\nu}(x) \simeq g'_{\mu\nu}(x') - \xi^\rho \partial_\rho g'_{\mu\nu} \quad (1.210)$$

Combining this with eq. (1.209) we get

$$g'_{\mu\nu}(x) = g_{\mu\nu}(x) - (D_\mu \xi_\nu + D_\nu \xi_\mu) \quad (1.211)$$

where the covariant derivative is

$$D_\mu \xi_\nu = \partial_\mu \xi_\nu - \Gamma_{\mu\nu}^\rho \xi_\rho \quad (1.212)$$

Equation (1.211) gives the lowest-order term in the small parameter $|D_\mu \xi_\nu|$. Similar to what we have done in linearized theory, we restrict ourselves to $|D_\mu \xi_\nu| \lesssim h$, where $h = O(|h_{\mu\nu}|)$. Since $\Gamma_{\mu\nu}^\rho = O(\partial g_{\mu\nu}) = O(1/L_B)$, where L_B is the typical variation scale of the background metric, the condition $|D_\mu \xi_\nu| \lesssim h$ means that both

$$|\partial \xi| \lesssim h, \quad \text{and} \quad \xi \lesssim h L_B \quad (1.213)$$

must be satisfied. (We use the notation $\xi = O(|\xi^\mu|)$.) In the case of a flat background metric, discussed in Section 1.1, we have $L_B = \infty$ and therefore we found only the condition $|\partial \xi| \lesssim h$.

A generic function ξ^μ has both low-frequency and high-frequency modes, without a clear separation between them, and therefore in the transformed metric $g'_{\mu\nu}(x)$ the separation between the background and the GW in general disappears. It is therefore more useful to restrict ourselves to functions ξ^μ which maintain a clear-cut separation between low- and high-frequencies.

In particular, we can consider a function ξ^μ that has only high-frequency modes. We observe that

$$D_\mu \xi_\nu + D_\nu \xi_\mu = (\bar{D}_\mu \xi_\nu + \bar{D}_\nu \xi_\mu) + \xi^\rho (\bar{D}_\mu h_{\nu\rho} + \bar{D}_\nu h_{\mu\rho} - \bar{D}_\rho h_{\mu\nu}) \quad (1.214)$$

where we have used the expansion (1.205) for the Christoffel symbol. Since $\bar{\Gamma}_{\nu\rho}^\mu$ is a purely low-frequency term, and ξ^μ has only high frequencies, the terms $\bar{D}_\mu \xi_\nu$ are purely high-frequency, and therefore contribute to the transformation of $h_{\mu\nu}$ rather than of $\bar{g}_{\mu\nu}$.³⁹ Therefore under such a transformation we have $\bar{g}'_{\mu\nu}(x) = \bar{g}_{\mu\nu}(x)$ and

$$h'_{\mu\nu}(x) \simeq h_{\mu\nu}(x) - (\bar{D}_\mu \xi_\nu + \bar{D}_\nu \xi_\mu) \quad (1.215)$$

under the condition that ξ^μ contains only high-frequencies and that it satisfies

$$|\bar{D}_\mu \xi_\nu| \leq |h_{\mu\nu}| \quad (1.216)$$

³⁹ Note that, since $|\partial\xi| \lesssim h$ and $|\xi| \lesssim hL_B$, the terms $\xi\bar{D}h$ in eq. (1.214) are at most $O(h^2L_B/)$. When the curvature is dominated by matter we found in Section 1.4.2 that $h \ll \lambda/L_B$, and therefore in eq. (1.214) the term $\xi\bar{D}h$ is negligible with respect to the term $\bar{D}_\mu \xi_\nu$. In the opposite limit of curvature dominated by GWs we have $h \sim 1/L_B$ and the term $\xi\bar{D}h$ becomes of the same order as $D_\mu \xi_\nu$, but not larger; therefore, even in this limit, the order-of-magnitude estimates given below are not affected when neglecting the term $\xi\bar{D}h$.

Equation (1.215) has the form of a gauge transformation for a symmetric tensor field $h_{\mu\nu}$ on a curved space described by $\bar{g}_{\mu\nu}$. In this sense, we have a gauge theory for a spin-2 field in a curved background. We next compute how the linearized Riemann tensor transforms under this gauge transformation. We write

$$R_{\mu\nu\rho\sigma} = \bar{R}_{\mu\nu\rho\sigma} + R_{\mu\nu\rho\sigma}^{(1)} + O(h^2) \quad (1.217)$$

where $\bar{R}_{\mu\nu\rho\sigma}$ is the Riemann tensor of the background and we raise and lower indices with $\bar{g}_{\mu\nu}$. The explicit calculation performed in Problem 1.1 gives

$$2R_{\mu\nu\rho\sigma}^{(1)} = \bar{D}_\rho \bar{D}_\nu h_{\mu\sigma} + \bar{D}_\sigma \bar{D}_\mu h_{\nu\rho} - \bar{D}_\rho \bar{D}_\mu h_{\nu\sigma} - \bar{D}_\sigma \bar{D}_\nu h_{\mu\rho} + h_\mu{}^\tau \bar{R}_{\tau\nu\rho\sigma} - h_\nu{}^\tau \bar{R}_{\tau\mu\rho\sigma},$$

which generalizes eq. (1.13) to a curved background. Under the gauge transformation (1.215) the variation of $R_{\mu\nu\rho\sigma}^{(1)}$ is given by

$$\begin{aligned} \delta R_{\mu\nu\rho\sigma}^{(1)} = & \xi^\tau \bar{D}_\tau \bar{R}_{\mu\nu\rho\sigma} + \bar{R}_{\tau\nu\rho\sigma} \bar{D}_\mu \xi^\tau - \bar{R}_{\tau\mu\rho\sigma} \bar{D}_\nu \xi^\tau \\ & + \bar{R}_{\mu\nu\tau\rho} \bar{D}_\sigma \xi^\tau - \bar{R}_{\mu\nu\tau\sigma} \bar{D}_\rho \xi^\tau \end{aligned} \quad (1.219)$$

Therefore, if the background is not flat, $R_{\mu\nu\rho\sigma}^{(1)}$ is no longer gauge invariant under the gauge transformations of linearized theory. However, let us estimate the order of magnitude of the various terms. For the background we have

$$\bar{R}_{\mu\nu\rho\sigma} \sim \partial^2 \bar{g}_{\mu\nu} \sim \frac{1}{L_B^2}, \quad \bar{D}_\tau \bar{R}_{\mu\nu\rho\sigma} \sim \partial^3 \bar{g}_{\mu\nu} \sim \frac{1}{L_B^3} \quad (1.220)$$

Equation (1.216) gives instead $\xi \sim hL_B$, $\bar{D}\xi \sim h$ and we therefore see that

$$\delta R_{\mu\nu\rho\sigma}^{(1)} \sim \frac{h}{L_B^2} \quad (1.221)$$

This means that the variation of $R_{\mu\nu\rho\sigma}^{(1)}$ is much smaller than $R_{\mu\nu\rho\sigma}^{(1)}$ itself, since

$$R_{\mu\nu\rho\sigma}^{(1)} \sim \partial^2 h \sim \frac{h}{\star^2} \quad (1.222)$$

and therefore

$$\delta R_{\mu\nu\rho\sigma}^{(1)} \sim \frac{\lambda^2}{L_B^2} R_{\mu\nu\rho\sigma}^{(1)} \quad (1.223)$$

We conclude that $R_{\mu\nu\rho\sigma}^{(1)}$ is approximately gauge-invariant in the limit $\lambda/L_B \ll 1$. More precisely, its leading term in an expansion in powers of λ/L_B is gaugeinvariant. Therefore, in the limit $\lambda/L_B \ll 1$, which was used from the very beginning to define $h_{\mu\nu}$, we can see $h_{\mu\nu}$ as a gauge field, with a gauge-invariant field-strength tensor given by the leading terms of $R_{\mu\nu\rho\sigma}^{(1)}$, as obtained from eq. (1.218).

6.0.23 Further reading

- Classical textbooks on general relativity are Weinberg (1972), Misner, Thorne and Wheeler (1973), and Landau and Lifshitz, Vol. II (1979). Among the more recent books, we suggest Hartle (2003) (at a rather introductory level, and with a very physical approach), and Straumann (2004) (more advanced).
- For discussions of freely falling frames, Riemann and Fermi normal coordinates, TT frame and the proper detector frame, see Misner, Thorne and Wheeler (1973), Sections 13.6 and 37.2, Hartle (2003), Section 8.4, and Thorne (1983, 1987). The metric of an accelerated, rotating observer is computed to quadratic order in x^i by Ni and Zimmermann (1978).
- The energy-momentum tensor of GWs and the short-wave expansion are discussed by Isaacson (1968a, 1968b), Misner, Thorne and Wheeler (1973), and Thorne (1987). The fact that performing a space-time average one obtains a gaugeinvariant energy-momentum tensor was already discussed in Arnowitt, Deser and Misner (1961).
- The geometric optics approximation in curved space-time is discussed in Isaacson (1968a, 1968b) and Misner, Thorne and Wheeler (1973) (see in particular Section 22.5 for photons, and Section 35.14 and Exercise 35.15 for gravitational waves). See also Thorne (1983, 1987). Diffraction and lensing of GWs is discussed in Bontz and Haugan (1980) and in Section 2.6.1 of Thorne (1983).
- A definition of GWs based on the asymptotics of the gravitational field at null infinity was given by Bondi, van der Burg and Metzner (1962) and Sachs (1962), and was also important historically for resolving the controversy on the reality of gravitational radiation. A geometric description of the asymptotic fall-off of radiative solutions, using the notion of asymptotically simple space-time, was given by Penrose (1963, 1965).
- The development of the concept of gravitational wave has a very interesting history. The notion even predates Einstein general relativity (the term "gravitational wave" was used by Poincaré as early as 1905, referring to the fact that, even in a gravitational theory, the interaction must propagate at a finite speed). With the advent of general relativity, gravitational waves were introduced by Einstein in 1916. However, the existence of physical effects associated with them has been questioned many times, with Einstein himself changing his mind more than once. Eddington is associated to the ironic remark that "gravitational waves propagate at the speed of thought", implying that they are gauge artifact. (Actually, he was referring only to the transverse-longitudinal and longitudinal-longitudinal components of h_{ij} , which, indeed, are pure gauge modes. Concerning the transverse-transverse part, he rather showed that it carries energy, and even corrected an

erroneous factor of two in Einstein's early version of the quadrupole formula.) The controversy on the existence of GWs was not settled until the early 1960s. A very interesting book on the history of the research in GWs is Kennefick (2007), see also the review article Kennefick (1997).

6.0.24 2

2.1 Linearized gravity and classical field theory

6.0.25 The field-theoretical approach to GWs

In the previous chapter we investigated GWs using the geometric interpretation which is at the core of general relativity. This geometrical perspective emphasizes that $g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}$ is the metric of spacetime, and therefore an incoming GW $h_{\mu\nu}$ induces perturbations in the space-time geometry. In this approach, the interaction of GWs with test masses is described by geometric tools such as the equation of the geodesic deviation, and the energy-momentum tensor of GWs is determined by examining how $h_{\mu\nu}$ contributes to the curvature of the background space-time.

General relativity can also be seen as a classical field theory, to which we can apply all standard field-theoretical methods. In this chapter we therefore go back to linearized gravity, writing $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, and we treat it as a classical field theory of the field $h_{\mu\nu}$ living in flat space-time with Minkowski metric $\eta_{\mu\nu}$. In this approach we are actually forgetting that $h_{\mu\nu}$ has an interpretation in terms of a space-time metric, and instead we treat it as any other field living in flat Minkowski space.

The fact that the beautiful geometric interpretation of $h_{\mu\nu}$ is hidden is compensated by the fact that we can put the conceptual issues discussed in Chapter 1 into the broader theoretical framework of classical field theory, and compare it to what happens in other field theories, such as classical electrodynamics. The geometric and the field-theoretical perspectives are indeed complementary; some aspects of GW physics can be better understood from the former perspective, some from the latter, and to study GWs from both vantage points results in a deeper overall understanding.

We will begin in Section 2.1.1 by recalling a basic tools of classical field theory, the Noether theorem, and we will see how we can reobtain the results of Chapter 1 using this standard field-theoretical tool. Besides providing a complementary understanding of various conceptual issues, the Noether theorem is also a very practical tool for explicit computations, and in particular we will see that it also provides the simplest way of deriving another important result, namely the general expression for the angular momentum carried by GWs.

In Section 2.2 we will pursue this field-theoretical approach further, discussing linearized gravity from the point of view of quantum field theory, and we will see how the notion of the graviton emerges. Actually, all astrophysical mechanisms of GW production, as well as the interaction of GWs with a detector, are fully accounted for by classical general relativity. In actual calculations, the notion of the graviton will not surface until we examine some cosmological production mechanisms (in particular, mechanisms related to the amplification of vacuum fluctuations), in Vol. 2. Nevertheless, we will see that also at the quantum level, the field-theoretical approach is illuminating for many conceptual aspects.

We will conclude this chapter with a more advanced section, which investigates whether gravitons can have a small mass. This, from the field-theoretical point of view, seems to be one of the most natural generalizations of Einstein gravity. We will see however that a field theory describing massive gravitons can have problems of internal consistency which, to date, are not yet fully understood.

6.0.26 2.1 Linearized gravity as a classical field theory

6.0.27 2.1.1 Noether's theorem

We begin by recalling some basic facts of classical field theory and in particular Noether's theorem.¹ We consider a field theory living in flat space-time, with fields ϕ_i , labeled by an index i . The action S is the integral of the Lagrangian density $\mathcal{L}$,

$$S = \int dt d^3x \mathcal{L}(\phi_i, \partial\phi_i) \quad (2.1)$$

In our case the fields ϕ_i will be the independent components of the metric $h_{\mu\nu}$, but it is useful to be more general, since we will also be interested in comparing with for instance classical electrodynamics, or any other classical field theory. We will denote collectively the fields ϕ_i simply by ϕ .

A transformation of the coordinates and of the fields is an operation that transforms the coordinates x^μ into new coordinates x'^μ , i.e. $x^\mu \rightarrow x'^\mu$, while at the same time the fields, denoted collectively by $\phi(x)$, are transformed into new functions of the new coordinates, $\phi(x) \rightarrow \phi'(x')$. To define the transformation means to state how x' is related to x and how $\phi'(x')$ is related to $\phi(x)$. For an infinitesimal transformation, we can write

$$x^\mu \rightarrow x'^\mu = x^\mu + \epsilon^a A_a^\mu(x) \quad (2.2)$$

$$\phi_i(x) \rightarrow \phi'_i(x') = \phi_i(x) + \epsilon^a F_{i,a}(\phi, \partial\phi) \quad (2.3)$$

and the transformation is specified assigning $A_a^\mu(x)$ and $F_{i,a}(\phi, \partial\phi)$. In general, the function $F_{i,a}$ depends on the collection of fields ϕ and on their derivatives, and will also mix different fields, so the transformation of a single field ϕ_i can depend also on all other fields ϕ_j with $j \neq i$. The above transformation is parametrized by a set of infinitesimal parameters ϵ^a , with $a = 1, \dots, N$.

Equations (2.2) and (2.3) define a symmetry transformation if they leave the action $S(\phi)$ invariant, for any ϕ . A symmetry transformation is called global if it leaves the action invariant when the parameters ϵ^a are constant, and local if it leaves invariant the action even when ϵ^a are allowed to be arbitrary functions of x .

Noether's theorem states that, for each generator of a global symmetry transformation, that is, for each of the parameters ϵ^a with $a = 1, \dots, N$, there is a current j_a^μ (which is a functional of the fields ϕ) which, when evaluated on a classical solution of the equations of motion ϕ^{cl} , is conserved, i.e. satisfies

$$(\partial_\mu j_a^\mu)|_{\phi=\phi^{\text{cl}}} = 0 \quad (2.4)$$

The corresponding charges Q_a are defined as the spatial integral of the $\mu = 0$ component of the currents

$$Q_a \equiv \int d^3x j_a^0(\mathbf{x}, t) \quad (2.5)$$

Current conservation (in the sense of eq. (2.4)) implies that Q_a is conserved (in the sense that it is time-independent). In fact

¹ We refer the reader to the textbook Maggiore (2005), Section 3.2, for further details and proofs. Note that in the present book, in order to follow the most common convention in general relativity, we are using the signature $\eta_{\mu\nu} = (-, +, +, +)$, while in Maggiore (2005) we use $\eta_{\mu\nu} = (+, -, -, -)$, which is the most common convention in quantum field theory. This is the origin of some sign differences between that textbook and the following equations.

$$\begin{aligned}
\partial_0 Q_a &= \int d^3x \partial_0 j_a^0(\mathbf{x}, t) \\
&= - \int d^3x \partial_i j_a^i(\mathbf{x}, t)
\end{aligned} \tag{2.6}$$

This is the integral of a total divergence, and it vanishes if we assume a sufficiently fast decrease of the fields at infinity. More generally, in a finite volume the variation of the charge is given by a boundary term representing the incoming or outgoing flux.

The explicit form of the current can be written in full generality in terms of the Lagrangian density $\mathcal{L}$ of the theory and of the functions $A_a^\mu(x)$ and $F_{i,a}(\phi, \partial\phi)$ that define the symmetry transformation (2.2), (2.3), and is given by

$$j_a^\mu = \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi_i)} [A_a^\nu(x) \partial_\nu \phi_i - F_{i,a}(\phi, \partial\phi)] - A_a^\mu(x) \mathcal{L} \tag{2.7}$$

The simplest application of this very general machinery is to the symmetry under space-time translations, and leads us to energy-momentum tensor. Under space-time translations we have $x^\mu \rightarrow x'^\mu = x^\mu + \epsilon^\mu$, and by definition all fields transform as $\phi'_i(x') = \phi_i(x)$, independently of their properties under Lorentz transformations; that is, a point P has the coordinate x in a frame and the coordinate x' in the translated frame, but the functional form of the fields changes so that the numerical values of the fields at the point P is invariant. So we write the transformation as

$$\begin{aligned}
x^\mu &\rightarrow x'^\mu = x^\mu + \epsilon^\mu \\
&= x^\mu + \epsilon^\nu \delta_\nu^\mu
\end{aligned} \tag{2.8}$$

$$\phi_i(x) \rightarrow \phi'_i(x') = \phi_i(x) \tag{2.9}$$

Observe that the index a appearing in ϵ^a in this case is a Lorentz index. Comparing with eqs. (2.2) and (2.3), we see that $A_\nu^\mu = \delta_\nu^\mu$ and $F_{i,a} = 0$.

The four conserved currents $\theta^\mu{}_\nu \equiv -j_{(\nu)}^\mu$ form a Lorentz tensor, known as the energy-momentum tensor. Using eq. (2.7) and raising the ν index, $\theta^{\mu\nu} = \eta^{\nu\rho} \theta^\mu{}_\rho$ we get ²

$$\theta^{\mu\nu} = - \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi_i)} \partial^\nu \phi_i + \eta^{\mu\nu} \mathcal{L} \tag{2.10}$$

Equation (2.4) becomes

$$\partial_\mu \theta^{\mu\nu} = 0 \tag{2.11}$$

on the solutions of the classical equations of motion. The conserved charge associated to space-time translations is, by definition, the fourmomentum P^ν , and therefore ³

$$cP^0 \equiv \int d^3x \theta^{00} \tag{2.12}$$

$$cP^i \equiv \int d^3x \theta^{0i} \tag{2.13}$$

This is the definition of four-momentum in classical field theory. A field configuration, solution of the equations of motion, carries an energy $E = cP^0$ and a spatial momentum P^i which can be calculated using eqs. (2.10), (2.12) and (2.13).

Observe that in general relativity the energy-momentum tensor of matter, defined by eq. (1.2), is automatically symmetric in the two indices μ, ν , since it is obtained by taking the functional derivative of the action with respect to the symmetric tensor $g_{\mu\nu}$. In contrast, the energy-momentum tensor defined from Noether's theorem, eq. (2.10), is not necessarily symmetric in the two indices μ, ν .

In fact, it is important to understand that the formal machinery of the Noether theorem, without some further physical input, is unable to uniquely fix the energy-momentum tensor, and more generally the Noether currents. For instance, consider what happens if we add a total four-divergence to the Lagrangian density,

$$\mathcal{L}' = \mathcal{L} + \partial_\mu K^\mu(\phi) \quad (2.14)$$

A total divergence, when integrated over d^4x , gives a boundary term. The equations of motion of the classical theory are obtained from a variation of the action, holding fixed the value of the fields on the boundaries. Therefore the equations of motion obtained from the variation of the action $S' = \int d^4x \mathcal{L}'$ are the same as those obtained from the variation of $S = \int d^4x \mathcal{L}$, so these two Lagrangians define the same classical field theory. However, the currents obtained from eq. (2.7) using $\mathcal{L}'$ or using $\mathcal{L}$ are in general different, and their difference is such that j^0 changes by a total spatial divergence, so that the charge in eq. (2.5) changes by a boundary term. Therefore the Noether currents are not uniquely defined; however, the Noether charges, computed integrating over a spatial volume V , are well defined if, and only if, the fields inside V go to zero sufficiently fast when we approach the boundaries of V , so that ² In order to minimize the number of factors of c in the equations, we have defined the flat-space Lagrangian density $\mathcal{L}$ from

$$S = \int dt d^3x \mathcal{L}$$

rather than from $S = \int d^4x \mathcal{L}$, as done, instead, in Landau and Lifshitz, Vol. II (1979). Recalling that, dimensionally, an action is (energy) $\times$ (time), in this way $\mathcal{L}$ has the same dimensions as $\theta^{\mu\nu}$, i.e. energy/volume. With our notation, for instance, the Lagrangian density of the electromagnetic field is $(-1/4)F_{\mu\nu}F^{\mu\nu}$, rather than $(-1/4c)F_{\mu\nu}F^{\mu\nu}$.

³ The factor c provides the correct dimensions, recalling that $P^0 = E/c$, see the Notation. all boundary terms can be neglected, and the ambiguity in the currents becomes irrelevant.

To illustrate this point, before computing the energy-momentum tensor of GWs it is instructive to recall what happens in the more familiar case of classical electrodynamics. In this case the Lagrangian density is

$$\mathcal{L}_{\text{em}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (2.15)$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. With our signature $\eta_{\mu\nu} = (-, +, +, +)$, the relation of $F_{\mu\nu}$ to the electric and magnetic fields is $F^{0i} = E^i$ and $F^{ij} = \epsilon^{ijk}B^k$. Equation (2.10) gives

$$\theta_{\text{em}}^{\mu\nu} = -\frac{\partial \mathcal{L}_{\text{em}}}{\partial (\partial_\mu A_\rho)} \partial^\nu A_\rho + \eta^{\mu\nu} \mathcal{L}_{\text{em}} \quad (2.16)$$

The functional derivative is easily computed,

$$\begin{aligned}
\frac{\partial}{\partial (\partial_\mu A_\rho)} \left(-\frac{1}{4} F_{\alpha\beta} F^{\alpha\beta} \right) &= -\frac{1}{2} F^{\alpha\beta} \frac{\partial F_{\alpha\beta}}{\partial (\partial_\mu A_\rho)} \\
&= -F^{\alpha\beta} \frac{\partial (\partial_\alpha A_\beta)}{\partial (\partial_\mu A_\rho)} \\
&= -F^{\mu\rho}
\end{aligned} \tag{2.17}$$

Therefore

$$\theta_{\text{em}}^{\mu\nu} = F^{\mu\rho} \partial^\nu A_\rho - \frac{1}{4} \eta^{\mu\nu} F^2 \tag{2.18}$$

At first sight, this result is surprising. Recall in fact that classical electrodynamics is invariant under gauge transformations,

$$A_\mu \rightarrow A_\mu - \partial_\mu \theta \tag{2.19}$$

as we can see observing that under eq. (2.19) $F_{\mu\nu}$ is invariant, and therefore also the Lagrangian (2.15) is invariant. However, the energymomentum tensor (2.18) depends on A_μ not only through the gaugeinvariant combination $F_{\mu\nu}$, but also through the term $\partial^\nu A_\rho$, which is not invariant, and under gauge transformations $\theta_{\text{em}}^{\mu\nu}$ changes as

$$\theta_{\text{em}}^{\mu\nu} \rightarrow \theta_{\text{em}}^{\mu\nu} - F^{\mu\rho} \partial^\nu \partial_\rho \theta \tag{2.20}$$

Apparently, we seem to be driven to the conclusion that the energy density θ^{00} of the electromagnetic field (as well as the momentum density θ^{0i}) is not gauge invariant. To deal with this problem, one first of all rewrites the energy-momentum tensor as follows,

$$\begin{aligned}
\theta_{\text{em}}^{\mu\nu} &= F^{\mu\rho} (\partial^\nu A_\rho - \partial_\rho A^\nu + \partial_\rho A^\nu) - \frac{1}{4} \eta^{\mu\nu} F^2 \\
&= \left(F^{\mu\rho} F_\rho^\nu - \frac{1}{4} \eta^{\mu\nu} F^2 \right) + F^{\mu\rho} \partial_\rho A^\nu \\
&= \left(F^{\mu\rho} F_\rho^\nu - \frac{1}{4} \eta^{\mu\nu} F^2 \right) + \partial_\rho (F^{\mu\rho} A^\nu)
\end{aligned} \tag{2.21}$$

(In the last line we used the equation of motion $\partial_\rho F^{\mu\rho} = 0$.) Therefore we have

$$\theta_{\text{em}}^{\mu\nu} = T_{\text{em}}^{\mu\nu} + \partial_\rho C^{\rho\mu\nu} \tag{2.22}$$

where

$$T_{\text{em}}^{\mu\nu} = F^{\mu\rho} F_\rho^\nu - \frac{1}{4} \eta^{\mu\nu} F^2 \tag{2.23}$$

while $C^{\rho\mu\nu} = F^{\mu\rho} A^\nu$ is a tensor antisymmetric in the indices (ρ, μ) . Now, $T_{\text{em}}^{\mu\nu}$ (which is sometimes called the "improved" energy-momentum tensor) is a gauge-invariant quantity, and its 00 component gives the usual form of the energy density,⁴

$$T_{\text{em}}^{00}(x) = \frac{1}{2} (\mathbf{E}^2 + \mathbf{B}^2)(x) \tag{2.24}$$

The term $\partial_\rho C^{\rho\mu\nu}$ instead is not gauge invariant, and we would like to get rid of it. The argument that is used to discard it is based on the following observations:

- If $\theta_{\text{em}}^{\mu\nu}$ is conserved, also $T_{\text{em}}^{\mu\nu}$ is conserved: in fact

$$\partial_\mu \partial_\rho C^{\rho\mu\nu} = 0 \quad (2.25)$$

automatically, whenever $C^{\rho\mu\nu}$ is antisymmetric under $\rho \leftrightarrow \mu$, as is our case.

- The difference between the charge cP^ν computed with $\theta_{\text{em}}^{\mu\nu}$ and that computed with $T_{\text{em}}^{\mu\nu}$ is given by

$$\int_V d^3x \partial_\rho C^{\rho 0\nu} = \int_V d^3x \partial_i C^{i 0\nu} \quad (2.26)$$

where we used the fact that $C^{00\nu} = 0$, which follows again from the antisymmetry of $C^{\rho\mu\nu}$ under $\rho \leftrightarrow \mu$. This is the integral of a divergence, and vanishes if the fields go to zero sufficiently fast at the boundaries of the volume V . Therefore the four-momentum P^ν computed with $\theta_{\text{em}}^{\mu\nu}$ is equal to that computed with $T_{\text{em}}^{\mu\nu}$, and is gauge invariant.

What we learn from this example is that the expression for the energymomentum tensor derived from the Noether theorem, eq. (2.10), is not necessarily a physical observable (in the case of electromagnetism it is not even gauge-invariant!).⁵ Rather, it is just a mathematical expression that, when integrated over space, gives unambiguously the total energy and momentum of a classical field configuration, as long as this field configuration goes to zero sufficiently fast on the boundaries of the integration region.

Equivalently, instead of speaking of the total energy of a localized object, we can divide by the volume and say that the expression in eq. (2.10) is a quantity that can be used to compute the average value of the energy-momentum tensor over a region of space sufficiently large, so that all boundary terms vanish, and any ambiguity related to total divergences disappears. Then, for instance, from eq. (2.22) we have

$$\langle \theta_{\text{em}}^{00} \rangle = \langle T_{\text{em}}^{00} \rangle + \langle \partial_i C^{i 00} \rangle \quad (2.27)$$

⁵ Another way of understanding the existence of such an ambiguity in a gauge theory is the fact that, in principle, one can allow that the gauge field, under space-time translations, does not go simply into itself, as in eq. (2.9), but into a configuration related by an arbitrary gauge transformation.

where the bracket represents the average. On a volume such that boundary terms give zero, we have $\langle \partial_i C^{i 00} \rangle = 0$ and therefore the average is unambiguously defined,

$$\begin{aligned} \langle \theta_{\text{em}}^{00} \rangle &= \langle T_{\text{em}}^{00} \rangle \\ &= \frac{1}{2} \langle \mathbf{E}^2 + \mathbf{B}^2 \rangle \end{aligned} \quad (2.28)$$

Whether one of the many equivalent integrands, in the expression for cP^0 , can be promoted to a physical observable, thereby providing a definition of a local energy density, is a physical question that cannot be answered using only the mathematics of the Noether theorem, without any additional physical input. We will discuss this issue at the end of the next section, and we will see that in fact the answer is in general different in electromagnetism and in general relativity.

⁴ The use of the Lagrangian (2.15) implies that we are using HeavisideLorentz units (also called rationalized c.g.s. unit) for the electric charge; in unrationalized units the factor $(-1/4)$ in eq. (2.15) becomes $(-1/16\pi)$, and the factor $1/2$ in eq. (2.24) becomes $1/(8\pi)$

6.0.28 2.1.2 The energy-momentum tensor of GWs

Now we can return to our original problem, which was the computation of the energy carried by GWs. We consider a wave-packet with reduced wavelengths centered around a value λ . In this case, according to the discussion above, the Noether theorem can give us an unambiguous answer for the energy density of the wave-packet, averaged over a box centered on the peak of the wave-packet, and with size $L \gg \lambda$. In this case the field is negligible on the boundaries and, using eq. (2.10),

$$t^{\mu\nu} = \left\langle -\frac{\partial \mathcal{L}}{\partial (\partial_\mu h_{\alpha\beta})} \partial^\nu h_{\alpha\beta} + \eta^{\mu\nu} \mathcal{L} \right\rangle \quad (2.29)$$

where $\langle \dots \rangle$ is a spatial average over several reduced wavelength (which, for plane waves, is the same as a temporal average over several periods), and $\mathcal{L}$ is the Lagrangian that governs the dynamics of $h_{\mu\nu}$.

As discussed in the previous subsection, the Noether theorem instead gives us an ambiguous answer if we ask what is the local energy and momentum density. Actually, we already saw in Section 1.4 that the final form of the energy-momentum tensor of GWs is indeed expressed as an average over several reduced wavelengths (or over several periods), and that this comes from a very fundamental reason, i.e. in order to discuss the back-reaction of GWs on the background, we need to perform a coarse-graining of the Einstein equations. Thus, we already know that it will not be possible to do better than this, and we cannot define a local expression for the energy and momentum density. Nevertheless, it is interesting to understand the reason also from a purely field-theoretical point of view; this will be discussed at the end of this section. First, we compute $t^{\mu\nu}$ from eq. (2.29), and we check that it agrees with the result that we derived in Section 1.4.

In order to use eq. (2.29) (and also to derive the angular momentum of GWs from Noether's theorem), we need the Lagrangian $\mathcal{L}$ or, equivalently, the action governing the dynamics of the field $h_{\mu\nu}$. To reproduce the Einstein equations to linear order in $h_{\mu\nu}$ we must expand the Einstein action to quadratic order in $h_{\mu\nu}$, while the linear term in the action vanishes, as always when we expand around a classical solution (in this case around the flat metric $\eta_{\mu\nu}$, since we consider Einstein equations in vacuum). We therefore start from the full Einstein action,

$$S_E = \frac{c^3}{16\pi G} \int d^4x \sqrt{-g} R \quad (2.30)$$

and we expand $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$. We observe that

$$R = g^{\mu\nu} R_{\mu\nu} = (\eta^{\mu\nu} - h^{\mu\nu} + O(h^2)) (R_{\mu\nu}^{(1)} + R_{\mu\nu}^{(2)} + O(h^3)) \quad (2.31)$$

where $R_{\mu\nu}^{(1)}$ is linear in h and $R_{\mu\nu}^{(2)}$ is quadratic in h . $R_{\mu\nu}^{(1)}$ and $R_{\mu\nu}^{(2)}$ can be obtained by specializing eqs. (1.113) and (1.114) to a flat background metric $\eta_{\mu\nu}$. The expansion of $\sqrt{-g}$ can be computed by writing $g_{\nu}^{\mu} = \delta_{\nu}^{\mu} + h_{\nu}^{\mu} \equiv (I + H)_{\nu}^{\mu}$, where I is the identity matrix and H a matrix whose elements are h_{ν}^{μ} . Since $g_{\mu\nu} = \eta_{\mu\rho} g_{\nu}^{\rho}$ and $\det \eta_{\mu\rho} = -1$, we have $-g = \det(I + H)$. Using the identity $\log(\det A) = \text{Tr}(\log A)$, valid for any non-degenerate matrix A , and expanding the logarithm,

$$\begin{aligned}
\det(I + H) &= \exp\{\log \det(I + H)\} \\
&= \exp\{\text{Tr} \log(I + H)\} \\
&= \exp\{\text{Tr}[H + O(H^2)]\} \\
&= 1 + \text{Tr} H + O(H^2) \\
&= 1 + h + O(h^2_{\mu\nu})
\end{aligned} \tag{2.32}$$

where $h = h^\mu_\mu = \eta^{\mu\nu} h_{\mu\nu}$ is the trace of $h_{\mu\nu}$. Since we are expanding over flat space, the lowest non-zero term in R is already $O(h)$, see eq. (2.31), so the terms $O(h^2)$ in $\sqrt{-g}$ give a contribution to the action $O(h^3)$, and can be neglected. Performing straightforward algebra we then obtain, after some integration by parts,

$$\begin{aligned}
S_E = -\frac{c^3}{64\pi G} \int d^4x & [\partial_\mu h_{\alpha\beta} \partial^\mu h^{\alpha\beta} - \partial_\mu h \partial^\mu h \\
& + 2\partial_\mu h^{\mu\nu} \partial_\nu h - 2\partial_\mu h^{\mu\nu} \partial_\rho h^\rho_\nu]
\end{aligned} \tag{2.33}$$

and the corresponding Lagrangian density is (see Note 2)

$$\mathcal{L} = -\frac{c^4}{64\pi G} [\partial_\mu h_{\alpha\beta} \partial^\mu h^{\alpha\beta} - \partial_\mu h \partial^\mu h + 2\partial_\mu h^{\mu\nu} \partial_\nu h - 2\partial_\mu h^{\mu\nu} \partial_\rho h^\rho_\nu] \tag{2.34}$$

We can now compute $t^{\mu\nu}$, using eq. (2.29). We evaluate the result directly in the gauge

$$\partial^\mu h_{\mu\nu} = 0, \quad h = 0 \tag{2.35}$$

Therefore, after computing the derivative $\delta\mathcal{L}/(\partial_\mu h_{\alpha\beta})$ in eq. (2.29), we impose the gauge condition (2.35). We observe that the second, third and fourth terms in brackets in eq. (2.33) are quadratic in quantities that will be set to zero by the gauge fixing and, after taking the functional derivative, they give contributions which are linear in h or in $\partial^\mu h_{\mu\nu}$, and therefore vanish when we impose eq. (2.35). So the only non-vanishing contribution comes from the term $\partial_\mu h_{\alpha\beta} \partial^\mu h^{\alpha\beta}$, and we get

$$\left. \frac{\partial\mathcal{L}}{\partial(\partial_\mu h_{\alpha\beta})} \right|_{\partial^\mu h_{\mu\nu}=h=0} = -\frac{c^4}{32\pi G} \partial^\mu h^{\alpha\beta} \tag{2.36}$$

We next evaluate the term $\langle\mathcal{L}\rangle$ in eq. (2.29), in our gauge. We recall that inside the average we are free to perform integration by parts (compare with Note 23 on page 35). Then, since in our gauge $h_{\mu\nu}$ satisfies the equations of motion $\square h_{\mu\nu} = 0$, even the term $\partial_\mu h_{\alpha\beta} \partial^{\mu\nu} h^{\alpha\beta}$ gives zero because, after an integration by parts, it becomes $-h_{\alpha\beta} \square h^{\alpha\beta}$ and vanishes upon use of the equations of motion, so $\langle\mathcal{L}\rangle = 0$.

In conclusion, we obtain

$$t^{\mu\nu} = \frac{c^4}{32\pi G} \langle \partial^\mu h^{\alpha\beta} \partial^\nu h_{\alpha\beta} \rangle \tag{2.37}$$

As expected, the result agrees with eq. (1.133).

We can now come back to the problem of the localization of the energy of GWs. In the field-theoretical description, the issue is whether one of the many equivalent integrands, in the expression for P^0 , can be promoted to a physical observable, thereby providing a definition of a local energy density. In electromagnetism there is a very natural candidate, the tensor $T^{\mu\nu}_{\text{em}}$ of eq. (2.23), which is conserved and gauge invariant, so it can be sensible to identify

$(1/2)(\mathbf{E}^2 + \mathbf{B}^2)$ with the local energy density of the electromagnetic field. No such answer is possible for the gravitational field. The quantity $\partial^\mu h^{\alpha\beta} \partial^\nu h_{\alpha\beta}$ which appears inside the average in eq. (2.37) is not gauge-invariant. It is futile to search for another local expression, whose integral gives the energy, but which is already gauge invariant before integration: the equivalence principle tells us that, at a given point, we can always find a locally inertial frame (see Section 1.3.2), such that at the point in question the gravitational field vanishes. Therefore any candidate expression for a local energy density can always be set to zero at a given point with a coordinate transformation, so it cannot be gauge invariant. This is an important difference between gravity and electromagnetism. In electromagnetism, at least if we consider a slowly varying electromagnetic field, it makes sense to assign to each point a local energy density $(1/2)(\mathbf{E}^2 + \mathbf{B}^2)$.

When we consider waves, however, concerning energy localization there is no real difference between gravity and electromagnetism. In both cases, all that can be really measured is the energy averaged over several wavelengths or periods. This can be understood even more easily at first by looking at the problem quantum-mechanically even if, as we will see below, the argument is really classical.

From the quantum point of view, a plane wave describes a collection of massless quanta (gravitons with helicity $h = \pm 2$ for gravitation, and photons with helicity ± 1 in electrodynamics). Consider a collection of such free particles. To determine the energy of this system in a volume V we must know how many quanta of the field are within V at a given time, and the energy of each. If we take a volume with sides smaller than the reduced wavelength λ of a photon or of a graviton, in order to know whether a given photon (or graviton) is inside the box we must measure its position with an error $\Delta x < \lambda$ and then, by the uncertainty principle, we have $\Delta p > \hbar/\lambda$, which is larger than the momentum $p = \hbar/\lambda$ of a quantum with reduced wavelength λ . Thus, we have completely lost information about the momentum $\mathbf{p}$ of the particle, and therefore (given that we are considering free particles described by simple plane waves), we lost information about its energy $E = c|\mathbf{p}|$, which means that we cannot localize the energy density better than to a few wavelengths. Alternatively, we can localize the energy in space if we delocalize it in time, according to $\Delta E \Delta t \gtrsim \hbar$. Clearly photons and gravitons do not show any difference in this respect.⁶ In fact, even if we phrased it in a quantum language, this argument is really classical, and follows from the fact that the position x of the peak of a classical wave-packet, and its typical wave-vector k satisfy $\Delta x \Delta k \geq 1$, and in this form is simply a property of the Fourier transform. In quantum theory, one identifies the momentum $\mathbf{p}$ with $\hbar \mathbf{k}$, and $\Delta x \Delta k \geq 1$ becomes $\Delta x \Delta p \geq \hbar$.

In conclusion the field-theoretical approach, based on Noether's theorem, gives us an unambiguous recipe for computing a spatial (or a temporal) average of the energy-momentum tensor of GWs. The fact that, for GWs, we cannot do better than this, i.e. that the energy of GWs cannot be localized in space (or in time) with a precision better than a few wavelengths (or a few periods) can be understood as a consequence of the equivalence principle, but it is in fact also a general property of any wave governed by a massless wave equation, which at the quantum level translates into a limitation required by the uncertainty principle.

6.0.29 2.1.3 The angular momentum of GWs

We next compute the angular momentum carried by GWs. Angular momentum is the conserved charge associated to invariance under spatial rotations. A symmetric tensor $h_{\mu\nu}$, from the point of view of spatial rotations, decomposes into h_{00} and the spatial trace h^i_i , which are both scalars under rotation and therefore are spin-0 fields, h_{0i} which is a spatial vector and therefore has spin 1, and a traceless symmetric tensor h_{ij} , which is a spin-2 field and has five degrees of freedom. To describe the GW we go to the TT gauge, so we have $h_{0\mu} = 0$ and we are

left only with the field h_{ij}^{TT} , which satisfies $(h^{\text{TT}})_i^i = 0$ and $\partial^i h_{ij}^{\text{TT}} = 0$, compare with eq. (1.31). Observe that h_{ij}^{TT} has only two degrees of freedom, corresponding to a massless particle with helicity ± 2 .⁷ As before, the last three terms in eq. (2.33) give a contribution to the Noether current that vanishes when, after taking the functional derivatives, we impose the gauge fixing condition. Then, for the purpose of computing the Noether current in this gauge, we can use as Lagrangian the first term in eq. (2.33), and we can keep only the physical degrees of freedom h_{ij}^{TT} ,

$$\mathcal{L} = -\frac{c^4}{64\pi G} \partial_\mu h_{ij}^{\text{TT}} \partial^\mu h_{ij}^{\text{TT}} \quad (2.38)$$

A rotation of three-dimensional space is described by a 3×3 orthogonal matrix $\mathcal{R}$, which transforms the coordinates according to $x^i \rightarrow \mathcal{R}^{ij} x^j$ (observe that, since our signature is $(-, +, +, +)$, we do not need to be careful about raising or lowering spatial indices). An infinitesimal rotation can be written as

$$\mathcal{R}^{ij} = \delta^{ij} + \omega^{ij} \quad (2.39)$$

and the condition that $\mathcal{R}$ is an orthogonal matrix gives $\omega^{ij} = -\omega^{ji}$. Therefore rotations can be parametrized by the three independent quantities ω^{12}, ω^{13} and ω^{23} , which play the role of the parameters ϵ^a in eqs. (2.2) and (2.3). So, for rotations eqs. (2.2) and (2.3) become

$$x^i \rightarrow x^i + \sum_{k < l} \omega^{kl} A_{kl}^i \quad (2.40)$$

$$h_{ij}^{\text{TT}} \rightarrow h_{ij}^{\text{TT}} + \sum_{k < l} \omega^{kl} F_{ij,kl}, \quad (2.41)$$

and the sum goes over the independent parameters ω^{kl} with $k < l$. Observe that, in $F_{ij,kl}$, the first pair of indices, (i, j) , is the label that identifies the field h_{ij}^{TT} , and therefore plays the role that the index i had in eq. (2.3), while the second pair (k, l) plays the role of the index a in eq. (2.3). In particular, $F_{ij,kl}$ is symmetric with respect to the pair (i, j) and can be taken to be antisymmetric with respect to the pair (k, l) .

The total angular momentum carried by the GWs is then given by

$$J^i = \frac{1}{2} \epsilon^{ijk} J_{kl} \quad (2.42)$$

where J_{kl} is the conserved charge associated to rotations in the (k, l) plane,

$$J_{kl} = \frac{1}{c} \int d^3x j_{kl}^0 \quad (2.43)$$

⁶ The fact that, as a quantum field theory, general relativity is not renormalizable plays no role here. As long as we are at energies much smaller than the Planck mass, linearized theory can be promoted to a well defined effective quantum field theory, describing weakly interacting gravitons, as we discuss in more detail in Section 2.2.

⁷ As we will recall in Section 2.2.2 and in Problem 2.1, the physical representation of the Poincar group are of two types: massive representations, characterized by their spin j and having $2j + 1$ states, and massless representations, which have a quantum number j but only two states, corresponding to helicities $h = \pm j$. We will use the name "spin- j field" to denote generically a field that can describe either a massive particle with spin j or a massless particle with helicities $h = \pm j$ (in which case, it will be subject to constraints that eliminate the extra states, see Section 2.2). The name "spin- j particle" will however be reserved to a massive particle with spin j and therefore $2j + 1$ degrees of freedom. A massless particle with quantum number j will be referred to as "a massless particles with helicity $\pm j$ "

(the factor $1/c$ provides the correct dimensions) and Noether's theorem gives

$$j_{kl}^0 = \frac{\partial \mathcal{L}}{\partial (\partial_0 h_{ij}^{\text{TT}})} [A_{kl}^\nu \partial_\nu h_{ij}^{\text{TT}} - F_{ij,kl}] - A_{kl}^0 \mathcal{L} \quad (2.44)$$

To find the explicit form of A_{kl}^ν and $F_{ij,kl}$ we compare the generic formulas (2.40) and (2.41) with the actual transformation properties of x^i and h_{ij}^{TT} under infinitesimal rotations; A_{kl}^i is easily computed observing that

$$\begin{aligned} x^i &\rightarrow \mathcal{R}^{ij} x^j = x^i + \omega^{ij} x^j \\ &= x^i + \sum_{k < l} \omega^{kl} (\delta^{ik} x^l - \delta^{il} x^k) \end{aligned} \quad (2.45)$$

so, comparing with eq. (2.40),

$$A_{kl}^i = \delta^{ik} x^l - \delta^{il} x^k \quad (2.46)$$

As for A_{kl}^0 , since time is unchanged under spatial rotations, we have $A_{kl}^0 = 0$

The quantity $F_{ij,kl}$ is determined by the properties of h_{ij}^{TT} under rotations. Since h_{ij}^{TT} is a spatial tensor, it transforms as

$$\begin{aligned} h_{ij}^{\text{TT}} &\rightarrow R_i^k R_j^l h_{kl}^{\text{TT}} \\ &= h_{ij}^{\text{TT}} + \omega_j^l h_{il}^{\text{TT}} + \omega_i^k h_{kj}^{\text{TT}} \\ &= h_{ij}^{\text{TT}} + \sum_{k < l} \omega^{kl} (\delta_{ik} h_{jl}^{\text{TT}} - \delta_{il} h_{jk}^{\text{TT}} + \delta_{jk} h_{il}^{\text{TT}} - \delta_{jl} h_{ik}^{\text{TT}}) \end{aligned} \quad (2.47)$$

Comparing eqs. (2.41) and (2.47), we see that

$$F_{ij,kl} = \delta_{ik} h_{jl}^{\text{TT}} - \delta_{il} h_{jk}^{\text{TT}} + \delta_{jk} h_{il}^{\text{TT}} - \delta_{jl} h_{ik}^{\text{TT}} \quad (2.48)$$

We can now plug these results into eq. (2.44). We also observe that

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial (\partial_0 h_{ij}^{\text{TT}})} &= -\frac{c^4}{32\pi G} \partial^0 h_{ij}^{\text{TT}} \\ &= +\frac{c^3}{32\pi G} \dot{h}_{ij}^{\text{TT}} \end{aligned} \quad (2.49)$$

and (renaming the indices $(i, j) \rightarrow (a, b)$) we get

$$j_{kl}^0 = \frac{c^3}{32\pi G} \left[-\dot{h}_{ab}^{\text{TT}} (x^k \partial^l - x^l \partial^k) h_{ab}^{\text{TT}} + 2\dot{h}_{ab}^{\text{TT}} (\delta_{bl} h_{ak}^{\text{TT}} + \delta_{al} h_{bk}^{\text{TT}}) \right] \quad (2.50)$$

From eqs. (2.43), (2.44) and (2.42), the total angular momentum of the GW is therefore

$$J^i = \frac{c^2}{32\pi G} \int d^3x \left[-\epsilon^{ikl} \dot{h}_{ab}^{\text{TT}} x^k \partial^l h_{ab}^{\text{TT}} + 2\epsilon^{ikl} h_{ak}^{\text{TT}} \dot{h}_{al}^{\text{TT}} \right] \quad (2.51)$$

To understand the physical meaning of the two terms in bracket, it is useful to recall the analogous results for spin-0 and spin-1 fields. ⁸ For a real scalar field, applying Noether's theorem, one finds that the angular momentum carried by a field configuration ϕ is

$$J^i = -\epsilon^{ikl} \int d^3x (\partial_0 \phi) x^k \partial^l \phi \quad (2.52)$$

We see that this has the same structure as the first term in eq. (2.51) (after rescaling h_{ab}^{TT} by a factor $(32\pi G/c^3)^{1/2}$; as we will see in Section 2.2, this rescaling gives to h_{ab}^{TT} the standard field-theoretical normalization). The physical meaning of this term can be understood observing that, for a real field ϕ satisfying the massless Klein-Gordon equation $\square\phi = 0$,⁸ For explicit proofs, see Maggiore (2005), Section 3.3.1 for spin-0 fields, and Section 4.3.1 for the electromagnetic field. Pay attention to the fact that this reference uses the opposite metric signature (following the most common convention used in field theory, while here we are following the most common convention used in general relativity), and units $\hbar = c = 1$. one can define the scalar product between two field configurations ϕ and ϕ' ,

$$\langle \phi | \phi' \rangle = \frac{i}{2} \int d^3x \phi \overleftrightarrow{\partial}_0 \phi' \quad (2.53)$$

where, on any two functions f and g , we define $\overleftrightarrow{\partial}_\mu$ by $f \overleftrightarrow{\partial}_\mu g \equiv f \partial_\mu g - (\partial_\mu f) g$. Since ϕ and ϕ' are functions of t and $\mathbf{x}$, and in eq. (2.53) we integrate over d^3x , the result is a priori still a function of time. However, this scalar product is actually time-independent, if ϕ_1, ϕ_2 are solutions of the Klein-Gordon equation. It is then not so surprising that the conserved charges of the scalar field theory can be expressed as expectation values of suitable operators, with respect to this scalar product.

Consider in particular $\hat{L}^i = -i\epsilon^{ikl} x^k \partial^l$, which is the orbital angular momentum operator (in units of $\hbar$). The expectation value of this operator, with respect to the scalar product (2.53), is

$$\begin{aligned} \langle \phi | \hat{L}^i | \phi \rangle &= \frac{i}{2} \int d^3x \left[\phi \hat{L}^i \partial_0 \phi - (\partial_0 \phi) \hat{L}^i \phi \right] \\ &= \frac{1}{2} \epsilon^{ikl} \int d^3x \left[\phi x^k \partial^l \partial_0 \phi - (\partial_0 \phi) x^k \partial^l \phi \right] \\ &= -\epsilon^{ikl} \int d^3x (\partial_0 \phi) x^k \partial^l \phi \end{aligned} \quad (2.54)$$

where, going from the second to the third line, we integrated by parts ∂^l in the first term. Comparing with eq. (2.52) we see that the total angular momentum carried by a scalar field configuration ϕ is equal to the expectation value of the orbital angular momentum operator $\hat{L}^i$, with respect to this scalar product,

$$J^i = \langle \phi | \hat{L}^i | \phi \rangle \quad (2.55)$$

Observe also that, for a scalar field, there is no additional contribution, and the total angular momentum is given by the expectation value of the orbital angular momentum operator.

Consider now the spin-1 case. If one computes, using Noether's theorem, the angular momentum carried by the electromagnetic field in the radiation gauge, where the electromagnetic field is transverse, $\partial_i A^i = 0$, and satisfies $\square A^i = 0$, one finds

$$J^i = \int d^3x \left[-\epsilon^{ikl} (\partial_0 A_j) x^k \partial^l A_j + \epsilon^{ikl} A_k \partial_0 A_l \right] \quad (2.56)$$

Again, since the equation of motion is $\square A_i = 0$, we can define a scalar product

$$\langle A | A' \rangle = \frac{i}{2} \int d^3x A_i \overleftrightarrow{\partial}_0 A'_i \quad (2.57)$$

which is conserved on the solutions of the equation of motion. Then, we see that the first term in eq. (2.56) is again the expectation value of $\hat{L}^i = -i\epsilon^{ikl} x^k \partial^l$ with respect to this scalar

product, so it is the contribution from the orbital angular momentum. In the derivation based on

the Noether theorem, this term is determined by the transformation of the coordinates x^i under spatial rotations, (i.e. it comes from the term proportional to A_a^ν in eq. (2.7)), so it is clear that it has the same structure for all fields, independently of their spin. The second term instead depends on the specific properties of the field under spatial rotations (i.e. it comes from the term proportional to $F_{i,a}$ in eq. (2.7)) and therefore it is the spin part. Defining

$$S^i = \epsilon^{ikl} \int d^3x A_k \partial_0 A_l \quad (2.58)$$

the explicit computation⁹ shows that, in second quantization, the circular polarizations states of the photons are eigenvectors of the helicity operator $\mathbf{S} \cdot \hat{\mathbf{p}}$ (where $\hat{\mathbf{p}}$ is the unit vector in the direction of propagation), with eigenvalues ± 1 , as required for a massless particle described by a vector field.

We can now understand the meaning of the two terms in eq. (2.51). Since h_{ab}^{TT} satisfies $\square h_{ab}^{\text{TT}} = 0$, the conserved scalar product is defined as in the Klein-Gordon or in the electromagnetic case,

$$\langle h | h' \rangle = \frac{i}{2} \int d^3x h_{ab}^{\text{TT}} \overleftrightarrow{\partial}_0 h_{ab}^{\text{TT}} \quad (2.59)$$

After rescaling $h_{ab}^{\text{TT}} \rightarrow (32\pi G/c^3)^{1/2} h_{ab}^{\text{TT}}$ (which, as we will see in Section 2.2.2, is the rescaling needed to give to $h_{\mu\nu}$ the canonical normalization to the kinetic term in the action) the first term in eq. (2.51) is the expectation value of the orbital angular momentum operator with respect to this scalar product, while the term $2\epsilon^{ikl} h_{ak}^{\text{TT}} \dot{h}_{al}^{\text{TT}}$ is the spin contribution. The factor of 2 in this term correctly reproduces the fact that the gravitational field has spin two, and therefore gravitons are eigenvectors of the helicity with eigenvalues ± 2 .

In the previous section we learned that the Noether currents cannot be localized better than a few wavelengths, so the physical density of angular momentum, j^i/c , is the integrand of eq. (2.51), averaged over a few wavelengths,

$$\frac{1}{c} j^i = \frac{c^2}{32\pi G} \left\langle -\epsilon^{ikl} \dot{h}_{ab}^{\text{TT}} x^k \partial^l h_{ab}^{\text{TT}} + 2\epsilon^{ikl} \dot{h}_{al}^{\text{TT}} h_{ak}^{\text{TT}} \right\rangle \quad (2.60)$$

Consider now a GW propagating outward from a source. At time t we consider a portion of the wave front, at radial distance r from the source, and covering a solid angle $d\Omega$. At time $t + dt$ this portion of the wave front has swept the volume $d^3x = r^2 dr d\Omega = r^2 (cdt) d\Omega$. Since the angular momentum per unit volume is (j^i/c) , the angular momentum carried away by the GW is $dJ^i = r^2 (cdt) d\Omega (j^i/c)$. Therefore the rate of angular momentum emission due to GWs is

$$\frac{dJ^i}{dt} = \frac{c^3}{32\pi G} \int r^2 d\Omega \left\langle -\epsilon^{ikl} \dot{h}_{ab}^{\text{TT}} x^k \partial^l h_{ab}^{\text{TT}} + 2\epsilon^{ikl} \dot{h}_{al}^{\text{TT}} h_{ak}^{\text{TT}} \right\rangle \quad (2.61)$$

This section is intended for readers with some knowledge of quantum field theory. In this section and in the next we use units

$$\hbar = c = 1$$

It can be useful to recall how dimensional analysis works in these units. Since $c = 1$, velocities are pure numbers while, dimensionally,

energy = momentum = mass.

Recalling, from the uncertainty principle, that a length times a momentum has dimensions of $\hbar$, in units $\hbar = c = 1$ we also have

length = (momentum)⁻¹ = (mass)⁻¹.

Then the dimensions of any quantity can be expressed as positive or negative powers of mass. For example, and energy density is energy/volume = (mass)⁴.

6.0.30 2.2 Gravitons

All astrophysical processes which generate GWs, as well as the interaction of GWs with detectors, are adequately described in the framework of a classical field theory of gravitation. Nevertheless, it can be instructive to discuss some conceptual issues using the vantage point provided by modern quantum field theory.

In this section we will see that, at the quantum level, gravity must be mediated by a massless particle with helicity ± 2 , the graviton. Its free action, in flat space-time, is fixed by field-theoretical considerations, and reproduces the linearization of the Einstein action. Consistency with gauge-invariance then requires the introduction of non-linear couplings between the graviton, with three-graviton vertices, four-graviton vertices, etc., and a full non-linear structure emerges. Finally, we will briefly mention some of the issues involved in the quantization of gravity and we will discuss why, even in the absence of a full quantum theory of gravitation, general relativity makes sense as an effective field theory below the Planck scale, and the notion of graviton is well defined.

6.0.31 2.2.1 Why a spin-2 field?

In quantum field theory all interactions are mediated by the exchange of bosons, i.e. particles with integer spin. Since a flat space-time background is an excellent approximation in many situations, it make sense to look for a relativistic quantum field theory living in flat space-time, that in the non-relativistic limit reduces to Newtonian gravity. Such a theory should be mediated by a boson which propagates in this flat space-time.

Gravity is the interaction that in the non-relativistic limit couples to the mass. To obtain such a coupling from a local quantum field theory, we need to couple the field that mediates gravity to a local quantity (i.e. a quantity which is function of the space-time point x) whose spatial integral, in the non-relativistic limit, becomes the mass.

Mathematics alone is not sufficient to obtain uniquely a consistent theory that gives back Newton's law. We will also use as guiding principle the deep insight of Einstein that in the full theory of gravitation not only the mass is a source for gravitation but, more generally, all forms of energy. In a local field theory, we therefore look for a coupling between the field that mediates gravity and the energy density.

Since, from the point of view of Lorentz transformations, energy density is the (00) component of the energy-momentum tensor, consistency with special relativity requires that we couple our gravitational field to the energy-momentum tensor $T_{\mu\nu}(x)$. As long as we work at linearized level, we can neglect the contribution of the gravitational field itself to $T^{\mu\nu}$, so we can take as $T^{\mu\nu}$ the energy-momentum tensor of the matter fields only. Since we are assuming a flat background, at linearized level $T^{\mu\nu}$ obeys the flat-space conservation law

$$\partial_\mu T^{\mu\nu} = 0 \quad (2.62)$$

We will come back later to the inclusion of the non-linear terms.

The simplest possibility that one can examine is that gravity is mediated by a spin-0 boson, described by a scalar field ϕ . A scalar field ϕ carries no Lorentz index, so the only possibility is that it couples to the trace of the energy-momentum tensor, $T = T^\mu_\mu$. Therefore the Lagrangian density which describes the dynamics of this hypothetical scalar gravitational field and its coupling to matter, in linearized theory, must be of the form

$$\mathcal{L} = -\frac{1}{2} (\partial_\mu \phi \partial^\mu \phi + \mu^2 \phi^2) + g\phi T \quad (2.63)$$

where g is a coupling constant, and μ is the mass of the scalar field (more generally, we could also add a potential $V(\phi)$, that however does not influence the following discussion). In order to see if this theory reduces correctly to Newtonian gravity in the non-relativistic limit, we can compute the potential induced by the exchange of a ϕ boson between two static particles of masses m_1 and m_2 . Let us recall that in quantum field theory the static interaction potential $V(\mathbf{x})$ is a derived concept, which makes sense only in the non-relativistic limit. To obtain it, one must compute the $2 \rightarrow 2$ scattering amplitude M_{fi} at tree level, taking the initial state equal to the final state, and then the potential is given by ¹⁰

$$V(\mathbf{x}) = - \int \frac{d^3 q}{(2\pi)^3} M_{fi}(\mathbf{q}) e^{i\mathbf{q} \cdot \mathbf{x}} \quad (2.64)$$

The energy-momentum tensor $T^{\mu\nu}$ is quadratic in the matter fields, so the vertex $g\phi T$ involves two matter-field lines and one ϕ boson, and to compute M_{fi} we must evaluate a Feynman diagram of the type given in Fig. 2.1 (plus a possible exchange graph for identical particles), with $q = (q^0, \mathbf{q})$, setting $q^0 = 0$ because the potential is obtained in the static limit. In general, the details of the computation depend on the form of $T(x)$, i.e. on the specific type of matter field considered. However, if the matter field is massive, we can take the non-relativistic limit, and in this limit we can treat $T(x)$ as an external classical field. In this case, the interaction vertex of this theory becomes as shown in Fig. 2.2, where the cross denotes the insertion of $\hat{T}(\mathbf{q})$. Then the diagram of Fig. 2.1 reduces to that shown in Fig. 2.3, and the amplitude M_{fi} is given simply by

$$iM_{fi}(\mathbf{q}) = (-ig)^2 \tilde{T}_1(\mathbf{q}) \tilde{D}(\mathbf{q}) \tilde{T}_2(-\mathbf{q}) \quad (2.65)$$

where

$$D(q) = \frac{-i}{q^2 + \mu^2} \quad (2.66)$$

is the propagator of the scalar field ϕ (recall that we are using the signature $\eta_{\mu\nu} = (-, +, +, +)$), and T_1, T_2 are the traces of the energymomentum tensor for particles of masses m_1 and m_2 , respectively. In the static limit, $q^2 = -(q^0)^2 + \mathbf{q}^2 \rightarrow \mathbf{q}^2$, and the propagator becomes

$$D(\mathbf{q}) = \frac{-i}{\mathbf{q}^2 + \mu^2} \quad (2.67)$$

Spin-0

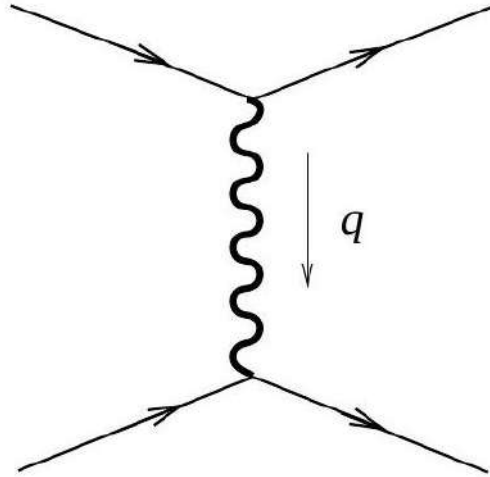


Fig. 2.1 The Feynman diagram that gives the scattering amplitude at tree level.

¹⁰ See e.g. Maggiore (2005) (Section 6.6 and eqs. (7.56) – (7.59)), for the derivation of this result and for explicit computations of the potential induced by the exchange of scalar or vector particles.

Fig. 2.2 The Feynman diagram that gives the vertex, when $T(x)$ is treated as an external field.

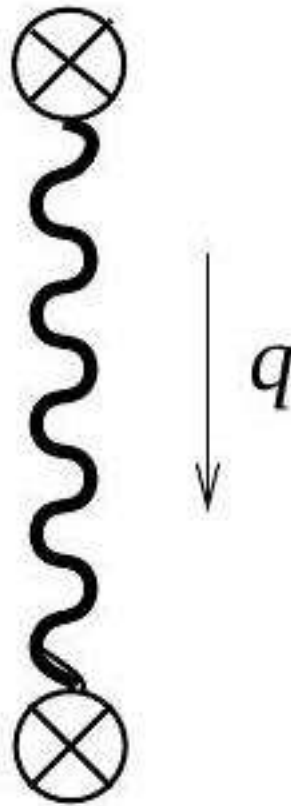


Fig. 2.3 The Feynman diagram that gives the scattering amplitude, when $T(x)$ is treated as an external field.

¹¹ More precisely, for a particle in $\mathbf{x}_1$, $T(\mathbf{x}) = -m\delta^{(3)}(\mathbf{x} - \mathbf{x}_1)$, so $\tilde{T}(\mathbf{q}) = -m \exp\{-i\mathbf{q} \cdot \mathbf{x}_1\}$. In the field theoretical language, the factors $\exp\{-i\mathbf{q} \cdot \mathbf{x}_1\}$ from $\tilde{T}_1(\mathbf{q})$ and $\exp\{+i\mathbf{q} \cdot \mathbf{x}_2\}$ from

$\tilde{T}_2(-\mathbf{q})$ comes from the wavefunctions of the external legs, and in eq. (2.68) they have already been taken into account by the term $\exp\{i\mathbf{q} \cdot (\mathbf{x}_2 - \mathbf{x}_1)\}$, where $\mathbf{x}_2 - \mathbf{x}_1 \equiv \mathbf{x}$.¹² Obviously, this rules out the possibility that the main contribution to gravity comes from spin-0 fields. It says nothing about the possibility that there are additional gravitationally interacting scalar fields, which give small corrections to general relativity either because they are massive or because their gravitational coupling is suppressed with respect to Newton constant G . Indeed, additional scalar fields enter in most extensions of general relativity.

Equation (2.64) then gives

$$V(\mathbf{x}) = -ig^2 \int \frac{d^3q}{(2\pi)^3} \tilde{T}_1(\mathbf{q}) \tilde{D}(\mathbf{q}) \tilde{T}_2(-\mathbf{q}) e^{i\mathbf{q} \cdot \mathbf{x}} \quad (2.68)$$

For a relativistic classical particle moving on the trajectory $\mathbf{x}_0(t)$, the energy-momentum tensor is given by (see, e.g. Landau and Lifshitz, Vol. II 1979)

$$T^{\mu\nu}(\mathbf{x}, t) = \frac{p^\mu p^\nu}{p^0} \delta^{(3)}(\mathbf{x} - \mathbf{x}_0(t)) \quad (2.69)$$

where p^μ is the four-momentum. Using $p^\mu p_\mu = -m^2$ and, for a particle at rest, $p^0 = m$, the trace of the energy-momentum tensor of a heavy source becomes $T(\mathbf{x}) = -m\delta^{(3)}(\mathbf{x})$, which in turn gives $\tilde{T}(\mathbf{q}) = -m$.¹¹ Therefore

$$V(\mathbf{x}) = -ig^2 m_1 m_2 D(\mathbf{x}) \quad (2.70)$$

If the mass μ vanishes,

$$\begin{aligned} D(\mathbf{x}) &= \int \frac{d^3q}{(2\pi)^3} \frac{-i}{\mathbf{q}^2} e^{i\mathbf{q} \cdot \mathbf{x}} \\ &= \frac{-i}{4\pi r} \end{aligned} \quad (2.71)$$

where $r = |\mathbf{x}|$, and therefore we get the correct Newtonian potential

$$V(r) = -\frac{Gm_1 m_2}{r} \quad (2.72)$$

once we make the identification $g^2/(4\pi) = G$. If instead the mass μ is non-vanishing, we get a Yukawa potential

$$V(\mathbf{x}) = -\frac{Gm_1 m_2}{r} e^{-\mu r} \quad (2.73)$$

This result shows that, as far as the non-relativistic Newtonian limit is concerned, a spin-0 massless scalar field is a viable possibility.

However, a spin-0 field fails when we come to the new predictions of this theory of gravitation in the relativistic regime. In particular, we see from eq. (2.23) that the energy-momentum tensor of the electromagnetic field is traceless,

$$T_{\text{em}} = F^{\mu\rho} F_{\mu\rho} - \frac{1}{4} \delta^\mu_\mu F^2 = 0 \quad (2.74)$$

Therefore, in this theory, photons do not couple to gravity. Experimentally, the gravitational bending of light rays from massive objects is very well established (it was in fact the first experimental confirmation of general relativity and nowadays, in the form of gravitational lensing, it is a beautiful routine tool in astrophysics). Therefore a spin-0 theory of gravitation is ruled out.¹²

The next possibility is a spin-1 field, just like in electrodynamics. In order to get a long-range potential, we need again a massless field, but a massless vector field A_μ can be coupled consistently only if we respect gauge invariance. In electrodynamics this can be obtained with a coupling $A_\mu j^\mu$, imposing that j^μ is a conserved current. In fact, under a gauge transformation, $A_\mu \rightarrow A_\mu - \partial_\mu \theta$ and, after an integration by parts, the term $-(\partial_\mu \theta) j^\mu \rightarrow \theta \partial_\mu j^\mu = 0$, so the action is invariant.

Therefore, a coupling between such a vector field A_μ and the energymomentum tensor of the form $A_\mu A_\nu T^{\mu\nu}$ is immediately ruled out, because it is not gauge invariant (and furthermore, the simultaneous exchange of two gauge bosons gives a potential $1/r^3$ rather than $1/r$.) A derivative coupling $(\partial_\mu A_\nu) T^{\mu\nu}$ is also not viable, since after integration by parts it gives zero, because of energy-momentum conservation.

If we limit ourselves to the level of first quantization, we could write down a coupling of a vector field to a point-like particles in the form

$$\int d^4x A_\mu(x) m \frac{dx_0^\mu}{d\tau} \delta^{(3)}(\mathbf{x} - \mathbf{x}_0(t)) \quad (2.75)$$

where $x_0^\mu(\tau)$ is the particle world-line. However, in quantum field theory the four-vector $j^\mu(x)$ that, in the limit of point-like particle, reduces to $m(dx_0^\mu/d\tau)\delta^{(3)}(\mathbf{x} - \mathbf{x}_0(t))$, is a $U(1)$ current. Then $Q \equiv \int d^3x j^0$ is equal to the mass m times (number of particles minus number of antiparticles), so it is not positive definite, and $\int d^3x j^0$ cannot be interpreted as the mass (unless we assign a negative gravitational mass to antiparticles). Furthermore, even if we ignore this problem and interpret j^0 has a mass density, still this attempt to construct a spin-1 theory of gravity fails because, as we know from classical electromagnetism, the interaction mediated by the photon between two particles of the same charge is repulsive. Technically, this comes out because the term $\tilde{T}_1(\mathbf{q})\tilde{D}(\mathbf{q})\tilde{T}_2(-\mathbf{q})$ in eq. (2.68) is now replaced by

$$\tilde{j}^\mu(\mathbf{q})\tilde{D}_{\mu\nu}(\mathbf{q})\tilde{j}^\nu(-\mathbf{q}), \quad (2.76)$$

where $\tilde{D}_{\mu\nu}(q)$ is the propagator of the massless vector field A_μ . In momentum space,

$$\tilde{D}_{\mu\nu}(q) = \frac{-i}{q^2} \eta_{\mu\nu} \quad (2.77)$$

In the static limit $q^2 = -(q^0)^2 + \mathbf{q}^2 \rightarrow \mathbf{q}^2$, and the propagator becomes

$$\tilde{D}_{\mu\nu}(\mathbf{q}) = \frac{-i}{\mathbf{q}^2} \eta_{\mu\nu} \quad (2.78)$$

Because of the factor $\eta_{\mu\nu}$, the propagator of the spatial components A_i is equal to that of a scalar field, but the propagator of A_0 has the opposite sign. In the non-relativistic limit $j^\mu \rightarrow (j^0, 0)$. Then in eq. (2.76) only the component D_{00} contributes, so we get the opposite sign compared to the scalar case, i.e. a repulsive potential between positive masses. In conclusion, also spin-1 is ruled out. Spin $j \geq 3$

Values of the spin $j \geq 3$ are also ruled out: the need for a long-range force requires again a massless field, which can be coupled consistently only to a conserved tensor. Except possibly for total derivative terms, there is no conserved tensor with three or more indices, so massless particles with $j \geq 3$ cannot produce long-range forces, neither gravitational nor of any other type.

The only possibility which is left is $j = 2$, and we examine it in the next section.

6.0.32 2.2.2 The Pauli-Fierz action

The considerations of the previous section lead us to study the action for a spin-2 massless field. To identify the field that describes a massless particle with helicity ± 2 , let us first recall some elementary facts about the representations of the Lorentz and Poincar groups, and their decomposition under representations of the rotation group.

6.0.33 Massless particles in field theory

The irreducible tensor representations of the Lorentz group are given by tensors that, with respect to any pair of indices, are either symmetric and traceless, or antisymmetric. An irreducible representation of the Lorentz group provides of course also a representation of the rotation subgroup $SO(3)$. However, a representation that is irreducible with respect to the full Lorentz group, will be reducible if we limit ourselves to the rotation subgroup (except for the trivial case of the scalar representation), so it decomposes into the direct sum of irreducible representations of the rotation group. For instance a four-vector A_μ is an irreducible representation of the Lorentz group, but from the point of view of rotations it decomposes into a scalar A_0 and a vector $\mathbf{A}$, or

$$A_\mu \in \mathbf{0} \oplus \mathbf{1} \quad (2.79)$$

where we denote by $\oplus$ the direct sum of representations, and by $\mathbf{s}$ the representation of the rotation group corresponding to spin s (so $\mathbf{0}$ is the scalar and $\mathbf{1}$ is the vector representation). The representation $\mathbf{s}$ has dimension $2s + 1$, so in particular the scalar is one-dimensional and the vector is a three-dimensional representation. When we consider tensors with two indices, an antisymmetric tensor $A^{\mu\nu}$ decomposes as

$$A_{\mu\nu} \in \mathbf{1} \oplus \mathbf{1} \quad (2.80)$$

while a traceless symmetric tensor $S^{\mu\nu}$ decomposes as

$$S_{\mu\nu} \in \mathbf{0} \oplus \mathbf{1} \oplus \mathbf{2} \quad (2.81)$$

Therefore, the simplest tensor that contains a spin-2 is the traceless symmetric tensor, and a spin-2 can be described using $S_{\mu\nu}$ and imposing conditions that eliminate the extra degrees of freedom. Equation (2.81) states that the nine independent components of a traceless symmetric tensor with two indices decompose into a scalar, the three components of a spin-1, and the five ($2s + 1$ with $s = 2$) components of a spin-2.

A further complication arises if we want to describe a massless particle. As we recall in more detail in Problem 2.1, particles are representations of the Poincar group, and the physically interesting representations of the Poincar group are of two type. (1) Massive representations, with $-P_\mu P^\mu = m^2 > 0$, labeled by the mass m and by the spin j , which takes integer and half-integer values, $j = 0, 1/2, 1, \dots$. The dimension of these representations is $2j + 1$. (2) Massless representations, $P_\mu P^\mu = 0$, which are two-dimensional (actually, one-dimensional, but become twodimensional if we also require that they are representations of parity) and are characterized by two helicity states $h = \pm j$.

In particular, a massive spin-1 particle has three degrees of freedom, and a massive spin-2 particle has five degrees of freedom. In contrast, a massless particle with $j = 1$ and a massless particle with $j = 2$ both have only two degrees of freedom, the former with helicities $h = \pm 1$, and the latter with $h = \pm 2$. This means that, in the case of a massless spin-1 field, a four-vector field A_μ contains two redundant degrees of freedom, while for a massless spin-2 field the nine components of $S^{\mu\nu}$ contain seven redundant degrees of freedom. The way to eliminate

these spurious degrees of freedom is to introduce a gauge-invariance. For electromagnetism, one imposes that the theory is invariant under

$$A_\mu \rightarrow A_\mu - \partial_\mu \theta \quad (2.82)$$

It is a standard textbook exercise to show that we can choose $\theta(x)$ so to set $A_0 = 0$. A residual gauge-invariance remains, due to the possibility of performing a further transformation with a function $\theta(\mathbf{x})$ independent of time and (making use of the Maxwell equation in vacuum $\nabla \cdot \mathbf{E} = 0$) it can be used to set $\nabla \cdot \mathbf{A} = 0$, so we eliminate also the longitudinal component of a plane wave solution, and we are left with only two degrees of freedom, at least at the classical level. Then, one can quantize the free theory and verify that one obtains a massless particle with two helicity states, the photon.

6.0.34 The graviton and its action

We want to do the same for a massless particle with $j = 2$, that we call the graviton. The strategy is therefore to start from a traceless symmetric tensor, to impose a local invariance, and to write down a Lagrangian that respects this local symmetry. Then (using the equations of motion derived from this Lagrangian) we can see how many degrees of freedom remain in the theory. In fact, it is even technically simpler to start from a tensor $h_{\mu\nu}$ which is symmetric, but not traceless, so from the point of view of Lorentz symmetry it decomposes into the trace and the symmetric traceless part, and from the point of view of rotations $h_{\mu\nu} \in \mathbf{0} \oplus (\mathbf{0} \oplus \mathbf{1} \oplus \mathbf{2})$. We therefore want to impose an invariance that eliminates eight spurious degrees of freedom.

In order to generalize eq. (2.82) to the case of a field $h_{\mu\nu}$ with two indices, we must assign a Lorentz index to the function that parametrizes the gauge transformation. Let us call $\xi_\mu(x)$ this function. Since we must

respect the fact that $h_{\mu\nu}$ is symmetric in (μ, ν) , the natural generalization of eq. (2.82) is

$$h_{\mu\nu}(x) \rightarrow h_{\mu\nu}(x) - (\partial_\mu \xi_\nu + \partial_\nu \xi_\mu) \quad (2.83)$$

that we still call a gauge transformation. Observe that this is nothing but the symmetry (1.8) of linearized Einstein gravity, so we are on the good track for recovering the linearization of general relativity.

Next, we want to construct a gauge-invariant action, for the free theory. Remarkably, the condition of gauge-invariance fixes this action uniquely. In fact, by inspection we see that the possible terms that one can write down, quadratic in $h_{\mu\nu}$ and with two derivatives, are

$$\partial_\rho h_{\mu\nu} \partial^\rho h^{\mu\nu}, \quad \partial_\rho h_{\mu\nu} \partial^\nu h^{\mu\rho}, \quad \partial_\nu h^{\mu\nu} \partial^\rho h_{\mu\rho}, \quad \partial_\nu h^{\mu\nu} \partial_\mu h, \quad \partial^\mu h \partial_\mu h \quad (2.84)$$

where $h = h^\mu_\mu$ is the trace of $h_{\mu\nu}$. Terms of the schematic form $h \partial \partial h$ are related to those written above by a single integration by parts. Furthermore, the second and third term in eq. (2.84) are related by two integrations by parts, that swaps the two derivatives. Therefore, the most general form of the free action is

$$S_2 = \int d^4x [a_1 \partial_\rho h_{\mu\nu} \partial^\rho h^{\mu\nu} + a_2 \partial_\rho h_{\mu\nu} \partial^\nu h^{\mu\rho} + a_3 \partial_\nu h^{\mu\nu} \partial_\mu h + a_4 \partial^\mu h \partial_\mu h],$$

where the label in S_2 stresses that this quantity is quadratic in $h_{\mu\nu}$. We now impose invariance under the gauge transformation (2.83). This fixes all the coefficients $a_1, \dots, a_4$, except of course for an overall normalization. We then obtain (choosing the normalization $a_1 = -1/2$; the sign is fixed requiring that the energy is positive definite),

$$S_2 = \frac{1}{2} \int d^4x [-\partial_\rho h_{\mu\nu} \partial^\rho h^{\mu\nu} + 2\partial_\rho h_{\mu\nu} \partial^\nu h^{\mu\rho} - 2\partial_\nu h^{\mu\nu} \partial_\mu h + \partial^\mu h \partial_\mu h] \quad (2.86)$$

This is the Pauli-Fierz action. Comparison with eq. (2.33) shows that we have indeed recovered the Einstein action of linearized theory, after a rescaling

$$h_{\mu\nu} \rightarrow (32\pi G)^{-1/2} h_{\mu\nu} \quad (2.87)$$

and taking into account that the last term in eq. (2.33) is equal to the second in eq. (2.86) after swapping the two derivatives with integrations by parts. We have therefore found that the linearized Einstein action is the unique action that describes a free massless particle with helicities ± 2 , propagating in flat space.

We can now repeat the considerations already made in Sections 1.1 and 1.2, see in particular the discussion around eqs. (1.26) and (1.27): we can use the gauge-invariance (2.83) to choose the Lorentz gauge (1.18). This eliminates four of the 10 degrees of freedom in $h_{\mu\nu}$, and still leaves a residual gauge invariance, i.e. the transformations (2.83) with functions ξ_μ that satisfy $\square \xi_\mu = 0$. In the vacuum, where $T^{\mu\nu} = 0$, using the Lorentz gauge, the equations of motion derived from the linearized

Einstein action are $\square \bar{h}_{\mu\nu} = 0$, and therefore four functions ξ_μ that satisfy $\square \xi_\mu = 0$ can be used to eliminate four components of $\bar{h}_{\mu\nu}$, so we remain with two degrees of freedom, and we arrive at the TT gauge, eq. (1.31). The requirements of gauge invariance therefore fixes uniquely the linearized action, and leaves us with a massless spin-2 field with two transverse degrees of freedom, the graviton.

For the interaction term, we can write

$$S_{\text{int}} = \frac{\kappa}{2} \int d^4x h_{\mu\nu} T^{\mu\nu} \quad (2.88)$$

The coupling constant κ will be fixed below. Observe that this interaction term is invariant under the gauge transformation (2.83) because, after an integration by parts, the term $(\partial_\mu \xi_\nu + \partial_\nu \xi_\mu) T^{\mu\nu}$ vanishes, using $\partial_\mu T^{\mu\nu} = 0$

Just as in electrodynamics, to find the graviton propagator we must add a gauge-fixing term to eq. (2.86), since otherwise the quadratic form is not invertible. The Lorentz gauge can be obtained adding the gaugefixing term

$$\begin{aligned} S_{\text{gf}} &= - \int d^4x (\partial^\nu \bar{h}_{\mu\nu})^2 \\ &= - \int d^4x \left(\partial^\nu h_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} \partial^\nu h \right)^2 \\ &= \int d^4x \left(-\partial_\rho h_{\mu\nu} \partial^\nu h^{\mu\rho} + \partial_\nu h^{\mu\nu} \partial_\mu h - \frac{1}{4} \partial^\mu h \partial_\mu h \right) \end{aligned} \quad (2.89)$$

where in the first term we swapped the derivatives integrating by parts. The overall numerical coefficient in eq. (2.89) has been chosen so that the terms $\partial_\rho h_{\mu\nu} \partial^\nu h^{\mu\rho}$ and $\partial_\nu h^{\mu\nu} \partial_\mu h$ cancel between S_2 and S_{gf} , so this choice corresponds to the Feynman gauge in electrodynamics. Putting everything together, we find

$$\begin{aligned} S &= S_2 + S_{\text{gf}} + S_{\text{int}} \\ &= \int d^4x \left[-\frac{1}{2} \partial_\rho h_{\mu\nu} \partial^\rho h^{\mu\nu} + \frac{1}{4} \partial^\mu h \partial_\mu h + \frac{\kappa}{2} h_{\mu\nu} T^{\mu\nu} \right] \end{aligned} \quad (2.90)$$

The equations of motion obtained performing the variation of this action are

$$\square \bar{h}_{\mu\nu} = -\frac{\kappa}{2} T_{\mu\nu} \quad (2.91)$$

or, equivalently

$$\square h_{\mu\nu} = -\frac{\kappa}{2} \left(T_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} T \right) \quad (2.92)$$

Comparing with eq. (1.24) and taking into account the rescaling (2.87), we get (in the units $c = 1$ that we are using in this section),

$$\kappa = (32\pi G)^{1/2} \quad (2.93)$$

¹³ Alternatively, we can define the 10 fields $\phi_i, i = 1, \dots, 10$, by $\phi_1 = h_{00}, \phi_2 = h_{01}, \dots, \phi_{10} = h_{33}$. We then write the kinetic term in the form $A_{ij} \partial_\mu \phi_i \partial^\mu \phi_j$, and we invert the 10×10 matrix A_{ij} .

6.0.35 The graviton propagator

We now find the graviton propagator. Integrating by parts, we have

$$\int d^4x \left[-\frac{1}{2} \partial_\rho h_{\mu\nu} \partial^\rho h^{\mu\nu} + \frac{1}{4} \partial^\mu h \partial_\mu h \right] = \frac{1}{2} \int d^4x h^{\mu\nu} A_{\mu\nu\rho\sigma} \partial^2 h^{\rho\sigma} \quad (2.94)$$

where

$$A_{\mu\nu\rho\sigma} = \frac{1}{2} (\eta_{\mu\rho} \eta_{\nu\sigma} + \eta_{\mu\sigma} \eta_{\nu\rho} - \eta_{\mu\nu} \eta_{\rho\sigma}) \quad (2.95)$$

The graviton propagator in this gauge is obtained inverting this matrix. Observe that

$$A_{\mu\nu\alpha\beta} A^{\alpha\beta}_{\rho\sigma} = \frac{1}{2} (\eta_{\mu\rho} \eta_{\nu\sigma} + \eta_{\mu\sigma} \eta_{\nu\rho}) . \quad (2.96)$$

The right-hand side is nothing but the identity matrix, in the space of tensors symmetric in (μ, ν) and in (ρ, σ) . Therefore in this space the inverse of A is A itself, ¹³ so the propagator is given by

$$\tilde{D}_{\mu\nu\rho\sigma}(k) = \frac{1}{2} (\eta_{\mu\rho} \eta_{\nu\sigma} + \eta_{\mu\sigma} \eta_{\nu\rho} - \eta_{\mu\nu} \eta_{\rho\sigma}) \left(\frac{-i}{k^2 - i\epsilon} \right) \quad (2.97)$$

where, as usual, the $i\epsilon$ prescription selects the Feynman propagator. In particular, $\tilde{D}_{0000}(k) = -i/(2k^2)$ and $D_{0000}(r) = -i/(8\pi r)$.¹⁴ Comparing with eqs. (2.66) and (2.78) we see that the propagator of h_{00} has the same sign as the propagator of the scalar field (and the opposite sign of the propagator of A_0), since $(\eta_{00})^2 = +1$. Therefore, in the static limit h_{00} mediates an attractive gravitational potential. Using the interaction terms (2.88) and repeating the same steps performed for the scalar field, we get

$$\begin{aligned} V(\mathbf{x}) &= -i \frac{\kappa^2}{4} \int \frac{d^3q}{(2\pi)^3} \tilde{T}_1^{00}(\mathbf{q}) \tilde{D}_{0000}(\mathbf{q}) \tilde{T}_2^{00}(-\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{x}} \\ &= -i \frac{\kappa^2}{4} m_1 m_2 D_{0000}(\mathbf{x}) \\ &= -\frac{\kappa^2}{32\pi} \frac{m_1 m_2}{r} \end{aligned} \quad (2.98)$$

¹⁴ The overall factor depends of course on the overall normalization of the action, i.e. on the normalization of $h_{\mu\nu}$. With our choice the field h_{00} is not canonically normalized, since its propagator has an extra factor $1/2$ with respect to the usual normalization. In contrast, writing $\phi_{\mu\nu} = h_{\mu\nu} - (1/4)\eta_{\mu\nu}h$, the field $\phi_{\mu\nu}$ is canonically normalized. Observe that $\phi_{\mu\nu}$ is a pure spin-2 field since it is traceless, while h is a spin-0 field.

This gives again $\kappa = (32\pi G)^{1/2}$, in agreement with eq. (2.93) (of course, this numerical value depends on the choice of normalization made for $h_{\mu\nu}$), and we have recovered the Newtonian limit.

6.0.36 2.2.3 From gravitons to gravity

We can now understand, from a field-theoretical point of view, that the simple action (2.90) cannot be the whole story, and that the correct field theory of gravitation must develop a full non-linear structure. The reason is that, as we have seen, the theory of a massless particle with helicities ± 2 must necessarily be gauge invariant. We have shown that we can use gauge-invariance to impose the Lorentz condition

$$\partial^\mu \bar{h}_{\mu\nu} = 0 \quad (2.99)$$

and that in this gauge the equations of motion read

$$\square \bar{h}_{\mu\nu} = -\frac{\kappa}{2} T_{\mu\nu} \quad (2.100)$$

Equations (2.99) and (2.100) together imply that the energy-momentum tensor of matter satisfies the flat-space conservation law $\partial_\mu T^{\mu\nu} = 0$. In integrated form, this conservation laws reads

$$\frac{d}{dt} \int_V d^3x T^{00} = - \int_V d^3x \partial_i T^{0i} \quad (2.101)$$

and states that any change in the energy of the matter field in a volume V is due uniquely to the flux of matter field flowing inside or outside this volume. However, this can be true only as long as we consider $T^{\mu\nu}$ as a given classical external source. As soon as we replace it with the energy-momentum tensor of dynamical matter fields, there will necessarily be an exchange of energy and of momentum between matter and the gravitational field. For instance, there will be gravitational radiation emitted by the matter and going off to infinity, draining energy from the matter sources. Therefore the equation $\partial_\mu T^{\mu\nu} = 0$ is untenable in a full dynamical theory, and eq. (2.100) cannot be exact.

In the full theory what should be conserved is not the energy and momentum of the matter alone, but that of matter plus gravitational field. To remedy this we could try to perform, on the right-hand side of eq. (2.100), the replacement

$$T_{\mu\nu} \rightarrow T_{\mu\nu} + t_{\mu\nu}^{(2)} \quad (2.102)$$

where $t_{\mu\nu}^{(2)}$ is the energy-momentum tensor of the gravitons, obtained from the Pauli-Fierz action using the Noether theorem. Using eq. (2.10), we see that a Lagrangian quadratic in $h_{\mu\nu}$, such as the Pauli-Fierz Lagrangian, produces an energy-momentum tensor quadratic in $h_{\mu\nu}$, of the type, symbolically, $\partial h \partial h$. In eq. (2.102) we added to $t_{\mu\nu}$ the superscript (2) to emphasize that it is quadratic in $h_{\mu\nu}$. Then, we are led to try

$$\square \bar{h}_{\mu\nu} = -\frac{\kappa}{2} (T_{\mu\nu} + t_{\mu\nu}^{(2)}) \quad (2.103)$$

and the gauge condition $\partial^\mu \bar{h}_{\mu\nu}$ is consistent with the conservation of the graviton plus matter energy-momentum tensor,

$$\partial^\mu (T_{\mu\nu} + t_{\mu\nu}^{(2)}) = 0 \quad (2.104)$$

However, this cannot yet be the end of the story. The equation of motion (2.103) has a term, $(-1/2)\kappa t_{\mu\nu}^{(2)}$, which is quadratic in $h_{\mu\nu}$ (and linear in κ) and to derive it from an action principle, we must add to the action

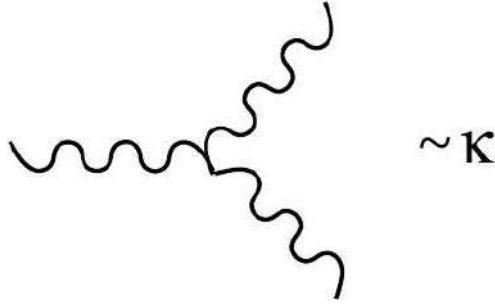


Fig. 2.4 The three-graviton vertex. Two lines carry a contribution proportional to their momentum. a term cubic in $h_{\mu\nu}$ and proportional to κ . Symbolically (i.e. omitting Lorentz indices), $t_{\mu\nu}^{(2)} \sim \partial h \partial h$ and a term of this type in the equations of motion can be obtained adding to the action a term of the form $h \partial h \partial h$, or, restoring Lorentz indices,

$$S_3 = \frac{\kappa}{2} \int d^4x h_{\mu\nu} \mathcal{S}^{\mu\nu}(\partial h) \quad (2.105)$$

where $\mathcal{S}^{\mu\nu}(\partial h)$ should be of the general form

$$\mathcal{S}^{\mu\nu}(\partial h) = A^{\mu\nu\rho\sigma\alpha\beta\gamma\delta} \partial_\rho h_{\alpha\beta} \partial_\sigma h_{\gamma\delta} \quad (2.106)$$

and the tensor $A^{\mu\nu\rho\sigma\alpha\beta\gamma\delta}$ is a product of flat metric factors. At this stage, the action of the matter plus gravitational field must therefore be of the general form

$$\begin{aligned} S &= S_2 + S_{\text{gf}} + S_{\text{int}} + S_3 \\ &= \int d^4x \left[-\frac{1}{2} \partial_\rho h_{\mu\nu} \partial^\rho h^{\mu\nu} + \frac{1}{4} \partial^\mu h \partial_\mu h + \frac{\kappa}{2} h_{\mu\nu} T^{\mu\nu} + \frac{\kappa}{2} h_{\mu\nu} \mathcal{S}^{\mu\nu}(\partial h) \right] \end{aligned} \quad (2.107)$$

Observe that $\mathcal{S}^{\mu\nu}(\partial h)$ is not equal to the energy-momentum tensor $t_{\mu\nu}^{(2)}$. In fact, the variation of a term $h^{\mu\nu} t_{\mu\nu}^{(2)}(h)$ in the action would produce, in the equations of motion, a term

$$\frac{\delta}{\delta h^{\mu\nu}} \left[h^{\alpha\beta} t_{\alpha\beta}^{(2)}(h) \right] = t_{\mu\nu}^{(2)}(h) + h^{\alpha\beta} \frac{\delta}{\delta h^{\mu\nu}} \left[t_{\alpha\beta}^{(2)}(h) \right] \quad (2.108)$$

so there would be an extra term due to the variation of $t_{\alpha\beta}^{(2)}(h)$. Therefore, in the action $h_{\mu\nu}$ couples to matter through the energy-momentum tensor of matter, but the coupling to itself is through a different tensor. In other words, the equivalence principle means that the energymomentum tensor of GWs enters on the same footing as the energymomentum tensor of matter in the equations of motion, not in the action.

The cubic term in the action means that gravitons have a non-linear coupling to themselves and, in the language of Feynman graphs, corresponds to a vertex as shown in Fig. 2.4. It is instructive to compare this situation with electrodynamics and with Yang-Mills theories. In electrodynamics, the photon mediates the interaction, but carries no electric charge. Therefore it does not couple directly to itself, and it does not contribute to the electric current. In Yang-Mills theories, instead, the gauge bosons are charged with respect to the gauge group, and therefore they couple non-linearly to themselves, and in the theory there are three-boson vertices and four-boson vertices. The situation for gravitons is analogous. Here the role of

the current is played by the energy-momentum tensor, and the gravitons couple to their own energymomentum tensor (in the equations of motion; or to the corresponding functional $\mathcal{S}^{\mu\nu}(\partial h)$ in the action). We see that, with the inclusion of the cubic term in the action, our gauge theory of a massless particle with helicities ± 2 begins to resemble to a non-Abelian gauge theory.

This suspicion is confirmed observing that the action (2.107) is no longer invariant under the linear gauge transformation (2.83) that was our starting point, since the term $h_{\mu\nu}T^{\mu\nu}$ is invariant only if $\partial_\mu T^{\mu\nu} = 0$, which is no longer true. Rather, under the transformation (2.83),

$$\begin{aligned}\delta \int d^4x \frac{\kappa}{2} h_{\mu\nu} T^{\mu\nu} &= \kappa \int d^4x (\partial_\mu \xi_\nu) T^{\mu\nu} \\ &= -\kappa \int d^4x \xi_\nu \partial_\mu T^{\mu\nu} \\ &= +\kappa \int d^4x \xi_\nu \partial_\mu t^{\mu\nu(2)} \\ &= -\kappa \int d^4x (\partial_\mu \xi_\nu) t^{\mu\nu(2)}\end{aligned}\tag{2.109}$$

However, we have seen that the presence of a local gauge invariance is crucial to eliminate the spurious degrees of freedom in $h_{\mu\nu}$, and we cannot afford to lose it at higher orders in κ . To remedy this, one can observe that this extra term can be canceled, at $O(\kappa)$, promoting the linear gauge transformation (2.83) to a non-linear transformation of the generic form

$$h_{\mu\nu} \rightarrow h_{\mu\nu} - (\partial_\mu \xi_\nu + \partial_\nu \xi_\mu) + \kappa O(h\partial\xi)\tag{2.110}$$

Then the transformation of the graviton kinetic term produces a term $O(\kappa)$, and it is possible to choose the tensorial structure of the $O(h\partial\xi)$ term so that it cancels the extra term coming from eq. (2.109). Thus, eq. (2.83) is the gauge symmetry of the theory only at the infinitesimal level, and at the finite level a non-linear gauge transformation emerges, just as in non-Abelian gauge theories. With the hindsight coming from the fact that we know already that the full theory of gravitation is general relativity, we recognize in eq. (2.110) the expansion up to next-to-leading order of a finite diffeomorphism.

It is clear that the iteration procedure does not stop here, neither in the action nor in the gauge transformation. Once we add to the action the term S_3 , which is cubic in $h_{\mu\nu}$ and proportional to κ , this produces (through Noether's theorem) a contribution to the energy-momentum tensor of the graviton again cubic in $h_{\mu\nu}$ and proportional to κ . We find useful to display the powers of κ explicitly, so we denote it by $\kappa t_{\mu\nu}^{(3)}$. Noether theorem, applied to the action (2.107), now gives

$$\partial^\mu (T_{\mu\nu} + t_{\mu\nu}^{(2)} + \kappa t_{\mu\nu}^{(3)}) = 0\tag{2.111}$$

Therefore, consistency with $\partial^\mu \bar{h}_{\mu\nu} = 0$ now requires to replace eq. (2.103) by

$$\square \bar{h}_{\mu\nu} = -\frac{\kappa}{2} (T_{\mu\nu} + t_{\mu\nu}^{(2)} + \kappa t_{\mu\nu}^{(3)})\tag{2.112}$$

In turn, the action which generates this equation of motion, that has a term cubic in $h_{\mu\nu}$ and proportional to κ^2 , must have an additional term quartic in $h_{\mu\nu}$ and proportional to κ^2 , so also its associated energy-

¹⁵ Nor the general covariance of the theory. Since boundary terms in the action do not affect classical physics, a classical theory is invariant under a given transformation even if its action changes by a boundary term. This typically happens, for instance, in supersymmetric theories,

where under a supersymmetry transformation the action changes by a boundary term. momentum tensor has a further term quartic in the fields and proportional to κ^2 , that we denote by $\kappa^2 t_{\mu\nu}^{(4)}$, and then

$$\square \bar{h}_{\mu\nu} = -\frac{\kappa}{2} (T_{\mu\nu} + t_{\mu\nu}^{(2)} + \kappa t_{\mu\nu}^{(3)} + \kappa^2 t_{\mu\nu}^{(4)} + \dots) \quad (2.113)$$

where the dots indicate that the iteration continues to all orders. We recognize the full non-linear structure typical of Einstein gravity, with arbitrarily large powers of $h_{\mu\nu}$, and a non-linear gauge invariance. Indeed, there exists a simple and explicit resummation algorithm, due to Deser, which gives back the Einstein equations. This algorithm uses the first order Palatini formalism, where the action becomes a cubic polynomial in the variable $\sqrt{-g}g_{\mu\nu}$ and in the Christoffel symbol (which in a first order formalism are varied independently), and the iteration stops at finite order, see the Further Reading section. So, general relativity can be inferred "bottom up", i.e. starting from gravitons. However, some aspects of this reconstruction procedure are not unique. In particular the full Einstein action includes a boundary term, since it can be written as

$$\begin{aligned} S_E &= \frac{1}{16\pi G} \int d^4x \sqrt{-g} R \\ &= \frac{1}{16\pi G} \int d^4x [\sqrt{-g} \mathcal{L}_2 - \partial_\mu K^\mu] \end{aligned} \quad (2.114)$$

where

$$\mathcal{L}_2 = g^{\mu\nu} \left(\Gamma_{\beta\mu}^\alpha \Gamma_{\alpha\nu}^\beta - \Gamma_{\mu\nu}^\alpha \Gamma_{\alpha\beta}^\beta \right) \quad (2.115)$$

is quadratic in the first derivative of the metric (it is usually called the ' Γ ' Lagrangian), and

$$K^\mu = \sqrt{-g} (g^{\mu\nu} \Gamma_{\alpha\nu}^\alpha - g^{\alpha\beta} \Gamma_{\alpha\beta}^\mu) \quad (2.116)$$

The latter term is a total divergence and therefore does not affect the equations of motion.¹⁵ However, boundary terms are relevant in the quantum theory, and in particular the boundary term in eq. (2.114) becomes physically relevant in semiclassical quantum gravity, in connection with black hole thermodynamics. It is clear that boundary terms are beyond the reach of the iterative procedure that starts from the Pauli-Fierz action, since at each stage of the procedure we have ambiguities connected with the possibility of dropping boundary terms. For instance, from the very beginning, we might have chosen to retain the third term in eq. (2.84) rather than the second term. More formally, expanding $g_{\mu\nu} = \eta_{\mu\nu} + \kappa h_{\mu\nu}$, $\mathcal{L}_2$ has the structure (we drop all Lorentz indices, since we are only interested in the dependence on κ)

$$\begin{aligned} \frac{1}{32\pi G} \int d^4x \sqrt{-g} \mathcal{L}_2 &\sim \frac{1}{\kappa^2} \int d^4x [\kappa^2 (\partial h)^2 + \kappa^3 h \partial h \partial h + \dots] \\ &\sim \int d^4x [(\partial h)^2 + \kappa h \partial h \partial h + O(\kappa^2)] \end{aligned} \quad (2.117)$$

and therefore is analytic in κ . On the contrary,

$$\begin{aligned} \frac{1}{32\pi G} \int d^4x \partial_\mu K^\mu &\sim \frac{1}{\kappa^2} \int d^4x \partial^2 (\eta + \kappa h) \\ &\sim \frac{1}{\kappa} \int d^4x \partial^2 h \end{aligned} \quad (2.118)$$

Therefore this term is non-analytic in κ and cannot be obtained unambiguously from the resummation of an expansion in powers of κ , without further physical input.

Alternatively we can proceed top-down, starting from Einstein action and expanding it in powers of $h_{\mu\nu}$, obtaining all non-linear interaction terms. Observe that in this section we have given to $h_{\mu\nu}$ the dimensions of mass that are canonical in field theory, and the dimensionless metric $h_{\mu\nu}$ is recovered with the rescaling (2.87). Therefore, the expansion that we are performing in this section can be written as

$$\begin{aligned} g_{\mu\nu} &= \eta_{\mu\nu} + (32\pi G)^{1/2} h_{\mu\nu} \\ &= \eta_{\mu\nu} + \kappa h_{\mu\nu} \end{aligned} \quad (2.119)$$

where $h_{\mu\nu}$ has the canonical dimension of mass and κ , dimensionally, is the inverse of a mass. Given the cubic, quartic, and higher terms in the action, one can read the corresponding vertices and compute scattering amplitudes using the Feynman rules. For instance, the $2 \rightarrow 2$ graviton scattering amplitude is obtained from the s -channel graph of Fig. 2.5 (together with the corresponding u -channel and t -channel graphs), and from the four-graviton vertex of Fig. 2.6, so this amplitude is $O(\kappa^2)$, i.e. $O(G)$.¹⁶

6.0.37 2.2.4 Effective field theories and the Planck scale

In units $\hbar = c = 1$, the Newton constant has the dimensions of the inverse of a mass squared, and therefore $\kappa \sim G^{1/2}$ is the inverse of a mass. The fact that the coupling constant κ has dimensions of the inverse of mass means that the quantization of gravity should give rise to a non-renormalizable theory. This is expected first of all on the basis of simple dimensional considerations. For instance, we have seen that the three-graviton vertex has the generic form $\kappa h \partial h \partial h$. The fact that there are two derivatives is dictated by dimensional analysis. In fact, the dimensions of any term in the Lagrangian density $\mathcal{L}$ must be $(\text{mass})^4$, so that the action $\int d^4x \mathcal{L}$ is dimensionless (recall that in this section we are using units $\hbar = c = 1$, and then in particular a length has the dimensions of $(\text{mass})^{-1}$, see Note 2). The kinetic term is, symbolically, $(\partial h)^2$, and since $\partial_\mu = 1/\text{length} = \text{mass}$, we see that $h_{\mu\nu}$ must have dimensions of mass.

The three-graviton vertex therefore must necessarily carry two powers of the momentum: in fact $\kappa h h h \sim (\text{mass})^2$, and to get something trilinear in h and proportional to κ we must add two derivatives, so we must have something of the form $\kappa h \partial h \partial h$. In momentum space each

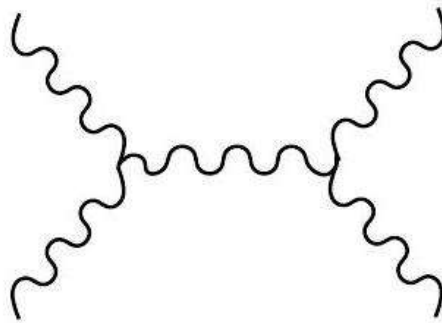


Fig. 2.5 A diagram contributing to the $2 \rightarrow 2$ graviton scattering amplitude.

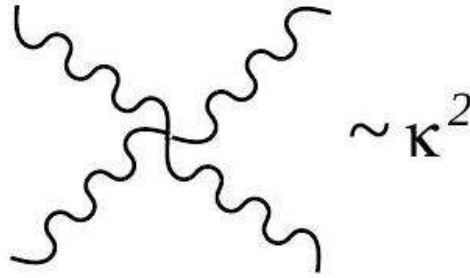


Fig. 2.6 The four-graviton vertex.

¹⁶ The explicit expression for the three-graviton vertex, and the (discouragingly long) four-graviton vertex can be found in eqs. (2.6) and (2.7) of DeWitt (1967). To compute a $2 \rightarrow 2$ graviton scattering amplitude in the Born approximation it is however sufficient to know the three-graviton vertex with two gravitons on-shell and only one offshell, while in the four-graviton vertex all lines are on-shell. Then in the on-shell lines we can impose $\partial^\nu h_{\mu\nu} = 0$ and $h = 0$, and the vertices becomes more manageable, see Grisaru, van Nieuwenhuizen and Wu (1975).

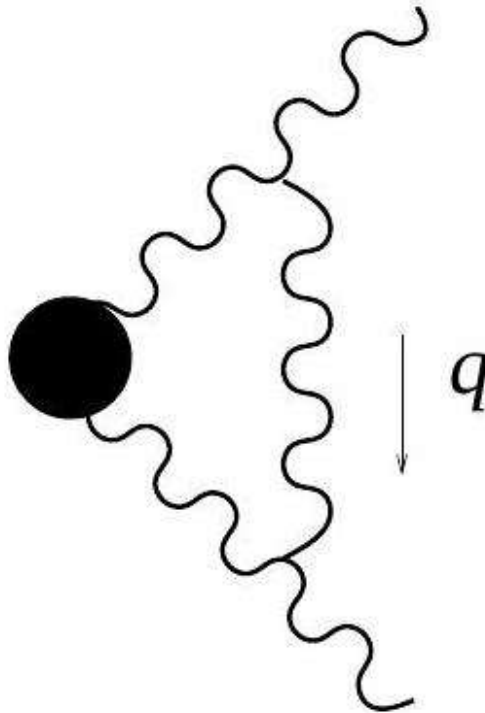


Fig. 2.7 The insertion of an internal graviton line among two external lines. The blob denotes all the remaining (unspecified) part of the Feynman graph.

¹⁷ More precisely, in the background field method one expands $g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}$, where $\bar{g}_{\mu\nu}$ is a classical solution of the equations of motion and $h_{\mu\nu}$ is a quantum field which, in the path integral formulation, is integrated over. After integrating over the quantum fluctuations $h_{\mu\nu}$, we are left with the counterterms evaluated on $\bar{g}_{\mu\nu}$. For pure gravity, i.e. when we neglect matter fields, $R_{\mu\nu}$ evaluated on any classical solution $\bar{g}_{\mu\nu}$ gives zero, and therefore the one-loop divergences disappear. This miracle does not repeat at higher orders, and does not take place even at one loop if we also have matter fields (in pure supergravity the divergencies cancel even at two-loop order, but not to higher orders). derivative gives a power of momentum, and we can see, writing explicitly some Feynman graphs, that with more and more insertions of this vertex we get Feynman graphs which (barring cancellations) are more and more divergent. Consider for instance the insertion of an internal graviton line in a Feynman graph, obtained adding two

three-graviton vertices on two external lines, and connecting them by a graviton propagator, as in Fig. 2.7. Denote by q^μ the momentum of this internal line. Adding this internal line we have created a new closed loop, which gives an integration over d^4q . There are three propagators in the loop, and at large q each of them is $\sim 1/q^2$. Finally, each of the two vertices gives a factor $q_\mu q_\nu$. Then, barring cancellations, the insertion of two vertices brings a further factor that, in the ultraviolet, is of order

$$\kappa^2 \int d^4q \frac{1}{(q^2)^3} q_\mu q_\nu q_\rho q_\sigma \quad (2.120)$$

and so is quadratically divergent. Therefore, adding two vertices we have introduced a new divergence in the graph. Inserting more vertices worsens the situation further and we get stronger and stronger divergencies. The same happens with insertions of the four-graviton vertex, five-graviton vertex, etc.

This lack of renormalizability, however, is not at all a problem as long as we study processes taking place at sufficiently small energies. The problem with non-renormalizable theories is a matter of predictivity, not of mathematical consistency: in the language of counterterms, the divergences are canceled adding to the Lagrangian terms that have a different functional form, compared to the only term $\sqrt{-g}R$ present in Einstein gravity and, as we increase the number of loops, more and more counterterms are required. For instance, to one-loop order one must add to the Lagrangian a term proportional to $R_{\mu\nu}R^{\mu\nu}$ and a term proportional to R^2 , to two-loop order we have for instance terms proportional to $R_{\mu\nu\rho\sigma}R^{\rho\sigma\alpha\beta}R_{\alpha\beta}{}^{\mu\nu}$, etc.¹⁷ Therefore the divergences cannot be reabsorbed into a renormalization of the Newton constant and of the fields. Rather, the coefficients of all these terms have a divergent part which is chosen so that it cancels the divergencies coming from the loops, and a finite part that must be fixed by comparison with experiments. Thus, we end up with a theory which, by definition, is finite, but apparently is left with very little predictive power, since we must fix an infinite number of amplitudes by comparison with experiments.

However, this loss of predictivity is not important at low energies. In quantum gravity, we have seen that the coupling constant is $\kappa = (32\pi G)^{1/2}$. Defining the Planck mass M_{Pl} (in our units $\hbar = c = 1$) by

$$G = \frac{1}{M_{\text{Pl}}^2} \quad (2.121)$$

we have

$$\kappa = \frac{(32\pi)^{1/2}}{M_{\text{Pl}}} \quad (2.122)$$

The perturbative expansion in quantum gravity, which is an expansion in κ^2 , is therefore an expansion in powers of $1/M_{\text{Pl}}^2$. After renormalizing the theory, each scattering amplitude with N external legs, computed up to order $(\kappa^2)^n \sim 1/M_{\text{Pl}}^{2n}$, has the generic form

$$A_N(E) = A_N^0(E) \left(1 + c_1 \frac{E^2}{M_{\text{Pl}}^2} + c_2 \frac{E^4}{M_{\text{Pl}}^4} + \dots + c_n \frac{E^{2n}}{M_{\text{Pl}}^{2n}} \right) \quad (2.123)$$

For simplicity we assumed that there is only one relevant energy scale E in the amplitude (which in principle depends on all the Lorentz-invariant quantities that one can make with the external momenta). The nonrenormalizability of the theory means that, whatever the value of N , we can always find a sufficiently large order n in perturbation theory, where a genuinely new divergence appears, which is not automatically cured by the renormalization of Green's functions with a smaller number of external legs. Therefore, the coefficients $c_1, c_2, \dots, c_{n-1}$ are

finite and calculable once we have renormalized the Green's functions A_M with $M < N$, but the coefficient c_n must be fixed by comparison with experiment, and this is the origin of the lack of predictivity.

However, as long as $E \ll M_{\text{Pl}}$, terms suppressed by powers of E/M_{Pl} are completely irrelevant, and the lack of predictivity due to the fact that c_n should be fixed by comparison with experiment is more apparent than real. As long as we study, say, a graviton-graviton scattering process at a center-of-mass energy $E \ll M_{\text{Pl}}$, all higher-loop corrections are totally insignificant. Given the huge value of the Planck mass,

$$M_{\text{Pl}} \simeq 1.2 \times 10^{19} \text{GeV} \quad (2.124)$$

this means that the lowest-order effective action, i.e. classical general relativity, is completely adequate in all "normal" situations. To see the effect of higher-order terms in the perturbative expansion, we should for instance perform a scattering experiment at center-of-mass energies of order M_{Pl} , or, equivalently, we should probe the structure of spacetime at length-scale of the order of the Planck length, $l_{\text{Pl}} = 1/M_{\text{Pl}} \sim 10^{-33}$ cm. In accelerator physics, this is beyond any conceivable future development. Possibly, our best chances of exploring this extremely energetic region is through the observation of some relic of the Big Bang, and in particular stochastic backgrounds of relic GWs, as we will discuss in Vol. 2. However, as long as we consider "normal" astrophysical situations, the lack of predictivity at the Planck scale becomes irrelevant, and general relativity is a totally adequate low-energy effective field theory. The fact that the expansion (2.123) blows up as E approaches M_{Pl} signals that at this point a new theory must set in.

6.0.38 2.3 Massive gravitons

In Section 2.2 we introduced the notion of graviton, and we saw that it is described by a massless spin-2 field. From a particle physicist's point of view, one of the most natural extensions of Einstein gravity consists in adding to the graviton a small mass term. However, as we will see in this section, the introduction of a mass term for the graviton turns out to be This section lies outside the main theme of the book, and can be omitted in a first reading. As in the previous section, we use units $\hbar = c = 1$.

¹⁸ See Binney and Tremaine (1994), Table 10.1. Section 10.2.1. quite subtle from a field-theoretical point of view. In particular the limit $m_g \rightarrow 0$ is very delicate, up to the point that one is led to discuss whether the graviton mass should be identically zero. Before entering into these considerations, however, we discuss at a simpler phenomenological level the bounds on the graviton mass.

6.0.39 2.3.1 Phenomenological bounds

In general, we expect that a boson with a mass m_g should mediate a short-range force which, compared to the massless case, is suppressed by a factor proportional to $\exp\{-m_g r\}$. In the case of the graviton, such an exponential would cut off the gravitational interaction at a distance r larger than the reduced Compton wavelength $\lambda_g = 1/m_g$ (or, reinstating $\hbar$ and c , at distances larger than $\lambda_g = \hbar/(m_g c)$).

However, we know experimentally that our Galaxy is held together by gravitation, which means that, at least up to a scale $r_{\text{gal}} \sim 10\text{kpc}$, there is no sensible attenuation of the Newton's law, so λ_g cannot be much smaller than 10kpc . Taking, for definiteness, $\lambda_g > 2\text{kpc}$, this already gives a bound $m_g = 1/\lambda_g < 3 \times 10^{-27} \text{eV}$.

¹⁹ See Binney and Tremaine (1994),

Actually, the experimental bound on the mass of the graviton is even stronger, since we know that the gravitational interaction is not exponentially suppressed even at the intergalactic scale. Our Galaxy has a number of small satellite galaxies, bound by the gravitational force, at distances up to 260kpc¹⁸ The Andromeda galaxy, at a distance of 730kpc, is the nearest giant spiral galaxy, and is approaching the center of mass of our galaxy with a speed $v = -119$ km/s. The most natural explanation is that the relative Hubble expansion between our Galaxy and Andromeda has been halted and reversed by the mutual gravitational attraction.¹⁹ This tells us that λ_g cannot be much smaller than, say, $O(100)$ kpc. On the scale of hundreds of kpc to 1Mpc, galaxies are seen to be distorted gravitationally by their reciprocal interaction, creating bridges and tails in their shapes. On the scale $(1-10) h_0^{-1}$ Mpc, where h_0 is the Hubble expansion rate in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (the current value is $h_0 = 0.73 \pm 0.03$), clusters of galaxies are held together by the gravitational attraction, so we can infer that λ_g cannot be much smaller than a few Mpc. Taking, conservatively,

$$\lambda_g > 300 h_0^{-1} \text{kpc} \quad (2.125)$$

results in a bound

$$m_g < 2 \times 10^{-29} h_0 \text{eV} \quad (2.126)$$

This bound, of course, refers only to the lightest particle that mediates the gravitational interaction. In some extensions of general relativity, and in particular in theories with extra dimensions, the graviton is accompanied by a whole family of massive excitations (the Kaluza-Klein modes). The bound (2.126) only refers to the lowest lying state, and says nothing about the possibility of further massive gravitons, whose effect vanishes on much shorter length-scales.

There is however a potential loophole in the above arguments. We saw in eq. (2.98) that the static gravitational potential is determined by D_{0000} , i.e. by the propagator of the component h_{00} of $h_{\mu\nu}$, which is a scalar under spatial rotation. Gravitational waves, instead, are described by h_{ij}^{TT} , which is a spin-2 tensor under rotations. It is possible to construct consistent models where Lorentz invariance is broken and the masses of scalar, vector and tensor perturbations are different. In particular, h_{ij}^{TT} can be massive while scalar perturbations (obtained from gauge-invariant combinations of h_{00} and of the trace h) remains massless, see the Further Reading. The bounds that we discussed above really refer to the mass of the scalar perturbations, and is the same as the mass of h_{ij}^{TT} only if Lorentz invariance holds.

A direct bound on the mass of the tensor mode h_{ij}^{TT} can be obtained from binary pulsars. A binary neutron star system loses energy because it radiates GWs, and this changes its orbital period. The remarkable agreement between the prediction of general relativity for the orbital change, and the measured value for the binary pulsar PSR B1913+16, is in fact one of the great experimental triumphs of general relativity, and also constitutes the first experimental confirmation of the emission of GWs, and will be discussed in great detail in Chapter 6.

We can understand qualitatively how a bound on the graviton mass emerges from the study of binary pulsars, as follows. The system emits GWs at frequencies of order of its orbital frequency ω_s .²⁰ Then, first of all we must have $m_g c^2 < O(\hbar \omega_s)$ or, in the units $\hbar = c = 1$ that we are using in this chapter, $m_g < O(\omega_s)$, otherwise such massive gravitons could not even be emitted. For m/ω_s small, one finds that the correction to the energy radiated in GWs by the source are of order $(m_g/\omega_s)^2$, with respect to the massless case.²¹ Since the agreement between theory and experiment is of order of 0.1% (see Chapter 6) we must actually have, in order of magnitude, $(m_g/\omega_s)^2 < O(10^{-3})$. Given that the orbital period of PSR B1913+16 is about 8hr, we immediately get an order-of-magnitude estimate of the bound, $m_g < O(10^{-20})$ eV. A more quantitative analysis of the orbital decay rate of PSR B1913+16 (and also of another binary system that will be discussed in Chapter 6, PSR B1534+12) including a mass term for the graviton gives²²

$$m_g < 7.6 \times 10^{-20} \text{eV} \quad (2.127)$$

corresponding to a value of λ_g of the order of the size of the solar system. Similar bounds on m_g would also come from the direct observation of the waveform of inspiraling compact binaries with interferometric detectors.²³

The issue of the graviton mass has potentially important cosmological implications, e.g. in connection with attempts of modifying Einstein gravity at cosmological distances (which is largely motivated by the problem of dark matter and dark energy). Furthermore, the whole subject is quite interesting from a field-theoretical point of view. We therefore discuss, in the rest of this section, the field-theoretical problems that arise when we attempt to give a mass to the graviton.²⁰ As we will compute in Section 4.1.2, for PSR B1913+16 the spectrum of the radiation emitted is actually peaked toward high harmonics of the orbital frequency, because of the large eccentricity of this binary system.

²¹ This comes out repeating the same steps that we will do in Section 3.1, using a massive rather than a massless wave equation for $h_{\mu\nu}$. Since a massive Klein-Gordon type equation depends on m_g^2 , the first correction to the radiated energy will also be proportional to the second power of m_g .

²² This analysis has been done by Finn and Sutton (2002) using a mass term for the graviton which actually implies the existence of six degrees of freedom, the five associated to a massive spin-2 graviton plus an additional scalar field, which however is a ghost, i.e. it has the wrong sign of the kinetic energy (see the discussion in Section 2.3.2 below). Different mass terms, and in particular Lorentz-violating mass terms, should however give similar results.

²³ The basic idea is that, for a massive graviton, the dispersion relation is $\hbar\omega = \gamma mc^2$, where $\gamma = (1 - v^2/c^2)^{-1/2}$ is the usual relativistic factor. Inverting this for v , and using units $\hbar = c = 1$, we get

$$v(\omega) = \sqrt{1 - \frac{m^2}{\omega^2}} \quad (2.128)$$

In the inspiral amplitude (that will be studied in Section 4.1) the radiation emitted earlier in the inspiral phase is at lower frequencies, and therefore travels slightly slower than the radiation emitted at later times, resulting in a potentially observable distortion of the waveform observed at the detector, see Will (1998).

²⁴ For the purpose of our discussion, it will not be important whether this mass term has been added by hand, or if it results from a Higgs mechanism.²⁵ Actually, even if $\partial_\nu j^\nu \neq 0$, we find that A^0 is fixed in terms of $\partial_\nu j^\nu$, so we have anyway eliminated one degree of freedom. Current conservation is however crucial to ensure a smooth limit as $m_\gamma \rightarrow 0$, as we will see below.

6.0.40 2.3.2 Field theory of massive gravitons

6.0.41 A warm-up: massive photons

We want to understand how to construct a consistent field theory for a massive graviton. As a warm-up exercise, let us first recall what happens in electrodynamics if we add by hand a mass term to the photon. The Lagrangian of a massive photon interacting with a conserved current j_μ is

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}m_\gamma^2 A_\mu A^\mu - j_\mu A^\mu \quad (2.129)$$

This is known as the Proca Lagrangian and is not gauge-invariant, because of the mass term.²⁴ This is as it should be, because in the massless case gauge invariance reduces the

four degrees of freedom of A_μ to only two degrees of freedom, the two helicity states of the photon, while we want to describe a massive particle, and a massive particle with spin $j = 1$ has $2j + 1 = 3$ degrees of freedom. Still, in order to describe the three physical degrees of freedom of a massive photon with the four components of A_μ , we need to eliminate one degree of freedom. Of course, we cannot do it by imposing the condition $\partial_\mu A^\mu = 0$ in the above Lagrangian with a gauge-fixing procedure, since there is no gauge symmetry to be fixed. Rather, the condition $\partial_\mu A^\mu = 0$ is recovered as follows. The equations of motion obtained from (2.129) are

$$\partial_\mu F^{\mu\nu} - m_\gamma^2 A^\nu = j^\nu \quad (2.130)$$

Acting with ∂_ν on both sides we find $\partial_\nu \partial_\mu F^{\mu\nu} = 0$, because $\partial_\nu \partial_\mu$ is symmetric while $F^{\mu\nu}$ is antisymmetric, and $\partial_\nu j^\nu = 0$ because we have coupled A^μ to a conserved current. Then, eq. (2.130) implies

$$m_\gamma^2 \partial_\nu A^\nu = 0 \quad (2.131)$$

and, if $m_\gamma \neq 0$, we get the Lorentz condition $\partial_\nu A^\nu = 0$ dynamically, and we have eliminated one degree of freedom. In momentum space this gives $k_\mu \tilde{A}^\mu(k) = 0$. Since for a massive particle the rest frame exists, we can choose the frame where $k^\mu = (m_\gamma, 0, 0, 0)$, so we have eliminated the component A_0 , and we remain with the three components of the vector A that describe a massive spin-1 particle, as it should be.²⁵

Still, it can be disturbing to observe that, apparently, the zero mass limit is discontinuous, because the number of physical degrees of freedom seems to change abruptly from three, for $m_\gamma \neq 0$, to two for $m_\gamma = 0$. To understand this point, let us see what are the polarization states of a massive photon and their coupling to the current j^μ . Consider first the propagation of a free massive photon, i.e. eq. (2.130) with $j^\nu = 0$,

$$\partial_\mu (\partial^\mu A^\nu - \partial^\nu A^\mu) - m_\gamma^2 A^\nu = 0 \quad (2.132)$$

Using $\partial_\mu A^\mu = 0$, which follows from eq. (2.131), this becomes a massive wave equation,

$$(\square - m_\gamma^2) A^\mu = 0 \quad (2.133)$$

whose solution is a superposition of plane waves of the form $\epsilon^\mu(k) e^{ikx}$ with $k^2 = -m_\gamma^2$, and of their complex conjugates. The condition (2.131) implies

$$\epsilon_\mu(k) k^\mu = 0 \quad (2.134)$$

We choose a frame where $k^\mu = (\omega, 0, 0, k_3)$, with $\omega^2 = k_3^2 + m_\gamma^2$. In this frame two solutions of eq. (2.134) are given by the transverse vectors

$$\epsilon_\mu^{(1)}(k) = (0, 1, 0, 0), \quad \epsilon_\mu^{(2)}(k) = (0, 0, 1, 0) \quad (2.135)$$

which are the same as the usual transverse degrees of freedom of a massless photons. For a massive photon there is however also a third physical solution

$$\epsilon_\mu^{(3)}(k) = \frac{1}{m_\gamma} (-k_3, 0, 0, \omega) \quad (2.136)$$

All three polarization vectors are normalized so that $\epsilon_\mu \epsilon^\mu = 1$. To understand what happens to the polarization state (2.136) in the limit $m_\gamma \rightarrow 0$, we observe that we can rewrite it in terms of $k_\mu = (-\omega, 0, 0, k_3)$, as

$$\epsilon_\mu^{(3)}(k) = \frac{1}{m_\gamma} k_\mu + \frac{\omega - k_3}{m_\gamma} (1, 0, 0, 1) \quad (2.137)$$

The interaction of this state with the current j^μ is proportional to $\epsilon_\mu^{(3)}(k)\tilde{j}^\mu(k)$. If the current is conserved, $k_\mu\tilde{j}^\mu(k) = 0$ and the first term in eq. (2.137) does not contribute. In the limit $m_\gamma \rightarrow 0$, we expand $\omega = (k_3^2 + m_\gamma^2)^{1/2} \simeq k_3 + m_\gamma^2/(2k_3)$, and therefore

$$\begin{aligned}\epsilon_\mu^{(3)}(k)\tilde{j}^\mu(k) &= \frac{\omega - k_3}{m_\gamma} [\tilde{j}^0(k) + \tilde{j}^3(k)] \\ &\simeq \frac{m_\gamma}{2k_3} [\tilde{j}^0(k) + \tilde{j}^3(k)]\end{aligned}\quad (2.138)$$

Therefore, in the massless limit, the longitudinal mode of a massive photon decouples, and the limit $m_\gamma \rightarrow 0$ is continuous.²⁶ Observe that current conservation has been crucial to show the decoupling of the longitudinal state.

The continuity of the limit $m_\gamma \neq 0$, as far as physical observables are concerned, can also be seen from the propagator, as follows. The propagator of the massive photon is found from the quadratic term in the action which, after an integration by parts, reads

$$S = \frac{1}{2} \int d^4x A_\mu [\eta^{\mu\nu} (\partial^2 - m_\gamma^2) - \partial^\mu \partial^\nu] A_\nu \quad (2.139)$$

In momentum space, $\partial_\mu \rightarrow ik_\mu$ and we must therefore invert the matrix

$$M^{\mu\nu}(k) = -\eta^{\mu\nu} (k^2 + m_\gamma^2) + k^\mu k^\nu \quad (2.140)$$

The inversion is easily performed writing M^{-1} in the general form

$$(M^{-1})_{\mu\nu} = a(k)\eta_{\mu\nu} + b(k)k_\mu k_\nu \quad (2.141)$$

and fixing $a(k)$ and $b(k)$ from $M^{\mu\rho}(M^{-1})_{\rho\nu} = \delta_\nu^\mu$. This gives $a =$ ²⁶ To illustrate the continuity of the physics as $m_\gamma \rightarrow 0$, consider the following example. In a thermal ensemble, each degree of freedom contributes to the thermodynamical properties, such as the internal energy or specific heat. For instance, in a relativistic Bose gas at equilibrium, the average energy density is $\rho = (\pi^2/30)gT^4$, where g is the number of polarization degrees of freedom. Therefore there is a finite difference between the energy density for massless photons ($g = 2$) and massive photons ($g = 3$). Apparently, from a measure of thermal properties of such a gas we should be able to decide whether $m_\gamma = 0$ exactly, or $m_\gamma \neq 0$. However, in the limit $m_\gamma \rightarrow 0$, the longitudinal mode takes an infinite time to reach equilibrium with the thermal bath, because its interactions goes to zero. Therefore any experiment lasting a finite time will only be able either to discover a positive mass or to put an upper limit on it, but not to state that $m_\gamma = 0$ exactly.

²⁷ We do not write explicitly the $i\epsilon$ prescription. In all cases, with our choice of signature, it is obtained replacing $k^2 \rightarrow k^2 - i\epsilon$ in the denominators. $-1/(k^2 + m_\gamma^2)$, $b = a/m_\gamma^2$. The propagator is defined by $\tilde{D}_{\mu\nu}(k) = i(M^{-1})_{\mu\nu}$. Then²⁷

$$\tilde{D}_{\mu\nu}(k) = \frac{-i}{k^2 + m_\gamma^2} \left(\eta_{\mu\nu} + \frac{k_\mu k_\nu}{m_\gamma^2} \right) \quad (2.142)$$

The propagator of a massive photon is therefore singular in the limit $m_\gamma \rightarrow 0$. However, the singularity disappears from physical observables, because the amplitudes are proportional to

$$\tilde{j}_\mu^*(k)\tilde{D}^{\mu\nu}(k)\tilde{j}_\nu(k). \quad (2.143)$$

Since the current is conserved, $k^\mu\tilde{j}_\mu(k) = 0$, and therefore the terms $k_\mu k_\nu/m_\gamma^2$ in the propagator give zero. We can therefore simply take as propagator of the massive photon the expression

$$\tilde{D}_{\mu\nu}(k) = \frac{-i}{k^2 + m_\gamma^2} \eta_{\mu\nu} \quad (2.144)$$

and in the limit $m_\gamma \rightarrow 0$ we recover smoothly the massless photon propagator (2.77). The whole procedure required that the massive photon is coupled to a conserved current j^μ .

6.0.42 Massive gravitons

Now that we have understood how to describe a massive photon, we can come back to our original problem, the construction of a field theory for massive gravitons. We start from the Pauli-Fierz action (2.86), and we add a mass term. The most general Lorentz-invariant mass term that one can add to the Pauli-Fierz action is a combination of the two scalars $h_{\mu\nu}h^{\mu\nu}$ and h^2 . Of course both terms break the gauge invariance (2.83) of the massless theory, as it should, since we want to describe the $2j+1=5$ degrees of freedom of a massive spin-2 particle, while we have seen that gauge invariance reduces them to only two. Nevertheless, a generic symmetric tensor $h_{\mu\nu}$ has 10 degrees of freedom so we expect that the appropriate number of conditions to reduce it to five emerges dynamically, similarly to what we have found for massive photons.

Of course, for massive gravitons we must start from the Pauli-Fierz action (2.86) before gauge-fixing, since now there is no gauge symmetry to fix. Adding the mass term, and writing also the source term, we have

$$S_2 = \frac{1}{2} \int d^4x \left[-\partial_\rho h_{\mu\nu} \partial^\rho h^{\mu\nu} + 2\partial_\rho h_{\mu\nu} \partial^\nu h^{\mu\rho} - 2\partial_\nu h^{\mu\nu} \partial_\mu h + \partial^\mu h \partial_\mu h \right. \\ \left. + c_1 h^2 + c_2 h_{\mu\nu} h^{\mu\nu} + \tilde{\kappa} h_{\mu\nu} T^{\mu\nu} \right] \quad (2.145)$$

We denote the coupling of the massive theory by $\tilde{\kappa}$ since, as we will see below, it need not be the same as the coupling κ introduced in the massless case. We can now obtain the propagator of the massive graviton repeating the same steps performed above for the photon, and we discover that, if we add to the Pauli-Fierz Lagrangian a mass term with coefficients c_1 and c_2 arbitrary, the resulting propagator in general has ghosts, i.e. poles with the "wrong" sign of the residue, corresponding to degrees of freedom with negative kinetic energy, which generates an instability of the vacuum. The only combination that does not introduces ghosts turns out to be $h_{\mu\nu}h^{\mu\nu} - h^2$, which is called the Pauli-Fierz mass term. Then, in linearized theory we are led to consider, as the action describing massive gravitons,

$$S_2 = \frac{1}{2} \int d^4x \left[-\partial_\rho h_{\mu\nu} \partial^\rho h^{\mu\nu} + 2\partial_\rho h_{\mu\nu} \partial^\nu h^{\mu\rho} - 2\partial_\nu h^{\mu\nu} \partial_\mu h \right. \\ \left. + \partial^\mu h \partial_\mu h + m_g^2 (h^2 - h_{\mu\nu} h^{\mu\nu}) + \tilde{\kappa} h_{\mu\nu} T^{\mu\nu} \right]$$

Before discussing the propagator obtained from this action, it is instructive to count the number of dynamical degrees of freedom of the theory, to see if they match with the five degrees of freedom of a massive spin-2 particle. We proceed in parallel to what we have done for a massive photon, and we write the equations of motions derived from the action (2.146),

$$\square h^{\mu\nu} - (\partial^\nu \partial_\rho h^{\mu\rho} + \partial^\mu \partial_\rho h^{\nu\rho}) + \eta^{\mu\nu} \partial_\rho \partial_\sigma h^{\rho\sigma} + \partial^\mu \partial^\nu h - \eta^{\mu\nu} \square h \\ = -\frac{\tilde{\kappa}}{2} T^{\mu\nu} + m_g^2 (h^{\mu\nu} - \eta^{\mu\nu} h) \quad (2.147)$$

Introducing on the left-hand side $\bar{h}_{\mu\nu} = h_{\mu\nu} - (1/2)\eta_{\mu\nu}h$, we can also write this as

$$\square \bar{h}^{\mu\nu} + \eta^{\mu\nu} \partial^\rho \partial^\sigma \bar{h}_{\rho\sigma} - \partial_\rho \partial^\nu \bar{h}^{\mu\rho} - \partial_\rho \partial^\mu \bar{h}^{\nu\rho} = -\frac{\tilde{\kappa}}{2} T^{\mu\nu} + m_g^2 (h^{\mu\nu} - \eta^{\mu\nu} h).$$

In this form we immediately see that this equation reduces to eq. (1.17) for $m_g = 0$. We can now apply ∂_μ to both sides. The left-hand side gives zero identically, while on the right-hand side $\partial_\mu T^{\mu\nu} = 0$, since we are working at linear order and then, as we saw in the previous section, $T^{\mu\nu}$ is an external conserved source. Then we get

$$m_g^2 \partial_\mu (h^{\mu\nu} - \eta^{\mu\nu} h) = 0 \quad (2.149)$$

which is analogous to eq. (2.131). Furthermore, taking the trace of both sides of eq. (2.147), we get

$$2\partial_\nu \partial_\mu (h^{\mu\nu} - \eta^{\mu\nu} h) = -\frac{\tilde{\kappa}}{2} T - 3m_g^2 h \quad (2.150)$$

When $m_g \neq 0$, the left-hand side vanishes because of eq. (2.149), and we get

$$-3m_g^2 h = \frac{\tilde{\kappa}}{2} T \quad (2.151)$$

In particular in the vacuum, where $T_{\mu\nu} = 0$, eq. (2.151) gives $h = 0$ and then eq. (2.149) gives $\partial^\mu h_{\mu\nu} = 0$ or, equivalently (since $h = 0$), $\partial^\mu \bar{h}_{\mu\nu} = 0$. So we have five conditions that reduce the 10 components of the symmetric matrix $h_{\mu\nu}$ to the five components expected for a massive spin-2 particle. The Pauli-Fierz mass term is actually the only combination for which there is no term proportional to $\square h$ on the left-hand side of eq. (2.151). For all other choices, such a term does appear, so h becomes a propagating degree of freedom, which furthermore turns out to be ghost-like.

The limit $m_g \rightarrow 0$

As soon as we switch on the interaction, i.e. a non-vanishing $T_{\mu\nu}$, the limit $m_g \rightarrow 0$ becomes quite peculiar. In fact, in the limit $m_g \rightarrow 0$, for generic matter field with a non-vanishing trace of the energy-momentum tensor, eq. (2.151) gives $h \rightarrow \infty$ rather than $h = 0$ as in the free theory. To understand the physical consequences of this result we observe that, inserting eqs. (2.149) and (2.151) (i.e. $\partial_\mu h^{\mu\nu} = \partial^\nu h$ and $h = -\tilde{\kappa}T/(6m_g^2)$) into eq. (2.147) we get

$$\begin{aligned} (\square - m_g^2) h_{\mu\nu} &= -\frac{\tilde{\kappa}}{2} \left(T_{\mu\nu} - \frac{1}{3} \eta_{\mu\nu} T + \frac{1}{3m_g^2} \partial_\mu \partial_\nu T \right) \\ &\equiv S_{\mu\nu} \end{aligned} \quad (2.152)$$

Comparing with eq. (2.92), we see that in the limit $m_g \rightarrow 0$ the left-hand side goes smoothly into the massless limit, but on the right-hand side there are two differences: the coefficient of $\eta^{\mu\nu} T$ is $-1/3$ rather than $-1/2$, and there is an apparently singular term $\sim (1/m_g^2) \partial^\mu \partial^\nu T$. Both find their origin in the fact that, according to eq. (2.151), $m_g^2 h$ stays finite as $m_g \rightarrow 0$.

To see the effect produced by these differences, we can consider the effective matter-matter interaction, which is given by

$$S_{\text{eff}} = \frac{\tilde{\kappa}}{2} \int d^4x h_{\mu\nu}(x) T^{\mu\nu}(x) \quad (2.153)$$

where $h_{\mu\nu}(x)$ is the solution of eq. (2.152),

$$h_{\mu\nu}(x) = \int d^4x' \Delta(x - x') S_{\mu\nu}(x') \quad (2.154)$$

and $\Delta(x - x')$ is a Green's function of the massive Klein-Gordon equation, defined by

$$(\square_x - m_g^2) \Delta(x - x') = \delta^{(4)}(x - x') \quad (2.155)$$

This gives

$$S_{\text{eff}} = \frac{\tilde{\kappa}}{2} \int d^4x d^4x' T^{\mu\nu}(x) \Delta(x-x') S_{\mu\nu}(x') \quad (2.156)$$

Using the conservation of the energy-momentum tensor we can show that the singular term in $S_{\mu\nu}$ drops from this expression. In fact, integrating twice by parts,

$$\begin{aligned} & \int d^4x d^4x' T^{\mu\nu}(x) \Delta(x-x') \frac{\partial}{\partial x'^\mu} \frac{\partial}{\partial x'^\nu} T(x') \\ &= - \int d^4x d^4x' T^{\mu\nu}(x) \left[\frac{\partial}{\partial x'^\mu} \Delta(x-x') \right] \frac{\partial}{\partial x'^\nu} T(x') \\ &= + \int d^4x d^4x' T^{\mu\nu}(x) \left[\frac{\partial}{\partial x^\mu} \Delta(x-x') \right] \frac{\partial}{\partial x'^\nu} T(x') \\ &= - \int d^4x d^4x' \left[\frac{\partial}{\partial x^\mu} T^{\mu\nu}(x) \right] \Delta(x-x') \frac{\partial}{\partial x'^\nu} T(x') \\ &= 0 \end{aligned} \quad (2.157)$$

However, the fact that the coefficient of $\eta^{\mu\nu}T$ is $-1/3$ rather than $-1/2$ gives a genuine difference in the physical amplitude,

$$\begin{aligned} S_{\text{eff}} &= \frac{\tilde{\kappa}}{2} \int d^4x d^4x' T^{\mu\nu}(x) \Delta(x-x') \left[T_{\mu\nu}(x') - \frac{1}{3} \eta_{\mu\nu} T(x') \right] \\ &= \frac{\tilde{\kappa}}{2} \int d^4x d^4x' T^{\mu\nu}(x) \Delta_{\mu\nu\rho\sigma}(x-x'; m_g) T^{\rho\sigma}(x') \end{aligned} \quad (2.158)$$

where

$$\Delta_{\mu\nu\rho\sigma}(x-x'; m_g) = \Delta(x-x') \left[\frac{1}{2} (\eta_{\mu\rho} \eta_{\nu\sigma} + \eta_{\mu\sigma} \eta_{\nu\rho}) - \frac{1}{3} \eta_{\mu\nu} \eta_{\rho\sigma} \right] \quad (2.159)$$

The same result, of course, could have been obtained directly inverting the quadratic part of the action to find the propagator, just as we did for the massless graviton and for the massive photons. Repeating steps analogous to those leading to eq. (2.142), we get

$$\tilde{D}_{\mu\nu\rho\sigma}(k; m_g) = \left[\frac{1}{2} (\Pi_{\mu\rho} \Pi_{\nu\sigma} + \Pi_{\mu\sigma} \Pi_{\nu\rho}) - \frac{1}{3} \Pi_{\mu\nu} \Pi_{\rho\sigma} \right] \left(\frac{-i}{k^2 + m_g^2 - i\epsilon} \right) \quad (2.160)$$

where

$$\Pi_{\mu\nu} \equiv \eta_{\mu\nu} + \frac{k_\mu k_\nu}{m_g^2} \quad (2.161)$$

Since the energy-momentum tensor is conserved, $k_\mu \tilde{T}^{\mu\nu}(k) = 0$ and, when contracted with $T^{\mu\nu}(-k)T^{\rho\sigma}(k)$, the above propagator is equivalent to

$$\tilde{D}_{\mu\nu\rho\sigma}(k; m_g) = \left[\frac{1}{2} (\eta_{\mu\rho} \eta_{\nu\sigma} + \eta_{\mu\sigma} \eta_{\nu\rho}) - \frac{1}{3} \eta_{\mu\nu} \eta_{\rho\sigma} \right] \left(\frac{-i}{k^2 + m_g^2 - i\epsilon} \right)$$

Comparing eqs. (2.160) or (2.162) with the massless graviton propagator found in eq. (2.97) we see the following. The terms $k_\mu k_\nu / m_g^2$,

²⁸ Actually, it was discovered independently by Iwasaki (1970), van Dam and Veltman (1970) and Zakharov (1970). that are singular in the massless limit, give a vanishing contribution

to the physical amplitudes, because $h_{\mu\nu}$ is coupled to a conserved energymomentum tensor. This is completely analogous to the situation for a massive photon, where the terms $k_\mu k_\nu / m_\gamma^2$ disappeared because the photon is coupled to a conserved current. However, there is difference in the numerical coefficient in front of $\eta_{\mu\nu}\eta_{\rho\sigma}$ and, unlikely the case of the photon, the propagator (2.162) does not reduce to the massless propagator as $m_g \rightarrow 0$. Writing

$$\begin{aligned} \frac{1}{2}(\eta_{\mu\rho}\eta_{\nu\sigma} + \eta_{\mu\sigma}\eta_{\nu\rho}) - \frac{1}{3}\eta_{\mu\nu}\eta_{\rho\sigma} &= \frac{1}{2}(\eta_{\mu\rho}\eta_{\nu\sigma} + \eta_{\mu\sigma}\eta_{\nu\rho} - \eta_{\mu\nu}\eta_{\rho\sigma}) \\ &\quad + \frac{1}{6}\eta_{\mu\nu}\eta_{\rho\sigma} \end{aligned} \quad (2.163)$$

we see that

$$\begin{aligned} \lim_{m_g \rightarrow 0} T^{\mu\nu}(-k) \tilde{D}_{\mu\nu\rho\sigma}(k; m_g) T^{\rho\sigma}(k) &= T^{\mu\nu}(-k) \tilde{D}_{\mu\nu\rho\sigma}(k) T^{\rho\sigma}(k) \\ &\quad + \frac{1}{6} T(-k) \frac{-i}{k^2} T(k) \end{aligned} \quad (2.164)$$

where $\tilde{D}_{\mu\nu\rho\sigma}(k)$ is the propagator computed directly in the massless theory, eq. (2.97). The second term on the right-hand side corresponds to the exchange of a massless scalar particle, coupled to the trace of the energy-momentum tensor, and it is because of this additional term that the limit $m_g \rightarrow 0$ of the massive theory does not reproduce the result obtained in the massless theory. This is usually referred to as the van Dam-Veltman-Zakharov (vDVZ) discontinuity. 28

Having identified the problem in the existence of an unexpected scalar degree of freedom, we can now understand what happened. Recall that a massive spin- j particles has $2j + 1$ degrees of freedom, while a massless particle with quantum number j has only two degrees of freedom if $j > 0$ (and if we demand that it is also a representation of parity, beside of Poincar transformations), and one degree of freedom if $j = 0$. For a massive photon, we saw that the three degrees of freedom, in the massless limit, decompose into the two transverse degrees of freedom of a massless photon, which have helicities ± 1 , plus the longitudinal component. The latter, in the limit $m_\gamma \rightarrow 0$, becomes a scalar particle and decouples, since its coupling becomes proportional to $\partial_\mu j^\mu$, which vanishes. The fact that the coupling is to $\partial_\mu j^\mu$ is dictated by the fact that this is the only scalar that we can make with j^μ and with derivatives, which is linear in the current. The decoupling of the spurious scalar mode is therefore ensured by current conservation.

Similarly, in the limit $m_g \rightarrow 0$, the five degrees of freedom of the massive graviton decompose into two states with helicities ± 2 , which make up a massless graviton, two states with helicities ± 1 , often termed the graviphoton, and a scalar, called the graviscalar. The graviphoton, being a vector, must be coupled to a four-vector made with $T_{\mu\nu}$ and possibly with derivatives, and linear in $T_{\mu\nu}$. The only possibility is $\partial_\mu T^{\mu\nu}$. However, this quantity vanishes and therefore the graviphoton

decouples. The graviscalar, on the other hand, can couple to the trace of the energy-momentum tensor. This is in general non-zero, and therefore the graviscalar does not decouple. It is, in fact, responsible for the additional term in eq. (2.164) and therefore for the vDVZ discontinuity.

We can now compare some predictions of the massless and massive theory. Consider first the Newtonian potential in the non-relativistic limit. We found in eq. (2.98) that

$$V(\mathbf{x}) = -i \frac{\kappa^2}{4} m_1 m_2 D_{0000}(\mathbf{x}) \quad (2.165)$$

In the massless theory, using eq. (2.97), we saw that

$$\tilde{D}_{0000}(k) = \frac{1-i}{2} \frac{1}{k^2} \quad (2.166)$$

This gives $D_{0000}(r) = -i/(8\pi r)$, so

$$V(\mathbf{x}) = -\frac{\kappa^2}{32\pi} \frac{m_1 m_2}{r} \quad (2.167)$$

and we recovered the Newtonian potential setting $\kappa^2 = 32\pi G$. In the massive theory, instead, we see from eq. (2.162) that

$$\tilde{D}_{0000}(k) = \frac{2}{3} \frac{-i}{k^2 + m_g^2} \quad (2.168)$$

and we get

$$V(\mathbf{x}) = -\frac{4}{3} \left(\frac{\tilde{\kappa}^2}{32\pi} \right) \frac{m_1 m_2}{r} \exp\{-m_g r\} \quad (2.169)$$

The Yukawa potential was of course expected because a massive particle mediates a short-range force. The result however differs also by an overall factor $4/3$ from the massless result, and the difference is due to the additional attractive contribution of the graviscalar.

As far as the Newtonian limit is concerned, one can simply reabsorb this difference setting, in the massive theory,

$$\frac{4}{3} \left(\frac{\tilde{\kappa}^2}{32\pi} \right) = G \quad (2.170)$$

Thus, at $m_g r \ll 1$, the correct Newtonian potential is obtained, at the price that the coupling $\tilde{\kappa}^2$ of the massive theory is smaller than the coupling κ^2 of the massless theory by a factor $3/4$. The problem however comes when we consider the predictions of the massive theory in the relativistic regime, in particular when we compute the deflection of light by a massive object. In the massless theory, the deflection angle of light skimming the surface of the Sun (and first detected, during a total eclipse, by Eddington in 1919), is

$$\Delta\theta = \frac{4GM_\odot}{R_\odot} \quad (2.171)$$

²⁹ We follow here the treatment of Deffayet, Dvali, Gabadadze and Vainshtein (2002), to whom we refer for more details. where $M_\odot$ and $R_\odot$ are the mass and radius of the Sun. Since the energymomentum tensor of the electromagnetic field is traceless, when we repeat the computation in the massive theory, the additional term proportional to $T(-k)(-i/k^2)T(k)$ in eq. (2.164) vanishes. This means that the result for the deflection of light in the massive theory, in the limit $m_g \rightarrow 0$, is the same as in the massless theory except that, instead of κ^2 , we have $\tilde{\kappa}^2$. Since $\tilde{\kappa}^2 = (3/4)\kappa^2$, the prediction for light bending in the limit $m_g \rightarrow 0$ of the massive theory is smaller by a factor $3/4$ than the prediction of the massless theory, i.e. is $\Delta\theta = 3GM_\odot/R_\odot$. Nowadays the experimental precision on the measurement of the bending of light by the Sun is better than one part in 10^4 , and confirms the prediction of the massless theory. Apparently, one then arrives at the amazing conclusion that the mass of the graviton must be exactly zero.

A loophole in the above argument was found by Vainshtein, considering the validity of the linearized approximation in the massive theory. To study the expansion in $h_{\mu\nu}$ systematically, we should start from the full Einstein action plus the Pauli-Fierz mass term,

$$S = \frac{1}{16\pi G} \int d^4x \sqrt{-g} \left[R + \frac{1}{4} m_g^2 (h^2 - h_{\mu\nu} h^{\mu\nu} + O(h_{\mu\nu}^3)) \right] \quad (2.172)$$

We then expand $g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}$ where $\bar{g}_{\mu\nu}$ is the appropriate background metric and, to have the canonical normalization for $h_{\mu\nu}$, we finally rescale $h_{\mu\nu} \rightarrow (32\pi G)^{1/2} h_{\mu\nu} = \kappa h_{\mu\nu}$. Bosonic matter fields, as usual, are coupled replacing in the matter action all ordinary derivative with covariant derivatives. Actually, one can use any form for the mass term that, in the linearized limit, reduces to the Pauli-Fierz mass term (in particular, we can decide to raise and lower the indices in the mass term with $\eta_{\mu\nu}$ or with $g_{\mu\nu}$. The difference shows up only at cubic and higher-order terms, that are not fixed anyway), so we also included the possibility of non-linear corrections to the Pauli-Fierz mass term.

To compute the gravitational scattering by a fixed heavy mass M , we expand around a metric $\bar{g}_{\mu\nu}$ which is a generalization of the Schwarzschild metric generated by the heavy mass M , so it is computed from the action (2.172), which includes the graviton mass. Naively one would expect that, if $m_g \rightarrow 0$, this metric goes smoothly into the usual Schwarzschild metric computed in standard general relativity, with massless graviton, for all values of r . However, the explicit computation shows that this is not true. We can search for the metric generated by an heavy mass in the theory (2.172), writing ²⁹

$$ds^2 = -e^{\nu(\rho)} dt^2 + e^{\sigma(\rho)} d\rho^2 + e^{\mu(\rho)} \rho^2 (d\theta^2 + \sin^2 \theta d\phi^2) \quad (2.173)$$

When the graviton is massless, the function $\mu(\rho)$ can be set to zero using coordinate invariance. In the theory (2.172), however, the reparametrization invariance is broken, and $\mu(\rho)$ must be kept. One then performs the substitutions

$$r \equiv \rho e^{\mu/2}, \quad e^\lambda \equiv \left(1 + \frac{\rho}{2} \frac{d\mu}{d\rho}\right)^{-2} e^{\sigma-\mu} \quad (2.174)$$

The standard Schwarzschild solution of the massless theory corresponds to

$$\begin{aligned} \nu(r) = -\lambda(r) &= \log \left(1 - \frac{R_S}{r}\right) \\ &= -\frac{R_S}{r} - \frac{1}{2} \left(\frac{R_S}{r}\right)^2 + \dots \end{aligned} \quad (2.175)$$

$$\mu(r) = 0 \quad (2.176)$$

In the theory with massive graviton one rather finds, up to next-to-leading order in G (therefore in R_S),

$$\nu(r) = -\frac{R_S}{r} \left[1 + O\left(\frac{R_S}{m_g^4 r^5}\right) + \dots\right] \quad (2.177)$$

$$\lambda(r) = \frac{1}{2} \frac{R_S}{r} \left[1 + O\left(\frac{R_S}{m_g^4 r^5}\right) + \dots\right] \quad (2.178)$$

$$\mu(r) = -\frac{1}{2} \frac{R_S}{m_g^2 r^3} \left[1 + O\left(\frac{R_S}{m_g^4 r^5}\right) + \dots\right] \quad (2.179)$$

If we limit ourselves to leading order, we observe that $\lambda(r)$ is smaller by a factor $1/2$ compared to the result in the massless theory. This is the origin of the vDVZ discontinuity. However, the surprise comes looking at the corrections, since they blow up if $m_g \rightarrow 0$! In other words, in the massive theory the linearized expansion becomes invalid if we send $m_g \rightarrow 0$ at fixed r . This does not mean that linearized theory is completely useless: if we define the Vainshtein radius R_V ,

$$R_V = (R_S \lambda_g^4)^{1/5} \quad (2.180)$$

where $\lambda_g = 1/m_g$, we see that the corrections are proportional to $(R_V/r)^5$. Therefore linearized theory is valid at $r \gg R_V$. We take $\lambda_g > 200\text{kpc}$, in agreement with eq. (2.125) (with $h_0 \simeq 0.7$), and we consider the scattering of light from the Sun, which has $R_S \sim 3\text{ km}$. Then we get $R_V > 40\text{pc}$, i.e. R_V is at least 10^7 times larger than the Earth-Sun distance. Therefore, the Newtonian potential found in eq. (2.169), and the result for the light deflection in the massive theory discussed below eq. (2.169), are simply not applicable at the solar system scale.

On the other hand, in the opposite limit $r \ll R_V$, it is possible to find a consistent expansion of the Schwarzschild solution in powers of m_g that, to lowest order, reproduces the Schwarzschild solution of the massless theory, of the form ³⁰

$$\nu(r) = -\frac{R_S}{r} + O\left(m_g^2 \sqrt{R_S r^3}\right) \quad (2.181)$$

$$\lambda(r) = +\frac{R_S}{r} + O\left(m_g^2 \sqrt{R_S r^3}\right) \quad (2.182)$$

$$\mu(r) = \sqrt{\frac{8R_S}{13r}} + O\left(m_g^2 r^2\right) \quad (2.183)$$

³¹ For example, consider the function $f(\epsilon, x) = e^{-\epsilon/x}$. If we expand it in powers of ϵ at fixed x we get $f(\epsilon, x) = 1 - (\epsilon/x) + (1/2)(\epsilon/x)^2 + \dots$. Of course, this expansion is not suitable for studying the limit $x \rightarrow 0$, since the various terms are more and more singular. However, if one knows the full resummed expression $e^{-\epsilon/x}$, one realizes that, for $\epsilon > 0$, the limit $x \rightarrow 0^+$ exists, and is in this case zero. In our case the role of ϵ is played by R_S and the role of x by m_g .

Therefore, at $r \ll R_V$, there is no mass discontinuity. To prove that there is no mass discontinuity altogether, one should be able to resum the whole perturbative expansion (2.177-2.179), which is valid at $r \gg R_V$, and it is in principle possible that, in the resummed expression, there is no singularity as $m_g \rightarrow 0$.³¹ If one can show that such a resummed solution, as we approach $r \sim R_V$, matches smoothly the solution (2.181)(2.183), which is valid at $r \ll R_V$, we have constructed a solution with a smooth limit $m_g \rightarrow 0$, valid for all r . Observe also that, since the linearized theory does not apply at the solar system scale, there is no need to require $\tilde{\kappa}^2 = (3/4)\kappa^2$. On the contrary, since at $r \ll R_V$ the expansion in m_g reproduces smoothly the massless limit, we must choose $\tilde{\kappa} = \kappa$, and all the results of the massive theory, from the Newtonian potential to the light deflection, go smoothly into those of the massless theory. At $r \gg R_V$, instead, where linearized theory can be trusted, we get a gravitational potential

$$V(r) = -\frac{4}{3} \frac{Gm_1 m_2}{r} \exp\{-m_g r\} \quad (2.184)$$

since we have fixed $\tilde{\kappa} = \kappa$. In conclusion, in this scenario, inside the Vainstein radius a tiny graviton mass has negligible effects, while at $r \gg R_V$ the graviscalar becomes effective, and gives a further attractive contribution to the gravitational potential.

To prove that this what actually happens requires to show that the solutions found in the regimes $r \ll R_V$ and $r \gg R_V$ do match. This is a non-trivial problem, because the solution (2.181)-(2.183) is obtained from an expansion in m_g while the solution (2.177)-(2.179) is non-analytic in m_g , since we have seen that it diverges as $m_g \rightarrow 0$. Conversely, the solution (2.177)-(2.179) is obtained performing an expansion in G , while the solution (2.181)-(2.183) is

³⁰ The existence of such a solution depends on the specific form of the nonlinear corrections to the Pauli-Fierz mass term, see Damour, Kogan and Papazoglou (2003).

non-analytic in G , since $\mu(r) \sim \sqrt{R_S} \sim \sqrt{G}$. Put it differently, the difficulty of the problem is that, as we approach R_V from the large distance region, the graviscalar becomes strongly coupled, and perturbation theory breaks down.

Numerical studies of the inward continuation of asymptotically flat solutions indicate that, for small m_g , they end up in a singularity at finite radius, rather than matching a continuous solution inside the Vainshtein radius. It is possible however that the matching takes place not for asymptotically flat solution but for asymptotically De Sitter solution. This would be physically acceptable, given the experimental evidence for a cosmological constant, and also because for m_g sufficiently small the value of R_V can be larger than the Hubble radius, in which case the form of the solution at $r \gg R_V$ is not physically relevant.

A further complication is that, in curved space, the trace degree of freedom h , which in linearized theory is eliminated through eq. (2.151), becomes dynamical again, so we have six degrees of freedom rather than five, and furthermore it is a ghost. In fact, in the theory with action (2.172), i.e. full Einstein gravity supplemented by a mass term, the linearized equation of motion (2.147) is replaced by

$$G_{\mu\nu} = \frac{\tilde{\kappa}}{4} T_{\mu\nu} - \frac{1}{2} m_g^2 [a h_{\mu\nu} + b h \eta_{\mu\nu} + O(h_{\mu\nu}^2)] \quad (2.185)$$

where $G_{\mu\nu}$ is the full Einstein tensor, rather than its linearization that appears in eq. (2.147), and we also allowed for a generic mass term, including higher-order corrections (the Pauli-Fierz mass term corresponds to $a = -b = 1$). Using the Bianchi identity $D^\mu G_{\mu\nu} = 0$ as well as the covariant conservation of the energy-momentum tensor, we get the four conditions

$$m_g^2 D^\mu [a h_{\mu\nu} + b h \eta_{\mu\nu} + O(h_{\mu\nu}^2)] = 0 \quad (2.186)$$

which replace their linearized version (2.149) and again allow us to eliminate four degrees of freedom. The elimination of these degrees of freedom is therefore a consequence of the Bianchi identity, or, equivalently, of the invariance of the Einstein-Hilbert action under diffeomorphisms. On the other hand, the elimination of h in linearized theory, eq. (2.151), is not ensured by any symmetry, but is a consequence of the fine-tuning in the mass term that leads to the Pauli-Fierz combination. Then, this constraint does not survive in curved space, and h becomes dynamical again and it can be shown to be ghost-like, see the Further Reading section for discussions.

Thus, presently the issue of the consistency of a field theory of massive gravitons is not settled. If, in some form, a continuous solution indeed exists, then the limit $m_g \rightarrow 0$ is smooth and it make sense to deform Einstein gravity by adding a small mass term for the graviton, with a value bounded experimentally by (2.126). Otherwise, one should accept the (rather odd) conclusion that the graviton mass must be identically zero. We finally observe that it is possible to add mass terms that break Lorentz invariance (which indeed emerge quite naturally if one breaks spontaneously the diffeomorphism invariance of general relativity), and in this case there are neither ghosts nor the vDVZ discontinuity. Again, we refer the reader to the Further Reading section.

6.0.43 2.4 Solved problems

6.0.44 Problem 2.1. The helicity of gravitons

We have seen that, for a given propagation direction $\hat{\mathbf{n}}$, a GW is described by a 2×2 matrix in the plane orthogonal to $\hat{\mathbf{n}}$, with matrix elements given in terms of the amplitudes h_+ and $h_\times$ of the two polarization. The Lorentz transformations that leave invariant the propagation direction $\hat{\mathbf{n}}$ are the rotations around the $\hat{\mathbf{n}}$ axis and the boosts in the $\hat{\mathbf{n}}$ direction. Under these operations h_+ and $h_\times$ will therefore transform between themselves. In this problem, we compute

explicitly the transformation of h_+ and $h_\times$ under rotations around the $\hat{\mathbf{n}}$ axis and under boosts in the $\hat{\mathbf{n}}$ direction.

In general, under a Lorentz transformation $x^\mu \rightarrow x'^\mu = \Lambda^\mu{}_\nu x^\nu$, a tensor $h_{\mu\nu}$ transforms as

$$h_{\mu\nu}(x) \rightarrow h'_{\mu\nu}(x') = \Lambda_\mu{}^\rho \Lambda_\nu{}^\sigma h_{\rho\sigma}(x). \quad (2.187)$$

Choosing $\hat{\mathbf{n}} = \hat{\mathbf{z}}$, a rotation around the z axis and a boost along z are written respectively as

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \psi & -\sin \psi & 0 \\ 0 & \sin \psi & \cos \psi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{rotation}) \quad (2.188)$$

and

$$\Lambda = \begin{pmatrix} \cosh \eta & 0 & 0 & -\sinh \eta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh \eta & 0 & 0 & \cosh \eta \end{pmatrix} \quad (\text{boost}) \quad (2.189)$$

where Λ denote the matrix whose elements are $\Lambda_\mu{}^\nu$; ψ is the rotation angle and η is related to the velocity v of the boost by $v = \tanh \eta$. Writing eq. (2.187) in the TT gauge we have

$$(h_{ab}^{\text{TT}})'(x') = \begin{pmatrix} h'_+ & h'_\times \\ h'_\times & -h'_+ \end{pmatrix}_{ab} e^{ikx} \quad (2.190)$$

where

$$\begin{pmatrix} h'_+ & h'_\times \\ h'_\times & -h'_+ \end{pmatrix}_{ab} = \Lambda_a{}^c \Lambda_b{}^d \begin{pmatrix} h_+ & h_\times \\ h_\times & -h_+ \end{pmatrix}_{cd} \quad (2.191)$$

and a, b take the value 1, 2. Since $kx = k'x'$, where k' is the four-momentum in the new frame, we can also rewrite eq. (2.190) as

$$(h_{ab}^{\text{TT}})'(x') = \begin{pmatrix} h'_+ & h'_\times \\ h'_\times & -h'_+ \end{pmatrix}_{ab} e^{ik'x'} \quad (2.192)$$

or, since x' is generic,

$$(h_{ab}^{\text{TT}})'(x) = \begin{pmatrix} h'_+ & h'_\times \\ h'_\times & -h'_+ \end{pmatrix}_{ab} e^{ik'x} \quad (2.193)$$

Using eq. (2.191), with $\Lambda_a{}^c$ given by the 2×2 submatrix made by the second and third rows and columns of eq. (2.188), and performing the matrix multiplication, we get

$$\begin{aligned} h'_+ &= h_+ \cos 2\psi - h_\times \sin 2\psi \\ h'_\times &= h_+ \sin 2\psi + h_\times \cos 2\psi \end{aligned} \quad (2.194)$$

Under boosts, the matrix $\Lambda_a{}^c$ (i.e. the 2×2 submatrix made by the second and third rows and columns of eq. (2.189)) is just the 2×2 identity matrix, so $h'_+ = h_+$ and $h'_\times = h_\times$. The GW amplitudes h_+ and $h_\times$ are therefore invariant under boosts.

From eq. (2.194) we see that, under rotations around the z axis, the combinations $h_\times \pm ih_+$ transform as

$$(h_{\times} \pm ih_{+}) \rightarrow e^{\mp 2i\psi} (h_{\times} \pm ih_{+}) \quad (2.195)$$

To understand the meaning of this transformation law, we recall some basic results from the theory of representations of the Poincar group (see, e.g. Maggiore 2005, Chapter 2). The Poincar group has two types of physically interesting representations:

- massive representation, which are labeled by the mass m , with $-P_{\mu}P^{\mu} = m^2 > 0$ (where P^{μ} is the four-momentum) and by the spin j , which (in units of $\hbar$) can take integer or half-integer values, $j = 0, 1/2, 1, \dots$. The representation with spin j has dimension $2j + 1$. Physically, this follows from the fact that for a massive particle exists the rest frame, and in the rest frame the component along the z axis of the spin, for a particle with spin j , can take the $2j + 1$ possible values $j_z = -j, -j + 1, \dots, j$. So, in particular, a massive spin-1 particle has three degrees of freedom and a massive spin-2 particle has five degrees of freedom.
- massless representation, which are characterized by $P_{\mu}P^{\mu} = 0$ and by a quantum number j , which again can be integer or half-integer. For massless particles the rest frame does not exist and the previous argument about the existence of $2j + 1$ states does not go through. Rather, these representations are one-dimensional, and are characterized by a definite value of the helicity, which is defined as the projection of the total angular momentum on the direction of motion,³²

$$h = \mathbf{J} \cdot \hat{\mathbf{n}}. \quad (2.196)$$

The total angular momentum is the sum of orbital and spin angular momenta, $\mathbf{J} = \mathbf{L} + \mathbf{S}$. Of course $\mathbf{L} \cdot \hat{\mathbf{n}} = (\mathbf{x} \times \mathbf{p}) \cdot \hat{\mathbf{n}} = 0$ since $\mathbf{p} = |\mathbf{p}|\hat{\mathbf{n}}$, and therefore the helicity is equal to the projection of the spin on the direction of motion, $h = \mathbf{S} \cdot \hat{\mathbf{n}}$.

Under a rotation by an angle ψ around the direction of motion a helicity eigenstate $|h\rangle$ transforms as

$$|h\rangle \rightarrow e^{ih\psi} |h\rangle \quad (2.197)$$

Massless representations can have either $h = +j$ or $h = -j$. Each of these possibilities, separately, provides a one-dimensional representation of the Poincar group: the states with $h = \pm j$ do not mix between them under (proper) spatial rotations, boosts nor translations. However, under a parity transformation $\hat{\mathbf{n}}$ changes sign, while the angular momentum is a pseudovector and is unchanged. Therefore the helicity is a pseudoscalar, i.e. it changes sign under parity, $h \rightarrow -h$. For this reason, in a theory which conserves parity (like gravity or electromagnetism) it is more convenient to define particles as representations of the Poincar group and of parity, that is, to assemble the two Poincar representations with $h = \pm j$ and to consider them as two polarization states of the same particle. The photon is then defined as a massless particle with two helicity states $h = \pm 1$, while the graviton is defined as a massless particle with two helicity states $h = \pm 2$. On the contrary, since weak interactions violate parity, in the limit in which the neutrinos can be taken as massless (there is nowadays evidence for a small neutrino mass from oscillation experiments), we reserve the name neutrino to the one-dimensional massless Poincar representation with $h = -1/2$, while the antineutrino is defined as the representation with $h = +1/2$.³³

In the light of these results, we see that eq. (2.195), together with the fact that h_{ij}^{TT} satisfies the massless Klein-Gordon equation $\square h_{ij}^{\text{TT}} = 0$, means that the quanta of the gravitational field are massless particles, and the combinations $h_{\times} \mp ih_{+}$ are the helicity eigenstates and have helicities ± 2 , respectively.³² The symbol h is traditionally used for the helicity, and of course

it should not be confused with the GW amplitudes h_+ and $h_\times$.³⁴ In general, we can define parity either changing the sign of vectors with respect to a fixed reference frame, or reversing the orientation of the axes of the reference frame while keeping the vectors fixed. Here we adopt the latter point of view.

6.0.45 Problem 2.2. Angular momentum and parity of graviton states

In this problem we examine the possible angular momentum states of the graviton. Let us first recall how such an analysis works for the photon, following Landau and Lifshitz, Vol. IV (1982), Section 6. After choosing the radiation gauge $A_0 = 0$ and $\nabla \cdot \mathbf{A} = 0$, a photon is described by a vector $\mathbf{A}(\mathbf{x})$ or, in momentum space, by $\tilde{\mathbf{A}}(\mathbf{k})$, subject to the transversality condition $\mathbf{k} \cdot \tilde{\mathbf{A}}(\mathbf{k}) = 0$. Let us at first neglect the transversality condition. The vector character of $\tilde{\mathbf{A}}(\mathbf{k})$ then corresponds to spin $s = 1$, and the total angular momentum j of a photon is given by the combination of $s = 1$ and of the orbital angular momentum l , with the usual composition rule of angular momenta in quantum mechanics. This means that a state with $j = 0$ can be obtained in only one possible way, i.e. combining the spin $s = 1$ with $l = 1$, while there are three states for each $j \neq 0$, which are obtained with $l = j, j \pm 1$ (for the purpose of this counting we consider a state with momentum j as one single state, regardless of the $2j + 1$ possible components of j_z).

To understand what is the parity of these states it is convenient to write explicitly their wavefunction. Just as the angular dependence of a scalar function can be expanded in terms of the (scalar) spherical harmonics $Y_{lm}(\theta, \phi)$, the angular dependence of a vector function $\mathbf{A}(\mathbf{x})$ can be expressed in vector spherical harmonics. As we will discuss in more detail in Section 3.5.2, see in particular eq. (3.247), these can be written as

$$\mathbf{Y}_{jj_z}^l(\theta, \phi) = \sum_{l_z=-l}^l \sum_{s_z=0, \pm 1} \langle 1s_z l_z | jj_z \rangle Y_{l_z}(\theta, \phi) \boldsymbol{\xi}^{(s_z)} \quad (2.198)$$

where $\boldsymbol{\xi}^{(s_z)}$ is the wavefunction of a spin-1 particle with a given value s_z of the projection of the spin on the z axis, and $\langle 1s_z l_z | jj_z \rangle$ are the Clebsch-Gordan coefficient necessary to combine a spin state $|ss_z\rangle$ with $s = 1$ and an orbital angular momentum state $|ll_z\rangle$, so to obtain a total angular momentum $|jj_z\rangle$. In terms of the unit vectors $\mathbf{e}_x, \mathbf{e}_y$ and $\mathbf{e}_z$ of a Cartesian reference frame, the explicit form of the spin wavefunction is

$$\boldsymbol{\xi}^{(\pm 1)} = \mp \frac{1}{\sqrt{2}} (\mathbf{e}_x \pm i\mathbf{e}_y), \quad \boldsymbol{\xi}^{(0)} = \mathbf{e}_z \quad (2.199)$$

Consider a parity transformation, defined so that it changes the sign of the orientation of the axes of the reference frame, $\mathbf{e}_i \rightarrow -\mathbf{e}_i$.³⁴ Under this transformation, the state (2.198) picks a factor $P = (-1)^{l+1}$, where the $(-1)^l$ comes from the transformation of the scalar spherical harmonics and the further minus sign comes from the spin wavefunction. Thus, we can summarize as follows the possible states before imposing the condition $\mathbf{k} \cdot \tilde{\mathbf{A}}(\mathbf{k}) = 0$,

$$\begin{aligned} j = 0 : & \quad \text{one state, } (l = 1, P = +) \\ j = 1 : & \quad \text{three states, } (l = 0, P = -), (l = 1, P = +), (l = 2, P = -), \\ j = 2 : & \quad \text{three states, } (l = 1, P = +), (l = 2, P = -), (l = 3, P = +), \end{aligned} \quad (2.200)$$

³⁴ The fact that, for each massless representation with a given helicity, there is a corresponding representation with the opposite helicity is a consequence of the CPT symmetry, which is present in any Lorentz-invariant quantum field theory with a hermitian Hamiltonian.

and similarly we have three states for all higher values of j . We now impose the transversality condition $\mathbf{k} \cdot \tilde{\mathbf{A}}(\mathbf{k}) = 0$. This means that we remove a longitudinal state of the form $\tilde{\mathbf{A}}(\mathbf{k}) = \phi(\mathbf{k})\mathbf{k}$. The number of states of this form is therefore the same as the number of states of a scalar particles with wavefunction $\phi(\mathbf{k})$ (or, equivalently, of the scalar degree of freedom described by $\nabla \cdot \mathbf{A}$). When we develop ϕ in spherical harmonics, the total angular momentum j of

such a state is equal to the order l of the spherical harmonic, and its parity is $P = (-1)^j$. Thus, at each level j we must remove from eq. (2.200) one spurious state, with $P = (-1)^j$. Therefore at level $j = 0$ we end up with no physical states, while at all higher levels we end up with two physical states of opposite parity. This shows that for the photon there can be no monopole radiation, because there is no physical photon state with $j = 0$, while for all other values of j we have two physical states. For instance, the states with $j = 1$ correspond to an electric dipole photon ($P = -$) and a magnetic dipole photon ($P = +$).

Having understood the argument for the photon, we can adapt it to the graviton. The graviton is described by a 2×2 traceless symmetric tensor $\tilde{h}_{ij}(\mathbf{k})$ subject to the transversality condition $k^i \tilde{h}_{ij}(\mathbf{k}) = 0$. Again, we neglect at first the transversality condition. A symmetric traceless tensor corresponds to spin $s = 2$, while the parity on a true tensor is $P = (-1)^l$. Then, combining the orbital angular momentum with the spin $s = 2$, we have the following states

$$\begin{aligned} j = 0 : & \quad \text{one state,} & (l = 2, P = +) \\ j = 1 : & \quad \text{three states,} & (l = 1, P = -), (l = 2, P = +), (l = 3, P = -) \\ j = 2 : & \quad \text{five states,} & (l = 0, P = +), (l = 1, P = -), (l = 2, P = +) \\ & & (l = 3, P = -), (l = 4, P = +). \end{aligned} \quad (2.201)$$

Similarly for all higher j there are five states, two with $P = +$, two with $P = -$ and one more with $P = (-1)^j$. We now impose the transversality condition. The most general traceless symmetric tensor which does not satisfy the transversality condition has the form

$$\tilde{h}_{ij}(\mathbf{k}) = a_i(\mathbf{k})k_j + a_j(\mathbf{k})k_i + b(\mathbf{k}) \left(k_i k_j - \frac{1}{3} \delta_{ij} |\mathbf{k}|^2 \right) \quad (2.202)$$

with $k^i a_i(\mathbf{k}) = 0$, in order to respect the condition of zero trace. The most general spurious state is therefore parametrized by a scalar b and by a transverse vector a_i . Exactly as with the scalar ϕ found above, expanding b in spherical harmonics we have one state for each j , with parity $P = (-1)^j$. This eliminates the state with $j = 0$ in eq. (2.201), while at level $j = 1$ it leaves us with one state with $P = +1$ and one with $P = -1$, and at all higher j levels we are left with two states with $P = +1$ and two with $P = -1$. Finally, we must remove the spurious states described by $a_i(\mathbf{k})$. However this is a vector, transverse to $\mathbf{k}$, and therefore its states are the same as the photon states discussed above. There is no state at $j = 0$, and two states, with opposite parity, at all other j level. This remove the two states which were left at $j = 1$, and leaves us with two states, with opposite parity, at all higher levels. In conclusion, for the graviton,

$$\begin{aligned} j = 0 : & \quad \text{no state,} \\ j = 1 : & \quad \text{no state,} \\ j = 2, 3, \dots : & \quad \text{two states, one with } P = +, \text{ one with } P = -. \end{aligned} \quad (2.203)$$

Therefore, for gravitational wave there can be no monopole nor dipole radiation, since these would correspond to gravitons with $j = 0$ and $j = 1$, respectively. We will come back to the multipole expansion of gravitational waves in Section 3.5.2, where we will show how to express the two states allowed for $j \geq 2$ in terms of tensor spherical harmonics, and we will verify again that states with $j = 0$ or $j = 1$ are not allowed.

6.0.46 Further reading

- For the quantum field-theoretical approach to gravitation see the Feynman Lectures on Gravitation by Feynman, Morinigo, and Wagner (1995) (which collects lectures given by Feynman in 1962-63), and also DeWitt (1967) and Veltman (1976). For explicit computations of graviton-graviton scattering see Grisar, van Nieuwenhuizen and Wu (1975).
- The possibility of deriving Einstein equation from an iteration of linearized theory is discussed, among others, by Gupta (1954), Kraichnan (1955), Feynman, Morinigo, and Wagner (1995), and Ogievetsky and Polubarinov (1965). An explicit and elegant iteration leading from the equations of motion of linearized theory to the full Einstein equations was performed by Deser (1970) using a first order Palatini formalism. The ambiguity concerning boundary terms is discussed by Padmanabhan (2004).
- Phenomenological limits on the graviton mass are discussed by Goldhaber and Nieto (1974). The discontinuity as the graviton mass goes to zero was found by Iwasaki (1970), van Dam and Veltman (1970) and Zakharov (1970). Massive gravitons have been further discussed by Boulware and Deser (1972). The fact that linearized theory becomes singular as $m_g \rightarrow 0$ was discovered by Vainshtein (1972). The radiation of massive gravitons in linearized theory is discussed by van Nieuwenhuizen (1973). Discussions of the fate of the discontinuity are given in Deffayet, Dvali, Gabadadze and Vainshtein (2002) and in Arkani-Hamed, Georgi and Schwartz (2003). The difficulty of performing the matching to an asymptotically flat solution, and the possibility of matching to a De Sitter solution, is discussed in Damour, Kogan and Papazoglou (2003). The fact that beyond linearized theory the trace h becomes a ghost is discussed by Boulware and Deser (1972) and, in full generality, by Creminelli, Nicolis, Papucci and Trincherini (2005).
- Lorentz-violating mass terms for $h_{\mu\nu}$ are discussed in Arkani-Hamed, Cheng, Luty and Mukohyama (2004), Rubakov (2004) and Dubovsky, Tinyakov and Tkachev (2005). In this case the mass of the scalar perturbations can be zero while the mass of the graviton h_{ij}^{TT} can be non-zero, and the bounds on the graviton mass derived from the Yukawa falloff of the gravitational potential only refer to the scalar sector. Furthermore, these models do not suffer of the vDVZ discontinuity and do not have ghosts.

A bound on the mass that refers directly to h_{ij}^{TT} can be obtained from pulsar timing, as recognized in Damour and Taylor (1991) and discussed quantitatively in Finn and Sutton (2002). The possibility of bounding the mass of h_{ij}^{TT} from the observation of inspiraling compact binaries is discussed in Will (1998) and Larson and Hiscock (2000).

6.0.47 Generation of GWs in linearized theory

We now consider the generation of GWs in the context of linearized theory. This means that we assume that the gravitational field generated by the source is sufficiently weak, so that the background space-time can be taken as flat. This will allow us to understand, in the simple setting of a flat background space-time (and therefore Newtonian or at most special-relativistic dynamics for the sources), how GWs are produced. In Section 3.1 we will derive the formulas for GW production valid in flat space-time, but exact in v/c . Then, expanding the exact result in powers of v/c , we will see that for small velocities the GW production can be organized in a multipole expansion (Section 3.2). In Section 3.3 we discuss in detail the lowest order term, which is the quadrupole radiation. In Section 3.4 we discuss the next-to-leading terms, i.e. the mass-octupole and the current quadrupole radiation, and in Section 3.5 we present the

systematic multipole expansion to all orders, using first the formalism of symmetric-trace-free (STF) tensors, and then the spherical tensors formalism. Finally, in a Solved Problems section we discuss some applications of this formalism and we collect additional technical material.

For a system whose dynamics is determined by non-gravitational forces, the weak-field expansion and the low-velocity expansion are independent, and in this case it makes sense to consider weak-field sources with arbitrary velocities, as we do in this chapter. The application of this formalism to a system held together by gravitational forces is more subtle. For a gravitationally-bound two-body system with reduced mass μ and total mass m , we have $E_{\text{kin}} = -(1/2)U$, i.e.

$$\frac{1}{2}\mu v^2 = \frac{1}{2} \frac{G\mu m}{r} \quad (3.1)$$

and therefore

$$\frac{v^2}{c^2} = \frac{R_S}{2r} \quad (3.2)$$

where $R_S = 2Gm/c^2$ is the Schwarzschild radius associated to a mass m . Therefore, the low-velocity expansion is not independent of the expansion in powers of G .

The most interesting astrophysical sources of GWs, such as neutron stars, black holes or compact binaries, are self-gravitating systems. In this case, if we want to compute corrections in v/c , we must take into

3

3.1 Weak-field sources with arbitrary velocity 102

3.2 Low-velocity expansion 105

3.3 Mass quadrupole 109

3.4 Mass octupole and current quadrupole

3.5 Systematic multipole expansion 131

3.7 Solved problems 156

¹ More precisely, the retarded Green's function is selected by imposing the Kirchoff-Sommerfeld "no-incoming radiation" boundary conditions, i.e. one imposes

$$\lim_{t \rightarrow -\infty} \left[\frac{\partial}{\partial r} + \frac{\partial}{c \partial t} \right] (r \bar{h}_{\mu\nu})(\mathbf{x}, t) = 0$$

where the limit is taken along any surface $ct + r = \text{constant}$, together with the condition that $r \bar{h}_{\mu\nu}$ and $r \partial_\rho \bar{h}_{\mu\nu}$ be bounded in this limit. Physically, this means that there is no incoming radiation at past null infinity. account that, because of eq. (3.2), space-time cannot be considered flat beyond lowest order, and therefore the dynamics of the sources can no longer be described by Newtonian gravity. The corresponding formalism is the GW generation from post-Newtonian sources, which will be the subject of Chapter 5. However we will see in Sect. 3.3.5 that, even for self-gravitating system, some lowest-order computations for the mass quadrupole and even mass octupole and current quadrupole terms can be consistently performed within linearized theory.

6.0.48 3.1 Weak-field sources with arbitrary velocity

In linearized theory the starting point is eq. (1.24), that we recall here,

$$\square \bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4} T_{\mu\nu} \quad (3.3)$$

where $T_{\mu\nu}$ is the energy-momentum tensor of matter. Recall also that we are in the Lorentz gauge, $\partial^\mu \bar{h}_{\mu\nu} = 0$, and that $T_{\mu\nu}$ satisfies the flat-space conservation law $\partial^\mu T_{\mu\nu} = 0$. Equation (3.3) is linear in $h_{\mu\nu}$ and can be solved by the method of Green's function: if $G(x - x')$ is a solution of the equation

$$\square_x G(x - x') = \delta^4(x - x') \quad (3.4)$$

(where $\square_x$ is the d'Alembertian operator with derivatives taken with respect to the variable x), then the corresponding solution of eq. (3.3) is

$$\bar{h}_{\mu\nu}(x) = -\frac{16\pi G}{c^4} \int d^4x' G(x - x') T_{\mu\nu}(x') \quad (3.5)$$

The solution of eq. (3.4) depends of course on the boundary conditions that we impose. Just as in electromagnetism, for a radiation problem the appropriate solution is the retarded Green's function,¹

$$G(x - x') = -\frac{1}{4\pi |\mathbf{x} - \mathbf{x}'|} \delta(x_{\text{ret}}^0 - x'^0) \quad (3.6)$$

where $x^0 = ct'$, $x_{\text{ret}}^0 = ct_{\text{ret}}$, and

$$t_{\text{ret}} = t - \frac{|\mathbf{x} - \mathbf{x}'|}{c} \quad (3.7)$$

is called retarded time. Then the solution of eq. (3.3) is

$$\bar{h}_{\mu\nu}(t, \mathbf{x}) = \frac{4G}{c^4} \int d^3x' \frac{1}{|\mathbf{x} - \mathbf{x}'|} T_{\mu\nu} \left(t - \frac{|\mathbf{x} - \mathbf{x}'|}{c}, \mathbf{x}' \right) \quad (3.8)$$

Outside the source we can put this solution in the TT gauge using eq. (1.40), $h_{ij}^{\text{TT}} = \Lambda_{ij,kl} h_{kl} = \Lambda_{ij,kl} \bar{h}_{kl}$ (in the last equality we used the property (1.38) of the Lambda tensor). Therefore, outside the source,

$$h_{ij}^{\text{TT}}(t, \mathbf{x}) = \frac{4G}{c^4} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \int d^3x' \frac{1}{|\mathbf{x} - \mathbf{x}'|} T_{kl} \left(t - \frac{|\mathbf{x} - \mathbf{x}'|}{c}, \mathbf{x}' \right)$$

where we use the notation $\hat{\mathbf{x}} = \hat{\mathbf{n}}$, and we will also denote $|\mathbf{x}| = r$. Observe that h_{ij}^{TT} depends only the integrals of the spatial components T_{kl} . The underlying reason that allowed us to eliminate T_{0k} and T_{00} is that they are related to T_{kl} by the conservation of the energy-momentum tensor.² If we denote by d the typical radius of the source, at $r \gg d$ we can expand

$$|\mathbf{x} - \mathbf{x}'| = r - \mathbf{x}' \cdot \hat{\mathbf{n}} + O\left(\frac{d^2}{r}\right) \quad (3.10)$$

see Fig. 3.1. We are particularly interested in the value of h_{ij}^{TT} at large distances from the source, where the detector is located, so we take the limit $r \rightarrow \infty$ at fixed t ,³ and we retain only the leading term in eq. (3.9). This is a term $O(1/r)$, obtained setting $|\mathbf{x} - \mathbf{x}'| = r$ in the denominator of eq. (3.9), so at large distances

$$h_{ij}^{\text{TT}}(t, \mathbf{x}) = \frac{1}{r} \frac{4G}{c^4} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \int d^3x' T_{kl} \left(t - \frac{r}{c} + \frac{\mathbf{x}' \cdot \hat{\mathbf{n}}}{c}, \mathbf{x}' \right) \quad (3.11)$$

plus terms $O(1/r^2)$ that we neglect. We now write T_{kl} in terms of its Fourier transform,⁴

$$T_{kl}(t, \mathbf{x}) = \int \frac{d^4k}{(2\pi)^4} \tilde{T}_{kl}(\omega, \mathbf{k}) e^{-i\omega t + i\mathbf{k} \cdot \mathbf{x}} \quad (3.12)$$

Then

$$\begin{aligned} & \int d^3x' T_{kl} \left(t - \frac{r}{c} + \frac{\mathbf{x}' \cdot \hat{\mathbf{n}}}{c}, \mathbf{x}' \right) \\ &= \int d^3x' \int \frac{d\omega}{2\pi c} \frac{d^3k}{(2\pi)^3} \tilde{T}_{kl}(\omega, \mathbf{k}) e^{-i\omega(t-r/c)} e^{i(\mathbf{k} - \omega \hat{\mathbf{n}}/c) \cdot \mathbf{x}'} \\ &= \int \frac{d\omega}{2\pi c} \frac{d^3k}{(2\pi)^3} \tilde{T}_{kl}(\omega, \mathbf{k}) e^{-i\omega(t-r/c)} (2\pi)^3 \delta^{(3)}(\mathbf{k} - \omega \hat{\mathbf{n}}/c) \\ &= \int \frac{d\omega}{2\pi c} \tilde{T}_{kl}(\omega, \omega \hat{\mathbf{n}}/c) e^{-i\omega(t-r/c)} \end{aligned} \quad (3.13)$$

and we get

$$h_{ij}^{\text{TT}}(t, \mathbf{x}) = \frac{1}{r} \frac{4G}{c^5} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \tilde{T}_{kl}(\omega, \omega \hat{\mathbf{n}}/c) e^{-i\omega(t-r/c)}$$

² Indeed, when performing the multipole expansion below, the lowest order result will be re-expressed in terms of T^{00} only, using again energy-momentum conservation.

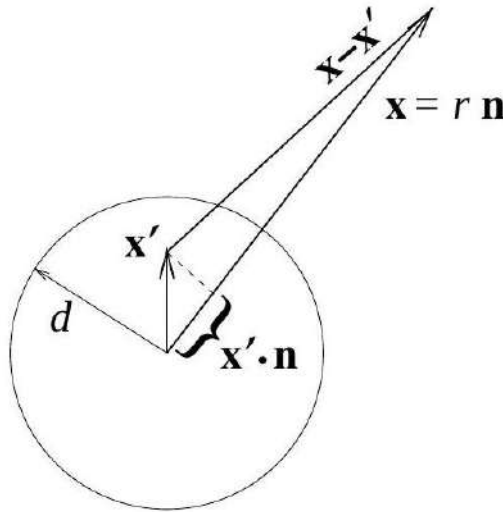


Fig. 3.1 A graphical illustration of the relation given in eq. (3.10).

³ In linearized theory, GWs are studied at spatial infinity, i.e. $r \rightarrow \infty$ at fixed t . We will see in Chapter 5 that, beyond linearized theory, it can be more convenient to work at future null infinity, i.e. $r \rightarrow \infty$ with $t - r/c$ fixed.

⁴ Our convention on the factors of c is that the four-dimensional wave-vector is $k^\mu = (\omega/c, \mathbf{k})$, and therefore $d^4k = (1/c)d\omega d^3k$. Since $x^\mu = (ct, \mathbf{x})$, we then have $k_\mu x^\mu = -\omega t + \mathbf{k} \cdot \mathbf{x}$. Observe that $\mathbf{k}$ has dimensions of the inverse of length. The spatial momentum of a particle with wave-vector $\mathbf{k}$ is $\mathbf{p} = \hbar \mathbf{k}$.

⁵ To compare with Weinberg (1972), Section 10.4, observe that we define the Fourier transform with respect to frequency using $d\omega/(2\pi)$ (see eq. (3.12), or the Notation section) while Weinberg uses $d\omega$; on the other hand, we both use $d^3k/(2\pi)^3$ in the spatial Fourier transform. Therefore our $\tilde{T}(\omega, \mathbf{k})$ is equal to $2\pi\tilde{T}^{\text{Weinberg}}(\omega, \mathbf{k})$.

⁶ Of course, at the latest this divergence is cutoff by the back-reaction due to GW emission. For example, a spinning neutron star with non-vanishing ellipticity emits GWs. The energy of these waves is taken from the rotational energy of the star which therefore gradually slows down, as we will compute in detail in Section 4.2. (Actually, in this case electromagnetic effects dominate and slow down the neutron star even earlier.)

⁷ More precisely, we have seen in Chapter 1 that the GW energy is only defined by averaging over a few periods, so this is really the average power radiated over one period of the source motion.

In general, the Fourier components of the energy-momentum tensor of the source will be large around a typical value ω_s , and the characteristic speed at which there is a bulk movement of mass across the source is $v \sim \omega_s d$. For the moment we have made no assumption on the relative values of ω_s and d , and in particular we have not assumed $\omega_s d \ll c$. Therefore eq. (3.14) is valid both for relativistic and for non-relativistic sources, as long as linearized theory applies, and we are at a sufficiently large distance r from the source.

From eq. (1.156), setting $dA = r^2 d\Omega$, we see that the total energy radiated per unit solid angle is

$$\frac{dE}{d\Omega} = \frac{r^2 c^3}{32\pi G} \int_{-\infty}^{\infty} dt \dot{h}_{ij}^{\text{TT}} \dot{h}_{ij}^{\text{TT}} \quad (3.15)$$

Inserting here the expression (3.14), using $\tilde{T}(-\omega, -\mathbf{k}) = \tilde{T}^*(\omega, \mathbf{k})$ and the property (1.37) of the Lambda tensor, we find ⁵

$$\frac{dE}{d\Omega} = \frac{G}{2\pi^2 c^7} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \int_0^{\infty} d\omega \omega^2 \tilde{T}_{ij}(\omega, \omega \hat{\mathbf{n}}/c) \tilde{T}_{kl}^*(\omega, \omega \hat{\mathbf{n}}/c)$$

and the energy spectrum is therefore

$$\frac{dE}{d\omega} = \frac{G\omega^2}{2\pi^2 c^7} \int d\Omega \Lambda_{ij,kl}(\hat{\mathbf{n}}) \tilde{T}_{ij}(\omega, \omega \hat{\mathbf{n}}/c) \tilde{T}_{kl}^*(\omega, \omega \hat{\mathbf{n}}/c) \quad (3.17)$$

A typical source will radiate for a characteristic time Δt . In the idealized case of an exactly monochromatic source, the radiation lasts for $\Delta t = \infty$ and the total radiated energy is formally divergent. ⁶ Thus, for a monochromatic source the instantaneously radiated power is a more useful quantity. For such a source, radiating at a frequency ω_0 , we write $\tilde{T}_{ij}(\omega, \mathbf{k})$ (for positive ω) as

$$\tilde{T}_{ij}(\omega, \mathbf{k}) = \theta_{ij}(\omega, \mathbf{k}) 2\pi \delta(\omega - \omega_0) \quad (3.18)$$

and eq. (3.16) becomes

$$\frac{dE}{d\Omega} = \frac{G\omega_0^2}{\pi c^7} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \theta_{ij}(\omega_0, \omega_0 \hat{\mathbf{n}}/c) \theta_{kl}^*(\omega_0, \omega_0 \hat{\mathbf{n}}/c) T \quad (3.19)$$

We have used $2\pi\delta(\omega = 0) = T$, where T is the total (formally infinite) time. Dividing by T we obtain the power radiated instantaneously, ⁷

$$\frac{dP}{d\Omega} = \frac{G\omega_0^2}{\pi c^7} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \theta_{ij}(\omega_0, \omega_0 \hat{\mathbf{n}}/c) \theta_{kl}^*(\omega_0, \omega_0 \hat{\mathbf{n}}/c) \quad (3.20)$$

The total power is obtained by integrating over $d\Omega$. To perform the integration one can use the identities

$$\int \frac{d\Omega}{4\pi} n_i n_j = \frac{1}{3} \delta_{ij} \quad (3.21)$$

$$\int \frac{d\Omega}{4\pi} n_i n_j n_k n_l = \frac{1}{15} (\delta_{ij} \delta_{kl} + \delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}) \quad (3.22)$$

These identities, as well as their generalization to an arbitrary number of n 's, can be found as follows. For an odd number of n_i the integral vanishes because the integrand is odd under parity. For an even number of n , we use the fact that the tensor $n_{i_1} n_{i_2} \dots n_{i_{2l}}$ is totally symmetric and therefore its integral can only depend on the totally symmetrized product of Kronecker deltas. Once the tensor structure is fixed, the overall constant is obtained by contracting all indices. This gives

$$\int \frac{d\Omega}{4\pi} n_{i_1} \dots n_{i_{2l}} = \frac{1}{(2l+1)!!} (\delta_{i_1 i_2} \delta_{i_3 i_4} \dots \delta_{i_{2l-1} i_{2l}} + \dots) \quad (3.23)$$

where the final dots denote all possible pairing of indices.

6.0.49 3.2 Low-velocity expansion

Just as in electrodynamics, the equations for the generation of radiation are greatly simplified if the typical velocities inside the source are small compared to the speed of light. If ω_s is the typical frequency of the motion inside the source and d is the source size, the typical velocities inside the source are $v \sim \omega_s d$. The frequency ω of the radiation will also be of order ω_s and therefore $\omega \sim \omega_s \sim v/d$. In terms of $\lambda = c/\omega$,

$$\lambda \sim \frac{c}{v} d \quad (3.24)$$

In a non-relativistic system, $v \ll c$ and the reduced wavelength of the radiation generated is much bigger than the size of the system:

$$\text{non-relativistic sources} \implies \lambda \gg d \quad (3.25)$$

When the reduced wavelength is much bigger than the size of the system, it is physically clear that we do not need to know the internal motions of the source in all its details, but only the coarse features matter, so the emission of radiation is governed by the lowest multipole moments.⁹

To perform the multipole expansion for gravitational radiation we start from the expression of h_{ij}^{TT} at spatial infinity given in eq. (3.11), that we recall here,

$$h_{ij}^{\text{TT}}(t, \mathbf{x}) = \frac{1}{r} \frac{4G}{c^4} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \int d^3x' T_{kl} \left(t - \frac{r}{c} + \frac{\mathbf{x}' \cdot \hat{\mathbf{n}}}{c}, \mathbf{x}' \right) \quad (3.26)$$

and we write T_{kl} in terms of its Fourier transform,

$$T_{kl} \left(t - \frac{r}{c} + \frac{\mathbf{x}' \cdot \hat{\mathbf{n}}}{c}, \mathbf{x}' \right) = \int \frac{d^4k}{(2\pi)^4} \tilde{T}_{kl}(\omega, \mathbf{k}) e^{-i\omega(t-r/c + \mathbf{x}' \cdot \hat{\mathbf{n}}/c) + i\mathbf{k} \cdot \mathbf{x}'} \quad (3.27)$$

For a non-relativistic source, $\tilde{T}_{kl}(\omega, \mathbf{k})$ is peaked around a typical frequency ω_s (or around a range of frequencies, with maximum value ω_s),

⁹ A typical example is the electromagnetic radiation from the hydrogen atom. The velocity of the electron inside the hydrogen atom is $v/c \sim \alpha$, where $\alpha \simeq 1/137$ is the fine-structure constant, and the reduced wavelengths of the transitions between the levels of the hydrogen atom are of order $\lambda \sim r_B/\alpha$, where r_B is the Bohr radius. Since $\alpha \ll 1$, we have $\lambda \gg r_B$, and the multipole expansion is adequate.

¹⁰ One could have directly written the expansion (3.30), as a formal Taylor expansion in the parameter $\mathbf{x}' \cdot \hat{\mathbf{n}}/c$. However, the above derivation emphasizes that the assumption behind this expansion is the condition $\omega_s d \ll c$, with ω_s the typical source frequency.

¹¹ Observe that, contrary to most of the literature on general relativity, we never use commas to denote derivatives (nor semicolons to denote covariant derivatives). with $\omega_s d \ll c$. On the other hand, the energy-momentum tensor is non-vanishing only inside the source, so the integral in eq. (3.26) is restricted to $|\mathbf{x}'| \leq d$. Then the dominant contribution to h_{ij}^{TT} comes from frequencies ω that satisfy

$$\frac{\omega}{c} \mathbf{x}' \cdot \hat{\mathbf{n}} \lesssim \frac{\omega_s d}{c} \ll 1 \quad (3.28)$$

and therefore we can expand the exponential in eq. (3.27),

$$e^{-i\omega(t-r/c+\mathbf{x}' \cdot \hat{\mathbf{n}}/c)} = e^{-i\omega(t-r/c)} \times \left[1 - i\frac{\omega}{c} x'^i n^i + \frac{1}{2} \left(-i\frac{\omega}{c} \right)^2 x'^i x'^j n^i n^j + \dots \right] \quad (3.29)$$

This is equivalent to expanding

$$T_{kl} \left(t - \frac{r}{c} + \frac{\mathbf{x}' \cdot \hat{\mathbf{n}}}{c}, \mathbf{x}' \right) = T_{kl} \left(t - \frac{r}{c}, \mathbf{x}' \right) + \frac{x'^i n^i}{c} \partial_t T_{kl} + \frac{1}{2c^2} x'^i x'^j n^i n^j \partial_t^2 T_{kl} + \dots \quad (3.30)$$

where all derivatives are evaluated at the point $(t - r/c, \mathbf{x}')$.¹⁰ We now define the momenta of the stress tensor T^{ij} ,

$$S^{ij}(t) = \int d^3x T^{ij}(t, \mathbf{x}) \quad (3.31)$$

$$S^{ij,k}(t) = \int d^3x T^{ij}(t, \mathbf{x}) x^k \quad (3.32)$$

$$S^{ij,kl}(t) = \int d^3x T^{ij}(t, \mathbf{x}) x^k x^l \quad (3.33)$$

and similarly for all higher order momenta. In this notation, a comma separates the spatial indices which originates from T^{ij} from the indices coming from $x^{i_1} \dots x^{i_N}$.¹¹ The energy-momentum tensor of matter that appears in eq. (3.3) is the one obtained from the variation of the matter action with respect to the metric, so it is in its symmetric form, $T^{ij} = T^{ji}$.

⁸ Apart from numerical factors which depend on the multipole moment involved and on the details of the motion of the source. We will see below that for a non-relativistic system the dominant contributions come from the lowest multipoles, and for these the numerical factors are $O(1)$; for instance, we will see that a source performing a simple harmonic oscillation at frequency ω_s emits quadrupole radiation $\omega = 2\omega_s$.

Then, its momenta are symmetric separately in the first type of indices and in the second, e.g. $S^{ij,k} = S^{ji,k}$ or $S^{ij,kl} = S^{ij,lk}$, but not necessarily under the exchange of two indices of different type, e.g. in general $S^{ij,k} \neq S^{ik,j}$.

Inserting the expansion (3.30) into eq. (3.26) we get

$$h_{ij}^{\text{TT}}(t, \mathbf{x}) = \frac{1}{r} \frac{4G}{c^4} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \times \left[S^{kl} + \frac{1}{c} n_m \dot{S}^{kl,m} + \frac{1}{2c^2} n_m n_p \ddot{S}^{kl,mp} + \dots \right]_{\text{ret}}$$

where the subscript "ret" means that the quantities S^{kl} , $\dot{S}^{kl,m}$, $\ddot{S}^{kl,mp}$, etc. are evaluated at retarded time $t - r/c$. This equation is the basis for

the multipole expansion. From the definitions (3.31)-(3.32) we see that, with respect to S^{kl} , $\dot{S}^{kl,m}$ has an additional factor $x^m \sim O(d)$, while each time derivative brings a factor $O(\omega_s)$. So, with respect to S^{kl} , the tensor $\dot{S}^{kl,m}$ has an additional factor $O(\omega_s d)$, i.e. $O(v)$, where v is a typical velocity inside the source. Then the term $(1/c)n_m \dot{S}^{kl,m}$ is a correction $O(v/c)$ to the term S^{kl} , and similarly the term $(1/c^2)n_m n_p \ddot{S}^{kl,mp}$ is a correction $O(v^2/c^2)$, etc.

The physical meaning of the various terms in this expansion becomes more clear if we eliminate the momenta of T^{ij} in favor of the momenta of the energy density T^{00} , and of the momenta of the linear momentum, T^{0i}/c . We define the momenta of T^{00}/c^2 by ¹²

$$M = \frac{1}{c^2} \int d^3x T^{00}(t, \mathbf{x}) \quad (3.35)$$

$$M^i = \frac{1}{c^2} \int d^3x T^{00}(t, \mathbf{x}) x^i \quad (3.36)$$

$$M^{ij} = \frac{1}{c^2} \int d^3x T^{00}(t, \mathbf{x}) x^i x^j \quad (3.37)$$

$$M^{ijk} = \frac{1}{c^2} \int d^3x T^{00}(t, \mathbf{x}) x^i x^j x^k \quad (3.38)$$

and so on, while the momenta of the momentum density $(1/c)T^{0i}$ are denoted by

$$P^i = \frac{1}{c} \int d^3x T^{0i}(t, \mathbf{x}) \quad (3.39)$$

$$P^{i,j} = \frac{1}{c} \int d^3x T^{0i}(t, \mathbf{x}) x^j \quad (3.40)$$

$$P^{i,jk} = \frac{1}{c} \int d^3x T^{0i}(t, \mathbf{x}) x^j x^k \quad (3.41)$$

and similarly for the higher momenta. The time derivatives of these quantities and of the momenta of T^{ij} satisfy relations which follow from energy-momentum conservation. Recall that we are working within linearized theory, which means that the energy-momentum tensor of matter $T^{\mu\nu}$ satisfies the flat-space equation $\partial_\mu T^{\mu\nu} = 0$, as we have seen in eq. (1.25), while non-linearities such as those written schematically in eq. (2.113) are neglected. This means that we are also neglecting the back-action of the GWs on the source. ¹³

To obtain these identities, we take a box of volume V larger than the source, and we denote its boundary by ∂V (so $T_{\mu\nu}$ vanishes on ∂V). Using $\partial_\mu T^{\mu 0} = 0$, that is

$$\partial_0 T^{00} = -\partial_i T^{0i} \quad (3.42)$$

(and recalling that $\dot{M} \equiv \partial M / \partial t = c \partial_0 M$) we have

$$\begin{aligned}
c\dot{M} &= \int_V d^3x \partial_0 T^{00} \\
&= - \int_V d^3x \partial_i T^{0i}
\end{aligned}$$

¹⁴ In particular, the momenta of T^{ij} , i.e. $\{S^{ij}, S^{ij,k}, \dots\}$ depend on the distribution of the stresses inside the body, which might be difficult to determine, while the total mass of a body, its mass quadrupole, etc. can be measured more easily. Furthermore, as we will discuss in Sect. 3.3.5, for a set of interacting point masses T^{00} is $O(v^0)$ and T^{0i} is $O(v)$, so to lowest order in v/c they can be consistently computed using the free-particle energy-momentum tensor and neglecting the contribution of the potential and the relativistic corrections, which start from $O(v^2)$. This is not the case for T^{ij} , which is itself $O(v^2)$.

$$\begin{aligned}
&= - \int_{\partial V} dS^i T^{0i} \\
&= 0
\end{aligned} \tag{3.43}$$

The last equality follows from the fact that T^{0i} vanishes on the boundary ∂V , since we have taken the volume V larger than the volume of the source. Of course, a physical system that radiates GWs loses mass. The conservation of the mass M of the radiating body, expressed by eq. (3.43), is due to the fact that in the linearized approximation the back action of the source dynamics due to the energy carried away by the GWs is neglected. Similarly, we obtain the identity

$$\begin{aligned}
c\dot{M}^i &= \int_V d^3x x^i \partial_0 T^{00} \\
&= - \int_V d^3x x^i \partial_j T^{0j} \\
&= \int_V d^3x (\partial_j x^i) T^{0j} \\
&= \int_V d^3x \delta_j^i T^{0j} \\
&= cP^i
\end{aligned} \tag{3.44}$$

In the same way one derives similar identities for the higher momenta of T^{00} and of T^{0i} . For the first few lowest-order momenta of T^{00} we get

¹² Dimensionally T^{00}/c^2 is a mass density but of course, besides the contribution due to the rest mass of the source, it contains also all contributions to T^{00} coming from the kinetic energy of the particles which make up the source, contributions from the potential energy, etc. For sources that generate a strong gravitational field, such as neutron stars, the gravitational binding energy will also be important. Only for weak-field sources and in the nonrelativistic limit, T^{00}/c^2 becomes the mass density. However, since the multipole expansion of the linearized theory assumes weak fields and is a nonrelativistic expansion, to lowest order in v/c we can actually replace T^{00}/c^2 with the mass density.

¹³ The inclusion of these non-linearities will be discussed in Chapter 5.

$$\dot{M} = 0 \quad (3.45)$$

$$\dot{M}^i = P^i \quad (3.46)$$

$$\dot{M}^{ij} = P^{i,j} + P^{j,i} \quad (3.47)$$

$$\dot{M}^{ijk} = P^{i,jk} + P^{j,ki} + P^{k,ij} \quad (3.48)$$

while the lowest-order momenta of T^{0i} satisfy

$$\dot{P}^i = 0 \quad (3.49)$$

$$\dot{P}^{i,j} = S^{ij} \quad (3.50)$$

$$\dot{P}^{i,jk} = S^{ij,k} + S^{ik,j} \quad (3.51)$$

The equations $\dot{M} = 0$ and $\dot{P}^i = 0$ are the conservation of the mass and of the total momentum of the source. Another interesting identity is $\dot{P}^{i,j} - \dot{P}^{j,i} = S^{ij} - S^{ji} = 0$, which follows from eq. (3.50) using the fact that S^{ij} is a symmetric tensor, and is the conservation of the angular momentum of the source.

We can now combine these identities to express the momenta S^{ij} , $\dot{S}^{ij,k}$, etc., that appear in the multipole expansion, in terms of the two sets of momenta $\{M, M^i, M^{ij}, \dots\}$ and $\{P^i, P^{i,j}, \dots\}$, which have a more immediate physical interpretation.¹⁴ Taking the time derivative of eq. (3.47) and using eq. (3.50), as well as the fact that $S^{ij} = S^{ji}$, we obtain the identity

$$S^{ij} = \frac{1}{2} \ddot{M}^{ij} \quad (3.52)$$

If we combine eq. (3.48) with eq. (3.51) instead, we get

$$\ddot{M}^{ijk} = 2 \left(\dot{S}^{ij,k} + \dot{S}^{ik,j} + \dot{S}^{jk,i} \right) \quad (3.53)$$

From eq. (3.51) it also follows that $\ddot{P}^{i,jk} = \dot{S}^{ij,k} + \dot{S}^{ik,j}$. Using this relation and eq. (3.53) we can verify that

$$\dot{S}^{ij,k} = \frac{1}{6} \ddot{M}^{ijk} + \frac{1}{3} \left(\ddot{P}^{i,jk} + \ddot{P}^{j,ik} - 2\ddot{P}^{k,ij} \right) \quad (3.54)$$

Equations (3.52) and (3.54) relate S^{ij} and $\dot{S}^{ij,k}$, which are the two lowest-order momenta appearing in the multipole expansion (3.34), to the momenta of T^{00} and of T^{0i} . One can proceed similarly with the higher-order terms. In the next two sections, we examine the leading and the next-to-leading terms, while in Section 3.5 we discuss systematically the expansion to all orders.

6.0.50 3.3 Mass quadrupole radiation

6.0.51 3.3.1 Amplitude and angular distribution

Using eq. (3.52), the leading term of the expansion (3.34) is

$$[h_{ij}^{\text{TT}}(t, \mathbf{x})]_{\text{quad}} = \frac{1}{r} \frac{2G}{c^4} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \ddot{M}^{kl}(t - r/c) \quad (3.55)$$

From the point of view of the rotation group the tensor M_{kl} , as any symmetric tensor with two indices, decomposes into irreducible representations as

$$M^{kl} = \left(M^{kl} - \frac{1}{3} \delta^{kl} M_{ii} \right) + \frac{1}{3} \delta^{kl} M_{ii} \quad (3.56)$$

where M_{ii} is the trace of M_{ij} . The first term is traceless by construction, and is a pure spin-2 operator, while the trace part is a scalar. Since the Lambda tensor $\Lambda_{ij,kl}$ gives zero when contracted with δ_{kl} , only the traceless term contributes. We use the notation

$$\rho = \frac{1}{c^2} T^{00} \quad (3.57)$$

To lowest order in v/c , ρ becomes the mass density, see Note 12. We also introduce the quadrupole moment

$$\begin{aligned} Q^{ij} &\equiv M^{ij} - \frac{1}{3} \delta^{ij} M_{kk} \\ &= \int d^3x \rho(t, \mathbf{x}) \left(x^i x^j - \frac{1}{3} r^2 \delta^{ij} \right) \end{aligned} \quad (3.58)$$

and eq. (3.55) becomes

$$\begin{aligned} [h_{ij}^{\text{TT}}(t, \mathbf{x})]_{\text{quad}} &= \frac{1}{r} \frac{2G}{c^4} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \ddot{Q}_{kl}(t - r/c) \\ &\equiv \frac{1}{r} \frac{2G}{c^4} \ddot{Q}_{ij}^{\text{TT}}(t - r/c) \end{aligned} \quad (3.59)$$

6.0.52 Angular distribution of quadrupole radiation

¹⁵ We write the result in terms of the second mass moment M_{ij} , rather than in terms of the quadrupole moment Q_{ij} . Since $\Lambda_{ij,kl} Q_{kl} = \Lambda_{ij,kl} M_{kl}$ (because $\Lambda_{ij,kl} \delta_{kl} = 0$, see eq. (1.38)), in the equations below we could use M_{ij} or Q_{ij} equivalently. Typically, it is slightly more practical to use M_{ij} when one makes explicit computations.

In order to obtain the waveform emitted into an arbitrary direction $\hat{\mathbf{n}}$, we could in principle plug the explicit expression (1.39) of the Lambda tensor into eq. (3.59), and perform the contraction with $\ddot{Q}_{kl}$. It is however more instructive to proceed as follows. First we observe that, when the direction of propagation $\hat{\mathbf{n}}$ of the GW is equal to $\hat{\mathbf{z}}$, P_{ij} is the diagonal matrix $\text{diag}(1, 1, 0)$, i.e. P_{ij} is a projector on the (x, y) plane. Writing $\Lambda_{ij,kl}$ in terms of P_{ij} using eq. (1.36) we have, for an arbitrary 3×3 matrix A_{kl} ,

$$\begin{aligned} \Lambda_{ij,kl} A_{kl} &= \left[P_{ik} P_{jl} - \frac{1}{2} P_{ij} P_{kl} \right] A_{kl} \\ &= (PAP)_{ij} - \frac{1}{2} P_{ij} \text{Tr}(PA) \end{aligned} \quad (3.60)$$

When P has the form

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (3.61)$$

we get

$$PAP = \begin{pmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (3.62)$$

while $\text{Tr}(PA) = A_{11} + A_{22}$. Therefore

$$\begin{aligned} \Lambda_{ij,kl} A_{kl} &= \begin{pmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix}_{ij} - \frac{A_{11} + A_{22}}{2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}_{ij} \\ &= \begin{pmatrix} (A_{11} - A_{22})/2 & A_{12} & 0 \\ A_{21} & -(A_{11} - A_{22})/2 & 0 \\ 0 & 0 & 0 \end{pmatrix}_{ij} \end{aligned} \quad (3.63)$$

Thus, when $\hat{\mathbf{n}} = \hat{\mathbf{z}}^{15}$

$$\Lambda_{ij,kl} \ddot{M}_{kl} = \begin{pmatrix} (\ddot{M}_{11} - \ddot{M}_{22})/2 & \ddot{M}_{12} & 0 \\ \ddot{M}_{21} & -(\ddot{M}_{11} - \ddot{M}_{22})/2 & 0 \\ 0 & 0 & 0 \end{pmatrix}_{ij} \quad (3.64)$$

From this we directly read the two polarization amplitudes, for a GW propagating in the z direction,

$$h_+ = \frac{1}{r} \frac{G}{c^4} (\ddot{M}_{11} - \ddot{M}_{22}) \quad (3.65)$$

$$h_\times = \frac{2}{r} \frac{G}{c^4} \ddot{M}_{12} \quad (3.66)$$

where it is understood that the right-hand side is computed at the retarded time $t - r/c$. To compute the amplitudes for a wave that, in a frame with axes (x, y, z) , propagates in a generic direction $\hat{\mathbf{n}}$, we introduce two unit vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$, orthogonal to $\hat{\mathbf{n}}$ and to each other, chosen so that $\hat{\mathbf{u}} \times \hat{\mathbf{v}} = \hat{\mathbf{n}}$ (so, when $\hat{\mathbf{n}} = \hat{\mathbf{z}}$, we can take $\hat{\mathbf{u}} = \hat{\mathbf{x}}$ and $\hat{\mathbf{v}} = \hat{\mathbf{y}}$), see Fig. 3.2. Then in the (x', y', z') frame, whose axes are in the directions $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{n}})$, the wave propagates along the z' axis and we can use the previous result to read h_+ and $h_\times$,

$$h_+(t, \hat{\mathbf{n}}) = \frac{1}{r} \frac{G}{c^4} (\ddot{M}'_{11} - \ddot{M}'_{22}) \quad (3.67)$$

$$h_\times(t, \hat{\mathbf{n}}) = \frac{2}{r} \frac{G}{c^4} \ddot{M}'_{12} \quad (3.68)$$

where M'_{ij} , are the components of the second mass moment in the frame (x', y', z') .¹⁶ These can be related to the components M_{ij} in the (x, y, z) frame observing that in the (x', y', z') frame the vector $\hat{\mathbf{n}}$ has coordinates $n'_i = (0, 0, 1)$, while in the (x, y, z) frame it has coordinates

$$n_i = (\sin \theta \sin \phi, \sin \theta \cos \phi, \cos \theta) \quad (3.69)$$

Then the components n_i and n'_i are related by a rotation matrix $\mathcal{R}$ such that $n_i = \mathcal{R}_{ij} n'_j$, whose explicit expression is

$$\mathcal{R} = \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{pmatrix} \quad (3.70)$$

Similarly, a tensor $\mathbf{M}$ with two indices has components M_{ij} in the (x, y, z) frame and M'_{ij} in the (x', y', z') frame, related by

$$M_{ij} = \mathcal{R}_{ik} \mathcal{R}_{jl} M'_{kl} \quad (3.71)$$

or, solving for M' , $M'_{ij} = (\mathcal{R}^T M \mathcal{R})_{ij}$, where $\mathcal{R}^T$ is the transpose matrix. Inserting $\mathcal{R}$ from eq. (3.70), and plugging the resulting values of M'_{ij} into eqs. (3.67) and (3.68), we get

$$\begin{aligned} h_+(t; \theta, \phi) &= \frac{1}{r} \frac{G}{c^4} \left[\ddot{M}_{11} (\cos^2 \phi - \sin^2 \phi \cos^2 \theta) \right. \\ &\quad + \ddot{M}_{22} (\sin^2 \phi - \cos^2 \phi \cos^2 \theta) \\ &\quad - \ddot{M}_{33} \sin^2 \theta \\ &\quad - \ddot{M}_{12} \sin 2\phi (1 + \cos^2 \theta) \\ &\quad + \ddot{M}_{13} \sin \phi \sin 2\theta \\ &\quad \left. + \ddot{M}_{23} \cos \phi \sin 2\theta \right] \\ h_\times(t; \theta, \phi) &= \frac{1}{r} \frac{G}{c^4} \left[(\ddot{M}_{11} - \ddot{M}_{22}) \sin 2\phi \cos \theta \right. \\ &\quad + 2\ddot{M}_{12} \cos 2\phi \cos \theta \\ &\quad - 2\ddot{M}_{13} \cos \phi \sin \theta \\ &\quad \left. + 2\ddot{M}_{23} \sin \phi \sin \theta \right]. \end{aligned} \quad (3.72)$$

This equation allows us to compute the angular distribution of the quadrupole radiation, once M_{ij} is given.¹⁶ Recall that h_+ and $h_\times$ are defined in terms of the components of h_{ij} in the plane transverse to the propagation direction. Therefore these are "the" polarization amplitudes, and are denoted by h_+ and $h_\times$, rather than h'_+ and $h'_\times$.

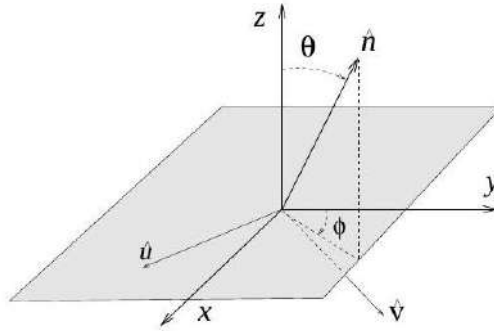


Fig. 3.2 The relation between the $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$ frame and the $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{n}})$ frame. The vector $\hat{\mathbf{u}}$ is in the $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ plane, while $\hat{\mathbf{v}}$ points downward, with respect to the $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ plane.

6.0.53 Absence of monopole and dipole gravitational radiation

We see from eq. (3.59) that the leading term of the multipole expansion is the mass quadrupole. There is neither monopole nor dipole radiation for GWs. This can be understood in two different ways. First of all, observe that a monopole term would depend on M and a dipole term on P^i (the mass dipole moment M^i can be set to zero with a shift of the origin of the coordinate system). Furthermore, h_{ij}^{TT} depends on derivatives of the multipole moments, since a static source does not radiate. However, M and P^i are conserved quantities, so any contribution from M or P^i must vanish.

Actually, M and P^i are conserved only at the level of linearized theory: a radiating system loses mass and, in general, also linear momentum (see page 130). However, the absence of monopole and dipole radiation holds more generally, and is not restricted to linearized theory. One can verify this observing that, even when we include all non-linear terms, as in eq. (2.113), the right-hand side of the wave equation must still be conserved, to be consistent with the Lorentz gauge condition. We will see explicitly in Section 5.2 how to write the equations in such a form. Using that expression, see in particular eqs. (5.69), (5.71) and (5.72), one can verify that the derivation of the absence of monopole and dipole radiation goes through even in the full non-linear theory. The difference with linearized theory is that the lowest-order multipole that contributes, rather than being the quadrupole moment of the energy density of matter, T^{00} , is the quadrupole moment of a more general quantity τ^{00} that includes also the contribution of the gravitational field.

However, there is no need to enter into the details of the non-linear theory: the absence of monopole and dipole radiation is simply the expression of the fact that the graviton is a massless particle with helicity ± 2 . We already showed in Problem 2.2 that it is impossible to put a graviton in a state with total angular momentum $j = 0$ or $j = 1$. This emerged as a consequence of the fact that the graviton is a massless particle with helicities ± 2 , and it therefore obeys gauge conditions that eliminate the spurious degrees of freedom. Indeed, these conditions allowed us to reduce the five degrees of a traceless symmetric tensor h_{ij} , which would be appropriate for describing a massive spin-2 particle, to the two degrees of freedom of a massless particle, as we discussed in Section 2.2.2.

Since it is impossible to put a graviton in a state with total angular momentum $j = 0$ or $j = 1$, we can have neither monopole nor dipole radiation, since they correspond to a collection of quanta with $j = 0$ and $j = 1$, respectively. The situation is completely analogous to electrodynamics, where the photon is massless and has helicity ± 1 , so it is impossible to put it in a state with total angular momentum $j = 0$ (see Problem 2.2, or Landau and Lifshitz, Vol. IV (1982), Section 6), and therefore monopole radiation is forbidden. In electromagnetism, the leading term of the multipole expansion is therefore dipole radiation.

6.0.54 3.3.2 Radiated energy

Inserting eq. (3.59) in eq. (1.153) and using the property (1.37) of the Lambda tensor we find the power radiated per unit solid angle, in the quadrupole approximation,

$$\begin{aligned} \left(\frac{dP}{d\Omega} \right)_{\text{quad}} &= \frac{r^2 c^3}{32\pi G} \langle \dot{h}_{ij}^{\text{TT}} \dot{h}_{ij}^{\text{TT}} \rangle \\ &= \frac{G}{8\pi c^5} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \langle \ddot{Q}_{ij} \ddot{Q}_{kl} \rangle \end{aligned} \quad (3.73)$$

where, as usual, the average is a temporal average over several characteristic periods of the GW, and it is understood that $\ddot{Q}_{ij}$ must be evaluated at the retarded time $t - r/c$. The angular integral can be performed observing that the dependence on $\hat{\mathbf{n}}$ is only in $\Lambda_{ij,kl}$. Using eqs. (3.21) and (3.22) we find

$$\int d\Omega \Lambda_{ij,kl} = \frac{2\pi}{15} (11\delta_{ik}\delta_{jl} - 4\delta_{ij}\delta_{kl} + \delta_{il}\delta_{jk}) \quad (3.74)$$

Then, the total radiated power (or, in notation used in astrophysics, the total gravitational luminosity $\mathcal{L}$ of the source) is, in the quadrupole approximation,

$$P_{\text{quad}} = \frac{G}{5c^5} \langle \ddot{Q}_{ij} \ddot{Q}_{ij} \rangle \quad (3.75)$$

where, again, $\ddot{Q}_{ij}$ must be evaluated at the retarded time $t - r/c$. This is the famous quadrupole formula, first derived by Einstein.¹⁷ Sometimes, in explicit computations, it is more practical to use M_{ij} rather than Q_{ij} . Substituting $Q_{ij} = M_{ij} - (1/3)\delta_{ij}M_{kk}$ in eq. (3.75) we have

$$P_{\text{quad}} = \frac{G}{5c^5} \left\langle \ddot{M}_{ij}\ddot{M}_{ij} - \frac{1}{3} \left(\ddot{M}_{kk} \right)^2 \right\rangle \quad (3.76)$$

The same result could be obtained by observing that eq. (3.73) is still valid if we replace $\Lambda_{ij,kl}\ddot{Q}_{ij}\ddot{Q}_{kl}$ by $\Lambda_{ij,kl}\ddot{M}_{ij}\ddot{M}_{kl}$, since the contraction of $\Lambda_{ij2,kl}$ with δ_{ij} (or with δ_{kl}) gives zero. However, when we use $\Lambda_{ij,kl}\ddot{M}_{ij}\ddot{M}_{kl}$, after integrating in $d\Omega$, on the right-hand side of eq. (3.74) the term $-4\delta_{ij}\delta_{kl}$ (which gave zero when contracted with $\ddot{Q}_{ij}\ddot{Q}_{kl}$) now contributes, since M_{ij} is not traceless, and we find eq. (3.76) again.

The energy radiated per unit solid angle is obtained by integrating the power, eq. (3.73), with respect to time. We write the quadrupole moment in terms of its Fourier transform,

$$Q_{ij}(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \tilde{Q}_{ij}(\omega) e^{-i\omega t} \quad (3.77)$$

and, integrating eq. (3.73) with respect to time, we get

$$\left(\frac{dE}{d\Omega} \right)_{\text{quad}} = \frac{G}{8\pi^2 c^5} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \int_0^{\infty} d\omega \omega^6 \tilde{Q}_{ij}(\omega) \tilde{Q}_{kl}^*(\omega)$$

¹⁸ To compare this equation with the results in the literature, beside checking the factor of 3 in the normalization of Q_{ij} , one must also check whether or not the Fourier transform is defined using $d\omega/(2\pi)$, as we do, or simply $d\omega$, as, for instance, in Weinberg (1972) or in Straumann (2004). where the integral in $d\omega$ from $-\infty$ to $+\infty$ has been written as twice an integral from zero to ∞ using $\tilde{Q}_{ij}(-\omega) = \tilde{Q}_{ij}^*(\omega)$. Integrating over the solid angle we find the total radiated energy,

$$E_{\text{quad}} = \frac{G}{5\pi c^5} \int_0^{\infty} d\omega \omega^6 \tilde{Q}_{ij}(\omega) \tilde{Q}_{ij}^*(\omega) \quad (3.79)$$

and therefore the energy spectrum, integrated over the solid angle, is ¹⁸

$$\left(\frac{dE}{d\omega} \right)_{\text{quad}} = \frac{G}{5\pi c^5} \omega^6 \tilde{Q}_{ij}(\omega) \tilde{Q}_{ij}^*(\omega) \quad (3.80)$$

For a monochromatic source, radiating at a frequency $\omega_0 > 0$, we proceed as in Section 3.1: we write, for positive ω ,

$$\tilde{Q}_{ij}(\omega) = q_{ij} 2\pi \delta(\omega - \omega_0) \quad (3.81)$$

insert this into eq. (3.78), and again use $2\pi\delta(\omega - \omega_0) = T$, where T is the total (infinite) time interval. The instantaneous power generated by the monochromatic source is obtained by dividing by T , so

$$\left(\frac{dP}{d\Omega d\omega} \right)_{\text{quad}} = \frac{G\omega_0^6}{4\pi c^5} (\Lambda_{ij,kl} q_{ij} q_{kl}^*) \delta(\omega - \omega_0) \quad (3.82)$$

As for the linear momentum, inserting eq. (3.59) into eq. (1.164) we get

$$\frac{dP^i}{dt} = -\frac{G}{8\pi c^5} \int d\Omega \ddot{Q}_{ab}^{\text{TT}} \partial^i \ddot{Q}_{ab}^{\text{TT}} \quad (3.83)$$

Under reflection, $\mathbf{x} \rightarrow -\mathbf{x}$, the quadrupole moment is invariant while $\partial^i \rightarrow -\partial^i$. Therefore the integrand is odd, and the angular integral vanishes. There is no loss of linear momentum in the quadrupole approximation. A non-vanishing result can be obtained by going beyond the quadrupole approximation, from the interference between multipoles of different parity, as we will see in Section 3.4.

6.0.55 3.3.3 Radiated angular momentum

The angular momentum carried away per unit time by GWs can be obtained by plugging the expression for h_{ij}^{TT} in the quadrupole approximation, eq. (3.59), into the general formula for the rate of angular momentum loss, eq. (2.61). Recalling that the first term in eq. (2.61) is the contribution from the orbital angular momentum L^i of the GWs while the second comes from the spin S^i of the field configuration, we write

$$\frac{dJ^i}{dt} = \frac{dL^i}{dt} + \frac{dS^i}{dt} \quad (3.84)$$

For the orbital part we have

$$\left(\frac{dL^i}{dt} \right)_{\text{quad}} = -\frac{c^3}{32\pi G} \epsilon^{ikl} \int d\Omega r^2 \left\langle \dot{h}_{ab}^{\text{TT}} x^k \partial^l h_{ab}^{\text{TT}} \right\rangle \quad (3.85)$$

We then substitute $h_{ab}^{\text{TT}}(t, \mathbf{x}) = (2G/rc^4) \Lambda_{ab,cd}(\hat{\mathbf{n}}) \ddot{Q}_{cd}(t - r/c)$ and we perform the angular integral. The explicit computation is slightly involved, but we find it useful to perform it in detail. The uninterested reader can jump to the result, eq. (3.93).

When we compute $\partial^l h_{ab}^{\text{TT}}$, the derivative ∂^l acts on $\Lambda_{ab,cd}(\hat{\mathbf{n}})$ (since $n^i = x^i/r$) as well as on $\ddot{Q}_{cd}(t - r/c)$. However,

$$\begin{aligned} \frac{\partial}{\partial x^l} \ddot{Q}_{cd}(t - r/c) &= \left(\frac{\partial r}{\partial x^l} \right) \frac{d}{dr} \ddot{Q}_{cd}(t - r/c) \\ &= -\frac{x^l}{r} \frac{1}{c} \ddot{Q}_{cd}(t - r/c). \end{aligned} \quad (3.86)$$

In eq. (3.85) this therefore gives a contribution proportional to $\epsilon^{ikl} x^k x^l = 0$. The only non-vanishing term is obtained when ∂^l acts on $\Lambda_{ab,cd}(\hat{\mathbf{n}})$, and

$$\left(\frac{dL^i}{dt} \right)_{\text{quad}} = -\frac{G}{2c^5} \epsilon^{ikl} \left\langle \ddot{Q}_{cd} \ddot{Q}_{fg} \right\rangle \int \frac{d\Omega}{4\pi} \Lambda_{ab,cd} x^k \partial^l \Lambda_{ab,fg} \quad (3.87)$$

where it is understood that the derivatives of the quadrupole moment are evaluated at the retarded time $t - r/c$. We observe that

$$\begin{aligned} \frac{\partial n^m}{\partial x^l} &= \frac{\partial}{\partial x^l} \left(\frac{x^m}{r} \right) \\ &= \frac{\delta^{lm}}{r} - \frac{x^l x^m}{r^3} \\ &= \frac{1}{r} P^{lm} \end{aligned} \quad (3.88)$$

where $P^{lm} = \delta^{lm} - n^l n^m$ is the projector first introduced in eq. (1.35).¹⁹ Then

$$\begin{aligned}\partial^l \Lambda_{ab,fg} &= \frac{\partial n^m}{\partial x^l} \frac{\partial}{\partial n^m} \Lambda_{ab,fg} \\ &= \frac{1}{r} P^{lm} \frac{\partial}{\partial n^m} \Lambda_{ab,fg}\end{aligned}\quad (3.89)$$

Using $\Lambda_{ab,fg}$ in the form (1.36) together with

$$\frac{\partial P^{ij}}{\partial n^m} = -(\delta^{im} n^j + \delta^{jm} n^i) \quad (3.90)$$

which follows from the definition of P^{ij} , we find the identity

$$P^{lm} \frac{\partial}{\partial n^m} \Lambda_{ab,fg} = -(n_f \Lambda_{ab,lg} + n_g \Lambda_{ab,lf} + n_a \Lambda_{lb,fg} + n_b \Lambda_{la,fg}) \quad (3.91)$$

The last two terms give zero when contracted with the factor $\Lambda_{ab,cd}$ in eq. (3.87), since the Lambda tensor is transverse on all indices, $n_a \Lambda_{ab,cd} = n_b \Lambda_{ab,cd} = 0$. Then

$$\begin{aligned}\left(\frac{dL^i}{dt}\right)_{\text{quad}} &= \frac{G}{2c^5} \epsilon^{ikl} \langle \ddot{Q}_{cd} \ddot{Q}_{fg} \rangle \int \frac{d\Omega}{4\pi} \Lambda_{ab,cd} n_k (n_f \Lambda_{ab,lg} + n_g \Lambda_{ab,lf}) \\ &= \frac{G}{2c^5} \epsilon^{ikl} \langle \ddot{Q}_{cd} \ddot{Q}_{fg} \rangle \int \frac{d\Omega}{4\pi} n_k (n_f \Lambda_{cd,lg} + n_g \Lambda_{cd,lf})\end{aligned}\quad (3.92)$$

¹⁹ As always, we do not need to be careful about raising and lowering spatial indices, since the spatial metric is δ_{ij} . Otherwise, we should write $\partial n^m / \partial x^l = (1/r) P_l^m$.

The angular integral can now be computed by inserting the explicit form of the Λ tensor (1.39) and using the identities (3.21) and (3.22) (the term with six factors $\hat{\mathbf{n}}$ is proportional to $n^k n^l$ and vanishes when contracted with ϵ^{ikl} , so we only need the integrals with two and with four factors $\hat{\mathbf{n}}$).

The result is

$$\left(\frac{dL^i}{dt}\right)_{\text{quad}} = \frac{2G}{15c^5} \epsilon^{ikl} \langle \ddot{Q}_{ka} \ddot{Q}_{la} \rangle \quad (3.93)$$

The calculation of the spin part gives

$$\begin{aligned}\left(\frac{dS^i}{dt}\right)_{\text{quad}} &= \frac{c^3}{16\pi G} \epsilon^{ikl} \int r^2 d\Omega \langle \dot{h}_{al}^{\text{TT}} \dot{h}_{ak}^{\text{TT}} \rangle \\ &= \frac{G}{c^5} \epsilon^{ikl} \langle \ddot{Q}_{mn} \ddot{Q}_{cd} \rangle \int \frac{d\Omega}{4\pi} \Lambda_{al,mn} \Lambda_{ak,cd}\end{aligned}\quad (3.94)$$

Using eq. (1.36) we can prove the identity

$$\Lambda_{al,mn} \Lambda_{ak,cd} = P_{ln} \Lambda_{mk,cd} - \frac{1}{2} P_{mn} \Lambda_{kl,cd} \quad (3.95)$$

The second factor gives zero contracted with ϵ^{ikl} and, again using the identities eqs. (3.21) and (3.22), the remaining angular integration is straightforward,

$$\begin{aligned}
\left(\frac{dS^i}{dt}\right)_{\text{quad}} &= \frac{G}{c^5} \epsilon^{ikl} \ddot{Q}_{mn} \ddot{Q}_{cd} \int \frac{d\Omega}{4\pi} P_{ln} \Lambda_{mk,cd} \\
&= \frac{4G}{15c^5} \epsilon^{ikl} \left\langle \ddot{Q}_{ka} \ddot{Q}_{la} \right\rangle
\end{aligned} \tag{3.96}$$

Summing the spin and orbital contribution, we finally get the angular momentum carried away, per unit time, by the GWs,

$$\left(\frac{dJ^i}{dt}\right)_{\text{quad}} = \frac{2G}{5c^5} \epsilon^{ikl} \left\langle \ddot{Q}_{ka} \ddot{Q}_{la} \right\rangle \tag{3.97}$$

where we recall again that the derivatives of the quadrupole moment are evaluated at the retarded time $t - r/c$.

6.0.56 3.3.4 Radiation reaction on non-relativistic sources

We have seen that gravitational radiation carries away energy and angular momentum. Ultimately, this energy and angular momentum must come from the source. We therefore expect that the energy and angular momentum carried by the GWs, at a large distance r from the source and at time t , were drained at retarded time $t - r/c$ from the energy and the angular momentum of the source. In the full non-linear theory of gravity (so, in particular, when dealing with self-gravitating sources), one must however take into account non-linear effects in the GW propagation from the source to infinity (such as back-scattering of gravitons on the background space-time, graviton-graviton scattering, etc.). We will see in Section 5.3.4 that, as a result, at higher orders in the postNewtonian expansion part of the gravitational radiation is delayed, and the total GW consists of a wavefront, which travels at the speed of light, and a "tail", which arrives later. Thus, it is not at all obvious that there is an exact equality, to all orders in the post-Newtonian expansion, between the instantaneous power radiated at large distances at a given time t , and the rate of energy loss of the source at the corresponding retarded time. We will discuss the issue in Section 5.3.5. However, as long as we are in linearized theory, the background space-time is flat, the wave propagates at the speed of light, and this energy balance argument is inescapable.

For $v/c \ll 1$, the leading term is given by quadrupole radiation, so the instantaneous rate of decreases of energy and orbital ²⁰ angular momentum of the source must be given by eqs. (3.75) and (3.97), i.e.

$$\frac{dE_{\text{source}}}{dt} = -\frac{G}{5c^5} \left\langle \ddot{Q}_{ij} \ddot{Q}_{ij} \right\rangle \tag{3.98}$$

$$\frac{dL_{\text{source}}^i}{dt} = -\frac{2G}{5c^5} \epsilon^{ikl} \left\langle \ddot{Q}_{ka} \ddot{Q}_{la} \right\rangle \tag{3.99}$$

We have required that dE_{source}/dt , computed at retarded time $t - r/c$, be the negative of the power radiated at a large distance r in GWs, at time t . Since the latter is expressed in terms of $\langle \ddot{Q}_{ij} \ddot{Q}_{ij} \rangle$ evaluated at retarded time, as in eq. (3.75), we have obtained an equality between dE_{source}/dt and $-(G/5c^5) \langle \ddot{Q}_{ij} \ddot{Q}_{ij} \rangle$, both evaluated at the same value of retarded time or, equivalently, at the same value of time (and similarly for the angular momentum).

The multipole expansion assumes that the sources are non-relativistic so, at least to lowest order, the dynamics of the source is governed by Newtonian mechanics, and it should be possible to describe the backreaction of the GWs on the source in terms of a force $\mathbf{F}$. Then, we expect that it should be possible to write eq. (3.98) in the form

$$\frac{dE_{\text{source}}}{dt} = \langle F_i v_i \rangle \quad (3.100)$$

or more precisely, for an extended body,

$$\frac{dE_{\text{source}}}{dt} = \left\langle \int d^3x \frac{dF_i}{dV} \dot{x}_i \right\rangle \quad (3.101)$$

where dF_i/dV is the force per unit volume. Similarly, we expect that eq. (3.99) can be written as

$$\frac{dL_{\text{source}}^i}{dt} = \langle T^i \rangle \quad (3.102)$$

where T^i is the torque associated to the force per unit volume dF_i/dV . The average $\langle \dots \rangle$ takes into account the fact that the energy and angular momentum of GWs are notions well defined only if we average over ²⁰ Gravitational waves, as any field, carry away a total angular momentum J^i that, as we saw in Section 2.1.2, is made of a spin contribution and of an orbital angular momentum contribution. This total angular momentum is drained from the total angular momentum of the source which, for a macroscopic source, is a purely orbital angular momentum.

several periods. The expression for this force, in the quadrupole approximation, can be found as follows. Since inside $\langle \dots \rangle$ we can integrate by parts (compare with Note 23 on page 35), we rewrite eq. (3.98) as

$$\frac{dE_{\text{source}}}{dt} = -\frac{G}{5c^5} \left\langle \frac{dQ_{ij}}{dt} \frac{d^5 Q_{ij}}{dt^5} \right\rangle \quad (3.103)$$

From eq. (3.58) we have

$$\frac{dQ_{ij}}{dt} = \int d^3x' \partial_t \rho(t, \mathbf{x}') \left(x'_i x'_j - \frac{1}{3} r'^2 \delta_{ij} \right) \quad (3.104)$$

The term proportional to δ_{ij} gives zero when contracted with $d^5 Q_{ij}/dt^5$. For a Newtonian source with $T^{00}/c^2 = \rho$ and $T^{0i}/c = \rho v^i$, the conservation of the energy-momentum tensor gives the continuity equation

$$\partial_t \rho + \partial_k (\rho v_k) = 0 \quad (3.105)$$

We then replace $\partial_t \rho$ by $-\partial_k (\rho v_k)$ in eq. (3.104) and we integrate by parts. The boundary term at infinity vanishes, because the body has a finite extent, so $\rho = 0$ beyond some value $r > a$. Therefore

$$\int d^3x' \partial_t \rho(t, \mathbf{x}') x'_i x'_j = \int d^3x' \rho(t, \mathbf{x}') \dot{x}'_k (\delta_{ik} x'_j + \delta_{jk} x'_i) \quad (3.106)$$

and

$$\frac{dE_{\text{source}}}{dt} = -\frac{2G}{5c^5} \left\langle \frac{d^5 Q_{ij}}{dt^5} \int d^3x' \rho(t, \mathbf{x}') \dot{x}'_i x'_j \right\rangle \quad (3.107)$$

Of course, $d^5 Q_{ij}/dt^5$ is a function only of time, and does not depend on the dummy integration variable x' . Then it can be carried inside the integral, and eq. (3.107) can be written as

$$\frac{dE_{\text{source}}}{dt} = \left\langle \int d^3x' \frac{dF_i}{dV'} \dot{x}'_i \right\rangle \quad (3.108)$$

with a force per unit volume

$$\frac{dF_i}{dV'} = -\frac{2G}{5c^5} \rho(t, \mathbf{x}') x'_j \frac{d^5 Q_{ij}}{dt^5}(t) \quad (3.109)$$

Therefore the total force is

$$F_i = -\frac{2G}{5c^5} \frac{d^5 Q_{ij}}{dt^5} \int d^3 x' \rho(t, \mathbf{x}') x'_j \quad (3.110)$$

Defining the center-of-mass coordinates by

$$x_j(t) \equiv \frac{1}{m} \int d^3 x' \rho(t, \mathbf{x}') x'_j \quad (3.111)$$

we find

$$F_i = -\frac{2G}{5c^5} m x_j \frac{d^5 Q_{ij}}{dt^5} \quad (3.112)$$

We have deduced this gravitational force on the source considering the GWs at infinity and imposing the energy balance. However, the motion of a particle under the effect of gravitational forces is completely determined by the value of the metric and its derivatives at the particle location. Thus, it should be possible to deduce the back-reaction force (3.112) also looking directly at the metric in the near-source region, without invoking the energy balance. In other words, if one solves for the gravitational field everywhere in space, in correspondence to a GW solution in the far region, there must be terms in the metric in the near region which, acting directly on the source motion through the geodesic equation, produce exactly the effect that we have inferred from the energy balance argument, i.e. provide the force (3.112). This correspondence between the GWs in the far region and the metric in the near region will be discussed in Chapter 5 in the appropriate context, which is the post-Newtonian formalism. The result is that, indeed, in the near region, among other terms that describe post-Newtonian corrections to the static potential, there is also a correction to the metric coefficient h_{00} of the form

$$\Delta h_{00} = -\frac{2\Phi}{c^2} \quad (3.113)$$

with

$$\Phi(t, \mathbf{x}) = \frac{G}{5c^5} x_i x_j \frac{d^5 Q_{ij}}{dt^5}(t) \quad (3.114)$$

This is known as the Burke-Thorne potential,²¹ and generates a Newtonian force

$$F_i = -m \partial_i \Phi \quad (3.115)$$

in agreement with eq. (3.112).²² As we will discuss in detail in Chapter 5, this term is singled out, and associated to radiation reaction, thanks to the fact that it is odd under time reversal. Terms associated with conservative forces are invariant under time reversal. In contrast, the emission of radiation breaks time-reversal invariance through the boundary conditions, since we impose that there is no incoming radiation (technically, at minus null infinity, see Note 1 on page 102), while at plus null infinity we can have outgoing radiation.

We now check that the force (3.112) also correctly accounts for the angular momentum loss. The torque on a unit volume located at the position x_i is

$$\begin{aligned}\frac{dT_i}{dV} &= \epsilon_{ijk} x_j \frac{dF_k}{dV} \\ &= -\frac{2G}{5c^5} \epsilon_{ijk} \rho(t, \mathbf{x}) x_j x_l \frac{d^5 Q_{kl}}{dt^5}\end{aligned}\quad (3.116)$$

where we used eq. (3.109). Then

$$T_i = -\frac{2G}{5c^5} \epsilon_{ijk} \frac{d^5 Q_{kl}}{dt^5} \int d^3x \rho(t, \mathbf{x}) x_j x_l \quad (3.117)$$

²¹ Higher-order corrections to the backreaction force are given in eqs. (5.190) and (5.191). When comparing these results with eq. (3.114), observe that eqs. (5.190) and (5.191) are written in terms of the variable $h_{\mu\nu}$ defined in eq. (5.69) (which in the linearized limit reduces to $-\bar{h}_{\mu\nu}$), rather than in terms of $h_{\mu\nu}$.

²² Observe that there is here an abuse of notation. In eq. (3.114), x^i is the generic spatial variable. After taking the derivative with respect to x^i in eq. (3.115), we evaluate the force on the actual location of the particle, i.e. on the position $x^i(t)$ defined by eq. (3.111), and this gives eq. (3.112). For instance, for a point-like mass μ the quadrupole moment is $Q_{ij}(t) = \mu [x_i(t)x_j(t) - (1/3)\delta_{ij}|\mathbf{x}(t)|^2]$, where $x_i(t)$ is the actual trajectory of the particle, rather than the generic spatial variable x_i . So Q_{ij} is a given function of time only, and ∂_i in eq. (3.115) does not act on it.

In the last integral we can replace $x_j x_l$ by $x_j x_l - (1/3)\delta_{jl}r^2$, since the term $\sim \delta_{jl}$ gives zero contracted with $\epsilon_{ijk} d^5 Q_{kl}/dt^5$, so

$$T_i = -\frac{2G}{5c^5} \epsilon_{ijk} Q_{jl} \frac{d^5 Q_{kl}}{dt^5} \quad (3.118)$$

In eq. (3.102) actually enters $\langle T_i \rangle$. Inside the average, we can integrate by parts twice, and we get

$$\langle T_i \rangle = -\frac{2G}{5c^5} \epsilon_{ikl} \left\langle \ddot{Q}_{ka} \ddot{Q}_{la} \right\rangle \quad (3.119)$$

Comparison with eq. (3.99) shows that we have indeed correctly reproduced the expression for the angular momentum loss of the source.

It should be observed that this is also by far the quickest way to derive eq. (3.97), without going through the formalism of Noether's theorem and the more complicated algebra of Section 3.3.3. However, the derivation from the Noether theorem is more general, since it holds independently of the multipole expansion, and of whether or not the back-reaction of GWs on the source can be described by Newtonian mechanics. It is also conceptually more satisfying, since it stresses that the angular momentum carried by GWs is an intrinsic property of the gravitational field, independent of the description of the source, that can be computed by applying the standard methods of classical field theory, as is usually done in all other field theories.

Finally, we expect that the change in linear momentum should be given by $dP^i/dt = \langle F^i \rangle$. This is proportional to $\langle x_j(t) d^5 Q_{ij}/dt^5(t) \rangle$. Since the quadrupole moment is even under parity while x_j is odd, the integrand is odd and, for a periodic motion, its average over one orbital period vanishes and therefore $\langle F_i \rangle = 0$. This is in agreement with the fact that, in the quadrupole approximation, linear momentum is conserved, see the discussion below eq. (3.83).

6.0.57 Gravitational radiation and the equivalence principle

The above results also allow us to clarify an apparent paradox related to the equivalence principle. Consider, for simplifying the setting, a mass μ orbiting a mass M , in the limit

$M/\mu \rightarrow \infty$. Thus, the light mass μ is accelerated by the gravitational field of the heavy mass M and, according to our computations, it radiates GWs (while M is static and does not radiate.) An observer at large distance from the source, well into the far region, would then in principle be able to detect the waves emitted, and would conclude that the mass μ indeed emits gravitational radiation.

Examine now the same situation from the point of view of an observer freely falling together with the mass μ . According to the equivalence principle, for this observer, in a sufficiently small region around the mass μ , the gravitational field vanishes. We have indeed seen explicitly in Section 1.3.2 how to construct such a freely falling frame all along a geodesic. In this frame the mass μ is not accelerated, and the corresponding observer should then conclude that the mass μ does not radiate, contrary to the findings of the observer at infinity.

This apparent paradox can be understood recalling that the equivalence principle holds only locally, i.e. in a region around the mass μ much smaller than the typical scales of spatial variation of the gravitational field. One such scale is the length λ , over which retardation effects become important (and which determines the wavelength of the GWs detected by the observer at infinity.) Then, conclusions based on the equivalence principle can be valid only up to a distance r from the mass μ , much smaller than λ .²³ This means that the equivalence principle at most gives us informations on what happens in the near zone $r \ll \lambda$; GWs rather appear in the far zone $r \gg \lambda$, so there is no paradox in the fact that, using arguments valid only for $r \ll \lambda$, one does not see them. The presence of gravitational radiation at infinity is reflected, in the near zone, in the existence of the force given by eqs. (3.114) and (3.115). However, in the near region retardation effects are negligible, so this term just gives a correction to the static gravitational force, which furthermore is hidden behind other, much larger, corrections. We will see in fact in Chapter 5 that, in an expansion in v/c , the radiation-reaction force is of order $(v/c)^5$ (as it is already clear from the factor $1/c^5$ in eq. (3.114)), while the Newtonian gravitational field receives generalrelativistic corrections, corresponding to conservative forces, already at orders $(v/c)^2$ and $(v/c)^4$. All these tidal-like terms, however, in the far region decrease much faster than $1/r$, leaving us with the radiation field only.

6.0.58 3.3.5 Radiation from a closed system of point masses

For a free point-like particle moving on a trajectory $x_0(t)$ in flat spacetime, the energy-momentum tensor is ²⁴

$$T^{\mu\nu}(t, \mathbf{x}) = \frac{p^\mu p^\nu}{\gamma m} \delta^{(3)}(\mathbf{x} - \mathbf{x}_0(t)) \quad (3.120)$$

where $\gamma = (1 - v^2/c^2)^{-1/2}$, and $p^\mu = \gamma m (dx_0^\mu/dt) = (E/c, \mathbf{p})$ is the fourmomentum. If we have a set of free point particles labeled by an index A , moving on trajectories $x_A^\mu(t)$, the total energy-momentum tensor of the system is therefore

$$\begin{aligned} T^{\mu\nu}(t, \mathbf{x}) &= \sum_A \frac{p_A^\mu p_A^\nu}{\gamma_A m_A} \delta^{(3)}(\mathbf{x} - \mathbf{x}_A(t)) \\ &= \sum_A \gamma_A m_A \frac{dx_A^\mu}{dt} \frac{dx_A^\nu}{dt} \delta^{(3)}(\mathbf{x} - \mathbf{x}_A(t)) \end{aligned} \quad (3.121)$$

and in particular

$$T^{00}(t, \mathbf{x}) = \sum_A \gamma_A m_A c^2 \delta^{(3)}(\mathbf{x} - \mathbf{x}_A(t)) \quad (3.122)$$

²⁵ The explicit form of these terms will be given in Section 5.3.2, when we study systematically the relativistic corrections. We will see in eqs. (5.111)(5.113) that, including the first postNewtonian correction, T^{00} must be replaced by

$$\tau^{00} = \left(1 + \frac{4V}{c^2}\right) T^{00} - \frac{7}{8\pi G} \partial_k V \partial_k V, \text{ while } T^{0i} \text{ is unchanged and } T^{ij} \text{ is replaced by}$$

$$\begin{aligned} \tau^{ij} &= T^{ij} \\ &+ \frac{1}{4\pi G} \left(\partial_i V \partial_j V - \frac{1}{2} \delta_{ij} \partial_k V \partial_k V \right) \end{aligned}$$

where V is given in eq. (5.39) and, to lowest order, reduces to the Newtonian potential. Observe that this energy-momentum tensor is conserved only if the particles follow the geodesics of flat space-time, $\dot{p}_A^\mu = 0$ for all A . Thus, a priori it is not legitimate to use it to compute the radiation emitted by a system of interacting particles moving on generic trajectories $\mathbf{x}_A(t)$. In a consistent treatment we should include in the energy-momentum tensor all the interaction terms among the particles, and possibly with external sources, that cause them to deflect from a straight-line trajectory. However, for a non-relativistic self-gravitating system it is still possible to use the energy-momentum tensor (3.121) to compute both the leading term in eq. (3.34) (i.e. the mass quadrupole radiation) as well as the next-to-leading term, i.e. the term proportional to $\dot{S}^{kl,m}$ (which, as we will discuss in more detail in Sect. 3.4, is the sum of mass octupole and current quadrupole radiation). In this way, using only linearized theory, we can obtain the correct lowest-order results that will be derived, with much more effort, with the full non-linear formalism described in Chapter 5. To this purpose we need to observe that the full energy-momentum tensor also has interaction terms responsible for binding the particles in a orbit. For a self-gravitating system these terms are $O(v^2/c^2)$, as it is clear from the fact the gravitational potential energy $-Gm_1m_2/r$ is of order v^2/c^2 .²⁵ Therefore, for a gravitationally-bound two-particle system in the non-relativistic limit,

$$T^{\mu\nu} = \sum_{A=1,2} m_A \frac{dx_A^\mu}{dt} \frac{dx_A^\nu}{dt} \delta^{(3)}(\mathbf{x} - \mathbf{x}_A(t)) + O(v^2/c^2) \quad (3.123)$$

Since $T^{00} = O(v^0)$, $T^{0i} = O(v/c)$ and $T^{ij} = O(v^2/c^2)$, T^{00} and T^{0i} can be computed consistently, to lowest order, ignoring the interaction term, while T^{ij} cannot. Observe also that the use of the lowest-order expressions

$$T^{00} = \sum_A m_A c^2 \delta^{(3)}(\mathbf{x} - \mathbf{x}_A(t)) \quad (3.124)$$

$$T^{0i} = \sum_A m_A c \dot{x}_A^i(t) \delta^{(3)}(\mathbf{x} - \mathbf{x}_A(t)) \quad (3.125)$$

is consistent with the conservation equation $\partial_0 T^{00} + \partial_i T^{0i} = 0$ since, with these expressions, $\partial_0 T^{00} = -\sum_A m_A c \dot{x}_A^i \partial_i \delta^{(3)}(\mathbf{x} - \mathbf{x}_A(t))$ and $\partial_i T^{0i} = +\sum_A m_A c \dot{x}_A^i \partial_i \delta^{(3)}(\mathbf{x} - \mathbf{x}_A(t))$. So, to lowest

²³ In fact, r must also be much smaller than the scale of spatial variation of the quasi-static tidal gravitational fields near the mass μ , which in turn is much smaller than λ .

²⁴ See, e.g. Landau and Lifshitz, Vol. II (1979), eq. (33.5), or Weinberg (1972), Section 2.8. The generalization to curved space is more easily obtained writing $p^\mu = m dx_0^\mu/d\tau$ where τ is the proper time of the particle. In flat space $c^2 d\tau^2 = -\eta_{\mu\nu} dx_0^\mu dx_0^\nu$, so $d/d\tau = \gamma d/dt$ and we get $p^\mu = \gamma m (dx_0^\mu/dt)$. In curved space, instead, $c^2 d\tau^2 = -g_{\mu\nu} dx_0^\mu dx_0^\nu$. Furthermore, in flat space $(1/\gamma)\delta^{(3)}(\mathbf{x} - \mathbf{x}_0(t)) = (d\tau/dt)\delta^{(3)}(\mathbf{x} - \mathbf{x}_0(t))$, which can be rewritten as $\int d\tau \delta^{(4)}(x - x_0(\tau))$. In curved space $\delta^{(4)}(x - x_0(\tau))$ becomes $(1/\sqrt{-g})\delta^{(4)}(x - x_0(\tau))$, so in the end in curved space $(1/\gamma)\delta^{(3)}(\mathbf{x} - \mathbf{x}_0(t))$ becomes $(d\tau/dt)(1/\sqrt{-g})\delta^{(3)}(\mathbf{x} - \mathbf{x}_0(t))$. In this way we obtain the curved-space expression given in eq. (5.47).

order $\partial_0 T^{00} = -\partial_i T^{0i}$, independently of the trajectory used. In contrast, the conservation equation $\partial_0 \tau^{0j} + \partial_i \tau^{ij} = 0$ (with τ^{ij} given in Note 25 and $\tau^{0j} = T^{0j}$ to lowest order) is only satisfied if $\mathbf{x}_A(t)$ satisfies the equation of motion in the potential V , as can be checked with the explicit computation.

Thus, if we want to compute directly S^{kl} or $\dot{S}^{kl,m}$ in eq. (3.34), we need T^{ij} , and therefore we need to include also the interaction terms, which are $O(v^2/c^2)$, just as the leading term in the T^{ij} components of the freeparticle energy-momentum tensor. However, using energy-momentum conservation, we can transform S^{kl} into $\ddot{M}^{kl}$, see eq. (3.52). Similarly, using eq. (3.54), we can trade $\dot{S}^{ij,k}$ for $\ddot{M}^{ijk}$ and $\ddot{P}^{i,jk}$. The derivation of eq. (3.54) uses the conservation of energy-momentum tensor so, when

we write the GW amplitude in terms of $\ddot{M}^{ijk}$ and $\ddot{P}^{i,jk}$, we are already implicitly using the correct $T^{\mu\nu}$, with all the necessary interaction terms. The important point is that, to trade $\dot{S}^{ij,k}$ for $\ddot{M}^{ijk}$ and $\ddot{P}^{i,jk}$, we do not need to know the explicit form of $T^{\mu\nu}$, including the interaction terms, but only that it satisfies energy-momentum conservation. The advantage of this procedure is that M^{ij} and M^{ijk} only depend on T^{00} , and $P^{i,jk}$ only depends on T^{0i} . Therefore, to lowest order they can be consistently computed neglecting the potential terms in $T^{\mu\nu}$. In conclusion, both the leading term (i.e. the quadrupole radiation) and the next-to-leading term (i.e. the mass octupole plus current quadrupole radiation) can be consistently computed using eq. (3.124) to evaluate M^{ij} and M^{ijk} , and eq. (3.125) to evaluate $P^{i,jk}$, and then using eqs. (3.52) and (3.54) to evaluate S^{ij} and $\dot{S}^{ij,k}$. In contrast, if one wants to compute S^{ij} and $\dot{S}^{ij,k}$ directly from T^{ij} , which is $O(v^2)$, even to lowest order one needs to include the interaction terms, and one cannot use the free-particle energy-momentum tensor.

As will be shown in Section 5.1, the relativistic corrections to the Newtonian orbit start from order v^2/c^2 . Therefore, the computation of the GW amplitude to leading and next-to-leading order in v/c can be performed evaluating the components T^{00} and T^{0i} of free-particle energy momentum tensor on the Newtonian orbit, and using them to compute M^{ij} , M^{ijk} and $P^{i,jk}$. We will perform these computations explicitly in Problems 3.2 and 3.3, in the Solved Problems section. Observe however that, in the radiated power, the corrections to the quadrupole amplitude give corrections to the leading term in the power of order $|1 + O(v^2/c^2)|^2 = 1 + O(v^2/c^2)$, while the mass octupole and current quadrupole give a correction $|O(v/c)|^2$, which is again $O(v^2/c^2)$, see the discussion below eq. (3.156).

For a non-relativistic two-body system it is convenient to define as usual the relative coordinate $\mathbf{x}_0 = \mathbf{x}_1 - \mathbf{x}_2$ and the center-of-mass coordinate

$$\mathbf{x}_{\text{CM}} = \frac{m_1 \mathbf{x}_1 + m_2 \mathbf{x}_2}{m_1 + m_2} \quad (3.126)$$

(Starting from $O(v^2/c^2)$ this must actually be replaced by a center-of-energy, which also receives contributions from the interaction potential). We also denote by $m = m_1 + m_2$ the total mass and by $\mu = m_1 m_2 / m$ the reduced mass. For a non-relativistic system, the second mass moment can be written as

$$M^{ij} = m_1 x_1^i x_1^j + m_2 x_2^i x_2^j = m x_{\text{CM}}^i x_{\text{CM}}^j + \mu x_0^i x_0^j \quad (3.127)$$

If the system is isolated, x_{CM}^i is not accelerating and does not contribute to the GW production. It is then convenient to choose the frame such that $x_{\text{CM}}^i = 0$, and we are left with a single effective particle of mass μ and coordinate $x_0^i(t)$. In the CM frame, the mass density is then

$$\rho(t, \mathbf{x}) = \mu \delta^{(3)}(\mathbf{x} - \mathbf{x}_0(t)) \quad (3.128)$$

the second mass moment is

$$M^{ij}(t) = \mu x_0^i(t) x_0^j(t) \quad (3.129)$$

²⁶ In a one-dimensional motion this example would be quite unrealistic, since the two bodies would go through each other whenever $\cos \omega_s t = 0$. However, this is just an example to illustrate what happens to a typical oscillatory mode of a system. For instance, one can consider an elliptic motion on a plane, which is the combination of two simple oscillations along the two axes. and the mass quadrupole is

$$Q^{ij}(t) = \mu \left(x_0^i(t) x_0^j(t) - \frac{1}{3} r_0^2(t) \delta^{ij} \right) \quad (3.130)$$

We can now study the radiation emitted by a two-body system whose relative coordinate performs a given periodic motion, say simple harmonic oscillations. Suppose that the relative coordinate $\mathbf{x}_0(t)$ performs a simple periodic motion with frequency ω_s , say along the z direction,

$$z_0(t) = a_1 \cos \omega_s t \quad (3.131)$$

Then

$$\begin{aligned} M^{ij}(t) &= \delta^{i3} \delta^{j3} \mu z_0^2(t) \\ &= \delta^{i3} \delta^{j3} \frac{\mu a_1^2}{2} (1 + \cos 2\omega_s t) \end{aligned} \quad (3.132)$$

Since the GW amplitude depends on the second derivative of M^{ij} , the constant term does not contribute and the only contribution to h_{ij}^{TT} comes from the term proportional to $\cos 2\omega_s t$. From eq. (3.55), we see that the corresponding waveform h_{ij}^{TT} oscillates as $\cos 2\omega_s t$. This shows that a non-relativistic source performing simple harmonic oscillations with a frequency ω_s emits monochromatic quadrupole radiation at $\omega = 2\omega_s$.

However, the fact that the quadrupole radiation is at twice the source frequency is only true if the source performs a simple harmonic motion. For instance, if the motion of the source is a superposition of a periodic motion and of its higher harmonics, e.g. if

$$z_0(t) = a_1 \cos \omega_s t + a_2 \cos 2\omega_s t + \dots \quad (3.133)$$

then $z_0^2(t)$ contains the term

$$a_1^2 \cos^2 \omega_s t = a_1^2 \frac{1 + \cos 2\omega_s t}{2} \quad (3.134)$$

considered above, plus a term

$$a_2^2 \cos^2 2\omega_s t = a_2^2 \frac{1 + \cos 4\omega_s t}{2} \quad (3.135)$$

which gives radiation at $\omega_{\text{gw}} = 4\omega_s$, etc., but also mixed terms such as

$$2a_1 a_2 \cos(\omega_s t) \cos(2\omega_s t) = a_1 a_2 (\cos \omega_s t + \cos 3\omega_s t) \quad (3.136)$$

Therefore in this case quadrupole radiation is emitted at all frequencies $n\omega_s$ for all integers n , both even and odd, including $n = 1$. We will see an example of this type in Section 4.1.2, when we study the radiation emitted from a mass in a Keplerian elliptic orbit.

An even simpler example is given by a system of two masses connected by a spring with rest length L (see Problem 3.1). In this case the relative coordinate satisfies

$$z_0(t) = L + a \cos \omega_s t \quad (3.137)$$

and in $z_0^2(t)$, besides a constant and a term $(a^2/2) \cos 2\omega_s t$, we also have a term $2La \cos \omega_s t$, so the spectrum of gravitational radiation has two lines, one at $\omega = \omega_s$, and one at $\omega = 2\omega_s$.

6.0.59 3.4 Mass octupole and current quadrupole

We now examine the next-to-leading term of the expansion (3.34),

$$(h_{ij}^{\text{TT}})_{\text{next-to-leading}} = \frac{1}{r} \frac{4G}{c^5} \Lambda_{ij,kl}(\hat{\mathbf{n}}) n_m \dot{S}^{kl,m}(t - r/c) \quad (3.138)$$

By definition $\dot{S}^{kl,m}$ is symmetric in kl , and it has no special symmetry with respect to the exchange of k with m . We have seen in eq. (3.54) that

$$\dot{S}^{kl,m} = \frac{1}{6} \ddot{M}^{klm} + \frac{1}{3} \left(\ddot{P}^{k,lm} + \ddot{P}^{l,km} - 2\ddot{P}^{m,kl} \right) \quad (3.139)$$

Therefore $\dot{S}^{kl,m}$ separates into a totally symmetric tensor $\ddot{M}^{klm}$ plus a tensor with mixed symmetry. The meaning of this separation from the point of view of group theory is explained in Problem 3.4, in the Solved Problems section. The totally symmetric term generates mass octupole radiation, while the term with mixed symmetry is called the current quadrupole.

6.0.60 Mass octupole

We consider first the mass octupole radiation. The mass octupole $\mathcal{O}^{klm}$ is defined removing all traces from M^{ijk} ,

$$\mathcal{O}^{klm} = M^{klm} - \frac{1}{5} \left(\delta^{kl} M^{k'k'm} + \delta^{km} M^{k'lk'} + \delta^{lm} M^{kk'k'} \right) \quad (3.140)$$

Using the fact that $\Lambda_{ij,kl}(\hat{\mathbf{n}})$ is transverse and traceless, we see that the contraction of the trace part with $\Lambda_{ij,kl}(\hat{\mathbf{n}}) n_m$ gives zero, and we can use M^{klm} or $\mathcal{O}^{klm}$ interchangeably in the expression for h_{ij}^{TT} . Therefore the mass octupole contribution to h_{ij}^{TT} can be written as

$$(h_{ij}^{\text{TT}})_{\text{oct}} = \frac{1}{r} \frac{2G}{3c^5} \Lambda_{ij,kl}(\hat{\mathbf{n}}) n_m \ddot{\mathcal{O}}^{klm} \quad (3.141)$$

Similarly to the case of quadrupole radiation, the use of $\mathcal{O}^{klm}$ is nicer from a group-theoretical point of view, since it is a pure spin-3 tensor, see Problem 3.4, while the use of M^{ijk} is simpler in actual computations. We will use $\ddot{M}^{klm}$ and $\ddot{\mathcal{O}}^{klm}$ interchangeably in eq. (3.141).

Observe that, when we consider quantities quadratic in h_{ij}^{TT} , as for instance the radiated energy, there is no interference between the mass quadrupole and the mass octupole terms because they have different parity. Under a parity operation, $\mathbf{x} \rightarrow -\mathbf{x}$, the mass density is a true scalar, and therefore the quadrupole is invariant while the octupole changes sign (for the same reason, in electrodynamics there is no interference between dipole and quadrupole radiation.) More generally, we will see in Section 3.5 how to systematically organize the multipole expansion so that there are no interference terms to all orders.

Comparing the mass quadrupole and the mass octupole we see that, while the contribution to the GW amplitude from the mass quadrupole is proportional to the second time derivative of M^{ij} , the contribution from the mass octupole is proportional to $(1/c)$ times the third time derivative of M^{ijk} . If d is the typical dimension of the source, the tensor M^{ijk} differ from M^{ij} by a factor $O(d)$. Each time derivative carries a

factor $O(\omega_s)$, where ω_s is the typical frequency of the movement of matter inside the source, so $(1/c)\ddot{M}^{ijk} = O(\omega_s d/c) \ddot{M}^{ij}$. Since $\omega_s d \sim v$ is the typical velocity inside the source, the octupole contribution to h_{ij}^{TT} is smaller than the quadrupole contribution by a factor $O(v/c)$, and the contribution of the mass octupole to the radiated power is smaller than the contribution of the mass quadrupole by a factor $O(v^2/c^2)$.

Consider now a two-body non-relativistic system whose relative coordinate in the center-of-mass frame is described by a function $x_0^i(t) = x_1^i(t) - x_2^i(t)$, with a reduced mass μ , and define $\delta m = m_2 - m_1$. Then (as we show in Problem 3.3) to lowest order in v/c

$$M^{ijk}(t) = \mu \frac{\delta m}{m} x_0^i(t) x_0^j(t) x_0^k(t) \quad (3.142)$$

Suppose now that $x_0(t)$ performs simple harmonic oscillations with frequency ω_s . Then, each factor $x_0^i(t)$ is the superposition of a term oscillating as $e^{i\omega_s t}$ and of a term oscillating as $e^{-i\omega_s t}$, so M^{ijk} is the sum of terms oscillating as $e^{\pm i\omega_s t}$ and of terms oscillating as $e^{\pm 3i\omega_s t}$. Therefore a source performing simple harmonic oscillations at a frequency ω_s emits octupole radiation at $\omega = \omega_s$ and at $\omega = 3\omega_s$.

If a non-relativistic source performs simple harmonic oscillations, then its energy spectrum is made of a stronger line due to quadrupole radiation at $\omega = 2\omega_s$, plus two smaller "lateral bands" due to the octupole at $\omega = \omega_s$ and at $\omega = 3\omega_s$, as we will see in more detail in Problem 3.3. Recall however from Section 3.3.5 that periodic trajectories that are not simple harmonic motions produce a more complicated spectrum, in which the quadrupole can already contribute to all frequencies $n\omega_s$, for all integer n .

The power emitted per unit solid angle by the octupole moment is computed inserting $\dot{h}_{ij}^{\text{TT}}$, obtained from eq. (3.141), into the expression for the power given in eq. (1.153). This gives

$$\begin{aligned} P_{\text{oct}} &= \frac{c^3 r^2}{32\pi G} \int d\Omega \left\langle \left(\dot{h}_{ij}^{\text{TT}} \right)_{\text{oct}} \left(\dot{h}_{ij}^{\text{TT}} \right)_{\text{oct}} \right\rangle \\ &= \frac{c^3}{32\pi G} \frac{4G^2}{9c^{10}} \left\langle \frac{d^4 \mathcal{O}^{klm}}{dt^4} \frac{d^4 \mathcal{O}^{k'l'm'}}{dt^4} \right\rangle \int d\Omega \Lambda_{ij,kl} \Lambda_{ij,k'l'} n_m n_{m'} \end{aligned} \quad (3.143)$$

Using the property (1.37) of the Lambda tensor, together with $\Lambda_{ij,kl} = \Lambda_{kl,ij}$, we have $\Lambda_{ij,kl} \Lambda_{ij,k'l'} = \Lambda_{kl,k'l'}$. Then, renaming the indices, we get

$$P_{\text{oct}} = \frac{G}{72\pi c^7} \left\langle \frac{d^4 \mathcal{O}^{ijm}}{dt^4} \frac{d^4 \mathcal{O}^{klp}}{dt^4} \right\rangle \int d\Omega \Lambda_{ij,kl}(\hat{\mathbf{n}}) n_m n_p \quad (3.144)$$

To integrate over the solid angle we need the identities (3.21) and (3.22), together with eq. (3.23) with $l = 3$,

$$\int \frac{d\Omega}{4\pi} n_i n_j n_k n_l n_m n_p = \frac{1}{105} (\delta_{ij} \delta_{kl} \delta_{mp} + \dots) \quad (3.145)$$

where the dots denote the other 14 possible pairings of indices. Using the fact that $\mathcal{O}^{ijk}$ is totally symmetric and traceless, the contractions are straightforward, and we get

$$P_{\text{oct}} = \frac{G}{189c^7} \left\langle \frac{d^4 \mathcal{O}^{ijk}}{dt^4} \frac{d^4 \mathcal{O}^{ijk}}{dt^4} \right\rangle \quad (3.146)$$

6.0.61 Current quadrupole

The current quadrupole is given by the second term in eq. (3.139). Its physical meaning can be understood observing that, from the definition (3.41)

$$\begin{aligned}
& P^{k,lm} + P^{l,km} - 2P^{m,kl} \\
&= \frac{1}{c} \int d^3x [T^{0k} x^l x^m + T^{0l} x^k x^m - 2T^{0m} x^k x^l] \\
&= \frac{1}{c} \int d^3x [x^l (x^m T^{0k} - x^k T^{0m}) + x^k (x^m T^{0l} - x^l T^{0m})] \\
&= \int d^3x [x^l j^{mk} + x^k j^{ml}]
\end{aligned} \tag{3.147}$$

where

$$j^{jk} = \frac{1}{c} (x^j T^{0k} - x^k T^{0j}) \tag{3.148}$$

This is the angular momentum density associated to the (j, k) plane. We write $j^{jk} = \epsilon^{jkl} j^l$, where j^l is the l -th component of the angular momentum density vector, and we define

$$J^{i,j} = \int d^3x j^i x^j \tag{3.149}$$

so $J^{i,j}$ is the first moment of the angular momentum density (i.e. the dipole moment of the angular momentum distribution). Then we get

$$P^{k,lm} + P^{l,km} - 2P^{m,kl} = \epsilon^{mkp} J^{p,l} + \epsilon^{mlp} J^{p,k} \tag{3.150}$$

Therefore the current quadrupole is the first moment of the angular momentum density of the source, symmetrized over the indices k, l . Its contribution to the GW amplitude can be written as

$$(h_{ij}^{\text{TT}})_{\text{curr quad}} = \frac{1}{r} \frac{4G}{3c^5} \Lambda_{ij,kl}(\hat{\mathbf{n}}) n_m \left(\epsilon^{mkp} \ddot{J}^{p,l} + \epsilon^{mlp} \ddot{J}^{p,k} \right) \tag{3.151}$$

We will indeed see in the next section that the full multipole expansion can be organized systematically so that it is determined by two types of objects: the momenta of the energy density of the source (which, to leading order in v/c , are the same as the momenta of the mass density), like the mass quadrupole Q^{ij} , the mass octupole $\mathcal{O}^{ijk}$, etc., and the momenta of the angular momentum density of the source, such as the current quadrupole.

The power associated to the current quadrupole is computed just as we have done for the mass quadrupole and mass octupole: we write

$$\begin{aligned}
P_{\text{curr quad}} &= \frac{c^3 r^2}{32\pi G} \int d\Omega \left\langle \left(\dot{h}_{ij}^{\text{TT}} \right)_{\text{curr quad}} \left(\dot{h}_{ij}^{\text{TT}} \right)_{\text{curr quad}} \right\rangle \\
&= \frac{c^3}{32\pi G} \frac{16G^2}{9c^{10}} \int d\Omega \Lambda_{kl,k'l'} n_m n_{m'} \\
&\quad \times \left\langle \left(\epsilon^{mkp} \ddot{J}^{p,l} + \epsilon^{mlp} \ddot{J}^{p,k} \right) \left(\epsilon^{m'k'p'} \ddot{J}^{p',l'} + \epsilon^{m'l'p'} \ddot{J}^{p',k'} \right) \right\rangle
\end{aligned} \tag{3.152}$$

where we used again $\Lambda_{ij,kl} \Lambda_{ij,k'l'} = \Lambda_{kl,k'l'}$. The angular integration is performed using the identities (3.21) and (3.22). Observe also that the term in the Lambda tensor with four factors of n , i.e. $n_k n_l n_{k'} n_{l'}$ (see eq. (1.39)) does not contribute since, together with the factors $n_m n_{m'}$ in eq. (3.152), it gives a totally symmetric tensor $n_k n_l n_m n_{k'} n_{l'} n_{m'}$, which vanishes upon contraction with ϵ^{mkp} or ϵ^{mlp} .

When performing contractions, we also make use of the fact that $J^{i,j}$ is traceless, since

$$\begin{aligned}
J^{i,i} &= \int d^3x x^i j^i \\
&= \frac{1}{c} \int d^3x x^i \epsilon^{ijk} x^j T^{0k} = 0
\end{aligned} \tag{3.153}$$

(the sum over i is understood). Then, we get

$$P_{\text{curr quad}} = \frac{16G}{45c^7} \langle \ddot{\mathcal{J}}^{ij} \ddot{\mathcal{J}}^{ij} \rangle \tag{3.154}$$

where we introduced the traceless symmetric matrix

$$\mathcal{J}^{ij} \equiv \frac{J^{i,j} + J^{j,i}}{2} \tag{3.155}$$

that is, the symmetrization of the dipole moment of the angular momentum density. Putting together the power radiated by the mass quadrupole, current quadrupole and mass octupole, we therefore get

$$\begin{aligned}
P &= \frac{G}{c^5} \left[\frac{1}{5} \langle \ddot{Q}_{ij} \ddot{Q}_{ij} \rangle + \frac{1}{c^2} \frac{16}{45} \langle \ddot{\mathcal{J}}_{ij} \ddot{\mathcal{J}}_{ij} \rangle \right. \\
&\quad \left. + \frac{1}{c^2} \frac{1}{189} \left\langle \frac{d^4 \mathcal{O}_{ijk}}{dt^4} \frac{d^4 \mathcal{O}_{ijk}}{dt^4} \right\rangle + O\left(\frac{v^4}{c^4}\right) \right]
\end{aligned} \tag{3.156}$$

where $O(v^4/c^4)$ denotes the contributions from higher orders in the multipole expansion. In this formula, one must be careful to include in Q_{ij} also its first corrections $O(v^2/c^2)$, since it gives a contribution to the power of the same order as that due to the mass octupole and current quadrupole. This means that Q_{ij} here is not simply the quadrupole of the mass distribution. Rather, we must go back to its original definition in terms of T^{00} , see eq. (3.37) and include in T^{00} not only the terms of order ρc^2 , where ρ is the mass density, but also the terms of order ρv^2 . Furthermore, the time-dependence of the leading term must be computed using the relativistic equations of motion, including corrections up to $O(v^2/c^2)$.

It is quite difficult to imagine a realistic physical system for which the time derivative of the mass quadrupole vanishes, but still the mass octupole is time-varying. For this reason, the mass octupole is always a correction to the leading term. However, there are interesting physical situations where both the mass quadrupole and the mass octupole are constant in time and therefore do not contribute to GW production, while the current quadrupole still contributes, so it becomes the leading term. To have M^{ij} and M^{ijk} constant, in fact, it suffices that the energy density T^{00} be constant. Then all the mass momenta are constant, see e.g. eqs. (3.35)-(3.38). Still, the angular momentum density and its momenta, such as its dipole $J^{i,j}$, are not necessarily constant. Consider for example a ring in the (x, y) plane with a uniform mass density ρ , rotating around the z axis. Even if any single volume element of the ring rotates, this rotation does not induce any temporal variation in the density ρ since the ring is spatially uniform, so all mass moments are constant.²⁷ However, this ring has non-vanishing angular momentum along z . To obtain a system with a dipole moment of angular momentum $J^{i,j}$, we can simply take two rings, both rotating around the z axis, but one counterclockwise and one clockwise, as in Fig. 3.3. The upper ring has a positive J^z while the lower one has a negative J^z , so it is clear that the whole system has a non-vanishing dipole moment $J^{z,z}$.²⁸

If the rotational velocities of the rings are not constant, this dipole moment of the angular momentum is time-dependent. A physical example of a system of this type is provided by

the torsional oscillations of a neutron star. It is possible that, either because of some external perturbation, or because of a "corequake", the upper hemisphere of a NS suffers a clockwise torsion while the lower hemisphere receives a counterclockwise torsion (with the equator which, by symmetry reasons, stays fixed). Then the two hemispheres will start oscillating back and forth in opposite directions, so that when one rotates clockwise the other rotates counterclockwise, and vice versa. This system will then emit current quadrupole radiation, but not mass quadrupole nor mass octupole radiation. Another important example of this type is the r -mode of neutron stars, which will be discussed in Vol. 2.

It is therefore worthwhile to study the current quadrupole radiation in more detail, and derive its angular distribution. First, we compute the amplitude of the GWs radiated in the z direction. This is quite simple, since we just have to substitute $n = (0, 0, 1)$ into eq. (3.151). Using the explicit expression for the Lambda tensor, eq. (1.39), and recalling that, for a wave traveling along z , $h_{11} = h_+$ and $h_{12} = h_\times$, we obtain

$$h_+ = \frac{1}{r} \frac{4G}{3c^5} \left(\ddot{J}_{1,2} + \ddot{J}_{2,1} \right) \quad (3.157)$$

$$h_\times = \frac{1}{r} \frac{4G}{3c^5} \left(\ddot{J}_{2,2} - \ddot{J}_{1,1} \right) \quad (3.158)$$

²⁷ One might wonder how it is possible that there is no quadrupole radiation from the whole ring, given that every single mass element is in rotation and therefore radiates GWs. From this "microscopic" point of view, the answer is that the total amplitude is the sum over the contribution of all the mass elements, and these contributions interfere destructively, so that the total GW amplitude h_{ij} vanishes.

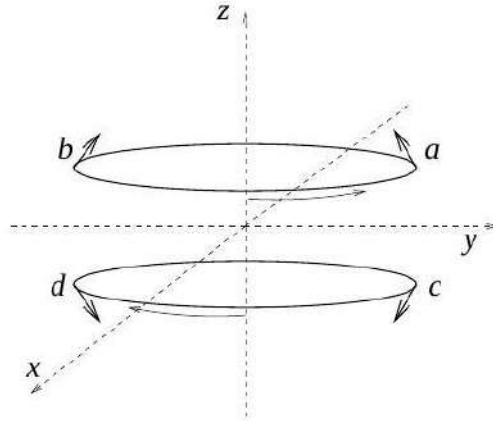


Fig. 3.3 Two rings, both with uniform mass density, one rotating counterclockwise and one clockwise. The arrows show the direction of the angular momentum $\delta \mathbf{J} = (\delta m) \mathbf{r} \times \mathbf{v}$ (with $\mathbf{r}$ measured from the origin of the coordinate system, rather than from the center of each disk) of four infinitesimal mass elements, labeled a, b, c, d . ²⁹ Comparing with eq. (3.72) we see that the angular dependence is the same as for the mass quadrupole, if we replace M_{ij} by $J_{i,j}$, and we exchange the roles of h_+ and $h_\times$ (with an additional minus sign for $(h_\times)_{\text{curr quad}}$). This was already clear from a comparison of eqs. (3.157) and (3.158) with eqs. (3.65) and (3.66), and

²⁸ It also has a non-zero $J^{x,x}$ and $J^{y,y}$ (consistent with the condition $J^{i,i} = 0$). This can be seen by drawing the direction of the angular momentum vector $\delta \mathbf{J} = (\delta m) \mathbf{r} \times \mathbf{v}$ of various mass elements, where $\mathbf{r}$ is measured from the origin of the coordinate system, see Fig. 3.3. We see for instance that both the infinitesimal mass elements labeled a and c have $J^y < 0$ and are at a coordinate $y > 0$, therefore their contribution to $J^{y,y}$ is negative. Similarly, for the mass elements b and d we have $J^y > 0$, but they have a coordinate $y < 0$, so their contribution to $J^{y,y}$ is again negative.

from the fact that both M_{ij} and $J_{i,j}$ transform as rank-2 tensors under rotations. To obtain the amplitudes h_+ and $h_\times$ for a GW propagating into an arbitrary direction, we proceed as we did for the mass quadrupole, between eqs. (3.67) and (3.72). We first write

$$h_+ = \frac{1}{r} \frac{4G}{3c^5} \left(\ddot{J}'_{1,2} + \ddot{J}'_{2,1} \right) \quad (3.159)$$

$$h_\times = \frac{1}{r} \frac{4G}{3c^5} \left(\ddot{J}'_{2,2} - \ddot{J}'_{1,1} \right) \quad (3.160)$$

where $J'_{i,j}$ are the components of the angular momentum dipole in the $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{n}})$ frame defined in Fig. 3.2. Since $J_{i,j}$ is a spatial tensor with two indices, it transforms under rotation just as we found for M_{ij} in eq. (3.71), i.e. $J'_{i,j} = \mathcal{R}_{ki} \mathcal{R}_{lj} J_{k,l}$, with $\mathcal{R}$ given by eq. (3.70). Performing the matrix multiplication, we get

$$\begin{aligned} h_+(t; \theta, \phi)|_{\text{curr quad}} &= \frac{1}{r} \frac{4G}{3c^5} \left[\left(\ddot{J}_{1,1} - \ddot{J}_{2,2} \right) \sin 2\phi \cos \theta \right. \\ &+ \left(\ddot{J}_{1,2} + \ddot{J}_{2,1} \right) \cos 2\phi \cos \theta \\ &- \left(\ddot{J}_{1,3} + \ddot{J}_{3,1} \right) \cos \phi \sin \theta \\ &\left. + \left(\ddot{J}_{2,3} + \ddot{J}_{3,2} \right) \sin \phi \sin \theta \right] \end{aligned} \quad (3.161)$$

$$\begin{aligned} h_\times(t; \theta, \phi)|_{\text{curr quad}} &= \frac{1}{r} \frac{4G}{3c^5} \left[-\ddot{J}_{1,1} (\cos^2 \phi - \sin^2 \phi \cos^2 \theta) \right. \\ &- \ddot{J}_{2,2} (\sin^2 \phi - \cos^2 \phi \cos^2 \theta) \\ &+ \ddot{J}_{3,3} \sin^2 \theta \\ &+ \left(\ddot{J}_{1,2} + \ddot{J}_{2,1} \right) \sin \phi \cos \phi (1 + \cos^2 \theta) \\ &- \left(\ddot{J}_{1,3} + \ddot{J}_{3,1} \right) \sin \phi \sin \theta \cos \theta \\ &\left. - \left(\ddot{J}_{2,3} + \ddot{J}_{3,2} \right) \cos \phi \sin \theta \cos \theta \right] \end{aligned} \quad (3.162)$$

This gives the angular distribution of the current quadrupole radiation,²⁹ for $J^{i,j}$ arbitrary (but satisfying the zero-trace condition $J_{1,1} + J_{2,2} + J_{3,3} = 0$). To check these equations we can plug them into the expression (1.154) for the angular distribution of the power in terms of h_+ and $h_\times$,

$$\frac{dP}{d\Omega} = \frac{c^3 r^2}{16\pi G} \left\langle \dot{h}_+^2 + \dot{h}_\times^2 \right\rangle \quad (3.163)$$

Performing explicitly the integral over $d\Omega = d\phi d\theta |\sin \theta|$, we verify that we get eq. (3.154) back.

6.0.62 Linear momentum losses

Finally, it is interesting to observe that the leading term contributing to the loss of linear momentum comes from the interference between the quadrupole term and the next-to-leading term (i.e. the octupole plus current quadrupole). In fact, recall from eq. (1.164) that dP^k/dt is proportional to

$$\int d\Omega \dot{h}_{ij}^{\text{TT}} \partial_k h_{ij}^{\text{TT}} \quad (3.164)$$

Using eqs. (3.59) and (3.138) we write

$$\begin{aligned} h_{ij}^{\text{TT}} &= [h_{ij}^{\text{TT}}]_{\text{quad}} + [h_{ij}^{\text{TT}}]_{\text{next-to-leading}} \\ &= \frac{1}{r} \frac{2G}{c^4} \left[\ddot{Q}_{ij}^{\text{TT}} + \frac{2}{c} n_m \dot{S}_{ij,m}^{\text{TT}} \right] \end{aligned} \quad (3.165)$$

where $\dot{S}_{ij,m}^{\text{TT}} = \Lambda_{ij,kl} \dot{S}_{kl,m}$. In the product $\dot{h}_{ij}^{\text{TT}} \partial_k h_{ij}^{\text{TT}}$ we have diagonal terms and interference terms between the quadrupole and the next-to-leading term. The diagonal terms are proportional to

$$\int d\Omega \ddot{Q}_{ij}^{\text{TT}} \partial_k \ddot{Q}_{ij}^{\text{TT}} \quad (3.166)$$

for the quadrupole, and to

$$\int d\Omega \left(n_l \ddot{S}_{ij,l}^{\text{TT}} \right) \partial_k \left(n_m \dot{S}_{ij,m}^{\text{TT}} \right) \quad (3.167)$$

for the next-to-leading term. Because of parity, these angular integrals vanish if the integrand is the product of an odd number of factors n_i . Therefore, we need to count the number of $\hat{\mathbf{n}}$ in these expressions. The TT projection is performed with the Lambda tensor, which has an even number of factors $\hat{\mathbf{n}}$. As for the derivative ∂_k which appears in eqs. (3.166) and (3.167), recall that, on a function of r , $\partial_k f(r) = (\partial_k r) df/dr = n^k df/dr$, while $\partial_k n^i = \partial_k (x^i/r) = (1/r) (\delta^{ik} - n^i n^k)$. Therefore the effect of ∂_k is always to lower or to raise by one the number of factors n^i , i.e. to transform a term with an even number of $\hat{\mathbf{n}}$ in a term with an odd number, and vice versa.

Then, we see that the diagonal terms in eqs. (3.166) and (3.167) have an odd number of factors $\hat{\mathbf{n}}$, so their angular integral vanish. In the interference terms, such as

$$\int d\Omega \ddot{Q}_{ij}^{\text{TT}} \partial_k \left(n_m \dot{S}_{ij,m}^{\text{TT}} \right) \quad (3.168)$$

the integrand is even in $\hat{\mathbf{n}}$ and in general is non-vanishing. So, while the radiated energy only gets contributions from the diagonal terms (as we will verify to all orders in the multipole expansion in the next section), the radiated momentum only gets contributions from the interference between multipoles of different parity, in order to compensate for the minus sign acquired by the derivative ∂_k under $\hat{\mathbf{n}} \rightarrow -\hat{\mathbf{n}}$.

6.0.63 3.5 Systematic multipole expansion

In eq. (3.34) the multipole expansion has been organized in terms of tensors such as S^{kl} , $\dot{S}^{kl,m}$, $\ddot{S}^{kl,m_1 m_2}$, etc. which have two sets of indices (separated by a comma), the first always made by two indices k, l and the second made by a generic number of indices, $m_1, \dots, m_N$. These tensors are symmetric under the exchange of k and l , and are also symmetric under the exchange of indices in the $m_1, \dots, m_N$ set. However, they have no special symmetry property under the exchange of indices between the two sets (and they are in general not traceless under contraction of pairs of indices). At next-to-leading order, we separated by hand the term $\dot{S}^{kl,m}$ into two contributions, one corresponding to mass octupole and one to current quadrupole. The reason underlying this separation is group-theoretical. As we discuss in Problem 3.4, this corresponds to a decomposition into irreducible representations of the rotation group. To generalize such a construction to arbitrary orders in the multipole expansion, we must introduce a complete set of representations of the rotation group, for all

multipoles. There are two particularly convenient ways of doing that. One is to consider tensors which are symmetric and trace-free (STF) with respect to all pairs of indices. This formalism will be introduced in Section 3.5.1, and we will use it extensively in particular in Chapter 5. A second possibility is to introduce the spherical components of tensors and the tensor spherical harmonics, which is the approach that we discuss in Section 3.5.2

To illustrate these two different approaches, it can be useful to begin our discussion by recalling how the multipole expansion works in the simpler case of a static situation, governed by a Poisson equation of the form

$$\nabla^2 \phi = -4\pi\rho \quad (3.169)$$

which would be the case, e.g. in electrostatics or in Newtonian gravity. We consider a stationary source with density $\rho(\mathbf{x})$ localized in space, so $\rho(\mathbf{x}) = 0$ if $r > d$, where d is the source size. The most general solution in the external region $r > d$ can be written as

$$\phi(\mathbf{x}) = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^l \frac{Q_{lm}}{2l+1} \frac{Y_{lm}(\theta, \phi)}{r^{l+1}} \quad (3.170)$$

In fact, for $r > 0$,

$$\nabla^2 \left[\frac{Y_{lm}(\theta, \phi)}{r^{l+1}} \right] = 0 \quad (3.171)$$

as can be seen immediately from the expression of the Laplacian in spherical coordinates,

$$\nabla^2 \psi(\mathbf{x}) = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) - \frac{\mathbf{L}^2}{r^2} \quad (3.172)$$

together with the property $\mathbf{L}^2 Y_{lm} = l(l+1)Y_{lm}$ of the spherical harmonics. Then, in the external region $r > d$, all terms in the sum in eq. (3.170) are separately solution of the vacuum equation $\nabla^2 \phi = 0$. The fact that eq. (3.170) is the most general vacuum solution follows from the fact that the spherical harmonics provide a complete set of representations of the rotation group. On the other hand, the solution of eq. (3.169) (subject to the boundary condition that ϕ approaches zero as $r \rightarrow \infty$) can be written in terms of the Green's function of the Laplacian, as

$$\phi(\mathbf{x}) = \int d^3y \frac{1}{|\mathbf{x} - \mathbf{y}|} \rho(\mathbf{y}) \quad (3.173)$$

which is valid both for $\mathbf{x}$ inside and outside the source. Outside the source we have $|\mathbf{x}| > |\mathbf{y}|$, and we can use the addition theorem for spherical harmonics³⁰ in the form

$$\frac{1}{|\mathbf{x} - \mathbf{y}|} = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^l \frac{1}{2l+1} \frac{r'^l}{r^{l+1}} Y_{lm}^*(\theta', \phi') Y_{lm}(\theta, \phi) \quad (3.174)$$

where we used the notation $|\mathbf{x}| = r$ and $|\mathbf{y}| = r'$; (θ, ϕ) are the polar angles of $\hat{\mathbf{x}}$ and (θ', ϕ') of $\hat{\mathbf{y}}$. Inserting this identity in eq. (3.173) and comparing with eq. (3.170) we find the expression for the multipole coefficients Q_{lm} in terms of the source density $\rho(\mathbf{x})$,

$$Q_{lm} = \int d^3y Y_{lm}^*(\theta', \phi') r'^l \rho(\mathbf{y}) \quad (3.175)$$

So, eq. (3.170) gives the most general solution of the vacuum equation, and eq. (3.175) fixes the coefficients Q_{lm} in terms of the density of the source.

An alternative way of performing the multipole expansion is to write

$$\begin{aligned}
\frac{1}{|\mathbf{x} - \mathbf{y}|} &= \frac{1}{|\mathbf{x}|} - y^i \partial_i \frac{1}{|\mathbf{x}|} + \frac{1}{2} y^i y^j \partial_i \partial_j \frac{1}{|\mathbf{x}|} + \dots \\
&= \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} y^{i_1} \dots y^{i_l} \partial_{i_1} \dots \partial_{i_l} \frac{1}{|\mathbf{x}|}
\end{aligned} \tag{3.176}$$

We then make use of the fact that, for $r > 0$, and therefore in particular outside the source, we have $\nabla^2(1/|\mathbf{x}|) = 0$, as can be checked again from eq. (3.172).³¹ Then, in eq. (3.176) we can replace $y^i y^j$ with the traceless combination $y^i y^j - (1/3)\delta^{ij}|\mathbf{y}|^2$, and similarly we can remove all traces from the tensors $y^{i_1} \dots y^{i_l}$. Then, inserting eq. (3.176) into eq. (3.173), we get

$$\phi(\mathbf{x}) = \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} Q_{i_1 \dots i_l} \partial_{i_1} \dots \partial_{i_l} \frac{1}{|\mathbf{x}|} \tag{3.178}$$

where

$$Q_{i_1 \dots i_l} = \int d^3y y^{i_1} \dots y^{i_l} \rho(\mathbf{y}) \tag{3.179}$$

and the brackets in $y^{i_1} \dots y^{i_l}$ means that we must take the symmetric-trace-free (STF) part of the tensor $y^{i_1} \dots y^{i_l}$. We therefore have two equivalent formalisms for the multipole expansion of a scalar field that satisfies the Poisson equation (3.169), either in terms of spherical harmonics, or in terms of STF tensors.

For applications to electrodynamics or to gravitational waves, we need to generalize the above results in two ways. First, we do not have a scalar field but rather a vector field A_μ or a tensor field $\bar{h}_{\mu\nu}$. Second, we do not have static fields governed by a Poisson equation, but rather time-dependent fields whose wave equation involve a d'Alembertian operator, such as $\square A_\mu = -(4\pi/c)J_\mu$ in electromagnetism, and eq. (3.3) for GWs. For dealing with non-stationary fields, a possible route is to³⁰ See, e.g. Jackson (1975), Section 3.6.

³¹ More generally,

$$\nabla^2 \frac{1}{|\mathbf{x}|} = -4\pi \delta^{(3)}(\mathbf{x})$$

i.e. $G(\mathbf{x} - \mathbf{y}) = -1/(4\pi|\mathbf{x} - \mathbf{y}|)$ is the Green's function of the Laplacian.

6.0.64 Multi-index notation

³² In general, it can be shown that the irreducible tensor representations of a group are those that give zero when contracted with the invariant tensors (i.e. with the tensors whose form is unchanged by a group transformation). For $SO(3)$ the invariant tensors are δ_{ij} and ϵ_{ijk} . In fact, if $\mathcal{R}_{ij}$ is rotation matrix, δ_{ij} transforms as $\delta_{ij} \rightarrow \mathcal{R}_{ik} \mathcal{R}_{jl} \delta_{kl}$, as for any tensor with two indices. However, $\mathcal{R}_{ik} \mathcal{R}_{jl} \delta_{kl} = \mathcal{R}_{ik} \mathcal{R}_{jk} = \delta_{ij}$ because $\mathcal{R}$ is an orthogonal matrix. Similarly ϵ_{ijk} is invariant because the determinant of $\mathcal{R}$ is equal to one. The condition that the contraction of any two indices of a tensor with an epsilon tensor gives zero singles out totally symmetric tensor, and the condition that the contraction with δ_{ij} gives zero gives the condition of zero trace, for any pair of indices. perform a Fourier transform in time (but not in space), thus reducing the d'Alembertian operator $\square \phi(t, \mathbf{x})$ to $(\nabla^2 + k^2) \tilde{\phi}_k(\mathbf{x})$, so that outside the source the field satisfies a Helmholtz equation $(\nabla^2 + k^2) \tilde{\phi}_k(\mathbf{x}) = 0$. For vector or tensor fields this can be combined with the use of vector and tensor spherical harmonics, respectively. In electrodynamics, this is the approach discussed in Chapter 16 of Jackson (1975). However, this approach does not make explicit the time integration linking the multipole moments to the actual evolution of the source, and furthermore yields somewhat complicated final expressions. Thus, in Section 3.5.1 we rather follow an elegant approach, based on STF tensors, which gives

a simple and unified treatment of scalar, vector and tensor fields. This formalism will be used extensively in Chapter 5, in the study of postNewtonian sources.

The generalization of the approach based on spherical harmonics to the vector and tensor case leads to a more complicated formalism, which nevertheless can be useful in various instances (e.g. for classifying the extra polarization states of GWs in alternative theories of gravity), and we study it in Section 3.5.2.

6.0.65 3.5.1 Symmetric-trace-free (STF) form

We begin by introducing a useful multi-index notation, due to Blanchet and Damour, where a tensor with l indices $i_1 i_2 \dots i_l$ is labeled simply using a capital letter L ,

$$F_L \equiv F_{i_1 i_2 \dots i_l} \quad (3.180)$$

Similarly, G_{iL} denotes a tensor with $l + 1$ indices, $G_{iL} \equiv G_{i i_1 i_2 \dots i_l}$, while, e.g. F_{iL-1} is a notation for $F_{i i_1 i_2 \dots i_{l-1}}$. Furthermore, ∂_L is a notation for $\partial_{i_1} \dots \partial_{i_l}$, and we will also use the notation $x_L \equiv x_{i_1} x_{i_2} \dots x_{i_l}$ and $n_L \equiv n_{i_1} n_{i_2} \dots n_{i_l}$, where $n_i = x_i/r$ is the unit vector in the radial direction. In expressions such as $F_L G_L$, with repeated L indices, the summation over all indices $i_1, i_2, \dots, i_l$ is understood, so

$$F_L G_L = \sum_{i_1 \dots i_l} F_{i_1 \dots i_l} G_{i_1 \dots i_l} \quad (3.181)$$

We use round brackets around indices to denote the symmetrization, e.g. $a_{(ij)} \equiv (1/2)(a_{ij} + a_{ji})$, and we denote by a hat the symmetric-trace-free (STF) projection. That is, the notation $\hat{K}_L$ means that, starting from the tensor with l indices $K_{i_1 \dots i_l}$, we symmetrize it over all indices, and remove all the traces. The operation of taking the STF projection can also be denoted by brackets around the indices, so $\hat{K}_L$ can be equivalently written as $K_{\langle L \rangle}$. The latter notation allows us to write compactly the STF operation between indices belonging to different tensors, e.g. $\epsilon_{ij\langle k} A_{L-1 \rangle i_l}$ means that we perform the STF operation among the index k of ϵ_{ijk} and the first $l - 1$ indices of $A_{i_1 \dots i_{l-1} i_l}$.

A STF tensor with l indices (i.e. of rank l) $A_{i_1 \dots i_l}$ has $2l + 1$ independent components and is therefore a representation of dimension $2l + 1$ of the rotation group $SO(3)$. The crucial point is that these representations are irreducible.³² On the other hand, we know from the theory of representation of the rotation group that the irreducible representations of $SO(3)$ are labeled by an integer value $l = 0, 1, \dots$, and have dimension $2l + 1$, so we see that the set of all STF tensor, for all possible ranks l , gives a complete set of representations of $SO(3)$.

A generic tensor does not provide an irreducible representation of the rotation group, and can be decomposed in irreducible representation, i.e. expressed in terms of STF tensors and factors δ_{ij} and ϵ_{ijk} . The simplest example is the decomposition of a generic tensor with two indices T_{ij} . Writing

$$\begin{aligned} T_{ij} &= \frac{1}{2}(T_{ij} + T_{ji}) + \frac{1}{2}(T_{ij} - T_{ji}) \\ &\equiv S_{ij} + A_{ij} \end{aligned} \quad (3.182)$$

we have decomposed T_{ij} into its symmetric part S_{ij} and its antisymmetric part A_{ij} . Defining $A_k = \epsilon_{ijk} A_{ij}$, we have $A_{ij} = (1/2)\epsilon_{ijk} A_k$. Furthermore, defining S as the trace of S_{ij} , i.e. $S = S_{ii}$, we can rewrite eq. (3.182) as

$$T_{ij} = \frac{1}{3} S \delta_{ij} + \frac{1}{2} \epsilon_{ijk} A_k + \left(S_{ij} - \frac{1}{3} S \delta_{ij} \right) \quad (3.183)$$

which shows explicitly the decomposition of a generic rank-2 tensor T_{ij} into a scalar S , a vector A_k and a rank-2 STF tensor $S_{ij} - (1/3)S\delta_{ij}$.³³

We now examine separately the application of the STF formalism to the multipole expansion of relativistic scalar, vector and tensor fields.

6.0.66 Scalar fields

Consider a scalar field ϕ satisfying the relativistic wave equation

$$\square\phi = -4\pi\rho \quad (3.184)$$

where the source $\rho(t, \mathbf{x})$ is in general time-dependent, but is localized in space, so it vanishes if $|\mathbf{x}| > d$. In the region outside the source, the most general solution can be written as

$$\phi(t, \mathbf{x}) = \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} \partial_L \left[\frac{F_L(t - r/c)}{r} \right] \quad (3.185)$$

where we have used the multi-index notation explained above. This result follows from the fact that, for $r > 0$ and F_L an arbitrary function of retarded time $u = t - r/c$,

$$\square \left[\frac{F_L(t - r/c)}{r} \right] = 0 \quad (3.186)$$

as can be checked immediately with the help of eq. (3.172). Therefore each term of the sum in eq. (3.185) is separately a solution of the vacuum equation $\square\phi = 0$. The fact that this is the most general solution follows from the fact that the set of tensors F_L , with all possible ranks l , provides³⁴ The computation is performed in detail in Appendix B of Blanchet and Damour (1989).

a complete set of representations of the rotation group. So, eq. (3.185) is the generalization of eq. (3.178) to fields governed by a relativistic wave equation, with a non-stationary source. On the other hand, eq. (3.184) can be solved exactly using the retarded Green's function (3.6), which is the appropriate Green's function for a radiation problem, so

$$\phi(t, \mathbf{x}) = \int d^3x' \frac{1}{|\mathbf{x} - \mathbf{x}'|} \rho \left(t - \frac{|\mathbf{x} - \mathbf{x}'|}{c}, \mathbf{x}' \right) \quad (3.187)$$

Comparing this expression, which holds everywhere and therefore in particular outside the source, with eq. (3.185), we can obtain the relativistic multipoles F_L in terms of the source density ρ . The result is³⁴

$$F_L(u) = \int d^3y \hat{y}_L \int_{-1}^1 dz \delta_l(z) \rho(u + z|\mathbf{y}|/c, \mathbf{y}) \quad (3.188)$$

where, according to the multi-index notation discussed above, $\hat{y}_L$ is the STF projection of y_L . The function $\delta_l(z)$ is defined as

$$\delta_l(z) = \frac{(2l+1)!!}{2^{l+1}l!} (1 - z^2)^l \quad (3.189)$$

and satisfies the identities

³³ Observe that scalar and vectors are trivially STF tensors of rank zero and one, respectively.

³⁵ Observe that the argument $u + z|\mathbf{y}|/c$ in eq. (3.188) can be changed to $u - z|\mathbf{y}|/c$, since $\delta_l(z) = \delta_l(-z)$.

$$\int_{-1}^1 dz \delta_l(z) = 1 \quad (3.190)$$

and

$$\lim_{l \rightarrow \infty} \delta_l(z) = \delta(z) \quad (3.191)$$

where $\delta(z)$ is the usual Dirac delta. Physically, the integration over dz in eq. (3.188) performs a weighted time average, different for each multipole moment l , and originates in the different time delay of the radiation emitted from different points inside the source. Equation (3.191) implies that, for sufficiently large l , this time delay becomes negligible.³⁵

6.0.67 Vector field

We next consider the electromagnetic field A^μ which, in the Lorentz gauge $\partial_\mu A^\mu = 0$ (and unrationalized units for the electric charge), satisfies the wave equation

$$\square A^\mu = -\frac{4\pi}{c} J^\mu \quad (3.192)$$

and again we consider a source $J^\mu = (c\rho, \mathbf{J})$ which is time-dependent but localized, so it vanishes if $|\mathbf{x}| > d$. Each component of A^μ can be treated just like the scalar field of the previous subsection, and therefore in the external source region we find

$$A^0(t, \mathbf{x}) = \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} \partial_L \left[\frac{F_L(u)}{r} \right] \quad (3.193)$$

$$A^i(t, \mathbf{x}) = \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} \partial_L \left[\frac{G_{iL}(u)}{r} \right] \quad (3.194)$$

where $u = t - r/c$. The explicit expression of $F_L(u)$ in terms of the source density is given again by eq. (3.188), while

$$G_{iL}(u) = \int d^3y \hat{y}_L \int_{-1}^1 dz \delta_l(z) J_i(u + z|\mathbf{y}|/c, \mathbf{y}) \quad (3.195)$$

This is not yet the most convenient final form of the result, because the tensor G_{iL} is symmetric and traceless with respect to the indices $i_1, \dots, i_l$, since it depends on $\hat{y}_L$, but not under exchange of the index i with one of the indices $i_1, \dots, i_l$, so it is not irreducible. It can however be decomposed in irreducible STF tensors as follows,

$$G_{iL} = U_{iL} + \frac{l}{l+1} \epsilon_{ai\langle i_l} C_{L-1\rangle a} + \frac{2l-1}{2l+1} \delta_{i\langle i_l} D_{L-1\rangle} \quad (3.196)$$

where $U_{L+1} \equiv G_{\langle L+1 \rangle}$, $C_L \equiv G_{ab\langle L-1 \rangle} \epsilon_{i_l \rangle ab}$ and $D_{L-1} \equiv G_{aaL-1}$. Then, $A^i(t, \mathbf{x})$ is given by the sum of three STF terms. It is convenient to perform a gauge transformation $A_\mu \rightarrow A_\mu - \partial_\mu \theta$ (with $\square \theta = 0$ in order not to spoil the Lorentz gauge) which removes the last term in eq. (3.196) from A^i , at the price of adding a new contribution to A^0 . The final result is³⁶

$$A^0(t, \mathbf{x}) = \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} \partial_L \left[\frac{Q_L(u)}{r} \right]$$

$$A^i(t, \mathbf{x}) = -\frac{1}{c} \sum_{l=1}^{\infty} \frac{(-1)^l}{l!} \partial_{L-1} \left[\frac{Q_{iL-1}^{(1)}(u)}{r} + \frac{l}{l+1} \epsilon_{iab} \partial_a \left(\frac{M_{bL-1}(u)}{r} \right) \right],$$

where $Q_{iL-1}^{(1)}$ denotes the first derivative of Q_{iL-1} with respect to retarded time.³⁷ The explicit expression of the moments Q_L and M_L in terms of the source is

$$Q_L(u) = \int d^3y \int_{-1}^1 dz [\delta_l(z) \hat{y}_L \rho(u + z|\mathbf{y}|/c, \mathbf{y}) - \frac{1}{c^2} \frac{2l+1}{(l+1)(2l+3)} \delta_{l+1}(z) \hat{y}_{iL} J_i^{(1)}(u + z|\mathbf{y}|/c, \mathbf{y})] \quad (3.199)$$

where $l \geq 0$, and

$$M_L(u) = \int d^3y \int_{-1}^1 dz \delta_l(z) \hat{y}_{\langle L-1} m_{i\rangle}(u + z|\mathbf{y}|/c, \mathbf{y}) \quad (3.200)$$

where $l \geq 1$ and $m_i = \epsilon_{ijk} y_j J_k$ is the "magnetization density". These results show that the electromagnetic field outside the source can be expressed in terms of two families of STF time-dependent multipole moments, the "electric moments" $Q_L(u)$ and the "magnetic moments" $M_L(u)$.

6.0.68 Gravitational field

We now consider the linearized gravitational field $\bar{h}_{\mu\nu}$ that, in the Lorentz gauge (1.18), satisfies eq. (1.24). Again, we assume that the source $T_{\mu\nu}$ ³⁶ The computation is performed in detail in Damour and Iyer (1991a).

³⁷ For a function $f(u)$ of retarded time, we will use the notation

$$f^{(n)}(u) \equiv \frac{d^n f}{du^n} \quad (3.198)$$

³⁸ See Damour and Iyer (1991a), where it is performed in all detail. has compact support. Then, in the exterior source region the same argument used above for scalar and electromagnetic fields allows us to write the most general solution of the vacuum equations, as

$$\bar{h}^{00}(t, \mathbf{x}) = \frac{4G}{c^4} \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} \partial_L \left[\frac{F_L(u)}{r} \right] \quad (3.201)$$

$$\bar{h}^{0i}(t, \mathbf{x}) = \frac{4G}{c^4} \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} \partial_L \left[\frac{G_{iL}(u)}{r} \right] \quad (3.202)$$

$$\bar{h}^{ij}(t, \mathbf{x}) = \frac{4G}{c^4} \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} \partial_L \left[\frac{H_{ijL}(u)}{r} \right] \quad (3.203)$$

The computation proceeds similarly to the electromagnetic case, but is technically more involved.³⁸ We first decompose G_{iL} in STF tensors, just as we have done in eq. (3.196). Similarly, we decompose H_{ijL} in STF tensors. After performing a suitable gauge transformation (that preserves the gauge condition $\partial_\mu \bar{h}^{\mu\nu} = 0$), the result can be written as

$$\bar{h}^{00} = +\frac{4G}{c^2} \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} \partial_L \left[\frac{M_L(u)}{r} \right]$$

$$\begin{aligned}\bar{h}^{0i} &= -\frac{4G}{c^3} \sum_{l=1}^{\infty} \frac{(-1)^l}{l!} \partial_{L-1} \left[\frac{M_{iL-1}^{(1)}(u)}{r} + \frac{l}{l+1} \epsilon_{iab} \partial_a \left(\frac{S_{bL-1}(u)}{r} \right) \right], \\ \bar{h}^{ij} &= +\frac{4G}{c^4} \sum_{l=2}^{\infty} \frac{(-1)^l}{l!} \partial_{L-2} \left[\frac{1}{r} M_{ijL-2}^{(2)}(u) + \frac{2l}{l+1} \partial_a \left(\frac{1}{r} \epsilon_{ab(i} S_{j)bL-2}^{(1)}(u) \right) \right].\end{aligned}$$

The result therefore depends again on two families of STF tensors, M_L and S_L . Their explicit expression in terms of the energy-momentum tensor of the source can be written introducing the "active gravitational mass density" σ ,

$$\sigma \equiv \frac{1}{c^2} [T^{00} + T^{ii}] \quad (3.205)$$

and the "active mass-current density",

$$\sigma_i \equiv \frac{1}{c} T^{0i} \quad (3.206)$$

as well as $\sigma_{ij} = T^{ij}$. Then

$$\begin{aligned}M_L(u) &= \int d^3x \int_{-1}^1 dz \left\{ \delta_l(z) \hat{x}_L \sigma - \frac{4(2l+1)\delta_{l+1}(z)}{c^2(l+1)(2l+3)} \hat{x}_{iL} \sigma_i^{(1)} \right. \\ &\quad \left. + \frac{2(2l+1)\delta_{l+2}(z)}{c^4(l+1)(l+2)(2l+5)} \hat{x}_{ijL} \sigma_{ij}^{(2)} \right\} (u + z|\mathbf{x}|/c, \mathbf{x})\end{aligned} \quad (3.207)$$

$$\begin{aligned}S_L(u) &= \int d^3x \int_{-1}^1 dz \epsilon_{ab\langle i_l} \left\{ \delta_l(z) \hat{x}_{L-1\rangle a} \sigma_b \right. \\ &\quad \left. - \frac{(2l+1)\delta_{l+1}(z)}{c^2(l+2)(2l+3)} \hat{x}_{L-1\rangle ac} \sigma_{bc}^{(1)} \right\} (u + z|\mathbf{x}|/c, \mathbf{x}).\end{aligned} \quad (3.208)$$

This result holds at the level of linearized theory. In Chapter 5 we will study the full non-linear theory and we will discover that, remarkably, the exact solution of the full general-relativistic problem is constructed using a quantity $h_1^{\mu\nu}$ that is obtained from eqs. (3.204)-(3.208) by means of a very simple modification, that is, with the replacement of the energy momentum tensor of matter $T_{\mu\nu}$ by an effective energy-momentum tensor $\tau_{\mu\nu}$ that includes also the non-linearities of the gravitational field,³⁹ see eqs. (5.135) and (5.136). Thus, eqs. (3.207) and (3.208) already contain the blueprint of the solution to the full non-linear problem.

The integration over z can be computed, in an expansion in powers of $1/c$, using the formula

$$\int_{-1}^1 dz \delta_l(z) f(u + z|\mathbf{x}|/c, \mathbf{x}) = \sum_{k=0}^{+\infty} \frac{(2l+1)!!}{2^k k! (2l+2k+1)!!} \left(\frac{|\mathbf{x}|}{c} \frac{\partial}{\partial u} \right)^{2k} f(u, \mathbf{x}) \quad (3.209)$$

From this we see that, in eqs. (3.207) and (3.208) (as well as in the analogous formulas for the scalar and vector fields), the integration over z allows us to take into account, in a compact way, an infinite series of derivatives.

Finally, we can use these multipolar expressions for $\bar{h}_{\mu\nu}$ to compute the total power radiated in GWs. We use eq. (1.40), plug it into eq. (1.153) retaining only the terms $O(1/r)$ in h_{ij}^{TT} , and we perform the angular integration.⁴⁰ The result is

$$\begin{aligned} \frac{dE}{dt} = & \sum_{l=2}^{+\infty} \frac{G}{c^{2l+1}} \left\{ \frac{(l+1)(l+2)}{(l-1)l!(2l+1)!!} \left\langle M_L^{(l+1)}(u) M_L^{(l+1)}(u) \right\rangle \right. \\ & \left. + \frac{4l(l+2)}{c^2(l-1)(l+1)!(2l+1)!!} \left\langle S_L^{(l+1)}(u) S_L^{(l+1)}(u) \right\rangle \right\} \end{aligned} \quad (3.210)$$

Similarly, for the linear momentum losses one finds

$$\begin{aligned} \frac{dP_i}{dt} = & \sum_{l=2}^{+\infty} \frac{G}{c^{2l+2}} \left\{ \frac{2(l+2)(l+3)}{l(l+1)!(2l+3)!!} \left\langle M_{iL}^{(l+2)}(u) M_L^{(l+1)}(u) \right\rangle \right. \\ & + \frac{8(l+3)}{c^2(l+1)!(2l+3)!!} \left\langle S_{iL}^{(l+2)}(u) S_L^{(l+1)}(u) \right\rangle \\ & \left. + \frac{8(l+2)}{c^2(l-1)(l+1)!(2l+1)!!} \left\langle \epsilon_{ijk} S_{jL-1}^{(l+1)}(u) S_{kL-1}^{(l+1)}(u) \right\rangle \right\} \end{aligned} \quad (3.211)$$

Observe that the linear momentum losses come from the interference between multipoles of different rank, such as the mass quadrupole/mass octupole mixed term, as we already saw on pages 130 – 131.⁴¹

6.0.69 3.5.2 Spherical tensor form

In this section we discuss an alternative formalism for performing the multipole expansion to all orders, which is based on the generalization of the notion of spherical harmonics to a spin-2 field.³⁹ Together with a prescription for rendering finite the integral, since the source $\tau_{\mu\nu}$ no longer has compact support.

⁴⁰ The power is quadratic in $\dot{h}_{ij}^{\text{TT}}$ and therefore in the multipole moments. However, mixed terms of the form $M_L S_{L'}$ give vanishing contribution, after the angular integration, because of parity, while terms $M_L M_{L'}$ or $S_L S_{L'}$ contribute only if $L = L'$. In fact, if $l > l'$, the indices in $M_{i_1 \dots i_l}$ cannot be all contracted with the indices of $M_{i_1 \dots i_{l'}}$, and the remaining indices of the $i_1 \dots i_l$ group are necessarily contracted among them, via the Kronecker deltas that come of from the angular integration, see eq. (3.23).

This section is quite technical, and can be omitted at a first reading.

6.0.70 Spherical components of tensors

To introduce spherical tensors we first consider the quadrupole moment or, more generally, any traceless symmetric tensor with two indices, whose Cartesian components we denote by Q_{ij} . As a first step we introduce a basis in the space of traceless symmetric tensors with two indices, which is chosen so to have a simple relation with the $l = 2$ spherical harmonics.

We recall that the spherical harmonics $Y^{lm}(\theta, \phi)$ with $l = 2$ are⁴²

⁴¹ See Thorne (1980), eq. (4.23') for the corresponding expression for the angular momentum losses.

⁴² Here, according to the standard definition of spherical harmonics, the angle ϕ is measured from the x axis, so for instance the unit vector in the radial direction $\hat{\mathbf{n}}$ has components $n_x = \sin \theta \cos \phi$, $n_y = \sin \theta \sin \phi$ and $n_z = \cos \theta$. Observe that this definition differs from that used, e.g. in Fig. 3.2 and in equations such as eq. (3.72), where ϕ is measured from the y axis.

$$Y^{22}(\theta, \phi) = \left(\frac{15}{32\pi}\right)^{1/2} (e^{i\phi} \sin \theta)^2 \quad (3.212)$$

$$Y^{21}(\theta, \phi) = -\left(\frac{15}{8\pi}\right)^{1/2} e^{i\phi} \sin \theta \cos \theta \quad (3.213)$$

$$Y^{20}(\theta, \phi) = \left(\frac{5}{16\pi}\right)^{1/2} (3 \cos^2 \theta - 1) \quad (3.214)$$

while the expressions for negative values of m are obtained using

$$Y^{l,-m} = (-1)^m Y^{lm*} \quad (3.215)$$

Consider now the unit radial vector $\hat{\mathbf{n}}$. In polar coordinates we have $n_x = \sin \theta \cos \phi$, $n_y = \sin \theta \sin \phi$, $n_z = \cos \theta$, and therefore

$$e^{i\phi} \sin \theta = n_x + i n_y, \quad \cos \theta = n_z \quad (3.216)$$

Plugging this into the explicit expressions for the spherical harmonics Y^{2m} , and using the fact that $n_i n_i = 1$, we see that we can write

$$Y^{2m}(\theta, \phi) = \mathcal{Y}_{ij}^{2m} n_i n_j \quad (3.217)$$

where $\mathcal{Y}_{ij}^{2m}$ is independent of θ, ϕ , and the sum over i, j on the right-hand side is understood. The above equation fixes the part of $\mathcal{Y}_{ij}^{2m}$ which is symmetric in (i, j) , and we complete the definition of $\mathcal{Y}_{ij}^{2m}$ requiring that the antisymmetric part vanishes. The explicit form of the tensors $\mathcal{Y}_{ij}^{2m}$ is then

$$\begin{aligned} \mathcal{Y}_{ij}^{22} &= \sqrt{\frac{15}{32\pi}} \begin{pmatrix} 1 & i & 0 \\ i & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}_{ij} \\ \mathcal{Y}_{ij}^{21} &= -\sqrt{\frac{15}{32\pi}} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & i \\ 1 & i & 0 \end{pmatrix}_{ij} \\ \mathcal{Y}_{ij}^{20} &= \sqrt{\frac{5}{16\pi}} \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}_{ij} \end{aligned} \quad (3.218)$$

together with $\mathcal{Y}_{ij}^{2,-m} = (-1)^m (\mathcal{Y}_{ij}^{2m})^*$. We see from the explicit expressions that the five matrices $\mathcal{Y}_{ij}^{2,m}$ are traceless in the (i, j) indices. This could have also been understood by integrating eq. (3.217) over the solid angle and using $\int d\Omega Y^{2m} = 0$ and $\int d\Omega n_i n_j \sim \delta_{ij}$. From the explicit expressions, we also see that the five tensors $\mathcal{Y}_{ij}^{2m}$ are an orthogonal basis, in the sense that

$$\sum_{ij} \mathcal{Y}_{ij}^{2m} (\mathcal{Y}_{ij}^{2m'})^* = \frac{15}{8\pi} \delta^{mm'} \quad (3.219)$$

It is sometimes useful to invert eq. (3.217). The result is

$$n_i n_j - \frac{1}{3} \delta_{ij} = \sum_{m=-2}^2 c_{ij}^m Y^{2m}(\theta, \phi) \quad (3.220)$$

where

$$c_{ij}^m = \frac{8\pi}{15} (\mathcal{Y}_{ij}^{2m})^* \quad (3.221)$$

This can be proved multiplying both sides of eq. (3.220) by $\mathcal{Y}_{ij}^{2m}$, summing over i, j and using eqs. (3.221) and (3.219), which gives back eq. (3.217). Since $\mathcal{Y}_{ij}^{2m}$ is traceless, the coefficient of the term proportional to δ_{ij} on the left-hand side of eq. (3.220) is not yet fixed in this way, but we can fix it by observing that the right-hand side of eq. (3.220) is traceless, so also the left-hand side must be traceless.

The five symmetric and traceless matrices $\mathcal{Y}_{ij}^{2m}$, with $m = -2, \dots, 2$, are linearly independent and therefore are a basis for the five-dimensional space of traceless symmetric tensors Q_{ij} . This means that we can expand an arbitrary traceless symmetric tensors Q_{ij} as

$$Q_{ij} = \sum_{m=-2}^2 Q_m \mathcal{Y}_{ij}^{2m} \quad (3.222)$$

The quantities Q_m are called the spherical components of Q_{ij} . Multiplying by $n_i n_j$ and using eq. (3.217) we obtain

$$Q_{ij} n_i n_j = \sum_{m=-2}^2 Q_m Y^{2m}(\theta, \phi) \quad (3.223)$$

(As always, the summation over the repeated i, j indices is understood; instead, we write explicitly the sum over m .) This equation could also have been obtained directly by observing that $Q_{ij} n_i n_j$ is a function of θ, ϕ (with the dependence hidden in $n_i(\theta, \phi)$, while Q_{ij} is a constant tensor) and therefore can be expanded in spherical harmonics as $\sum_{l,m} Q_{lm} Y^{lm}(\theta, \phi)$. However, Q_{ij} is symmetric and traceless, so it is a spin-2 operator and therefore in the expansion in spherical harmonics, only $l = 2$ contributes.

The five independent components of the symmetric traceless tensor Q_{ij} are therefore expressed in terms of the five independent quantities Q_m , with $m = -2, \dots, 2$. If Q_{ij} is real, as in the case of the mass

⁴³ In this section we prefer to write explicitly the indices $i_1 \dots i_l$, rather than using the multi-index notation defined on page 134. quadrupole, the five complex quantities Q_m satisfy $Q_m^* = (-1)^m Q_{-m}$, because of eq. (3.215). Using eq. (3.219) we can invert eq. (3.222),

$$Q_m = \frac{8\pi}{15} Q_{ij} (\mathcal{Y}_{ij}^{2m})^* \quad (3.224)$$

or, explicitly,

$$\begin{aligned} Q_{\pm 2} &= \left(\frac{2\pi}{15}\right)^{1/2} (Q_{11} - Q_{22} \mp 2iQ_{12}) \\ Q_{\pm 1} &= \mp \left(\frac{8\pi}{15}\right)^{1/2} (Q_{13} \mp iQ_{23}) \\ Q_0 &= -\left(\frac{4\pi}{5}\right)^{1/2} (Q_{11} + Q_{22}) \end{aligned} \quad (3.225)$$

We can now write the power emitted by the quadrupole radiation, given in eq. (3.75), in terms of the spherical components Q_m . Using eq. (3.223) we write

$$\ddot{Q}_{ij}n_in_j = \sum_{m=-2}^2 \ddot{Q}_m Y^{2m}(\theta, \phi) \quad (3.226)$$

and we take the squared modulus,

$$\ddot{Q}_{ij}\ddot{Q}_{kl}n_in_jn_kn_l = \sum_{m,m'} \ddot{Q}_m^* \ddot{Q}_{m'} Y^{2m*}(\theta, \phi) Y^{2m'}(\theta, \phi). \quad (3.227)$$

Integrating over $d\Omega$ with the help of eq. (3.22) and using the orthogonality of the spherical harmonics,

$$\int d\Omega Y^{lm*}(\theta, \phi) Y^{l'm'}(\theta, \phi) = \delta^{ll'} \delta^{mm'} \quad (3.228)$$

we get

$$\frac{8\pi}{15} \ddot{Q}_{ij} \ddot{Q}_{ij} = \sum_{m=-2}^2 |\ddot{Q}_m|^2 \quad (3.229)$$

Therefore, eq. (3.75) becomes

$$P_{\text{quad}} = \frac{3G}{8\pi c^5} \sum_{m=-2}^2 \langle |\ddot{Q}_m|^2 \rangle \quad (3.230)$$

We can now generalize the above construction to traceless symmetric tensors with an arbitrary number of indices. We consider a (real) STF tensor with l indices, $T_{i_1 \dots i_l}$ ⁴³ A basis in this tensor space can be obtained by observing that the spherical harmonics $Y^{lm}(\theta, \phi)$ with $m \geq 0$ are given explicitly by

$$\begin{aligned} Y^{lm}(\theta, \phi) &= C^{lm} e^{im\phi} P^{lm}(\cos \theta) \\ &= C^{lm} (e^{i\phi} \sin \theta)^m \sum_{k=0}^{[(l-m)/2]} a_k^{lm} (\cos \theta)^{l-m-2k} \end{aligned} \quad (3.231)$$

and $Y^{lm} = (-1)^m (Y^{l,-m})^*$ for $m < 0$. The notation $[(l-m)/2]$ denotes the largest integer smaller or equal to $(l-m)/2$, and the coefficients are given by

$$C^{lm} = (-1)^m \left(\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!} \right)^{1/2} \quad (3.232)$$

$$a_k^{lm} = \frac{(-1)^k}{2^l k! (l-k)!} \frac{(2l-2k)!}{(l-m-2k)!} \quad (3.233)$$

Comparing with eq. (3.216) we see that Y^{lm} is the sum of a term containing l factors n_i , a term containing $l-2$ factors n_i , a term containing $l-4$ factors n_i , etc. Using $n_i n_i = 1$, a term with $l-2$ factors $n_{i_1} \dots n_{i_{l-2}}$ can be rewritten trivially as a term with l factors n_i , as $\delta_{ij} n_{i_1} \dots n_{i_{l-2}} n_i n_j$, and similarly for all terms with $l-2k$ factors. Then we can write

$$Y^{lm}(\theta, \phi) = \mathcal{Y}_{i_1 \dots i_l}^{lm} n_{i_1} \dots n_{i_l} \quad (3.234)$$

where the tensors $\mathcal{Y}_{i_1 \dots i_l}^{lm}$ are independent of θ, ϕ , and the sum over the l indices $i_1 \dots i_l$ is understood. We will not need the explicit form of $\mathcal{Y}_{i_1 \dots i_l}^{lm}$, which anyway can be read from eqs.

(3.231) and (3.216). Just as in the case $l = 2$ discussed above, one can show that the tensors $\mathcal{Y}_{i_1 \dots i_l}^{lm}$ are a basis in the space of traceless symmetric tensors with l indices. This means that we can expand

$$T_{i_1 \dots i_l} = \sum_{m=-l}^l T_{lm} \mathcal{Y}_{i_1 \dots i_l}^{lm} \quad (3.235)$$

and this defines the spherical components T_{lm} of the tensor $T_{i_1 \dots i_l}$. Multiplying by $n_{i_1} \dots n_{i_l}$ we have the identity

$$T_{i_1 \dots i_l} n_{i_1} \dots n_{i_l} = \sum_{m=-l}^l T_{lm} Y^{lm}(\theta, \phi) \quad (3.236)$$

which expresses the fact that in the expansion in spherical harmonics of the left-hand side contribute only the spherical harmonics whose angular momentum l is equal to the number of indices of $T_{i_1 \dots i_l}$. This is a consequence of the fact that both a STF tensor with l indices, and the spherical harmonics Y^{lm} , provide an irreducible representation of dimension $2l + 1$ of the rotation group.

Using the orthogonality of spherical harmonics we can invert eq. (3.236) and we obtain the spherical components T_{lm} in terms of the Cartesian components $T_{i_1 \dots i_l}$,

$$\begin{aligned} T_{lm} &= T_{i_1 \dots i_l} \int d\Omega (Y^{lm})^* n_{i_1} \dots n_{i_l} \\ &= T_{i_1 \dots i_l} (\mathcal{Y}_{j_1 \dots j_l}^{lm})^* \int d\Omega n_{i_1} \dots n_{i_l} n_{j_1} \dots n_{j_l} \end{aligned} \quad (3.237)$$

where in the second line we used eq. (3.234). The integral can be performed using eq. (3.23). Since $\mathcal{Y}_{j_1 \dots j_l}^{lm}$ and $T_{i_1 \dots i_l}$ are traceless, in the

⁴⁴ For this reasons, the spherical components find a typical application in quantum mechanics, for writing selection rules in atomic transitions. sum over permutations in eq. (3.23) the terms with Kronecker deltas of the type $\delta_{i_k i_{k'}}$ or $\delta_{j_k j_{k'}}$ give zero, and the only contributions come from the term $\delta_{i_1 j_1} \dots \delta_{i_l j_l}$ and from its permutations. Since $\mathcal{Y}_{j_1 \dots j_l}^{lm}$ is totally symmetric, these $l!$ permutations all give the same result, so

$$T_{lm} = 4\pi \frac{l!}{(2l+1)!!} T_{i_1 \dots i_l} (\mathcal{Y}_{i_1 \dots i_l}^{lm})^* \quad (3.238)$$

For $l = 2$, we recover the result obtained in eq. (3.224). Finally, a useful identity which generalizes eq. (3.229) is obtained taking the modulus squared of eq. (3.236) and integrating over $d\Omega$ with the help of eq. (3.23). This gives

$$4\pi \frac{l!}{(2l+1)!!} T_{i_1 \dots i_l} T^{i_1 \dots i_l} = \sum_{m=-l}^l |T_{lm}|^2 \quad (3.239)$$

The transformation properties under rotations of the spherical components of tensors are fixed by the transformation properties of the spherical harmonics. For instance, consider a rotation by an angle φ around the z -axis, $\phi \rightarrow \phi + \varphi$. The left-hand side of eq. (3.236) is a scalar so it is invariant, while on the right-hand side $Y_{lm} \rightarrow e^{im\varphi} Y_{lm}$. Therefore T_{lm} transforms into itself, as ⁴⁴

$$T_{lm} \rightarrow e^{-im\varphi} T_{lm} \quad (3.240)$$

More generally, under arbitrary rotations the $2l + 1$ components of T_{lm} , with $m = -l, \dots, l$ and l given, transform among themselves in the same way as $Y_{lm}^*(\theta, \phi)$.

6.0.71 Vector and tensor spherical harmonics

The spherical components of tensors, introduced above, are one of the tools useful for the construction of a systematic multipole expansion. Here we introduce another necessary ingredient of this formalism, the tensor spherical harmonics. In the same way as the usual (scalar) spherical harmonics are useful to describe the angular dependence of a scalar field, tensor spherical harmonics are useful for describing the angular dependence of a field with spin.

We denote by $\mathbf{L}$ the orbital angular momentum operator, by $\mathbf{S}$ the spin operator and by $\mathbf{J} = \mathbf{L} + \mathbf{S}$ the total angular momentum. All these quantities are measured in units of $\hbar$, so for instance, as operator acting on functions, $\mathbf{L} = \mathbf{r} \times (-i\nabla)$. Since the operators $\mathbf{J}^2$, J_z^2 , $\mathbf{L}^2$ and $\mathbf{S}^2$ commute, we can diagonalize them simultaneously. The eigenfunctions are the tensor spherical harmonics and are denoted by $Y_{jj_z}^{ls}(\theta, \phi)$. Therefore, by definition, the functions $Y_{jj_z}^{ls}(\theta, \phi)$ are the solutions of

$$\mathbf{J}^2 Y_{jj_z}^{ls} = j(j+1) Y_{jj_z}^{ls} \quad (3.241)$$

$$J_z Y_{jj_z}^{ls} = j_z Y_{jj_z}^{ls} \quad (3.242)$$

$$\mathbf{L}^2 Y_{jj_z}^{ls} = l(l+1) Y_{jj_z}^{ls} \quad (3.243)$$

$$\mathbf{S}^2 Y_{jj_z}^{ls} = s(s+1) Y_{jj_z}^{ls} \quad (3.244)$$

Their explicit form can be obtained coupling the (scalar) spherical harmonics Y_{lm} to the spin function χ_{ss_z} , with the appropriate ClebschGordan coefficients which gives a state with total angular momentum $|j, j_z\rangle$

$$Y_{jj_z}^{ls}(\theta, \phi) = \sum_{l_z=-l}^l \sum_{s_z=-s}^s \langle sls_z l_z | jj_z \rangle Y_{l_z}(\theta, \phi) \chi_{ss_z} \quad (3.245)$$

It is easy to check that this expression indeed satisfies eqs. (3.241)(3.244). For instance, the operators L_i act only on the variables θ, ϕ of $Y_{l_z}(\theta, \phi)$, and then eq. (3.243) follows from $\mathbf{L}^2 Y_{l_z} = l(l+1) Y_{l_z}$. Similarly, eq. (3.244) follows from the fact that the spin operator acts only on the spin wavefunction χ , with $\mathbf{S}^2 \chi_{ss_z} = s(s+1) \chi_{ss_z}$. Finally, the Clebsch-Gordan coefficients $\langle sls_z l_z | jj_z \rangle$ couple a state with orbital angular momentum $|l_z\rangle$ to a state with spin $|ss_z\rangle$ to give a state with total angular momentum $|jj_z\rangle$, so that eqs. (3.241) and (3.242) follow.

Depending on the value of s , one has spinor spherical harmonics ($s = 1/2$), vector spherical harmonics ($s = 1$), spin-2 tensor spherical harmonics, etc. ⁴⁵

Tensor spherical harmonics describe the angular distribution and polarization of particles of spin s , in a state with definite values of the total angular momentum j , of j_z , and of the orbital angular momentum l . For gravitational waves, we are interested in spin-2 tensor spherical harmonics. Observe that, beside the indices l, s, j, j_z written explicitly, $Y_{jj_z}^{ls}$ carries also an index which depends on the nature of the spin wavefunction χ ; e.g. a spinor index for $s = 1/2$, a vector index for $s = 1$, a pair of spatial indices (i, i') for $s = 2$ (with $(Y_{jj_z}^{l2})_{ii'}$ symmetric and traceless in i, i'), etc.

Let us first examine the vector spherical harmonics. The spin wave function χ in this case is a vector, and we denote it by $\boldsymbol{\xi}$. The wavefunctions with a definite value of $s_z = 0, \pm 1$ can be constructed from the unit vectors $\mathbf{e}_x, \mathbf{e}_y$ and $\mathbf{e}_z$ as

$$\boldsymbol{\xi}^{(\pm 1)} = \mp \frac{1}{\sqrt{2}} (\mathbf{e}_x \pm i\mathbf{e}_y), \quad \boldsymbol{\xi}^{(0)} = \mathbf{e}_z \quad (3.246)$$

Then the vector spherical harmonics are

$$\mathbf{Y}_{jj_z}^l(\theta, \phi) = \sum_{l_z=-l}^l \sum_{s_z=0, \pm 1} \langle 1l s_z l_z | jj_z \rangle Y_{ll_z}(\theta, \phi) \boldsymbol{\xi}^{(s_z)} \quad (3.247)$$

Observe that in $\boldsymbol{\xi}^{(s_z)}$ the index $s_z = \pm 1, 0$ tells us which vector we must consider, according to eq. (3.246); the spatial components of the vector are instead denoted by $\xi_i^{(s_z)}$. Correspondingly, $\mathbf{Y}_{jj_z}^l$ is a vector with components $(Y_{jj_z}^l)_i$. Note also that we have written $\mathbf{Y}_{jj_z}^{l1}$ simply as $\mathbf{Y}_{jj_z}^l$, since the fact that $s = 1$ is already implicit in the vector notation.

⁴⁵ One should not be misled by this nomenclature. Of course, the properties of $Y_{jj_z}^{ls}$ under rotations depend on the value of the total angular momentum j , not on the spin s , so for instance a vector spherical harmonics $Y_{jj_z}^{l1}$ with $j = 2$ (and therefore $l = 1, 2$ or 3) has the transformation properties of a spin-2 operator, not of a vector, just as the usual scalar spherical harmonics Y_{lm} transform of course as a spin- l operator, not as a scalar.

Vector spherical harmonics

⁴⁶ Using the explicit expression of the spherical harmonics, we can verify immediately that $Y_{10}\boldsymbol{\xi}^{(0)} - Y_{11}\boldsymbol{\xi}^{(-1)} - Y_{1-1}\boldsymbol{\xi}^{(+1)}$ is proportional to $\hat{\mathbf{n}} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$. On the other hand, apart from a proportionality factor, the combination $Y_{10}\boldsymbol{\xi}^{(0)} - Y_{11}\boldsymbol{\xi}^{(-1)} - Y_{1-1}\boldsymbol{\xi}^{(+1)}$ is just the combination (3.247) with the ClebschGordan coefficients necessary to produce a state with $j = 0$ from $l = 1$ and $s = 1$. Thus, the vector spherical harmonics $\mathbf{Y}_{00}^1$ is proportional to $\hat{\mathbf{n}}$, and is therefore proportional to $\mathbf{Y}_{00}^R$ (observe that $Y_{00} = 1/(4\pi)^{1/2}$ is just a constant.) In this way we have taken into account the only vector harmonics that exists for $j = 0$ while, for each $j \geq 1$, we have three vector harmonics, with $l = j, j-1, j+1$. The one with $l = j$ is rewritten as $\mathbf{Y}_{jj_z}^B$ while the two with $l = j \pm 1$ are combined to form $\mathbf{Y}_{jj_z}^E$ and $\mathbf{Y}_{jj_z}^R$, with $j \geq 1$.

By construction, the vector spherical harmonics are eigenfunctions of $\mathbf{L}^2$,

$$\mathbf{L}^2 (Y_{jj_z}^l)_i = l(l+1) (Y_{jj_z}^l)_i \quad (3.248)$$

and therefore are useful in solving an equation of the form $\square \mathbf{V} = 0$, where $\square = -\partial_0^2 + \nabla^2$ is the flat space d'Alembertian and $\mathbf{V}$ a vector field. In fact, using the expression for the Laplacian in spherical coordinates, eq. (3.172), we can separate the radial and the angular dependence writing

$$\mathbf{V}(r, \theta, \phi) = \sum_{l,j,j_z} f_{ljj_z}(r) \mathbf{Y}_{jj_z}^l(\theta, \phi) \quad (3.249)$$

Observe that the vector spherical harmonics are orthonormal,

$$\int d\Omega \mathbf{Y}_{jj_z}^l \cdot (\mathbf{Y}_{j'j'_z}^{l'})^* = \delta_{ll'} \delta_{jj'} \delta_{j_z j'_z} \quad (3.250)$$

The vectors $\mathbf{Y}_{jj_z}^l(\theta, \phi)$ can have $j = l-1, l, l+1$ (if $l \neq 0$), or $j = 1$ if $l = 0$, (that is, the possible quantum combinations of spin $s = 1$ and orbital angular momentum l). For generic values of l and j within this range, they have no special property with respect to the radial unit vector $\hat{\mathbf{n}}$, that is, they are neither purely transverse nor purely longitudinal. We can however observe that the full set of vectors $\mathbf{Y}_{jj_z}^l$, with $j = l-1, l, l+1$ (if $l \neq 0$), or $j = 1$ if $l = 0$, can be expressed in terms of the following combinations,

⁴⁷ In eqs. (3.254) – (3.256), in order to conform to the most common notation used in electrodynamics (for vector spherical harmonics) and in the gravitational-wave literature (for the spin-2 tensor harmonics), we have changed the labeling of the indices of $\mathbf{Y}^E$, $\mathbf{Y}^B$ and $\mathbf{Y}^R$ from j, j_z to l, m . It is however important to understand that these indices refer to the total angular momentum, and not to the orbital angular momentum.

$$\mathbf{Y}_{jjz}^E = (2j+1)^{-1/2} [(j+1)^{1/2} \mathbf{Y}_{jjz}^{j-1} + j^{1/2} \mathbf{Y}_{jjz}^{j+1}] \quad (3.251)$$

$$\mathbf{Y}_{jjz}^B = i \mathbf{Y}_{jjz}^j \quad (3.252)$$

$$\mathbf{Y}_{jjz}^R = (2j+1)^{-1/2} [j^{1/2} \mathbf{Y}_{jjz}^{j-1} - (j+1)^{1/2} \mathbf{Y}_{jjz}^{j+1}] \quad (3.253)$$

with $j \geq 1$, together with $\mathbf{Y}_{00}^R = Y_{00} \hat{\mathbf{n}}$. Observe that, since $s = 1$, a given value of $j \geq 1$ can be obtained with $l = j-1, j$ or $j+1$. In eqs. (3.251) and (3.253) we have combined the vector harmonic with $l = j-1$ and the vector harmonic with $l = j+1$, while $\mathbf{Y}_{jjz}^B$ is made with $l = j$.⁴⁶

Since $\mathbf{Y}_{jjz}^E$ and $\mathbf{Y}_{jjz}^R$ are superposition of vector harmonics with different values of l , they are no longer eigenfunctions of $\mathbf{L}^2$. However, using the properties of spherical harmonics, we can rewrite the above definitions in terms of the scalar spherical harmonics Y_{lm} as⁴⁷

$$\mathbf{Y}_{lm}^E = [l(l+1)]^{-1/2} r \nabla Y_{lm} \quad (l \geq 1) \quad (3.254)$$

$$\mathbf{Y}_{lm}^B = [l(l+1)]^{-1/2} i \mathbf{L} Y_{lm} \quad (l \geq 1) \quad (3.255)$$

$$\mathbf{Y}_{lm}^R = Y_{lm} \hat{\mathbf{n}} \quad (l \geq 0) \quad (3.256)$$

From these expressions we see that $\mathbf{Y}_{lm}^R$ is a longitudinal vector, since it is proportional to $\hat{\mathbf{n}}$, while $\mathbf{Y}_{lm}^E$ and $\mathbf{Y}_{lm}^B$ are transverse. In fact, the operator $\mathbf{L}$ has only components in the $\hat{\theta}$ and $\hat{\phi}$ directions and, since Y_{lm} depends only on θ, ϕ and not on r , also ∇Y_{lm} has only components in the $\hat{\theta}$ and $\hat{\phi}$ directions. Furthermore, since $\mathbf{L} = \mathbf{r} \times (-i \nabla) = -ir \hat{\mathbf{n}} \times \nabla$, we have

$$\mathbf{Y}_{lm}^B = \hat{\mathbf{n}} \times \mathbf{Y}_{lm}^E \quad (3.257)$$

Therefore $\mathbf{Y}_{lm}^E$ and $\mathbf{Y}_{lm}^B$ are transverse with respect to $\hat{\mathbf{n}}$ and are orthogonal to each other. Under a parity transformation $\mathbf{Y}_{lm}^E$ and $\mathbf{Y}_{lm}^R$ pick a factor $\pi_l = (-1)^l$. This is the transformation property of the electric field, so this is called an electric-type parity. Instead $\mathbf{Y}_{lm}^B$ picks a factor $\pi_l = (-1)^{l+1}$, so it has a magnetic type parity.⁴⁸

The vector functions $\mathbf{Y}_{lm}^E, \mathbf{Y}_{lm}^B$ and $\mathbf{Y}_{lm}^R$ are called "pure-spin vector harmonics" because they are appropriate for describing the polarization states of a vector field, while the vector functions $\mathbf{Y}_{lm}'$ given in eq. (3.247) are called "pure-orbital vector harmonics" because they are eigenfunctions of the orbital angular momentum.

The pure-spin vector harmonics are orthonormal,

$$\int d\Omega \mathbf{Y}_{lm}^J \cdot (\mathbf{Y}_{l'm'}^{J'})^* = \delta_{JJ'} \delta_{ll'} \delta_{mm'} \quad (3.258)$$

where $J = E, B, R$. The angular dependence of an arbitrary vector field can be expanded in pure-spin vector harmonics as

$$\begin{aligned} \mathbf{V}(t, r, \theta, \phi) &= \sum_{l=0}^{\infty} \sum_{m=-l}^l R_{lm}(t, r) \mathbf{Y}_{lm}^R(\theta, \phi) \\ &+ \sum_{l=1}^{\infty} \sum_{m=-l}^l [E_{lm}(t, r) \mathbf{Y}_{lm}^E(\theta, \phi) + B_{lm}(t, r) \mathbf{Y}_{lm}^B(\theta, \phi)] \end{aligned} \quad (3.259)$$

Observe that, in the second line, the sum over l runs only over $l \geq 1$, since the corresponding pure-spin vector harmonics start from $l = 1$.⁴⁹

A massive spin-1 particle has three degrees of freedom, and we see that these degrees of freedom are described by E_{lm} , B_{lm} and R_{lm} , respectively. If we want to describe a massless vector particle, however, the situation is different. A massless vector particle has only two physical degrees of freedom (see the discussion of the Poincaré representations in Problem 2.1), with helicities $h = \pm 1$. If we perform a rotation by an angle θ around the $\hat{\mathbf{n}}$ axis the two transverse vectors $\mathbf{Y}_{lm}^E$ and $\mathbf{Y}_{lm}^B$ transform among themselves (see eq. (3.257)) so $\mathbf{Y}_{lm}^E \pm i\mathbf{Y}_{lm}^B$ are multiplied by $\exp\{\pm i\theta\}$. Comparing with eq. (2.197), we understand that they describe the two components of a massless particle with helicities $h = \pm 1$. Instead $\mathbf{Y}_{lm}^R$, being proportional to $\hat{\mathbf{n}}$, is invariant under rotations around the $\hat{\mathbf{n}}$ axis, and therefore, again from eq. (2.197), we see that it has $h = 0$ which, for a massless particle, implies $s = 0$; so $\mathbf{Y}_{lm}^R$ describes a spin-zero massless particle.

In electrodynamics, the longitudinal degree of freedom described by $\mathbf{Y}_{lm}^R$ is eliminated by gauge invariance, and electromagnetic radiation is purely transverse. Therefore in the expansion of the vector potential $\mathbf{A}$ ⁴⁸ We define parity reversing the orientation of the axes of the reference frame while keeping the vectors fixed (compare with Note 34 on page 98). With respect to the Cartesian basis vectors $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$, a vector spherical harmonic $\mathbf{Y}_{lm}$ has components $(\mathbf{Y}_{lm})_i$ that is $\mathbf{Y}_{lm} = (\mathbf{Y}_{lm})_i \mathbf{e}_i$. Under parity, the components $(\mathbf{Y}_{lm}^E)_i$ and $(\mathbf{Y}_{lm}^R)_i$ pick a factor $(-1)^{l+1}$, with $(-1)^l$ coming from the scalar spherical harmonic Y_{lm} and a further minus sign from ∂_i and from n_i , respectively. However, $\mathbf{Y}_{lm}^E = (\mathbf{Y}_{lm}^E)_i \mathbf{e}_i$ transform with a factor $(-1)^l$ because of the further minus sign from the transformation of the base vectors $\mathbf{e}_i$. Instead, $(\mathbf{Y}_{lm}^B)_i$ picks only the factor $(-1)^l$ from the scalar spherical harmonics, because angular momentum is a pseudovector and its components L_i are unchanged under parity, and then $\mathbf{Y}_{lm}^B = (\mathbf{Y}_{lm}^B)_i \mathbf{e}_i$ is multiplied by $\pi_l = (-1)^{l+1}$.

6.0.72 Spin-2 tensor harmonics

Table 3.1 The coefficients which enter in the definition of pure-spin $s = 2$ spherical harmonics (from Thorne 1980).

⁴⁹ Recall that we have changed notation between eq. (3.253) and eq. (3.254), see Note 47, and the quantity that we are now labeling by l is the total angular momentum, previously denoted by j .

$$\begin{aligned}
a_{11} &+ \left(\frac{(j+1)(j+2)}{(2j+1)(2j+3)} \right)^{1/2} \\
a_{12} &- \left(\frac{2j(j+1)}{3(2j-1)(2j+3)} \right)^{1/2} \\
a_{13} &+ \left(\frac{(j-1)j}{(2j-1)(2j+1)} \right)^{1/2} \\
a_{21} &- \left(\frac{2j(j+2)}{(2j+1)(2j+3)} \right)^{1/2} \\
a_{22} &- \left(\frac{(2j-1)(2j+3)}{(2j-2)} \right)^{1/2} \\
a_{23} &+ \left(\frac{2(j-1)(j+1)}{(2j-1)(2j+1)} \right)^{1/2} \\
a_{31} &+ \left(\frac{(j-1)j}{2(2j+1)(2j+3)} \right)^{1/2} \\
a_{32} &+ \left(\frac{3(j-1)(j+2)}{(2j-1)(2j+3)} \right)^{1/2} \\
a_{33} &+ \left(\frac{(j+1)(j+2)}{2(2j-1)(2j+1)} \right)^{1/2} \\
b_{11} &+ \left(\frac{j+2}{2j+1} \right)^{1/2} \\
b_{12} &- \left(\frac{j-1}{2j+1} \right)^{1/2} \\
b_{21} &- \left(\frac{j-1}{2j+1} \right)^{1/2} \\
b_{22} &- \left(\frac{j+2}{2j+1} \right)^{1/2}
\end{aligned}$$

we have $R_{lm} = 0$,

$$\mathbf{A}(t, r, \theta, \phi) = \sum_{l=1}^{\infty} \sum_{m=-l}^l [E_{lm}(t, r) \mathbf{Y}_{lm}^E(\theta, \phi) + B_{lm}(t, r) \mathbf{Y}_{lm}^B(\theta, \phi)] \quad (3.260)$$

For each value of the total (recall Note 47) angular momentum $l = 1, 2, \dots$ the electromagnetic field is therefore characterized by two polarization states with opposite parity, described by the wavefunctions E_{lm} and B_{lm} , called electric and magnetic photons, respectively. Their linear combinations give rise to the two helicity states of the photon. The state with total angular momentum $l = 0$ is instead absent.

The fact that, for a massless particle with helicities ± 1 , such as the photon, the state with total angular momentum $l = 0$ is absent while for all other values of the total angular momentum we have two states with opposite parity agrees with the analysis that we performed in Problem 2.2.

We can now introduce the spin-2 tensor harmonics, which are relevant for the description of gravitational radiation. First of all we need the spin wavefunction for $s = 2$ with a definite value of s_z ; this wavefunction is a traceless symmetric tensor, that we denote by $t_{ik}^{(s_z)}$, and is obtained taking two spin-1 wavefunctions $\xi_i^{(m_1)}$ and $\xi_k^{(m_2)}$, and combining them with the appropriate Clebsch-Gordan coefficients,

$$t_{ik}^{(s_z)} = \sum_{m_1, m_2=-1}^1 \langle 11m_1m_2 | 2s_z \rangle \xi_i^{(m_1)} \xi_k^{(m_2)} \quad (3.261)$$

The five tensors $t_{ik}^{(s_z)}$, with $s_z = 0, \pm 1, \pm 2$, are symmetric and traceless, and play the role that the three vectors $\xi_i^{(s_z)}$ with $s_z = 0, \pm 1$ played for vector spherical harmonics, that is, we combine them with the scalar spherical harmonics to obtain the spin-2 tensor spherical harmonics,

$$\begin{aligned}
(\mathbf{T}_{jjz}^l)_{ik} &\equiv (Y_{jjz}^{l2})_{ik} \\
&= \sum_{l_z=-l}^l \sum_{s_z=-2}^2 \langle 2ls_zl_z | jjz \rangle Y_{l_z}(\theta, \phi) t_{ik}^{(s_z)}
\end{aligned} \tag{3.262}$$

Just as in the vector case, $(\mathbf{T}_{jjz}^l)_{ik}$ are by construction eigenfunctions of the $\mathbf{L}^2$ operator (and for this reason are called pure-orbital $s = 2$ tensor spherical harmonics), but have no special property with respect to the radial unit vector $\hat{\mathbf{n}}$. Similarly to the case of vector harmonics, we can however observe that the full set of tensors $\mathbf{T}_{jjz}^l$, with $j \geq 0$ and $j = l \pm 2, l \pm 1, l$ (if $l \geq 2$) or $j = 1, 2, 3$ if $l = 1$, or $j = 2$ if $l = 0$ (that is, the possible quantum combinations of spin $s = 2$ and orbital angular momentum l) can be expressed in terms of combinations, called the pure-spin $s = 2$ tensor spherical harmonics, with definite properties under rotations along the radial directions. For $j \geq 2$, these are given by

$$\mathbf{T}_{jjz}^{S0} = a_{11} \mathbf{T}_{jjz}^{j+2} + a_{12} \mathbf{T}_{jjz}^j + a_{13} \mathbf{T}_{jjz}^{j-2} \tag{3.263}$$

$$\mathbf{T}_{jjz}^{E1} = a_{21} \mathbf{T}_{jjz}^{j+2} + a_{22} \mathbf{T}_{jjz}^j + a_{23} \mathbf{T}_{jjz}^{j-2} \tag{3.264}$$

$$\mathbf{T}_{jjz}^{E2} = a_{31} \mathbf{T}_{jjz}^{j+2} + a_{32} \mathbf{T}_{jjz}^j + a_{33} \mathbf{T}_{jjz}^{j-2} \tag{3.265}$$

$$\mathbf{T}_{jjz}^{B1} = b_{11} i \mathbf{T}_{jjz}^{j+1} + b_{12} i \mathbf{T}_{jjz}^{j-1}, \tag{3.266}$$

$$\mathbf{T}_{jjz}^{B2} = b_{21} i \mathbf{T}_{jjz}^{j+1} + b_{22} i \mathbf{T}_{jjz}^{j-1}, \tag{3.267}$$

where the coefficients are given in Table 3.1. These combinations can be expressed in terms of the scalar spherical harmonics as follows⁵⁰

$$(\mathbf{T}_{lm}^{S0})_{ij} = [n_i n_j - (1/3) \delta_{ij}] Y_{lm} \tag{3.268}$$

$$(\mathbf{T}_{lm}^{E1})_{ij} = c_l^{(1)} (r/2) (n_i \partial_j + n_j \partial_i) Y_{lm} \tag{3.269}$$

$$(\mathbf{T}_{lm}^{B1})_{ij} = c_l^{(1)} (i/2) (n_i L_j + n_j L_i) Y_{lm} \tag{3.270}$$

$$(\mathbf{T}_{lm}^{E2})_{ij} = c_l^{(2)} r^2 \Lambda_{ij,i'j'}(\hat{\mathbf{n}}) \partial_{i'} \partial_{j'} Y_{lm} \tag{3.271}$$

$$(\mathbf{T}_{lm}^{B2})_{ij} = c_l^{(2)} r \Lambda_{ij,i'j'}(\hat{\mathbf{n}}) (i/2) (\partial_{i'} L_{j'} + \partial_{j'} L_{i'}) Y_{lm} \tag{3.272}$$

where, as usual, $\Lambda_{ij,i'j'}$ is the tensor that implements the TT projection, see eq. (1.36), and

$$c_l^{(1)} = \left(\frac{2}{l(l+1)} \right)^{1/2}, \quad c_l^{(2)} = \left(2 \frac{(l-2)!}{(l+2)!} \right)^{1/2} \tag{3.273}$$

A complete set of $s = 2$ spherical harmonic is given by eqs. (3.268)-(3.272) where $\mathbf{T}_{lm}^{S0}$ has $l \geq 0$, $\mathbf{T}_{lm}^{E1}$ and $\mathbf{T}_{lm}^{B1}$ have $l \geq 1$, while $\mathbf{T}_{lm}^{E2}$ and $\mathbf{T}_{lm}^{B2}$ have $l \geq 2$.⁵¹ The above tensors are all symmetric and traceless by construction. On a traceless-symmetric tensor h_{ij} , the transversality condition $n_i h_{ij} = 0$ eliminates three degrees of freedom, and indeed we see from the explicit expressions that only $\mathbf{T}_{lm}^{E2}$ and $\mathbf{T}_{lm}^{B2}$ are transverse,

$$n_i (\mathbf{T}_{lm}^{E2})_{ij} = 0, \quad n_i (\mathbf{T}_{lm}^{B2})_{ij} = 0 \tag{3.274}$$

The five pure-spin $s = 2$ tensor harmonics are appropriate for describing the five independent components of a massive spin-2 particle. However, as we discussed in Problem 2.1 a massless

particle with quantum number s has only two components rather than $2s + 1$, with helicities $h = \pm s$. Under a rotation by an angle θ around the $\hat{\mathbf{n}}$ axis, $(\mathbf{T}_{lm}^{S0})_{ij}$ is invariant since it depends only on n_i, n_j . Therefore in the massless case it describes a spin-0 particle. Instead $(\mathbf{T}_{lm}^{E1})_{ij}$ and $(\mathbf{T}_{lm}^{B1})_{ij}$ have one index (i or j) proportional to n_i or n_j , which is invariant, while the other index is carried by a vector in the transverse plane. Therefore they combine to give the two eigenvectors of helicity with $h = \pm 1$ and describe a massless vector particle. Finally, the transverse and traceless tensors $(\mathbf{T}_{lm}^{E2})_{ij}$ and $(\mathbf{T}_{lm}^{B2})_{ij}$ have two transverse indices and combine to give rise to the states with $h = \pm 2$ which make up a massless particle with $s = 2$. Therefore, even if we use the name "spin-2 tensor harmonics", ⁵⁰As in the vector case, we now switch notation from j, j_z to l, m , but one should be aware that these indices refer to the total angular momentum; we reserve instead the notation (i, j) for the spatial indices of vectors.

⁵¹ Various useful ways of rewriting eqs. (3.268)-(3.272) can be found in eq. (2.30) of Thorne (1980). Observe also that, using the explicit form of the spherical harmonics Y_{lm} , $\mathbf{T}_{lm}^{E1}$ and $\mathbf{T}_{lm}^{B1}$, as defined in eqs. (3.269) and (3.270), vanish for $l = 0$, while $\mathbf{T}_{lm}^{E2}$ and $\mathbf{T}_{lm}^{B2}$, as defined in eqs. (3.271) and (3.272), vanish for $l = 0$ and for $l = 1$.

it is important to understand that they can be used to describe either the five polarization states of a massive spin-2 particle, or to describe a massless field; in the latter case these five degrees of freedom decompose into the two degrees of freedom of a massless particle with $h = \pm 2$, the two degrees of freedom of a massless particle with $h = \pm 1$, and one degree of freedom corresponding to a massless scalar particle. Observe that the labels 0, 1 or 2 in $S0, E1, B1, E2, B2$ refer to the (absolute value of the) helicity carried in the massless case.

In particular, in standard general relativity the graviton is a massless particle with helicity ± 2 or, equivalently, the tensor h_{ij}^{TT} that describes GWs in the TT gauge, beside being symmetric and traceless, is also transverse. Therefore in its expansion enter only $\mathbf{T}_{lm}^{E2}$ and $\mathbf{T}_{lm}^{B2}$, while the other components are eliminated by gauge invariance, as discussed in Sections 1.2 and 2.2.2. In the wave zone, where h_{ij}^{TT} decreases as $1/r$, the most general form of $h_{ij}^{\text{TT}}(t, r, \theta, \phi)$ is then

$$h_{ij}^{\text{TT}} = \frac{1}{r} \frac{G}{c^4} \sum_{l=2}^{\infty} \sum_{m=-l}^l \left[u_{lm} (\mathbf{T}_{lm}^{E2})_{ij}(\theta, \phi) + v_{lm} (\mathbf{T}_{lm}^{B2})_{ij}(\theta, \phi) \right]$$

where u_{lm} and v_{lm} are functions of retarded time $t - r/c$, and the factor G/c^4 in front is a useful normalization of u_{lm}, v_{lm} . The other pure-spin $s = 2$ tensor harmonics can enter in extensions of general relativity in which further degrees of freedom are present, and the condition of transversality no longer holds. Furthermore, in scalar-tensor extensions of general relativity, h_{ij} is not even traceless and there is a sixth degree of freedom corresponding to the trace of h_{ij} , which is a scalar field. So, in the most general case we must include all five spin-2 tensor harmonics (3.268)-(3.272), and we must further add $\delta_{ij} Y_{lm}$, which is not traceless and accounts for the scalar field corresponding to the trace part. The function $\delta_{ij} Y_{lm}$ can be combined with $(\mathbf{T}_{lm}^{S0})_{ij}$ to give a purely longitudinal and a purely transverse (but not traceless) tensor harmonic,

$$(\mathbf{T}_{lm}^{L0})_{ij} = n_i n_j Y_{lm}, \quad (\mathbf{T}_{lm}^{T0})_{ij} = \frac{1}{\sqrt{2}} (\delta_{ij} - n_i n_j) Y_{lm} \quad (3.276)$$

with $l \geq 0$. The coefficients in eqs. (3.268)-(3.272) and in eq. (3.276) are chosen so that the pure-spin harmonics are orthonormal,

$$\int d\Omega (\mathbf{T}_{lm}^J)_{ij} (\mathbf{T}_{l'm'}^{J'})_{ij}^* = \delta^{JJ'} \delta_{ll'} \delta_{mm'} \quad (3.277)$$

where the label J takes the values $L0, T0, E1, B1, E2, B2$. Finally, we observe from the explicit expressions that $\mathbf{T}_{lm}^{L0}, \mathbf{T}_{lm}^{T0}, \mathbf{T}_{lm}^{E1}$ and $\mathbf{T}_{lm}^{E2}$ have "electric-type" parity $\pi_l = (-1)^l$ while

$\mathbf{T}_{lm}^{B1}$ and $\mathbf{T}_{lm}^{B2}$ have "magnetic-type" parity $\pi_l = (-1)^{l+1}$.

Equation (3.275) is the main result of this subsection. It is the generalization to the spin-2 field h_{ij} of the more usual expansion of the solution of the wave equation for a relativistic scalar field in the wave zone, in terms of (scalar) spherical harmonics,

$$\phi(t, r, \theta, \phi) = \frac{1}{r} \sum_{l=0}^{\infty} \sum_{m=-l}^l c_{lm} Y_{lm}(\theta, \phi) \quad (3.278)$$

with c_{lm} functions of $t - r/c$, in the wave zone. Comparing eqs. (3.275) and (3.278) we see the following important differences: (1) a scalar field has only one spin degree of freedom and therefore, at each angular momentum level l, m , it is described by a single function $c_{lm}(t - r/c)$. A gravitational wave, instead, has two helicity states. Therefore there are two sets of tensor spherical harmonics, $\mathbf{T}_{lm}^{E2}$ and $\mathbf{T}_{lm}^{B2}$, and correspondingly two sets of functions $u_{lm}(t - r/c)$ and $v_{lm}(t - r/c)$. These states are transverse, see eq. (3.274), one with electric-type parity and the other with magnetic-type parity. Because of eq. (3.277), the pure-spin harmonics provides an orthonormal basis for these modes. (2) The expansion of h_{ij}^{TT} starts from total angular momentum $l = 2$. It is impossible to construct a GW with total angular momentum $l = 0$ or $l = 1$. This counting of degrees of freedom is in full agreement with the discussion of graviton states in Problem 2.2.

Our next task is to relate the coefficients u_{lm}, v_{lm} in eq. (3.275) to the appropriate multipole moments of the source, as we do in the next subsection. ⁵²

6.0.73 Relation with the source moments

In eq. (3.34) we found the solution of the equation of motion for h_{ij}^{TT} in the wave zone $r \gg d$, in the form

$$\begin{aligned} h_{ij}^{\text{TT}} &= \frac{1}{r} \frac{4G}{c^4} \Lambda_{ij,pq} \left[S^{pq} + \frac{1}{c} n_{i_1} \dot{S}^{pq, i_1} + \frac{1}{2c^2} n_{i_1} n_{i_2} \ddot{S}^{pq, i_1 i_2} + \dots \right] \\ &= \frac{1}{r} \frac{4G}{c^4} \Lambda_{ij,pq} \sum_{\alpha=0}^{\infty} \frac{1}{\alpha!} (\partial_0^\alpha S^{pq, i_1 \dots i_\alpha}) n_{i_1} \dots n_{i_\alpha} \end{aligned} \quad (3.279)$$

where it is understood that all the $S^{pq, i_1 \dots i_\alpha}$ are functions of $t - r/c$. On the other hand, in the previous section we have seen that the most general expansion for h_{ij}^{TT} in the wave zone is given by

$$h_{ij}^{\text{TT}} = \frac{1}{r} \frac{G}{c^4} \sum_{l=2}^{\infty} \sum_{m=-l}^l \left[u_{lm} (\mathbf{T}_{lm}^{E2})_{ij} + v_{lm} (\mathbf{T}_{lm}^{B2})_{ij} \right] \quad (3.280)$$

where again it is understood that u_{lm} and v_{lm} are functions of $t - r/c$. Comparing the two expressions, we can determine u_{lm} and v_{lm} . To obtain u_{lm} , we multiply both sides of eq. (3.280) by $(\mathbf{T}_{lm}^{E2})_{ij}^*$ and we integrate over $d\Omega$. On the right-hand side, using the orthonormality condition (3.277), we single out u_{lm} , while on the left-hand side we insert the expression (3.279) for h_{ij}^{TT} . Similarly, to obtain v_{lm} we multiply by ⁵² Recall that all our discussion holds in the context of linearized theory, i.e. in a flat background space-time. The determination of the GW at infinity in terms of the source moments in the full non-linear theory is a much more difficult problem, whose solution will be given in Chapter 5. In the non-linear case, we will find that the STF formalism is much more convenient, and in fact the final result that will be given in eqs. (5.135) and (5.136) is a simple generalization of eqs. (3.207) and (3.208).

⁵³ In general, if one wishes to compute the total radiated power consistently to a given order in v^2/c^2 , one cannot stop to the leading order in v^2/c^2 in the computations of u_{lm} . For instance, the subleading terms in the mass quadrupole u_{2m} are of the same order, in the radiated power, as the leading term of the mass octupole u_{3m} . Furthermore, for self-gravitating sources, we cannot use the expansion over flat space as we have done in this chapter, and the computation must be performed using the post-Newtonian formalism presented in Chapter 5. $(\mathbf{T}_{lm}^{B2})_{ij}^*$. Then we get

$$u_{lm} = \sum_{\alpha=0}^{\infty} \frac{4}{\alpha!} (\partial_0^\alpha S^{pq, i_1 \dots i_\alpha}) \int d\Omega (\mathbf{T}_{lm}^{E2})_{ij}^* \Lambda_{ij,pq} n_{i_1} \dots n_{i_\alpha} \quad (3.281)$$

$$v_{lm} = \sum_{\alpha=0}^{\infty} \frac{4}{\alpha!} (\partial_0^\alpha S^{pq, i_1 \dots i_\alpha}) \int d\Omega (\mathbf{T}_{lm}^{B2})_{ij}^* \Lambda_{ij,pq} n_{i_1} \dots n_{i_\alpha}$$

Since the Lambda tensor projects on the transverse and traceless part of a tensor, and $\mathbf{T}_{lm}^{E2}$ and $\mathbf{T}_{lm}^{B2}$ are already transverse and traceless, we have

$$(\mathbf{T}_{lm}^J)_{ij}^* \Lambda_{ij,pq} = (\mathbf{T}_{lm}^J)_{pq}^* \quad (3.282)$$

(with $J = E2, B2$), and we can simplify the above expressions,

$$u_{lm} = \sum_{\alpha=0}^{\infty} \frac{4}{\alpha!} (\partial_0^\alpha S^{ij, i_1 \dots i_\alpha}) \int d\Omega (\mathbf{T}_{lm}^{E2})_{ij}^* n_{i_1} \dots n_{i_\alpha} \quad (3.283)$$

$$v_{lm} = \sum_{\alpha=0}^{\infty} \frac{4}{\alpha!} (\partial_0^\alpha S^{ij, i_1 \dots i_\alpha}) \int d\Omega (\mathbf{T}_{lm}^{B2})_{ij}^* n_{i_1} \dots n_{i_\alpha} \quad (3.284)$$

The computation of the integrals is involved, but can be performed order by order in v/c . To leading order in v/c , we perform it in detail in Problem 3.5. ⁵³ The result is

$$u_{lm} = \frac{16\pi}{(2l+1)!!} \left[\frac{l}{2}(l-1)(l+1)(l+2) \right]^{1/2} \mathcal{Y}_{i_1 \dots i_l}^{lm*} \partial_0^{l-2} S^{i_1 i_2, i_3 \dots i_l} \times \left[1 + O\left(\frac{v^2}{c^2}\right) \right] \quad (3.285)$$

We now want to write the time derivatives of $S^{i_1 i_2, i_3 \dots i_l}$ in terms of the derivatives of the momenta of T^{00} , as follows. From the conservation of energy-momentum tensor, $\partial_\mu T^{\mu\nu} = 0$, we have the relations $\partial_0 T^{00} = -\partial_i T^{0i}$ and $\partial_0 T^{0i} = -\partial_j T^{ji}$. We can combine them to get $\partial_0^2 T^{00} = \partial_i \partial_j T^{ij}$. Then, integrating twice by parts,

$$\begin{aligned} \ddot{M}^{i_1 \dots i_l} &= \int d^3x (\partial_0^2 T^{00}) x^{i_1} \dots x^{i_l} \\ &= \int d^3x (\partial_i \partial_j T^{ij}) x^{i_1} \dots x^{i_l} \\ &= \int d^3x T^{ij} \partial_i \partial_j (x^{i_1} \dots x^{i_l}) \end{aligned} \quad (3.286)$$

For $l \geq 2$, as in our case,

$$\begin{aligned}\partial_i \partial_j (x^{i_1} \dots x^{i_l}) &= (\partial_i \partial_j x^{i_1} x^{i_2}) x^{i_3} \dots x^{i_l} + \dots \\ &= (\delta_i^{i_1} \delta_j^{i_2} + \delta_i^{i_2} \delta_j^{i_1}) x^{i_3} \dots x^{i_l} + \dots\end{aligned}\quad (3.287)$$

where the dots denote the other similar terms; in total there are $l(l-1)/2$ terms of this type. Therefore

$$\ddot{M}^{i_1 \dots i_l} = 2 (S^{i_1 i_2, i_3, \dots, i_l} + \dots) \quad (3.288)$$

where the dots denote all other $l(l-1)/2$ pairing of indices (the permutation of the first two indices i, j in $S^{ij, kl \dots}$ is already taken into account by the overall factor of 2). This is the generalization to arbitrary l of the relation $\ddot{M}^{ij} = 2S^{ij}$ found in eq. (3.52). Once we contract the left- and right-hand sides of this equation with $\mathcal{Y}_{i_1 \dots i_l}^{lm*}$, which is totally symmetric, all these permutations give the same result, so

$$\mathcal{Y}_{i_1 \dots i_l}^{lm*} S^{i_1 i_2, i_3, \dots, i_l} = \frac{1}{l(l-1)} \mathcal{Y}_{i_1 \dots i_l}^{lm*} \ddot{M}^{i_1 \dots i_l} \quad (3.289)$$

Therefore, in eq. (3.285),

$$\mathcal{Y}_{i_1 \dots i_l}^{lm*} \partial_0^{l-2} S^{i_1 i_2, i_3, \dots, i_l} = \frac{1}{c^{l-2}} \frac{1}{l(l-1)} \mathcal{Y}_{i_1 \dots i_l}^{lm*} \frac{d^l}{dt^l} M^{i_1 \dots i_l} \quad (3.290)$$

Then we get

$$u_{lm} = \frac{d^l}{dt^l} I_{lm} \quad (3.291)$$

where, to leading order in $(v/c)^2$,

$$I_{lm} = \frac{1}{c^{l-2}} \frac{16\pi}{(2l+1)!!} \left[\frac{(l+1)(l+2)}{2l(l-1)} \right]^{1/2} \mathcal{Y}_{i_1 \dots i_l}^{lm} M^{i_1 \dots i_l} \quad (3.292)$$

Similarly, repeating the same analysis for v_{lm} (and using results discussed in Problem 3.5), we get

$$v_{lm} = \frac{d^l}{dt^l} S_{lm} \quad (3.293)$$

where, again to leading order in $(v/c)^2$,

$$S_{lm} = \frac{1}{c^{l-1}} \frac{32\pi}{(2l+1)!!} \left[\frac{l(l+2)}{2(l-1)(l+1)} \right]^{1/2} \mathcal{Y}_{i_1 \dots i_{l-1}}^{lm*} \epsilon_{ijk} P^{j, ki_1 \dots i_{l-1}} \quad (3.294)$$

where $P^{j, ki_1 \dots i_{l-1}}$ are the momenta of the linear momentum. ⁵⁴

Comparing with eq. (3.238) we see that I_{lm} and S_{lm} are just the spherical components of the tensors $M^{i_1 \dots i_l}$ and $\epsilon_{ijk} P^{j, ki_1 \dots i_{l-1}}$, respectively, apart from an l -dependent normalization. The tensor $M^{i_1 \dots i_l}$ represents the moments of T^{00}/c^2 ; if the source is non-relativistic and has a negligible self-gravity, T^{00}/c^2 is the same as the mass density. Instead, from eqs. (3.40) and (3.41), we see that $\epsilon_{ijk} P^{j, k}$ is the angular momentum, and that $\epsilon_{ijk} P^{j, ki_1 \dots i_{l-1}}$ are the momenta of the angular momentum. Writing explicitly

$$\begin{aligned}M^{i_1 \dots i_l} &= \frac{1}{c^2} \int d^3x T^{00} x^{i_1} \dots x^{i_l} \\ &= \frac{1}{c^2} \int d^3x T^{00} r^l n^{i_1} \dots n^{i_l}\end{aligned}\quad (3.295)$$

and using eq. (3.234), we can rewrite eq. (3.292) as

$$I_{lm} = \frac{1}{c^l} \frac{16\pi}{(2l+1)!!} \left[\frac{(l+1)(l+2)}{2l(l-1)} \right]^{1/2} \int d^3x r^l T^{00} Y_{lm}^* \quad (3.296)$$

again to leading order in v^2/c^2 . Similarly we find

$$S_{lm} = \frac{1}{c^l} \frac{32\pi}{(2l+1)!!} \left[\frac{l(l+2)}{2(l-1)(l+1)} \right]^{1/2} \times \epsilon_{j p q} \int d^3x r^l T^{0p} \left(\mathcal{Y}_{j i_1 \dots i_{l-1}}^{lm} \right)^* n^q n^{i_1} \dots n^{i_{l-1}} \quad (3.297)$$

This expression can be written in terms of the vector spherical harmonic $\mathbf{Y}_{lm}^B$ defined in eq. (3.255). In fact, inserting eq. (3.234) in the definition of this vector harmonic, one finds

$$(\mathbf{Y}_{lm}^B)_i = \left(\frac{l}{l+1} \right)^{1/2} \epsilon_{ijk} \mathcal{Y}_{k i_1 \dots i_{l-1}}^{lm} n^j n^{i_1} \dots n^{i_{l-1}} \quad (3.298)$$

and therefore, to leading order in $(v/c)^2$,

$$S_{lm} = \frac{1}{c^l} \frac{32\pi}{(2l+1)!!} \left[\frac{(l+2)}{2(l-1)} \right]^{1/2} \int d^3x r^l T^{0i} (\mathbf{Y}_{lm}^B)_i^* \quad (3.299)$$

The coefficients of the expansion of h_{ij}^{TT} have therefore been written as integrals over the source of quantities which depend on the energymomentum tensor.

6.0.74 Radiated power

We can finally write the radiated power to all orders in the multipole expansion. In eq. (3.275) we have expressed h_{ij}^{TT} in the basis of $\mathbf{T}_{lm}^{E2}, \mathbf{T}_{lm}^{B2}$. Inserting eq. (3.275) into eq. (1.153) and using the orthonormality relation (3.277), we find

$$\int d\Omega \dot{h}_{ij}^{\text{TT}} \dot{h}_{ij}^{\text{TT}} = \frac{1}{r^2} \frac{G^2}{c^8} \sum_{l=2}^{\infty} \sum_{m=-l}^l [|\dot{u}_{lm}|^2 + |\dot{v}_{lm}|^2] \quad (3.300)$$

Therefore

$$\frac{dE}{dt} = \frac{G}{32\pi c^5} \sum_{l=2}^{\infty} \sum_{m=-l}^l \left\langle \left| \frac{d^{l+1} I_{lm}}{dt^{l+1}} \right|^2 + \left| \frac{d^{l+1} S_{lm}}{dt^{l+1}} \right|^2 \right\rangle \quad (3.301)$$

where I_{lm} and S_{lm} are functions of retarded time $t - r/c$. The power is therefore a sum of terms each one associated with a single mass multipole or angular momentum multipole, and there are no mixed terms.

As a check of the above result, we can verify that we reproduce the mass quadrupole and the current quadrupole radiation. We start from the mass quadrupole. Equation (3.292), with $l = 2$, gives

⁵⁴ The expressions that we have computed in this section for I_{lm} and S_{lm} are valid only to leading order in v/c , i.e. they are Newtonian expressions. The full expansion in v/c is studied in Thorne (1980), Section V.

$$\begin{aligned}
I_{2m} &= \frac{16\pi}{5\sqrt{3}} \mathcal{Y}_{ij}^{2m*} M_{ij} \\
&= \frac{16\pi}{5\sqrt{3}} \mathcal{Y}_{ij}^{2m*} Q_{ij}
\end{aligned} \tag{3.302}$$

where we could replace M_{ij} by Q_{ij} since $\mathcal{Y}_{ij}^{2m*}$ is traceless in the (i, j) indices. From eq. (3.301), the power radiated by the quadrupole is

$$P_{\text{quad}} = \frac{G}{32\pi c^5} \left(\frac{16\pi}{5\sqrt{3}} \right)^2 \langle \ddot{Q}_{ij} \ddot{Q}_{kl} \rangle \sum_{m=-2}^2 \mathcal{Y}_{ij}^{2m*} \mathcal{Y}_{kl}^{2m} \tag{3.303}$$

We now use the identity ⁵⁵

$$\sum_{m=-2}^2 \mathcal{Y}_{ij}^{2m*} \mathcal{Y}_{kl}^{2m} = \frac{15}{16\pi} \left(\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk} - \frac{2}{3} \delta_{ij} \delta_{kl} \right) \tag{3.304}$$

and we correctly recover eq. (3.75). The current quadrupole is checked similarly: from eq. (3.294) we have

$$S_{2m} = \frac{1}{c} \frac{64\pi}{15\sqrt{3}} \mathcal{Y}_{il}^{2m*} \epsilon_{ijk} P^{j,kl} \tag{3.305}$$

Since

$$\begin{aligned}
\epsilon_{ijk} P^{j,kl} &= \epsilon_{ijk} \int d^3x T^{0j} x^k x^l \\
&= - \int d^3x (\epsilon_{ikj} x^k T^{0j}) x^l \\
&= - \int d^3x j^i x^l \\
&= -J^{i,l}
\end{aligned} \tag{3.306}$$

we get

$$S_{2m} = -\frac{1}{c} \frac{64\pi}{15\sqrt{3}} \mathcal{Y}_{ij}^{2m*} J_{i,j} \tag{3.307}$$

Then, using again the identity (3.304), and recalling that $J_{i,j} \delta_{ij} = 0$,

$$\begin{aligned}
P_{\text{curr quad}} &= \frac{G}{32\pi c^7} \left(\frac{64\pi}{15\sqrt{3}} \right)^2 \langle \ddot{J}_{i,j} \ddot{J}_{k,l} \rangle \sum_{m=-2}^2 \mathcal{Y}_{ij}^{2m*} \mathcal{Y}_{kl}^{2m*} \\
&= \frac{8G}{45c^7} \langle \ddot{J}_{i,j} \ddot{J}_{k,l} \rangle (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}) \\
&= \frac{8G}{45c^7} \langle \ddot{J}_{i,j} (\ddot{J}_{i,j} + \ddot{J}_{j,i}) \rangle \\
&= \frac{16G}{45c^7} \langle \ddot{J}_{ij} \ddot{J}_{ij} \rangle
\end{aligned} \tag{3.308}$$

in agreement with eq. (3.154). ⁵⁵ This identity could be proved using the explicit expressions (3.218) for $\mathcal{Y}_{ij}^{2m}$. However, the simplest way to derive it is to observe that, after summing over

$m = -2, \dots, 2$, there is no longer a dependence on the direction chosen as quantization axis for j_z , so the right-hand side of eq. (3.304) can only depend on combinations of Kronecker deltas, i.e. it must be of the form $a\delta_{ik}\delta_{jl} + b\delta_{il}\delta_{jk} + c\delta_{ij}\delta_{kl}$. Since the left-hand side of eq. (3.304) gives zero when contracted with δ_{ij} , we must have $\delta_{ij}(a\delta_{ik}\delta_{jl} + b\delta_{il}\delta_{jk} + c\delta_{ij}\delta_{kl}) = 0$, which fixes $c = -(1/3)(a+b)$. Then a and b are fixed comparing the left and right-hand side of eq. (3.304) for two different values of the indices, using the explicit expressions (3.218).

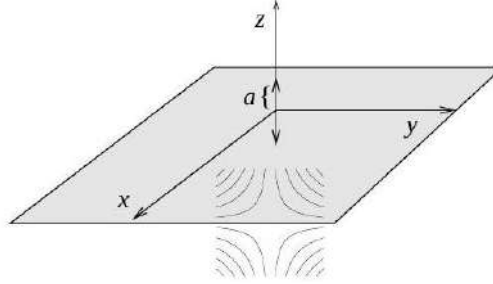


Fig. 3.4 A source oscillating along the z axis (double arrow), and the lines of force of the GW emitted in a direction with $\theta = \pi/2$.

6.0.75 3.6 Solved problems

6.0.76 Problem 3.1. Quadrupole radiation from an oscillating mass

As a first simple application, we compute the quadrupole radiation emitted by a non-relativistic degree of freedom z_0 that performs harmonic oscillations along the z axis,

$$z_0(t) = a \cos \omega_s t \quad (3.309)$$

with $a\omega_s \ll c$ and $\omega_s > 0$. The actual physical system could be made for instance by two masses connected by a spring, with $z_0(t)$ the relative coordinate in the center-of mass system. For the moment we consider the case where the rest length of the spring is zero (which is not realistic for a one-dimensional spring, but is representative of a number of situations where some degree of freedom performs a simple harmonic motion). As discussed in Section 3.3.5, the whole formalism that we have developed for computing the emission of gravitational radiation is consistent only if we include in the energy-momentum tensor also the interaction terms responsible for the actual trajectory, which however are subleading when computing the mass quadrupole radiation to lowest order in v/c . The mass density is then

$$\rho(t, \mathbf{x}) = \mu \delta(x) \delta(y) \delta(z - z_0(t)) \quad (3.310)$$

where μ is the reduced mass of the system, and the second mass moment is

$$\begin{aligned} M^{ij}(t) &= \int d^3x \rho(t, \mathbf{x}) x^i x^j \\ &= \mu z_0^2(t) \delta^{i3} \delta^{j3} \\ &= \mu a^2 \frac{1 + \cos 2\omega_s t}{2} \delta^{i3} \delta^{j3} \end{aligned} \quad (3.311)$$

Inserting this into eq. (3.72), we obtain

$$\begin{aligned}
h_+(t; \theta, \phi) &= -\frac{1}{r} \frac{G}{c^4} \ddot{M}_{33}(t_{\text{ret}}) \sin^2 \theta \\
&= \frac{2G\mu a^2 \omega_s^2}{rc^4} \sin^2 \theta \cos(2\omega_s t_{\text{ret}})
\end{aligned} \tag{3.312}$$

$$h_\times(t; \theta, \phi) = 0 \tag{3.313}$$

Therefore we have monochromatic radiation at a frequency $\omega = 2\omega_s$, with a purely plus polarization, see Fig. 3.4.

The angular distribution is independent of ϕ (reflecting the cylindric symmetry of the source), and has a maximum at $\theta = \pi/2$, i.e. in the direction orthogonal to the axis along which the source oscillates. Observe that the radiation vanishes along the z axis. This reflects the fact that only the components of the motion of the source transverse to the line-of-sight contribute to the production of GWs. This is a general result, which follows from the fact that $\Lambda_{ij,kl}n^k = \Lambda_{ij,kl}n^l = 0$. That is, the Lambda tensor projects the motion of the source onto the plane transverse to the propagation direction.

Observe also that the pattern of lines of force of the GW shown in Fig. 3.4 is a physical result, independent of our conventions. The fact that we call it a "plus" polarization, instead, is related to our choice of axes $(\hat{\mathbf{u}}, \hat{\mathbf{v}})$ with respect to which the plus and cross polarizations are defined. With our definition, $(\hat{\mathbf{u}}, \hat{\mathbf{v}})$ are obtained from the $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ axes applying the rotation matrix $\mathcal{R}$ given

in eq. (3.70), see also Fig. 3.2. Thus, in Fig. 3.4, for the particular propagation direction for which the lines of force are shown (i.e. for a direction $\hat{\mathbf{n}}$ such that $\theta = \pi/2$), we have $\hat{\mathbf{v}} = -\hat{\mathbf{z}}$, while $\hat{\mathbf{u}}$ is along the intersection of this transverse plane with the (x, y) plane. If, in this transverse plane, we instead used $(\hat{\mathbf{u}}, \hat{\mathbf{v}})$ axes rotated by 45 degrees to define h_+ and $h_\times$, we would rather call the pattern of Fig. 3.4 a purely cross polarization, compare with Figs. 1.2 and 1.3, while for axes rotated by a generic angle ψ we would have a mixture of plus and cross polarizations, according to eq. (2.194).

The radiated power is computed from eq. (3.73),

$$\begin{aligned}
\left(\frac{dP}{d\Omega}\right)_{\text{quad}} &= \frac{r^2 c^3}{16\pi G} \langle \dot{h}_+^2 \rangle \\
&= \frac{G\mu^2 a^4 \omega_s^6}{2\pi c^5} \sin^4 \theta
\end{aligned} \tag{3.314}$$

where we used $\langle \cos^2(2\omega_s t) \rangle = 1/2$. Alternatively, we can recover the same result using eq. (3.73) in the form

$$\left(\frac{dP}{d\Omega}\right)_{\text{quad}} = \frac{G}{8\pi c^5} \Lambda_{33,33}(\hat{\mathbf{n}}) \langle \ddot{M}_{33}^2 \rangle \tag{3.315}$$

and observing that, from eq. (1.39),

$$\begin{aligned}
\Lambda_{33,33} &= \frac{1}{2} (1 - n_3^2)^2 \\
&= \frac{1}{2} \sin^4 \theta
\end{aligned} \tag{3.316}$$

since $n_3 = \cos \theta$. In Fig. 3.5 we show this angular distribution, in the (x, z) plane. The integration over the solid angle gives

$$P_{\text{quad}} = \frac{16}{15} \frac{G\mu^2}{c^5} a^4 \omega_s^6 \quad (3.317)$$

The total energy radiated over one period $T = 2\pi/\omega_s$ of the source motion is therefore

$$\langle E_{\text{quad}} \rangle_T = \frac{32\pi}{15} \frac{G\mu^2}{c^5} a^4 \omega_s^5 \quad (3.318)$$

This result becomes physically more transparent if we rewrite it in terms of $v = a\omega_s$ (which is the maximum speed of the source),

$$\langle E_{\text{quad}} \rangle_T = \frac{32\pi}{15} \frac{G\mu^2}{a} \left(\frac{v}{c}\right)^5 \quad (3.319)$$

Observe that $G\mu^2/a$ is of order of the gravitational self-energy of an object of mass μ and size a . In the quadrupole approximation, the energy radiated over a cycle is suppressed, with respect to this energy scale, by a factor $(v/c)^5$.

Finally, it is instructive to consider the case of two masses connected by a spring with a rest length L , so their relative coordinate obeys

$$z_0(t) = L + a \cos \omega_s t \quad (3.320)$$

Observe that L has an invariant meaning, because $z_0(t)$ is the relative position between the two masses, so we cannot set it to zero with a choice of the origin. Now we get

$$z_0^2(t) = \frac{a^2}{2} \cos 2\omega_s t + 2La \cos \omega_s t + \text{const.} \quad (3.321)$$

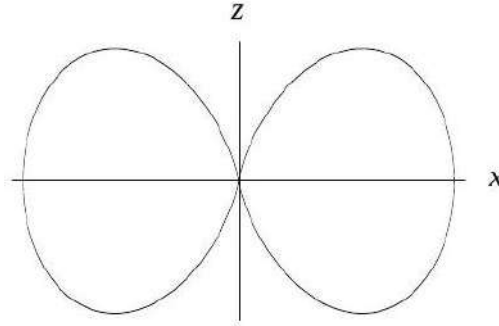


Fig. 3.5 The angular distribution of the quadrupole radiation, for a mass oscillating along the z axis. We represent the angular distribution plotting the function $\rho = \sin^4 \theta$, where (ρ, θ) are the polar coordinates in the (x, z) plane and θ is measured from the z axis. The full three-dimensional pattern has cylindrical symmetry around the z axis. and eq. (3.312) becomes

$$h_+(t; \theta, \phi) = \frac{2G\mu\omega_s^2}{rc^4} \sin^2 \theta \left[a^2 \cos(2\omega_s t_{\text{ret}}) + La \cos(\omega_s t_{\text{ret}}) \right] \quad (3.322)$$

while still $h_\times = 0$. Thus, beside having gravitational radiation at $\omega_{\text{gw}} = 2\omega_s$, we also have radiation at $\omega_{\text{gw}} = \omega_s$. Observe that, in the power, there is no interference between these two terms because $\langle \cos^2(2\omega_s t) \rangle = 1/2 = \langle \cos^2(\omega_s t) \rangle$, but

$$\langle 2 \cos(2\omega_s t) \cos(\omega_s t) \rangle = \langle \cos(3\omega_s t) + \cos(\omega_s t) \rangle = 0 \quad (3.323)$$

Problem 3.2. Quadrupole radiation from a mass in circular orbit

In this problem we consider a binary system with masses m_1 and m_2 , and we assume that the relative coordinate is performing a circular motion. We assume for the moment that the orbital motion is given, and we neglect any back-reaction on the motion due to GW emission. In this form, this is just a simple exercise, propedeutic for understanding a real self-gravitating binary system. In Section 4.1 we will include the effect of the GW back-reaction within linearized theory. Furthermore, at higher orders in v/c we cannot keep the space-time as flat when describing a self-gravitating system, and the correct formalism for computing the v/c corrections will be the subject of Chapter 5.

So, for the moment, we rather assign ourselves the trajectory. We choose the (x, y, z) frame so that the orbit lies in the (x, y) plane, and is given by

$$\begin{aligned} x_0(t) &= R \cos\left(\omega_s t + \frac{\pi}{2}\right) \\ y_0(t) &= R \sin\left(\omega_s t + \frac{\pi}{2}\right) \\ z_0(t) &= 0 \end{aligned} \quad (3.324)$$

(The phase $\pi/2$ is a useful choice of the origin of time.) We denote by $\mu = m_1 m_2 / (m_1 + m_2)$ the reduced mass of the system. From eq. (3.129), in the CM frame the second mass moment is $M^{ij} = \mu x_0^i(t) x_0^j(t)$, so

$$M_{11} = \mu R^2 \frac{1 - \cos 2\omega_s t}{2} \quad (3.325)$$

$$M_{22} = \mu R^2 \frac{1 + \cos 2\omega_s t}{2} \quad (3.326)$$

$$M_{12} = -\frac{1}{2} \mu R^2 \sin 2\omega_s t \quad (3.327)$$

while the other components vanish. Therefore we have

$$\ddot{M}_{11} = 2\mu R^2 \omega_s^2 \cos 2\omega_s t, \quad (3.328)$$

$$\ddot{M}_{12} = 2\mu R^2 \omega_s^2 \sin 2\omega_s t, \quad (3.329)$$

and $\ddot{M}_{22} = -\ddot{M}_{11}$. Plugging these expressions into eq. (3.72) we get

$$h_+(t; \theta, \phi) = \frac{1}{r} \frac{4G\mu\omega_s^2 R^2}{c^4} \left(\frac{1 + \cos^2 \theta}{2} \right) \cos(2\omega_s t_{\text{ret}} + 2\phi) \quad (3.330)$$

$$h_\times(t; \theta, \phi) = \frac{1}{r} \frac{4G\mu\omega_s^2 R^2}{c^4} \cos \theta \sin(2\omega_s t_{\text{ret}} + 2\phi) \quad (3.331)$$

Thus, the quadrupole radiation is at twice the frequency ω_s of the source. It is interesting to observe that the dependences $h_+ \sim (1 + \cos^2 \theta)$ and $h_\times \sim \cos \theta$ are a general consequence of eq. (3.72), whenever $M_{13} = M_{23} = M_{33} = 0$ and $\ddot{M}_{22} = -\ddot{M}_{11}$. These conditions are satisfied also in other problems (e.g. a rigid body rotating around one of its principal axis), so we will meet again this type of angular dependence.

As for the dependence on ϕ , it can be clearly understood physically: contrary to the oscillating mass of the previous problem, now the source is not invariant under rotations around the z axis, since at any given value of t the mass μ is at a specific point along the orbit, which is changed by a rotation around the z axis; therefore h_+ and $h_\times$ have a dependence on ϕ . However,

since the orbit is circular, a rotation of the source by an angle $\Delta\phi$ around the z axis is the same as a time translation Δt with $\omega_s \Delta t = \Delta\phi$, and therefore the dependence of h_+ and $h_\times$ on ϕ is only through the combination $\omega_s t_{\text{ret}} + \phi$.

From the observational point of view, we have only access to the radiation that a binary star emits in the direction which points from the star toward us. The angle θ is therefore equal to the angle ι between the normal to the orbit and the line-of-sight (see Fig. 3.6). The distance r to an astrophysical source is, for most practical purposes, a constant.⁵⁶ As long as, during the observation, we can neglect the proper motion of the source (which however sometimes is not the case, as we will discuss in Section 7.6), also the angle ϕ is fixed, so we have $\omega_s t_{\text{ret}} + \phi = \omega_s t + \alpha$, with $\alpha = \phi - \omega_s r/c$ a fixed constant. Then we can shift the origin of time so that $2\omega_s t + 2\alpha \rightarrow 2\omega_s t$ plus an integer multiple of 2π , so $\cos(2\omega_s t + 2\alpha) \rightarrow \cos(2\omega_s t)$ and $\sin(2\omega_s t + 2\alpha) \rightarrow \sin(2\omega_s t)$. Therefore, this observer can write the GWs received by a binary system (as long as the approximation of a fixed circular orbit is valid, see Section 4.1) as

$$\begin{aligned} h_+(t) &= \frac{1}{r} \frac{4G\mu\omega_s^2 R^2}{c^4} \left(\frac{1 + \cos^2 \iota}{2} \right) \cos(2\omega_s t) \\ h_\times(t) &= \frac{1}{r} \frac{4G\mu\omega_s^2 R^2}{c^4} \cos \iota \sin(2\omega_s t) \end{aligned} \quad (3.332)$$

If we see the orbit edge-on, $\iota = \pi/2$, then $h_\times$ vanishes and therefore the GW is linearly polarized. Instead, at $\iota = 0$, h_+ and $h_\times$ have the same amplitude; in this case, since the former is proportional to $\sin(2\omega_s t_{\text{ret}})$ while the latter to $\cos(2\omega_s t_{\text{ret}})$, in the plane $(h_+, h_\times)$ the radiation describes a circle parametrized by t , that is, the radiation is circularly polarized. To understand what this means in more physical terms, we consider the pattern of lines of force corresponding to a circular polarization. If we have a purely plus polarization, $h_+ = A_+ \cos 2\omega_s t$, according to eq. (1.96) we have a force field

$$\mathbf{F}_+ = -2\mu\omega_s^2 A_+ \mathbf{v}_+ \cos 2\omega_s t \quad (3.333)$$

where the vector field $\mathbf{v}_+(x, y)$ has components $(x, -y)$. This force field is shown in Fig. 1.2. If instead we have a purely cross polarization of the form $h_\times = A_\times \sin 2\omega_s t$, according to eq. (1.96) we have a force field

$$\mathbf{F}_\times = -2\mu\omega_s^2 A_\times \mathbf{v}_\times \sin 2\omega_s t \quad (3.334)$$

where the vector field $\mathbf{v}_\times(x, y)$ has components (y, x) . This force field is shown in Fig. 1.3. If we have both $h_+ = A_+ \cos 2\omega_s t$ and $h_\times = A_\times \sin 2\omega_s t$,⁵⁶ As we will discuss in Section 7.6 there are situations in which we must take into account corrections due to motion of the Earth around the Sun, or more precisely around the Solar System Barycenter (SSB). So, more generally, we can take r to be the distance from the source to the SSB. This however is only important for observations lasting at least few months. We will see in Section 4.1 that, in ground-based interferometers, the gravitational waves emitted by a binary system are observable only for about the last 15 minutes before the system coalesces.

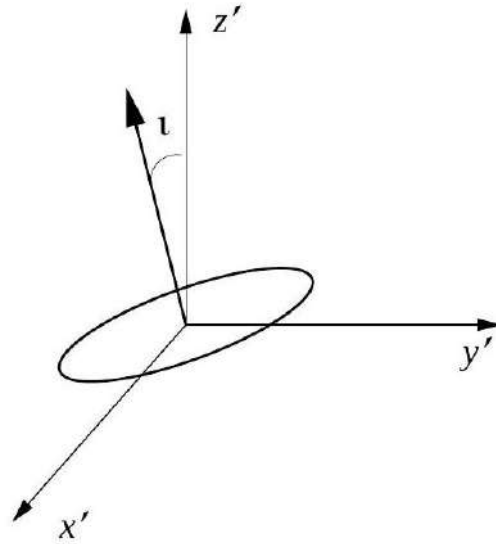


Fig. 3.6 The geometry of the problem in a frame (x', y', z') where a fixed observer is at large distance along the positive z' axis. The normal to the orbit makes an angle ι with the z' axis.

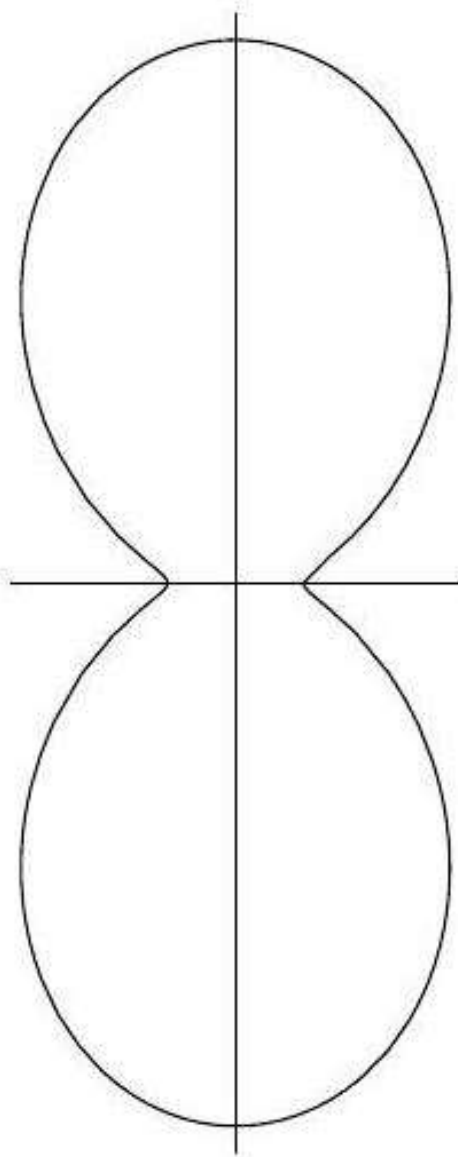


Fig. 3.7 The function $g(\theta)$ in polar coordinates. The angle θ is measured from the vertical axis.

⁵⁷ If one is interested only in the total power, rather than in the angular distribution, it is actually simpler to derive it directly from eq. (3.75): we use again the reference frame where the orbit is given by eq. (3.324), so

$$\ddot{M}_{ij} = 4\mu\omega_s^3 R^2 \times \begin{pmatrix} -\sin 2\omega_s t & \cos 2\omega_s t & 0 \\ \cos 2\omega_s t & \sin 2\omega_s t & 0 \\ 0 & 0 & 0 \end{pmatrix}_{ij}$$

Since this is traceless, it is also equal to $\ddot{Q}_{ij}$. Plugging this into eq. (3.75), we recover eq. (3.339). and furthermore $A_+ = A_\times \equiv A$, we have

$$\mathbf{F} = -2\mu\omega_s^2 A [\mathbf{v}_+ \cos 2\omega_s t + \mathbf{v}_\times \sin 2\omega_s t] \quad (3.335)$$

Observe that $\mathbf{v}_+$ and $\mathbf{v}_\times$ are orthogonal, $\mathbf{v}_+ \cdot \mathbf{v}_\times = 0$, and therefore the pattern of lines of forces described by eq. (3.335) is the same as the pattern of Fig. 1.2, that rotates uniformly so that at $\omega_s t = 0$ it is the same as Fig. 1.2, at $\omega_s t = \pi/4$ it is the same as Fig. 1.3, and so on.

At intermediate values of ι the amplitudes for h_+ and $h_\times$ are different and therefore we have elliptic polarization, i.e. in the plane $(h_+, h_\times)$ the radiation describes an ellipse parametrized by t . We see that, from a measurement of the degree of polarization, i.e. of the relative amplitude of h_+ and $h_\times$, we can deduce the inclination ι of the orbit.

The angular distribution of the radiated power, in the quadrupole approximation, is obtained as usual using eq. (3.73),

$$\left(\frac{dP}{d\Omega}\right)_{\text{quad}} = \frac{r^2 c^3}{16\pi G} \langle \dot{h}_+^2 + \dot{h}_\times^2 \rangle \quad (3.336)$$

Inserting here eq. (3.332), and using $\langle \cos^2(2\omega_s t) \rangle = \langle \sin^2(2\omega_s t) \rangle = 1/2$, we get

$$\left(\frac{dP}{d\Omega}\right)_{\text{quad}} = \frac{2G\mu^2 R^4 \omega_s^6}{\pi c^5} g(\theta) \quad (3.337)$$

where

$$g(\theta) = \left(\frac{1 + \cos^2 \theta}{2}\right)^2 + \cos^2 \theta \quad (3.338)$$

The radiation is maximum at $\theta = 0$, i.e. in the direction normal to the plane of the orbit. A polar plot of $g(\theta)$ is shown in Fig. 3.7. Observe that, contrary to the angular distribution found in Problem 3.1, $g(\theta)$ never vanishes since, whatever the angle θ at which an observer is located, there is always a component of the source motion orthogonal to the observer's line-of-sight.

Integrating eq. (3.337) over the solid angle we get the total power radiated in the quadrupole approximation,

$$\begin{aligned} P_{\text{quad}} &= \frac{32}{5} \frac{G\mu^2}{c^5} R^4 \omega_s^6 \\ &= \frac{1}{10} \frac{G\mu^2}{c^5} R^4 \omega_s^6 \end{aligned} \quad (3.339)$$

where $\omega = 2\omega_s$ is the frequency of the GW. ⁵⁷ The energy radiated in one period $T = 2\pi/\omega_s$ of the source motion is therefore, writing $v = \omega_s R$,

$$\langle E_{\text{quad}} \rangle_T = \frac{64\pi}{5} \frac{G\mu^2}{R} \left(\frac{v}{c}\right)^5 \quad (3.340)$$

Similarly to the result found in eq. (3.319), this is suppressed by a factor $(v/c)^5$ with respect to the energy scale $G\mu^2/R$.

Problem 3.3. Mass octupole and current quadrupole radiation from a mass in circular orbit

In this problem we compute the mass octupole and current quadrupole radiation generated by a binary system of reduced mass μ , whose center-of-mass coordinate describes a circular trajectory. As discussed in Section 3.3.5, to lowest order in v/c the mass octupole and the current quadrupole radiation emitted by a self-gravitating system can be consistently computed taking the free-particle energy-momentum tensor, evaluating it on a Newtonian circular orbit, and plugging the resulting values of T^{00} and T^{0i} into the expression for $\ddot{M}^{ijk}$ and $\ddot{P}^{i,jk}$ (while the same procedure is not correct if applied to $\dot{S}^{ij,k}$). Defining as usual the center-of-mass coordinate $\mathbf{x}_{\text{CM}}$ by $m\mathbf{x}_{\text{CM}} = m_1\mathbf{x}_1 + m_2\mathbf{x}_2$ (where $m = m_1 + m_2$) and the relative coordinate $\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$, we have $\mathbf{x}_1 = \mathbf{x}_{\text{CM}} + (m_2/m)\mathbf{x}$ and $\mathbf{x}_2 = \mathbf{x}_{\text{CM}} - (m_1/m)\mathbf{x}$. Thus, in the CM frame where $\mathbf{x}_{\text{CM}} = 0$, $T^{00}(t, \mathbf{x}) = m_1 c^2 \delta^{(3)}(\mathbf{x} - \frac{m_2}{m}\mathbf{x}_0(t)) + m_2 c^2 \delta^{(3)}(\mathbf{x} + \frac{m_1}{m}\mathbf{x}_0(t))$, and

$$M^{ijk}(t) = \frac{1}{c^2} \int d^3x T^{00}(t, \mathbf{x}) x^i(t) x^j(t) x^k(t) = \mu \frac{\delta m}{m} x_0^i(t) x_0^j(t) x_0^k(t)$$

where $\delta m = m_2 - m_1$. To compute the radiation emitted from the star in the direction of the observer it is simpler to use the geometrical setting of Fig. 3.6 (labeling now the axes of this figure as (x, y, z) rather than (x', y', z')), in which the observer is along the z axis. The equation of the orbit of the relative coordinate in the center-of-mass frame is then

$$\begin{aligned} x_0(t) &= R \cos \omega_s t \\ y_0(t) &= R \cos \iota \sin \omega_s t \\ z_0(t) &= R \sin \iota \sin \omega_s t \end{aligned} \quad (3.341)$$

and is obtained from an orbit lying in the (x, y) plane performing a rotation by an angle ι around the x axis. We set the observer in the z direction, so we compute the radiation emitted along $\hat{\mathbf{n}} = (0, 0, 1)$. For the octupole radiation, eq. (3.141) gives

$$(h_{ij}^{\text{TT}})_{\text{oct}} = \frac{1}{r} \frac{2G}{3c^5} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \ddot{M}_{kl3} \quad (3.342)$$

(As usual, in actual computations, it is more convenient to use M^{ijk} rather than $\mathcal{O}^{ijk}$.) We found in eq. (3.63) that, when $\hat{\mathbf{n}} = (0, 0, 1)$, in the multiplication of a matrix $\ddot{M}_{kl3}$ by the Lambda tensor the components of $\ddot{M}_{kl3}$ with $k = 3$ or $l = 3$ do not contribute, so we just need to compute M_{ab3} with $a, b = 1, 2$. With the trajectory given in eq. (3.341) we find, for $a, b = 1, 2$

$$M_{ab3} = \mu \frac{\delta m}{m} R^3 \sin \iota \sin \omega_s t \begin{pmatrix} \cos^2 \omega_s t & \cos \iota \sin \omega_s t \cos \omega_s t \\ \cos \iota \sin \omega_s t \cos \omega_s t & \cos^2 \iota \sin^2 \omega_s t \end{pmatrix}_{ab}.$$

Computing $\Lambda_{ab,cd}(\hat{\mathbf{n}}) M_{cd3}$ using eq. (3.63) we get

$$\begin{aligned} \Lambda_{ab,cd}(\hat{\mathbf{n}}) M_{cd3} &= \mu \frac{\delta m}{m} R^3 \sin \iota \sin \omega_s t \times \\ &\times \begin{pmatrix} \frac{1}{2} (\cos^2 \omega_s t - \cos^2 \iota \sin^2 \omega_s t) & \cos \iota \sin \omega_s t \cos \omega_s t \\ \cos \iota \sin \omega_s t \cos \omega_s t & -\frac{1}{2} (\cos^2 \omega_s t - \cos^2 \iota \sin^2 \omega_s t) \end{pmatrix}_{ab} \end{aligned} \quad (3.344)$$

and taking the third time derivative we find

$$\begin{aligned} (h_+)_{\text{oct}} &= \frac{1}{r} \frac{G\mu R^3 \omega_s^3}{12c^5} \frac{\delta m}{m} \sin \iota [(3 \cos^2 \iota - 1) \cos \omega_s t - 27 (1 + \cos^2 \iota) \cos 3\omega_s t], \\ (h_\times)_{\text{oct}} &= \frac{1}{r} \frac{G\mu R^3 \omega_s^3}{12c^5} \frac{\delta m}{m} \sin 2\iota [\sin \omega_s t - 27 \sin 3\omega_s t]. \end{aligned}$$

(*) The apparent difference by an overall minus sign is due to a sign convention explained in Note 19 on page 250. Observe also that, in the full treatment of the problem, the argument of the trigonometric function is not $\omega_s t$ but $\phi(t) = \int_0^t dt' \omega_s(t')$, since the frequency ω_s changes because of the energy loss due to GW emission. Further higherorder effects (tail terms) are included in the phase, see eq. (5.265), as we will see in Chapter 5.

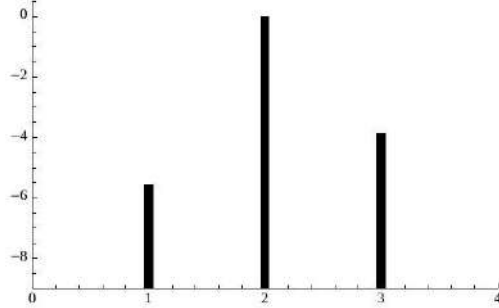


Fig. 3.8 $\log_{10} [P(\omega)/P(2\omega_s)]$, as a function of ω/ω_s , for $v/c = 10^{-2}$ and $\delta m/m = 1$, including the contributions of the mass quadrupole, of the mass octupole, and of the current quadrupole. The line at $\omega = 2\omega_s$ is due to the mass quadrupole, the line at $\omega = \omega_s$ is due to the mass octupole and current quadrupole, while that at $\omega = 3\omega_s$ is due only to the mass octupole. As expected, we have radiation both at $\omega = \omega_s$ and at $\omega = 3\omega_s$, since the mass octupole is trilinear in $x_0^i(t)$. The current quadrupole radiation can be computed similarly. Plugging $T^{0i}(t, \mathbf{x}) = m_1 c v_1^i \delta^{(3)}(\mathbf{x} - \frac{m_2}{m} \mathbf{x}_0(t)) + m_2 c v_2^i \delta^{(3)}(\mathbf{x} + \frac{m_1}{m} \mathbf{x}_0(t))$ into eqs. (3.148) and (3.149) we get

$$J^{i,j}(t) = \mu \frac{\delta m}{m} \epsilon^{ikl} x_0^k(t) \dot{x}_0^l(t) x_0^j(t) \quad (3.346)$$

Observe that $l^i \equiv \epsilon^{ikl} x_0^k(t) \dot{x}_0^l(t)$ is the angular momentum of a unit mass moving on the circular orbit $x_0^i(t)$, and is therefore a constant vector of modulus $R^2 \omega_s$ and direction normal to the plane of the orbit. Therefore $J^{i,j}(t) = \mu(\delta m/m) l^i x_0^j(t)$ depends on time only through $x_0^j(t)$ and oscillates only at the frequency ω_s , rather than at ω_s and $3\omega_s$ as the mass octupole. Inserting this expression for $J^{i,j}(t)$ into eq. (3.151) and setting $\hat{\mathbf{n}} = (0, 0, 1)$ we get the current quadrupole radiation

$$(h_{ij}^{\text{TT}})_{\text{cq}} = \frac{1}{r} \frac{4G}{3c^5} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \left(\epsilon^{3kp} \ddot{J}^{p,l} + \epsilon^{3lp} \ddot{J}^{p,k} \right) \quad (3.347)$$

Performing the contraction with the Lambda tensor we get

$$\begin{aligned} (h_+)_{\text{cq}} &= \frac{1}{r} \frac{4G\mu R^3 \omega_s^3}{3c^5} \frac{\delta m}{m} \sin \iota \cos \omega_s t \\ (h_\times)_{\text{cq}} &= \frac{1}{r} \frac{2G\mu R^3 \omega_s^3}{3c^5} \frac{\delta m}{m} \sin 2\iota \sin \omega_s t \end{aligned}$$

Summing the mass octupole and current quadrupole radiation, using Kepler's law $\omega_s^2 R^3 = Gm$, and introducing the notation $x = (Gm\omega_s/c^3)^{2/3}$ that is useful to make contact with the results of Chapter 5, we finally get

$$\begin{aligned} (h_+)_{\text{oct+cq}} &= \frac{G\mu}{4rc^2} \frac{\delta m}{m} x^{3/2} \sin \iota \left[(\cos^2 \iota + 5) \cos \omega_s t - 9(1 + \cos^2 \iota) \cos 3\omega_s t \right] \\ (h_\times)_{\text{oct+cq}} &= \frac{3G\mu}{4rc^2} \frac{\delta m}{m} x^{3/2} \sin 2\iota \left[\sin \omega_s t - 3 \sin 3\omega_s t \right] \end{aligned} \quad (3.348)$$

This result agrees with the one that we will find in Chapter 5 from a complete post-Newtonian treatment, see eqs. (5.262), (5.266) and (5.267).^(*)

The contribution to the total radiated power from the mass octupole and the current quadrupole is

$$\begin{aligned} P_{\text{oct+cq}} &= \frac{r^2 c^3}{16\pi G} 2\pi \int_{-1}^1 d\cos\iota \left\langle \dot{h}_+^2 + \dot{h}_\times^2 \right\rangle_{\text{oct+cq}} \\ &= \frac{62}{7} \frac{G\mu^2}{c^7} \left(\frac{\delta m}{m} \right)^2 R^6 \omega_s^8 \end{aligned} \quad (3.349)$$

It is interesting to compare the power at $\omega = \omega_s, 3\omega_s$ (generated by the mass octupole plus the current quadrupole), with the power at $\omega = 2\omega_s$, generated by the mass quadrupole. From eq. (3.348) we find

$$P(\omega_s) = \frac{25}{896} \left(\frac{v}{c} \right)^2 \left(\frac{\delta m}{m} \right)^2 P(2\omega_s) \quad (3.350)$$

and

$$P(3\omega_s) = \frac{1215}{896} \left(\frac{v}{c} \right)^2 \left(\frac{\delta m}{m} \right)^2 P(2\omega_s) \quad (3.351)$$

where $P(2\omega_s)$ is the leading-order quadrupole result, eq. (3.339). In Fig. 3.8 we show the relative intensity of the three spectral lines at $\omega = \omega_s, 2\omega_s$ and $3\omega_s$, for $v/c = 10^{-2}$ and an extreme mass ratio $m_2/m_1 \rightarrow 0$ and therefore $\delta m/m = 1$. Observe that the vertical scale is logarithmic.

Problem 3.4. Decomposition of $\dot{S}^{kl,m}$ into irreducible representations of $SO(3)$

We have seen that the next-to-leading term in the multipole expansion is proportional to $\dot{S}^{kl,m}$. In this problem we discuss the decomposition of $\dot{S}^{kl,m}$ into irreducible representations of the rotation group $SO(3)$, and we will understand the group-theoretical origin of the mass octupole and current quadrupole terms.

Let us recall that the irreducible representations of Lie groups, as for instance $O(N)$ or $U(N)$, are conveniently expressed in terms of Young diagrams. A Young diagram is a set of n boxes, organized into r lines of length $n_1, n_2, \dots, n_r$, with $n_1 \geq n_2 \geq \dots \geq n_r$ and $n_1 + \dots + n_r = n$. In each box we put an index $i_1, \dots, i_n$. For $O(N)$ each index takes the values $1, \dots, N$, which is the dimension of the vector representation. The irreducible tensor representations can be obtained antisymmetrizing first over the indices in the columns and then symmetrizing over the indices in the lines.⁵⁸ For $O(N)$ we must also remove the traces on all pairs of symmetric indices. For example, a generic tensor with three indices k, l, m is decomposed into irreducible representations of $O(N)$ as follows. First of all, we remove from the tensor all the traces. Let us call T^{klm} the resulting traceless tensor, so that $T^{kmm} = T^{mlm} = T^{kkm} = 0$ (repeated indices are summed over). Then the decomposition of T^{klm} is shown in Fig. 3.9. The Young diagram (a) represents the tensor obtained symmetrizing T^{klm} over all indices; the diagram (b) represents the tensor obtained antisymmetrizing first over (k, m) , which gives $T^{klm} - T^{mlk}$, and then symmetrizing over the pair (k, l) . This gives

$$\text{diagram(b)} : \quad T^{klm} + T^{lkm} - T^{mlk} - T^{mkl} \quad (3.352)$$

The tensor corresponding to the diagram (c) is obtained similarly, antisymmetrizing first over (k, l) and then symmetrizing over (k, m) , and gives

$$\text{diagram}(c) : T^{klm} + T^{mlk} - T^{lkm} - T^{lmk} \quad (3.353)$$

Finally, the Young diagram (d) represents the tensor obtained antisymmetrizing T^{klm} over all indices.⁵⁹ Counting the independent components we see that a traceless, but otherwise generic, tensor with three indices k, l, m , each taking the values $1, 2, 3$, has 18 components, and that the diagram (a) represents a tensor with seven components, (b) and (c) with five components each, and (d) has one component, so we get $18 = 7 + 5 + 5 + 1$. Recall that the representations of the rotation group can also be labeled by the spin s , and the representation labeled by s has dimension $2s + 1$. Then, the representation with dimension seven, i.e. the diagram (a), corresponds to $s = 3$, while the representations with dimension five corresponds to $s = 2$. Denoting by $\mathbf{s}$ the representation with spin s , and by $\oplus$ the direct sum of representations, the decomposition of Fig. 3.9 reads

$$T^{klm} \in \mathbf{3} \oplus \mathbf{2} \oplus \mathbf{2} \oplus \mathbf{0} \quad (3.354)$$

We can now understand the decomposition given in eq. (3.139), in terms of Fig. 3.9. Our starting point is the tensor $\dot{S}^{kl,m}$. This is symmetric in (k, l) , and has no special symmetry with respect to the other indices. Therefore, it is not an irreducible representation of $SO(N)$.

First of all observe that, in eq. (3.138), we can replace $\dot{S}^{kl,m}$ by the tensor in which all the traces have been removed. In fact, subtracting from $\dot{S}^{kl,m}$ a term⁵⁸ It is more common to first symmetrize over the lines and then antisymmetrize over the columns (see, e.g. Hamermesh (1962), Section 10.6). The two procedures are however equivalent and for our purposes the former is more convenient.

⁵⁹ If, rather than $O(N)$, we consider $SO(N)$, we must also take into account that there is a globally defined antisymmetric tensor $\epsilon^{i_1 \dots i_N}$. In this case the representations obtained one from the other contracting antisymmetric indices with the ϵ tensor are equivalent. In particular, for $SO(3)$ we have the tensor ϵ^{ijk} . So, for instance, an antisymmetric tensor with two indices A^{kl} is equivalent to a vector $A_i = \epsilon_{ikl} A^{kl}$ while a totally antisymmetric tensor A^{ijk} is equivalent to a scalar A through the relation $A = \epsilon_{ijk} A^{ijk}$. Since the indices take the values $i = 1, \dots, N$, we cannot antisymmetrize over more than N indices, and there are no diagrams with more than N lines, for $O(N)$. For $SO(N)$ a column with N box is equivalent to a scalar, as we see contracting with the totally antisymmetric tensor, and can be eliminated. Therefore a Young diagram of $SO(N)$ can always be reduced to an equivalent diagram with no more than $N - 1$ rows.

$$T^{klm} = (\boxed{k} \times \boxed{l}) \times \boxed{m} = (\boxed{k \ l} + \boxed{\begin{smallmatrix} k \\ l \end{smallmatrix}}) \times \boxed{m} =$$

$$= \boxed{k \ l \ m} + \boxed{\begin{smallmatrix} k & 1 \\ m & \end{smallmatrix}} + \boxed{\begin{smallmatrix} k & m \\ 1 & \end{smallmatrix}} + \boxed{\begin{smallmatrix} k \\ 1 \\ m \end{smallmatrix}}$$

(a)

(b)

(c)

(d)

Fig. 3.9 The Young diagrams corresponding to irreducible tensor representations with three indices k, l, m .

proportional to δ_{kl} we get in eq. (3.138) a factor proportional to $\Lambda_{ij,kl} n_m \delta_{kl}$, which vanishes because $\Lambda_{ij,kk} = 0$, while subtracting from $\dot{S}^{kl,m}$ a term proportional to δ_{km} or to δ_{lm} we get a factor proportional to $\Lambda_{ij,kl} n_l = 0$. We can therefore consider $\dot{S}^{kl,m}$ as a tensor from which all

traces have already been removed (this is analogous to the fact that, in the quadrupole term, we could substitute M^{ij} with Q^{ij}).

Figure 3.9 gives the decomposition in irreducible representations of a generic traceless tensor T^{klm} . However, $\dot{S}^{kl,m}$ is not generic, but is symmetric in (k, l) (and of course remains symmetric also after we remove all its traces). Then, when we decompose $\dot{S}^{kl,m}$ in irreducible representations as in Fig. 3.9, the diagrams (c) and (d) do not contribute, since they are obtained antisymmetrizing first over (k, l) which, applied to $\dot{S}^{kl,m}$, gives zero. Therefore, in the decomposition of $\dot{S}^{kl,m}$, only two irreducible representations appear:

- diagram (a), which is the totally symmetric and traceless combination, and therefore has the symmetries of the mass octupole;
- diagrams (b), which has the structure given in eq. (3.352). For a tensor T^{klm} symmetric in (k, l) , the structure of indices given in eq. (3.352) simplifies to

$$2T^{klm} - T^{mlk} - T^{mkl} \quad (3.355)$$

Identifying T^{klm} with $\ddot{P}^{m,kl}$, (which is indeed symmetric in (k, l)), we see that this is precisely the structure of indices of the current quadrupole term in eq. (3.139).

We therefore understand that the algebraic identity (3.139) expresses the decomposition of $\dot{S}^{kl,m}$ in irreducible representations of $SO(3)$. The mass octupole corresponds to the Young diagram (a) in Fig. 3.9, and is a spin-3 representation, while the current quadrupole corresponds to the Young diagram (b) in Fig. 3.9, and is a spin-2 representation. The representations corresponding to the Young diagrams (c) and (d) instead do not appear, because $\dot{S}^{kl,m}$ is symmetric in (k, l) .

The fact that $\ddot{P}^{k,lm} + \ddot{P}^{l,km} - 2\ddot{P}^{m,kl}$ is a spin-2 tensor is the origin of its name "current quadrupole": the term "quadrupole" refers to the spin-2 nature of the operator, and the term "current" refers to the fact that it is obtained from the momenta of the momentum P^i .

Problem 3.5. Computation of $\int d\Omega \left(\mathbf{T}_{lm}^{E2,B2} \right)_{ij}^* n_{i_1} \cdots n_{i_\alpha}$

In Section 3.5 we have seen that the coefficients of the multipole expansion, u_{lm} and v_{lm} , are given in terms of integrals over the solid angle of the quantity $(\mathbf{T}_{lm}^{E2})_{ij}^* n_{i_1} \cdots n_{i_\alpha}$, for u_{lm} , and $(\mathbf{T}_{lm}^{B2})_{ij}^* n_{i_1} \cdots n_{i_\alpha}$, for v_{lm} , see eqs. (3.283) and (3.284). In this problem we compute explicitly these integrals. As a first step, we consider

$$\int d\Omega Y_{lm}^* n_{i_1} \cdots n_{i_\alpha} \quad (3.356)$$

with l, α arbitrary integers. Using eq. (3.234), we can write it as

$$(\mathcal{Y}_{j_1, \dots, j_l}^{lm})^* \int d\Omega n_{j_1} \cdots n_{j_l} n_{i_1} \cdots n_{i_\alpha} \quad (3.357)$$

We can now perform this integral using eq. (3.23). If $\alpha < l$, in the product of Kronecker delta of eq. (3.23) there are not enough indices of the type $i_1, \dots, i_\alpha$ to be contracted with the indices of the type $j_1, \dots, j_l$, and therefore necessarily we have at least one Kronecker delta involving two indices of the group $j_1, \dots, j_l$, e.g. terms containing $\delta_{j_1 j_2}$. Then, since $\mathcal{Y}_{j_1, \dots, j_l}^{lm}$ is traceless, the result is zero. Therefore the integral in eq. (3.356) can be non-vanishing only if $\alpha \geq l$. If $\alpha = l$, we found below eq. (3.237) that

$$\int d\Omega Y_{lm}^* n_{i_1} \cdots n_{i_l} = 4\pi \frac{l!}{(2l+1)!!} (\mathcal{Y}_{i_1 \dots i_l}^{lm})^* \quad (3.358)$$

Now, observe that u_{lm} is expressed as a sum over α of terms containing the α -th time derivatives of $S^{ij,i_1\dots i_\alpha}$, see eq. (3.283). In order of magnitude, if d is the size of the source,

$$\begin{aligned}\partial_0^{\alpha+1} S^{ij,i_1\dots i_\alpha i_{\alpha+1}} &\sim \frac{\omega_s d}{c} \partial_0^\alpha S^{ij,i_1\dots i_\alpha} \\ &= \frac{v}{c} \partial_0^\alpha S^{ij,i_1\dots i_\alpha}\end{aligned}\quad (3.359)$$

since $\partial_0 = (1/c)\partial/\partial t$ and each time derivative brings down a factor ω_s , where ω_s is the typical frequency of the source, while the addition of the index $i_{\alpha+1}$ corresponds to the insertion of a factor $x^{i_{\alpha+1}}$ inside the integral in d^3x over the source volume, which therefore gives a contribution $O(d)$. Therefore, in the limit $v/c \ll 1$ in which the multipole expansion is useful, the dominant term in eqs. (3.283) and (3.284) is the one with the smallest value of α for which the integral is non-vanishing. Recall from eqs. (3.265) and (3.267) that

$$\mathbf{T}_{lm}^{E2} = a_{31} \mathbf{T}_{lm}^{l+2} + a_{32} \mathbf{T}_{lm}^l + a_{33} \mathbf{T}_{lm}^{l-2} \quad (3.360)$$

$$\mathbf{T}_{lm}^{B2} = b_{21} i \mathbf{T}_{lm}^{l+1} + b_{22} i \mathbf{T}_{lm}^{l-1} \quad (3.361)$$

where the coefficients are given in Table 3.1. Since $\mathbf{T}_{jjz}^l$ is proportional to Y_{llz} , see eq. (3.262), the integral $\int d\Omega (\mathbf{T}_{lm}^{l-2})_{ij}^* n_{i_1} \dots n_{i_\alpha}$ is proportional to $\int d\Omega Y_{l-2,m}^* n_{i_1} \dots n_{i_\alpha}$ and therefore the lowest value of α for which it is nonvanishing is $\alpha = l - 2$, while for the integral of $\mathbf{T}_{lm}^l$ we need at least $\alpha = l$ and for the integral of $\mathbf{T}_{lm}^{l+2}$ we need at least $\alpha = l + 2$. We conclude that in eq. (3.283) the smallest possible value of α which gives a non-vanishing contribution is obtained from $\mathbf{T}_{lm}^{l-2}$ and is $\alpha_{\min} = l - 2$. Similarly, for v_{lm} the leading contribution is obtained from $\mathbf{T}_{lm}^{l-1}$ and is $\alpha_{\min} = l - 1$. Therefore

$$\begin{aligned}u_{lm} &\simeq a_{33} \frac{4}{(l-2)!} (\partial_0^{l-2} S^{ij,i_1\dots i_{l-2}}) \int d\Omega (\mathbf{T}_{lm}^{l-2})_{ij}^* n_{i_1} \dots n_{i_{l-2}} \\ v_{lm} &\simeq -ib_{22} \frac{4}{(l-1)!} (\partial_0^{l-1} S^{ij,i_1\dots i_{l-1}}) \int d\Omega (\mathbf{T}_{lm}^{l-1})_{ij}^* n_{i_1} \dots n_{i_{l-1}}\end{aligned}\quad (3.362)$$

times a factor $[1 + O(v^2/c^2)]$. Now we use the identities

$$(\mathbf{T}_{lm}^{l-2})_{ij} = \left[\frac{l(l-1)}{(2l-1)(2l+1)} \right]^{1/2} \mathcal{Y}_{ij i_1 \dots i_{l-2}}^{lm} n_{i_1} \dots n_{i_{l-2}} \quad (3.363)$$

$$(\mathbf{T}_{lm}^{l-1})_{ij} = i \left[\frac{2l(l-1)}{(l+1)(2l+1)} \right]^{1/2} \epsilon_{pq(i} \mathcal{Y}_{j) q i_1 \dots i_{l-2}}^{lm} n_p n_{i_1} \dots n_{i_{l-2}} \quad (3.364)$$

where the parentheses on the indices, in eq. (3.364), denotes the symmetrization over the indices i, j , i.e. $A_{(ij)} \equiv (1/2)(A_{ij} + A_{ji})$. These identities can be obtained (with quite some work) inserting eq. (3.234) into the definition of the spin-2 tensor harmonics. Inserting the explicit values of a_{33}, b_{22} from Table 3.1 (and recalling the change of notation $j \rightarrow l$ that we made in between, see Note 47), we get

$$\begin{aligned}
u_{lm} &= \left[\frac{(l+1)(l+2)}{2(2l-1)(2l+1)} \right]^{1/2} \frac{4}{(l-2)!} \left[\frac{l(l-1)}{(2l-1)(2l+1)} \right]^{1/2} \\
&\quad \times (\partial_0^{l-2} S^{ij,i_1 \dots i_{l-2}}) \mathcal{Y}_{ij,i_1 \dots i_{l-2}}^{lm*} \int d\Omega n_{i_1} \dots n_{i_{l-2}} n_{j_1} \dots n_{j_{l-2}} \\
&= \left[\frac{1}{2} l(l-1)(l+1)(l+2) \right]^{1/2} \frac{4}{(l-2)!} \frac{1}{(2l-1)(2l+1)} \\
&\quad \times \frac{4\pi}{(2l-3)!!} (l-2)! (\partial_0^{l-2} S^{ij,i_1 \dots i_{l-2}}) \mathcal{Y}_{ij,i_1 \dots i_{l-2}}^{lm*} \\
&= \frac{16\pi}{(2l+1)!!} \left[\frac{1}{2} l(l-1)(l+1)(l+2) \right]^{1/2} (\partial_0^{l-2} S^{ij,i_1 \dots i_{l-2}}) \mathcal{Y}_{ij,i_1 \dots i_{l-2}}^{lm*} \quad (3.365)
\end{aligned}$$

where the final integral as been performed using eq. (3.23). The integral for v_{lm} is performed similarly using eqs. (3.363) and (3.364), with the final result given in the text.

6.0.77 Further reading

- The quadrupole radiation is discussed in all general relativity textbooks, see in particular Weinberg (1972), Misner, Thorne and Wheeler (1973), Landau and Lifshitz, Vol. II (1979) and Straumann (2004).
- The radiation from sources with arbitrary velocity is discussed in Weinberg (1972), Section 10.4. Gravitational wave generation is also discussed in detail in the reviews Thorne (1983) and (1987).
- Radiation reaction for slow-motion sources is discussed in Misner, Thorne and Wheeler (1973), Sections 36.8 and 36.11.
- The multipole expansion for time-dependent fields in terms of STF tensors was introduced by Sachs (1961) and Pirani (1964). Thorne (1980) derived the slow-motion expansion of the mass and spin multipole moments, both in STF and in spherical tensor form. The closed-form expression for these moments in STF form is derived in Damour and Iyer (1991a). A detailed review of the multipole expansion for GWs, as well as a historical overview of the relevant literature, is Thorne (1980). A physical discussion of current quadrupole radiation is given in Schutz and Ricci (2001).

6.0.78 Applications

In this chapter we apply the formalism that we have developed to various instructive problems. The systems that we examine here are still somewhat idealized, compared to real astrophysical sources. This allows us to understand the essence of the physical mechanisms with a minimum of complications, and forms the basis for a more detailed study of realistic sources, which will be the subject of Vol. 2.

We begin, in Section 4.1, with the study of binary systems, taking the bodies as point-like and moving at first on a Newtonian trajectory. We will compute how the back-reaction of GWs affects the motion of the sources, inducing the inspiral and coalescence of the binary system, and we will see how this, in turn, affects the emission of the GWs themselves.

In Section 4.2 we will compute the radiation emitted by spinning rigid bodies, which are a first idealization of rotating neutron stars.

In Section 4.3 we compute the radiation emitted by a body falling radially into a black hole. A full resolution of this problem requires expansion over the Schwarzschild metric, rather than over flat spacetime of linearized theory, and will be deferred to Vol. 2. However, we will see that the low-frequency part of the spectrum can be computed using a flat space-time background, so we can perform here this part of the computation. We will also compare the situation in which the infalling particle is point-like with that of a real star, which can be disrupted by the tidal gravitational force of the black hole. This is particularly instructive because it allows us to compare the coherent and the incoherent emission of GWs.

In Section 4.4 we study the radiation emitted by a mass accelerated by an external force. It will be interesting to compare the results with the electromagnetic radiation from an accelerated charge. We will see that, while the electromagnetic field of a relativistic charge is beamed into a small angle in the forward direction, this does not happen in the gravitational case. Finally, some computational details are collected in a Solved Problems section, at the end of the chapter.

6.0.79 4.1 Inspiral of compact binaries

In this section we consider a binary system made of two compact stars, such as neutron stars or black holes, and we treat them as point-like, with masses m_1, m_2 , and positions $\mathbf{r}_1$ and $\mathbf{r}_2$. In the Newtonian approximation, and in the center-of-mass frame (CM), the dynamics reduces to a one-body problem with mass equal to the reduced mass



4.1 Inspiral of compact binaries

167

4.2 Radiation from rotating rigid bodies

4.3 Infall into a black hole 215

4.4 Radiation from accelerated masses

4.5 Solved problems

$\mu = m_1 m_2 / (m_1 + m_2)$, and equation of motion $\ddot{\mathbf{r}} = -(Gm/r^3)\mathbf{r}$, where $m = m_1 + m_2$ is the total mass and $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ the relative coordinate. We consider first the case of circular orbits. Then the orbital frequency ω_s is related to the orbital radius R by $v^2/R = Gm/R^2$ with $v = \omega_s R$, so we have Kepler's law

$$\omega_s^2 = \frac{Gm}{R^3} \quad (4.1)$$

We already studied this system in Problem 3.2, and the corresponding gravitational wave amplitudes are given in eqs. (3.330) and (3.331). In these expressions, we eliminate R in favor of ω_s using eq. (4.1), and we introduce the chirp mass

$$M_c = \mu^{3/5} m^{2/5} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} \quad (4.2)$$

Then eqs. (3.330) and (3.331) become

$$h_+(t) = \frac{4}{r} \left(\frac{GM_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}}{c} \right)^{2/3} \frac{1 + \cos^2 \theta}{2} \cos(2\pi f_{\text{gw}} t_{\text{ret}} + 2\phi)$$

$$h_\times(t) = \frac{4}{r} \left(\frac{GM_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}}{c} \right)^{2/3} \cos \theta \sin(2\pi f_{\text{gw}} t_{\text{ret}} + 2\phi)$$

where we expressed the result in terms of $f_{\text{gw}} = \omega_{\text{gw}}/(2\pi)$, with $\omega_{\text{gw}} = 2\omega_s$. Observe that, in this lowest-order Newtonian approximation, the amplitudes h_+ and $h_\times$ of the GWs emitted depend on the masses m_1, m_2 only through the combination M_c .

The factors in eq. (4.3) are possibly more expressive if we write them in terms of ratios of quantities with dimension of length, introducing the Schwarzschild radius associated to the chirp mass,

$$R_c \equiv \frac{2GM_c}{c^2} \quad (4.4)$$

and using the reduced wavelength of the GW, $\lambda = c/\omega_{\text{gw}}$. Then, eq. (4.3) reads

$$h_+(t) = \mathcal{A} \frac{1 + \cos^2 \theta}{2} \cos(\omega_{\text{gw}} t_{\text{ret}} + 2\phi)$$

$$h_\times(t) = \mathcal{A} \cos \theta \sin(\omega_{\text{gw}} t_{\text{ret}} + 2\phi) \quad (4.6)$$

where

$$\mathcal{A} = \frac{1}{2^{1/3}} \left(\frac{R_c}{r} \right) \left(\frac{R_c}{\star} \right)^{2/3}$$

From eq. (3.337) we can write the power radiated in GWs, per unit solid angle, as

$$\frac{dP}{d\Omega} = \frac{2}{\pi} \frac{c^5}{G} \left(\frac{GM_c \omega_{\text{gw}}}{2c^3} \right)^{10/3} g(\theta) \quad (4.8)$$

where

$$g(\theta) = \left(\frac{1 + \cos^2 \theta}{2} \right)^2 + \cos^2 \theta \quad (4.9)$$

Observe that $\langle \cos^2(2\omega t + 2\phi) \rangle = 1/2$ is independent of ϕ , so the angular distribution of the radiated power, which is proportional to $\langle \dot{h}_+^2 + \dot{h}_\times^2 \rangle$, is also independent of ϕ . The angular average of $g(\theta)$ is

$$\int \frac{d\Omega}{4\pi} g(\theta) = \frac{4}{5} \quad (4.10)$$

where, as usual, $d\Omega = d\cos\theta d\phi$. The total radiated power P is therefore ¹

$$P = \frac{32}{5} \frac{c^5}{G} \left(\frac{GM_c \omega_{\text{gw}}}{2c^3} \right)^{10/3} \quad (4.12)$$

6.0.80 4.1.1 Circular orbits. The chirp amplitude

Equation (4.3) gives the amplitude of the GWs, assuming that the motion of the sources is on a given, fixed, circular Keplerian orbit. However, the emission of GWs costs energy. The source for the radiated energy is the sum of kinetic plus potential energy of the orbit, which is

$$\begin{aligned} E_{\text{orbit}} &= E_{\text{kin}} + E_{\text{pot}} \\ &= -\frac{Gm_1m_2}{2R} \end{aligned} \quad (4.13)$$

and therefore, to compensate for the loss of energy to GWs, R must decrease in time, so that E_{orbit} becomes more and more negative.²

According to eq. (4.1), if R decreases, ω_s increases. On the other hand, if ω_s increases, also the power radiated in GWs increases, as we see from eq. (4.12). Then R must decrease further, and we have a runaway process which, on a sufficiently long time-scale, leads to the coalescence of the binary system. As long as

$$\dot{\omega}_s \ll \omega_s^2 \quad (4.14)$$

we are in the regime called of quasi-circular motion. In fact, using eq. (4.1) we see that the radial velocity $\dot{R}$ can be written as

$$\begin{aligned} \dot{R} &= -\frac{2}{3}R\frac{\dot{\omega}_s}{\omega_s} \\ &= -\frac{2}{3}(\omega_s R)\frac{\dot{\omega}_s}{\omega_s^2} \end{aligned} \quad (4.15)$$

Then, as long as the condition $\dot{\omega}_s \ll \omega_s^2$ is fulfilled, $|\dot{R}|$ is much smaller than the tangential velocity $\omega_s R$, and the approximation of a circular orbit with a slowly varying radius is applicable. In the following, we compute the back-reaction of GWs when eq. (4.14) applies.

Using eq. (4.1), we eliminate R in favor of ω_{gw} in eq. (4.13), and we get

$$E_{\text{orbit}} = -\left(G^2 M_c^5 \omega_{\text{gw}}^2 / 32\right)^{1/3} \quad (4.16)$$

$$P = \frac{32}{5} \frac{c^5}{G} \nu^2 x^5 \quad (4.11)$$

The corrections to this result in the post-Newtonian expansion are given in eq. (5.257).³ If we denote by $(t_{\text{coal}})_{\text{ret}}$ the value of retarded time t_{ret} at coalescence, and by t_{coal} the value of the

¹ In Section 5.6, where we apply the post-Newtonian formalism to inspiraling binaries, we will see that, rather than using the source frequency ω_s , it is convenient to use the variable $x = (Gm\omega_s/c^3)^{2/3}$, which is dimensionless and of order v^2/c^2 , where v is the typical orbital velocity. We also introduce the "symmetric mass ratio" $\nu = \mu/m$, so $M_c = \nu^{3/5}m$. Since here we are computing in the quadrupole approximation and we are considering a circular orbit, we have $\omega_{\text{gw}} = 2\omega_s$, and the result (4.12) can be written as

² In our idealized setting of point-like particles, the two masses have no internal structure, so no degrees of freedom that can relax supplying, at least partially, the required energy, and the only possible source of energy is the orbital energy of the system. Even in a realistic system of two stars, however, at least in the early phase of the coalescence, the orbital frequency is much smaller than the frequencies of the normal modes of the star, and therefore the internal dynamics of the stars is decoupled from the orbital motion, and all the energy supplied to GWs again comes from the orbital energy of the system. We will see in Section 5.5 that, for compact objects, corrections that depend on the internal structure of the bodies enter only at the very high order 5PN in the postNewtonian expansion, i.e. they are corrections of order $(v/c)^{10}$ to the equations of motion.

observer time t at coalescence, we have $\tau = (t_{\text{coal}})_{\text{ret}} - t_{\text{ret}} = t_{\text{coal}} - t$, since retarded time and the observer time differ only by a constant quantity, the time taken by the GW to propagate, at the speed of light, from the source (whose center-of-mass, to an excellent approximation, can be taken to be at a fixed position during the observation time) to the observer. Therefore, even if what we observe today on Earth depends on what happened millions of years ago in a far galaxy, when a quantity like f_{gw} depends on the difference τ we can express it simply in terms of the observer time t .

Observe that, at fixed ω_{gw} , the dependence on the masses is again only through the chirp mass. Equating P in eq. (4.12) to $-dE_{\text{orbit}}/dt$ we find

$$\dot{\omega}_{\text{gw}} = \frac{12}{5} 2^{1/3} \left(\frac{GM_c}{c^3} \right)^{5/3} \omega_{\text{gw}}^{11/3} \quad (4.17)$$

or, in terms of $f_{\text{gw}} = \omega_{\text{gw}}/(2\pi)$

$$\dot{f}_{\text{gw}} = \frac{96}{5} \pi^{8/3} \left(\frac{GM_c}{c^3} \right)^{5/3} f_{\text{gw}}^{11/3} \quad (4.18)$$

Integrating eq. (4.18) we see that f_{gw} formally diverges at a finite value of time, that we denote t_{coal} . In terms of $\tau \equiv t_{\text{coal}} - t$ (the time to coalescence), the solution of eq. (4.18) reads³

$$f_{\text{gw}}(\tau) = \frac{1}{\pi} \left(\frac{5}{256} \frac{1}{\tau} \right)^{3/8} \left(\frac{GM_c}{c^3} \right)^{-5/8} \quad (4.19)$$

The divergence is cut off by the fact that, when their separation becomes smaller than a critical distance, the two stars merge, as we discuss in more detail below, see eq. (4.38). Inserting in eq. (4.19) the numerical values we find

$$f_{\text{gw}}(\tau) \simeq 134 \text{ Hz} \left(\frac{1.21 M_{\odot}}{M_c} \right)^{5/8} \left(\frac{1 \text{ s}}{\tau} \right)^{3/8} \quad (4.20)$$

where, as reference value for M_c , we have taken $1.21 M_{\odot}$, which is the chirp mass of a system of two stars, each one with a mass of $1.4 M_{\odot}$. Equivalently, we can write eq. (4.20) as

$$\tau \simeq 2.18 \text{ s} \left(\frac{1.21 M_{\odot}}{M_c} \right)^{5/3} \left(\frac{100 \text{ Hz}}{f_{\text{gw}}} \right)^{8/3} \quad (4.21)$$

From this we find that (when $M_c = 1.21 M_{\odot}$) at 10 Hz (which is of the order of the lowest frequencies accessible to ground-based interferometers) we get the radiation emitted at about $\tau = 17$ min to coalescence; at 100 Hz we get the radiation from the last two seconds, and at 1 kHz we get the radiation from the last few milliseconds. From Kepler's law (4.1) we find that, when $f_{\text{gw}} = 1 \text{ kHz}$, the separation between two bodies with $m_1 = m_2 = 1.4 M_{\odot}$ is $R \simeq 33 \text{ km}$. Such a small separation can be reached only by very compact bodies like neutron stars and black holes. Since the radius of a neutron star with $m = 1.4 M_{\odot}$ is about 10 km, for neutron stars and black holes the point-like approximation at this stage becomes inaccurate, but still not meaningless.

A useful quantity for assessing the sensitivity of detectors to inspiraling binaries is the number of cycles spent in the detector bandwidth $f \in [f_{\text{min}}, f_{\text{max}}]$. When the period $T(t)$ of the GW is a slowly varying function of time, the number of cycles in a time interval dt is given by $d\mathcal{N}_{\text{cyc}} = dt/T(t) = f_{\text{gw}}(t)dt$, so

$$\begin{aligned}
\mathcal{N}_{\text{cyc}} &= \int_{t_{\min}}^{t_{\max}} f_{\text{gw}}(t) dt \\
&= \int_{f_{\min}}^{f_{\max}} df_{\text{gw}} \frac{f_{\text{gw}}}{\dot{f}_{\text{gw}}}
\end{aligned} \tag{4.22}$$

For the inspiral of a binary system we can express $\dot{f}_{\text{gw}}$ as a function of f_{gw} using eq. (4.18), and we get

$$\begin{aligned}
\mathcal{N}_{\text{cyc}} &= \frac{1}{32\pi^{8/3}} \left(\frac{GM_c}{c^3} \right)^{-5/3} \left(f_{\min}^{-5/3} - f_{\max}^{-5/3} \right) \\
&\simeq 1.6 \times 10^4 \left(\frac{10 \text{ Hz}}{f_{\min}} \right)^{5/3} \left(\frac{1.2 M_{\odot}}{M_c} \right)^{5/3}
\end{aligned} \tag{4.23}$$

where in the second line we assumed $f_{\min}^{-5/3} - f_{\max}^{-5/3} \simeq f_{\min}^{-5/3}$, as is typically the case. This means that ground-based interferometers can follow the evolution of the signal for thousands of cycles, and a space-borne interferometer sensitive to the mHz region can follow it for millions of cycles.⁴ For this reason, accurate predictions of the waveform, going well beyond the Newtonian approximation used here, are very important, and will be discussed in Section 5.6.

As the frequency increases, the orbital radius shrinks. From eqs. (4.1) and (4.19) (recalling that $\tau = t_{\text{coal}} - t$, so that $d/d\tau = -d/dt$, and using a dot denotes d/dt)

$$\begin{aligned}
\frac{\dot{R}}{R} &= -\frac{2}{3} \frac{\dot{\omega}_{\text{gw}}}{\omega_{\text{gw}}} \\
&= -\frac{1}{4\tau}
\end{aligned} \tag{4.24}$$

which integrates to

$$\begin{aligned}
R(\tau) &= R_0 \left(\frac{\tau}{\tau_0} \right)^{1/4} \\
&= R_0 \left(\frac{t_{\text{coal}} - t}{t_{\text{coal}} - t_0} \right)^{1/4}
\end{aligned} \tag{4.25}$$

where R_0 is the value of R at the initial time t_0 , and $\tau_0 = t_{\text{coal}} - t_0$. The function $R(t)$ is shown in Fig. 4.1. We see that there is a long phase where R decreases smoothly, followed by a plunge phase, where our approximation of quasi-circular orbit is no longer valid.⁵

Inserting eq. (4.19), evaluated at an initial time t_0 when $\tau = \tau_0$, into eq. (4.1), we find the relation between the initial radius R_0 and the time to coalescence τ_0

$$\tau_0 = \frac{5}{256} \frac{c^5 R_0^4}{G^3 m^2 \mu} \tag{4.26}$$

⁴ More precisely, a space interferometer such as LISA has $f_{\min} \sim 10^{-4}$ Hz, and signals in this frequency band are produced by the inspiral into supermassive BHs with a mass $O(10^6) M_{\odot}$. In a coalescence of two supermassive BHs with $m_1 = m_2 = 10^6 M_{\odot}$, we have $M_c \simeq 0.9 \times 10^6 M_{\odot}$ and $\mathcal{N}_c$ is only of order 600. However, in the infall of a stellar mass black hole into a supermassive black hole, taking $m_1 = 10^6 M_{\odot}$ and $m_2 = 10 M_{\odot}$, we have $M_c \simeq 10^3 M_{\odot}$ and $\mathcal{N}_c \simeq 5 \times 10^7$.

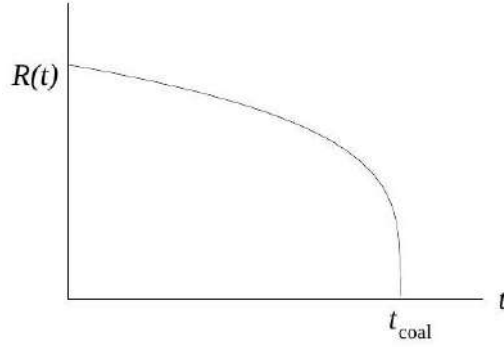


Fig. 4.1 The evolution of the separation $R(t)$ between the two bodies, in the lowest-order Newtonian approximation.

Expressing the initial radius R_0 in terms of the initial orbital period $T_0 = 2\pi/\omega_s(\tau_0)$ through Kepler's law, $R_0^2 = Gm(T_0/2\pi)^2$, and plugging in the numerical values, we find

$$\tau_0 \simeq 9.829 \times 10^6 \text{yr} \left(\frac{T_0}{1\text{hr}} \right)^{8/3} \left(\frac{M_\odot}{m} \right)^{2/3} \left(\frac{M_\odot}{\mu} \right) \quad (4.27)$$

Therefore, under our assumption of circular orbits and for masses of order $M_\odot$, only binaries which at formation had an initial orbital period of less than about one day can have coalesced by emission of GWs, within the present age of galaxies. We will discuss in Section 4.1.2 how this result is modified for elliptic orbits.

Until now we have studied how the GW frequency evolves in time, in a binary system. We next examine how the corresponding waveform of the GW changes. A particle that moves on a quasi-circular orbit in the (x, y) plane with a radius $R = R(t)$ and an angular velocity $\omega_s = \omega_s(t)$ has Cartesian coordinates $x(t) = R(t) \cos(\Phi(t)/2)$ and $y(t) = R(t) \sin(\Phi(t)/2)$, where we have defined

$$\begin{aligned} \Phi(t) &= 2 \int_{t_0}^t dt' \omega_s(t') \\ &= \int_{t_0}^t dt' \omega_{\text{gw}}(t') \end{aligned} \quad (4.28)$$

When we compute the GW production in the quadrupole approximation, in the computation of the second time derivative of the quadrupole moment there are therefore three differences, compared to the case when ω_s and R constants, i.e. to the computation which gives eq. (4.3):

- in the argument of the trigonometric functions, $\omega_{\text{gw}}t$ must be replaced by $\Phi(t)$
- in the factors in front of the trigonometric functions, ω_{gw} is replaced by $\dot{\Phi}(t)$

⁵ Furthermore, in compact binaries made of black holes or neutron stars, the flat-space approximation is totally inadequate in the plunge phase, see the discussion around eq. (4.38) below. We will discuss the plunge phase for systems with BH or NS in Vol. 2.

⁶ As discussed in Problem 3.2, we are interested in the radiation that a binary star emits in the direction which points from the star toward us. The angle θ is therefore equal to the angle ι between the normal to the orbit and the line-of-sight, while we can orient the axes so that $\phi = 0$ or, equivalently, we can reabsorb the fixed value of ϕ into a shift of the origin of time.

- we should also include in the computation the contributions coming from the derivatives of $R(t)$ and of $\omega_{\text{gw}}(t)$.

However, as we have seen above, the radial velocity $\dot{R}$ is negligible as long as $\dot{\omega}_s \ll \omega_s^2$. Using eq. (4.17), the condition $\dot{\omega}_s \ll \omega_s^2$ translates into $GM_c \omega_s / c^3 \ll 0.5$. In terms of $f_{\text{gw}} = \omega_{\text{gw}} / (2\pi)$, this means that $\dot{R}$ is negligible as long as $f_{\text{gw}} \ll 13\text{kHz}$ ($1.2M_\odot / M_c$). As we will see in eq. (4.40) below, the transition to the plunge phase takes place earlier, thus as long as we assume that we are in the inspiral phase we can simply neglect the terms proportional to $\dot{R}$ in the computation of the waveform, at least to lowest order, and similarly for the terms involving the derivative of $\omega_{\text{gw}}(t)$. The only changes, therefore, are the replacement of $\omega_{\text{gw}}t$ by $\Phi(t)$ in the argument of the sine and cosine, and of ω_s by $\omega_s(t)$ in the prefactor, all evaluated at the retarded time t_{ret} .⁶ Then,

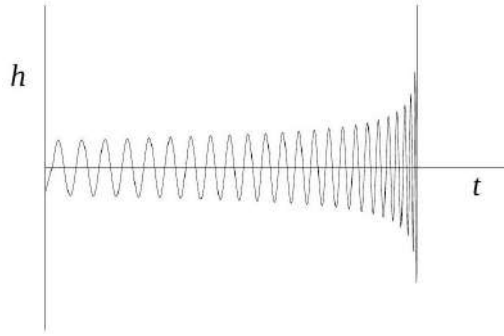


Fig. 4.2 The time evolution of the GW amplitude in the inspiral phase of a binary system.

$$\begin{aligned} h_+(t) &= \frac{4}{r} \left(\frac{GM_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}(t_{\text{ret}})}{c} \right)^{2/3} \left(\frac{1 + \cos^2 \iota}{2} \right) \cos[\Phi(t_{\text{ret}})] \\ h_\times(t) &= \frac{4}{r} \left(\frac{GM_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}(t_{\text{ret}})}{c} \right)^{2/3} \cos \iota \sin[\Phi(t_{\text{ret}})] \end{aligned} \quad (4.29)$$

Using the explicit expression (4.19) we find (recall that $d\tau = -dt$)

$$\Phi(\tau) = -2 \left(\frac{5GM_c}{c^3} \right)^{-5/8} \tau^{5/8} + \Phi_0 \quad (4.30)$$

where $\Phi_0 = \Phi(\tau = 0)$ is an integration constant, equal to the value of Φ at coalescence. Then the GW amplitude can be expressed directly in terms of the time to coalescence τ measured by the observer,

$$h_+(t) = \frac{1}{r} \left(\frac{GM_c}{c^2} \right)^{5/4} \left(\frac{5}{c\tau} \right)^{1/4} \left(\frac{1 + \cos^2 \iota}{2} \right) \cos[\Phi(\tau)] \quad (4.31)$$

$$h_\times(t) = \frac{1}{r} \left(\frac{GM_c}{c^2} \right)^{5/4} \left(\frac{5}{c\tau} \right)^{1/4} \cos \iota \sin[\Phi(\tau)] \quad (4.32)$$

where

$$\tau = t_{\text{coal}} - t \quad (4.33)$$

t is the observer time (rather than retarded time) and t_{coal} is the value of t when the coalescence takes place (compare with Note 3). We see from the above equations that both

the frequency and the amplitude increase as the coalescence is approached; this behavior is referred to as "chirping" (for its similarity with the chirp emitted by a bird). The functional dependence of h_+ (or of $h_\times$) on t is shown in Fig. 4.2.

As we will see in detail in Chapter 7, to compare the theoretical waveforms with the experimental sensitivities it is necessary to have the Fourier transform of the GW amplitudes. To compute the Fourier transform of the chirp signal is not completely straightforward, since $h_{+,\times}(t)$

⁷ More precisely, this is valid in the test mass limit, when one of the two stars is much lighter than the other. For finite masses, this result receives corrections that can be expressed as an expansion in powers of the symmetric mass ratio $\nu = \mu/m$. The position of the ISCO is also affected by the spin of the bodies. We will discuss these effects in Vol. 2. are defined only on the interval $-\infty < t < t_{\text{coal}}$, and we compute their Fourier transform explicitly in Problem 4.1. The result is

$$\tilde{h}_+(f) = A e^{i\Psi_+(f)} \frac{c}{r} \left(\frac{GM_c}{c^3} \right)^{5/6} \frac{1}{f^{7/6}} \left(\frac{1 + \cos^2 \iota}{2} \right) \quad (4.34)$$

$$\tilde{h}_\times(f) = A e^{i\Psi_\times(f)} \frac{c}{r} \left(\frac{GM_c}{c^3} \right)^{5/6} \frac{1}{f^{7/6}} \cos \iota \quad (4.35)$$

where the constant A has the value

$$A = \frac{1}{\pi^{2/3}} \left(\frac{5}{24} \right)^{1/2} \quad (4.36)$$

The phases are given by $\Psi_\times = \Psi_+ + (\pi/2)$ and

$$\Psi_+(f) = 2\pi f (t_c + r/c) - \Phi_0 - \frac{\pi}{4} + \frac{3}{4} \left(\frac{GM_c}{c^3} 8\pi f \right)^{-5/3} \quad (4.37)$$

where Φ_0 is the value of the phase Φ at coalescence, see eq. (4.30). We will see in Sections 5.6 and 7.7 that an accurate computation of $\Psi_{+,\times}(f)$, going well beyond the Newtonian approximation, is crucial for discriminating the signal of a coalescing binary from the detector noise, and we will give in eq. (5.275) the post-Newtonian corrections to eq. (4.37).

The above computations have been performed in a flat background. For binary systems made of black holes or neutron stars, the gravitational field close to the stars is however strong, and this has important consequences on the dynamics of the binary system when the two objects get close. A full discussion of a realistic coalescence of compact binaries will be deferred to Vol. 2. However, the most important qualitative modification to the dynamics comes from the fact that, in the Schwarzschild geometry, there is a minimum value of the radial distance beyond which stable circular orbits are no longer allowed, i.e. an Innermost Stable Circular Orbit (ISCO). In Schwarzschild coordinates this is located at $r = r_{\text{ISCO}}$, with

$$r_{\text{ISCO}} = \frac{6Gm}{c^2} \quad (4.38)$$

where $m = m_1 + m_2$ is the total mass of the binary. ⁷ Therefore, for binaries made of BH or NS, a phase of slow adiabatic inspiral, going through a succession of quasi-circular orbit and driven by the emission of gravitational radiation, can only take place at distances $r \gtrsim r_{\text{ISCO}}$. When the ISCO is approached, the dynamics is rather dominated by strong-field effects (namely, by the fact that the Schwarzschild geometry no longer allows for circular orbits), and the two stars plunge toward each other. Thus, the waveform computed above is only valid

up to a maximum frequency $f_{\max}$, beyond which the inspiral phase ends and the two stars plunge toward each other and coalesce. From Kepler's law

(4.1), this means that the inspiral phase ends when the source frequency f_s approaches the value

$$(f_s)_{\text{ISCO}} = \frac{1}{6\sqrt{6}(2\pi)} \frac{c^3}{Gm} \quad (4.39)$$

Inserting the numerical values, this gives

$$(f_s)_{\text{ISCO}} \simeq 2.2\text{kHz} \left(\frac{M_\odot}{m} \right) \quad (4.40)$$

For instance, for a NS-NS system with the typical masses $m_1 = m_2 \simeq 1.4M_\odot$, so that $m \simeq 2.8M_\odot$, we have $(f_s)_{\text{ISCO}} \sim 800$ Hz, while for a BH–BH binary with a total mass $m = 10M_\odot$ we have $(f_s)_{\text{ISCO}} \sim 200$ Hz. The coalescence of two supermassive BHs with $m \sim 10^6 M_\odot$ takes place when f_s is in the mHz region.

Plugging eqs. (4.34) and (4.35) into eq. (1.160) and performing the angular integration we get the energy spectrum in the inspiral phase, in the Newtonian approximation,

$$\frac{dE}{df} = \frac{\pi^{2/3}}{3G} (GM_c)^{5/3} f^{-1/3} \quad (4.41)$$

Integrating up to the maximum GW frequency $f_{\max}$ for which we are still in the inspiral phase we can estimate the total energy radiated during the inspiral phase,

$$\Delta E_{\text{rad}} \sim \frac{\pi^{2/3}}{2G} (GM_c)^{5/3} f_{\max}^{2/3} \quad (4.42)$$

or, inserting the numerical values,

$$\Delta E_{\text{rad}} \sim 4.2 \times 10^{-2} M_\odot c^2 \left(\frac{M_c}{1.21 M_\odot} \right)^{5/3} \left(\frac{f_{\max}}{1\text{kHz}} \right)^{2/3} \quad (4.43)$$

Setting $f_{\max} = 2(f_s)_{\text{ISCO}}$ (as appropriate for quadrupole radiation) and using eq. (4.40) we see that the total energy radiated during the inspiral phase actually depends only on the reduced mass μ of the system, and is

$$\Delta E_{\text{rad}} \sim 8 \times 10^{-2} \mu c^2 \quad (4.44)$$

For stellar mass objects this is a huge amount of energy, and this makes coalescing binaries one of the most interesting sources of GWs. Actually, the above numerical value only provides an order-of-magnitude estimate, since we performed a flat-space computation, and we cut off the result by hand at the value of $(f_s)_{\text{ISCO}}$ obtained from the Schwarzschild geometry. A better approach consists in observing that, in the Schwarzschild metric, the binding energy of the ISCO is given by ⁸

$$E_{\text{binding}} = (1 - \sqrt{8/9})\mu c^2 \simeq 5.7 \times 10^{-2} \mu c^2 \quad (4.45)$$

⁸ See e.g. Landau and Lifshitz, Vol. II (1979), Section 102, Problem 1.

and this is the total energy radiated in GWs when the binary system is slowly inspiraling from an orbit with large relative separation down to the ISCO. We will see in Vol. 2 that the post-Newtonian corrections to this binding energy are of order of a few per cent. Thus, the order-of-magnitude estimate given in eq. (4.44) turns out to be reasonably close to the correct result.

6.0.81 4.1.2 Elliptic orbits. (I) Total power and frequency spectrum of the radiation emitted

We now consider the radiation emitted by masses in an elliptic Keplerian orbit. As before, we denote by m_1, m_2 the masses of the two stars, by $m = m_1 + m_2$ the total mass, and by μ the reduced mass. As usual, in the CM the problem reduces to a one-body problem for a particle of mass μ , subject to an acceleration $\ddot{\mathbf{r}} = -(Gm/r^2) \hat{\mathbf{r}}$. We first recall, from elementary mechanics, the solution of this equation of motion for an elliptic orbit, and then we compute the total power radiated in GWs and its frequency spectrum.

6.0.82 Elliptic Keplerian orbits

The general solution of the equation of motion is obtained by observing that there are two first integral of motion, the angular momentum $\mathbf{L}$, and the energy E . The conservation of $\mathbf{L}$ implies that the orbit lies in a plane; we then introduce polar coordinates (r, ψ) in the plane of the orbit, with origin on the position of the center-of-mass (we use ψ to denote the angular position along the orbit, since we reserve θ and ϕ to describe the angular distribution of the radiation emitted). Then, in terms of r and ψ , the modulus of the angular momentum is given by

$$L = \mu r^2 \dot{\psi} \quad (4.46)$$

while the energy is given by

$$\begin{aligned} E &= \frac{1}{2} \mu \left(\dot{r}^2 + r^2 \dot{\psi}^2 \right) - \frac{G\mu m}{r} \\ &= \frac{1}{2} \mu \dot{r}^2 + \frac{L^2}{2\mu r^2} - \frac{G\mu m}{r} \end{aligned} \quad (4.47)$$

From eq. (4.47) we get dr/dt as a function of r while from eq. (4.46) we get $d\psi/dt$; combining the two expressions we find $dr/d\psi$ as a function of r and, integrating this expression, we obtain the equation of the orbit,

$$\frac{1}{r} = \frac{1}{R}(1 + e \cos \psi) \quad (4.48)$$

The length-scale R and the eccentricity e are constants of motion, and are related to the energy E of the system (with $E < 0$ for a bound orbit) and to the orbital angular momentum L by

$$R = \frac{L^2}{Gm\mu^2} \quad (4.49)$$

and

$$e^2 = 1 + \frac{2EL^2}{G^2 m^2 \mu^3} \quad (4.50)$$

The eccentricity e for an ellipse satisfies $0 \leq e < 1$; for $e = 0$ the ellipse becomes a circle, while $e = 1$ corresponds to a parabola. The two semiaxes of the ellipse are given by

$$a = \frac{R}{1 - e^2} \quad (4.51)$$

$$b = \frac{R}{(1 - e^2)^{1/2}} \quad (4.52)$$

The geometry is shown in Fig. 4.3. Inserting the explicit expression for e given in eq. (4.50), we see that

$$a = \frac{Gm\mu}{2|E|} \quad (4.53)$$

Observe that a is independent of L , so orbits with the same energy have the same value of the semimajor axis a . In terms of a and e , we can rewrite the equation of the orbit (4.48) in the form

$$r = \frac{a(1 - e^2)}{1 + e \cos \psi} \quad (4.54)$$

Combining eqs. (4.46) and (4.49), we see also that $r(t)$ and $\dot{\psi}(t)$ satisfy the relation

$$\dot{\psi} = \frac{(GmR)^{1/2}}{r^2} \quad (4.55)$$

The explicit time dependence of $r(t)$ and $\psi(t)$ is obtained by integrating eqs. (4.46) and (4.47) (making use also of the equation of the orbit) and is given in parametric form by

$$r = a[1 - e \cos u] \quad (4.56)$$

$$\cos \psi = \frac{\cos u - e}{1 - e \cos u} \quad (4.57)$$

where u is called the eccentric anomaly, and is related to t by the famous Kepler equation

$$\beta \equiv u - e \sin u = \omega_0 t \quad (4.58)$$

with

$$\omega_0^2 = \frac{Gm}{a^3} \quad (4.59)$$

We have chosen the origin of time such that, at $t = 0$, we have $\psi = 0$. Making use of trigonometric identities, eq. (4.57) can also be rewritten as

$$\tan \frac{\psi}{2} = \left(\frac{1+e}{1-e} \right)^{1/2} \tan \frac{u}{2} \quad (4.60)$$

i.e.

$$= A_e(u) \equiv 2 \arctan \left[\left(\frac{1+e}{1-e} \right)^{1/2} \tan \frac{u}{2} \right] \quad (4.61)$$

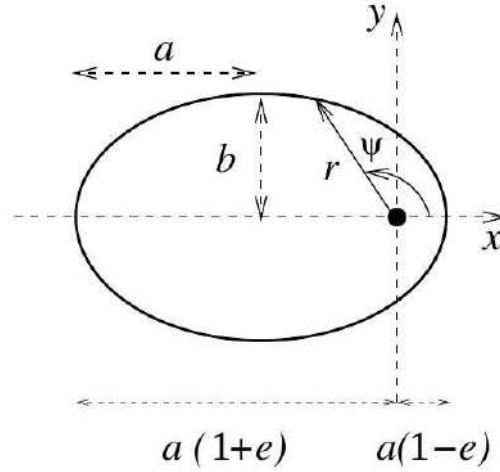


Fig. 4.3 The definitions used for an elliptic orbit. The polar coordinates (r, ψ) , as well as the Cartesian coordinates (x, y) , are centered on a focus of the ellipse (dark blob). The angle ψ is measured counterclockwise, from the x axis. The semiaxes a, b are indicated. The focus splits the major axis into two segments of length $a(1+e)$ and $a(1-e)$, respectively.

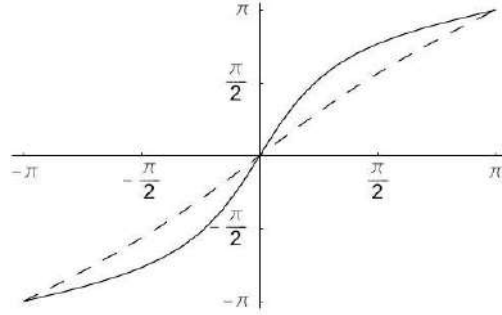


Fig. 4.4 The function $\psi(u)$ for $e = 0.2$ (dashed line) and for $e = 0.75$ (solid line).

⁹ The equation of the ellipse in Cartesian coordinates would be slightly simpler if we choose the center of the ellipse, rather than a focus, as the origin, since in this case we simply have $x = a \cos u(t)$ and $y = b \sin u(t)$. However, (x, y) are already relative coordinates in the CM frame, so we cannot change them with a shift of the origin.

The function $A_e(u)$ is called the true anomaly. Observe from eq. (4.58) that, if $t \rightarrow t + 2\pi/\omega_0$, we have $\beta \rightarrow \beta + 2\pi$ and $u \rightarrow u + 2\pi$, so the coordinates r, ψ are periodic functions of t , with period

$$T = \frac{2\pi}{\omega_0} \quad (4.62)$$

As u runs between $-\pi$ and π , ψ also runs between $-\pi$ and π . A graph of $\psi(u)$ is shown in Fig. 4.4, for two different values of the ellipticity. Observe that, for $e = 0$, $\psi = u$. We will also use Cartesian coordinates (x, y) centered on the focus of the ellipse,

$$\begin{aligned} x &= r \cos \psi \\ &= a[\cos u(t) - e] \end{aligned} \quad (4.63)$$

$$\begin{aligned} y &= r \sin \psi \\ &= b \sin u(t) \end{aligned} \quad (4.64)$$

In terms of the original problem with two bodies of masses m_1 and m_2 , the focus of the ellipse is the point where $\mathbf{x}_{\text{CM}} = 0$, and the coordinates, measured from this point, are the relative coordinates in the CM system. ²

6.0.83 Radiated power

We first compute the total power radiated in GWs, integrated over all frequencies and over the solid angle. We choose a reference frame where the orbit is in the (x, y) plane. In this frame the second mass moment is given by the 2×2 matrix

$$M_{ab} = \mu r^2 \begin{pmatrix} \cos^2 & \sin \psi \cos \psi \\ \sin \psi \cos \psi & \sin^2 \end{pmatrix}_{ab} \quad (4.65)$$

where $a, b = 1, 2$ are indices in the (x, y) plane. To compute the total power emitted in the quadrupole approximation we must evaluate the third derivatives of M_{ab} , and we can then use the quadrupole formula in the form (3.76). In eq. (4.65) the time dependence is both in $r(t)$ and in $\psi(t)$. To compute these derivative, the simplest way is to write M_{ab} as a function of ψ only, eliminating r with the help of eq. (4.54), e.g.

$$\begin{aligned} M_{11} &= \mu r^2 \cos^2 \\ &= \mu a^2 (1 - e^2)^2 \frac{\cos^2}{(1 + e \cos \psi)^2} \end{aligned} \quad (4.66)$$

Now we can compute the time derivatives using

$$\begin{aligned} \dot{\psi} &= \frac{(GmR)^{1/2}}{r^2} \\ &= \left(\frac{Gm}{a^3} \right)^{1/2} (1 - e^2)^{-3/2} (1 + e \cos \psi)^2 \end{aligned} \quad (4.67)$$

where we used eqs. (4.55), (4.54) and (4.51). Then a simple computation gives

$$\ddot{M}_{11} = \beta (1 + e \cos \psi)^2 [2 \sin 2\psi + 3e \sin \psi \cos^2 \psi] \quad (4.68)$$

$$\ddot{M}_{22} = \beta (1 + e \cos \psi)^2 [-2 \sin 2\psi - e \sin \psi (1 + 3 \cos^2 \psi)] \quad (4.69)$$

$$\ddot{M}_{12} = \beta (1 + e \cos \psi)^2 [-2 \cos 2\psi + e \cos \psi (1 - 3 \cos^2 \psi)], \quad (4.70)$$

with

$$\beta^2 \equiv \frac{4G^3 \mu^2 m^3}{a^5 (1 - e^2)^5} \quad (4.71)$$

Plugging this into eq. (3.76) we get the power radiated, in the quadrupole approximation, as a function of the position along the orbit,

$$\begin{aligned} P(\psi) &= \frac{G}{5c^5} \left[\ddot{M}_{11}^2 + \ddot{M}_{22}^2 + 2\ddot{M}_{12}^2 - \frac{1}{3} (\ddot{M}_{11} + \ddot{M}_{22})^2 \right] \\ &= \frac{2G}{15c^5} \left[\ddot{M}_{11}^2 + \ddot{M}_{22}^2 + 3\ddot{M}_{12}^2 - \ddot{M}_{11}\ddot{M}_{22} \right] \\ &= \frac{8G^4}{15c^5} \frac{\mu^2 m^3}{a^5 (1 - e^2)^5} (1 + e \cos \psi)^4 [12(1 + e \cos \psi)^2 + e^2 \sin^2 \psi] \end{aligned} \quad (4.72)$$

As explained in Chapter 1, the GW energy is only well defined by taking a temporal average over several periods of the wave. As we will see below, a particle in a Keplerian elliptic orbit

emits GWs at frequencies which are integer multiple of the frequency ω_0 defined in eq. (4.59), and therefore the period of the GWs is a fraction of the orbital period T given in eq. (4.62). Then, a well defined quantity is the average of $P(\psi(t))$ over one period T . We can now perform this time average, writing

$$\begin{aligned}
 P &\equiv \frac{1}{T} \int_0^T dt P(\psi) \\
 &= \frac{\omega_0}{2\pi} \int_0^{2\pi} \frac{d\psi}{\dot{\psi}} P(\psi) \\
 &= (1 - e^2)^{3/2} \int_0^{2\pi} \frac{d\psi}{2\pi} (1 + e \cos \psi)^{-2} P(\psi) \\
 &= \frac{8G^4 \mu^2 m^3}{15c^5 a^5} (1 - e^2)^{-7/2} \\
 &\quad \times \int_0^{2\pi} \frac{d\psi}{2\pi} [12(1 + e \cos \psi)^4 + e^2(1 + e \cos \psi)^2 \sin^2]
 \end{aligned} \tag{4.73}$$

where we used eq. (4.67). The integral is elementary and one finally obtains the total radiated power ¹⁰

$$P = \frac{32G^4 \mu^2 m^3}{5c^5 a^5} f(e) \tag{4.74}$$

with

$$f(e) = \frac{1}{(1 - e^2)^{7/2}} \left(1 + \frac{73}{24} e^2 + \frac{37}{96} e^4 \right) \tag{4.75}$$

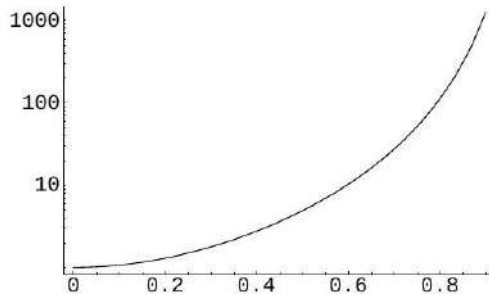


Fig. 4.5 The function $f(e)$, on a logarithmic scale.

Using eq. (4.59) we can eliminate m in favor of ω_0 , and rewrite the above result as

$$P = \frac{32}{5} \frac{G\mu^2}{c^5} a^4 \omega_0^6 f(e) \tag{4.76}$$

For $e = 0$, we have $f(e) = 1$, while a becomes the radius of the circular orbit and ω_0 becomes the same as ω_s , so we get back the result for circular orbits, eq. (3.339). The function $f(e)$ is plotted, on a logarithmic scale, in Fig. 4.5. For instance, for the Hulse-Taylor binary pulsar

¹⁰ This classic formula is due to Peters and Mathews (1963).

(that will be discussed in detail in Chapter 6) the eccentricity is quite large, $e \simeq 0.617$, and $f(e) \simeq 11.8$. The radiated power is therefore an order of magnitude larger than the power emitted in a circular orbit with radius a (i.e. with the same energy).

Combining eqs. (4.62), (4.59) and (4.53) we see that the orbital period T is related to the orbital energy E by $T = \text{const.} \times (-E)^{-3/2}$ and therefore

$$\frac{\dot{T}}{T} = -\frac{3}{2} \frac{\dot{E}}{E} \quad (4.77)$$

From eq. (4.74) we find the energy loss (averaged over one orbital period) $\dot{E} = -P$ and therefore (using again eq. (4.53) to express E in terms of a) we get

$$\frac{\dot{T}}{T} = -\frac{96}{5} \frac{G^3 \mu m^2}{c^5 a^4} f(e) \quad (4.78)$$

where the average over an orbital period is understood. Expressing a in terms of T using eqs. (4.62) and (4.59) we can also rewrite this as

$$\frac{\dot{T}}{T} = -\frac{96}{5} \frac{G^{5/3} \mu m^{2/3}}{c^5} \left(\frac{T}{2\pi} \right)^{-8/3} f(e) \quad (4.79)$$

This equation is of great importance, since it is at the basis of the first experimental evidence for gravitational radiation, as will be discussed in Chapter 6.

Coming back to eq. (4.74), we see that in the limit $e \rightarrow 1^-$ (with a fixed) the radiated power diverges. This is due to the fact that, if we send $e \rightarrow 1^-$ keeping a fixed, we get $R \rightarrow 0$ (see eq. (4.51)) and $b = a(1 - e^2)^{1/2} \rightarrow 0$. Therefore, as the ellipticity $e \rightarrow 1^-$ at fixed a , the motion of the relative coordinates approaches more and more the point $r = 0$, where the acceleration Gm/r^2 diverges, and therefore the GW production formally diverges as well. However, clearly at this moment the approximation of point-like masses ceases to be valid, and we must take into account the finite size of the bodies.

The limit $e \rightarrow 1^-$ at fixed R corresponds instead to parabolic motion,

$$r = \frac{R}{1 + \cos \psi} \quad (4.80)$$

When $\psi \rightarrow -\pi$ the particle is at $r = \infty$; by increasing ψ the value of r decreases down to $r = R/2$ (reached for $\psi = 0$), and then increases again, until again $r \rightarrow \infty$ as $\psi \rightarrow \pi$ (see Fig. 4.6). In this limit, eliminating a in favor of R using eq. (4.51), we obtain from eq. (4.72) the power radiated along the trajectory,

$$\begin{aligned} P(\psi) &= \frac{8G^4 \mu^2 m^3}{15c^5 R^5} (1 + \cos \psi)^4 [12(1 + \cos \psi)^2 + \sin^2 \psi] \\ &= \frac{16G^4 \mu^2 m^3}{15c^5} \frac{1}{r^5} \left(1 + \frac{11R}{2r} \right) \end{aligned} \quad (4.81)$$

Then, $P(\psi)$ goes to zero quite fast as $r \rightarrow \infty$ or, equivalently, as $\psi \rightarrow \pm\pi$, see Fig. 4.7. This is of course expected, since the acceleration vanishes as $1/r^2$. As a result, the total energy radiated in GWs is finite

$$\begin{aligned}
E_{\text{rad}} &= \int_{-\infty}^{\infty} dt P(\psi(t)) \\
&= \int_{-\pi}^{\pi} \frac{d\psi}{\dot{\psi}} P(\psi) \\
&= \frac{85\pi}{48} \frac{G\mu^2}{R} \left(\frac{v_0}{c} \right)^5
\end{aligned} \tag{4.82}$$

where $v_0 = 2(Gm/R)^{1/2}$ is the velocity at $\psi = 0$ (which corresponds to $r = R/2$), i.e. is the maximum velocity attained along the trajectory. In this form, the result is similar to that found for a periodic source, compare, e.g. with eq. (3.319). Since this finite energy is emitted in a time $T = \infty$, if we take the average value of the power over a time T we find zero, as we can also check from eq. (4.74) writing $(1/a^5) f(e) = (1/R^5) (1 - e^2)^5 f(e)$ and taking the limit $e \rightarrow 1^-$ at fixed R . Of course, this reflects simply the fact that the radiation is emitted basically between $-\pi/2 < \psi < \pi/2$, see Fig. 4.7; the radiation emitted when $\pi/2 < |\psi| < \pi$ is instead finite (and, indeed, even negligible), even if it takes an infinite time to get to $\psi = \pi$ or to come from $\psi = -\pi$.

6.0.84 Frequency spectrum

Above we have computed the total radiated power, integrated both over the frequency and over the solid angle. It is however very interesting to compute the frequency spectrum of the radiated power, $dP/d\omega$, for a Keplerian elliptic orbit.

Such a trajectory, given as a function of time by eqs. (4.63), (4.64) and (4.58), is of course not a simple harmonic motion, and therefore the first step is to perform the Fourier decomposition of the trajectory. This can be done by observing, first of all, that $x(t)$ and $y(t)$ are periodic functions of the variable β defined in eq. (4.58), with period 2π . Therefore we can restrict β to the range $-\pi \leq \beta \leq \pi$, and we can perform a discrete Fourier transform,

$$x(\beta) = \sum_{n=-\infty}^{\infty} \tilde{x}_n e^{-in\beta} \tag{4.83}$$

$$y(\beta) = \sum_{n=-\infty}^{\infty} \tilde{y}_n e^{-in\beta} \tag{4.84}$$

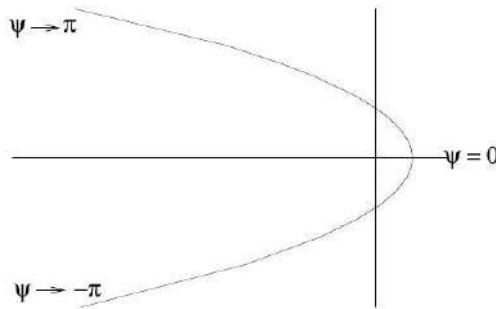


Fig. 4.6 The parabolic trajectory.

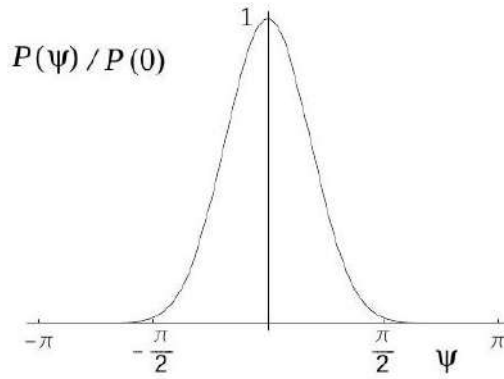


Fig. 4.7 The function $P(\psi)$, normalized to the value at $\psi = 0$, for parabolic motion.

¹¹ Actually, Bessel functions were introduced for the first time just in this context, by Lagrange and, half a century later, in the more general solution for the Fourier transform of the Kepler problem given by Bessel. See Watson (1966). with $\tilde{x}_n = \tilde{x}_{-n}^*$ and $\tilde{y}_n = \tilde{y}_{-n}^*$, since $x(\beta)$ and $y(\beta)$ are real functions. Furthermore, we have chosen the origin of time so that at $t = 0$ (i.e. at $\beta = 0$), we are at the point $x = a(1 - e)$, $y = 0$; with this choice, $x(-\beta) = x(\beta)$ while $y(-\beta) = -y(\beta)$. Therefore, writing $e^{-in\beta} = \cos(n\beta) - i\sin(n\beta)$, in the expansion of $x(\beta)$ contributes only $\cos(n\beta)$ while in the expansion of $y(\beta)$ contributes only $\sin(n\beta)$, and eqs. (4.83) and (4.84) simplify to

$$x(\beta) = \sum_{n=0}^{\infty} a_n \cos(n\beta) \quad (4.85)$$

$$y(\beta) = \sum_{n=1}^{\infty} b_n \sin(n\beta) \quad (4.86)$$

where, for $n \geq 1$, $a_n = 2\tilde{x}_n$ and $b_n = -2i\tilde{y}_n$, while $a_0 = \tilde{x}_0$. Since $\beta = \omega_0 t$, we can rewrite this as

$$x(t) = \sum_{n=0}^{\infty} a_n \cos \omega_n t \quad (4.87)$$

$$y(t) = \sum_{n=1}^{\infty} b_n \sin \omega_n t \quad (4.88)$$

where

$$\omega_n = n\omega_0 \quad (4.89)$$

We see that the Fourier decomposition of the Keplerian motion involves the fundamental frequency $\omega_0 = (Gm/a^3)^{1/2}$ and all its higher harmonics. The coefficients a_n and b_n are obtained by inverting eqs. (4.85) and (4.86), which gives, for $n \neq 0$,

$$a_n = \frac{2}{\pi} \int_0^{\pi} d\beta x(\beta) \cos(n\beta) \quad (4.90)$$

$$b_n = \frac{2}{\pi} \int_0^{\pi} d\beta y(\beta) \sin(n\beta) \quad (4.91)$$

while, for $n = 0$,

$$a_0 = \frac{1}{\pi} \int_0^\pi d\beta x(\beta) \quad (4.93)$$

With $x(\beta)$ and $y(\beta)$ given by eqs. (4.63), (4.64) and (4.58), these integrals are computed explicitly in Problem 4.2, and the result is given in terms of Bessel functions,¹¹

$$a_n = \frac{a}{n} [J_{n-1}(ne) - J_{n+1}(ne)] \quad (4.94)$$

$$b_n = \frac{b}{n} [J_{n-1}(ne) + J_{n+1}(ne)] \quad (4.95)$$

for $n \neq 0$, and $a_0 = -(3/2)ae$. The above result is interesting in itself but, to compute the spectrum of GWs, we are really interested in the

Fourier decomposition of the second mass moment, and therefore of $x^2(t)$, $y^2(t)$ and $x(t)y(t)$. This is again computed in Problem 4.2, where we show that

$$x^2(t) = \sum_{n=0}^{\infty} A_n \cos \omega_n t \quad (4.96)$$

$$y^2(t) = \sum_{n=0}^{\infty} B_n \cos \omega_n t \quad (4.97)$$

$$x(t)y(t) = \sum_{n=1}^{\infty} C_n \sin \omega_n t \quad (4.98)$$

where

$$A_n = \frac{a^2}{n} [J_{n-2}(ne) - J_{n+2}(ne) - 2e (J_{n-1}(ne) - J_{n+1}(ne))] \quad (4.99)$$

$$B_n = \frac{b^2}{n} [J_{n+2}(ne) - J_{n-2}(ne)] \quad (4.100)$$

$$C_n = \frac{ab}{n} [J_{n+2}(ne) + J_{n-2}(ne) - e (J_{n+1}(ne) + J_{n-1}(ne))] \quad (4.101)$$

The second mass moment (4.65) therefore has the Fourier decomposition

$$\begin{aligned} M_{ab}(t) &= \mu \sum_{n=0}^{\infty} \begin{pmatrix} A_n \cos \omega_n t & C_n \sin \omega_n t \\ C_n \sin \omega_n t & B_n \cos \omega_n t \end{pmatrix}_{ab} \\ &\equiv \sum_{n=0}^{\infty} M_{ab}^{(n)}(t) \end{aligned} \quad (4.102)$$

When we compute the radiated power, temporal averages such as

$$\langle \sin \omega_n t \sin \omega_m t \rangle$$

are non-vanishing only if $n = m$, so there is no interference term between the different harmonics, and

$$P = \sum_{n=1}^{\infty} P_n \quad (4.103)$$

where P_n is the power radiated in the n -th harmonics. To compute P_n we use the quadrupole formula, written in the form

$$P = \frac{2G}{15c^5} \left\langle \ddot{M}_{11}^2 + \ddot{M}_{22}^2 + 3\ddot{M}_{12}^2 - \ddot{M}_{11}\ddot{M}_{22} \right\rangle \quad (4.104)$$

(that takes into account that the orbit is in the (x, y) plane, compare with eq. (4.72)), and we replace M_{ab} by $M_{ab}^{(n)}$. Using

$$\ddot{M}_{ab}^{(n)} = \mu\omega_n^3 \begin{pmatrix} A_n \sin \omega_n t & -C_n \cos \omega_n t \\ -C_n \cos \omega_n t & B_n \sin \omega_n t \end{pmatrix}_{ab} \quad (4.105)$$

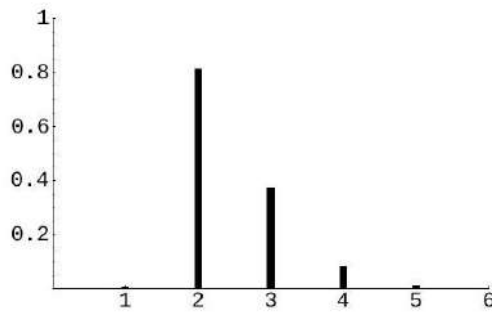


Fig. 4.8 The power P_n radiated at the GW frequency $\omega_n = n\omega_0$, as a function of n , for $e = 0.2$. P_n is normalized to the value for $e = 0$.

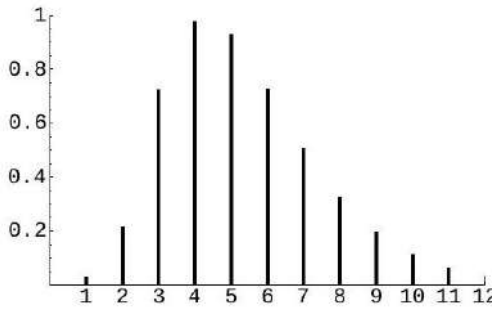


Fig. 4.9 The same as Fig. 4.8, for $e = 0.5$.

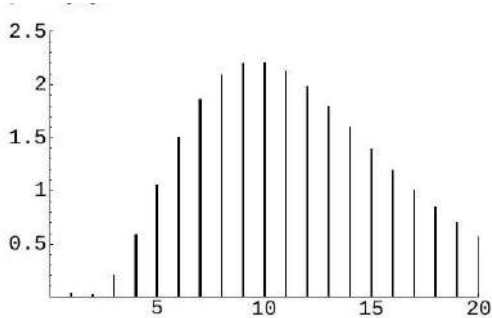


Fig. 4.10 The same as Fig. 4.8, for $e = 0.7$. we get

$$P_n = \frac{G\mu^2\omega_0^6}{15c^5} n^6 (A_n^2 + B_n^2 + 3C_n^2 - A_n B_n) \quad (4.106)$$

where we used $\langle \sin^2 \omega_n t \rangle = \langle \cos^2 \omega_n t \rangle = 1/2$. Recalling that $\omega_0^2 = Gm/a^3$, the above result can be rewritten as

$$P_n = \frac{32G^4 \mu^2 m^3}{5c^5 a^5} g(n, e) \quad (4.107)$$

where

$$g(n, e) = \frac{n^6}{96a^4} [A_n^2(e) + B_n^2(e) + 3C_n^2(e) - A_n(e)B_n(e)] \quad (4.108)$$

where the notation stresses that A_n, B_n and C_n depends on the eccentricity e (recall also that $b^2/a^2 = (1 - e^2)$).

First of all, we can check from this expression that, in the limit $e \rightarrow 0$, we get back the result for circular motion. If $e = 0$, the argument of the Bessel functions, in eqs. (4.99)-(4.101), vanishes. Since $J_m(0) = \delta_{m0}$, and since for us $n \geq 1$, we see from eqs. (4.99)-(4.101) that in this limit $A_n \rightarrow (a^2/2) \delta_{n2}$, $B_n \rightarrow (-a^2/2) \delta_{n2}$ (of course $b = a$ when $e = 0$) and $C_n \rightarrow (a^2/2) \delta_{n2}$. Then we get back the result that, for circular motion, only the harmonic with $n = 2$ contributes, and we find from eq. (4.108) that $g(0, n) = \delta_{n2}$, so we get back the correct value for the radiated power.

For a generic value of e (with $0 < e < 1$) all harmonics contribute, and we have radiation at all frequencies $\omega_n = n\omega_0$ for all integer values of n , including also $n = 1$. Increasing the ellipticity, increases also the value of $n = \bar{n}$ where $g(n, e)$ is maximum, as well as the value of P_n at $n = \bar{n}$, and the total radiated power. In Figs. 4.8 – 4.10 we show the power spectrum P_n , normalized to the power radiated when $e = 0$, for three different values of the ellipticity. Observe the change in the horizontal and vertical scales, among these figures.

6.0.85 4.1.3 Elliptic orbits. (II) Evolution of the orbit under back-reaction

6.0.86 The evolution equations

A binary system in a Keplerian orbit radiates both energy and angular momentum. In our approximation of point-like bodies (without an intrinsic spin) these are necessarily drained from the energy and angular momentum of the orbital motion, which therefore undergoes secular changes, both in its semimajor axis and in its ellipticity, until the system enters the merging phase and collapses. In this section we compute how the shape and size of the orbit evolve, for a generic elliptic orbit.

We already computed the energy radiated, in the quadrupole approximation, in eq. (4.74). We now compute the angular momentum radiated, again in the quadrupole approximation. We therefore start from eq. (3.99), which we write as

$$\frac{dL^i}{dt} = -\frac{2G}{5c^5} \epsilon^{ikl} \langle \ddot{M}_{ka} \ddot{M}_{la} \rangle \quad (4.109)$$

where L^i is the orbital angular momentum of the binary system, and we replaced $\ddot{Q}_{ka}$ by $\ddot{M}_{ka}$ (because the difference gives a contribution to eq. (4.109) proportional to $\epsilon^{ikl} \delta_{ka} \ddot{Q}_{la} = \epsilon^{ikl} \ddot{Q}_{kl}$, which vanishes because $\ddot{Q}_{kl}$ is symmetric in (k, l) and ϵ^{ikl} is antisymmetric), and similarly we replaced $\ddot{Q}_{la}$ by $\ddot{M}_{la}$. As in the computation of the radiated energy, we put the orbit in the (x, y) plane, so M_{ab} is given by eq. (4.65), and we write $L_z = L$. Then, recalling that inside $\langle \dots \rangle$ we can integrate by parts, eq. (4.109) gives

$$\begin{aligned}\frac{dL}{dt} &= -\frac{2G}{5c^5} \left\langle \ddot{M}_{1a} \ddot{M}_{2a} - \ddot{M}_{2a} \ddot{M}_{1a} \right\rangle \\ &= \frac{4G}{5c^5} \left\langle \ddot{M}_{12} \left(\ddot{M}_{11} - \ddot{M}_{22} \right) \right\rangle\end{aligned}\quad (4.110)$$

The third derivatives $\ddot{M}_{11}$ and $\ddot{M}_{22}$ have been computed in eqs. (4.68) and (4.69), while a similar computation gives

$$\begin{aligned}\ddot{M}_{12} &= \frac{G\mu m}{a(1-e^2)} \sin \psi \\ &\times \left[-4(1+e \cos \psi)^2 \cos \psi + 2e(3 \cos^2 \psi - 1 + 2e \cos^3 \psi) \right]\end{aligned}\quad (4.111)$$

For periodic motion, the average over several periods of the wave is the same as the average over one orbital period T and, as in eq. (4.73), we transform the temporal average over one period into an integration over ψ , using

$$\int_0^T \frac{dt}{T} (\dots) = (1-e^2)^{3/2} \int_0^{2\pi} \frac{d\psi}{2\pi} (1+e \cos \psi)^{-2} (\dots) \quad (4.112)$$

Then we get, for the angular momentum loss averaged over one period,

$$\begin{aligned}\frac{dL}{dt} &= \frac{8}{5} \frac{G^{7/2} \mu^2 m^{5/2}}{c^5 a^{7/2}} \frac{1}{(1-e^2)^2} \int_0^{2\pi} \frac{d\psi}{2\pi} \sin^2 \\ &\times \left[-4(1+e \cos \psi)^2 \cos \psi + 2e(3 \cos^2 \psi - 1 + 2e \cos^3 \psi) \right] \\ &\times \left[8 \cos \psi + 6e \cos^2 \psi + e \right]\end{aligned}\quad (4.113)$$

The integral over ψ is elementary, and gives $-4(1+7e^2/8)$. In conclusion, from this result and eq. (4.74), the energy E and the angular momentum L of the orbit evolve as

$$\frac{dE}{dt} = -\frac{32}{5} \frac{G^4 \mu^2 m^3}{c^5 a^5} \frac{1}{(1-e^2)^{7/2}} \left(1 + \frac{73}{24} e^2 + \frac{37}{96} e^4 \right) \quad (4.114)$$

$$\frac{dL}{dt} = -\frac{32}{5} \frac{G^{7/2} \mu^2 m^{5/2}}{c^5 a^{7/2}} \frac{1}{(1-e^2)^2} \left(1 + \frac{7}{8} e^2 \right) \quad (4.115)$$

and we recall again that these quantities are really averages over one period, rather than instantaneous energy and angular momentum losses. Using eqs. (4.50) and (4.53) we can rewrite these equations in terms of the evolution of the semimajor axis a and of the eccentricity e ,

$$\frac{da}{dt} = -\frac{64}{5} \frac{G^3 \mu m^2}{c^5 a^3} \frac{1}{(1-e^2)^{7/2}} \left(1 + \frac{73}{24} e^2 + \frac{37}{96} e^4 \right) \quad (4.116)$$

$$\frac{de}{dt} = -\frac{304}{15} \frac{G^3 \mu m^2}{c^5 a^4} \frac{e}{(1-e^2)^{5/2}} \left(1 + \frac{121}{304} e^2 \right) \quad (4.117)$$

From eq. (4.117) we see that, if $e = 0$, then $de/dt = 0$. Therefore a circular orbit remains circular. This could also have been seen directly from eqs. (4.114) and (4.115), observing that, for $e = 0$, they give

$$\frac{dE}{dt} = \omega_0 \frac{dL}{dt} \quad (4.118)$$

where $\omega_0 = (Gm/a^3)^{1/2}$ is the frequency of the circular motion, see eq. (4.59). Equation (4.118) is just what is needed to maintain the relation between energy and angular momentum that holds for circular motion. In fact, for circular orbits, using eq. (4.59),

$$E = -\frac{Gm\mu}{2a} = -\frac{1}{2}(Gm)^{2/3}\mu\omega_0^{2/3} \quad (4.119)$$

while

$$L = \mu a^2 \omega_0 = (Gm)^{2/3} \mu \omega_0^{-1/3} \quad (4.120)$$

and therefore

$$\frac{dE}{dt} = -\frac{1}{3}(Gm)^{2/3}\mu\omega_0^{-1/3}\dot{\omega}_0 \quad (4.121)$$

$$\frac{dL}{dt} = -\frac{1}{3}(Gm)^{2/3}\mu\omega_0^{-4/3}\dot{\omega}_0 \quad (4.122)$$

from which eq. (4.118) follows. For $e > 0$, eq. (4.117) gives $de/dt < 0$ instead, and therefore an elliptic orbit becomes more and more circular because of the emission of GWs.

A direct numerical integration of eqs. (4.116) and (4.117) is not as straightforward as one might think. The reason can be seen by putting the equations in dimensionless form. To do so, it is convenient to introduce a length-scale R_* , from $R_*^3 = 4G^3\mu m^2/c^6$. We have chosen the numerical factor so that when the masses of the two stars are equal, $m_1 = m_2 = M$, we have $R_* = 2GM/c^2$, so R_* becomes equal to the Schwarzschild radius of the stars. We then introduce the dimensionless variable $\tau = ct/R_*$, i.e. time measured in units of the light travel time across a distance R_* , and the dimensionless function $\tilde{a}(\tau) = a(\tau)/R_*$. Then eqs. (4.116) and (4.117) read

$$\frac{d\tilde{a}}{d\tau} = -\frac{16}{5} \frac{1}{\tilde{a}^3} \frac{1}{(1-e^2)^{7/2}} \left(1 + \frac{73}{24}e^2 + \frac{37}{96}e^4 \right) \quad (4.123)$$

$$\frac{de}{d\tau} = -\frac{76}{15} \frac{1}{\tilde{a}^4} \frac{e}{(1-e^2)^{5/2}} \left(1 + \frac{121}{304}e^2 \right) \quad (4.124)$$

This shows that $\tau = ct/R_*$ is the natural adimensional time-scale in the differential equation. However, for typical solar-mass stars, $R_* \sim 3$ km and $R_*/c = O(10^{-5})$ s. Therefore, to follow the evolution of the orbit for a time $t \sim 1$ yr, we need to push the integration up to a dimensionless value $\tau \sim 10^{12}$, which is numerically difficult.

A better approach is to combine eqs. (4.116) and (4.117) to get da/de ,

$$\frac{da}{de} = \frac{12}{19} a \frac{1 + (73/24)e^2 + (37/96)e^4}{e(1-e^2)[1 + (121/304)e^2]} \quad (4.125)$$

This equation can be integrated analytically, and gives

$$a(e) = c_0 \frac{e^{12/19}}{1-e^2} \left(1 + \frac{121}{304}e^2 \right)^{870/2299} \quad (4.126)$$

where c_0 is determined by the initial condition $a = a_0$ when $e = e_0$. It is convenient to define the function¹²

$$g(e) = \frac{e^{12/19}}{1 - e^2} \left(1 + \frac{121}{304}e^2\right)^{870/2299} \quad (4.127)$$

so that $a(e) = c_0 g(e)$; c_0 is fixed by $a_0 = c_0 g(e_0)$, and therefore

$$a(e) = a_0 \frac{g(e)}{g(e_0)} \quad (4.128)$$

In particular, for $e \ll 1$, we have $g(e) \simeq e^{12/19}$ while, for e close to 1,

$$g(e) \simeq \frac{g_1}{1 - e^2} \quad (4.129)$$

where $g_1 = (425/304)^{870/2299} \simeq 1.1352$. A plot of the function $g(e)$ is shown in Fig. 4.11.

6.0.87 Orbit circularization

A consequence of the above result is that the eccentricity decreases quite fast, so the effect of the back-reaction of GWs is to circularize the orbit. Consider a compact binary system, say a NS-NS binary, that at an initial time has a_0 very large compared to the radius of the neutron stars, so we are still very far from coalescence, and an eccentricity e_0 not particularly small, neither especially close to one, so that $g(e_0) = O(1)$; let a and e be the values of the semiaxis and eccentricity at a much later time, say when the system is approaching the coalescence phase. Then e will be small and, from eq. (4.128) and the small e limit of $g(e)$, we have

$$a(e) \simeq a_0 \frac{e^{12/19}}{g(e_0)} \quad (4.130)$$

i.e.

$$e \simeq \left[\frac{a}{a_0} g(e_0) \right]^{19/12} \quad (4.131)$$

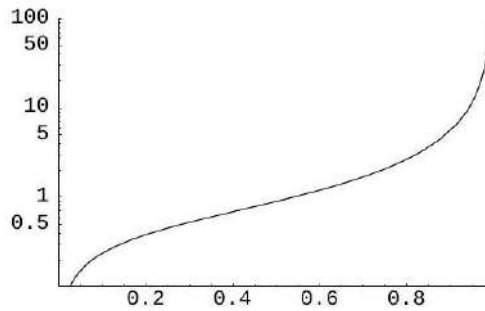


Fig. 4.11 The function $g(e)$, on a logarithmic scale, against e .

For example, consider the Hulse-Taylor binary pulsar (which will be discussed in detail in Chapter 6). Today it has a semimajor axis $a_0 \simeq 2 \times 10^9$ m and a rather large eccentricity $e_0 \simeq 0.617$. By the time that the two stars reach a short separation a , say of the order of a hundred times the radius of the neutron stars, $a = O(10^2 R_{\text{NS}}) \simeq 10^3$ km, we have $a/a_0 = O(5 \times 10^{-4})$ and, since $g(e_0) = O(1)$, the eccentricity has become $e \sim (5 \times 10^{-4})^{19/12} \sim 6 \times 10^{-6}$. The

¹² This has nothing to do with the function $g(n, e)$ defined in eq. (4.108). Both the notation $g(e)$ for this function and the notation $g(n, e)$ for the function defined by eq. (4.108) are standard in the literature.

conclusion is that (unless some external interaction perturbs the binary system), long before the two neutron stars approach the coalescence phase, the ellipticity has become zero to very high accuracy, and the two stars move on a circular orbit which shrinks adiabatically.

6.0.88 The time to coalescence

We now compute the time to coalescence, $\tau(a_0, e_0)$, of a binary system that, at an initial time $t_0 = 0$, has semimajor axis a_0 and eccentricity e_0 . When $e_0 = 0$, we found the result in eq. (4.26),

$$\tau(a_0, e_0 = 0) \equiv \tau_0(a_0) = \frac{5}{256} \frac{c^5 a_0^4}{G^3 m^2 \mu} \quad (4.132)$$

For an elliptic orbit, we can integrate eq. (4.116) requiring $a(t) = 0$ at $t = \tau(a_0, e_0)$ or, equivalently, we can integrate eq. (4.117) requiring $e(t) = 0$ at $t = \tau(a_0, e_0)$, since we have seen that at the coalescence e goes to zero. Since the analytic expression for $a(e)$ is simpler than the form of the inverse function $e(a)$, it is in fact better to use eq. (4.117), so we write

$$\int_0^{\tau(a_0, e_0)} dt = -\frac{15}{304} \frac{c^5}{G^3 m^2 \mu} \int_{e_0}^0 de \frac{a^4(e) (1 - e^2)^{5/2}}{e \left(1 + \frac{121}{304} e^2\right)} \quad (4.133)$$

that is,

$$\tau(a_0, e_0) = \frac{15}{304} \frac{c^5}{G^3 m^2 \mu} \int_0^{e_0} de \frac{a^4(e) (1 - e^2)^{5/2}}{e \left(1 + \frac{121}{304} e^2\right)} \quad (4.134)$$

Using eq. (4.128) for $a(e)$, we get

$$\tau(a_0, e_0) = \tau_0(a_0) \frac{48}{19} \frac{1}{g^4(e_0)} \int_0^{e_0} de \frac{g^4(e) (1 - e^2)^{5/2}}{e \left(1 + \frac{121}{304} e^2\right)} \quad (4.135)$$

Expressing $\tau_0(a_0)$ in terms of the orbital period as in eq. (4.27), we can write the result for $\tau_0(a_0, e_0)$ as

$$\tau_0(a_0, e_0) \simeq 9.829 \text{Myr} \left(\frac{T_0}{1 \text{hr}}\right)^{8/3} \left(\frac{M_\odot}{m}\right)^{2/3} \left(\frac{M_\odot}{\mu}\right) F(e_0)$$

where $m = m_1 + m_2$ is the total mass, μ is the reduced mass,

$$F(e_0) = \frac{48}{19} \frac{1}{g^4(e_0)} \int_0^{e_0} de \frac{g^4(e) (1 - e^2)^{5/2}}{e \left(1 + \frac{121}{304} e^2\right)} \quad (4.137)$$

and $g(e)$ is given in eq. (4.127). For example, the Hulse-Taylor binary pulsar has $e_0 \simeq 0.617$ and $F(e_0) \simeq 0.184$, so its time to coalescence is shorter by a factor $1/F(e_0) \simeq 5.4$, compared to a binary on a circular orbit with the same period. Using the values, $m_1 = m_2 \simeq 1.4 M_\odot$ and $T_0 \simeq 7.75 \text{hr}$ we obtain the time to coalescence of the Hulse-Taylor binary pulsar, $\tau(a_0, e_0) \simeq 300 \text{Myr}$.

It can be useful to write some approximate, but more handy expression, for $F(e_0)$. In the limit $e_0 \ll 1$ the integrand can be approximated as $g^4(e)/e$, and $g(e) \simeq e^{12/19}$, see eq. (4.127). Then

$$\begin{aligned}
F(e_0) &\simeq \frac{48}{19} \frac{1}{g^4(e_0)} \int_0^{e_0} de e^{48/19-1} \\
&= \frac{1}{g^4(e_0)} e_0^{48/19} \simeq 1
\end{aligned} \tag{4.138}$$

so for $e_0 = 0$ we have $F(0) = 1$ and we get back the result for circular orbit, eq. (4.27). In the opposite limit $e_0 \rightarrow 1^-$, the integral in eq. (4.134) is dominated by the region $e \simeq 1$, where $g(e)$ is given by eq. (4.129). Then we get

$$\begin{aligned}
F(e_0) &\simeq \frac{48}{19} \left(\frac{1 - e_0^2}{g_1} \right)^4 \int_0^{e_0} de \frac{1}{[1 + (121/304)]} \frac{g_1^4}{(1 - e^2)^4} (1 - e^2)^{5/2} \\
&\simeq \frac{768}{429} (1 - e_0^2)^4 \left[(1 - e_0^2)^{-1/2} + O(1) \right] \\
&\simeq \frac{768}{429} (1 - e_0^2)^{7/2}
\end{aligned} \tag{4.139}$$

Therefore, in this limit the time to coalescence is smaller than for a circular orbit with the same period, by a factor proportional to $(1 - e_0^2)^{7/2}$. This dependence is easily understood from the fact that the system spends most of its lifetime near $a = a_0, e = e_0$, where the radiated energy is enhanced by a factor proportional to $(1 - e_0^2)^{7/2}$, compared to the circular case, see eqs. (4.74) and (4.75). Comparing the limits $e \rightarrow 0$ and $e \rightarrow 1^-$, we see that it is convenient to define a function $G(e_0)$ from

$$F(e_0) \equiv G(e_0) (1 - e_0^2)^{7/2} \tag{4.140}$$

because $G(e_0)$ is a function everywhere close to one. A plot of $G(e_0)$ from the exact numerical integration of eq. (4.137) is shown in Fig. 4.12. We see that $G(e)$ is always very close to one, within a few per cent at least until $e = 0.6$, where $G(e) \simeq 0.982$, and then, approaching $e = 1$, it raises to reach the finite value $G(1) = 768/425 \simeq 1.80$.

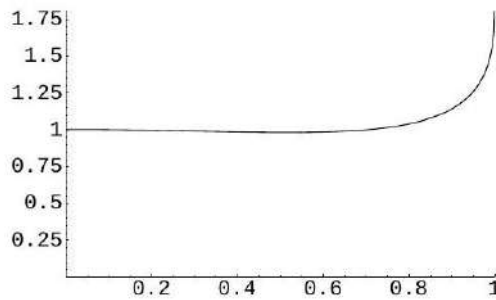


Fig. 4.12 The function $G(e)$, against e . At $e = 1$ $G(e)$ is finite, $G(1) = 768/425 \simeq 1.80$.

¹³ Here the word "flat" refers to the geometry of the spatial slices. As a fourdimensional space-time, the FRW metric with $k = 0$ is curved.

¹⁴ Equation (4.142) also implies a much stronger result. If at $t = t_0$ the velocity has an initial value $\mathbf{u}_0$, the integration of eq. (4.142) gives

$$|\mathbf{u}(t)| = \frac{a(t_0)}{a(t)} |\mathbf{u}_0| \tag{4.143}$$

so spatial velocities are redshifted by the cosmological expansion. In a Universe which expands forever, even test masses with a non-vanishing initial value of $\mathbf{u}$ will eventually come to rest with respect to the observer that uses comoving coordinates.

6.0.89 4.1.4 Binaries at cosmological distances

Until now, we have implicitly assumed that the binary system that coalesces is at a sufficiently small distance from us, so that the expansion of the Universe during the propagation of the GWs from the source to the detector can be neglected. However, some gravitational-wave detectors, in their advanced stage, have the potential of observing coalescing binaries out to cosmological distances: we will see that advanced groundbased interferometers could detect BH – BH coalescences out to distances of order 1-2 Gpc (corresponding to a redshift $z \sim 0.25 - 0.5$), while the space interferometer LISA could detect the coalescence of super-massive black-hole binaries out to redshifts $z \sim 5 - 10$. Actually, these will be among the most fascinating sources of GW astronomy. The rich cosmological information that could be extracted from such events will be discussed in Vol. 2. Here we prepare the necessary tools, discussing how the signal of a binary inspiral at a cosmological distance is affected by the expansion of the Universe.

First of all, we recall a few elementary notions of cosmology (see, e.g. Chapter 2 of Kolb and Turner (1990) for more details and derivations), also in order to fix our notation. The expert reader might wish to skip this part and move on to page 194.

6.0.90 A reminder of FRW kinematic

On the Gpc scale, the Universe is to a first approximation isotropic and homogeneous, and is described by the Friedmann-Robertson-Walker (FRW) metric,

$$ds^2 = -c^2 dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \right] \quad (4.141)$$

where the function $a(t)$ is called the scale factor and is determined by the Einstein equations. After an appropriate rescaling of the coordinates, k can take the values $k = 0$ (flat Universe¹³), $k = +1$ (closed Universe) or $k = -1$ (open Universe). We define $a(t)$ as adimensional, so the dimensions of length are carried by r . The coordinates (t, r, θ, ϕ) that appear above are called comoving coordinates. This name reflects the property that a test mass, initially at rest in the comoving frame, remains at a fixed value of its comoving coordinates r, θ, ϕ in spite of the expansion of the Universe. This can be proved writing down the geodesic equation in the metric (4.141) (just as we did in Section 1.3, when we understood the physical meaning of the TT gauge). From the $\mu = 0$ component, one finds

$$\frac{d|\mathbf{u}|}{dt} = -\frac{\dot{a}}{a}|\mathbf{u}| \quad (4.142)$$

where $|\mathbf{u}|^2 = g_{ij}u^i u^j$ is the squared modulus of the spatial part of the four-velocity u^μ . This equation shows that, in the comoving frame, if at $t = t_0$ the modulus of the four-velocity $|\mathbf{u}|$ is zero, then $d|\mathbf{u}|/dt$ also vanishes, so $|\mathbf{u}|$ remains zero at all times.¹⁴ This explains why these coordinates are called comoving: they "stretch" themselves, following the expansion of the Universe. If, at an initial time, a star has comoving spatial coordinates (r_1, θ_1, ϕ_1) , and a second star has comoving spatial coordinates (r_2, θ_2, ϕ_2) , and if both stars have zero velocity, then at any subsequent time their comoving spatial coordinates will still be (r_1, θ_1, ϕ_1) and (r_2, θ_2, ϕ_2) , respectively, and in particular their coordinate distance will be unchanged, in spite of the expansion of the Universe. Comparing with the discussion in Section 1.3.3, we see that comoving coordinates in a FRW space-time play the same role as TT coordinates in the space-time generated by a gravitational wave.

A star at $r = r_1$ and a star at $r = r_2$ (and with the same values of θ and ϕ) have a coordinate distance $r = r_2 - r_1$. Of course, this quantity has no physical meaning, since coordinates are

arbitrary in general relativity. The physical spatial distance is the proper spatial distance, $r_{\text{phys}}(t)$, given by

$$dr_{\text{phys}}^2 = g_{ij}dx^i dx^j \quad (4.144)$$

If the first star is at the origin and the second is at comoving radial coordinate r , then eq. (4.141) gives $dr_{\text{phys}}^2 = a^2(t)dr^2/(1 - kr^2)$, and therefore their physical distance is

$$r_{\text{phys}}(t) = a(t) \int_0^r \frac{dr}{(1 - kr^2)^{1/2}} \quad (4.145)$$

In particular, for a flat Universe ($k = 0$) we have $r_{\text{phys}}(t) = a(t)r$.

Consider now a source located at comoving distance r , that emits signals (such as electromagnetic or gravitational waves) which travel at the speed of light, and are later received by an observer located at $r = 0$. Suppose that the source emits a wavecrest at a time t_{emis} . The signal will be detected by the observer at a time t_{obs} , which is obtained by imposing $ds^2 = 0$ in eq. (4.141). This gives

$$\int_{t_{\text{emis}}}^{t_{\text{obs}}} \frac{cdt}{a(t)} = \int_0^r \frac{dr}{(1 - kr^2)^{1/2}} \quad (4.146)$$

Suppose that a second wavecrest is emitted at time $t_{\text{emis}} + \Delta t_{\text{emis}}$, and received at $t_{\text{obs}} + \Delta t_{\text{obs}}$. Then

$$\int_{t_{\text{emis}} + \Delta t_{\text{emis}}}^{t_{\text{obs}} + \Delta t_{\text{obs}}} \frac{cdt}{a(t)} = \int_0^r \frac{dr}{(1 - kr^2)^{1/2}} \quad (4.147)$$

Observe that the right-hand side is always the same, since the source is at a fixed comoving distance. Then, taking the difference between these two equations, the right-hand sides cancel and, to linear order in Δt_{emis} , one finds

$$\Delta t_{\text{obs}} = \frac{a(t_{\text{obs}})}{a(t_{\text{emis}})} \Delta t_{\text{emis}} \quad (4.148)$$

In an expanding Universe, there is therefore a dilatation of the time measured by the observer, with respect to the time of the source. The redshift of the source, z , is defined by

$$1 + z = \frac{a(t_{\text{obs}})}{a(t_{\text{emis}})} \quad (4.149)$$

Equation (4.148) means that the time t_{obs} measured by the observer's clock, and the time t_s measured by the source clock are related by

$$dt_{\text{obs}} = (1 + z)dt_s \quad (4.150)$$

As a consequence, any frequency measured by the observer, $f^{(\text{obs})}$, is related to the corresponding frequency measured in the source frame, $f^{(s)}$, by

$$f^{(\text{obs})} = \frac{f^{(s)}}{1 + z} \quad (4.151)$$

and the wavelengths are related by

$$\lambda^{(\text{obs})} = (1 + z)\lambda^{(s)} \quad (4.152)$$

To complete our quick tour of kinematics in a FRW space-time, we need the definition of luminosity distance and the relation between the luminosity distance and the redshift. Let $\mathcal{F}$ be the energy flux (energy per unit time per unit area) measured in the observer's frame, and let $\mathcal{L}$ be the absolute luminosity of the source, i.e. the power that it radiates in its rest frame,

$$\mathcal{L} = \frac{dE_s}{dt_s} \quad (4.153)$$

where E_s is the energy measured in the source proper frame. The luminosity distance d_L is then defined by

$$\mathcal{F} = \frac{\mathcal{L}}{4\pi d_L^2} \quad (4.154)$$

In the absence of redshift, the energy E_{obs} measured in the observer frame is the same as the energy E_s measured in the source frame, and $dt_{\text{obs}} = dt_s$, so

$$\frac{dE_{\text{obs}}}{dt_{\text{obs}}} = \frac{dE_s}{dt_s} \quad (4.155)$$

When it arrives at the detector, the energy radiated by an isotropic source is distributed over an area $A = 4\pi r^2$, and therefore

$$\begin{aligned} \mathcal{F} &= \frac{1}{4\pi r^2} \frac{dE_{\text{obs}}}{dt_{\text{obs}}} \\ &= \frac{1}{4\pi r^2} \frac{dE_s}{dt_s} \\ &= \frac{\mathcal{L}}{4\pi r^2} \end{aligned} \quad (4.156)$$

which shows that, in the absence of redshift, d_L is just equal to the distance r of the source. In an expanding Universe, however, the energy observed is redshifted,

$$E_{\text{obs}} = \frac{E_s}{1+z} \quad (4.157)$$

(the quickest derivation of this result makes use of eq. (4.151), together with the quantum relation $E = \hbar\omega$, but this is a kinematic property that can also be derived purely classically), while $dt_{\text{obs}} = (1+z)dt_s$. Therefore

$$\frac{dE_{\text{obs}}}{dt_{\text{obs}}} = \frac{1}{(1+z)^2} \frac{dE_s}{dt_s} \quad (4.158)$$

Furthermore, using the FRW metric (4.141), we find that, at time t , the surface of a sphere with comoving radius r is $4\pi a^2(t)r^2$ (independently of k), so when the radiation arrives at the detector it is spread over an area $A = 4\pi a^2(t_{\text{obs}})r^2$. Therefore

$$\mathcal{F} = \frac{\mathcal{L}}{4\pi a^2(t_{\text{obs}})r^2(1+z)^2} \quad (4.159)$$

which means that (setting t_{obs} equal to the present time t_0)

$$d_L = (1+z)a(t_0)r \quad (4.160)$$

For small values of the redshift, we can express d_L as a function of z as follows. We perform a Taylor expansion of $a(t)$ around the present epoch $t = t_0$,

$$\frac{a(t)}{a(t_0)} = 1 + H_0(t - t_0) - \frac{1}{2}q_0 H_0^2(t - t_0)^2 + \dots \quad (4.161)$$

where

$$H_0 \equiv \frac{\dot{a}(t_0)}{a(t_0)} \quad (4.162)$$

is the Hubble constant (or, more appropriately, the present value of the Hubble parameter $H(t) = \dot{a}/a$), while

$$\begin{aligned} q_0 &\equiv -\frac{\ddot{a}(t_0)}{a(t_0)} \frac{1}{H_0^2} \\ &= -\frac{a(t_0)\ddot{a}(t_0)}{\dot{a}^2(t_0)} \end{aligned} \quad (4.163)$$

is called the deceleration parameter.¹⁵ Using $a(t_0)/a(t) = 1+z$, eq. (4.161) can be inverted to give perturbatively $(t - t_0)$ as a function of z , while inserting the expansion (4.161) into eq. (4.146) gives $a(t_0)r$, as an expansion in powers of $(t - t_0)$ and therefore of z . The result is (see, e.g. Kolb and Turner (1990), pages 41-42 for the explicit calculation)

$$\frac{H_0 d_L}{c} = z + \frac{1}{2}(1 - q_0)z^2 + \dots \quad (4.164)$$

The first term of this expansion just gives the Hubble law $z \simeq (H_0/c)d_L$, which states that redshifts are proportional to distances. The term $O(z^2)$ is the correction to the linear law for moderate redshifts. For large redshifts, the Taylor series is no longer appropriate, and the whole expansion¹⁵ This name is somewhat unfortunate, since it suggests that presently the expansion of the Universe is decelerating. Rather, we will see below that the evidence from the observation of type Ia supernovae indicates that presently the expansion of the Universe is accelerating, i.e. $q_0 < 0$.

¹⁶ As we will discuss in Vol. II, Type Ia supernovae are modeled as carbon-oxygen white dwarfs, in a close binary system. They accrete matter from their companion until their mass reaches a critical value $\simeq 1.30M_\odot$, and then they explode as supernovae. Since the mass of the star at the moment of the explosion is always the same, the light curves of Type Ia supernovae are sufficiently similar to each other and, after applying corrections that take into account a relation between the shape of the light curve and the peak luminosity, they can be used as standard candles. history of the Universe is encoded in a function $d_L(z)$. As an example, for a flat Universe ($k = 0$), eq. (4.146) gives

$$\int_{t_{\text{emis}}}^{t_{\text{obs}}} \frac{cdt}{a(t)} = r \quad (4.165)$$

Differentiating the relation $1 + z(t) = a(t_0)/a(t)$ we get

$$\frac{dt}{a(t)} = -\frac{1}{a(t_0)} \frac{dz}{H(z)} \quad (4.166)$$

so eq. (4.165) becomes

$$a(t_0)r = c \int_0^z \frac{dz'}{H(z')} \quad (4.167)$$

where the upper limit in the integral is the redshift z corresponding to the emission time t_{emis} while, in the lower limit of the integral, we used the fact that at the present time $t = t_0$ we have $z = 0$. Then eq. (4.160) gives

$$d_L(z) = c(1+z) \int_0^z \frac{dz'}{H(z')} \quad (4.168)$$

or, taking a derivative with respect to z ,

$$\frac{c}{H(z)} = \frac{d}{dz} \left(\frac{d_L(z)}{1+z} \right) \quad (4.169)$$

Thus, from the knowledge of $d_L(z)$, we can get the Hubble parameter $H(z)$. This shows that the luminosity distance function $d_L(z)$ is a very important quantity, which encodes the whole expansion history of the Universe.

From the definition (4.154) we see that, in order to determine the luminosity distance of a source, we need to know both $\mathcal{F}$ and $\mathcal{L}$. The flux $\mathcal{F}$ is the quantity that is directly measured by the observer. The problem is to know the intrinsic luminosity $\mathcal{L}$ of the source. This is possible if we have a "standard candle", that is, a source whose absolute luminosity is known. The redshift of an electromagnetic source can be determined from the redshift of its spectral lines, and then we get the corresponding value of $d_L(z)$. Up to distances of order 25Mpc, standard candles are provided by Cepheid variable stars while, at larger cosmological distances, a standard candle is provided by Type Ia supernovae.¹⁶ The Hubble constant is usually written as $H_0 = h_0 \times 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, and the most recent determination of h_0 is $h_0 = 0.73 \pm 0.03$, while the value of the deceleration parameter is $q_0 = -0.74 \pm 0.18$.

6.0.91 Propagation of GWs in a FRW Universe

We can now discuss how the waveform produced by a binary inspiral is modified by the propagation across cosmological distances. First of all, we can define a local wave zone, as the region where the distance to the source is sufficiently large so that the gravitational field already has the $1/r$ behavior characteristic of waves, but still sufficiently small, so that the expansion of the Universe is negligible. During the propagation of the GW in the local wave zone, the scale factor $a(t)$ does not change appreciably, so in the local wave zone physical distances can be written as $r_{\text{phys}} = a(t_{\text{emis}}) r$, where r is the comoving distance and t_{emis} is (any) time of emission,¹⁷ so r_{phys} differs from r by just a constant normalization factor. Using eq. (4.29) we see that the GW produced by a binary inspiral, at a distance $r_{\text{phys}} = a(t_{\text{emis}}) r$ in the local wave zone, can be written as

$$h_+(t_s) = h_c(t_s^{\text{ret}}) \frac{1 + \cos^2 \iota}{2} \cos \left[2\pi \int_{t_s}^{t_s^{\text{ret}}} dt'_s f_{\text{gw}}^{(s)}(t'_s) \right] \quad (4.170)$$

$$h_\times(t_s) = h_c(t_s^{\text{ret}}) \cos \iota \sin \left[2\pi \int_{t_s}^{t_s^{\text{ret}}} dt'_s f_{\text{gw}}^{(s)}(t'_s) \right] \quad (4.171)$$

where

$$h_c(t_s^{\text{ret}}) = \frac{4}{a(t_{\text{emis}}) r} \left(\frac{GM_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}^{(s)}(t_s^{\text{ret}})}{c} \right)^{2/3} \quad (4.172)$$

Here time is the time t_s measured by the clock of the source (and t_s^{ret} is the corresponding value of retarded time) and the GW frequency f_{gw} is the one associated to this definition of time, that we denote by $f_{\text{gw}}^{(s)}$. They are related to the quantities measured by the observer which is at a cosmological distance, very far from the local wave zone, by eqs. (4.150) and (4.151). The dependence of $f_{\text{gw}}^{(s)}$ on t_s is given by eq. (4.19), that we rewrite as

$$f_{\text{gw}}^{(s)}(\tau_s) = \frac{1}{\pi} \left(\frac{5}{256} \frac{1}{\tau_s} \right)^{3/8} \left(\frac{GM_c}{c^3} \right)^{-5/8} \quad (4.173)$$

where τ_s is the time to coalescence measured by the source's clock.

To compute how this waveform propagates across cosmological distances to reach the observer, we should use eq. (1.179), that describes the propagation of GWs in a curved space-time, specializing it to the FRW metric. Actually, it is instructive to start from a simpler problem, namely the propagation of a scalar perturbation ϕ in the FRW metric. In this case, the propagation equation is simply $\square\phi = 0$ where, on scalar functions, the curved-space d'Alembertian $\square = D_\mu D^\mu$ can be written as

$$\square = \frac{1}{\sqrt{-g}} \partial_\mu (\sqrt{-g} g^{\mu\nu} \partial_\nu) \quad (4.174)$$

To solve this wave equation in the FRW metric, it is convenient to introduce the conformal time η , from

$$d\eta = \frac{dt}{a(t)} \quad (4.175)$$

¹⁸ More precisely, we are interested in the evolution in a matter-dominated Universe (since we need that stars already formed!). In this case the FRW scale factor evolves as $a(\eta) \sim \eta^2$ and $a''/a = 2/\eta^2$. Instead, in a radiation-dominated Universe, $a(\eta) \sim \eta$ and $a''/a = 0$

¹⁹ Obviously, since $a(\eta)$ evolves appreciably only on a cosmological timescale, once we fix $\eta = t$ at one moment of time in the present epoch, we have $\eta = t$, with exceedingly good accuracy, over the whole time-scale relevant for GW observation at a detector, so we can write $g(\eta - r/c) = g(t - r/c)$ for all these values of time. For the same reason, $a(\eta)$ can be written simply as a constant factor $a(t_0)$. i.e.

$$\eta = \int^t \frac{dt'}{a(t')} \quad (4.176)$$

so the FRW metric reads (limiting ourselves for simplicity to $k = 0$)

$$ds^2 = a^2(\eta) [-c^2 d\eta^2 + dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2] \quad (4.177)$$

We want to know how a solution which decreases as $1/r$ evolves in this space-time. We therefore search for spherically symmetric solutions of the form $\phi(r, \eta) = (1/r)f(r, \eta)$. The equation $\square\phi = 0$ in this metric then becomes

$$\begin{aligned} 0 &= \partial_\mu (\sqrt{-g} g^{\mu\nu} \partial_\nu) \phi \\ &= -\frac{1}{c^2} \partial_\eta [a^2(\eta) r^2 \partial_\eta \phi] + \partial_r [a^2(\eta) r^2 \partial_r \phi] \\ &= \partial_r^2 f - f'' - 2 \frac{a'}{a} f' \end{aligned} \quad (4.178)$$

¹⁷ We do not need to track the change of the scale factor during the observed part of the emission process, so we do not need to be more precise about t_{emis} .

where the prime denotes derivation with respect to $c\eta$, $f' = (1/c)\partial f/\partial\eta$. It is convenient to search for the solution in the form

$$f(r, \eta) = \frac{1}{a(\eta)} g(r, \eta) \quad (4.179)$$

Then $g(r, \eta)$ satisfies the equation

$$\partial_r^2 g - g'' + \frac{a''}{a} g = 0 \quad (4.180)$$

Now observe that $a''/a \sim \eta^{-2}$, for dimensional reasons.¹⁸ Then we see that eq. (4.180) has the approximate solutions

$$g(r, \eta) \simeq e^{\pm i\omega(\eta-r/c)} \quad (4.181)$$

as long as $\omega^2 \gg 1/\eta^2$, since in this case in eq. (4.180) the term $(a''/a)g \sim g/\eta^2$ is negligible with respect to $-g'' = \omega^2 g$, and we are left with a simple wave equation $\partial_r^2 g - g'' \simeq 0$. More generally, any function of the form $g(\eta - r/c)$ is a solution, as long as in Fourier space it contains only frequencies such that $\eta^2 \omega^2 \gg 1$. In conclusion, we have the approximate solution

$$\phi(r, \eta) \simeq \frac{1}{ra(\eta)} g(\eta - r/c) \quad (4.182)$$

We can normalize conformal time so that, at the present epoch, $\eta = t$. Then, the wave observed today at a detector reads¹⁹

$$\phi(r, t) \simeq \frac{1}{ra(t_0)} g(t - r/c) \quad (4.183)$$

Thus, the propagation of a scalar wave through a FRW background is very simple. Compared to the solution in the absence of cosmological expansion, we just need to replace the factor $1/r$ with $1/[ra(t)]$.

Now we can turn to the propagation equation of a tensor perturbation $h_{\mu\nu}$, eq. (1.179). It is in principle straightforward to write down this equation explicitly in the FRW metric (4.177); we then find that, once we disregard all terms $O(1/\eta^2) h_{\mu\nu}$ with respect to the terms $O(\omega^2) h_{\mu\nu}$, we get back the same equation that we discussed for scalar perturbations. Of course, this is not surprising. Simply, the condition $\eta^2 \omega^2 \gg 1$ demands that ω be large with respect to the typical scale of the background space-time and this is nothing but the condition that defines the geometrical optics approximation. In this approximation all massless particles follow null geodesics, independently of their spin, as we saw in Section 1.5.1.

Then we find that, to leading order:

- The two polarizations h_+ and $h_\times$ decouple, that is each one satisfies a wave equation which is independent of the other. This means that the propagation does not introduce a mixing among them. For instance, if we observe the binary edge-on ($\cos \iota = 0$), we see from eqs. (4.170) and (4.171) that in the local wave zone the GW has only the plus polarization; then, even after propagation across a cosmological distance, the wave will still have only the plus polarization. This is as expected, since we already saw in Section 1.5.1 that the polarization tensor of GWs is parallel-propagated along the null geodesics.
- Both h_+ and $h_\times$ satisfy the same equation that we discussed above for a scalar field.

The conclusion is that, after propagation from the source to the detector, the GW amplitude from an inspiraling binary is still given by eqs. (4.170) and (4.171), but with h_c in eq. (4.172) replaced by

$$h_c(t_s^{\text{ret}}) = \frac{4}{a(t_0)r} \left(\frac{GM_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}^{(s)}(t_s^{\text{ret}})}{c} \right)^{2/3} \quad (4.184)$$

as long as the geometrical optics approximation is valid. Today this condition is $2\pi f_{\text{gw}} \gg t_0^{-1}$, where t_0 is the present age of the Universe, and therefore is satisfied with extreme accuracy by all GWs in which we are interested (in fact, by all GWs whose wavelength is smaller than the present Hubble size of the Universe!). More generally, also the condition $2\pi f_{\text{gw}} \gg t_{\text{emis}}^{-1}$ is extremely well satisfied, so the geometrical optics approximation is excellent for the whole propagation of the GW from the source to the detector.

In principle eqs. (4.170) and (4.171), together with eq. (4.184), provide the final result. However, it is convenient to express them in terms of the time t_{obs} and the GW frequency $f_{\text{gw}}^{(\text{obs})}$ measured by the observer, rather than using the time t_s measured by the source and its associated frequency. Using eqs. (4.150) and (4.151), we see that, in eqs. (4.170) and (4.171)

$$\int_{t_s^{\text{ret}}}^{t_s^{\text{ret}}} dt' f_{\text{gw}}^{(s)}(t'_s) = \int_{t_{\text{obs}}^{\text{ret}}}^{t_{\text{obs}}^{\text{ret}}} dt'_{\text{obs}} f_{\text{gw}}^{(\text{obs})}(t'_{\text{obs}}) \quad (4.185)$$

since the redshift in dt cancels the redshift in f . Writing

$$f_{\text{gw}}^{(s)} = (1+z)f_{\text{gw}}^{(\text{obs})} \quad (4.186)$$

²⁰ Since $\tau = t_{\text{coal}}^{\text{ret}} - t^{\text{ret}} = t_{\text{coal}} - t$, we have $\tau = \tau^{\text{ret}}$, and the result can be written just as a function of the observer's time, rather than of the observer's retarded time. eq. (4.184) becomes

$$\begin{aligned} h_c(t_{\text{obs}}^{\text{ret}}) &= \frac{4}{a(t_0)r} (1+z)^{2/3} \left(\frac{GM_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}^{(\text{obs})}(t_{\text{obs}}^{\text{ret}})}{c} \right)^{2/3} \\ &= \frac{4}{d_L(z)} (1+z)^{5/3} \left(\frac{GM_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}^{(\text{obs})}(t_{\text{obs}}^{\text{ret}})}{c} \right)^{2/3}, \end{aligned} \quad (4.187)$$

where in the second line we expressed the result in terms of the luminosity distance, using eq. (4.160). We see that, if we define the quantity

$$\mathcal{M}_c \equiv (1+z)M_c = (1+z)\mu^{3/5}m^{2/5} \quad (4.188)$$

we get

$$h_c(t_{\text{obs}}^{\text{ret}}) = \frac{4}{d_L} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}^{(\text{obs})}(t_{\text{obs}}^{\text{ret}})}{c} \right)^{2/3} \quad (4.189)$$

and therefore the GW amplitude takes the same form as in the absence of cosmological expansion, with the replacements $r \rightarrow d_L$ and $M_c \rightarrow \mathcal{M}_c$. In the general case of non-vanishing redshift, we will reserve the name "chirp mass" for $\mathcal{M}_c$, rather than for M_c . From eqs. (4.151) and (4.173), we find that the dependence of $f_{\text{gw}}^{(\text{obs})}$ on t_{obs} is given by

$$\begin{aligned}
f_{\text{gw}}^{(\text{obs})}(t_{\text{obs}}^{\text{ret}}) &= \frac{1}{1+z} f_{\text{gw}}^{(s)}(t_s^{\text{ret}}) \\
&= \frac{1}{1+z} \frac{1}{\pi} \left(\frac{5}{256} \frac{1+z}{\tau_{\text{obs}}} \right)^{3/8} \left(\frac{GM_c}{c^3} \right)^{-5/8} \\
&= \frac{1}{\pi} \left(\frac{5}{256} \frac{1}{\tau_{\text{obs}}} \right)^{3/8} \left(\frac{G\mathcal{M}_c}{c^3} \right)^{-5/8}
\end{aligned} \tag{4.190}$$

where $\tau_{\text{obs}} = (1+z)\tau_s$ is the time to coalescence measured by the observer's clock.²⁰ Then, even the dependence of $f_{\text{gw}}^{(\text{obs})}$ on t_{obs} is formally the same as the relation at $z = 0$, once we replace $M_c \rightarrow \mathcal{M}_c$. This result is due to the fact that, when $z = 0$, both the GW amplitude and the evolution of the frequency with time are determined by the only time-scale of the problem, which is GM_c/c^3 . What we are actually finding is that, in a cosmological context, this time-scale is redshifted, $GM_c/c^3 \rightarrow (1+z)GM_c/c^3$. This is a very natural result, which formally amounts to the replacement $M_c \rightarrow (1+z)M_c$.

To summarize, the signal received by the observer from a binary inspiral at redshift z , when expressed in terms of the observer time t_{obs} , or equivalently in terms of the time to coalescence measured by the observer, τ_{obs} , is given by

$$h_+(\tau_{\text{obs}}) = h_c(\tau_{\text{obs}}) \frac{1 + \cos^2 \iota}{2} \cos[\Phi(\tau_{\text{obs}})] \tag{4.191}$$

$$h_\times(\tau_{\text{obs}}) = h_c(\tau_{\text{obs}}) \cos \iota \sin[\Phi(\tau_{\text{obs}})] \tag{4.192}$$

where (compare with eq. (4.30)),

$$\begin{aligned}
\Phi(\tau_{\text{obs}}) &= -2 \left(\frac{5G\mathcal{M}_c(z)}{c^3} \right)^{-5/8} \tau_{\text{obs}}^{5/8} + \Phi_0 \\
h_c(\tau_{\text{obs}}) &= \frac{4}{d_L(z)} \left(\frac{G\mathcal{M}_c(z)}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}^{(\text{obs})}(\tau_{\text{obs}})}{c} \right)^{2/3}
\end{aligned} \tag{4.194}$$

and

$$f_{\text{gw}}^{(\text{obs})}(\tau_{\text{obs}}) = \frac{1}{\pi} \left(\frac{5}{256} \frac{1}{\tau_{\text{obs}}} \right)^{3/8} \left(\frac{G\mathcal{M}_c(z)}{c^3} \right)^{-5/8} \tag{4.195}$$

The latter equation implies also that

$$\dot{f}_{\text{gw}}^{(\text{obs})} = \frac{96}{5} \pi^{8/3} \left(\frac{G\mathcal{M}_c(z)}{c^3} \right)^{5/3} [f_{\text{gw}}^{(\text{obs})}]^{11/3} \tag{4.196}$$

In other words, we have the following modifications compared to the case $z = 0$:

- The observed frequency is redshifted with respect to the frequency measured in the source frame, $f^{(\text{obs})} = f^{(s)}/(1+z)$. For instance, we saw below eq. (4.40) that the inspiral of a NS-NS binary system terminates, and the two stars merge, when the intrinsic orbital frequency of the source, in the source frame, is of order 800 Hz. This means that, from the point of view of the observer, the orbital frequency of the source will sweep up to a maximum value of order 800 Hz/(1+z). For example, for a NS-NS coalescence at $z \simeq 2$, the maximum value of the source orbital frequency, in the observer frame, is of order 270 Hz, and the quadrupole radiation emitted in the inspiral phase will be cutoff at twice this value.

- The overall factor $1/r$ in the GW amplitude is replaced by $1/d_L(z)$.
- M_c is replaced by $\mathcal{M}_c = (1+z)M_c$.

A very interesting consequence of the above results is the following. Suppose that we can measure the amplitudes of both polarizations h_+ , $h_\times$, as well as $\dot{f}_{\text{gw}}^{(\text{obs})}$. The amplitude of h_+ is $h_c(1 + \cos^2 \iota)/2$, while the amplitude of $h_\times$ is $h_c \cos \iota$. From their ratio, we can therefore obtain the value of $\cos \iota$, that is, the inclination of the orbit with respect to the line of sight. On the other hand, eq. (4.196) shows that, if we measure the value of $\dot{f}_{\text{gw}}^{(\text{obs})}$ corresponding to a given value of $f_{\text{gw}}^{(\text{obs})}$, we get the chirp mass $\mathcal{M}_c$. Now in the expression for h_+ and $h_\times$ all parameters have been fixed, except $d_L(z)$.²¹ This means that, from the measured value of h_+ (or of $h_\times$) we can now read d_L . If, at the same time, we can measure the redshift z of the source, we have found a gravitational standard candle, and we can use it to measure the Hubble parameter $H(z)$, compare with eq. (4.168). We will discuss in Vol. 2 the rich cosmological information that can be obtained from this type of measurements, as well as the methods proposed to have an associated measure of the redshift (either from optical observations or from GW observations themselves), and the associated experimental uncertainties.

6.0.92 4.2 Radiation from rotating rigid bodies

The production of GWs from a rotating rigid body is of great importance, in particular for application to isolated neutron stars. In Section 4.2.1 we examine the production of GWs in the simplest situation, i.e. the radiation emitted by a rigid body which rotates around one of its principal axes. Then, in Section 4.2.2, we will discuss the more complicated situation where the rotation axis does not coincide with a principal axis, and therefore there is a motion of precession.

Let us first recall, from elementary mechanics (see e.g. Landau and Lifshitz, Vol. I 1976), a few basic notions of kinematics of rigid bodies. A rigid body is characterized by its inertia tensor

$$I^{ij} = \int d^3x \rho(\mathbf{x}) (r^2 \delta^{ij} - x^i x^j) \quad (4.197)$$

where ρ is the mass density. Since any hermitian matrix can be diagonalized by an appropriate rotation, there exists an orthogonal frame where I_{ij} is diagonal. The corresponding axes are called the principal axes of the body, and the eigenvalues I_1, I_2, I_3 are called the principal moments of inertia. We will refer to the frame where I_{ij} is diagonal as the "body frame". Denoting by x'_i the coordinates in the body frame, we have

$$I_1 = \int d^3x' \rho(\mathbf{x}') (x_2'^2 + x_3'^2) \quad (4.198)$$

$$I_2 = \int d^3x' \rho(\mathbf{x}') (x_1'^2 + x_3'^2) \quad (4.199)$$

$$I_3 = \int d^3x' \rho(\mathbf{x}') (x_1'^2 + x_2'^2) \quad (4.200)$$

From these explicit expressions we see that $I_1 + I_2 \geq I_3$. Therefore each principal moment of inertia must be smaller or equal than the sum of the other two. The identity $I_1 + I_2 = I_3$ holds only if $\rho(\mathbf{x}) \sim \delta(x_3)$, that is, for a bidimensional configuration of matter.

²¹ Observe that the ellipticity of the orbit does not enter since, as we discussed in Section 4.1.3, by the time that the stars approach the coalescence stage, angular momentum losses have circularized the orbit to great accuracy. Therefore it is legitimate to use the results for circular orbits, as we have done in this section, and there is no free parameter associated to the ellipticity.

A simple geometry is that of an ellipsoid with semiaxes a, b, c , uniform density, and mass m . In this case, eqs. (4.198)-(4.200) give

$$I_1 = \frac{m}{5} (b^2 + c^2) \quad (4.201)$$

$$I_2 = \frac{m}{5} (a^2 + c^2) \quad (4.202)$$

$$I_3 = \frac{m}{5} (a^2 + b^2) \quad (4.203)$$

If the body rotates with angular velocity $\boldsymbol{\omega}$, its angular momentum is

$$J_i = I_{ij} \omega_j \quad (4.204)$$

We denote by (J'_1, J'_2, J'_3) and $(\omega'_1, \omega'_2, \omega'_3)$ the components of the angular momentum and of the angular velocity, respectively, in the body frame. Then $J'_1 = I_1 \omega'_1$, $J'_2 = I_2 \omega'_2$ and $J'_3 = I_3 \omega'_3$. Observe that the direction

of $\boldsymbol{\omega}$ is different from the direction of $\mathbf{J}$ unless either $I_1 = I_2 = I_3$ (which holds only for a spherical object) or the rotation is around one of the principal axes, e.g. when, $\omega'_1 = \omega'_2 = 0$. The rotational kinetic energy is

$$E_{\text{rot}} = \frac{1}{2} I_{ij} \omega_i \omega_j \quad (4.205)$$

so in the body frame it is given simply by

$$E_{\text{rot}} = \frac{1}{2} (I_1 \omega'^2_1 + I_2 \omega'^2_2 + I_3 \omega'^2_3) \quad (4.206)$$

If we denote by $\hat{\boldsymbol{\omega}}$ the unit vector in the direction of the axis of rotation, so that $\boldsymbol{\omega} = \omega \hat{\boldsymbol{\omega}}$, then $E_{\text{rot}} = (1/2) I \omega^2$, where $I = I_{ij} \hat{\omega}_i \hat{\omega}_j$ is called the moment of inertia about the axis of rotation.

6.0.93 4.2.1 GWs from rotation around a principal axis

We first consider the situation in which the body rotates rigidly about one of its principal axes. We denote by (x'_1, x'_2, x'_3) the coordinates in the body frame. This reference frame, by definition, is attached to the body and rotates with it. We take the x'_3 axis as the axis around which the body rotate, and we denote by ω_{rot} the corresponding angular velocity. We also introduce a fixed reference frame, with coordinates (x_1, x_2, x_3) , oriented so that $x'_3 = x_3$, see Fig. 4.13. In both frames, the origin of the axes coincides with the center-of-mass of the body. The two frames are related by a time-dependent rotation matrix $\mathcal{R}_{ij}$,

$$x'_i = \mathcal{R}_{ij} x_j \quad (4.207)$$

with

$$\mathcal{R}_{ij} = \begin{pmatrix} \cos \omega_{\text{rot}} t & \sin \omega_{\text{rot}} t & 0 \\ -\sin \omega_{\text{rot}} t & \cos \omega_{\text{rot}} t & 0 \\ 0 & 0 & 1 \end{pmatrix}_{ij} \quad (4.208)$$

We denote by $I'_{ij} = \text{diag}(I_1, I_2, I_3)$ the inertia tensor in the (x'_1, x'_2, x'_3) coordinate system, and by I_{ij} its components in the (x_1, x_2, x_3) frame. Thus, I'_{ij} is a constant matrix, while I_{ij} is time-dependent. The fact that the moment of inertia is a tensor implies that

$$\begin{aligned} I'_{ij} &= \mathcal{R}_{ik} \mathcal{R}_{jl} I_{kl} \\ &= (\mathcal{R} I \mathcal{R}^T)_{ij} \end{aligned} \quad (4.209)$$

where $\mathcal{R}^T$ is the transpose matrix, and therefore

$$I = \mathcal{R}^T I' \mathcal{R} \quad (4.210)$$

This gives

$$\begin{aligned} I_{11} &= I_1 \cos^2 \omega_{\text{rot}} t + I_2 \sin^2 \omega_{\text{rot}} t \\ &= \frac{I_1 + I_2}{2} + \frac{I_1 - I_2}{2} \cos 2\omega_{\text{rot}} t \end{aligned} \quad (4.211)$$

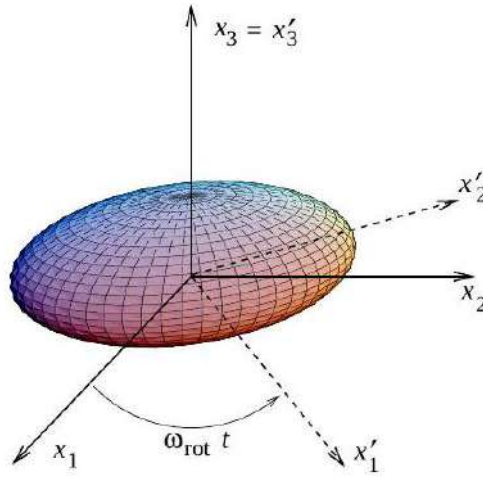


Fig. 4.13 The principal axes (x'_1, x'_2, x'_3) , which rotate with the rigid body, and the fixed axes (x_1, x_2, x_3) .

²² Actually, for a relativistic source (and in the approximation that the internal gravitational fields are not strong) M^{ij} is really the second moment of T^{00}/c^2 , i.e. the energy density over c^2 . For non-relativistic internal motions, T^{00}/c^2 becomes equal to the mass density ρ . Otherwise, all the arguments below go through if, in the definition of I^{ij} , we define ρ to be T^{00}/c^2 .

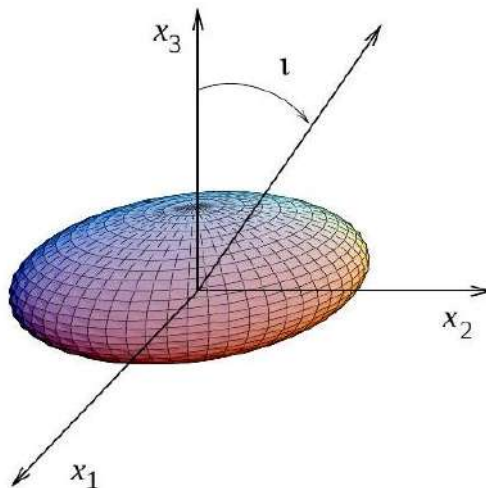


Fig. 4.14 The rigid body rotating around the x_3 axis; ι is the angle between the line-of-sight and the x_3 axis.

$$I_{12} = \frac{I_1 - I_2}{2} \sin 2\omega_{\text{rot}} t \quad (4.212)$$

$$\begin{aligned} I_{22} &= I_1 \sin^2 \omega_{\text{rot}} t + I_2 \cos^2 \omega_{\text{rot}} t \\ &= \frac{I_1 + I_2}{2} - \frac{I_1 - I_2}{2} \cos 2\omega_{\text{rot}} t \end{aligned} \quad (4.213)$$

$$I_{33} = I_3 \quad (4.214)$$

while $I_{13} = I_{23} = 0$. Recall that in the quadrupole approximation the GW amplitude depends on the second moment of the mass density M^{ij} . From eq. (4.197) we see that M^{ij} differ from I^{ij} by an overall minus sign, and for the absence of the trace term.²² However, the trace of a tensor is invariant under rotation; in fact, from eq. (4.209), using the cyclic property of the trace and the property of the orthogonal matrices that $\mathcal{R}\mathcal{R}^T$ is equal to the identity matrix, we have $\text{Tr } I = \text{Tr } [\mathcal{R}^T I' \mathcal{R}] = \text{Tr } [I' \mathcal{R} \mathcal{R}^T] = \text{Tr } I'$. Since $\text{Tr } I' = I_1 + I_2 + I_3$ is a constant, we have

$$M_{ij} = -I_{ij} + c_{ij} \quad (4.215)$$

where c_{ij} are constants. In the computation of the GW amplitude only the second time derivative of M_{ij} enters and therefore the explicit value of the constants is irrelevant, and we can write

$$M_{11} = -\frac{I_1 - I_2}{2} \cos 2\omega_{\text{rot}} t + \text{constant} \quad (4.216)$$

$$M_{12} = -\frac{I_1 - I_2}{2} \sin 2\omega_{\text{rot}} t + \text{constant} \quad (4.217)$$

$$M_{22} = +\frac{I_1 - I_2}{2} \cos 2\omega_{\text{rot}} t + \text{constant} \quad (4.218)$$

while M_{13}, M_{23} and M_{33} are constant. We observe that, in our setting in which the body rotates around its principal axis x'_3 , there is a time-varying second mass moment only if $I_1 \neq I_2$, which is quite clear geometrically. We also see that M_{ij} is a periodic function of $2\omega_{\text{rot}} t$, so we already understand that we have production of GWs with a frequency $\omega_{\text{gw}} = 2\omega_{\text{rot}}$.

We can now compute the GW amplitude received by an observer at a distance r , whose line-of-sight makes an angle ι with the direction of the spin of the star, i.e. with the x_3 axis, see Fig. 4.14. We use eq. (3.72), setting with $\theta = \iota$ and, without loss of generality, we orient the fixed frame (x_1, x_2, x_3) so that the observer is at $\phi = 0$.

Inserting the expressions (4.216)-(4.218) for M_{ij} and taking into account that $\ddot{M}_{13} = \ddot{M}_{23} = \ddot{M}_{33} = 0$, eq. (3.72) gives

$$h_+ = \frac{1}{r} \frac{4G\omega_{\text{rot}}^2}{c^4} (I_1 - I_2) \frac{1 + \cos^2 \iota}{2} \cos(2\omega_{\text{rot}} t) \quad (4.219)$$

$$h_\times = \frac{1}{r} \frac{4G\omega_{\text{rot}}^2}{c^4} (I_1 - I_2) \cos \iota \sin(2\omega_{\text{rot}} t) \quad (4.220)$$

We therefore have a periodic GW, with $\omega_{\text{gw}} = 2\omega_{\text{rot}}$. The fact that h_+ is proportional to $(1 + \cos^2 \iota)/2$, while $h_\times$ to $\cos \iota$, is the same result that we found for the GW amplitude of a binary system in a circular orbit,

see eq. (3.332). In fact, this is a generic property of eq. (3.72) whenever $\ddot{M}_{11} = -\ddot{M}_{22}$ and $\ddot{M}_{13} = \ddot{M}_{23} = \ddot{M}_{33} = 0$.

It is useful to define the ellipticity ϵ by

$$\epsilon \equiv \frac{I_1 - I_2}{I_3} \quad (4.221)$$

For instance, for a homogeneous ellipsoid with semiaxes a, b and c , in the limit of small asymmetry, i.e. $a \simeq b$, eqs. (4.198)-(4.200) give

$$\epsilon \simeq \frac{b - a}{a} + O(\epsilon^2) \quad (4.222)$$

We write the angular velocity of the source as $\omega_{\text{rot}} = 2\pi f_{\text{rot}}$ and we introduce $f_{\text{gw}} = 2f_{\text{rot}}$, which is the frequency of the GW. In terms of f_{gw} and of ϵ , the result for the GW amplitude found above can be rewritten as

$$\begin{aligned} h_+ &= h_0 \frac{1 + \cos^2 \iota}{2} \cos(2\pi f_{\text{gw}} t) \\ h_\times &= h_0 \cos \iota \sin(2\pi f_{\text{gw}} t) \end{aligned} \quad (4.223)$$

where

$$h_0 = \frac{4\pi^2 G}{c^4} \frac{I_3 f_{\text{gw}}^2}{r} \epsilon \quad (4.224)$$

Neutron stars typically have a mass $m \simeq 1.4M_\odot$ and a radius $a \simeq 10$ km, which gives $I_3 \simeq (2/5)ma^2 \simeq 1 \times 10^{38}$ kg m². The value of the ellipticity depends on the neutron star properties, and in particular on the maximum strain that can be supported by its crust. This is quite uncertain but, as we will discuss in Vol. 2, plausible values are in the range $\epsilon \leq 10^{-6}$, although values as large as $\epsilon \simeq 10^{-5}$ can be considered. Inserting these numerical reference values, and taking a typical galactic distance $r = 10$ kpc, eq. (4.224) gives

$$h_0 \simeq 1.06 \times 10^{-25} \left(\frac{\epsilon}{10^{-6}} \right) \left(\frac{I_3}{10^{38} \text{ kg m}^2} \right) \left(\frac{10 \text{ kpc}}{r} \right) \left(\frac{f_{\text{gw}}}{1 \text{ kHz}} \right)^2 \quad (4.225)$$

Observe that neutron stars that rotate more rapidly produce a stronger GW signal, since $h_0 \sim f_{\text{gw}}^2$.

We next compute the power P radiated in GWs, plugging eqs. (4.216)(4.218) into the quadrupole formula (3.76). Observing that $\ddot{M}_{11} = -\ddot{M}_{22}$, this gives

$$\begin{aligned} P &= \frac{2G}{5c^5} \left\langle \ddot{M}_{11}^2 + \ddot{M}_{12}^2 \right\rangle \\ &= \frac{32G}{5c^5} \epsilon^2 I_3^2 \omega_{\text{rot}}^6 \end{aligned} \quad (4.226)$$

and therefore the rotational energy of the star decreases, because of GW emission, as

$$\frac{dE_{\text{rot}}}{dt} = -\frac{32G}{5c^5} \epsilon^2 I_3^2 \omega_{\text{rot}}^6 \quad (4.227)$$

²³ In the literature, (α, β, γ) are usually denote by (θ, ϕ, ψ) , respectively. However, here we prefer to reserve the notation (θ, ϕ) to denote the angles that describe the angular distribution of the gravitational radiation.

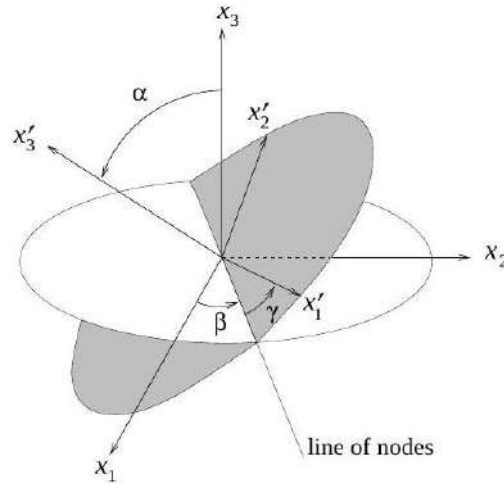


Fig. 4.15 The definition of the Euler angles (α, β, γ) .

Since the rotational energy of a star rotating around its principal axis x_3 is $E_{\text{rot}} = (1/2)I_3\omega_{\text{rot}}^2$ (see eq. (4.206)), if GW emission were the dominant mechanism for the loss of rotational energy, the rotational frequency of a neutron star should decrease as

$$\dot{\omega}_{\text{rot}} = -\frac{32G}{5c^5}\epsilon^2 I_3 \omega_{\text{rot}}^5 \quad (4.228)$$

Experimentally, from the electromagnetic signal emitted by neutron stars observed as pulsars (see Chapter 6), one rather finds $\dot{\omega}_{\text{rot}} \sim -\omega_{\text{rot}}^n$ where n , called the braking index, depends on the specific pulsar but typically ranges between 2 and 3, rather than $n = 5$. This means that GW emission is not the main energy loss mechanism for a rotating pulsar, and other mechanisms, of electromagnetic nature, dominate.

6.0.94 4.2.2 GWs from freely precessing rigid bodies

In general, in astronomical objects the rotation axis does not coincide with a principal axis and, as a consequence, the motion of the rigid body is a combination of rotation around a principal axis and precession. As we will see in this section, the precession motion introduces qualitatively new features in the GWs radiated.

To compute the GW production we proceed as in eqs. (4.207)-(4.209). We first introduce a fixed reference frame, with axes (x_1, x_2, x_3) . In this inertial frame the angular momentum $\mathbf{J}$ of the rigid body is conserved, and we choose the x_3 axis in the direction of $\mathbf{J}$. Next we introduce the body frame, i.e. a reference frame attached to the rotating body, with coordinates (x'_1, x'_2, x'_3) , whose axes coincide with the principal axes of the body. The relation between the two frames is given by the Euler angles (α, β, γ) defined in Fig. 4.15. ²³

To pass from the fixed frame to the body frame we first perform a counterclockwise rotation by an angle β around the x_3 axis. This brings the x_1 axis onto an axis which is called the line of nodes (which is the intersection of the (x_1, x_2) plane with the (x'_1, x'_2) plane). Next we perform a counterclockwise rotation by an angle α around the line of nodes. This brings the x_3 axis onto the x'_3 axis. Finally, we perform a counterclockwise rotation by an angle γ around the x'_3 axis, which brings the line of nodes onto the x'_1 axis. Therefore we have $x'_i = \mathcal{R}_{ij}x_j$, as in eq. (4.207), but now the rotation matrix is more complicated,

$$\mathcal{R} = \begin{pmatrix} \cos \gamma & \sin \gamma & 0 \\ -\sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha \\ 0 & -\sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} \cos \beta & \sin \beta & 0 \\ -\sin \beta & \cos \beta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The motion of the rigid body is specified once we know how α, β and γ evolve with time. Let us recall, from elementary mechanics (see again Landau and Lifshitz, Vol. I, 1976), how

this can be obtained. In the (x_1, x_2, x_3) frame, the angular momentum $\mathbf{J}$ is conserved (this, of course,

is not the case in the body frame, which is non-inertial, since it is rotating), and we oriented the x_3 axis so that $\mathbf{J} = (0, 0, J)$. In the (x'_1, x'_2, x'_3) frame, instead, we denote the components of the angular momentum by (J'_1, J'_2, J'_3) . From Fig. 4.15 we see that

$$\begin{aligned} J'_1 &= J \sin \alpha \sin \gamma \\ J'_2 &= J \sin \alpha \cos \gamma \\ J'_3 &= J \cos \alpha \end{aligned} \quad (4.230)$$

On the other hand, we can write (J'_1, J'_2, J'_3) in terms of $\dot{\alpha}, \dot{\beta}$ and $\dot{\gamma}$ using again Fig. 4.15 and observing that, as a vector, the angular velocity $d\boldsymbol{\alpha}/dt$ is directed along the line of nodes, so its components in the body frame are

$$\frac{d\boldsymbol{\alpha}}{dt} = (\dot{\alpha} \cos \gamma, -\dot{\alpha} \sin \gamma, 0) \quad (4.231)$$

Similarly, $d\boldsymbol{\beta}/dt$ is directed along the x_3 axis, so its components in the body frame are

$$\frac{d\boldsymbol{\beta}}{dt} = (\dot{\beta} \sin \alpha \sin \gamma, \dot{\beta} \sin \alpha \cos \gamma, \dot{\beta} \cos \alpha) \quad (4.232)$$

while $d\gamma/dt$ is along the x'_3 axis, so in the body frame $d\gamma/dt = (0, 0, \dot{\gamma})$. The total angular velocity $\boldsymbol{\omega}$ is the vector sum of these angular velocities, $\boldsymbol{\omega} = d\boldsymbol{\alpha}/dt + d\boldsymbol{\beta}/dt + d\gamma/dt$, so its components ω'_i in the body frame are

$$\begin{aligned} \omega'_1 &= \dot{\alpha} \cos \gamma + \dot{\beta} \sin \alpha \sin \gamma \\ \omega'_2 &= -\dot{\alpha} \sin \gamma + \dot{\beta} \sin \alpha \cos \gamma \\ \omega'_3 &= \dot{\gamma} + \dot{\beta} \cos \alpha \end{aligned} \quad (4.233)$$

In the body frame the inertia tensor is diagonal, with eigenvalues I_1, I_2 and I_3 , so $J'_1 = I_1 \omega'_1, J'_2 = I_2 \omega'_2$ and $J'_3 = I_3 \omega'_3$. Comparing eqs. (4.230) and (4.233) we therefore get

$$I_1(\dot{\alpha} \cos \gamma + \dot{\beta} \sin \alpha \sin \gamma) = J \sin \alpha \sin \gamma \quad (4.234)$$

$$I_2(-\dot{\alpha} \sin \gamma + \dot{\beta} \sin \alpha \cos \gamma) = J \sin \alpha \cos \gamma \quad (4.235)$$

$$I_3(\dot{\gamma} + \dot{\beta} \cos \alpha) = J \cos \alpha \quad (4.236)$$

These are first order equations in the variables (α, β, γ) , and are the first integral of the equations of motion provided by the conservation of angular momentum. ²⁴ One can now integrate these equations and obtain $\alpha(t), \beta(t)$ and $\gamma(t)$. In the most general case, the result can be written in terms of elliptic functions, and we will get back to it later in this section. However, we first limit ourselves to the simpler case of an axisymmetric body with $I_1 = I_2$.

6.0.95 "Wobble" radiation from an axisymmetric rigid body

We consider an axisymmetric body, whose longitudinal axis x'_3 makes an angle α with the angular momentum axis x_3 . The angle α is often ²⁴ They are in fact the first integral of the Euler equations for rigid bodies in the absence of external torques,

$$\begin{aligned}
I_1 \dot{\omega}'_1 &= \omega'_2 \omega'_3 (I_2 - I_3) \\
I_2 \dot{\omega}'_2 &= \omega'_1 \omega'_3 (I_3 - I_1) \\
I_3 \dot{\omega}'_3 &= \omega'_1 \omega'_2 (I_1 - I_2)
\end{aligned}$$

as we easily verify taking the first derivative of eqs. (4.234)-(4.236) and using eq. (4.233).

²⁵ Using the expression for $\dot{\gamma}$ given in eq. (4.240), we have

$$\begin{aligned}
\omega'_3 &= \dot{\gamma} + \dot{\beta} \cos \alpha \\
&= (I_1/I_3) \Omega \cos \alpha
\end{aligned}$$

and therefore we can also write

$$\omega_p = \frac{I_3 - I_1}{I_1} \omega'_3 \quad (4.241)$$

²⁶ Observe that, for an axisymmetric body, the orientation of the principal axes in the (x'_1, x'_2) plane is arbitrary. We choose the origin of time so that $\beta(t=0) = 0$, and then we choose the orientation of the (x'_1, x'_2) axes so that, at $t=0$, $\gamma(0) = \pi/2$. Therefore eq. (4.240) gives $\gamma(t) = (\pi/2) - \omega_p t$. called the "wobble" angle, and the corresponding GW emission is called "wobble radiation".

When $I_1 = I_2$, the analytic solution of eqs. (4.234)-(4.236) is very simple. Multiplying the first equation by $\cos \gamma$ and the second by $\sin \gamma$ and subtracting them, we get

$$\dot{\alpha} = 0 \quad (4.237)$$

This shows that the inclination of the x'_3 axis with respect to the angular momentum $\mathbf{J}$ is constant. Multiplying the first equation by $\sin \gamma$ and the second by $\cos \gamma$ and summing them, we get

$$I_1 \dot{\beta} \sin \alpha = J \sin \alpha \quad (4.238)$$

This shows that, if $\alpha \neq 0$, the x'_3 axis rotates with a constant angular velocity $\dot{\beta} = J/I_1$ about the direction of $\mathbf{J}$. We define

$$\Omega \equiv \dot{\beta} = \frac{J}{I_1} \quad (4.239)$$

Finally, having found that both $\cos \alpha$ and $\dot{\beta}$ are constants, eq. (4.236) shows that even $\dot{\gamma}$ is constant. Equation (4.236), together with $J = I_1 \dot{\beta}$, gives

$$\omega_p \equiv -\dot{\gamma} = \frac{I_3 - I_1}{I_3} \Omega \cos \alpha \quad (4.240)$$

The minus sign in the definition $\omega_p = -\dot{\gamma}$ is chosen so that, for an oblate body ($I_3 > I_1$), which is the normal shape of astrophysical objects, $\omega_p > 0$.²⁵ Using eq. (4.233) we see that the components of the angular velocity in the body frame evolve as²⁶

$$\begin{aligned}
\omega'_1 &= a \cos(\omega_p t) \\
\omega'_2 &= a \sin(\omega_p t) \\
\omega'_3 &= b
\end{aligned} \quad (4.242)$$

where $a = \Omega \sin \alpha$ and $b = -\omega_p + \Omega \cos \alpha$ are constants. This shows that, in the body frame, the angular velocity rotates in the (x'_1, x'_2) plane, i.e. precesses around the x'_3 axis, with angular

velocity ω_p . This precession is counterclockwise if $\omega_p > 0$ (which justifies the notation ω_p for $-\dot{\gamma}$). Observe that $|I_3 - I_1| \ll I_3$ (in a NS, a possible value could be $|I_3 - I_1|/I_3 \sim 10^{-7}$), and therefore $|\omega_p| \ll \Omega$. This motion is called free precession, since it takes place in the absence of external torques, just as a consequence of the deviation of the rigid body from spherical symmetry.

We can now compute how the inertia tensor of the rigid body evolves with time, in the fixed frame with coordinates (x_1, x_2, x_3) . As before we denote by I' the inertia tensor in the body frame, so $I' = \text{diag}(I_1, I_1, I_3)$ is a constant matrix, while we denote by I_{ij} the inertia tensor in the fixed frame. Then, as in eq. (4.210), $I_{ij} = (\mathcal{R}^T I' \mathcal{R})_{ij}$, but now $\mathcal{R}$ is given by eq. (4.229).

The explicit computation is simplified observing that, if we write the rotation matrix (4.229) as $\mathcal{R} = A(\gamma)B(\alpha)C(\beta)$, we see that

$$\begin{aligned} \mathcal{R}^T I' \mathcal{R} &= (C^T B^T A^T) I' (ABC) \\ &= (C^T B^T) I' (BC) \end{aligned} \quad (4.243)$$

because, when $I_1 = I_2$, the matrix A commutes with the matrix I' , and $A^T A = 1$ because it is an orthogonal matrix. Therefore, the angle γ drops out from I_{ij} , and the remaining matrix multiplication gives

$$\begin{aligned} I_{11} &= I_1 (\cos^2 \beta + \cos^2 \alpha \sin^2 \beta) + I_3 \sin^2 \alpha \sin^2 \beta \\ &= \frac{1}{2} (I_1 - I_3) \sin^2 \alpha \cos 2\beta + \text{constant} \\ I_{12} &= \frac{1}{2} (I_1 - I_3) \sin^2 \alpha \sin 2\beta \\ I_{22} &= I_1 (\sin^2 \beta + \cos^2 \alpha \cos^2 \beta) + I_3 \sin^2 \alpha \cos^2 \beta \\ &= -\frac{1}{2} (I_1 - I_3) \sin^2 \alpha \cos 2\beta + \text{constant} \\ I_{13} &= -(I_1 - I_3) \sin \alpha \cos \alpha \sin \beta \\ I_{23} &= (I_1 - I_3) \sin \alpha \cos \alpha \cos \beta \\ I_{33} &= I_1 \sin^2 \alpha + I_3 \cos^2 \alpha = \text{constant} \end{aligned} \quad (4.244)$$

Observe that, since γ dropped out and α is constant, the time dependence of these expressions comes uniquely from β which, from eq. (4.239), is $\beta(t) = \Omega t$ (recall that we have chosen the origin of time so that $\beta = 0$ at $t = 0$, see Note 26). As in eq. (4.215), $M_{ij} = -I_{ij}$, plus constant terms that give zero upon derivation. Therefore, we get

$$\begin{aligned} \ddot{M}_{11} &= 2 (I_1 - I_3) \Omega^2 \sin^2 \alpha \cos(2\Omega t) \\ \ddot{M}_{12} &= 2 (I_1 - I_3) \Omega^2 \sin^2 \alpha \sin(2\Omega t) \\ \ddot{M}_{22} &= -2 (I_1 - I_3) \Omega^2 \sin^2 \alpha \cos(2\Omega t) \\ \ddot{M}_{13} &= -(I_1 - I_3) \Omega^2 \sin \alpha \cos \alpha \sin(\Omega t) \\ \ddot{M}_{23} &= (I_1 - I_3) \Omega^2 \sin \alpha \cos \alpha \cos(\Omega t) \\ \ddot{M}_{33} &= 0 \end{aligned} \quad (4.245)$$

We see that some matrix elements of $\ddot{M}_{ij}$ oscillate as $\sin(2\Omega t)$ or $\cos(2\Omega t)$, while others as $\sin(\Omega t)$ or $\cos(\Omega t)$. From this, we already understand that we will have GW emission at two frequencies, $\omega_{\text{gw}} = 2\Omega$ and $\omega_{\text{gw}} = \Omega$. The origin of these factors is easily traced back: since the

only time dependence comes from $\beta(t)$, and β describes a rotation in the (x_1, x_2) plane, the time-dependent part of I_{ij} has a factor $\sin(\Omega t)$ or $\cos(\Omega t)$ for each index i or j that take the value 1 or 2, while the index 3 produces no further time dependence. Then, as far as the time-dependent part is concerned, we have $I_{11} \sim \sin^2 \Omega t \sim \cos(2\Omega t)$, $I_{22} \sim \cos^2 \Omega t \sim \cos(2\Omega t)$, $I_{12} \sim \sin(\Omega t) \cos(\Omega t) \sim \sin(2\Omega t)$, while $I_{13} \sim \sin(\Omega t)$ and $I_{23} \sim \cos(\Omega t)$. Thus, the fact that, beside the expected radiation at $\omega_{\text{gw}} = 2\Omega$, we also have radiation at $\omega_{\text{gw}} = \Omega$, is a consequence of the fact that the motion of precession produces, together

²⁷ Observe that the overall sign of the amplitude depends on the choice of the origin of time, and on the definition of the axes $(\hat{\mathbf{u}}, \hat{\mathbf{v}})$ in the plane transverse to the propagation direction, with respect to which the polarizations are defined. With our definition, when $\iota = 0$ these axes are $\hat{\mathbf{u}} = \hat{\mathbf{x}}$ and $\hat{\mathbf{v}} = \hat{\mathbf{y}}$, see page 111 and Fig. 3.2. In the literature, this computation has been performed choosing $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ so that, when $\iota = 0$, $\hat{\mathbf{u}} = \hat{\mathbf{y}}$ and $\hat{\mathbf{v}} = -\hat{\mathbf{x}}$. This is related to our convention by a rotation of the $(\hat{\mathbf{u}}, \hat{\mathbf{v}})$ axes in the transverse plane, by an angle $\psi = \pi/2$. Under this rotation $h_+ \rightarrow -h_+$ and $h_\times \rightarrow -h_\times$, see eq. (2.194), so $h'_0 \rightarrow -h'_0$. with a time-varying value of I_{ij} with $i, j = 1, 2$, also a variation in I_{23} and I_{13} .

The computation of the GW amplitude radiated in a direction corresponding to polar angles $(\theta = \iota, \phi = 0)$ is now performed using eq. (3.72). We get

$$\begin{aligned} h_+ &= \frac{1}{r} \frac{G}{c^4} \left[\ddot{M}_{11} - \ddot{M}_{22} \cos^2 \iota + \ddot{M}_{23} \sin 2\iota - \ddot{M}_{33} \sin^2 \iota \right] \\ &= A_{+,1} \cos(\Omega t) + A_{+,2} \cos(2\Omega t) \end{aligned} \quad (4.246)$$

$$\begin{aligned} h_\times &= \frac{2}{r} \frac{G}{c^4} \left[\ddot{M}_{12} \cos \iota - \ddot{M}_{13} \sin \iota \right] \\ &= A_{\times,1} \sin(\Omega t) + A_{\times,2} \sin(2\Omega t), \end{aligned} \quad (4.247)$$

where

$$A_{+,1} = h'_0 \sin 2\alpha \sin \iota \cos \iota \quad (4.248)$$

$$A_{+,2} = 2h'_0 \sin^2 \alpha (1 + \cos^2 \iota) \quad (4.249)$$

$$A_{\times,1} = h'_0 \sin 2\alpha \sin \iota \quad (4.250)$$

$$A_{\times,2} = 4h'_0 \sin^2 \alpha \cos \iota \quad (4.251)$$

and²⁷

$$h'_0 = -\frac{G}{c^4} \frac{(I_3 - I_1) \Omega^2}{r} \quad (4.252)$$

Of course it is understood that, on the right-hand side of eqs. (4.246) and (4.247), t is actually the retarded time. As we already anticipated, we have radiation at both $\omega_{\text{gw}} = \Omega$ and at $\omega_{\text{gw}} = 2\Omega$.

Observe that $A_{+,1}/A_{\times,1} = \cos \iota$. Therefore from this ratio (or, equivalently, from $A_{+,2}/A_{\times,2}$) we get the inclination angle ι (which, by definition, is in the range $0 \leq \iota \leq \pi$). Given ι , the ratio $A_{+,1}/A_{+,2}$ determines the angle α . The overall value of the amplitude then fixes $|h'_0|$. If the distance to the source is known by other means, we therefore get a measure of $|I_1 - I_3|$.

It is now straightforward to compute the total power radiated. Using eq. (3.76) and observing, from eq. (4.245), that $\ddot{M}_{33} = 0$ and $\sum_k \ddot{M}_{kk} = 0$, we get

$$\begin{aligned} P_{\text{quad}} &= \frac{G}{5c^5} \left\langle \ddot{M}_{11}^2 + \ddot{M}_{22}^2 + 2\ddot{M}_{12}^2 + 2\ddot{M}_{13}^2 + 2\ddot{M}_{23}^2 \right\rangle \\ &= \frac{2G}{5c^5} (I_1 - I_3)^2 \Omega^6 \sin^2 \alpha (\cos^2 \alpha + 16 \sin^2 \alpha) \end{aligned} \quad (4.253)$$

where the term $\sin^2 \alpha \cos^2 \alpha$ is from the radiation at $\omega_{\text{gw}} = \Omega$ and the term $16 \sin^4 \alpha$ is from the radiation at $\omega_{\text{gw}} = 2\Omega$. Observe that, in the limit of small α , the power radiated at $\omega_{\text{gw}} = \Omega$ is $O(\alpha^2)$ while the power radiated at $\omega_{\text{gw}} = 2\Omega$ is $O(\alpha^4)$, so for sufficiently small α the radiation at $\omega_{\text{gw}} = \Omega$ is dominant. The two contributions become equal when $\cos^2 \alpha = 16 \sin^2 \alpha$, i.e. for $\alpha \simeq 0.245 \text{ rad} \simeq 14^\circ$.

6.0.96 The back-reaction of GWs

We can now compute how the emission of GWs back-reacts on the motion of the rigid body. The energy radiated in GWs is supplied by the rotational energy E_{rot} of the rigid body, so

$$\frac{dE_{\text{rot}}}{dt} = -\frac{2G}{5c^5} (I_1 - I_3)^2 \Omega^6 \sin^2 \alpha (\cos^2 \alpha + 16 \sin^2 \alpha) \quad (4.254)$$

The angular momentum radiated (in the quadrupole approximation that we are using) is computed from eq. (3.97), and is supplied by the angular momentum $\mathbf{J} = (0, 0, J)$ of the rigid body. Setting $i = 3$ into eq. (3.97), we have

$$\begin{aligned} \frac{dJ}{dt} &= -\frac{2G}{5c^5} \left\langle \ddot{M}_{1a} \ddot{M}_{2a} - \ddot{M}_{2a} \ddot{M}_{1a} \right\rangle \\ &= -\frac{4G}{5c^5} \left\langle \ddot{M}_{1a} \ddot{M}_{2a} \right\rangle \end{aligned} \quad (4.255)$$

where we observed that we can replace Q_{ka} by M_{ka} since $\epsilon^{3kl} \delta_{ka} Q_{la} = \epsilon^{3kl} Q_{kl} = 0$, and similarly $Q_{la} \rightarrow M_{la}$, and in the last line we integrated by part a time derivative, inside the temporal average. Inserting the explicit expressions (4.245), we get

$$\frac{dJ}{dt} = -\frac{2G}{5c^5} (I_1 - I_3)^2 \Omega^5 \sin^2 \alpha (\cos^2 \alpha + 16 \sin^2 \alpha) \quad (4.256)$$

Comparing this result with eq. (4.254), we see that $dE_{\text{rot}}/dt = \Omega dJ/dt$. Since $\Omega = \dot{\beta}$ and $J = I_1 \dot{\beta}$ (see eq. (4.239)), the above equation gives

$$\ddot{\beta} = -\frac{2G}{5c^5} \frac{(I_1 - I_3)^2}{I_1} \dot{\beta}^5 \sin^2 \alpha (\cos^2 \alpha + 16 \sin^2 \alpha) \quad (4.257)$$

The equation governing the evolution of α can instead be obtained writing the rotational kinetic energy as in eq. (4.206), and using $\omega'_1 = J'_1/I_1 = (J/I_1) \sin \alpha \sin \gamma$, $\omega'_2 = J'_2/I_1 = (J/I_1) \sin \alpha \cos \gamma$, and $\omega'_3 = J'_3/I_3 = (J/I_3) \cos \alpha$, see eq. (4.230). This gives

$$E_{\text{rot}} = \frac{J^2}{2} \left(\frac{\sin^2 \alpha}{I_1} + \frac{\cos^2 \alpha}{I_3} \right) \quad (4.258)$$

Taking the time derivative and making use of eq. (4.256) and of $J = I_1 \dot{\beta}$ on the right-hand side, and of eq. (4.254) on the left-hand side, we get

$$\dot{\alpha} = -\frac{2G}{5c^5} \frac{(I_1 - I_3)^2}{I_1} \dot{\beta}^4 \sin \alpha \cos \alpha (\cos^2 \alpha + 16 \sin^2 \alpha) \quad (4.259)$$

Equations (4.257) and (4.259) determine the evolution of the angles α and β . We see that, because of the back-reaction of GWs, both the inclination angle α , and the angular velocity $\dot{\beta}$ decrease. However, using eqs. (4.256) and (4.259), we see that

$$\frac{d}{dt}(J \cos \alpha) = 0 \quad (4.260)$$

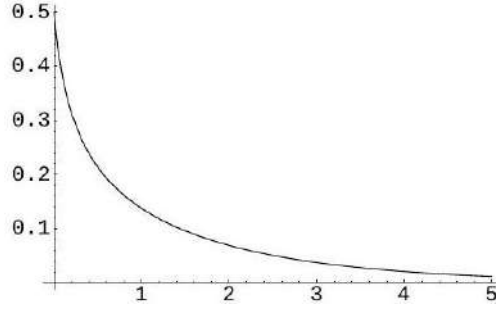


Fig. 4.16 The inclination angle $\alpha(t)$, plotted as a function of t/τ_0 , with initial conditions $\alpha(0) = 0.5, u(0) = 1$.

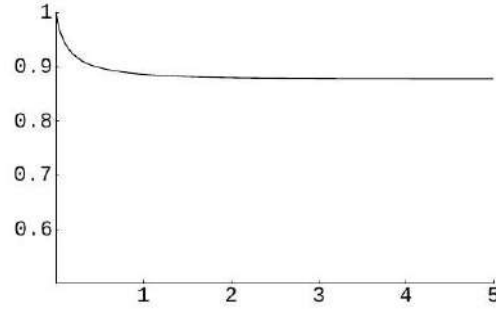


Fig. 4.17 The function $u(t) = \dot{\beta}/\dot{\beta}_0$, plotted against t/τ_0 , with initial conditions $\alpha(0) = 0.5, u(0) = 1$.

Therefore J decreases and α decreases (i.e. $\cos \alpha$ increases) in such a way that $J \cos \alpha$ stays constant. Since $J \cos \alpha$ is the projection of the angular momentum over the x'_3 axis of the body (see eq. (4.230)), we find that the rigid body rotates around its longitudinal axis with a constant angular velocity $\omega'_3 = (J/I_3) \cos \alpha$. The rotation of the body around its longitudinal axis is not affected by the GW back-reaction.

To study in more detail the coupled system of equations (4.257) and (4.259), we introduce the dimensionless function

$$u(t) = \dot{\beta}(t)/\dot{\beta}_0 \quad (4.261)$$

where $\dot{\beta}_0$ is the value of $\dot{\beta}(t)$ at some initial value $t = 0$, and we also introduce the time-scale

$$\begin{aligned} \tau_0 &\equiv \left[\frac{2G}{5c^5} \frac{(I_1 - I_3)^2}{I_1} \dot{\beta}_0^4 \right]^{-1} \\ &\simeq 1.8 \times 10^6 \text{yr} \left(\frac{10^{-7}}{(I_1 - I_3)/I_3} \right)^2 \left(\frac{1 \text{kHz}}{f_0} \right)^4 \left(\frac{10^{38} \text{ kg m}^2}{I_1} \right) \end{aligned} \quad (4.262)$$

where $f_0 = \dot{\beta}_0/(2\pi)$, and we used as reference for $(I_1 - I_3)/I_3$ and for I_3 , values that can be typical for neutron stars. Then eqs. (4.257) and (4.259) become

$$\dot{u} = -\frac{1}{\tau_0} u^5 \sin^2 \alpha (\cos^2 \alpha + 16 \sin^2 \alpha) \quad (4.263)$$

$$\dot{\alpha} = -\frac{1}{\tau_0} u^4 \sin \alpha \cos \alpha (\cos^2 \alpha + 16 \sin^2 \alpha) \quad (4.264)$$

with initial conditions $u(0) = 1$ and $\alpha(0) = \alpha_0$. In this form, it is clear that τ_0 is the characteristic time-scale of the problem, and the equations could be recast in a completely dimensionless form introducing a dimensionless variable $x = t/\tau_0$, so that $\tau_0 \dot{u} = du/dx$ and $\tau_0 \dot{\alpha} = d\alpha/dx$.

The asymptotic behavior of the solution can be easily read from eqs. (4.263) and (4.264), making use also of the fact that, as we shown above, $J \cos \alpha$ is a constant of motion. Since $J = I_1 \dot{\beta}$, this means that $\dot{\beta}_\infty \cos \alpha_\infty = \dot{\beta}_0 \cos \alpha_0$, where the subscripts ∞ denotes the value at $t = \infty$. In terms of u , this means

$$u_\infty \cos \alpha_\infty = \cos \alpha_0 \quad (4.265)$$

Equating to zero the right-hand sides of eqs. (4.263) and (4.264), we see that the possible fixed points of the evolution are either $u = 0$ or $\alpha = 0, \pi$. However, the value $u_\infty = 0$ is not consistent with eq. (4.265) (unless $\alpha_0 = \pi/2$; clearly, this is an unstable fixed point, in which an infinitesimal perturbation drives the evolution either toward $\alpha = 0$ or $\alpha = \pi$). Therefore the only asymptotic solution, for $\alpha_0 < \pi/2$, has $\alpha_\infty = 0$. Then, eq. (4.265) shows that $u_\infty = \cos \alpha_0$. Once we realize that u approaches a finite constant u_∞ and that $\alpha \rightarrow 0$, from eq. (4.264) we see that at large times $\dot{\alpha} \simeq -\alpha/\tau_\infty$, and therefore $\alpha(t)$ approaches zero exponentially, with a time-scale $\tau_\infty = \tau_0/u_\infty^4$, which is obtained simply replacing $\dot{\beta}_0$ with $\dot{\beta}_\infty$ in eq. (4.262).

The numerical integration of eqs. (4.263) and (4.264) is straightforward, and we show in Figs. 4.16 and 4.17 the solution for the initial condition $\alpha_0 = 0.5$, which confirms the asymptotic behavior found analytically.

In conclusion, on a time-scale given by eq. (4.262), the rigid body aligns its rotation axis with the direction of the angular momentum ($\alpha \rightarrow 0$), while the rotational angular velocity $\dot{\beta} = \Omega$ around the x_3 axis decreases toward the constant value $\Omega_0 \cos \alpha_0$, and the rotational velocity around its principal axis $\omega'_3 = \Omega \cos \alpha$ stays constant.

Finally, it is interesting to compare the frequency ω_{em} of the electromagnetic pulses of a pulsar, idealized as a rotating rigid body, ²⁸ with the frequency of the GWs that it emits. If a pulsar rotates around its principal axis with frequency Ω , and has $I_1 \neq I_2$, which is the case discussed in Section 4.2.1, then $\omega_{\text{em}} = \Omega$, so we can determine Ω very precisely by electromagnetic observations, and GW emission is at $\omega_{\text{gw}} = 2\Omega$. In this case, therefore, we have an accurate prediction of the frequency at which we expect GW emission.

Instead, in the case of "wobble" radiation discussed in this subsection, in the limit of small α , the axes x_3 and x'_3 almost coincide. If the source of the electromagnetic radiation is some "hot spot" fixed on the star surface then, for a far observer, the hot spot is basically rotating around the x_3 axis with a total angular velocity $\omega_{\text{rot}} = \dot{\beta} + \dot{\gamma} = \Omega - \omega_p$, see Fig. 4.15, so

$$\Omega = \omega_{\text{rot}} + \omega_p \quad (4.266)$$

The mean frequency ω_{em} of repetition of the electromagnetic pulses will be equal to ω_{rot} , so $\omega_{\text{em}} = \Omega - \omega_p$. Instead, we saw that wobble gravitational radiation is emitted at $\omega_{\text{gw}} = \Omega$ and at $\omega_{\text{gw}} = 2\Omega$, so these two lines are shifted, with respect to ω_{em} and to $2\omega_{\text{em}}$, by the unknown quantities ω_p and $2\omega_p$, respectively. Using eq. (4.240), for the line at $\omega_{\text{gw}} = \Omega$, we have

$$\omega_{\text{gw}} \simeq \left(1 + \frac{I_3 - I_1}{I_3} \cos \alpha\right) \omega_{\text{em}} \quad (4.267)$$

Since $|I_1 - I_3|/I_3$ is in general very small, e.g. could be of order 10^{-7} in a neutron star, the difference between the actual GW frequency and the value suggested by electromagnetic observation is small. An axisymmetric astrophysical object is normally expected to be oblate ($I_3 > I_1$), and in this case $\omega_{\text{gw}} > \omega_{\text{em}}$. For a prolate body ($I_3 < I_1$), instead, we have $\omega_{\text{gw}} < \omega_{\text{em}}$.

6.0.97 Rotating and precessing triaxial bodies

We now study the radiation emitted by a generic triaxial body. We choose the axes of the body frame so that $I_1 < I_2 < I_3$. The timedependent moment of inertia I_{ij} in the fixed frame is again computed²⁹ We assume for definiteness $J^2 > 2EI_2$; in this case, in the body frame, the apparent precession of $\mathbf{J}$ is around the x'_3 axis; if instead $J^2 < 2EI_2$, it is around the x'_1 axis.

³⁰ For the elliptic functions cn , sn and dn and for the Jacobi theta functions that will appear below, we follow the notation and conventions of Abramowitz and Stegun (1972). from $I_{ij} = (\mathcal{R}^T I' \mathcal{R})_{ij}$, with $\mathcal{R}$ given by eq. (4.229), but now the dependence on γ does not cancel, contrary to the case $I_1 = I_2$ discussed above, compare with eq. (4.243).

Once expressed I_{ij} in terms of α, β, γ , to compute the waveform we need the time-dependence of the Euler angles, which can be found for instance in Landau and Lifshitz, Vol. I (1976). The evolution of the components of the angular velocity in the body frame is²⁹

$$\begin{aligned}\omega'_1 &= a \text{cn}(\tau, m) \\ \omega'_2 &= a \left[\frac{I_1 (I_3 - I_1)}{I_2 (I_3 - I_2)} \right]^{1/2} \text{sn}(\tau, m) \\ \omega'_3 &= b \text{dn}(\tau, m)\end{aligned}\tag{4.268}$$

where cn , sn and dn are elliptic functions,³⁰ and τ is a rescaled time variable,

$$\tau = bt \left[\frac{(I_3 - I_2)(I_3 - I_1)}{I_1 I_2} \right]^{1/2}\tag{4.269}$$

The parameter m of the elliptic functions is given by

$$m = \frac{(I_2 - I_1) I_1 a^2}{(I_3 - I_2) I_3 b^2}\tag{4.270}$$

and a, b are the initial values of ω'_1 and ω'_3 ,

$$a = \omega'_1(0), \quad b = \omega'_3(0)\tag{4.271}$$

Observe that we have chosen the origin of time so that $\omega'_2(0) = 0$. In the limit $I_1 = I_2$ we have $m = 0$; since $\text{cn}(\tau, 0) = \cos \tau$, $\text{sn}(\tau, 0) = \sin \tau$, and $\text{dn}(\tau, 0) = 1$, in this limit we recover the result (4.242). The elliptic functions are periodic functions of τ , with periodicity $4K(m)$, where $K(m)$ is the complete elliptic integral of the first kind. The components of the angular velocity in the body frame are therefore periodic in t , with period

$$T = \frac{4K(m)}{b} \left[\frac{I_1 I_2}{(I_3 - I_2)(I_3 - I_1)} \right]^{1/2}\tag{4.272}$$

To compute I_{ij} , we need explicitly the time-dependence of the Euler angles. The solution for α and γ is

$$\cos \alpha(t) = \frac{I_3 b}{J} \text{dn}(\tau, m)\tag{4.273}$$

$$\tan \gamma(t) = \left[\frac{I_1 (I_3 - I_2)}{I_2 (I_3 - I_1)} \right]^{1/2} \frac{\text{cn}(\tau, m)}{\text{sn}(\tau, m)}\tag{4.274}$$

²⁸ The discussion below is really specific to our crude modelization of neutron stars as rigid bodies. Real neutron stars are rather mostly fluid in their interior. See the Further Reading section for the results with a more realistic modelization of neutron stars.

so they are both periodic with period T . The solution for β is instead

$$\beta(t) = \frac{2\pi}{T'}t + \beta_1(t) \quad (4.275)$$

where

$$\begin{aligned} \frac{2\pi}{T'} &= \frac{J}{I_1} - \frac{2i}{T} \frac{\vartheta'_4(i\pi c_1)}{\vartheta_4(i\pi c_1)} \\ &= \frac{J}{I_1} + \frac{2b}{K(m)} \left[\frac{(I_3 - I_2)(I_3 - I_1)}{I_1 I_2} \right]^{1/2} \sum_{n=1}^{\infty} \frac{q^n}{1 - q^{2n}} \sinh(2\pi n c_1) \end{aligned} \quad (4.276)$$

Here ϑ_4 is a Jacobi theta function, c_1 is defined to be any ³¹ solution of the equation $\text{sn}[2ic_1 K(m), m] = iI_3 b / (I_1 a)$, and

$$q = \exp\{-\pi K(1 - m)/K(m)\} \quad (4.277)$$

The function $\beta_1(t)$ can be written as a ratio of theta functions,

$$\beta_1(t) = \frac{\vartheta_4\left(\frac{2\pi t}{T} - i\pi c_1\right)}{\vartheta_4\left(\frac{2\pi t}{T} + i\pi c_1\right)} \quad (4.278)$$

and admits the series expansion

$$\beta_1(t) = \sum_{n=1}^{\infty} \left[\frac{-2q^n}{n(1 - q^{2n})} \right] \sin\left(\frac{4\pi n t}{T}\right) \quad (4.279)$$

Observe that $\beta_1(t)$ is periodic with period $T/2$. In the problem there are therefore two distinct periodicities, T and T' , and we can expect that the frequency spectrum of the gravitational radiation will have two sets of lines. The fact that γ is periodic with period T identifies $2\pi/T$ as the angular frequency associated to precession, so we define

$$\begin{aligned} \omega_p &= \frac{2\pi}{T} \\ &= \frac{\pi b}{2K(m)} \left[\frac{(I_3 - I_2)(I_3 - I_1)}{I_1 I_2} \right]^{1/2} \end{aligned} \quad (4.280)$$

The angular frequency associated to rotation around the x_3 axis is instead ³²

$$\omega_{\text{rot}} = \frac{2\pi}{T'} - \frac{2\pi}{T} \quad (4.281)$$

Given these complicated but explicit solutions for $\alpha(t)$, $\beta(t)$ and $\gamma(t)$, one could in principle plug them into the rotation matrix $\mathcal{R}_{ij}$ and grind, to obtain the moment of inertia in the fixed frame, and hence the waveform. However, the discrete nature of the spectrum becomes more evident if, rather than working with the exact expressions, we perform a series expansion for small wobble angle (which is equivalent to expanding in powers of m) and small deviations from axisymmetry. ³³ The result, up to second order in m , is

$$h_+ = A_{+,0} \cos(2\omega_{\text{rot}} t) \quad (4.282)$$

$$+ A_{+,1} \cos[(\omega_{\text{rot}} + \omega_p) t] + A_{+,2} \cos[2(\omega_{\text{rot}} + \omega_p) t]$$

$$h_{\times} = A_{\times,0} \sin(2\omega_{\text{rot}} t) \quad (4.283)$$

$$+ A_{\times,1} \sin[(\omega_{\text{rot}} + \omega_p) t] + A_{\times,2} \sin[2(\omega_{\text{rot}} + \omega_p) t]$$

³¹ Since the elliptic functions and the theta functions have the same periodicity, all solutions for c_1 are equivalent. ³² One can also check this result observing that, in the limit of small wobble angle, it gives $\omega_{\text{rot}} = \dot{\beta} - \omega_{\text{p}}$. Since $\dot{\gamma} = -\omega_{\text{p}}$, see eq. (4.240), we get $\omega_{\text{rot}} = \dot{\beta} + \dot{\gamma}$, in agreement with the discussion above eq. (4.266).

³³ This computation is performed in Zimmermann (1980) and Van Den Broeck (2005), see the Further Reading.

³⁴ Observe that in Section 4.2.1 we choose the origin of time so that, at $t = 0$, the axis x'_1 is aligned with the axis x_1 , i.e. $\gamma(0) = 0$, while here we have chosen $\gamma(0) = \pi/2$. This is the origin of the overall minus sign between eq. (4.224) and eq. (4.285). As discussed in Note 27, we can reabsorb the minus sign in both h_0 and h'_0 rotating by an angle $\pi/2$ in the transverse plane the axes $(\hat{\mathbf{u}}, \hat{\mathbf{v}})$ used to define the plus and cross polarizations. where

$$\begin{aligned} A_{+,0} &= h_0(1/2) (1 + \cos^2 \iota) \\ A_{+,1} &= 2h'_0 g(\alpha_0) \sin \iota \cos \iota \\ A_{+,2} &= 2h'_0 g^2(\alpha_0) (1 + \cos^2 \iota) \\ A_{\times,0} &= h_0 \cos \iota \\ A_{\times,1} &= 2h'_0 g(\alpha_0) \sin \iota \\ A_{\times,2} &= 4h'_0 g^2(\alpha_0) \cos \iota \end{aligned} \tag{4.284}$$

The amplitudes h_0 and h'_0 are defined as

$$h_0 = -\frac{1}{r} \frac{4G\omega_{\text{rot}}^2}{c^4} (I_1 - I_2) \tag{4.285}$$

$$h'_0 = -\frac{1}{r} \frac{G(\omega_{\text{rot}} + \omega_{\text{p}})^2}{c^4} \left(I_3 - \frac{I_1 + I_2}{2} \right) \tag{4.286}$$

and

$$g(\alpha_0) \equiv \frac{I_1 a}{I_3 b} \tag{4.287}$$

Using eqs. (4.234) and (4.236), and recalling that we have chosen the origin of time so that $\gamma(0) = \pi/2$, we see that $I_1 a / (I_3 b) = \tan \alpha_0$, where $\alpha_0 = \alpha(t = 0)$. Therefore, in the limit of small wobble angle α_0 , we have $g(\alpha_0) \simeq \alpha_0$.

From these expressions we see that, setting $\alpha_0 = 0$ (and therefore $g(\alpha_0) = 0$) we recover the correct limit of rotation around a principal axis, given in eqs. (4.219), (4.220) and (4.224).
³⁴

Instead, setting $I_1 = I_2$ (and recalling, from eq. (4.281) and Note 32, that $\Omega = \omega_{\text{rot}} + \omega_{\text{p}}$), we recover the correct limit of precession of an axisymmetric body when α is small, given in eqs. (4.246)-(4.252)

We see that, to zeroth-order in $g(\alpha_0)$, we find the line at $\omega_{\text{gw}} = 2\omega_{\text{rot}}$ due to rotation around a principal axis. To order $g(\alpha_0)$ emerges a line at $\omega_{\text{gw}} = \omega_{\text{rot}} + \omega_{\text{p}}$, and to order $g^2(\alpha_0)$ we get a further line at $\omega_{\text{gw}} = 2\omega_{\text{rot}} + 2\omega_{\text{p}}$. If we expand to higher order in $g(\alpha_0)$, we find two sets of lines at frequencies, respectively,

$$\omega_{\text{gw}} = 2\omega_{\text{rot}} + 2k\omega_{\text{p}} \tag{4.288}$$

and

$$\omega_{\text{gw}} = \omega_{\text{rot}} + (2k + 1)\omega_{\text{p}} \quad (4.289)$$

with $k = 0, 1, 2, \dots$, whose intensity is suppressed by increasing powers of $g(\alpha_0)$.

To conclude this section, we should stress that all results discussed above have been obtained for rigid bodies. For application to neutron stars, a more realistic modeling of the neutron star interior (e.g. a fluid body with an elastic crust) is necessary. We will come back to realistic neutron stars in Vol. 2.

Part III

Detection of GW and Noises

Обсудим подробно различные способы детектирования гравитационных волн.

7 Известные методы

7.1 Gravitational Wave Detectors and Their Sensitivity by Sathyaprakash

Detectors of gravitational waves generally divide into two classes: beam detectors and resonant mass detectors. In beam detectors, gravitational waves interact with a beam of electromagnetic radiation, which is monitored in some way to register the passage of the wave. In resonant mass detectors, the gravitational wave transfers energy to a massive body, from which the resultant oscillations are observed.

Both classes include a variety of systems. The principal beam detectors are the large groundbased laser interferometers currently operating in several locations around the globe, such as the LIGO system in the USA. The ESA-NASA LISA mission aims to put a laser interferometer into space to detect milliHertz gravitational waves. But beam detectors do not need to involve interferometry: the radio beams transponded to interplanetary spacecraft can carry the signature of a passing gravitational wave, and this method has been used to search for low-frequency gravitational waves. And radio astronomers have for many years monitored the radio beams of distant pulsars for evidence of gravitational waves; new radio instrumentation is turning this into a powerful and promising method of looking for stochastic backgrounds and individual sources. And at ultra-low frequencies, gravitational waves in the early universe may have left their imprint on the polarization of the cosmic microwave background.

Resonant mass detectors were the first kind of detector built in the laboratory to detect gravitational waves: Joseph Weber [387] built two cylindrical aluminum bar detectors and attempted to find correlated disturbances that might have been caused by a passing impulsive gravitational wave. His claimed detections led to the construction of many other bar detectors of comparable or better sensitivity, which never verified his claims. Some of those detectors were not developed further, but others had their sensitivities improved by making them cryogenic, and today there are two ultra-cryogenic detectors in operation (see Section 4.1).

In the following, we will examine the principal detection methods that hold promise today and in the near future.

7.1.1 Principles of the operation of resonant mass detectors

A typical "bar" detector consists of a cylinder of aluminum with a length $\ell \sim 3$ m, a very narrow resonant frequency between $f \sim 500$ Hz and 1.5kHz, and a mass $M \sim 1000$ kg. A short gravitational wave burst with $h \sim 10^{-21}$ will make the bar vibrate with an amplitude

$$\delta\ell_{gw} \sim h\ell \sim 10^{-21} \text{ m}$$

To measure this, one must fight against three main sources of noise.

1. Thermal noise. The original Weber bar operated at room temperature, but the most advanced detectors today, Nautilus [51] and Auriga [227], are ultra-cryogenic, operating at $T = 100\text{mK}$. At this temperature the root mean square (rms) amplitude of vibration is

$$\langle \delta \ell^2 \rangle_{th}^{1/2} = \left(\frac{kT}{4\pi^2 M f^2} \right)^{1/2} \sim 6 \times 10^{-18} \text{ m}$$

This is far larger than the gravitational wave amplitude expected from astrophysical sources. But if the material has a high Q (say, 10^6) in its fundamental mode, then that changes its thermal amplitude of vibration in a random walk with very small steps, taking a time $Q/f \sim 1000$ s to change by the full amount. However, a gravitational wave burst will cause a change in 1 ms. In 1 ms, thermal noise will have random-walked to an expected amplitude change $(1000 \text{ s}/1 \text{ ms})^{1/2} = Q^{1/2}$ times smaller, or (for these numbers)

$$\langle \delta \ell^2 \rangle_{th:1 \text{ ms}}^{1/2} = \left(\frac{kT}{4\pi^2 M f^2 Q} \right)^{1/2} \sim 6 \times 10^{-21} \text{ m}$$

So ultra-cryogenic bars can approach the goal of detection near $h = 10^{-20}$ despite thermal noise.

2. Sensor noise. A transducer converts the bar's mechanical energy into electrical energy, and an amplifier increases the electrical signal to record it. If sensing of the vibration could be done perfectly, then the detector would be broadband: both thermal impulses and gravitational wave forces are mechanical forces, and the ratio of their induced vibrations would be the same at all frequencies for a given applied impulsive force.

But sensing is not perfect: amplifiers introduce noise, and this makes small amplitudes harder to measure. The amplitudes of vibration are largest in the resonance band near f , so amplifier noise limits the detector sensitivity to gravitational wave frequencies near f . But if the noise is small, then the measurement bandwidth about f can be much larger than the resonant bandwidth f/Q . Typical measurement bandwidths are 10 Hz, about 10^4 times larger than the resonant bandwidths, and 100 Hz is not out of the question [59].

3. Quantum noise. The zero-point vibrations of a bar with a frequency of 1kHz are

$$\langle \delta \ell^2 \rangle_{\text{quant}}^{1/2} = \left(\frac{\hbar}{2\pi M f} \right)^{1/2} \sim 4 \times 10^{-21} \text{ m}$$

This is comparable to the thermal limit over 1 ms. So, as detectors improve their thermal limits, they run into the quantum limit, which must be breached before a signal at 10^{-21} can be seen with such a detector.

It is not impossible to do better than the quantum limit. The uncertainty principle only sets the limit above if a measurement tries to determine the excitation energy of the bar, or equivalently the phonon number. But one is not interested in the phonon number, except in so far as it allows one to determine the original gravitational wave amplitude. It is possible to define other observables that also respond to the gravitational wave and can be measured more accurately by squeezing their uncertainty at the expense of greater errors in their conjugate observable [110]. It is not yet clear whether squeezing will be viable for bar detectors, although squeezing is now an established technique in quantum optics and will soon be implemented in interferometric detectors (see below).

Reliable gravitational wave detection, whether with bars or with other detectors, requires coincidence observations, in which two or more detectors confirm each other's findings. The principal bar detector projects around the world formed the International Gravitational Event Collaboration (IGEC) [202] to arrange for long-duration coordinated observations and joint data analysis. A report in 2003 of an analysis of a long period of coincident observing over three

years found no evidence of significant events [50]. The ALLEGRO bar [243] at Louisiana State University made joint data-taking runs with the nearby LIGO interferometer, setting an upper limit on the stochastic gravitational-wave background at around 900 Hz of $h_{100}^2 \Omega_{gw}(f) \leq 0.53$ [17]. More recently, because funding for many of the bar detector projects has become more restricted, only two groups continue to operate bars at present (end of 2008): the Rome [367] and Auriga [227] groups. The latest observational results from IGEC may be found in [54].

It is clear from the above discussion that bars have great difficulty achieving the sensitivity goal of 10^{-21} . This limitation was apparent even in the 1970 s, and that motivated a number of groups to explore the intrinsically wide-band technique of laser interferometry, leading to the projects described in Section 4.3.1 below. However, the excellent sensitivity of resonant detectors within their narrow bandwidths makes them suitable for specialized, high-frequency searches, including cross-correlation searches for stochastic backgrounds [119]. Therefore, novel and imaginative designs for resonant-mass detectors continue to be proposed. For example, it is possible to construct large spheres of a similar size (1 to 3 m diameter) to existing cylinders. This increases the mass of the detector and also improves its direction-sensing. One can in principle push to below 10^{-21} with spheres [117]. A spherical prototype called MiniGRAIL [234] has been operated in the Netherlands [181]. A similar prototype called the Schenberg detector [203] is being built in Brazil [21]. Nested cylinders or spheres, or masses designed to sense multiple modes of vibration may also provide a clever way to improve on bar sensitivities [86].

While these ideas have interesting potential, funding for them is at present (2008) very restricted, and the two remaining bar detectors are likely to be shut down in the near future, when the interferometers begin operating at sensitivities clearly better than 10^{-21} .

7.1.2 Principles of the operation of beam detectors

Interferometers use laser light to measure changes in the difference between the lengths of two perpendicular (or nearly perpendicular) arms. Typically, the arm lengths respond differently to a given gravitational wave, so an interferometer is a natural instrument to measure gravitational waves. But other detectors also use electromagnetic radiation, for example, ranging to spacecraft in the solar system and even pulsar timing.

The basic equation we need is for the effect of a plane linear gravitational wave on a beam of light. Suppose the angle between the direction of the beam and the direction of the plane wave is θ . We imagine a very simple experiment in which the light beam originates at a clock, whose proper time is called t , and is received by a clock, whose proper time is t_f . The beam and gravitational-wave travel directions determine a plane, and we denote the polarization component of the gravitational wave that acts in this plane by $h_+(t)$, as measured at the location of the originating clock. The proper distance between the clocks, in the absence of the wave, is L . If the originating clock puts timestamps onto the light beam, then the receiving clock can measure the rate of arrival of the timestamps. If there is no gravitational wave, and if the clocks are ideal, then the rate will be constant, which can be normalized to unity. The effect of the gravitational wave is to change the arrival rate as a function of the emission rate by

$$\frac{dt_f}{dt} = 1 + \frac{1}{2}(1 + \cos \theta) \{h_+[t + (1 - \cos \theta)L] - h_+(t)\}$$

This is very simple: the beam of light leaves the emitter at the time when the gravitational wave of phase t passes the emitter, and it reaches the receiver at the time when the gravitational wave of phase $t + (1 - \cos \theta)L$ is passing the receiver. So in the plane wave case, only the amplitudes of the wave at the emitting and receiving events affect the time delay.

In order to use such an arrangement to detect gravitational waves, one needs two very stable clocks. The best clocks today are stable to a few parts in 10^{16} [40], which implies

that the minimum amplitude of gravitational waves that could be detected by such a two-clock experiment is $h \sim 10^{-15}$. However, this equation is also fundamental to the detection of gravitational waves by pulsar timing, in which the originating 'clock' is a pulsar. By correlating many pulsar signals, one can beat down the single-pulsar noise. This is described below in Section 4.4.2.

An arrangement that uses only one clock is one that sends a beam out to a receiver, which then reflects or retransmits (transponds) the beam back to the sender. The sender has the clock, which measures variations in the round-trip time. This method has been used with interplanetary spacecraft, which has the advantage that the only clock is on the ground, which can be made more stable than one carried in a spacecraft (see Section 4.4.1). For the same arrangement as above, the return time t_{return} varies at the rate

$$\begin{aligned} \frac{dt_{\text{return}}}{dt} = 1 + \frac{1}{2} \{ & (1 - \cos \theta)h_+(t + 2L) - (1 + \cos \theta)h_+(t) \\ & + 2 \cos \theta h_+[t + L(1 - \cos \theta)] \} \end{aligned}$$

This is known as the three-term relation, the third term being the wave strength at the time the beam returns back to the sender.

But the sensitivity of such a one-path system as a gravitational wave detector is still limited by the stability of the clock. For that reason, interferometers have become the most sensitive beam detectors: effectively one arm of the interferometer becomes the 'clock', or at least the time standard, that variations in the other arm are compared to. Of course, if both arms are affected by a gravitational wave in the same way, then the interferometer will not see the wave. But this happens only in very special geometries. For most wave arrival directions and polarizations, the arms are affected differently, and a simple interferometer measures the difference between the round-trip travel time variations in the two arms. For the triangular space array LISA, the measured signal is somewhat more complex (see Section 4.4.3 below), but it still preserves the principle that the time reference for one arm is a combination of the others.

The response of a ground-based interferometer

Ground-based interferometers are the most sensitive detectors operating today, and are likely to make the first direct detections [197]. The largest detectors operating today are the LIGO detectors [302], two of which have arm lengths of 4 km. This is much smaller than the wavelength of the gravitational wave, so the interaction of one arm with a gravitational wave can be well approximated by the small- L approximation to Equation (43), namely

$$\frac{dt_{\text{return}}}{dt} = 1 + \sin^2 \theta L \dot{h}_+(t)$$

(See [69] for first corrections to the short-arm approximation.) To analyze the full detector, where the second arm will normally point out of the plane we have been working in up till now, it is helpful to go over to a tensorial expression, independent of special coordinate orientations. The gravitational wave will act in the plane transverse to the propagation direction; let us call that direction $\hat{N}$ and let us set up radiation basis vectors $\hat{e}_x^R$ and $\hat{e}_y^R$ in the transverse plane, such that $\hat{e}_x^R$ lies in the plane formed by the wave propagation direction and the arm of our gravitational wave sensor, which lies along the x -axis of the detector plane, whose unit vector is $\hat{e}_x$. (For a picture of this geometry, see the left-hand panel of Figure 3, where for the moment we are ignoring the y -arm of the detector shown there.)

With these definitions, the wave amplitude h_+ is the one that has $\hat{e}_x^R$ and $\hat{e}_y^R$ as the axes of its ellipse. The full wave amplitude is described, as in Equation (6), by the wave tensor

$$\mathbf{h}(t) = h_+(t)\mathbf{e}_+ + h_\times(t)\mathbf{e}_\times$$

where $\mathbf{e}_+$ and $\mathbf{e}_\times$ are the polarization tensors associated with these basis vectors (compare Equation (4)):

$$\mathbf{e}_+ = (\hat{e}_x^R \otimes \hat{e}_x^R - \hat{e}_y^R \otimes \hat{e}_y^R), \quad \mathbf{e}_\times = (\hat{e}_x^R \otimes \hat{e}_y^R + \hat{e}_y^R \otimes \hat{e}_x^R)$$

The unique way of expressing Equation (44) in terms of $\mathbf{h}$ is

$$\left(\frac{dt_{\text{return}}}{dt} \right)_{\text{x-arm}} = 1 + L \hat{e}_x \cdot \dot{\mathbf{h}} \cdot \hat{e}_x$$

This does not depend on any special orientation of the arm relative to the wave direction, and does not depend on the basis we chose in the transverse plane, so we can use it as well for the second arm of the interferometer, no matter what its orientation. Let us assume it lies along the unit vector by $\hat{e}_y$. (We do not, in fact, have to assume that the two arms are perpendicular to each other, but it simplifies the diagram a little.) The return-time derivative along the second arm is then given by

$$\left(\frac{dt_{\text{return}}}{dt} \right)_{\text{y-arm}} = 1 + L \hat{e}_y \cdot \dot{\mathbf{h}} \cdot \hat{e}_y$$

. The interferometer responds to the difference between these times,

$$\left(\frac{d\delta t_{\text{return}}}{dt} \right) = \left(\frac{dt_{\text{return}}}{dt} \right)_{\text{x-arm}} - \left(\frac{dt_{\text{return}}}{dt} \right)_{\text{y-arm}} = L \hat{e}_x \cdot \dot{\mathbf{h}} \cdot \hat{e}_x - L \hat{e}_y \cdot \dot{\mathbf{h}} \cdot \hat{e}_y$$

. By analogy with the wave tensor, we define the detector tensor $\mathbf{d}$ by [146]

$$\mathbf{d} = L (\hat{e}_x \otimes \hat{e}_x - \hat{e}_y \otimes \hat{e}_y)$$

(If the arms are not perpendicular this expression would still give the correct tensor if the unit vectors lie along the actual arm directions.) Then we can express the differential return time rate in the simple invariant form

$$\left(\frac{d\delta t_{\text{return}}}{dt} \right) = \mathbf{d} : \dot{\mathbf{h}}$$

where the notation $\mathbf{d} : \mathbf{h} \equiv d_{lm} h^{lm}$ denotes the Euclidean scalar product of the tensors $\mathbf{d}$ and $\mathbf{h}$. Equation (49) can be integrated over time to give the instantaneous path-length (or time-delay, or phase) difference between the arms, as measured by the central observer's proper time clock:

$$\delta t_{\text{return}}(t) = \mathbf{d} : \mathbf{h}.$$

This is a robust and compact expression for the response of any interferometer to any wave in the long-wavelength (short-arm) limit. Its dependence on the wave direction is called its antenna pattern.

It is conventional to re-express this measurable in terms of the stretching of the arms of the interferometer. Within our approximation that the arms are shorter than a wavelength, this makes sense: it is possible to define a local inertial coordinate system that covers the entire interferometer, and within this coordinate patch (where light travels at speed 1) time differences measure proper length differences. The differential return time is twice the differential length change of the arms:

$$\delta L(t) = \frac{1}{2} \mathbf{d} : \mathbf{h}$$

For a bar detector of length L lying along the director $\hat{a}$, the detector tensor is

$$\mathbf{d} = L \hat{a} \otimes \hat{a}$$

although one must be careful that the change in proper length of a bar is not simply given by Equation (51), because of the restoring forces in the bar.

When dealing with observations by more than one detector, it is not convenient to tie the alignment of the basis vectors in the sky plane with those in the detector frame, as we have done in the left-hand panel of Figure 3, since the detectors will have different orientations. Instead it will usually be more convenient to choose polarization tensors in the sky plane according to some universal reference, e.g., using a convenient astronomical reference frame. The right-hand panel of

Figure 3 shows the general situation, where the basis vectors $\hat{a}$ and $\hat{\beta}$ are rotated by an angle ψ from the basis used in the left-hand panel. The polarization tensors on this new basis,

$$\epsilon_+ = (\hat{a} \otimes \hat{a} - \hat{\beta} \otimes \hat{\beta}), \quad \epsilon_\times = (\hat{a} \otimes \hat{\beta} + \hat{\beta} \otimes \hat{a})$$

are found by the following transformation from the previous ones:

$$\begin{aligned} \epsilon_+ &= \mathbf{e}_+ \cos 2\psi + \mathbf{e}_\times \sin 2\psi \\ \epsilon_\times &= -\mathbf{e}_+ \sin 2\psi + \mathbf{e}_\times \cos 2\psi \end{aligned}$$

Then one can write Equation (51) as

$$\frac{\delta L(t)}{L} = F_+(\theta, \phi, \psi) h_+(t) + F_\times(\theta, \phi, \psi) h_\times(t)$$

where F_+ and $F_\times$ are the antenna pattern functions for the two polarizations defined on the sky-plane basis by

$$F_+ \equiv \mathbf{d} : \mathbf{e}_+, \quad F_\times \equiv \mathbf{d} : \mathbf{e}_\times$$

Using the geometry in the right-hand panel of Figure 3, one can show that

$$\begin{aligned} F_+ &= \frac{1}{2} (1 + \cos^2 \theta) \cos 2\phi \cos 2\psi - \cos \theta \sin 2\phi \sin 2\psi \\ F_\times &= \frac{1}{2} (1 + \cos^2 \theta) \cos 2\phi \sin 2\psi + \cos \theta \sin 2\phi \cos 2\psi \end{aligned}$$

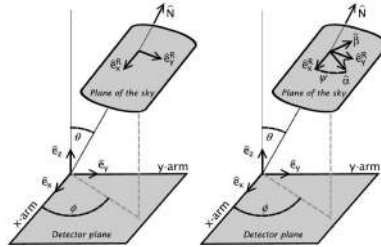


Figure 3: The relative orientation of the sky and detector frames (left panel) and the effect of a rotation by the angle ψ in the sky frame (left panel).

These are the antenna-pattern response functions of the interferometer to the two polarizations of the wave as defined in the sky plane [359]. If one wants the antenna pattern referred to the detector's own axes, then one just sets $\psi = 0$. If the arms of the interferometer are not

perpendicular to each other, then one defines the detector-plane coordinates x and y in such a way that the bisector of the angle between the arms lies along the bisector of the angle between the coordinate axes [334]. Note that the maximum value of either F_+ or $F_\times$ is 1.

The corresponding antenna-pattern functions of a bar detector whose longitudinal axis is aligned along the z direction, are

$$F_+ = \sin^2 \theta \cos 2\psi, \quad F_\times = \sin^2 \theta \sin 2\psi$$

Any one detector cannot directly measure both independent polarizations of a gravitational wave at the same time, but responds rather to a linear combination of the two that depends on the geometry of the detector and source direction. If the wave lasts only a short time, then the responses of three widely-separated detectors, together with two independent differences in arrival times among them, are, in principle, sufficient to fully reconstruct the source location and gravitational wave polarization. A long-lived wave will change its location in the antenna pattern as the detector moves, and it will also be frequency modulated by the motion of the detector; these effects are in principle sufficient to determine the location of the source and the polarization of the wave.

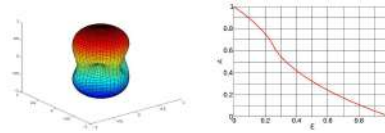


Figure 4: The antenna pattern of an interferometric detector (left panel) with the arms in the $x - y$ plane and oriented along the two axes. The response F for waves coming from a certain direction is proportional to the distance to the point on the antenna pattern in that direction. Also shown is the fractional area in the sky over which the response exceeds a fraction ϵ of the maximum (right panel).

Since the polarization angle of an incoming gravitational wave would generally be expected to be unrelated to its direction of arrival, depending rather on the internal orientations in the source, it is useful to characterize the directional sensitivity of a detector by averaging over the polarization angle ψ . If the wave has a given amplitude h and is linearly polarized, then, if we are interested in a single detector's response, it is always possible to align the polarization angle ψ in the sky plane with that of the wave, so that the wave has pure $+$ -polarization. Then the rms response function of the detector is

$$\bar{F} = \left(\int F_+^2 d\psi \right)^{1/2}$$

The function $\bar{F}$ is often simply called the antenna pattern. For a resonant bar, the antenna pattern is

$$\bar{F} = \sin^2 \theta$$

and for an interferometer, it is given by

$$\bar{F}^2 = \frac{1}{4} (1 + \cos^2 \theta)^2 \cos^2 2\phi + \cos^2 \theta \sin^2 2\phi$$

The antenna pattern of an interferometric detector is plotted in the left panel of Figure 4, which clearly shows the quadrupolar nature of the detector. Note that the response of an interferometer is the best for waves coming from a direction orthogonal to the plane containing the detector, and it is zero for waves in the plane of an interferometer's arms (i.e., $\theta = \pi/2$) that arrive from a direction bisecting the two arms (i.e., $\phi = \pi/4$) or from directions differing from this by a multiple of $\pi/2$. What is the response of an antenna to a linearly-polarized source at a random location in the sky? This is given by the rms value of $\bar{F}$ over the sky,

$$\left[\frac{1}{4\pi} \int \bar{F}^2 \sin \theta d\theta d\phi \right]^{1/2}$$

which is smaller than the maximum response by a factor of $2/\sqrt{15}$ (52%) for a bar detector and by $\sqrt{2/5}$ (63%) for an interferometer.

The polarization amplitudes of the radiation from an inspiraling binary, a rotating neutron star, or a ringing black hole, take a simple form as follows:

$$h_+ = \frac{h_0}{2} (1 + \cos^2 \iota) \cos \Phi(t), \quad h_\times = h_0 \cos \iota \sin \Phi(t)$$

where h_0 is an overall (possibly time-dependent) amplitude, $\Phi(t)$ is the signal's phase and ι is the angle made by the characteristic direction in the source (e.g., the orbital or the spin angular momentum) with the line of sight. In this case, the response takes a particularly simple form:

$$h(t) = F_+ h_+ + F_\times h_\times = A h_0 \cos(\Phi(t) - \Phi_0)$$

where

$$A = (A_+^2 + A_\times^2)^{1/2}, \quad \tan \Phi_0 = \frac{A_\times}{A_+}, \quad A_+ = \frac{1}{2} F_+ (1 + \cos^2 \iota), \quad A_\times = F_\times \cos \iota.$$

Note that A , just as F , takes values in the range $[0, 1]$. In this case the average response has to be worked out by considering all possible sky locations, polarizations (which drops out of the calculation) and source orientations. More precisely, the rms response is

$$\bar{A} = \frac{1}{8\pi^2} \int_0^\pi \sin \iota d\iota \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\varphi (A_+^2 + A_\times^2)$$

For an interferometer the above integral gives $2/5$. Thus, the rms response is still 40% of the peak response.

The right-hand panel of Figure 4 shows the percentage area of the sky over which the antenna pattern of an interferometric detector is larger than a certain fraction ϵ of the peak value. The response is better than the rms value over 40% of the sky, implying that gravitational wave detectors are fairly omni-directional. In comparison, the sky coverage of most conventional telescopes (radio, infrared, optical, etc.) is a tiny fraction of the area of the sky.

7.1.3 Practical issues of ground-based interferometers

A detector with an arm length of 4 km responds to a gravitational wave with an amplitude of 10^{-21} with

$$\delta l_{gw} \sim h l \sim 4 \times 10^{-18} \text{ m}$$

Light takes only about 10^{-5} s to go up and down one arm, much less than the typical period of gravitational waves of interest. Therefore, it is beneficial to arrange for the light to remain in an arm longer than this, say for 100 round trips. This increases its effective path length by 100 and hence the shift in the position of a given phase of the light beam will be of order 10^{-16} m. Most interferometers keep the light in the arms for this length of time by setting up optical cavities in the arms with low-transmissivity mirrors; these are called Fabry-Pérot cavities.

The main sources of noise against which a measurement must compete are:

1. Ground vibration. External mechanical vibrations must be screened out. These are a problem for bar detectors, too, but are more serious for interferometers, not least because interferometers bounce light back and forth between the mirrors, and so each reflection introduces further vibrational noise. Suspension/isolation systems are based on pendulums. A pendulum is a good mechanical filter for frequencies above its natural frequency. By hanging the mirrors on pendulums of perhaps 0.5 m length, one achieves filtering above a few Hertz. Since the spectrum of ground noise falls at higher frequencies, this provides suitable isolation. But these systems can be very sophisticated; the GEO600 [143] detector has a three-stage pendulum and other vibration isolation components [291]. The most ambitious multi-stage isolation system has been developed for the Virgo detector [175].
2. Thermal noise. Vibrations of the mirrors and of the suspending pendulum can mask gravitational waves. As with vibrational noise, this is increased by the bouncing of the light between the mirrors. Opposite to bars, interferometers perform measurements at frequencies far from the resonant frequency, where the amplitude of thermal vibration is largest. Thus, the pendulum suspensions have thermal noise at a few Hertz, so measurements will be made above 40 Hz in the first detectors. Internal vibrations of the mirrors have natural frequencies of several kHz, which sets an effective upper limit to the observing band. By ensuring that both kinds of oscillations have very high Q , one can confine most of the vibration energy to a small bandwidth around the resonant frequency, so that at the measurement frequencies the vibration amplitudes are extremely small. This allows interferometers to operate at room temperature. But mechanical Q_s of 10^7 or higher are required, and this is technically demanding.

Thermal effects produce other disturbances besides vibration. Some of the mirrors in interferometers are partly transmissive, as is the beam splitter. A small amount of light power is absorbed during transmission, which raises the temperature of the mirror and changes its index of refraction. The resulting "thermal lensing" can ruin the optical properties of the system, and random fluctuations in lensing caused by time-dependent thermal fluctuations (thermorefractive noise) can appear at measurement frequencies. These effects can limit the amount of laser power that can be used in the detector. Other problems can arise from heating effects in the multiple-layer coatings that are applied to the reflective surfaces of mirrors.

3. Shot noise. The photons that are used to do interferometry are quantized, and so they arrive at random and make random fluctuations in the light intensity that can look like a gravitational wave signal. The more photons one uses, the smoother the interference signal will be. As a random process, the error improves with the square root of the number N of photons. Using infrared light with a wavelength $\lambda \sim 1\mu\text{m}$, one can expect to measure to an accuracy of

$$\delta l_{\text{shot}} \sim \lambda / (2\pi\sqrt{N})$$

To measure at a frequency f , one has to make at least $2f$ measurements per second, so one can accumulate photons for a time $1/2f$. With light power P , one gets $N = P/(hc/\lambda)/(2f)$ photons. In order that δl_{shot} should be below 10^{-16} m, one needs high light power, far beyond the output of any continuous laser.

Light-recycling techniques overcome this problem, by using light efficiently. An interferometer actually has two places where light leaves. One is where the interference is measured, the difference port. The other is the sum of the two return beams on the beam splitter, which goes back towards the input laser. Normally one makes sure that no light hits the interference

sensor, so that only when a gravitational wave passes does a signal register there. This means that all the light normally returns toward the laser input, apart from small losses at the mirrors. Since mirrors are of good quality, only one part in 10^3 or less of the light is lost during a 1 ms storage time. By placing a power-recycling mirror in front of the laser, one can reflect this wasted light back in, allowing power to build up in the arms until the laser merely resupplies the mirror losses [149]. This can dramatically reduce the power requirement for the laser. The first interferometers work with laser powers of 5 – 10 W. Upgrades will use ten or more times this input power.

4. Quantum effects. Shot noise is a quantum noise, and like all quantum noises there is a corresponding conjugate noise. As laser power is increased to reduce shot noise, the position sensing accuracy improves, and one eventually comes up against the Heisenberg uncertainty principle: the momentum transferred to the mirror by the measurement leads to a disturbance that can mask a gravitational wave. To reduce this backaction pressure fluctuation, scientists are experimenting with a variety of interferometer configurations that modify the quantum state of the light, by "squeezing" the Heisenberg uncertainty ellipse, in order to reduce the effect of this uncertainty on the variable being measured, at the expense of its (unmeasured) conjugate. The key point here is that we are using a quantum field (light) to measure an effectively classical quantity (gravitational wave amplitude), so we do not need to know everything about our quantum system: we just need to reduce the uncertainty in that part of the quantum field that responds to the gravitational wave at the readout of our interferometer. The best results on squeezing so far [371] have been obtained during preparations for the GEO-HF upgrade of the GEO600 detector [395]. These techniques may be needed for the second-generation advanced detectors and will certainly be needed for advances beyond that.
5. Gravity gradient noise. One noise that cannot be screened out is that due to changes in the local Newtonian gravitational field on the timescale of the measurements. A gravitational wave detector will respond to tidal forces from local sources just as well as to gravitational waves. Environmental noise comes not only from man-made sources, but even more importantly from natural ones: seismic waves are accompanied by changes in the gravitational field, and changes in air pressure are accompanied by changes in air density. The spectrum falls steeply with increasing frequency, so for first-generation interferometers this will not be a problem, but it may limit the performance of more advanced detectors. And it is the primary reason that detecting gravitational waves in the low-frequency band below 1 Hz must be done in space.

Interferometers around the globe

The two largest interferometer projects are LIGO [302] and VIRGO [175]. LIGO has built three detectors at two sites. At Hanford, Washington, there is a 4 km and a 2 km detector in the same vacuum system. At Livingston, Louisiana, there is a single 4 km detector, oriented to be as nearly parallel to the Hanford detector as possible. After a series of "engineering" runs, which helped to debug the instruments, interspersed with several "science runs", which helped to test and debug the data acquisition system and various analysis pipelines, LIGO reached its design sensitivity goal in the final months of 2005. In November 2005, LIGO began a two-year data-taking run, called S5, which acquired a year's worth of triple coincidence data among the three LIGO detectors. S5 ended on 30 September 2007. Although interferometers are pretty stable detectors, environmental disturbances and instrumental malfunctions can cause them to lose lock during which the data quality will be either poor or ill defined. The typical duty cycle at one of the LIGO sites is about 80%, and hence about two years of operation was required

to accumulate a year's worth of triple coincident data. Up to date information on LIGO can be found on the project's website [103]. A recent review of LIGO's status is [303].

VIRGO finished commissioning its single 3-km detector at Cascina, near Pisa, in early 2007 and began taking data in coincidence with LIGO in May 2007, thus joining for the last part of S5. VIRGO is a collaboration among research laboratories in Italy and France, and its umbrella organization EGO looks after the operation of the site and planning for the future. There are websites for both VIRGO [380] and EGO [152]. A recent review of VIRGO's status is [20].

A smaller 600-m detector, GEO600, has been operational near Hanover, Germany, since 2001 [143]. It is a collaboration among research groups principally in Germany and Britain. Although smaller, GEO600 has developed and installed second-generation technology (primarily in its suspensions, mirror materials and interferometer configuration) that help it achieve a higher sensitivity. GEO600 technology is being transferred to LIGO and VIRGO as part of their planned upgrades, described below. Full information about GEO can be found on its website [261]. A recent review of GEO600's status is [396].

LIGO and GEO have worked together under the umbrella of the LIGO Scientific Collaboration (LSC) since the beginning of science data runs in 2001. The LSC contains dozens of groups from universities around the world, which contribute to data analysis and technology development. The two detector groups pool their data and analyze it jointly. The LSC has a website containing detailed information, and providing access to the publications and open-source software of the collaboration [236].

VIRGO has signed an agreement with the LSC to pool data and analyze it jointly, thereby creating a single worldwide network of long-baseline gravitational wave detectors. VIRGO is not, however, a member of the LSC.

The LSC has already published many papers on the analysis of data acquired during its science runs, and many more can be expected. The results from these science runs, which will be discussed later, are already becoming astrophysically interesting. The LSC maintains a public repository of its papers and contributions to conference proceedings [237].

For instance, although the search for continuous waves from known pulsars has not found any definitive candidates, it has been possible to set stringent upper limits $\epsilon \leq \text{few} \times 10^{-6}$ on the magnitude of the ellipticity of some of these systems [10]. In particular, in the case of the Crab pulsar, gravitational wave observations have begun to improve [12] the upper limit on the strength of radiation obtained by radio observations of the spin-down rate.

A yet smaller detector in Japan, TAMA300 [362], with 300 m arms, was the first large-scale interferometer to achieve continuous operation, at a sensitivity of about $10^{-19} - 10^{-20}$. TAMA is seen as a development prototype, and its sensitivity will be confined to higher frequencies (above ~ 500 Hz). An ambitious follow-on detector called the Large-scale Cryogenic Gravitational-Wave Telescope (LCGT) is being planned in Japan, and, as its name suggests, it will be the first to use cooled mirrors to reduce the effects of thermal noise. TAMA [269] and the LCGT [268] have websites where one can get more information. A recent review of TAMA's status is [130].

There are plans for a detector in Australia, and a small interferometer is operating in Western Australia [252]. The Australian Interferometric Gravitational Observatory (AIGO) [368] is a proposal of the Australian Consortium for Interferometric Gravitational Astronomy (ACIGA) [56]. The ACIGA collaboration is a member of the LSC and assists in mirror and interferometry development, but it is not yet clear whether and when a larger detector might be funded. From the point of view of extracting information from observations, it is very desirable to have large-scale detectors in Japan and Australia, because of their very long baselines to the USA and Europe. But the future funding of both LCGT and AIGO is not secure as of this writing (2008).

The initial sensitivity levels achieved by LIGO, VIRGO, and GEO are just the starting

point. Detailed plans exist for upgrades for all three projects. In October 2007, both LIGO and VIRGO began upgrading to enhanced detectors, which should improve on LIGO's S5 sensitivity by a factor of roughly two. These should come online in 2009. After a further observing run, called S6, the detectors will again shut down for a much more ambitious upgrade to advanced detectors, to operate around 2014. This will provide a further factor of five in sensitivity, and hence in range. Altogether the two upgrades will extend the volume of space that can be surveyed for gravitational waves by a factor of 1000, and this will make regular detections a virtual certainty. Advanced LIGO has a website giving the plans for the upgrade in the context of development from the initial sensitivity [235].

GEO600 will remain in science mode during the upgrade to enhanced detectors, just in case a nearby supernova or equally spectacular event should occur when the larger detectors are down. But, when the enhanced detectors begin operating, GEO will upgrade to GEO-HF [395], a modification designed to improve its sensitivity in the high-frequency region above 1kHz, where its short arm length does not prevent it being competitive with the larger instruments. GEO is also a partner in the Advanced LIGO project, contributing high-power lasers and high-Q suspensions for controlling thermal noise.

Beyond that, scientists are now studying the technologies that may be needed for a further large step in sensitivity to third-generation detectors. This may involve cooling mirrors, using ultra-massive substrates of special materials, using purely nontransmissive optics, and even circumventing the quantum limit in interferometers, as has been studied for bars. The goal of third-generation detectors would be to be limited just by gravity-gradient noise and quantum effects. A design study for a concept called the "Einstein Telescope" started in Europe in 2008.

Very-high-frequency detectors

The gravitational wave spectrum above the detection band of conventional interferometers, say above 10kHz, may not be empty, and stochastic gravitational waves from the Big Bang may be present up to megaHertz frequencies and beyond. It is exceedingly difficult to build sensitive detectors at these high frequencies, but two projects have made prototypes: a microwave-based detector that senses the change in polarization as the electromagnetic waves follow a waveguide circuit as a gravitational wave passes by [126], and a more conventional light-based interferometer [23].

7.1.4 Detection from space

Space offers two important ingredients for beam detectors: long arms and a free vacuum. In this section, we describe the three ways that space has been and will be used for gravitational wave detection: ranging to spacecraft (Section 4.4.1), pulsar timing (Section 4.4.2), and direct detection using space-based interferometers (Section 4.4.3).

Ranging to spacecraft

Both NASA and ESA perform experiments in which they monitor the return time of communication signals with interplanetary spacecraft for the characteristic effect of gravitational waves. For missions to Jupiter and Saturn, for example, the return times are of order $2 - 4 \times 10^3$ s. Any gravitational wave event shorter than this will, by Equation (43), appear three times in the time delay: once when the wave passes the Earth-based transmitter, once when it passes the spacecraft, and once when it passes the Earth-based receiver. Searches use a form of data analysis using pattern matching. Using two transmission frequencies and very stable atomic clocks, it is possible to achieve sensitivities for h of order 10^{-13} , and even 10^{-15} may soon be reached [40].

Pulsar timing

Many pulsars, particularly the old millisecond pulsars, are extraordinarily regular clocks when averaged over timescales of a few years, with random timing irregularities too small for the best atomic clocks to measure. If one assumes that they emit pulses perfectly regularly, then one can use observations of timing irregularities of single pulsars to set upper limits on the background gravitational-wave field. Here, the one-way formula Equation (42) is appropriate. The transit time of a signal to the Earth from the pulsar may be thousands of years, so we cannot look for correlations between the two terms in a given signal. Instead, the delay is a combination of the effects of waves at the pulsar when the signal was emitted and waves at the Earth when it is received. If one observes a single pulsar, then because not enough is known about the intrinsic irregularity in pulse emission, the most one can do is to set upper limits on a background of gravitational radiation at very low frequencies [242, 344].

If one simultaneously observes two or more pulsars, then the Earth-based part of the delay is correlated, and this offers, in addition, a means of actually detecting strong gravitational waves with periods of several years that pass the Earth (in order to achieve the long-period stability of pulse arrival times). Observations are currently underway at a number of observatories. The most stringent limits to date are from the Parkes Pulsar Timing Array [208], which sets an upper limit on a stochastic background of $\Omega_{\text{gw}} \leq 2 \times 10^{-8}$. Two further collaborations for timing have been formed: the European Pulsar Timing Array (EPTA) [345] and NanoGrav [39]. Astrophysical backgrounds in this frequency band are likely (see Section 8.2.2), so these arrays have a good chance of making early detections. Future timing experiments will be even more powerful, using new phased arrays of radio telescopes that can observe many pulsars simultaneously, such as the Low Frequency Array (LOFAR) [156] and the Square Kilometer Array [107].

Pulsar timing can also be used to search for individual events, not just a stochastic signal. The first example of an upper limit from such a search was the exclusion of a black-hole-binary model for 3C66B [209].

Space interferometry

Gravity-gradient noise on the Earth is much larger than the amplitude of any expected waves from astronomical sources at frequencies below about 1 Hz, but this noise falls off rapidly as one moves away from the Earth. A detector in space would not notice the Earth's noisy environment. The Laser Interferometer Space Antenna (LISA) project, currently being developed in collaboration by ESA and NASA with a view toward launching in 2018, would open up the frequency window between 0.1 mHz and 0.1 Hz for the first time [196, 144]. There are several websites that provide full information about this project [24, 153, 266].

We will see below that there are many exciting sources expected in this wave band, for example the coalescences of giant black holes in the centers of galaxies. LISA would see such events with extraordinary sensitivity, recording typical SNRs of 1000 or more for events at redshift one.

A space-based interferometer can have arm lengths much greater than a wavelength. LISA, for example, would have arms 5×10^6 km long, and that would be longer than half a wavelength for any gravitational waves above 30 mHz. In this regime, the response of each arm will follow the three-term formula we encountered earlier. The short-arm approximation we used for ground-based interferometers works for LISA only at the lowest frequencies in its observing band.

LISA will consist of three free-flying spacecraft, arranged in an array that orbits the sun at 1 AU, about 20 degrees behind the Earth in its orbit. They form an approximately equilateral triangle in a plane tilted at 60° to the ecliptic, and their simple Newtonian elliptical orbits

around the sun preserve this arrangement, with the array rotating backwards once per year as the spacecraft orbit the sun. By passing light along each of the arms, one can construct three different Michelson-type interferometers, one for each vertex. With this array one can measure the polarization of a gravitational wave directly. The spacecraft are too far apart to use simple mirrors to reflect light back along an arm: the reflected light would be too weak. Instead, LISA will have optical transponders: light from one spacecraft's on-board laser will be received at another, which will then send back light from its own laser locked exactly to the phase of the incoming signal.

The main environmental disturbances to LISA are forces from the sun: solar radiation pressure and pressure from the solar wind. To minimize these, LISA incorporates drag-free technology. Interferometry is referenced to an internal proof mass that falls freely, unattached to the spacecraft. The job of the spacecraft is to shield this mass from external disturbances. It does this by sensing the position of the mass and firing its own jets to keep itself (the spacecraft) stationary relative to the proof mass. To do this, it needs thrusters of very small thrust that have accurate control. The key technologies that have enabled the LISA mission are the availability of such thrusters, accelerometers needed to sense disturbances to the spacecraft, and lasers capable of continuously emitting 1 W of infrared light for many years. ESA is planning to launch a satellite called LISA Pathfinder to test all of these technologies in 2010 [230].

LISA is not the only proposal for an interferometer in space for gravitational wave detection. The DECIGO proposal is a more ambitious design, positioned at a higher frequency to fill the gap between LISA and ground-based detectors [213]. Even more ambitious, in the same frequency band, is the Big Bang Observer, a NASA concept study to examine what technology would be needed to reach the ultimate sensitivity of detecting a gravitational wave background from inflation at these frequencies [267].

7.1.5 Characterizing the sensitivity of a gravitational wave antenna

The performance of a gravitational wave detector is characterized by the power spectral density (henceforth denoted PSD) of its noise background. One can construct the noise PSD as follows; a gravitational wave detector outputs a dimensionless data train, say $x(t)$, which in the case of an interferometer is the relative strain in the two arms, scaled to represent the value of h that would produce that strain if the wave is optimally oriented with respect to the detector. In the absence of any gravitational wave signal, the detector output is just an instance of noise $n(t)$, that is, $x(t) = n(t)$. The noise auto-correlation function κ is defined as

$$\kappa \equiv \overline{n(t_1) n(t_2)}$$

where an overline indicates the average over an ensemble of noise realizations. In general, κ depends both on t_1 and t_2 . However, if the detector output is a stationary noise process, i.e., its performance is, statistically speaking, independent of time, then κ depends only on $\tau \equiv |t_1 - t_2|$.

The assumption of stationarity is not strictly valid in the case of real gravitational-wave detectors; however, if their performance doesn't vary greatly over time scales much larger than typical observation time scales, stationarity could be used as a working rule. While this may be good enough in the case of binary inspiral and coalescence searches, it is a matter of concern for the observation of continuous and stochastic gravitational waves. In this review, for simplicity, we shall assume that the detector noise is stationary. In this case the one-sided noise PSD, defined only at positive frequencies, is the Fourier transform of the noise auto-correlation function:

$$S_h(f) \equiv \frac{1}{2} \int_{-\infty}^{\infty} \kappa(\tau) e^{2\pi i f \tau} d\tau, \quad f \geq 0$$

$$\equiv 0, \quad f < 0$$

where a factor of $1/2$ is included by convention. By using the Fourier transform of $n(t)$, that is $\tilde{n}(f) \equiv \int_{-\infty}^{\infty} n(t) e^{2\pi i f t} dt$, in Equation (65) and substituting the resulting expression in Equation (66), it is easy to show that for a stationary noise process background

$$\overline{\tilde{n}(f) \tilde{n}^*(f')} = \frac{1}{2} S_h(f) \delta(f - f')$$

where $\tilde{n}^*(f)$ denotes the complex conjugate of $\tilde{n}(f)$. The above equation justifies the name PSD given to $S_h(f)$.

It is obvious that $S_h(f)$ has dimensions of time but it is conventional to use the dimensions of Hz^{-1} , since it is a quantity defined in the frequency domain. The square root of $S_h(f)$ is the noise amplitude, $\sqrt{S_h(f)}$, and has dimensions of $\text{Hz}^{-1/2}$. Both noise PSD and noise amplitude measure the noise in a linear frequency bin. It is often useful to define the power per logarithmic bin $h_n^2(f) \equiv f S_h(f)$, where $h_n(f)$ is called the effective gravitational-wave noise, and it is a dimensionless quantity. In gravitational-wave-interferometer literature one also comes across gravitational-wave displacement noise or gravitational-wave strain noise defined as $h_\ell(f) \equiv \ell h_n(f)$, and the corresponding noise spectrum $S_\ell(f) \equiv \ell^2 S_h(f)$, where ℓ is the arm length of the interferometer. The displacement noise gives the smallest strain $\delta\ell/\ell$ in the arms of an interferometer that can be measured at a given frequency.

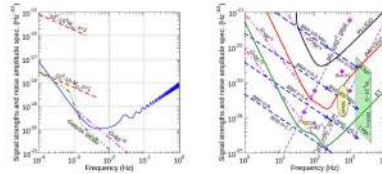


Figure 5: The right panel plots the noise amplitude spectrum, $\sqrt{f S_h(f)}$, in three generations of groundbased interferometers. For the sake of clarity, we have only plotted initial and advanced LIGO and a possible third generation detector sensitivities. VIRGO has similar sensitivity to LIGO at the initial and advanced stages, and may surpass it at lower frequencies. Also shown are the expected amplitude spectrum of various narrow and broad-band astrophysical sources. The left panel is the same as the right except for the LISA detector. The SMBH sources are assumed to lie at a redshift of $z = 1$, but LISA can detect these sources with a good SNR practically anywhere in the universe. The curve labelled "Galactic WDBs" is the confusion background from the unresolvable Galactic population of white dwarf binaries.

Noise power spectral density in interferometers

As mentioned earlier, the performance of a gravitational wave detector is characterized by the one-sided noise PSD. The noise PSD plays an important role in signal analysis. In this review we will only discuss the PSDs of interferometric gravitational-wave detectors.

The sensitivity of ground based detectors is limited at frequencies less than a Hertz by the timevarying local gravitational field caused by a variety of different noise sources, e.g., low frequency seismic vibrations, density variation in the atmosphere due to winds, etc. Thus, for data analysis purposes, the noise PSD is assumed to be essentially infinite below a certain lower cutoff f_s . Above

Table 1: Noise power spectral densities $S_h(f)$ of various interferometers in operation and under construction: GEO600, Initial LIGO (ILIGO), TAMA, VIRGO, Advanced LIGO (ALIGO),

Einstein Telescope (ET) and LISA (instrumental noise only). For each detector the noise PSD is given in terms of a dimensionless frequency $x = f/f_0$ and rises steeply above a lower cutoff f_s . The parameters in the ET design sensitivity curve are $\alpha = -4.1, \beta = -0.69, a_0 = 186, b_0 = 233, b_1 = 31, b_2 = -65, b_3 = 52, b_4 = -42, b_5 = 10, b_6 = 12, c_1 = 14, c_2 = -37, c_3 = 19, c_4 = 27$. (See also Figure 5.)

Detector	f_s/Hz	f_0/Hz	S_0/Hz^{-1}	$S_h(x)/S_0$
GEO	40	150	1.0×10^{-46}	$(3.4x)^{-30} + 34x^{-1} + \frac{20(1-x^2+0.5x^4)}{(1+0.5x^2)}$
ILIGO	40	150	9.0×10^{-46}	$(4.49x)^{-56} + 0.16x^{-4.52} + 0.52 + 0.32x^2$
TAMA	75	400	7.5×10^{-46}	$x^{-5} + 13x^{-1} + 9(1+x^2)$
VIRGO	20	500	3.2×10^{-46}	$(7.8x)^{-5} + 2x^{-1} + 0.63 + x^2$
ALIGO	20	215	1.0×10^{-49}	$x^{-4.14} - 5x^{-2} + \frac{111(1-x^2+0.5x^4)}{1+0.5x^2}$
ET	10	200	1.5×10^{-52}	$x^\alpha + a_0x^\beta + \frac{b_0(1+b_1x+b_2x^2+b_3x^3+b_4x^4+b_5x^5+b_6x^6)}{1+c_1x+c_2x^2+c_3x^3+c_4x^4}$
LISA	10^{-5}	10^{-3}	9.2×10^{-44}	$(x/10)^{-4} + 173 + x^2$

this cutoff, i.e., for $f \geq f_s$, Table 1 lists the noise PSD $S_h(f)$ for various interferometric detectors and some of these are plotted in Figure 5.

For LISA, Table 1 gives the internal instrumental noise only, taken from [162]. It is based on the noise budget obtained in the LISA Pre-Phase A Study [70]. However, in the frequency range $10^{-4} - 10^{-2}$ Hz, LISA will be affected by source confusion from astrophysical backgrounds produced by several populations of galactic binary systems, such as closed white-dwarf binaries, binaries consisting of Cataclysmic Variables, etc. At frequencies below about 1mHz, there are too many binaries for LISA to resolve in, say, a 10-year mission, so that they form a Gaussian noise. Above this frequency range, there will still be many resolvable binaries which can, in principle, be removed from the data.

Nelemans et al. [272] estimate that the effective noise power contributed by binaries in the galaxy is

$$S_h^{\text{gal}} = 2.1 \times 10^{-38} \left(\frac{f}{f_s} \right)^{7/3} \text{ Hz}^{-1}, \quad f_s = 10^{-3} \text{ Hz}$$

normalized to the same f_s as we use for LISA in Table 1. This power is a mean frequency average based on projections of the population LISA will find, but, of course, above about 1mHz, LISA will resolve many binaries and identify most of the members of this population. Barack and Cutler [67] have provided a prescription for including this effect when adding in the confusion noise. They make the conservative assumption that individual binaries contaminate the instrumental noise S_h^{instr} (see Table 1) in such a way that, effectively, one or a few frequency resolution bins need to be cut out and ignored when detecting other signals, including, of course, other binary signals. This would have approximately the same effect as if the overall instrumental noise at that frequency were raised by an amount obtained simply by dividing the noise by the fraction η of bins free of contamination. Of course, when this fraction reaches zero (below 1mHz), this approximation is not valid, and instead one should just add the full binary confusion noise in Equation (68) to the instrumental noise. A smooth way of merging these two regimes is to set

$$S_h^{\text{full}} = \min \left(\frac{1}{\eta} S_h^{\text{instr}}, \quad S_h^{\text{instr}} + S_h^{\text{gal}} \right)$$

where S_h^{instr} is from Table 1 and S_h^{gal} is from Equation (68). This prescription uses the contaminated instrumental noise, when it is below the total noise power from the binaries, but

then uses the total binary confusion power when the prescription for allowing for contamination breaks down.

The fraction η of uncontaminated frequency bins as a function of frequency remains to be specified. Let dN/df be the number of binaries in the galaxy per unit frequency. Since the size of the frequency bin for an observation that lasts a time T_{obs} is $1/T_{\text{obs}}$, the expected number of binaries per frequency bin is

$$\Delta N(f) = \frac{1}{T_{\text{obs}}} \frac{dN(f)}{df}$$

Barack and Cutler multiply this by a "fudge factor" $\kappa > 1$ to allow for the fact that any binary may contaminate several bins, so that $\kappa\Delta N(f)$ is the expected number of contaminated bins per binary. If this is small, then it will equal the fractional contamination at frequency f . In that case, the fraction of uncontaminated bins is just $1 - \kappa\Delta N(f)$. However, if the expected contamination per bin approaches or exceeds one, then we have to allow for the fact that the binaries are really randomly distributed in frequency, so that the expected fraction not contaminated comes from the Poisson distribution,

$$\eta = \exp(-\kappa\Delta N)$$

Inserting this into Equation (69) gives a reasonable approximation to the effective instrumental noise if binaries cannot be removed in a clean way from the data stream when looking for other signals.

Because LISA will observe binaries for several years, the accuracy with which it will know the frequency, say, of a binary, will be much better than the frequency resolution of LISA during the observation of a transient source, such as many of the IMBH events considered by Barack and Cutler. Therefore, there is a good chance that, in the global LISA data analysis, the effective noise can be reduced below the one-year noise levels that are normally used in projecting the sensitivity of LISA and the science it can do.

Sensitivity of interferometers in units of energy flux

In radio astronomy one talks about the sensitivity of a telescope in terms of the limiting detectable energy flux from an astronomical source. We can do the same here too. Given the gravitational wave amplitude h we can use Equation (17) to compute the flux of gravitational waves. One can translate the noise power spectrum $S_h(f)$, given in units of Hz^{-1} at frequency f , to Jy (Jansky), with the conversion factor $4c^3 f^2 / (\pi G)$. In Figure 6, the left panel shows the noise power spectrum in astronomical units of Jy and the right panel depicts the noise spectrum in units of Hz^{-1} together with lines of constant flux.

What is striking in Figure 6 is the magnitude of flux. While modern radio interferometers are sensitive to flux levels of milli and micro-Jy the gravitational wave interferometers need their sources to be 24 – 27 orders of magnitude brighter. Turning this argument around, the gravitational wave sources we expect to observe are not really weak, but rather extremely bright sources. The difficulty in detecting them is due to the fact that gravitation is the weakest of all known interactions.

7.1.6 Source amplitudes vs sensitivity

How does one compare the gravitational wave amplitude of astronomical sources with the instrumental sensitivity and assess what sort of sources will be observable against noise?

Comparisons

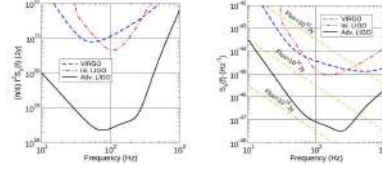


Figure 6: The sensitivity of interferometers in terms of the limiting energy flux they can detect, Jy/Hz, (left panel) and in terms of the gravitational wave amplitude with lines of constant flux levels (right panel).

are almost always made in the frequency domain, since stationary noise is most conveniently characterized by its PSD.

The simplest signal to characterize is a long-lasting periodic signal with a given fixed frequency f_0 . In an observation time T , all the signal power $|\tilde{h}(f_0)|^2$ is concentrated in a single frequency bin of width $1/T$. The noise against which it competes is just the noise power in the same bin, $S_h(f_0)/T$. The power SNR is then $T|\tilde{h}(f_0)|^2/S_h(f_0)$, and the amplitude SNR is $\sqrt{T}|\tilde{h}(f_0)|/|S_h(f_0)|^{1/2}$. This improves with observation time as the square root of the time. The reason for this is that the noise is stationary, but longer and longer observation times permit the signal to compete only with noise in smaller and smaller frequency windows.

Of course, no expected gravitational-wave signal would have a single fixed frequency in the detector frame, because the detector is attached to the Earth, whose motion produces frequency modulations. But the principle of this SNR increase with time can still be maintained if one has a signal model that allows one to exclude more and more noise from competing with the signal over longer and longer periods of time. This happens with matched filtering, which we return to in Section 5.

Short-lived signals have wider bandwidths, and long observation times are not relevant. To characterize their SNR, it is useful to define the dimensionless noise power per logarithmic bandwidth, $fS_h(f)$, which we earlier called $h_n^2(f)$. The signal Fourier amplitude $\tilde{h}(f) \equiv \int_{-\infty}^{\infty} dt h(t) e^{2\pi i f t}$ has dimensions of Hz^{-1} and so the Fourier amplitude per logarithmic frequency, which is called the characteristic signal amplitude $h_c = f|\tilde{h}(f)|$, is dimensionless. This quantity should be compared with $h_n(f)$ to obtain a rough estimate of the SNR of the signal: $\text{SNR} \sim h_c/h_n$.

7.1.7 Network detection

Gravitational wave detectors are almost omni-directional. As discussed in Section 4.2.1, both interferometers and bars have good sensitivity over a large area of the sky. In this regard, gravitational wave antennas are unlike conventional astronomical telescopes, e.g., optical, radio, or infrared bands, which observe only a very small fraction of the sky at any given time. The good news is that gravitational wave interferometers will have good sky coverage and therefore only a small number (around six) are enough to survey the sky. The bad news, however, is that gravitational wave observations will not automatically provide the location of the source in the sky. It will either be necessary to observe the same source in several non-co-located detectors and triangulate the position of the source using the information from the delay in the arrival times of the signals to different detectors, or observe for a long time and use the location-dependent Doppler modulation caused by the motion of the detector relative to the source to infer the source's position in the sky. The latter is a well-known technique in radio astronomy of synthesizing a long-baseline observation to gain resolution, and only possible for sources, such as rotating neutron stars or stochastic backgrounds, that last for a long enough

duration.

A network of detectors is, therefore, essential for source reconstruction. Network observation is not only powerful in identifying a source in the sky, but independent observation of the same source in several detectors adds to the detection confidence, especially since the noise background in the first generation of interferometers is not well understood and is plagued by nonstationarity and non-Gaussianity.

Coherent vs coincidence analysis

The availability of a network of detectors offers two different methods by which the data can be combined. One can either first bring the data sets together, combine them in a certain way, and then apply the appropriate filter to the network data and integrate the signal coherently, coherent detection [282, 87, 160, 41], or first analyze the data from each detector separately by applying the relevant filters and then look for coincidences in the multi-dimensional space of intrinsic (masses of the component stars, their spins, ...) and extrinsic (arrival times, a constant phase, source location, ...) parameters, coincidence detection [205, 207, 160, 42, 353, 2, 6, 7, 8].

A recent comparison of coherent analysis vis-a-vis coincidence analysis under the assumption that the background noise is Gaussian and stationary has concluded that coherent analysis, as one might expect, is far better than coincidence analysis [263]. These authors also explore, to a limited extent, the effect of nonstationary noise and reach essentially the same conclusion.

At the outset, coherent analysis sounds like a good idea, since in a network of N_D similar detectors the visibility of a signal improves by a factor of $\sqrt{N_D}$ over that of a single detector. One can take advantage of this enhancement in SNR to either lower the false alarm rate by increasing the detection threshold, while maintaining the same detection efficiency, or improve detection efficiency at a given false alarm rate.

However, there are two reasons that current data-analysis pipelines prefer coincidence analysis over coherent analysis. Firstly, since the detector noise is neither Gaussian nor stationary, coincidence analysis can potentially reduce the background rate far greater than one might think otherwise. Secondly, coherent analysis is computationally far more expensive than coincidence analysis and it is presently not practicable to employ coherent analysis.

Coincidence analysis is indeed a very powerful method to veto out spurious events. One can associate with each event in a given detector an ellipsoid, whose location and orientation depends on where in the parameter space and when the event was found, and the SNR can be used to fix the size of the ellipsoid [314]. One is associating with each event a 'sphere' of influence in the multi-dimensional space of masses, spins, arrival times, etc., and there is a stringent demand that the spheres associated with events from different detectors should overlap each other in order to claim a detection. Since random triggers from a network of detectors are less likely to be consistent with one another, this method serves as a very powerful veto.

It is probably not possible to infer beforehand which method might be more effective in detecting a source, as this might depend on the nature of the detector noise, on how the detection statistic is constructed, etc. An optimal approach might be a suitable combination of both of these methods. For instance, a coherent follow-up of a coincidence analysis (as is currently done by searches for compact binaries within the LSC) or to use coincidence criteria on candidate events from a coherent search.

Coherent addition of data improves the visibility of the signal, but 'coherent subtraction' of the data in a detector network should lead to data products that are devoid of gravitational wave signals. This leads us naturally to the introduction of the null stream veto.

Null stream veto

Data from a network of detectors, when suitably shifted in time and combined linearly with coefficients that depend on the source location, will yield a time series that, in the ideal case, will be entirely devoid of the gravitational signal. Such a combination is called a null stream. For instance, for a set of three misaligned detectors, each measuring a data stream $x_k(t)$, $k = 1, 2, 3$, the combination $x(t) = A_{23}(\theta, \varphi)h_1(t + \tau_1) + A_{31}(\theta, \varphi)h_2(t + \tau_2) + A_{12}(\theta, \varphi)h_3(t + \tau_3)$, where A_{ij} are functions of the responses of the antennas i and j , and τ_k 's, $k = 1, 2, 3$, are time delays that depend on the source location and the location of the antenna, is a null stream. If $x_k(t)$, $k = 1, 2, 3$, contain a gravitational wave signal from an astronomical source, then $x(t)$ will not contain the signature of this source. In contrast, if $x(t)$ and $x_k(t)$ both contain the signature of a gravitational wave event, then that is an indication that one of the detectors has a glitch.

The existence and usefulness of a null stream was first pointed out by Gürsel and Tinto [184]. Wen and Schutz [390] proposed implementing it in LSC data analysis as a veto, and this has been taken up now by several search groups.

Detection of stochastic signals by cross-correlation

Stochastic background sources and their detection is discussed in more detail in Section 8. Here we will briefly mention the problem in the context of detector networks. As mentioned in Section 3.6, the universe might be filled with stochastic gravitational waves that were either generated in the primeval universe or by a population of background sources. For point sources, although each source in a population might not be individually detectable, they could collectively produce a confusion background via a random superposition of the waves from that population. Since the waves are random in nature, it is not possible to use the techniques described in Sections 4.7.1, 4.7.2 and 5.1 to detect a stochastic background. However, we might use the noisy stochastic signal in one of the detectors as a "matched-filter" for the data in another detector [359, 163, 27, 91]. In other words, it should be possible to detect a stochastic background by cross-correlating the data from a pair of detectors; the common gravitational-wave background will grow in time more rapidly than the random backgrounds in the two instruments, thereby facilitating the detection of the background.

If two instruments with identical spectral noise density S_h are cross-correlated over a bandwidth Δf for a total time T , the spectral noise density of the output is reduced by a factor of $(T\Delta f)^{1/2}$. Since the noise amplitude is proportional to the square root of S_h , the amplitude of a signal that can be detected by cross-correlation improves only with the fourth root of the observing time. This should be compared with the square root improvement that matched filtering gives.

The cross-correlation technique works well when the two detectors are situated close to one another. When separated, only those waves whose wavelength is larger than or comparable to the distance between the two detectors, or which arrive from a direction perpendicular to the separation between the detectors, can contribute coherently to the cross-correlation statistic. Since the instrumental noise builds up rapidly at lower frequencies, detectors that are farther apart are less useful in cross-correlation. However, very near-by detectors (as in the case of two LIGO detectors within the same vacuum tube in Hanford) will suffer from common background noise from the control system and the environment, making it rather difficult to ascertain if any excess noise is due to a stochastic background of gravitational waves.

7.1.8 False alarms, detection threshold and coincident observation

Gravitational-wave event rates in initial interferometers is expected to be rather low: about a few per year. Therefore, one has to set a high threshold, so that the noise-generated false alarms mimicking an event are negligible.

For a detector output sampled at 1kHz and processed through a large number of filters, say 10^3 , one has $\sim 3 \times 10^{13}$ instances of noise in a year. If the filtered noise is Gaussian, then the probability $P(x)$ of observing an amplitude in the range of x to $x + dx$ is

$$P(x)dx = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-x^2}{2\sigma^2}\right) dx$$

where σ is the standard deviation. The above probability-distribution function implies that the probability that the noise amplitude is greater than a given threshold η is

$$P(x | x \geq \eta) = \int_{\eta}^{\infty} P(x)dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{\eta}^{\infty} \exp\left(\frac{-x^2}{2\sigma^2}\right) dx$$

Demanding that no more than one noise-generated false alarm occur in a year's observation means that $P(x | x \geq \eta) = 1/(3 \times 10^{13})$. Solving this equation for η , one finds that $\eta \simeq 7.5\sigma$ in order that false alarms are negligible in a year's observation. Therefore, a source is detectable only if its amplitude is significantly larger than the effective noise amplitude, i.e., $f\tilde{h}(f) \gg h_n(f)$.

The reason for accepting only such high-sigma events is that the event rate of a transient source, i.e., a source lasting for a few seconds to minutes, such as a binary inspiral, could be as low as a few per year, and the noise generated false alarms, at low SNRs $\sim 3 - 4$, over a period of a year, tend to be quite large. Setting higher thresholds for detection helps in removing spurious, noise generated events. However, signal enhancement techniques (cf. Section 5) make it possible to detect a signal of relatively low amplitude, provided there are a large number of wave cycles and the shape of the wave is known accurately.

Real detector noise is neither Gaussian nor stationary and therefore the filtered noise cannot be expected to obey these properties either. One of the most challenging problems is how to remove or veto the false alarm generated by a non-Gaussian and/or nonstationary background. There has been some effort to address the issue of non-Gaussianity [124] and nonstationarity [260]; more work is needed in this direction. However, it is expected that the availability of a network of gravitational wave detectors alleviates the problem to some extent. This is because a high amplitude gravitational wave event will be coincidentally observed in several detectors, although not necessarily with the same SNR, while false alarms are, in general, not coincident, as they are normally produced by independent sources located close to the detectors.

We have seen that coincident observations help to reduce the false alarm rate significantly. The rate can be further reduced, and possibly even nullified, by subjecting coincident events to further consistency checks in a detector network consisting of four or more detectors. As discussed in Section 2, each gravitational wave event is characterized by five kinematic (or extrinsic) observables: location of the source with respect to the detector (D, θ, φ) and the two polarizations ($h_+, h_\times$). Each detector in a network measures a single number, say the amplitude of the wave. In addition, in a network of N detectors, there are $N - 1$ independent time delays in the arrival times of the wave at various detector locations, giving a total of $2N - 1$ observables. Thus, the minimum number of detectors needed to reconstruct the wave and its source is $N = 3$. More than three detectors in a network will have redundant information that will be consistent with the quantities inferred from any three detectors, provided the event is a true coincident event and not a chance coincidence, and most likely a true gravitational

wave event. In a detector network consisting of $N(\geq 4)$ detectors, one can perform $2N - 6$ consistency checks. Such consistency checks further reduce the number of false alarms.

When the shape of a signal is known, matched filtering is the optimal strategy to pull out a signal buried in Gaussian, stationary noise (see Section 5.1). The presence of high-amplitude transients in the data can render the background nonstationary and non-Gaussian, therefore matched filtering is not necessarily an optimal strategy. However, the knowledge of a signal's shape, especially when it has a broad bandwidth, can be used beyond matched filtering to construct a χ^2 veto [28] to distinguish between triggers caused by a true signal from those caused by high-amplitude transients or other artifacts. One specific implementation of the χ^2 veto compares the expected signal spectrum with the real spectrum to quantify the confidence with which a trigger can be accepted to be caused by a true gravitational wave signal and has been the most powerful method for greatly reducing the false alarm rate. We shall discuss the χ^2 veto in more detail in Section 5.1.

7.2 Data Analysis by Sathyaprakash

Observing gravitational waves requires a data analysis strategy, which is in many ways different from conventional astronomical data analysis. There are several reasons why this is so:

- Gravitational wave antennas are essentially omni-directional, with their response better than 50% of the root mean square over 75% of the sky (see Figure 4, right panel, recalling that the rms response is 2/5 of the peak). Hence, data analysis systems will have to carry out all-sky searches for sources.
- Interferometers are typically broadband covering three to four orders of magnitude in frequency. While this is obviously to our advantage, as it helps to track sources whose frequency may change rapidly, it calls for searches to be carried out over a wide range of frequencies.
- In Einstein's theory, gravitational radiation has two independent states of polarization. Measuring polarization is of fundamental importance (as there are other theories of gravity in which the number of polarization states is more than two and in some theories even dipolar and scalar waves exist [392]) and has astrophysical implications too (for example, gravitational-wave-polarization measurement is one way to resolve the mass-inclination degeneracy of binary systems observed electromagnetically, as discussed in Section 7.1.1). Polarization measurements would be possible with a network of detectors, which means analysis algorithms that work with data from multiple antennas will have to be developed. This should also benefit coincidence analysis and the efficiency of event recognition.
- Unlike typical detection techniques for electromagnetic radiation from astronomical sources, most astrophysical gravitational waves are detected coherently, by following the phase of the radiation, rather than just the energy. That is, the SNR is built up by coherent superposition of many wave cycles emitted by a source. The phase evolution contains more information than the amplitude does and the signal structure is a rich source of the underlying physics. Nevertheless, tracking a signal's phase means searches will have to be made not only for specific sources but over a huge region of the parameter space for each source, placing severe demands both on the theoretical understanding of the emitted waveforms as well as on the data analysis hardware.
- Finally, gravitational wave detection is computationally intensive. Gravitational wave antennas acquire data continuously for many years at the rate of several megabytes per

second. About 1% of this data is signal data; the rest is housekeeping data that monitors the operation of the detectors. The large parameter space mentioned above requires that the signal data be filtered many times for different searches, and this puts big demands on computing hardware and algorithms.

Data analysis for broadband detectors has been strongly developed since the mid 1980s [359, 331, 330]. The field has a regular series of annual Gravitational Wave Data Analysis Workshops; the published proceedings are a good place to find current thinking and challenges. Early coincidence experiments with interferometers [273] and bars [29] provided the first opportunities to apply these techniques. Although the theory is now fairly well understood [206], strategies for implementing data analysis depend on available computer resources, data volumes, astrophysical knowledge, and source modeling, and so are under constant revision.

We will begin with a discussion of the matched filtering algorithm and next use it to estimate the SNRs for binary coalescences in various detectors. After that, we will develop the theory of matched filtering further to work out the computational costs to carry out online searches, that is to search at the same rate as the data is acquired. In the final section, we will use the formalism developed in earlier sections to discuss parameter estimation. The foundations of signal analysis lie in the statistics of making "best estimates" of whether a signal is present in noisy data or not. See the Living Review by Jaranowski and Królak [206] for a discussion of this in the gravitational wave context.

7.2.1 Matched filtering and optimal signal-to-noise ratio

Matched filtering is a data analysis technique that efficiently searches for a signal of known shape buried in noisy data [186]. The technique consists in correlating the output of a detector with a waveform, variously known as a template or a filter. Given a signal $h(t)$ buried in noise $n(t)$, the task is to find an 'optimal' template $q(t)$ that would produce, on the average, the best possible SNR. In this review, we shall treat the problem of matched filtering as an operational exercise. However, this intuitive picture has a solid basis in the theory of hypothesis testing. The interested reader may consult any standard text book on signal analysis, for example Helstrom [186], for details.

Let us first fix our notation. We shall use $x(t)$ to denote the detector output, which is assumed to consist of a background noise $n(t)$ and a useful gravitational wave signal $h(t)$. The Fourier transform of a quantity $x(t)$ will be denoted $\tilde{x}(f)$ and is defined as

$$\tilde{x}(f) = \int_{-\infty}^{\infty} x(t) e^{2\pi i f t} dt$$

Optimal filter

The detector output $x(t)$ is just a realization of noise $n(t)$, i.e., $x(t) = n(t)$, when no signal is present. In the presence of a signal $h(t)$ with an arrival time t_a , $x(t)$ takes the form,

$$x(t) = h(t - t_a) + n(t)$$

The correlation c of a template $q(t)$ with the detector output is defined as

$$c(\tau) \equiv \int_{-\infty}^{\infty} x(t) q(t + \tau) dt$$

In the above equation, τ is called the lag; it denotes the duration by which the filter function lags behind the detector output. The purpose of the above correlation integral is to concentrate all the signal energy at one place. The following analysis reveals how this is achieved; we shall work out the optimal filter $q(t)$ that maximizes the correlation $c(\tau)$ when a signal $h(t)$ is present in the detector output. To do this let us first write the correlation integral in the Fourier domain by substituting for $x(t)$ and $q(t)$, in the above integral, their Fourier transforms $\tilde{x}(f)$ and $\tilde{q}(f)$, i.e., $x(t) \equiv \int_{-\infty}^{\infty} \tilde{x}(f) \exp(-2\pi i f t) df$ and $q(t) \equiv \int_{-\infty}^{\infty} \tilde{q}(f) \exp(-2\pi i f t) df$, respectively. After some straightforward algebra, one obtains

$$c(\tau) = \int_{-\infty}^{\infty} \tilde{x}(f) \tilde{q}^*(f) e^{-2\pi i f \tau} df$$

where $\tilde{q}^*(f)$ denotes the complex conjugate of $\tilde{q}(f)$.

Since n is a random process, c is also a random process. Moreover, correlation is a linear operation and hence c obeys the same probability distribution function as n . In particular, if n is described by a Gaussian random process with zero mean, then c is also described by a Gaussian distribution function, although its mean and variance will, in general, differ from those of n . The mean value of c , denoted by $S \equiv \bar{c}$, is, clearly, the correlation of the template q with the signal h , since the mean value of n is zero:

$$S \equiv \bar{c}(\tau) = \int_{-\infty}^{\infty} \tilde{h}(f) \tilde{q}^*(f) e^{-2\pi i f(\tau - t_a)} df$$

The variance of c , denoted $N^2 \equiv \overline{(c - \bar{c})^2}$, turns out to be,

$$N^2 = \overline{(c - \bar{c})^2} = \int_{-\infty}^{\infty} S_h(f) |\tilde{q}(f)|^2 df$$

Now the SNR ρ is defined by $\rho^2 \equiv S^2/N^2$.

The form of integrals in Equations (77) and (78) leads naturally to the definition of the scalar product of waveforms. Given two functions, $a(t)$ and $b(t)$, we define their scalar product $\langle a, b \rangle$ to be [159, 161, 114, 128]

$$\langle a, b \rangle \equiv 2 \int_0^{\infty} \frac{df}{S_h(f)} \left[\tilde{a}(f) \tilde{b}^*(f) + \tilde{a}^*(f) \tilde{b}(f) \right]$$

Note that $S_h(f) \geq 0$ [cf. Equation (67)], consequently, the scalar product is real and positive definite.

Noting that the Fourier transform of a real function $h(t)$ obeys $\tilde{h}(-f) = \tilde{h}^*(f)$, we can write down the SNR in terms of the above scalar product:

$$\rho^2 = \frac{\langle h e^{2\pi i f(\tau - t_a)}, S_h q \rangle}{\sqrt{\langle S_h q, S_h q \rangle}}$$

From this it is clear that the template q that obtains the maximum value of ρ is simply

$$\tilde{q}(f) = \gamma \frac{\tilde{h}(f) e^{i 2\pi f(\tau - t_a)}}{S_h(f)}$$

where γ is an arbitrary constant. From the above expression for an optimal filter we note two important things. First, the SNR is maximized when the lag parameter τ is equal to the time of arrival of the signal t_a . Second, the optimal filter is not just a copy of the signal, but rather it is weighted down by the noise PSD. We will see below why this should be so.

Optimal signal-to-noise ratio

We can now work out the optimal SNR by substituting Equation (81) for the optimal template in Equation (80),

$$\rho_{\text{opt}} = \langle h, h \rangle^{1/2} = 2 \left[\int_0^\infty df \frac{|\tilde{h}(f)|^2}{S_h(f)} \right]^{1/2}$$

We note that the optimal SNR is not just the total energy of the signal (which would be $2 \int_0^\infty df |\tilde{h}(f)|^2$), but rather the integrated signal power weighted down by the noise PSD. This is in accordance with what we would guess intuitively: the contribution to the SNR from a frequency bin where the noise PSD is high is smaller than from a bin where the noise PSD is low. Thus, an optimal filter automatically takes into account the nature of the noise PSD.

The expression for the optimal SNR Equation (82) suggests how one may compare signal strengths with the noise performance of a detector. Note that one cannot directly compare $\tilde{h}(f)$ with $S_h(f)$, as they have different physical dimensions. In gravitational wave literature one writes the optimal SNR in one of the following equivalent ways

$$\rho_{\text{opt}} = 2 \left[\int_0^\infty \frac{df}{f} \frac{|\sqrt{f} \tilde{h}(f)|^2}{S_h(f)} \right]^{1/2} = 2 \left[\int_0^\infty \frac{df}{f} \frac{|f \tilde{h}(f)|^2}{f S_h(f)} \right]^{1/2}$$

which facilitates the comparison of signal strengths with noise performance. One can compare the dimensionless quantities, $f|\tilde{h}(f)|$ and $\sqrt{f S_h(f)}$, or dimensionful quantities, $\sqrt{f}|\tilde{h}(f)|$ and $\sqrt{S_h(f)}$.

Signals of interest to us are characterized by several (a priori unknown) parameters, such as the masses of component stars in a binary, their intrinsic spins, etc., and an optimal filter must agree with both the signal shape and its parameters. A filter whose parameters are slightly mismatched with that of a signal can greatly degrade the SNR. For example, even a mismatch of one cycle in 10^4 cycles can degrade the SNR by a factor two.

When the parameters of a filter and its shape are precisely matched with that of a signal, what is the improvement brought about, as opposed to the case when no knowledge of the signal is available? Matched filtering helps in enhancing the SNR in proportion to the square root of the number of signal cycles in the detector band, as opposed to the case in which the signal shape is unknown and all that can be done is to Fourier transform the detector output and compare the signal energy in a frequency bin to noise energy in that bin. We shall see below that, in initial interferometers, matched filtering leads to an enhancement of order 30-100 for compact binary inspiral signals.

Practical applications of matched filtering

Matched filtering is currently being applied to mainly two sources: detection of (1) chirping signals from compact binaries consisting of black holes and/or neutron stars and (2) continuous waves from rapidly-spinning neutron stars.

5.1.3.1 Coalescing binaries. In the case of chirping binaries, post-Newtonian theory (a perturbative approximation to Einstein's equations in which the relevant quantities are expanded

as a power-series in $1/c$, where c is the speed of light) has been used to model the dynamics of these systems to a very high order in v/c , where v is the relative speeds of the objects in the binary (see also Section 6.5, in which binaries are discussed in more detail). This is an approximation that can be effectively used to match filter the signal from binaries whose component bodies are of equal, or nearly equal, masses and the system is still "far" from coalescence. In reality, one takes the waveform to be valid until the last stable circular orbit (LSCO). In the case of binaries consisting of two neutron stars, or a neutron star and a black hole, tidal effects might affect the evolution significantly before reaching the LSCO. However, this is likely to occur at frequencies well-above the sensitivity band of the current ground-based detectors, so that for all practical purposes post-Newtonian waveforms are a good approximation to low-mass ($M < 10M_\odot$) binaries.

As elucidated in Section 6.5.2, progress in analytical and numerical relativity has made it possible to have a set of waveforms for the merger phase of compact binaries too. The computational cost in matched filtering the merger phase, however, will not be high, as there will only be on the order of a few 100 cycles in this phase. But it is important to have the correct waveforms to enhance signal visibility and, more importantly, to enable strong-field tests of general relativity.

In the general case of black-hole-binary inspiral the search space is characterized by 17 different parameters. These are the two masses of the bodies, their spins, eccentricity of the orbit, its orientation at some fiducial time, the position of the binary in the sky and its distance from the Earth, the epoch of coalescence and phase of the signal at that epoch, and the polarization angle.

However, not all these parameters are important in a matched filter search. Only those parameters that change the shape of the signal, such as the masses, orbital eccentricity and spins, or cause a modulation in the signal due to the motion of the detector relative to the source, such as the direction to the source, are to be searched for and others, such as the epoch of coalescence and the phase at that epoch, are simply reference points in the signal that can be determined without any significant burden on computational power.

For binaries consisting of nonspinning objects that are either observed for a short enough period that the detector motion can be neglected, or last for only a short time in the sensitive part of a detector's sensitivity band, there are essentially two search parameters - the component masses of the binary. It turns out that the signal manifold in this case is nearly flat, but the masses are curvilinear coordinates and are not good parameters for choosing templates. Chirp times, which are nonlinear functions of the masses, are very close to being Cartesian coordinates and template spacing is more or less uniform in terms of these parameters. Chirp times are post-Newtonian contributions at different orders to the duration of a signal starting from a time when the instantaneous gravitational-wave frequency has a fiducial value f_L to a time when the gravitational wave frequency formally diverges and system coalesces. For instance, the chirp times τ_0 and τ_3 at Newtonian and 1.5 PN orders, respectively, are

$$\tau_0 = \frac{5}{256\pi\nu f_L} (\pi M f_L)^{-5/3}, \quad \tau_3 = \frac{1}{8\nu f_L} (\pi M f_L)^{-2/3}$$

where M is the total mass and $\nu = m_1 m_2 / M^2$ is the symmetric mass ratio. The above relations can be inverted to obtain M and ν in terms of the chirp times:

$$M = \frac{5}{32\pi^2 f_L \tau_0} \tau_3, \quad \nu = \frac{1}{8\pi f_L \tau_3} \left(\frac{32\pi \tau_0}{5\tau_3} \right)^{2/3}$$

There is a significant amount of literature on the computational requirements to search for compact binaries [321, 145, 277, 279]. The estimates for initial detectors are not alarming and it is possible to search for these systems online. Searches for these systems by the LSC (see,

for example, [8]) employs a hexagonal lattice of templates [118] in the two-dimensional space of chirp times. For the best LIGO detectors we need several thousand templates to search for component masses in the range $[m_{\text{low}}, m_{\text{high}}] = [1, 100]M_{\odot}$ [279]. Decreasing the lower-end of the mass range leads to an increase in the number of templates that goes roughly as $m_{\text{low}}^{-8/3}$ and most current searches [2, 6] only begin at $m_{\text{low}} = 1M_{\odot}$, with the exception of one that looked for black hole binaries of primordial origin [7], in which the lower end of the search was $0.2M_{\odot}$.

Inclusion of spins is only important when one or both of the components is rapidly spinning [38, 95]. Spins effects are unimportant for neutron star binaries, for which the dimensionless spin parameter q , that is the ratio of its spin magnitude to the square of its mass, is tiny: $q = J_{\text{NS}}/M_{\text{NS}}^2 \ll 1$. For ground-based detectors, even after including spins, the computational costs, while high, are not formidable and it should be possible to carry out the search on large computational clusters in real time [95]. Recently, the LSC has successfully carried out such a search [15].

5.1.3.2 Searching for Continuous Wave Signals. In the case of continuous waves (CWs), the signal shape is pretty trivial: a sinusoidal oscillation with small corrections to take account of the slow spin-down of the neutron star/pulsar to account for the loss of angular momentum to gravitational waves and other radiation/particles. However, what leads to an enormous computational cost here is the Doppler modulation of the signal caused by Earth's rotation, the motion of the Earth around the solar system barycenter and the moon. The number of independent patches that we have to observe so as not to lose appreciable amounts of SNR can be worked out in the following manner. The baseline of a gravitational wave detector for CW sources is essentially $L = 2 \times 1AU \simeq 3 \times 10^{11}$ m. For a source that emits gravitational waves at 100 Hz, the wavelength of the radiation is $\lambda = 3 \times 10^6$ m, and the angular resolution $\Delta\theta$ of the antenna at an SNR of 1 is $\Delta\theta \simeq \lambda/L = 10^{-5}$, or a solid angle of $\Delta\Omega \simeq (\Delta\theta)^2 = 10^{-10}$. In other words, the number of patches one should search for is $N_{\text{patches}} \sim 4\pi/\Delta\Omega \simeq 10^{11}$. Moreover, for an observation that lasts for about a year ($T \simeq 3 \times 10^7$ s) the frequency resolution is $\Delta f = 1/T \simeq 3 \times 10^{-8}$. Searching over a frequency band of 300 Hz, around the best sensitivity of the detector, gives the number of frequency bins to be about 10^{10} . Thus, it is necessary to search over roughly 10^{11} patches in the sky for each of the 10^{10} frequency bins. This is a formidable task and one can only perform a matched filter search over a short period (days/weeks) of the data or over a restricted region in the sky, or just perform targeted searches for known objects such as pulsars, the galactic center, etc. [90].

The severe computational burden faced in the case of CW searches has led to the development of specialized searches that look for signals from known pulsars [5, 10, 12] using an efficient search algorithm that makes use of the known parameters [115, 150] and hierarchical algorithms that add power incoherently with the minimum possible loss in signal visibility [225, 4, 339, 11]. The most ambitious project in this regard is the Einstein@Home project [369]. The goal here is to carry out coherent searches for CW signals using wasted CPUs on idle computers at homes, offices and university departments around the world. The project has been successful in attracting a large number of subscriptions and provides the largest computational infrastructure to the LSC for the specific search of CW signals and the first scientific results from such are now being published by the LSC [13].

5.1.3.3 χ^2 veto. Towards the end of Section 4.8 we discuss a powerful way of rejecting triggers, whose root cause is not gravitational wave signals but false alarms due to instrumental and environmental artifacts. In this section we will further quantify the χ^2 veto [28] by using the scalar product introduced in the context of matched filtering.

The main problem with real data is that it can be glitchy in the form of high amplitude transients that might look like damped sinusoids. An inspiral signal and a template employed to detect it are both broadband signals. Therefore, the matched-filter SNR for such signals has contributions from a wide range of frequencies. However, the statistic of matched filtering,

namely the SNR, is an integral over frequency and it is not sensitive to contributions from different frequency regions. Imagine dividing the frequency range of integration into a finite number of bins $f_k \leq f < f_{k+1}$, $k = 1, \dots, p$, such that their union spans the entire frequency band, $f_1 = 0$ and $f_{p+1} = \infty$, and further that the contribution to the SNR from each frequency bin is the same, namely,

$$4 \int_{f_k}^{f_{k+1}} \frac{|\tilde{h}(f)|^2}{S_h(f)} df = \frac{4}{p} \int_0^\infty \frac{|\tilde{h}(f)|^2}{S_h(f)} df$$

Now, define the contribution to the matched filtering statistic coming from the k -th bin by [28]

$$z_k \equiv \langle q, x \rangle_k \equiv 2 \int_{f_k}^{f_{k+1}} [\tilde{q}^*(f)\tilde{x}(f) + \tilde{q}(f)\tilde{x}^*(f)] \frac{df}{S_h(f)},$$

where, as before, $\tilde{x}(f)$ and $\tilde{q}(f)$ are the Fourier transforms of the detector output and the template, respectively. Note that the sum $z = \sum_k z_k$ gives the full matched filtering statistic [28]:

$$z = \langle q, x \rangle \equiv 2 \int_0^\infty [\tilde{q}^*(f)\tilde{x}(f) + \tilde{q}(f)\tilde{x}^*(f)] \frac{df}{S_h(f)}$$

Having chosen the bins and quantities z_k as above, one can construct a statistic based on the measured SNR in each bin as compared to the expected value, namely

$$\chi^2 = p \sum_{k=1}^p \left(z_k - \frac{z}{p} \right)^2$$

When the background noise is stationary and Gaussian, the quantity χ^2 obeys the well-known chi-square distribution with $p-1$ degrees of freedom. Therefore, the statistical properties of the χ^2 statistic are known. Imagine two triggers with identical SNRs, but one caused by a true signal and the other caused by a glitch that has power only in a small frequency range. It is easy to see that the two triggers will have very different χ^2 values; in the first case the statistic will be far smaller than in the second case. This statistic has served as a very powerful veto in the search for signals from coalescing compact binaries and it has been instrumental in cleaning up the data (see, e.g., [2, 6]).

7.2.2 Suboptimal filtering methods

It is not always possible to compute the shape of the signal from a source. For instance, there is no computation, numerical or analytical, that reliably gives us the highly relativistic and nonlinear dynamics of gravitational collapse, the supernova that follows it and the emitted gravitational signal. The biggest problem here is the unknown physical state of the pre-supernova star and the complex physics that is involved in the collapse and explosion. Thus, matched filtering cannot be used to detect signals from supernovae.

Even when the waveform is known, the great variety in the shape of the emitted signals might render matched filtering ineffective. In binaries, in which one of the component masses is much smaller than the other, the smaller body will evolve on a highly precessing and in some cases eccentric orbit, due to strong spin-orbit coupling. Moreover, the radiation backreaction

effects, which in the case of equal mass binaries are computed in an approximate way by averaging over an orbital time scale, should be computed much more accurately. The resulting motion of the small body in the Kerr spacetime of the larger body is extremely complicated, leading to a waveform that is rather complex and matched filtering would not be a practical approach.

Suboptimal methods can be used in such cases and they have a twofold advantage: they are less sensitive to the shape of the signal and are computationally significantly cheaper than matched filtering. Of course, the price is a loss in the SNR. The best suboptimal methods are sensitive to signal amplitudes a factor of two to three larger than that required by matched filtering and a factor of 10 to 30 in volume.

Most suboptimal techniques are one form of time-frequency transform or the other. They determine the presence or absence of a signal by comparing the power over a small volume in the time-frequency plane in a given segment of data to the average power in the same volume over a large segment of data. The time-frequency transform $q(\tau, f)$ of data $x(t)$ using a window $w(t)$ is defined as

$$q(\tau, f) = \int_{-\infty}^{\infty} w(t - \tau)x(t)e^{2\pi ift} dt$$

The window function $w(t - \tau)$ is centered at $t = \tau$, and one obtains a time-frequency map by moving the window from one end of a data segment to the other. The window is not unique and the effectiveness of a window depends on the signal one is looking for. Once the time-frequency map is constructed, one can look for excess power (compared to average) in different regions [33], or look for certain patterns.

The method followed depends on the signal one is looking for. For instance, when looking for unknown signals, all that can be done is to look for a departure from averaged behavior in different regions of the map [33]. However, when some knowledge of the spectral and temporal content of the signal is known, it is possible to tune the algorithm to improve efficiency. The wavelet-based waveburst algorithm is one such example [217] that has been applied to search for unstructured bursts in LIGO data [9].

One can employ strategies that improve detection efficiency over a simple search for excess power. For example, chirping signals will leave a characteristic track in the time-frequency plane, with increasing frequency and power as a function of time. Time frequency map of a chirp signal buried in noisy data is shown in Figure 7. An algorithm that optimizes the search for specific shapes in the time-frequency plane is discussed in [187]. These and other methods have been applied to understand how to analyze LISA data [172, 389].

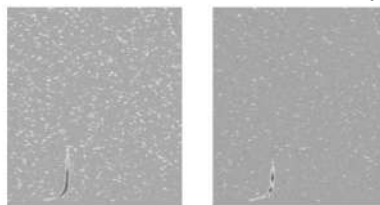


Figure 7: Time-frequency maps showing the track left by the inspiral of a small black hole falling into an SMBH as expected in LISA data. The left panel is for a central black hole without spin and the right panel is for a central black hole whose dimensionless spin parameter is $q = 0.9$.

More recently, there has been a lot of progress in extending burst search algorithms for a network of detectors [113, 218], as well as exploring new Bayesian-based methods to search for unknown transients [335].

7.2.3 Measurement of parameters and source reconstruction

We have so far focused on the problem of detection and have not discussed parameter estimation in any concrete way. In principle, parameter estimation should not be considered to be distinct from detection. From a practical point of view, however, certain methods might be computationally efficient in detecting a signal but not necessarily the best for parameter estimation, while the best parameter estimation methods might not be computationally efficient. Thus, the problem of parameter estimation is often treated separately from detection.

The first thing to realize is that we can never be absolutely certain that a signal is present in a data train [159, 161]; we can only give confidence levels about its presence, which could be close to 100% for high values of the SNR. The next thing to realize is that, whatever the SNR may be, we cannot be absolutely certain about the true parameters of the signal: at best we can make an estimate and these estimates are given in a certain range. The width of the range depends on the confidence level required, being larger for higher confidence levels [159].

Maximum likelihood estimates have long been used to measure the parameters of a known signal buried in noisy data. The method consists in maximizing the likelihood ratio - the ratio of the probability that a given signal is present in the data to the probability that the signal is absent [186, 159]. Maximum likelihood estimates are not always minimum uncertainty estimates, as has been particularly demonstrated in the case of binary inspiral signals by Balasubramanian, et al. [64, 65]. However, until recently, this is the method that has been very widely followed in the gravitational wave literature. But what is important to note is that maximum likelihood estimates are unbiased when the SNR is large ³, and the mean of the distribution of measured values of the parameters will be centered around the true parameter values. This is an important quality that will be useful in our discussion below.

Bayesian estimates, which take into account any prior knowledge that may be available about the distribution of the source parameters, often give much better estimates and do not rely on the availability of an ensemble of detector outputs [340, 274]. However, they are computationally a lot more expensive than maximum likelihood estimates.

In any one measurement, any estimated parameters, however efficient, robust and accurate, are unlikely to be the actual parameters of the signal, since, at any finite SNR, noise alters the input signal. In the geometric language, the signal vector is being altered by the noise vector and our matched filtering aims at computing the projection of this altered vector onto the signal space. The true parameters are expected to lie within an ellipsoid of p dimensions at a certain confidence level - the volume of the ellipsoid increasing with the confidence level at a given SNR but decreasing with the SNR at a given confidence level.

Ambiguity function

The ambiguity function, well known in the statistical theory of signal detection [186], is a very powerful tool in signal analysis: it helps one to assess the number of templates required to span the parameter space of the signal [321], to make estimates of variances and covariances involved in the measurement of various parameters, to compute biases introduced in using a family of templates whose shape is not the same as that of a family of signals intended to be detected, etc. We will see below how the ambiguity function can be used to compute the required number of templates. Towards the end of this section we will use the ambiguity function for the estimation of parameters.

The ambiguity function is defined (see Equation (91) below) as the scalar product of two normalized waveforms maximized over the initial phase of the waveform, in other words, the absolute value of the scalar product ⁴. A waveform e is said to be normalized if $\langle e, e \rangle^{1/2} = 1$, where the inner product is inversely weighted by the PSD, as in Equation (79). Among other things, normalized waveforms help in defining signal strengths: a signal is said to be of strength

h_0 if $h = h_0 e$. Note that the optimal SNR for such a signal of strength h_0 is, $\langle h, h \rangle^{1/2} = h_0$.

Let $e(t; \alpha)$, where $\alpha = \{\alpha^i \mid i = 0, \dots, p\}$ is the parameter vector comprised of $p+1$ parameters, denote a normalized waveform. It is conventional to choose the parameter α^0 to be the lag τ , which simply corresponds to a coordinate time when an event occurs and is therefore called an extrinsic parameter, while the rest of the p parameters are called the intrinsic parameters and characterize the gravitational wave source.

Given two normalized waveforms $e(t; \alpha)$ and $e(t; \beta)$, whose parameter vectors are not necessarily the same, the ambiguity $\mathcal{A}$ is defined as

$$\mathcal{A}(\alpha, \beta) \equiv |\langle e(\alpha), e(\beta) \rangle|$$

Since the waveforms are normalized, $\mathcal{A}(\alpha, \alpha) = 1$ and $\mathcal{A}(\alpha, \beta) < 1$, if $\alpha \neq \beta$. Here, α can be thought of as the parameters of a (normalized) template while β those of a signal. With the template parameters α fixed, the ambiguity function is a function of p signal parameters β^i , giving the SNR obtained by the template for different signals. The region in the signal parameter space for which a template obtains SNRs larger than a chosen value (called the minimal match [277]) is the span of that template. Template families should be chosen so that altogether they span the entire signal parameter space of interest with the least overlap of one other's spans. One can equally well interpret the ambiguity function as the SNR obtained for a given signal by filters of different parameter values.

It is clear that the ambiguity function is a local maximum at the "correct" set of parameters, $\beta = \alpha$. Search methods that vary β to find the best fit to the parameter values make use of this property in one way or another. But the ambiguity function will usually have secondary maxima as a function of β with fixed α . If these secondaries are only slightly smaller than the primary maximum, then noise can lead to confusion: it can, at random, sometimes elevate a secondary and suppress a primary. These can lead to false measurements of the parameters. Search methods need to be designed carefully to avoid this as much as possible. One way would be to fit the known properties of the ambiguity function to an ensemble of maxima. This would effectively average over the noise on individual peaks and point more reliably to the correct one.

It is important to note that in the definition of the ambiguity function there is no need for the functional forms of the template and signal to be the same; the definition holds true for any signal/template pair of waveforms. Moreover, the number of template parameters need not be identical (and usually aren't) to the number of parameters characterizing the signal. For instance, a binary can be characterized by a large number of parameters, such as the masses, spins, eccentricity of the orbit, etc., while we may take as a model waveform the one involving only the masses. In the context of inspiral waves, $e(t; \beta)$ is the exact general relativistic waveform emitted by a binary, whose form we do not know, while the template family is a post-Newtonian, or some other, approximation to it, that will be used to detect the true waveform. Another example would be signals emitted by spinning neutron stars, isolated or in binaries, whose time evolution is unknown, either because we cannot anticipate all the physical effects that affect their spin, or because the parameter space is so large that we cannot possibly take into account all of them in a realistic search.

Of course, in such cases we cannot compute the ambiguity function, since one of the ar-

³ How large the SNR should be to presume that there is no bias in the estimation of parameters depends on the number of parameter-space dimensions and strictly speaking the statement is true only in the limit as $\text{SNR} \rightarrow \infty$.

⁴ Working with analytic signals $h(t) = a(t)e^{\phi(t) + i\phi_0}$, where $a(t)$ and $\phi(t)$ are the time-varying amplitude and phase of the signal, respectively, we see that the initial phase ϕ_0 of the signal simply factors out as a constant phase in the Fourier domain and we can maximize over this initial phase by simply taking the absolute value of the scalar product of a template with a signal.

guments to the ambiguity function is unknown. These are, indeed, issues where substantial work is called for. What are all the physical effects to be considered so as not to miss out a waveform from our search? How to make a choice of templates when the functional form of templates is different from those of signals? For this review it suffices to assume that the signal and template waveforms are of identical shape and the number of parameters in the two cases is the same.

Metric on the space of waveforms

The computational cost of a search and the estimation of parameters of a signal afford a lucid geometrical picture developed by Balasubramanian et al. [65] and Owen [277]. Much of the discussion below is borrowed from their work.

Let $x_k, k = 1, 2, \dots, N$, denote the discretely sampled output of a detector. The set of all possible detector outputs satisfy the usual axioms of a vector space. Therefore, x_k can be thought of as an N -dimensional vector. It is more convenient to work in the continuum limit, in which case we have infinite dimensional vectors and the corresponding vector space. However, all the results are applicable to the realistic case in which detector outputs are treated as finite dimensional vectors.

Amongst all vectors, of particular interest are those corresponding to gravitational waves from a given astronomical source. While every signal can be thought of as a vector in the infinitedimensional vector space of the detector outputs, the set of all such signal vectors do not, by themselves, form a vector space. However, the set of all normed signal vectors (i.e., signal vectors of unit norm) form a manifold, the parameters of the signal serving as a coordinate system [64, 65, 277, 279]. Thus, each class of an astronomical source forms an n -dimensional manifold $\mathcal{S}_n$, where n is the number of independent parameters characterizing the source. For instance, the set of all signals from a binary on a quasi-circular orbit inclined to the line of sight at an angle ι , consisting of nonspinning black holes of masses m_1 , and m_2 , located a distance D from the Earth ⁵ initially in the direction (θ, φ) and expected to merge at a time t_C with the phase of the signal at merger φ_C , forms a nine-dimensional manifold with coordinates $\{D, \theta, \varphi, m_1, m_2, t_C, \varphi_C, \iota, \psi\}$, where ψ is the polarization angle of the signal. In the general case of a signal characterized by n parameters we shall denote the parameters by p^α , where $\alpha = 1, \dots, n$.

The manifold $\mathcal{S}_n$ can be endowed with a metric $g_{\alpha\beta}$ that is induced by the scalar product defined in Equation (79). The components of the metric in a coordinate system p^α are defined by ⁶

$$g_{\alpha\beta} \equiv \left\langle \partial_\alpha \hat{h}, \partial_\beta \hat{h} \right\rangle, \quad \partial_\alpha \hat{h} \equiv \frac{\partial \hat{h}}{\partial p^\alpha}$$

The metric can then be used on the signal manifold as a measure of the proper distance $d\ell$ between nearby signals with coordinates p^α and $p^\alpha + dp^\alpha$, that is signals $\hat{h}(p^\alpha)$ and $\hat{h}(p^\alpha + dp^\alpha)$,

$$d\ell^2 = g_{\alpha\beta} dp^\alpha dp^\beta$$

Now, by Taylor expanding $\hat{h}(p^\alpha + dp^\alpha)$ around p^α , and keeping only terms to second order in dp^α , it is straightforward to see that the overlap $\mathcal{O}$ of two infinitesimally close signals can be computed using the metric:

$$\begin{aligned} \mathcal{O}(dp^\alpha; p^\alpha) &\equiv \left\langle \hat{h}(p^\alpha), \hat{h}(p^\alpha + dp^\alpha) \right\rangle \\ &= 1 - \frac{1}{2} g_{\alpha\beta} dp^\alpha dp^\beta \end{aligned}$$

The metric on the signal manifold is nothing but the well-known Fisher information matrix usually denoted $\Gamma_{\alpha\beta}$, (see, e.g., [186, 283]) but scaled down by the square of the SNR, i.e., $g_{\alpha\beta} = \rho^{-2}\Gamma_{\alpha\beta}$. The information matrix is itself the inverse of the covariance matrix $C_{\alpha\beta}$ and is a very useful quantity in signal analysis.

Covariance matrix

Having defined the metric, we next consider the application of the geometric formalism in the estimation of statistical errors involved in the measurement of the parameters. We closely follow the notation of Finn and Chernoff [159, 161, 114].

Let us suppose a signal of known shape with parameters p^α is buried in background noise that is Gaussian and stationary. Since the signal shape is known, one can use matched filtering to dig the signal out of the noise. The measured parameters $\bar{p}^\alpha$ will, in general, differ from the true parameters of the signal⁷. Geometrically speaking, the noise vector displaces the signal vector and the process of matched filtering projects the (noise + signal) vector back on to the signal manifold. Thus, any nonzero noise will make it impossible to measure the true parameters of the signal. The best one can hope for is a proper statistical estimation of the influence of noise.

The posterior probability density function $\mathcal{P}$ of the parameters $\bar{p}^\alpha$ is given by a multivariate Gaussian distribution⁸:

$$\mathcal{P}(\Delta p^\alpha) d^n \Delta p = \frac{d^n \Delta p}{(2\pi)^{n/2} \sqrt{C}} \exp \left[-\frac{1}{2} C_{\alpha\beta}^{-1} \Delta p^\alpha \Delta p^\beta \right]$$

where n is the number of parameters, $\Delta p^\alpha = p^\alpha - \bar{p}^\alpha$, and $C_{\alpha\beta}$ is the covariance matrix, C being its determinant. Noting that $C_{\alpha\beta}^{-1} = \rho^2 g_{\alpha\beta}$, we can rewrite the above distribution as

$$\mathcal{P}(\Delta p^\alpha) d^n \Delta p = \frac{\rho^n \sqrt{g} d^n \Delta p}{(2\pi)^{n/2}} \exp \left[-\frac{\rho^2}{2} g_{\alpha\beta} \Delta p^\alpha \Delta p^\beta \right]$$

where we have used the fact that $C = 1/(\rho^{2n} g)$, g being the determinant of the metric $g_{\alpha\beta}$. Note that if we define new parameters $p^\alpha = \rho p^\alpha$, then we have exactly the same distribution function for all SNRs, except that the deviations Δp^α are scaled by ρ .

Let us first specialize to one dimension to illustrate the region of the parameter space with which one should associate an event at a given confidence level. In one dimension the distribution of the deviation from the mean of the measured value of the parameter p is given by

$$\mathcal{P}(\Delta p) d\Delta p = \frac{d\Delta p}{\sqrt{2\pi}\sigma} \exp \left(-\frac{\Delta p^2}{2\sigma^2} \right) = \frac{\rho \sqrt{g_{pp}} d\Delta p}{\sqrt{2\pi}} \exp \left(-\frac{\rho^2}{2} g_{pp} \Delta p^2 \right)$$

where, analogous to the n -dimensional case, we have used $\sigma^2 = 1/(\rho^2 g_{pp})$. Now, at a given SNR, what is the volume V_P in the parameter space, such that the probability of finding the measured parameters $\bar{p}$ inside this volume is P ? This volume is defined by

$$P = \int_{\Delta p \in V_P} \mathcal{P}(\Delta p) d\Delta p$$

⁷ Even though we deal with normed signals (which amounts to fixing D), astrophysical gravitational wave signals are characterized by this additional parameter.

⁸ We have followed the definition of the metric as is conventional in parameter estimation theory (see, e.g., [159, 161, 114, 65]), which differs from that used in template placement algorithms (see, e.g., [277]) by a factor of two. This difference will impact the relationship between the metric and the match, as will be apparent in what follows.

Although V_P is not unique, it is customary to choose it to be centered around $\Delta p = 0$:

$$P = \int_{(\Delta p/\sigma)^2 \leq r^2(P)} \frac{d\Delta p}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\Delta p^2}{2\sigma^2}\right) = \int_{\rho^2 g_{pp} \Delta p^2 \leq r^2(P)} \frac{\rho \sqrt{g_{pp}} d\Delta p}{\sqrt{2\pi}} \exp\left(-\frac{\rho^2 g_{pp} \Delta p^2}{2}\right)$$

where, given P , the above equation can be used to solve for $r(P)$ and it determines the range of integration: $-r\sigma \leq \Delta p \leq r\sigma$. For instance, the volumes V_P corresponding to $P \simeq 0.683, 0.954, 0.997, \dots$, are the familiar intervals $[-\sigma, \sigma], [-2\sigma, 2\sigma], [-3\sigma, 3\sigma], \dots$, and the corresponding values of r are 1, 2, 3. Since $\sigma = 1/\sqrt{\rho^2 g_{pp}}$, we see that in terms of g_{pp} the above intervals translate to

$$\frac{1}{\rho} \left[-\frac{1}{\sqrt{g_{pp}}}, \frac{1}{\sqrt{g_{pp}}} \right], \frac{1}{\rho} \left[-\frac{2}{\sqrt{g_{pp}}}, \frac{2}{\sqrt{g_{pp}}} \right], \frac{1}{\rho} \left[-\frac{3}{\sqrt{g_{pp}}}, \frac{3}{\sqrt{g_{pp}}} \right], \dots$$

Thus, for a given probability P , the volume V_P shrinks as $1/\rho$. The maximum distance $d_{\max}$ within which we can expect to find "triggers" at a given P depends inversely on the SNR ρ : $d\ell = \sqrt{g_{pp} \Delta p^2} = r/\rho$. Therefore, for $P \simeq 0.954, r = 2$ and at an SNR of 5 the maximum distance is 0.4, which corresponds to a match of $\epsilon = 1 - \frac{1}{2}d\ell^2 = 0.92$. In other words, in one dimension 95% of the time we expect our triggers to come from templates that have an overlap greater than or equal to 0.92 with the buried signal when the SNR is five. This interpretation in terms of the match is a good approximation as long as $d\ell \ll 1$, which will be true for large SNR events. However, for weaker signals and/or greater values of P we can't interpret the results in terms of the match, although Equation (98) can be used to determine $r(P)$. As an example, at $P \simeq 0.997, r = 3$ and at an SNR of $\rho = 4$, the maximum distance is $d\ell = 0.75$ and the match is $\epsilon = 23/32 \simeq 0.72$, which is significantly smaller than one and the quadratic approximation is not good enough to compute the match.

These results generalize to n dimensions. In n -dimensions the volume V_P is defined by

$$P = \int_{\Delta p^\alpha \in V_P} \mathcal{P}(\Delta p^\alpha) d^n \Delta p$$

Again, V_P is not unique but it is customary to center the volume around the point $\Delta p^\alpha = 0$:

$$P = \int_{\rho^2 g_{\alpha\beta} \Delta p^\alpha \Delta p^\beta \leq r^2(P,n)} \frac{\rho^n \sqrt{g} d^n \Delta p}{(2\pi)^{n/2}} \exp\left[-\frac{\rho^2}{2} g_{\alpha\beta} \Delta p^\alpha \Delta p^\beta\right]$$

Given P and the parameter space dimension n , one can iteratively solve the above equation for $r(P, n)$. The volume V_P is the surface defined by the equation

$$g_{\alpha\beta} \Delta p^\alpha \Delta p^\beta = \left(\frac{r}{\rho}\right)^2$$

This is the equation of an n -dimensional ellipsoid whose size is defined by r/ρ . For a given r (which determines the confidence level), the size of the ellipsoid is inversely proportional to the SNR, the volume decreasing as ρ^n . However, the size is not small enough for all combinations of P and ρ to interpret the distance from the center of the ellipsoid to its surface, in terms of the overlap or match of the signals at the two locations, except when the distance is close to zero.

⁷ In what follows we shall use an over-line to distinguish the measured parameters from the true parameters p^α .

⁸ A Bayesian interpretation of $\mathcal{P}(\Delta p^\alpha)$ is the probability of having the true signal parameters lie somewhere inside the ellipsoidal volume centered at the Maximum Likelihood point $\bar{p}^\alpha$.

This is because the expression for the match in terms of the metric is based on the quadratic approximation, which breaks down when the matches are small. However, the region defined by Equation (103) always corresponds to the probability P and there is no approximation here (except that the detector noise is Gaussian).

When the SNR ρ is large and $1 - P$ is not close to zero, the triggers are found from the signal with matches greater than or equal to $1 - \frac{r^2(P,n)}{2\rho^2}$. Table 2 lists the value of r for several values of P in one, two and three-dimensions and the minimum match ϵ_{MM} for SNRs 5, 10 and 20.

Table 2 should be interpreted in light of the fact that triggers come from an analysis pipeline in which the templates are laid out with a certain minimal match and one cannot, therefore, expect the triggers from different detectors to be matched better than the minimal match.

From Table 2, we see that, when the SNR is large (say greater than about 10), the dependence of the match ϵ_{MM} on n is very weak; in other words, irrespective of the number of dimensions, we expect the match between the trigger and the true signal (and for our purposes the match between triggers from different instruments) to be pretty close to 1, and mostly larger than a minimal match of about 0.95 that is typically used in a search. Even when the SNR is in the region of 5, for low P again there is a weak dependence of ϵ_{MM} on the number of parameters. For large P and low SNR, however, the dependence of ϵ_{MM} on the number of dimensions becomes important. At an SNR of 5 and $P \simeq 0.997$, $\epsilon_{\text{MM}} = 0.91, 0.87, 0.85$ for $n = 1, 2, 3$ dimensions, respectively.

Bounds on the estimation computed using the covariance matrix are called Cramér-Rao bounds. Cramér-Rao bounds are based on local analysis and do not take into consideration the effect of distant points in the parameter space on the errors computed at a given point, such as the

Table 2: The value of the (squared) distance $d\ell^2 = r^2/\rho^2$ for several values of P and the corresponding smallest match that can be expected between templates and the signal at different values of the SNR.

	$P = 0.683$		$P = 0.954$		$P = 0.997$	
ρ	$d\ell^2$	ϵ_{MM}	$d\ell^2$	ϵ_{MM}	$d\ell^2$	ϵ_{MM}
$n = 1$						
5	0.04	0.9899	0.16	0.9592	0.36	0.9055
10	0.01	0.9975	0.04	0.9899	0.09	0.9772
20	0.0025	0.9994	0.01	0.9975	0.0225	0.9944
$n = 2$						
5	0.092	0.9767	0.2470	0.9362	0.4800	0.8718
10	0.023	0.9942	0.0618	0.9844	0.1200	0.9695
20	0.00575	0.9986	0.0154	0.9961	0.0300	0.9925
$n = 3$						
5	0.1412	0.9641	0.32	0.9165	0.568	0.8462
10	0.0353	0.9911	0.08	0.9798	0.142	0.9638
20	0.00883	0.9978	0.02	0.9950	0.0355	0.9911

secondary maxima in the likelihood. Though the Cramér-Rao bounds are in disagreement with maximum likelihood estimates, global analysis, taking the effect of distant points on the estimation of parameters, does indeed give results in agreement with maximum likelihood estimation as shown by Balasubramanian and Dhurandhar [63].

Bayesian inference

A good example of an efficient detection algorithm that is not a reliable estimator is the timefrequency transform of a chirp. For signals that are loud enough, a time-frequency trans-

form of the data would be a very effective way of detecting the signal, but the transform contains hardly any information about the masses, spins and other information about the source. This is because the time-frequency transform of a chirp is a mapping from the multi-dimensional (17 in the most general case) space of chirps to just the two-dimensional space of time and frequency. Even matched filtering, which would use templates that are defined on the full parameter space of the signal, would not give the parameters at the expected accuracy. This is because the templates are defined only at a certain minimal match and might not resolve the signal well enough, especially for signals that have a high SNR.

In recent times Bayesian inference techniques have been applied with success in many areas in astronomy and cosmology. These techniques are probably the most sensible way of estimating the parameters, and the associated errors, but cannot be used to efficiently search for signals. Bayesian inference is among the simplest of statistical measures to state, but is not easy to compute and is often subject to controversies. Here we shall only discuss the basic tenets of the method and refer the reader for details to an excellent treatise on the subject (see, e.g., Sivia [340]).

To understand the chief ideas behind Bayesian inference, let us begin with some basic concepts in probability theory. Given two hypothesis or statements A and B about an observation, let $P(A, B)$ denote the joint probability of A and B being true. For the sake of clarity, let A denote a statement about the universe and B some observation that has been made. Now, the joint probability can be expressed in terms of the individual probability densities $P(A)$ and $P(B)$ and conditional probability densities $P(A | B)$ and $P(B | A)$ as follows:

$$P(A, B) = P(A)P(B | A) \quad \text{or} \quad P(A, B) = P(B)P(A | B)$$

The first of these equations says the joint probability of A and B both being true is the probability that A is true times the probability that B is true given that A is true and similarly the second. We can use the above equations to arrive at Bayes theorem:

$$P(B)P(A | B) = P(A)P(B | A) \quad \text{or} \quad P(A | B) = \frac{P(A)P(B | A)}{P(B)}$$

The left-hand side of the above equation can be interpreted as a statement about A (the universe) given B (the data). This is the posterior probability density. The right-hand side contains $P(B | A)$, which is the probability that B is obtained given that A is true and is called the likelihood, $P(A)$, which is the probability of A , called the prior probability of A , and $P(B)$ (the prior of B), which is simply a normalizing constant often ignored in Bayesian analysis.

For instance, if A denotes the statement it is going to rain and B the amount of humidity in the air then the above equation gives us the posterior probability that it rains when the air contains a certain amount of humidity. Clearly, the posterior depends on what is the likelihood of the air having a certain humidity when it rains and the prior probability of rain on a given day. If the prior is very small (as it would be in a desert, for example) then you would need a rather large likelihood for the posterior to be large. Even when the prior is not so small, say a 50% chance of rain on any given day (as it would be if you are in Wales), the likelihood has to be large for posterior probability to say something about the relationship between the level of humidity and the chance of rain.

As another example, and more relevant to the subject of this review, let s be the statement the data contains a chirp (signal), n the statement the data contains an instrumental transient, (noise), and let t be a test that is performed to infer which of the two statements above are true. Let us suppose t is a very good test, in that it discriminates between s and n very well, and say the detection probability is as high as $P(t | s) = 0.95$ with a low false alarm rate $P(t | n) = 0.05$ (note that $P(t | s)$ and $P(t | n)$ need not necessarily add up to 1). Also, the expected event rate of a chirp during our observation is low, $P(s) = 10^{-5}$, but the chance of an

instrumental transient is relatively large, $P(n) = 0.01$. We are interested in knowing what the posterior probability of the data containing a chirp is, given that the test has been passed. By Bayes theorem this is

$$P(s | t) = \frac{P(t | s)P(s)}{P(t)} = \frac{P(t | s)P(s)}{P(t | s)P(s) + P(t | n)P(n)},$$

where $P(t)$ (the probability of the test being positive) is taken to result from either the chirp or the instrumental transient. Substituting for various quantities in the above equation we find

$$P(s | t) = \frac{0.95 \times 10^{-5}}{0.95 \times 10^{-5} + 0.05 \times 0.01} \simeq 0.02$$

There is only a 2% chance that the data really contains a chirp when the test was taken. On the contrary, for the same data we find that the chance of an instrumental transient for a positive test result is $P(n | t) \sim 98\%$. Thus, though there is a high (low) probability for the test to be positive in the presence of a signal (noise) when the test is indeed positive, we cannot necessarily conclude that a signal is present. This is not surprising since the prior probability of the signal being present is very low. The situation can be remedied by designing a test that gives a far lower probability for the test to give a positive result in the case of an instrumental transient (i.e., a very low false alarm rate).

Thus, Bayesian inference neatly folds the prior knowledge about sources in the estimation process. One might worry that the outcome of a measurement process would be seriously biased by our preconception of the prior. To understand this better, let us rewrite Equation (106) as follows:

$$P(s | t) = \frac{1}{1 + P(t | n)P(n)/P(t | s)P(s)} = \frac{1}{1 + L_{\text{NS}}p_{\text{SN}}}$$

where $L_{\text{NS}} = P(t | n)/P(t | s)$ is the ratio of the two likelihoods and $p_{\text{SN}} = P(s)/P(n)$ is the ratio of the priors. The latter is not in the hands of a data analyst; it is determined by the nature of the source being searched for and the property of the instrument. The only way an analyst can make the posterior probability large is by choosing a test that gives a small value for the ratio of the two likelihoods. When $L_{\text{NS}} \ll p_{\text{SN}}$ (i.e., the likelihood of the test being positive when the signal is present is far larger, depending on the priors, than when the transient is present) the posterior will be close to unity.

The above example tells us why we have to work with unusually-small false-alarm probability in the case of gravitational wave searches. For instance, to search for binary coalescences in groundbased detectors we use a (power) SNR threshold of about 30 to 50. This is because the expected event rate is about 0.04 per year.

Computing the posterior involves multi-dimensional integrals and these are highly expensive computationally, when the number of parameters involved is large. This is why it is often not possible to apply Bayesian techniques to continuously streaming data; it would be sensible to reserve the application of Bayesian inference only for candidate events that are selected from inexpensive analysis techniques. Thus, although Bayesian analysis is not used in current detection pipelines, there has been a lot of effort in evaluating its ability to search for [115, 348, 122, 120] and measure the parameters of [116, 121, 377] a signal and in follow-up analysis [378].

8 Шумы при детектировании интерферометром (!?!?!?)

(прямо очень подробно - см. отдельную запись в статфизе, потому что это совсем ничего общего с грав волнами не имеет. но тут вмеру подробно напишу, что с шумами все-таки происходит и почему? я пока вообще не уверен, где лучше эту отодельную запись ставить, потому что там очень разные разделы нужны становятся.)

8.0.1 On thermal noise in interferometric due to dielectric optical coatings

Essence of Thermal noise in interferometric gravitational wave detectors due to dielectric optical coatings

У нас просто внутри материала разные происходят колебательные волны, так что меняется размер поверхности (???? пока так понял).

$$S_x(f) = \frac{2k_B T}{\pi^2 f} \frac{d}{w^2} \phi \left(\frac{Y_{\text{coat}}}{Y_{\text{sub}}^2} + \frac{1}{Y_{\text{coat}}} \right).$$

Here, T is the coating temperature, w is the beam radius at the coating, as described in section 2, ϕ is the mechanical loss angle of the coating, d is the coating thickness and Y_{coat} , Y_{sub} being the Young's modulus of the material coating and substrate respectively.

(а дальше указание, почему такая формула. это же в теории дальше будет написано!! доработаю теорию про это!!)

Mechanical admittance

$$\mathbb{Y}(f) := i2\pi f \frac{x(f)}{F}$$

$$p(\vec{r}, t) = p(r, t) = \frac{2F}{\pi w^2} \exp\left(\frac{-2r^2}{w^2}\right) \sin(2\pi f t)$$

Rewrite:

$$\text{Re}\{\mathbb{Y}(f)\} = \frac{4\pi f U(f)}{F^2} \phi$$

$U(f)$ - maximum elastic energy stored in the test mass as a result of the excitation, and ϕ is the loss angle of the response. This equation holds at frequencies far below the first resonance of the test mass, provided $\phi \ll 1$. It can be obtained as follows. Under the conditions stated

$$U(f) := \frac{1}{2} |F| |x(f)|$$

and the response $x(f)$ to the excitation is

$$x(f) \equiv |x(f)| \exp(-i\phi) \approx |x(f)| (1 - i\phi).$$

So we need to obtain $U(f)$ from pressure distribution.

$$\phi_{\text{readout}} = \frac{1}{U} (U_{\text{substrate}} \phi_{\text{substrate}} + U_{\text{coating}} \phi_{\text{coating}})$$

where U is the maximum elastic energy stored in the sample as a consequence of the applied pressure, $U_{\text{substrate}}$ is the portion of the energy stored in the substrate, U_{coating} is the portion of the energy stored in the coating, $\phi_{\text{substrate}}$ is the loss angle of the substrate, and ϕ_{coating} is the loss angle of the coating.

Допустим, что

$$p(r) = \frac{2F}{\pi w^2} \exp\left(\frac{-2r^2}{w^2}\right).$$

$$U_{\text{coating}} \approx \delta U d,$$

where δU is the energy density stored at the surface, integrated over the surface, and d is the thickness of the coating. (размерность δU не равна размерности U !) Similarly, $U_{\text{substrate}} \approx U$, so

$$\phi_{\text{readout}} = \phi_{\text{substrate}} + \frac{\delta U d}{U} \phi_{\text{coating}}$$

Плотность энергии можно разложить и написать через σ_{ij} - stresses and ϵ_{ij} - strains:

$$\begin{aligned} \rho U &= \rho_{rr} + \rho_{r\theta} + \dots \\ \rho_{rr} &\equiv \frac{1}{2} \sigma_{rr} \epsilon_{rr}, & \rho_{r\theta} &\equiv \frac{1}{2} \sigma_{r\theta} \epsilon_{r\theta} \end{aligned}$$

Это ключевая связь с теорией упругости. The associated loss angles are $\phi_{rr}, \phi_{r\theta}$, etc. (??? типа тут цилиндрические координаты в плоскости зеркала.) We shall always have $\epsilon_{rz} = \epsilon_{zr} = 0$. (???? не понимаю, что конкретно значит это????)

So

$$\begin{aligned} \rho'_{U\parallel} &= \frac{1}{2} (\epsilon'_{rr} \sigma'_{rr} + \epsilon'_{\theta\theta} \sigma'_{\theta\theta} + \epsilon'_{r\theta} \sigma'_{r\theta}) \\ \rho'_{U\perp} &= \frac{1}{2} \epsilon'_{zz} \sigma'_{zz} \end{aligned}$$

So

$$\delta U_{\parallel} = \int_S \rho'_{U\parallel} d^2 r, \quad \delta U_{\perp} = \int_S \rho'_{U\perp} d^2 r.$$

Полная энергия:

$$U = \pi \int_0^\infty dz \int_0^\infty r dr (\epsilon_{rr} \sigma_{rr} + \epsilon_{\theta\theta} \sigma_{\theta\theta} + \epsilon_{zz} \sigma_{zz} + 2\epsilon_{rz} \sigma_{rz})$$

(тут кстати Торн и др. умные люди занимались какими-то умными вычислениями.)

Due to axial symmetry, $\epsilon_{r\theta} = \epsilon_{\theta r} = \epsilon'_{r\theta} = \epsilon'_{\theta r} = 0$. Due to traction-free boundary conditions, $\epsilon'_{rz} = \epsilon'_{zr} = 0$ at the coating surface, and the same must therefore hold (to leading order) for the entire coating.

Далее нужно повозиться с теорией упругости и дойти до выражения энергии через модуль Юнга и коэффициент Пуассона.

$$\begin{aligned} U &= \frac{F^2 (1 - \sigma^2)}{2\sqrt{\pi} w Y} \\ \delta U_{\parallel}/U &= \frac{1}{\sqrt{\pi} w} \frac{Y'(1 + \sigma)(1 - 2\sigma)^2 + Y\sigma'(1 + \sigma')(1 - 2\sigma)}{Y(1 + \sigma')(1 - \sigma')(1 - \sigma)} \\ \delta U_{\perp}/U &= \frac{1}{\sqrt{\pi} w} \frac{Y(1 + \sigma')(1 - 2\sigma') - Y'\sigma'(1 + \sigma)(1 - 2\sigma)}{Y'(1 - \sigma')(1 + \sigma)(1 - \sigma)} \end{aligned}$$

where Y and σ are the Young's modulus and Poisson's ratio of the substrate, and Y' and σ' are the Young's modulus and Poisson's ratio of the coating. (про эту возню в выводе напишу, я

там слишком плохо разбираюсь, чтобы быстро понять это) Thus,

$$\phi_{\text{readout}} = \phi_{\text{substrate}} + \frac{1}{\sqrt{\pi}} \frac{d}{w} \left(\frac{Y'(1+\sigma)(1-2\sigma)^2 + Y\sigma'(1+\sigma')(1-2\sigma)}{Y(1+\sigma')(1-\sigma')(1-\sigma)} \phi_{\parallel} + \frac{Y(1+\sigma')(1-2\sigma') - Y'\sigma'(1+\sigma)(1-2\sigma)}{Y'(1-\sigma')(1+\sigma)(1-\sigma)} \phi_{\perp} \right) \quad (8.1)$$

$$S_x(f) = \frac{2k_B T}{\pi^{3/2} f} \frac{1-\sigma^2}{wY} \left\{ \phi_{\text{substrate}} + \frac{1}{\sqrt{\pi}} \frac{d}{w} \frac{1}{YY'(1-\sigma'^2)(1-\sigma^2)} \left[Y'^2(1+\sigma)^2(1-2\sigma)^2 \phi_{\parallel} + YY'\sigma'(1+\sigma)(1+\sigma')(1-2\sigma)(\phi_{\parallel} - \phi_{\perp}) + Y^2(1+\sigma')^2(1-2\sigma') \phi_{\perp} \right] \right\}$$

The limit $Y' = Y, \sigma' = \sigma$, and $\phi_{\perp} = \phi_{\parallel}$ gives

$$S_x(f) = \frac{2k_B T}{\pi^{3/2} f} \frac{1-\sigma^2}{wY} \left\{ \phi_{\text{substrate}} + \frac{2}{\sqrt{\pi}} \frac{(1-2\sigma)}{(1-\sigma)} \frac{d}{w} \phi_{\parallel} \right\}.$$

For the case of fused silica or sapphire substrates coated with alternating layers of Ta_2O_5 and SiO_2 , the Poisson's ratio of the coating may be small enough ($\lesssim 0.25$) so we approximate (within about 30%) by setting $\sigma = \sigma' = 0$

$$S_x(f) = \frac{2k_B T}{\pi^{3/2} f} \frac{1}{wY} \left\{ \phi_{\text{substrate}} + \frac{1}{\sqrt{\pi}} \frac{d}{w} \left(\frac{Y'}{Y} \phi_{\parallel} + \frac{Y}{Y'} \phi_{\perp} \right) \right\}.$$

We see that in order to estimate the thermal noise performance of a particular coating, we must know all of $Y, Y', \phi_{\parallel}$, and $\phi_{\perp}$. It also shows that if $\phi_{\parallel} \approx \phi_{\perp}$, then the lowest coating-induced thermal noise occurs when the Young's modulus of the coating is matched to that of the substrate. If $Y' \not\approx Y$, one of $\phi_{\parallel}$ or $\phi_{\perp}$ will be emphasized and the other de-emphasized. This is particularly worrisome for coatings on sapphire substrates, whose high Young's modulus means that for most coatings $\phi_{\perp}$ is likely to be the main contributor to the coating thermal noise.

Essence of Substrate noises

$$S^{\delta z}(f) = \frac{a\beta^2 k_B T^2}{2\pi\kappa} \times (e^{it} E_1(it) + e^{-it} E_1(-it)).$$

Here $\beta = \partial n / \partial T$ with n being the material refractive index, a is the test mass thickness, k_B is the Boltzmann constant, T is the substrate temperature and κ is the thermal conductivity. The term t here is defined as $t = \frac{2\pi f w^2 \rho C}{2\kappa}$. Inside this term, w is the laser beam size on the substrate surface, ρ is the density of the material and C is the specific heat. The functions $E_1(x)$ are exponential integrals which equal $E_1(x) = \int_x^{\infty} e^{-xy} y^{-1} dy$. For the effect on both ITMs, the noise is then projected to the differential arm-length signal as $S_{\text{DARM}}^{\delta z}(f) = S^{\delta z}(f) \times \frac{\pi}{F}$, with F being the arm cavity finesse value. The parameters involved in the computation are shown in table A1.

$$S_{\text{TE}}(f) = \frac{8}{\sqrt{2\pi}} \alpha^2 (1+\sigma)^2 \frac{k_B T^2 w}{\rho C a^2} J[\Omega],$$

$$J[\Omega] = \sqrt{\frac{2}{\pi}} \int_0^{\infty} du \int_{-\infty}^{\infty} dv \frac{u^3 e^{-u^2/2}}{(u^2+v^2)((u^2+v^2)^2+\Omega^2)}$$

In the above equations C is the specific heat of the material, α is the coefficient of linear expansion, ρ is the density of the substrate, w is the beam radius at the test mass and $a =$

$\sqrt{w^2 \cdot \omega_c}$ with ω_c being the angular frequency of the adiabatic limit given by $\omega_c = \frac{\kappa}{\rho C w^2}$. Figure 4 illustrates the low-frequency discrepancy between the adiabatic approximations and the full solutions, which makes the implementation of the complete solution strongly relevant for the parameters in our experiment.

Finally, the spectral density of the noise source associated with the Brownian fluctuations in the bulk material is estimated using an extension of Levin's direct method [51], developed by Bondu et al (1998) [52]:

$$S_x(f) = \frac{2k_B T}{\sqrt{\pi^3} f} \frac{1 - \sigma^2}{Y w} \phi_{\text{substrate}}(f, T).$$

Derivation of Thermal noise (?????)

(выгружу потом, пока не так уж нужно на самом деле)

Derivation by Levin

(мб разберу когда-то)

Derivation by Lui and Thorne

(мб разберу когда-то)

Оставшиеся вопросы

(там кстати другие методы есть вычислений, мб потом их напишу, пока забиваю.)

8.0.2 On mechanical loses

Essence

Theory

Оставшиеся вопросы

Так почему же квадрупольный маятник помогает? (?????)

(мб это для механических шумов, а не для тепловых!!! так что это в другом месте! пока так думаю!)

8.0.3 On other typical noises (???)

(там много еще есть, которые я просто укажу. разные супер гипотетические отнес пока в часть с дополнительными темами, ибо их может быть ну прямо очень много)

8.0.4 Общие вопросы о шумах

Как все-таки отличить шум от влияния гравитационной волны? (????)

(хочу четко на этот вопрос ответить)

9 Эксперименты

9.1 Метод LIGO

Laser Interferometric Gravitational Wave Observatory

самый популярный

(тут суть, подробнее - в записи про волны)

9.1.1 Общее описание метода

Вкратце описание метода

Gravitational Wave Detectors Gravitational wave detectors are interferometers. They bounce light back and forth between two arms, with the mirrors at either end playing the role of test masses.

If the gravitational wave travels perpendicular to the plane of the detector, it will shorten one arm and lengthen the other. With the arms aligned along the x and y axes, the maximum change in length can be read from (5.24),

$$L' = L \left(1 \pm \frac{H_+}{2} \right) \Rightarrow \frac{\delta L}{L} = \frac{H_+}{2}$$

To get a ballpark figure for this, we need to understand how large we expect H_+ to be from any plausible astrophysical source. We'll do this in Section 5.3.2. It turns out it's not really very large at all: typical sources have $H_+ \sim 10^{-21}$. The lengths of each arm in the LIGO detectors is around $L \sim 3$ km, meaning that we have to detect a change in length of $\delta L \sim 10^{-18}$ m. This seems like a crazy small number: it's smaller than the radius of a proton, and around 10^{12} times smaller than the wavelength of the light used in the interferometer. Nonetheless, the sensitivity of the detectors is up to the task and the LIGO observatories detected gravitational waves for the first time in 2015. For this, three members of the collaboration were awarded the 2017 Nobel prize. Subsequently,

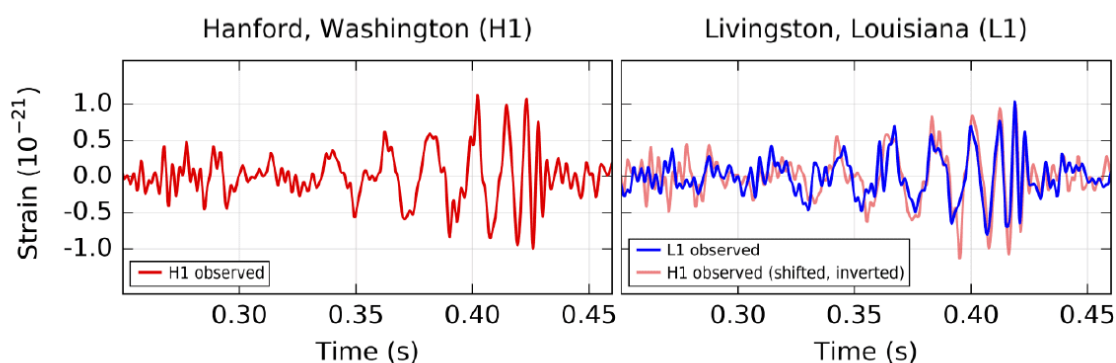


Figure 2

Параметры экспериментальной установки LIGO

часто про это возникает вопрос

Обзор критики и сомнений

она какая-то тоже есть

Обзор применений

где-то это нужно, подумаем, где.
пока хз.

О том, как можно вообще детектировать настолько малые изменения?

(типа даже протон больше по размеру. как такое детектировать??? пока не понимаю ответ.)

9.1.2 О способах обработки результатов

Обсудим вкратце обработку результатов в эксперименте LIGO
(Говорят, там отдельные хорошие способы, что ж, тут что-нибудь про них напишу.)

9.1.3 Результаты

9.2 VIRGO Experiment

(тут суть, подробнее - в записи про волны)

9.2.1 Общее описание метода

Вкратце описание метода

Параметры экспериментальной установки LIGO

часто про это возникает вопрос

9.2.2 Other about VIRGO

Обзор критики и сомнений

она какая-то тоже есть

Обзор применений

где-то это нужно, подумаем, где.
пока хз.

10 Другие методы детектирования волн

Соберем в этом разделе всевозможные методы, которые можно использовать для детектирования.

10.0.1 Detection by phase modulation of the light from a distant star

Обсудим детектирование за счет поиска задержки фазы света, проходящего вблизи источника гравитационных волн.

(Статья Лесовика, мб когда-то почитаю ее, просто такой-то вот метод, ну а также есть теория про шумы, и вообще эйконала, которая немного интересная.)

Обзор метода

Вкратце описание метода

Here a light signal emitted from a distant star interacts with a local (also distant) GW source and travels towards the Earth, where it is detected. While traveling in the field of the GW, the light acquires specific phase modulation,

(?????)

which can be considered as a coherent spreading of the given initial photons energy over a set of satellite lines, spaced at the frequency of GW (from quantum point of view it is multi-graviton absorption and emission processes).

This coherent state of photons with the energy distributed among the set of equidistant lines, can be analyzed and identified on Earth either by passing the signal through a Fabry-Perot filter or by monitoring the intensity-intensity correlations at different times.

Обзор Лесовика (из заключения)

Let us now summarize all results obtained through the article and give some estimations and limitations for the proposed gravitational waves detection method.

The primary effect we have dealt with is the modulation of the photon phase due to interaction with gravitational field have been found in the geometrical optics approximation and described by eq.

(26) for the photon wave function:

$$\phi_k(\xi) = \mathcal{C}(\xi)e^{-i\omega_k\xi - i\omega_k\tau_g \sin \Omega\xi}$$

where $\xi = t - r/c$ and $\mathcal{C}(\xi) = [1 + \Omega\tau_g \cos \Omega\xi]^{-1}$ is the pre-exponential factor.

This effect of phase modulation leads immediately to well known Shapiro effect of the time delay of the signal propagating near the source of gravitational field.

For the case of the source of gravitational waves this time delay is an alternating function of time $\tau_{ar} = r/c - \tau_g \sin \Omega\tau$, where τ_{ar} is an arrival time of the signal, r/c the traveling time of the signal in flat Minkovski space and

$$\tau_g = -\frac{h_{xx}(\Lambda)}{\pi} \left(\frac{\Lambda}{D}\right)^2 \Omega^{-1}$$

Here $h_{xx}(\Lambda)$ is the dimensionless strength of gravitational wave near the GW source, D impact parameter, $\Omega = 2\pi c/\Lambda$ is the frequency of gravitational waves.

The Shapiro effect in fact seems to be a very promising candidate for gravity wave detection experiment since the time delay (or the additional alternating in time phase) mainly accumulated near the source of gravitational waves where the strength of GW many order of magnitude greater when they reach the Earth.

However in the proposed experiment [16] it requires a pulsar on the line of sight which passes near the source of gravitational radiation.

In this paper we propose a gravitational waves detection method based on the FabryPerot interference analysis (or equivalently time correlation measurement) of the light signal from arbitrarily light source (not necessarily a pulsar) passing near the strong source of gravitational radiation.

Then the photodetector signal contains an alternating in time component (with GW frequency) proportional to the strength of gravitational radiation near the the GW source (see eq.

56,61) :

$$P(t) = P_0 (1 - \Omega\tau_g \cos \Omega t)$$

with P_0 a photodetector signal corresponding to a light coming to a FP filter from distant star.

To estimate an effectiveness of the proposed method let us first estimate a brightness of the light source needed to resolve an alternating component in the photo-detection signal.

Collecting the signal over long time τ_{obs} allow to better resolve the alternating component.

Assuming that during the observation time N_{obs} photons coming from a distant star hit the photodetector, in the limit of Poissonian statistics, $\langle \delta N_{\text{obs}}^2 \rangle = N_{\text{obs}}$ one has the following limiting requirement:

$$\Omega_{\tau_g}(D)\sqrt{N_{\text{obs}}} \gg 1$$

in the extreme case where the light passes in immediate proximity ($D \approx \Lambda$) to the GW source one has the condition:

$$h(\Lambda)\sqrt{N_{\text{obs}}} \gg 1$$

Then assuming that the distant star emits in the frequency band $\Delta\omega$ and that the diameter of the telescope equals to d one has:

$$N_{\text{obs}} \sim L \frac{d^2}{R^2} \frac{\Gamma}{\Delta\omega} \frac{\tau_{\text{obs}}}{\hbar\omega_0}$$

where L is the brightness of the distant star, R is the distance to the Earth, The most promising candidates of gravitational waves sources appropriate for proposed detection method are the periodic sources of gravitational radiation such as close binaries at the final stage of evolution or a asymmetric rotating neutron stars, allowing to collect the signal long enough time to resolve the tiny gravitational wave modulation on the stochastic Poisson noise.

Let us give some estimations, assuming the total mass, m of binary system of the order of the mass of the Sun, and rotating with a frequency $\nu \sim 10^{-2}$ Hz.

Then the dimensionless strength of gravitational waves at the distance r from the emitter, according to quadrupole approximation, is given by

$$h \approx 4 \frac{G^{5/3}}{c^4} \frac{1}{r} m^{5/3} \nu^{2/3}$$

assuming that the role of the distant star plays one of the components of the binary system $r = (Gm/\nu^2)^{1/3}$ one has $h(r) \approx 4(Gm\nu)^{4/3}/c^4 \sim 10^{-9}$. In order to resolve this modulation one needs about $N_{\text{obs}} \sim 10^{18}$ photons coming to a photodetector during τ_{obs} .

Assuming the the typical distance from the binary system to the Earth of the order 1 kpa with $d \approx 10$ m, $\omega_0 \sim 10^{14}$ Hz and $\Gamma/\Delta\omega \sim 1$, $\tau_{\text{obs}} \sim 10^3 \nu^{-1} \sim 10^5$ s one has the required brightness of the binary component, according to eq.

(80) of the order $L \sim 10^5 L_S$ where L_S is the brightness of the Sun.

Of course the requirement to have such a bright component of the binary system makes the observation of the proposed effect rather problematic in practice.

In reality the situation with binary systems is even worse due to the presence of the Doppler effect for the light emitted from one of the rotating components.

In fact the Doppler effect results in the same frequency modulation of the emitted light as the gravitational waves modulation and the amplitude of the frequency modulation due to the Doppler effect is of the order of

$$\left(\frac{\Delta\omega}{\omega} \right)_{\text{Doppler}} \sim \left(\frac{v}{c} \right)^2 \approx \frac{(Gm\nu)^{1/3}}{c} \frac{\nu}{c}$$

For parameters described above it results in an effect of the order 10^5 that is five orders of magnitude greater than the frequency modulation due to gravitational waves.

The situation however could be much better for very close neutron star binaries at the last stage of evolution.

In this case the rotational frequency can be of the order of 10^2 Hz and results in a dimensionless strength of emitted gravitational waves of the order $h(\Lambda) \sim 10^{-6}$.

Assuming that the distance to the binary neutron star system is of the order 1kpc and that there is a bright distant star situated on the line of sight to the binary from the Earth with impact parameter $D \sim \Lambda$ and collecting the signal during the time $\tau_{\text{obs}} \sim 10^3 \nu^{-1} \sim 10s$ one has the following lower limit for the brightness of distant star $L \sim 10^2 L_S$

Плюсы метода

хз

гравитационные эффекты вблизи источника волн заметно сильнее.

Минусы метода

хз

Про прошлые методы

(тут пара пар-ов про другие методы сперва, потом улучшу структуру)

The interferometric gravitational-wave detector like LIGO and VIRGO represents a Michelson interferometer with a laser beam split between two perpendicular arms of interferometer.

The action of gravitational waves on an interferometer can be presented as relative deformation of both interferometer arms.

A gravitational wave with dimensionless amplitude h induces the opposite length changes (?????? потом проверю)

$$\delta l/l = 1/2 h \cos \Omega t$$

in each arm of the Michelson interferometer, where l is the length of the arm, Ω is the gravitational wave frequency.

These length changes produce opposite phase shifts between two light beams in interferometer arms, when interference occurs at the beam splitter of Michelson interferometer.

The resulting phase shift of a single beam of light spending time τ in the interferometer can be written as

$$\delta\phi = h \frac{\omega}{\Omega} \sin \frac{\Omega\tau}{2}$$

where ω is the light frequency.

This phase shift results an intensity signal change of the light from interferometer beam splitter hitting the photodetector.

The main problem of the acoustic and interferometric methods that they both deal with gravitational waves with extremely small amplitudes of the order $h \sim 10^{-21}$ reached the Earth from deep space.

One can see from eq. (1) that for gravity wave frequencies in the 1kHz range, $\Omega \sim 10^3$ Hz, and for the light in visual frequency range, $\omega \sim 10^{14}$ Hz, one has the maximum phase shift of the order $\delta\phi \sim 10^{-10}$ for interferometer arms length of the order 150 kilometers.

Such extraordinarily weak effect requires a exceedingly high detector sensitivity both acoustic and interferometric detectors.

интро про детектирование с помощью прохождения света через источник волн

Alternatively, GW detection may be based on effects associated with propagation of light or electromagnetic waves in gravitational fields. There are two primary effects for the light in

constant gravitational field

- i) the deflection of light rays near massive bodies
- ii) the Shapiro effect accounting for integrated time delay of the signal passing near a strong source of gravitational field.

The same effects have to be observed for light propagating in gravity waves: the gravitational waves have to induce a weak time dependent deflection of light ray propagating through these waves and also lead to gravity-wave-induced variation in time delay.

The idea to use astrometry to detect periodic variation in apparent angular separation of appropriate light sources was proposed by Fakir (?? тут потом посмотрю ссылку). It is shown that for a gravity wave source located between the Earth and the light source (with line of sight close aligned to gravity wave source) a periodical variation of the order

$$\Delta\phi \sim \pi h(\Lambda)$$

in the angular position of the light source has to be observed. Here $h(\Lambda)$ is the dimensionless strength of the gravity wave at distances of the order gravitational wavelength Λ which is many orders of magnitude greater than the strength of the same waves when they reach the Earth.

(?? а что на первой hand? про один и тот же метод речь, нет?)

On the other hand one can directly measure the variation of the integrated time delay induced by gravity waves on the light emitted by distant star which passing through space region with strong gravity wave.

The idea to use timing observation for detection of the gravitational waves was suggested first by Sazhin and then this problem was studied in details by several authors.

The estimations carried out (в одной из статей см. Лесовика) give the following answer for the rate of change in the gravity-wave-induced time delay, $\dot{\tau}$: (?? хз, что это за параметр?)

$$|\dot{\tau}| \sim |h(r = D)|$$

where $h(r = D)$ is the gravity wave strength at distances of the order of impact parameter D for the light beam passing near the gravity wave source.

To put some numbers let us consider a dissymmetric rotating neutron star with spin frequency of the order 10^3 Hz and gravity wave amplitude of the order $H \sim 10^{-5}$ cm, then eq. $|\dot{\tau}| \sim |h(r = D)|$ gives $|\dot{\tau}| \sim 10^{-12}$ while the same effect measured near Earth results $|\dot{\tau}| \sim 10^{-26}$ assuming that the neutron star to be at a typical distance of a few kiloparsecs from the Earth.

One can see that all effects due to gravitational waves near the gravity wave source are several orders of magnitude stronger than the same ones on the Earth.

ИНТРО В МЕТОД, ОСНОВАННЫЙ НА ВЗАИМОДЕЙСТВИЕ ФОТОНА С ВОЛНОЙ

(? а в чем разница от методов выше?)

In the present paper we want to suggest a new gravitational waves detection method based on the interaction of the photon with gravitational waves.

Assuming that the photons from distant star passing near gravitational wave source, where the photon-gravity wave interaction assumed to be strong, the photon-gravity wave interaction leads to relatively strong modulation in time of photon frequency.

The latter allows to analyze this modulation of the photons reaching the Earth by means of standard optical methods including the Fabry-Perot analysis and quantum photon correlations measurements.

It is important that while the interaction of photons with gravitational wave is rather weak (proportional to strength of gravitational wave) the frequency modulation can be accumulated over large distances during the photon propagation that could result in an experimentally

measurable effect on the Earth. (? это не очевидно, почему при общем расстоянии эффект усиливается?)

In some aspects our treatment of the photon-gravity wave interaction resembles the effect of photon acceleration by gravitational wave, where the photon propagating in plane gravitational wave long enough time acquires a considerable increase of the frequency.

However while in Ref. [20] the photon-gravity wave interaction was treated in the frame of reference of the photon (?? про неё потом пропишу раздел в механике, уже заготовил), we have considered this interaction in the point of observation frame coordinate which seems to give rather the widening of the initial monochromatic photon wave packet than the increase of the photon frequency.

Распространение света вблизи источника гравитационных волн

(где-то потом сделаю это большой теорией)

In this section we consider how the plane electromagnetic waves interact with gravitational wave field emitted by some localized source.

The situation we have in mind is depicted in Fig. 1.

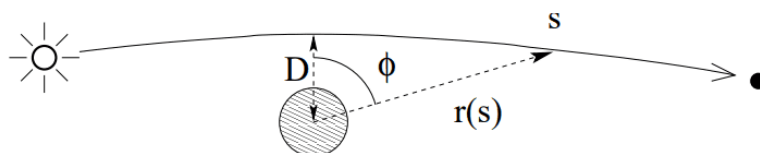


Figure 3

The light signal originates from a distant star S and travels towards the Earth to be detected by an observer O , but along the way it interacts with a gravitational body M , which also emits gravity waves.

We consider the trajectory of the light ray in the two dimensional plane formed by the star S , the body which emits the GW, M , and the Earth.

Provided that the wave length of the GW is large compared to the wave length of the light signal, we will use the eikonal or geometrical optics approximation to describe the interaction of the light in gravitational waves.

(? ниже про статьи по этой теме. посмотрю их тоже)

The propagation of the electromagnetic waves in the field of gravitational waves было изучено в [7, 21, 22, 23, 24, 25]. (???!! тут это всё посмотрю, ибо очень хочу почитать, как же это описывать!! сперва с Лесовиком закончу)

Our subsequent analysis of this problem will be carried out in close analogy with the paper of Sazhin and Maslova [25], where they have considered the structure of electromagnetic field in Fabry-Perot resonator in the field of gravitational waves.

However in the present paper we will be interested in a different aspect of the interaction of the light beam passing in immediate proximity to the gravitational wave source, where the amplitude of the gravitational waves is assumed to be strong in comparison with the amplitude of the same waves reaching the Earth.

(?? кст, так и не понял, а что же тогда в предыдущих работах рассматривалось? их тоже почитать бы нужно)

О связи с гравитационным линзированием (?)

(пока не понимаю, как конкретно устроена эта связь, но может быть, к моделям линзирования можно добавлять так вот ещё волны??? просто проблема в том, что линзирование описывается не совсем в таком формализме, чтобы можно было вставить волны. так что пока связи не понимаю)

Прохождение света вблизи источника гравитационных волн

(тут модель нахождения задержки по фазе)

Вывод в двух словах

(тут коротко пишу то, что подробно Лесовик описывает)

Уравнение эйконала

$$g^{ik} \frac{\partial \psi}{\partial x^i} \frac{\partial \psi}{\partial x^k} = 0$$

при $\psi = -\omega t + ky + \psi_1$ имеет вид:

$$\frac{\partial \psi_1}{\partial t} + c \frac{\partial \psi_1}{\partial y} = -\frac{\omega}{2} F(t, r),$$

где $F(t, r) = h^{00} + h^{yy} - 2h^{0y}$

(? проверю потом)

Тогда его решение есть (? впишу потом в УМФ, пока такие ходы не очевидны)

$$\psi_1(t, y) = -\frac{1}{2} \frac{\omega}{c} \int_{-\infty}^y F\left(t - \frac{y - y'}{c}, y'\right) dy'$$

The eikonal equation

)и тут я понимаю, что в теории поля это криво написано, так что я отлетаю на оптику и пишу там основы этого...)

In the geometrical optics approximation the propagation of the light ray is described by the eikonal equation:

$$g^{ik} \frac{\partial \psi}{\partial x^i} \frac{\partial \psi}{\partial x^k} = 0$$

where $g^{ik} = g_0^{ik} - h^{ik}$ the metric tensor associated with the Schwarzschildian static metric (?? так там же 2 тела, а не одно???) g_0^{ik} of the object M , which is perturbed by the tensor h^{ik} associated with the GW emitted from M .

Here we will neglect the true form of the static metric g_0^{ik} assuming that light ray propagates in the flat Minkovsky space $g_0^{ik} = \eta^{ik} = \text{diag}\{1, -1, -1, -1\}$. (? а так точно можно сделать?)

It is known that the static metric leads both to the deflection of the light ray near the massive body and time delay of the light signal passing along the ray (so called Shapiro effect).

It can be rigorously shown in the subsequent analysis that the former effect of the light deviation leads to the negligible correction to the additional phase accumulated by photons due to the interaction with GW and one can assume the light trajectory to be a straight line.

As for static Shapiro time delay it can be easily incorporated to the final answer for photon phase and actually does not affect on the our subsequent detection method of modulated photons. (? не понял, типа эффект не влияет?)

Let $\psi_0 = -\omega t + ky$ be the eikonal in the absence of gravitational waves perturbation with metric η^{ik} (we assumed the light ray to be propagating along the y axis in the plane formed by three bodies).

Assuming that the perturbation is small, in the presence of gravitational waves the eikonal becomes $\psi = \psi_0 + \psi_1$, where ψ_1 a small addition to the eikonal ψ_0 computed to first order in h^{ik} , it satisfies the equation:

$$\frac{\partial \psi_1}{\partial t} + c \frac{\partial \psi_1}{\partial y} = -\frac{\omega}{2} F(t, r)$$

where we have introduced the notation $F(t, r) = h^{00} + h^{yy} - 2h^{0y}$. Using the Green's function formalism one can find the general solution of this equation.

Then to first order in the gravitational wave perturbation, h^{ik} , the additional phase acquired by the photon due to the interaction with the GW, can be written for arbitrary time dependence of the perturbed metric:

$$\psi_1(t, y) = -\frac{1}{2} \frac{\omega}{c} \int_{-\infty}^y F\left(t - \frac{y - y'}{c}, y'\right) dy'$$

The emission of GW

(???!! а абзац ниже должен быть четко заготовлен в разделе про излучение грав волн! Собственно, параграф как раз об этом! пока слабая в этом плане теория...)

Assuming that the GW tensor, h^{ik} , is a small perturbation to the static flat metric in the quadrupole approximation one can write the coordinate components of the GW tensor as:

$$\tilde{h}_{\alpha\beta} = \frac{2G}{c^4} \frac{\ddot{Q}_{\alpha\beta}(t - r/c)}{r}$$

where G is the Newtonian gravitation constant, $\tilde{h}_{ik} = h_{ik} - \frac{1}{2}\eta_{ik}h$ (where $h \equiv h^i_i$), $Q_{\alpha\beta}$ is the quadrupole moment:

$$Q_{\alpha\beta}(t) = \int \tilde{T}_{00}(t, r) x^\alpha x^\beta d^3x.$$

where $\tilde{T}_{ik} = T_{ik} - \frac{1}{2}\eta_{ik}T$, T_{ik} is the energy-momentum tensor of the GW source, $T \equiv T^i_i$.

(??? калибровочные условия как ниже потом пропишу в первой части! пока что этим не владею?)

We use so called the harmonic gauge condition, $\partial_k \tilde{h}^k_i = 0$, that has been chosen for the components of GW tensor h_{ik} . Using the gauge condition one can find the $\tilde{h}_{00}$ and $\tilde{h}_{0\alpha}$ components of the GW tensor. To do this let us write the equation $\partial_k \tilde{h}^k_i = 0$ separately for coordinate and time components:

$$\frac{\partial \tilde{h}_{\alpha\beta}}{\partial x^\beta} = \frac{\partial \tilde{h}_{\alpha 0}}{\partial x^0}, \quad \frac{\partial \tilde{h}_{0\alpha}}{\partial x^\alpha} = \frac{\partial \tilde{h}_{00}}{\partial x^0}.$$

Подставим в первое уравнение значение $\tilde{h}$ через квадрупольный момент, и $\tilde{h}^{\alpha 0}$ components of GW tensor can be written as:

$$\tilde{h}_{\alpha 0} = \frac{2G}{c^3} \frac{\partial}{\partial x^\beta} \left[\frac{\dot{Q}_{\alpha\beta}(t - r/c)}{r} \right]$$

и во второе:

$$\tilde{h}_{00} = \frac{2G}{c^2} \frac{\partial^2}{\partial x^\alpha \partial x^\beta} \left[\frac{Q_{\alpha\beta}(t - r/c)}{r} \right]$$

Теперь мы можем написать точное выражение для функции $F(t, r) = h^{00} + h^{yy} - 2h^{0y}$, за счет которой найдется изменение фазы.

(?? дальше преобразования, которые я потом повторять буду, пока лишь верю)

Taking coordinate derivatives and keeping in mind that along the ray trajectory $\partial_x r = D/r$, $\partial_y r = y/r$, $\partial_z r = 0$ (where D is the impact parameter) the $F(t, r)$ takes the form

$$\begin{aligned} F(t, r) = & \frac{2G}{c^4} \left[\frac{\ddot{Q}_{xx}}{r} \frac{D^2}{r^2} + 2 \frac{\ddot{Q}_{xy}}{r} \frac{D}{r} \left(\frac{y}{r} - 1 \right) + \frac{\ddot{Q}_{yy}}{r} \left(\frac{y}{r} - 1 \right)^2 \right] + \\ & + \frac{2G}{c^3} \left[3 \frac{\dot{Q}_{xx}}{r^2} \frac{D^2}{r^2} + 2 \frac{\dot{Q}_{xy}}{r^2} \frac{D}{r} \left(3 \frac{y}{r} - 1 \right) + \frac{\dot{Q}_{yy}}{r^2} \left(3 \frac{y^2}{r^2} - 2 \frac{y}{r} - 1 \right) \right] + \\ & + \frac{2G}{c^2} \left[3 \frac{Q_{xx}}{r^3} \frac{D^2}{r^2} + 6 \frac{Q_{xy}}{r^3} \frac{yD}{r^2} + \frac{Q_{yy}}{r^3} \left(3 \frac{y^2}{r^2} - 1 \right) \right] \end{aligned}$$

where all quadrupole moments $Q_{\alpha\beta}$ are assumed to be the functions of the retarded variable $t - r/c$.

Solution of the eikonal equation

(? что ли будем расписывать формулу выше?)

For simplicity the source of GW is assumed to emit only one frequency Ω . We consider the two following situations: either the GW source has been emitting since $t = -\infty$, or the signal is bounded in time, with an exponential decay.

In the former case one can write $\ddot{Q}_{\alpha\beta}(t) = q_{\alpha\beta} \sin \Omega t$.

Then substituting the expression for $F(t, r)$, see eq.(11), into eq. (5) one could write the solution for ψ_1 as an integral over the variable $z = y'/D - \sqrt{1 + (y'/D)^2}$:

$$\begin{aligned} \psi_1(\xi) = & -2 \frac{\omega}{c} \int_{D/(2y)}^{\infty} \frac{z^3 H_{yy} - 2z^2 H_{xy} + z H_{xx}}{(z^2 + 1)^2} \sin \left[\Omega \xi - 2\pi \frac{D}{\Lambda} z \right] dz + \\ & + 4 \frac{\omega}{c} \left(\frac{\Lambda}{2\pi D} \right) \int_{D/(2y)}^{\infty} \frac{z^2 (z^2 - 2) H_{yy} - 2(2z^3 - z) H_{xy} + 3z^2 H_{xx}}{(z^2 + 1)^3} \cos \left[\Omega \xi - 2\pi \frac{D}{\Lambda} z \right] dz + \\ & + 4 \frac{\omega}{c} \left(\frac{\Lambda}{2\pi D} \right)^2 \int_{D/(2y)}^{\infty} \frac{z (z^4 - 4z^2 + 1) H_{yy} - 6z^2 (z^2 - 1) H_{xy} + 6z^3 H_{xx}}{(z^2 + 1)^4} \sin \left[\Omega \xi - 2\pi \frac{D}{\Lambda} z \right] dz \end{aligned}$$

where the variable $\xi = t - y/c$ accounts for the effect of retardation of the light signal during its propagation to the point of observation, $H_{\alpha\beta} = 2Gq_{\alpha\beta}/c^4$, Λ is the wavelength of gravitational waves.

(?? потом глубже подумаю, какой смысл параметра H ???)

Assuming that at the point of observation $y \gg D$, one can safely set the lower limit for all integrals in this expression equal to zero. Then for large values of the impact parameter $2\pi D/\Lambda \gg 1$ one can asymptotically estimate these integrals and the main contribution to ψ_1 is given by the following expression:

$$\begin{aligned} \psi_1(\xi) & \approx \frac{1}{2\pi^2} \left(\frac{\Lambda}{D} \right)^2 \frac{\omega}{c} H_{xx} \sin \Omega \xi \\ & \approx \frac{h_{xx}}{\pi} \frac{\Lambda}{\lambda} \left(\frac{\Lambda}{D} \right)^2 \sin \Omega \xi \end{aligned}$$

where in the last line the answer is rewritten in terms of the dimensionless strength of gravitational waves: $h_{xx} = H_{xx}\Lambda^{-1}$.

One could see that the main contribution to the photon phase ψ comes from the h_{xx} component of gravitational radiation confirming the well known fact that only gravitational waves propagating perpendicular to the photon wave vector (directed along y axis in our geometry) have a considerable effect on the electromagnetic radiation.

Случай экспоненциального спада волн (?)

(?? а такое может быть?)

Let us now consider the case where gravity waves are emitted starting at $t = 0$ decaying exponentially with characteristic time τ_0 , то есть пусть

$$\ddot{Q}_{\alpha\beta}(t) = q_{\alpha\beta}\theta(t) \exp\left(-\frac{t}{\tau_0}\right) \sin \Omega t$$

где $\theta(x) = 0$ for $x < 0$ and 1 for $x > 0$.

Then for large values of the impact parameter, $2\pi D/\Lambda \gg 1$, the solution for ψ_1 is given by the following formula:

$$\psi_1(\xi) \approx \frac{h_{xx}}{\pi} \left(\frac{\Lambda}{D}\right)^2 \frac{\Lambda}{\lambda} \theta(\xi) e^{-\xi/\tau_0} \sin \Omega \xi$$

One can see that in this case at the point of observation the modulated contribution to the eikonal also decaying exponentially with the same characteristic time τ_0 .

(??? хотелось бы понять, что с того, ну нашли мы ответ и как его дальше использовать?)

The pre-exponential factor

Let us now go beyond the geometrical optics or eikonal approximation and calculate the pre-exponential factor for the photon wave function.

(??? место подумать о связи волнового уравнения и уравнения эйконала, пока слабо эту связь вижу)

Амплитуда фотона $\varphi(t, y)$ with волновым вектором k satisfies the wave equation:

$$\frac{1}{\sqrt{-g}} \frac{\partial}{\partial x^i} \left(\sqrt{-g} g^{ik} \frac{\partial \varphi}{\partial x^k} \right) = 0$$

where $g = \det g_{ik}$.

(??? круто, теперь что же делать с этим?)

Up to the first order contribution in h^{ik} one could find $g = -(1+h)$ with $h = h_i^i$. Taking the derivative over x^i and keeping only the first order terms in h^{ik} one could arrive to expression: (?? проверю потом)

$$g^{ik} \frac{\partial^2 \varphi}{\partial x^i \partial x^k} - \frac{\partial \varphi}{\partial x^k} \frac{\partial}{\partial x^i} \left(h^{ik} - \frac{1}{2} \eta^{ik} h \right) = 0.$$

(пропишу потом роль калибровок, ведь именно это ниже и позволяет продвинутся! пока не так глубоко понимаю)

One could see that under the harmonic gauge condition $\partial_i \tilde{h}_k^i = 0$ with $\tilde{h}_k^i = h_k^i - \frac{1}{2} \eta_k^i h$ the second term in above equation vanishes and the photon wave equation reduces to

$$g^{ik} \frac{\partial^2 \varphi}{\partial x^i \partial x^k} = 0$$

Let us write the photon wave function as $\varphi(t, y) = c(t, y)e^{i\psi}$, where ψ is the eikonal of the photon interacting with gravitational waves.

Then assuming that $c = 1 + a$, where a is the small contribution of the first order in h^{ik} , one could rewrite the above equation as

$$g^{ik} \left(\frac{\partial^2 a}{\partial x^i \partial x^k} + 2i \frac{\partial \psi}{\partial x^i} \frac{\partial a}{\partial x^k} + i(1+a) \frac{\partial^2 \psi}{\partial x^i \partial x^k} - (1+a) \frac{\partial \psi}{\partial x^i} \frac{\partial \psi}{\partial x^k} \right) = 0$$

The last term in this equation vanishes due to the eikonal equation (3)). (???? а вот тут вот не понял, у нас ведь не случай эйконального приближения, так что оно не выполняется ведь (????))

Since the free eikonal $\psi_0 = ky - \omega t$ is the linear function of x^i the all second derivatives of ψ_0 are vanish and the remaining wave equation up to the first order terms in h^{ik} takes the form:

$$\eta^{ik} \frac{\partial^2 a}{\partial x^i \partial x^k} + 2i\eta^{ik} \frac{\partial \psi_0}{\partial x^i} \frac{\partial a}{\partial x^k} + i\eta^{ik} \frac{\partial^2 \psi_1}{\partial x^i \partial x^k} = 0.$$

Assuming that the $a(t, y) = a(t - y/c)$ thus the first term above vanishes and one could finally find

$$\frac{\partial a}{\partial t} + c \frac{\partial a}{\partial y} = \frac{1}{2} \frac{c^2}{\omega} \left(\frac{1}{c^2} \frac{\partial^2 \psi_1}{\partial t^2} - \frac{\partial^2 \psi_1}{\partial y^2} \right)$$

Общее решение ψ_1 известно как интеграл от $F(t, y)$, так что получим

$$\frac{\partial a}{\partial t} + c \frac{\partial a}{\partial y} = \frac{c}{4} \left(\frac{\partial F}{\partial y} - \frac{1}{c} \frac{\partial F}{\partial t} \right)$$

(? не догоняю, что там дальше, короче, как-то вышла ф-ла ниже, потом посмотрю еще раз)

According to our previous analysis the main contribution to $a(t, y)$ is given by the H_{xx} component of gravitational radiation:

$$F(t, r) \approx H_{xx} \frac{\sin \Omega(t - r/c)}{r} \frac{D^2}{r^2}$$

В этом случае the solution for $a(\xi)$ for large values of impact parameter $2\pi D/\Lambda \gg 1$ has the form:

$$a(\xi) \approx -\frac{h_{xx}}{\pi} \left(\frac{\Lambda}{D} \right)^2 \cos \Omega \xi$$

Combining together eq. (14) with eq. (25) one can finally write the photon wave function at the point of observation in the form:

$$\varphi_k(\xi) = \frac{e^{-i\omega(\xi + \tau_g \sin \Omega \xi)}}{1 + \tau_g \Omega \cos \Omega \xi}$$

where we have introduced the time τ_g associated with integrated Shapiro time delay (see below) (?? так пока и не прописал его): $\tau_g = -\frac{h_{xx}}{\pi} \left(\frac{\Lambda}{D} \right)^2 \Omega^{-1}$

Это же решение for the photon wave function is valid for the case where gravitational radiation are bounded exponentially in time, however in this case τ_g depends on time and equals to

$$\tau_g(\xi) = -\frac{h_{xx}}{\pi} \left(\frac{\Lambda}{D} \right)^2 \Omega^{-1} \theta(\xi) e^{-\xi/\tau_0}$$

(не пойму, абзац выше важен?)

Modulation due to alternating Newtonian gravitational field

(Вот это пока что оставил, не читал, потом почитаю)

The observation and detection of the gravitational waves undoubtedly provides us a new astrophysical tool which can give us a deeper insight on the processes like the neutron star formation, the processes happening in close binaries at the final stage of it evolution, etc....

However the observation of gravitational waves gives us also an additional proof of the General Theory of Relativity itself and it seems reasonable to consider the effect of photon phase modulation in Newtonian theory of gravitation where the gravitational potential instantly adjusts to the current configuration of moving bodies.

Consider for example an oscillating neutron star or a rotating double star system which in Newtonian theory of gravity causes an alternating Newtonian gravitational potential $\phi(\mathbf{r}, t)$.

This alternating field can also lead to photon phase modulation and results in principle in the same effect like for the case of photon in the field of gravitational waves.

The aim of this section is to compare the effect of photon phase modulation between these two theories.

As the rotation or the oscillation frequency of the stars is much less than the frequency of the photon we can work in the same eikonal approximation as in the previous sections.

The Newtonian gravitational field contributes only to the diagonal metric elements: $h^{00} = h^{ii} = 2\phi/c^2$, where ϕ — is a gravitational potential at the point of observation.

At large distances compared to the size of the star system we can leave only quadrupole term in the potential

$$\phi(\mathbf{r}, t) = GQ_{\alpha\beta} \frac{n_\alpha n_\beta}{2r^3}$$

Here r - is the distance from the star to the photon, the coordinate axis are chosen to be the same as before, so $r = \sqrt{y^2 + D^2}$, n_i — the direction vectors of $\mathbf{r}$: $n_x = D/\sqrt{y^2 + D^2}$, $n_y = y/\sqrt{D^2 + y^2}$, $n_z = 0$, $Q_{\alpha\beta}(t)$ is the alternating tensor of the mass quadrupole moment of the star.

The values of $Q_{\alpha\beta}$ are determined by the parameters of the star, like orientation of the angular velocity, eccentricity, etc.

We will further discuss the general case, assuming only, that due to rotation or oscillation $\ddot{Q}_{\alpha\beta}(t) = q_{\alpha\beta} \sin \Omega t$ The correction to the eikonal due to rotation is given by the same equation (5) as for the GW case solution of the eikonal equation, where in our case $F(t, y) = h^{00} + h^{yy} = 2h^{00}$

$$\Psi_n = -\frac{\omega}{c} \int_{-\infty}^y h^{00} \left(t - \frac{y - y'}{c}, y' \right) dy'$$

Substituting the expression for ϕ into the formula above and making the substitution $z = y'/D$ one obtains:

$$\Psi_n(\xi) = \frac{1}{2} \frac{\omega}{c} \left(\frac{\Lambda}{2\pi D} \right)^2 \int_{-\infty}^{y/D} \frac{H_{xx} + 2H_{xy}z + H_{yy}z^2}{(1 + z^2)^{5/2}} \sin \left[\Omega \xi + 2\pi \frac{D}{\Lambda} z \right] dz$$

where we have introduced $H_{\alpha\beta} = 2Gq_{\alpha\beta}/c^4$ Then for large values of the impact parameter $2\pi D/\Lambda \gg 1$ and large distances $y/D \gg 1$ one can obtain the following approximate expression for Ψ_n :

$$\Psi_n(\xi) = \frac{1}{6} \frac{\omega}{c} \sqrt{\frac{\Lambda}{D}} e^{-2\pi D/\Lambda} [(H_{xx} - H_{yy}) \sin \Omega \xi + 2H_{xy} \cos \Omega \xi]$$

One can see that the photon phase modulation due to the Newtonian alternating potential vanishes exponentially for large impact parameters in contrast to the GW modulation which vanishes as $1/D^2$ (see eq.

(14)).

Technically the difference between Newtonian and GW cases follows from the different y' dependence of the $F(t, y')$ function.

For the properly retarded GW radiation the solution for ψ_1 , see eq.

(14), involves the oscillating integral over $\sin \Omega [\xi + (y' - \sqrt{(y')^2 + D^2})/c]$, where the oscillating dependence on y' near the source $|y'| \sim D$ almost vanishes making the light-GW interaction more effective.

Oppositely for the instantly adjusted Newtonian radiation one has in eq.

(30) fast oscillating integral over y' with $\sin \Omega (\xi + y'/c)$ near vicinity of the gravitational wave source, which leads to a very small value of modulation.

О применении этих расчетов в других случаях (??!!!)

(очень важно, пойму - пропишу, как это использовать в других ситуациях!!! пока понять бы метод!!)

Эффект Шапиро

(если понадобится - буду развивать этот раздел, пока просто знаю, что такое есть и небольшая теория про него)

The Shapiro effect in GW

(наверное, ниже важные выкладки, но как-то я проморгал их...)

Let us calculate the integrated time delay of the light signal propagating in the field of gravitational waves.

Consider an arbitrary wave packet emitted by some distant light source traveling toward to Earth near a source of GW:

$$\varphi(\xi) = \int \varphi_{\omega} e^{-i\omega\xi} \frac{d\omega}{2\pi}$$

where $\xi = t - y/c$.

(интересно, что пакет так записывается, не знал, пропишу у себя потом это тоже)

Using the results for monochromatic, photons propagating in the field of gravitational radiation

(??!!! пока не вижу, какие конкретно это результаты??!!! потом перечитаю, что там выше?)

one can find the signal at the point of observation:

$$\begin{aligned} \varphi_{ob}(\xi) &= \int \varphi_{\omega} e^{-i\omega[\xi + \tau_g \sin \Omega \xi]} \frac{d\omega}{2\pi} \\ &= \varphi(\xi + \tau_g \sin \Omega \xi) \end{aligned}$$

One can see that the interaction with the gravitational radiation leads to alternating time delay of the signal

$$\delta t = \tau_g \sin \Omega,$$

(note that the τ_g is negative, see eq. (27))

This effect was suggested independently first by Sazhin and Detweiler [16] for detection of gravitational waves using the timing observation of a pulsar the line of sight to which passes near the source of GW.

О применениях эффекта Шапиро

(хз, но однажды узнаю, какая польза в нем)

Детектирование волн

(вторая часть статьи Лесовика, пока для просто кругозора. в оптике многое дополню, чтобы понимать, как это работает?)

Теория

Let us now turn to the discussion about how the photons modulated by the interaction with the gravitational waves can be measured in some realistic setup.

Consider first the simple case where the modulated photons (26) simply hit a photodetector which can react at all photon's frequencies.

However, any real photodetector reacts not on the vector potential $\mathbf{A}(t)$ of the photons in the photodetector directly, but rather on the electrical field, $\mathbf{E}(t)$, induced by the photons in this photodetector.

Then the one photon photo-detection probability, $P(t)\Delta t$, to observe a photon during time interval Δt can be expressed as [26] :

$$P(t)\Delta t = \alpha \frac{Sc\Delta t}{E_{det}} \left\langle \hat{\mathbf{E}}^{(-)}(t, \mathbf{r}) \cdot \hat{\mathbf{E}}^{(+)}(t, \mathbf{r}) \right\rangle$$

where S is the area of the photodetector, E_{dec} some characteristic energy describing the interaction of the photodetector with incoming photons, α is the quantum efficiency of the photo-detection, $\hat{\mathbf{E}}^{(+)}(t, \mathbf{r})$ is the positive frequency part of the electrical field operator, $\hat{\mathbf{E}}^{(-)}(t, \mathbf{r})$ is the Hermitian conjugate of $\hat{\mathbf{E}}^{(+)}(t, \mathbf{r})$:

$$\hat{\mathbf{E}}^{(+)}(t, \mathbf{r}) = -\frac{\sqrt{4\pi}}{L^{3/2}} \sum_{\mathbf{k}s} \sqrt{\frac{\hbar}{2\omega_k}} \dot{\varphi}_k(t, \mathbf{r}) \mathbf{e}_{ks} \hat{a}_{ks}$$

At the point of observation we set $\mathbf{r} = 0$ to shorten the notations.

Taking the time derivative of the photon wave function (26) one gets the photo-detection probability equals to

$$\begin{aligned} P(t)\Delta t &= \frac{4\pi\alpha Sc\Delta t}{E_{det}L^3} \sum_{\mathbf{k}s} \frac{\hbar\omega_k}{2} n_{ks} \\ &= 2\alpha \frac{S}{r^2} \Delta t \int d\omega \frac{\hbar\omega}{E_{det}} n(\omega) \end{aligned}$$

where r is the distance between the light source and the Earth, $n(\omega)d\omega$ is the number of the photons with frequency ω , emitted by the distant light source per unit time, into a unit angular domain.

One can see that the photo-detection probability does not feel any photon phase modulation due to GW and this probability is the same as it would be for the monochromatic photons emitted by a distant light source.

Thus for such a simple setup it is not possible to retrieve any information about the gravity wave from the photodetector signal. Consider however a more realistic setup where the

photodetector has a sensitivity edge ω_s which means that photodetector does not react on the photons with frequency less than ω_s .

One can treat this situation as if the photons passed through a filter with transparency $t(\omega) = \theta(\omega - \omega_s)$, before hitting the photodetector.

It is convenient to write this transparency as an operator acting on the photon wave function: $\hat{t} = \theta[i\partial_t - \omega_s]$.

Then after passing such a filter the photon has a wave function equals to $\hat{t}\varphi_k(t) \approx \theta[\omega_k(t) - \omega_s] \varphi_k(t)$, where $\omega_k(t) = \omega_k[1 + \Omega\tau_g \cos \Omega t] = \omega_k f(t)$.

Then for the photo-detection probability one has:

$$P(t)\Delta t \approx 2\alpha \frac{S\Delta t}{r^2} \int d\omega \frac{\hbar\omega}{E_{\text{det}}} \theta[\omega f(t) - \omega_s] n(\omega)$$

Consider now the case where photons coming to the photodetector from a sharp Lorentzian spectral line with maximum, ω_0 , near the photodetector sensitivity edge $\omega_0 \approx \omega_s$ and width Γ_0 :

$$n(\omega) = n(\omega_0) \frac{\Gamma_0^2}{(\omega - \omega_0)^2 + \Gamma_0^2}$$

then according to the eq.

(38) the photo-detection probability for the modulated photons (26) has an appreciable contribution which is periodic in time at the GW frequency, and which is proportional to the strength of gravitational waves:

$$P(t)\Delta t \approx P_0\Delta t \left(1 + \Omega\tau_g \frac{2\omega_0}{\pi\Gamma_0} \cos \Omega t \right)$$

where $P_0\Delta t$ is the photo-detection probability for non modulated photons. From the above analysis one can conclude that in order to detect the modulation of the photons by GW one should place before the photodetector some filter which has a finite frequency band.

Then the formula for photo-detection probability (38) with a sensitivity edge has a rather general sense.

In fact let us consider an arbitrary filter placed before the photodetector with transparency $T(\omega)$.

Then provided that the time spent by the photon in this filter (that is an inverse frequency band of the filter) is much smaller than frequency of the GW, Ω , one can treat the modulated photons as monochromatic photons with slowly varying frequency $\omega_k(t) = \omega_k[1 + \Omega\tau_g \cos \Omega t] = \omega_k f(t)$.

In this case the photo-detection probability can be written simply as

$$P(t)\Delta t \approx 2\alpha \frac{S\Delta t}{r^2} \int d\omega \frac{\hbar\omega}{E_{\text{det}}} T(\omega f(t)) n(\omega)$$

From this expression one can see that there are two different regimes of GW detection in this proposed setup.

Consider the first regime where the distribution function of the light source, $n(\omega)$, has a sharp peak of the width Γ_0 (or in more general case some sufficient irregularity like the edge of the spectrum) centered near the transmission window of the filter.

Then, provided that the amplitude of the photon's frequency modulation, $\omega_k\Omega\tau_g$, is much bigger than the width of the spectrum irregularity, $\omega_k\Omega\tau_g \gg \Gamma_0$, the resulting photo-detection signal will be a periodic sequence of peaks.

In the opposite case where the light spectrum $n(\omega)$ is a slowly varying function of the frequency within the transparency window of the filter one can consider $n(\omega)$ as a constant in eq.

(41).

Then the time dependence of the photo-detection probability can be written as:

$$P(t)\Delta t \approx 2\alpha \frac{S\Delta t}{r^2} \frac{1}{f(t)} \int d\omega \frac{\hbar\omega}{E_{det}} T(\omega) n(\omega)$$

where $P_0\Delta t$ is the photo-detection probability for non modulated photons. From the above analysis one can conclude that in order to detect the modulation of the photons by GW one should place before the photodetector some filter which has a finite frequency band.

Then the formula for photo-detection probability (38) with a sensitivity edge has a rather general sense.

In fact let us consider an arbitrary filter placed before the photodetector with transparency $T(\omega)$.

Then provided that the time spent by the photon in this filter (that is an inverse frequency band of the filter) is much smaller than frequency of the GW, Ω , one can treat the modulated photons as monochromatic photons with slowly varying frequency $\omega_k(t) = \omega_k [1 + \Omega\tau_g \cos \Omega t] = \omega_k f(t)$.

In this case the photo-detection probability can be written simply as

$$P(t)\Delta t \approx 2\alpha \frac{S\Delta t}{r^2} \int d\omega \frac{\hbar\omega}{E_{det}} T(\omega f(t)) n(\omega)$$

From this expression one can see that there are two different regimes of GW detection in this proposed setup.

Consider the first regime where the distribution function of the light source, $n(\omega)$, has a sharp peak of the width Γ_0 (or in more general case some sufficient irregularity like the edge of the spectrum) centered near the transmission window of the filter.

Then, provided that the amplitude of the photon's frequency modulation, $\omega_k \Omega \tau_g$, is much bigger than the width of the spectrum irregularity, $\omega_k \Omega \tau_g \gg \Gamma_0$, the resulting photo-detection signal will be a periodic sequence of peaks.

In the opposite case where the light spectrum $n(\omega)$ is a slowly varying function of the frequency within the transparency window of the filter one can consider $n(\omega)$ as a constant in eq.

(41).

Then the time dependence of the photo-detection probability can be written as:

$$P(t)\Delta t \approx 2\alpha \frac{S\Delta t}{r^2} \frac{1}{f(t)} \int d\omega \frac{\hbar\omega}{E_{det}} T(\omega) n(\omega)$$

Analysis of the light signal with an interferometer

In this section we consider in detail the setup where at the point of observation - the Earth - the modulated light signal before hitting the photodetector passes through an interferometer, which is characterized by a complex transmission amplitude $t(\omega)$.

We will assume that the light passing through a Fabry-Perot filter comes from a single Lorentzian spectral line of the type (39).

For simplicity, let us first assume that there is only one FP resonance within this spectral line.

The transparency of FP is written in the usual way:

$$t(\omega) = \frac{i\Gamma/2}{\omega - \omega_0 + i\Gamma/2}$$

Here a derivation of the photo-detection probability is provided.

It relies on the assumption $\Omega\tau_g \ll 1$, which is relevant for experimental situations (see below).

Consider the propagation of a photon wave-packet $\varphi_k(t)$ (26) through the FP interferometer.

The resulting wave packet after the FP can be written $\phi_k(t) = \hat{t}\varphi_k(t)$, where $\hat{t}$ is the transparency operator of the FP in real time representation:

$$\hat{t} = \frac{i\Gamma}{2} [i\partial_t - \omega_0 + i\Gamma/2]^{-1}$$

where Γ is the width of FP resonance.

However the real FP filter does not act directly on the vector potential of the incoming photons but rather on the electromagnetic field induced by the these photons.

Using similar arguments the positive frequency part of electrical field operator $\mathbf{E}^{(+)}(t, r)$, after passing the Fabry-Perot filter, can be written in terms of the transmission operator of the filter:

$$\hat{\mathbf{E}}^{(+)}(t, \mathbf{r}) = -\frac{\sqrt{4\pi}}{L^{3/2}} \sum_{\mathbf{k}s} \sqrt{\frac{\hbar}{2\omega_k}} [\hat{t}\dot{\varphi}_k(t, \mathbf{r})] \mathbf{e}_{ks} \hat{a}_{ks}$$

With these notations the photo-detection probability can be written as

$$P(t) = 2\alpha \frac{S}{r^2} \int \frac{\hbar d\omega}{\omega E_{det}} [\hat{t}\dot{\varphi}_k(t)]^* [\hat{t}\dot{\varphi}_k(t)] n(\omega) d\omega$$

The action of the operator $\hat{t}$ on the function $\dot{\varphi}_k(t)$ can be found using the Green's function formalism:

$$\hat{t}\dot{\varphi}_k(t) = \frac{\Gamma}{2} \int_{-\infty}^t e^{-i\omega_0(t-\tau) - \frac{\Gamma}{2}(t-\tau)} \dot{\varphi}_k(\tau) d\tau$$

Substituting this expression into eq.

(47) in the limit $\omega_0 \gg \Gamma$ one finally obtains the following expression for the photo-detection probability:

$$P(t) = 2\alpha \frac{S}{r^2} \frac{\hbar\omega_0}{E_{det}} \frac{\Gamma^2}{4} \int n(\omega) e^{-\Gamma t} d\omega \int_{-\infty}^t \int_{-\infty}^t e^{-i\omega_0(\tau-s) + \frac{\Gamma}{2}(\tau+s)} \dot{\varphi}_k^*(\tau) \dot{\varphi}_k(s) d\tau ds$$

10 Performing the integral over ω and introducing the relative time $\xi = \tau - s$, and the total time $\eta = (\tau + s)/2$, in the working limit $\Omega\tau_g \ll 1$ one can write the photo-detection probability as:

$$P(t) = \alpha\pi \frac{S}{r^2} \frac{\hbar\omega_0}{E_{det}} n(\omega_0) \Gamma_0 \frac{\Gamma^2}{4} \int_{-\infty}^t d\eta e^{-\Gamma(t-\eta)} \int_{-2(t-\eta)}^{2(t-\eta)} e^{i\xi[\omega_0\Omega\tau_g \cos \Omega\eta]} e^{-\frac{\Gamma_0(\eta)}{2}|\xi|} d\xi$$

Performing the integration over ξ in the limit $\Gamma_0 \gg \Gamma$ the leading contribution to the photo-detection probability is given by the following expression:

$$P(t) = \alpha\pi \frac{S}{r^2} \frac{\hbar\omega_0}{E_{det}} n(\omega_0) \Gamma_0 \frac{\Gamma^2}{4} \int_0^\infty \frac{\Gamma_0(t-z) e^{-\Gamma z} dz}{(\omega_0\Omega\tau_g)^2 \cos^2 \Omega(t-z) + \Gamma_0^2(t-z)/4}$$

where we have made the substitution $z = (t - \eta)$ and we introduced the notation $\Gamma_0(t) = \Gamma_0 |1 + \Omega\tau_g \cos \Omega t|$.

“Strong” GW or narrow spectrum

(кст, сильные грав волны мне так и не зашли, потом мб пропишу их)

Let us consider the limit where the frequency broadening of the initially monochromatic photon wave packet due to the interaction with gravitational waves, $\delta\Gamma = \omega_0\Omega\tau_g$, is much bigger than the width of the spectral line, Γ_0 .

The condition $\delta\Gamma \gg \Gamma_0$ can be achieved both in the case where the amplitude of GW is strong enough or when one has a very narrow spectral line.

In this limit one can safely neglect the dependence on $|1 + \Omega\tau_g \cos \Omega t|$ and put $\Gamma_0(t) = \Gamma_0$ in eq.

(51) assuming $\Omega\tau_g \ll 1$.

Then the leading contribution to the photo-detection probability can be written in terms of the occupation number, $n(\omega)$, of the spectral line:

$$P(t) = \alpha\pi \frac{S}{r^2} \frac{\hbar\omega_0}{E_{\text{det}}} \Gamma^2 \int_0^\infty n[\omega_0(t-z)] e^{-\Gamma z} dz$$

where $\omega_0(t) = \omega_0 (1 + \Omega\tau_g \cos \Omega t)$.

This result has a very simple qualitative explanation which is consistent with our previous simple arguments for the photodetector with sensitivity edge (see eq.

(41)).

Consider a photon wave packet coming to FP filter: $\varphi_k = \mathcal{C}(t)e^{-i\omega_k(t+\tau_g \sin \Omega t)}$.

Such a wave packet in the limit $\Omega \ll \omega_k$ can be interpreted as a photon with a frequency which is slowly varying in time $\omega_k(t) = \omega_k (1 + \Omega\tau_g \cos \Omega t)$.

If however the time spent by this photon in the FP filter, Γ^{-1} , much smaller than the characteristic time of the frequency change Ω^{-1} one can assume the photon at any instant of time to be monochromatic with frequency $\omega_k(t)$.

Then one can take advantage of our eq.

(41).

Performing the integration over ω in eq.

(41) with Lorentzian spectral line $n(\omega)$ (39) one has

$$P(t) = \alpha\pi \frac{S}{r^2} \frac{\hbar\omega_0}{E_{\text{det}}} n[\omega_0(t)] \Gamma$$

On the other hand this result can be derived directly from our elaborate analysis by putting $n(\omega_0(t-z)) = n(\omega_0(t))$ in eq.

(52).

The exponential factor in eq.

(52) describes a time delay of the photon wave packet propagating through the FP filter.

The result eq.

(53) describes the periodic sequence of symmetric peaks with a half period of gravitational waves π/Ω and widths equal to

$$\tau = \frac{\Gamma_0}{\delta\Gamma} \Omega^{-1}$$

If however the width of these peaks becomes smaller than the time needed for the photon to penetrate through the FP filter: $\tau \ll \Gamma^{-1}$ then the form of peaks becomes asymmetric and the more accurate expression (52) has to be used. It should be noted that the expression (52) is valid not only for a single spectral line (39) but also for arbitrary distribution function for the light source, where there is a sufficient irregularity in $n(\omega)$ like, for example, at the edge of the spectrum.

Indeed if one consider the arbitrary distribution function $n(\omega)$ which changes much at frequency scale $\Delta\omega$ around ω_0 then it follows from eq.

(52) that one again has a periodical sequence of peaks for photo-detection signal in regime $\delta\Gamma \gg \Delta\omega$.

In this sense the described effects (52), (53) have nothing to do with the interference of the different components of the photon wave packet (43) and deal rather with the effect of the photon frequency modulation $\omega_k(t) = \omega_k(1 + \Omega\tau_g \cos \Omega t)$ due to the interaction with the GW.

If however the amplitude of the frequency modulation $\delta\Gamma$ becomes smaller than the characteristic scale of the spectral function $n(\omega)$ (the case of a broad spectrum) the effects associated with frequency modulation completely disappear.

Слабая гравитационная волна и широкий спектр

Consider now the opposite limit to the previous case when the amplitude of GW is small or when the width of the spectral line is large enough so that the amplitude of the frequency modulation, $\delta\Gamma$, is much smaller than the width of the spectral line, Γ_0 .

It should be noted that this situation is the most common situation in practice since the amplitude of GW is extremely small even in the vicinity of GW source.

In the limit $\delta\Gamma \ll \Gamma_0$ one can safely expand the integral expression in eq.

(51) up to the lowest order in the small ratio $\delta\Gamma/\Gamma_0 \ll 1$ and perform the integration over z .

Keeping all terms of the type $1 + \Omega\tau_g \cos \Omega t$ in the working assumption $\Gamma \gg \Omega$ the photo-detection probability can be now written as:

$$P(t) \approx \alpha\pi \frac{S}{r^2} \frac{\hbar\omega_0}{E_{\text{det}}} n(\omega_0) \Gamma \left(1 - \Omega\tau_g \cos \Omega t - 2 \frac{\delta\Gamma^2}{\Gamma_0^2} \cos 2\Omega t \right)$$

One can see that in the case of a broad spectrum one obtains small oscillations against a huge constant background with frequencies Ω and 2Ω suppressed by small factors $\Omega\tau_g \ll 1$ and $\delta\Gamma^2/\Gamma_0^2 \ll 1$, correspondingly.

However as one can see from eq.

(55) even for the extreme case of infinitely broad spectral line still there is a time dependent contribution to the $P(t)$

$$P(t) \approx \alpha\pi \frac{S}{r^2} \frac{\hbar\omega_0}{E_{\text{det}}} n(\omega_0) \Gamma [1 - \Omega\tau_g \cos \Omega t]$$

It should be noted that this expression does not depend absolutely on the form of the spectra of the distant light source and the time modulation of the photo-detection signal defined only by the strength of the gravitational waves.

The amplitude of the photo-detection probability in this case proportional to the number of photons $n(\omega_0) \Gamma$ coming to the photodetector per unit time and practically one can use the FP filter with appropriate width Γ to collect the enough number of photons.

The second important point which has to be emphasized here is that the time dependent term in eq.

(56) has nothing to do with the effect of the frequency modulation mentioned above (see eq.

52,53) and presents an absolutely different effect associated with interference of the different components of the photon's wave packet (43) after passing the FP filter.

To clarify the nature of this interference effect we will consider in the next section an example where there are two extremely narrow FP resonances, $\Gamma \ll \Omega$, within the broad spectrum of the distant light source.

Two resonances case. Let us consider a more general case where two narrow resonances at frequencies $\omega_0 \pm \Delta/2$ are present within the spectral line (Δ is the separation between resonances).

In this case the transmission amplitude of the FP filter, $t(\omega)$ again can be written as an operator acting on the e.m.f.

induced by modulated photons: $\hat{t} = \hat{t}_1 + \hat{t}_2$, where $\hat{t}_{1(2)}$ corresponds to the first (second) resonance:

$$\hat{t}_{1(2)} = e^{\pm i\phi} \frac{i\Gamma}{2} \left[i\partial_t - \omega_0 \pm \frac{\Delta}{2} + i\frac{\Gamma}{2} \right]^{-1}$$

where 2ϕ is the relative phase difference between resonances. Using similar type of arguments and the same method of calculation as for the case of FP filter with a single resonance, in the limit $\Gamma_0 \gg \Omega, \Gamma_0 \gg \Gamma$ one arrives at the following expression for the photo-detection probability:

$$P(t) = \alpha \frac{\pi S}{r^2} \frac{\hbar\omega_0}{E_{\text{det}}} n(\omega_0) \Gamma_0 \frac{\Gamma^2}{4} [P_1(t) + 2P_{12}(t) + P_2(t)]$$

where

$$P_{1(2)}(t) = \int_0^\infty \frac{\Gamma_0(t-z)e^{-\Gamma z} dz}{\left(\delta\Gamma \cos \Omega(t-z) \pm \frac{\Delta}{2}\right)^2 + \frac{\Gamma_0^2(t-z)}{1}}$$

$$P_{12}(t) = \int_0^\infty \frac{\Gamma_0(t-z) \cos(\Delta z - \phi) e^{-\Gamma z} dz}{\delta\Gamma^2 \cos^2 \Omega(t-z) + \frac{\Gamma_0^2(t-z)}{4}}$$

Consider now the extreme limit of infinitely broad spectrum: $\Gamma_0 \rightarrow \infty$, then in the limit $\Gamma \gg \Omega, \Gamma \ll \Delta$ one has:

$$P(t) = \alpha \pi \frac{S}{r^2} \frac{\hbar\omega_0}{E_{\text{det}}} n(\omega_0) (2\Gamma) [1 - \Omega\tau_g \cos \Omega t]$$

Comparing this result with the expression for the photo-detection probability for the single resonance case (56) one can see that in the regime of a broad spectrum $\Gamma_0 \rightarrow \infty$ the photo-detection probability does not depend on the number of resonances in FP filter and depends rather on the total transparency of the filter (2Γ in present case). Consider however the opposite limit of extremely narrow resonances $\Gamma \ll \Omega$ (this is hardly possible in practice).

Then the resonance factor with respect to separation between resonances, Δ , appears in the expression for photodetection probability:

$$P(t) \approx \alpha \pi \frac{S}{r^2} \frac{\hbar\omega_0}{E_{\text{det}}} n(\omega_0) (2\Gamma) \left[1 - \frac{\Omega\tau_g \Gamma}{2} \frac{\Gamma \cos(\Omega t - \phi) + (\Omega - \Delta) \sin(\Omega t - \phi)}{\Gamma^2 + (\Omega - \Delta)^2} \right]$$

One can see that there is a parametric resonance for the photo-detection signal at $\Delta = \Omega$.

The photo-detection probability at the resonance equals to:

$$P_{\Delta=\Omega}(t) \approx \alpha \frac{\pi S}{r^2} \frac{\hbar\omega_0}{E_{\text{det}}} n(\omega_0) (2\Gamma) \left[1 - \frac{\Omega\tau_g}{2} \cos(\Omega t - \phi) \right]$$

while for $|\Delta - \Omega| \gg \Gamma$ (out of resonance) the term which depends on time in the photo-detection probability is saturated by the small factor $\Gamma/|\Delta - \Omega|$

A similar resonance appears in regime $\Gamma \ll \Omega$ for the case where the width of the spectrum $n(\omega)$ is broad but $(\omega_0/\Gamma_0)^2 \Omega\tau_g \gg 1$ so one can neglect all factors $|1 + \Omega\tau_g \cos \Omega t|$ in eq. (58)

$$P(t) \approx \alpha \frac{\pi S}{r^2} \frac{\hbar\omega_0}{E_{\text{det}}} n(\omega_0) (2\Gamma) \left[1 - \frac{\delta\Gamma^2}{\Gamma_0^2} \frac{\Gamma \cos(2\Omega t - \phi) + (2\Omega - \Delta) \sin(2\Omega t - \phi)}{\Gamma^2 + (2\Omega - \Delta)^2} \right]$$

One can see that the parametric resonance at $\Delta = 2\Omega$ appears in this case.

Now we argue that the resonances at $\Delta = \Omega$ and $\Delta = 2\Omega$ (62), (64) arises because of the interference of different components of the photon's wave packet (26).

To show it let us simply write the transparency of the FP filter as $t(\omega) = \delta_\Gamma(\omega - \omega_0) + \delta(\omega - \omega_0 - \Delta)$ where $\delta_\Gamma(\omega)$ is a narrow peaked function with finite width $\Gamma \ll \Omega$.

Then using the expansion of the photon wave packet (43) one can write the photo-detection probability in terms of the sums over different components of the photon's wave packet:

$$P(t) = 2\alpha \frac{S}{r^2} \sum_{n,m=-\infty}^{\infty} \int \frac{\hbar\omega}{E_{\text{det}}} n(\omega) d\omega e^{i(n-m)\Omega t} t^*(\omega + n\Omega) t(\omega + m\Omega) J_n(\omega\tau_g) J_m(\omega\tau_g)$$

Let us assume that the distance between resonances are exactly equal to $N\Omega = \Delta$ where N is an integer.

Then substituting the transmission amplitude $t(\omega)$ into this expression and making the integration over ω one gets:

$$P(t) \approx 4\alpha \frac{S}{r^2} \sum_{n=-\infty}^{\infty} n(\omega_0 - n\Omega) \Gamma \frac{\hbar(\omega_0 - n\Omega)}{E_{\text{det}}} [J_n^2[(\omega_0 - n\Omega)\tau_g] + J_n[(\omega_0 - n\Omega)\tau_g] J_{n+N}[(\omega_0 - n\Omega)\tau_g] \cos 2\Omega t]$$

Then for the special choice $N = 2$ assuming that the argument of the Bessel function is typically large one can put $J_{n+2}(x) \approx -J_n(x)$ for $x \gg 1$.

Then the photo-detection probability can be written as

$$P(t) \approx 4\alpha \frac{S}{r^2} (1 - \cos 2\Omega t) \sum_{n=-\infty}^{\infty} n(\omega_0 - n\Omega) \Gamma \frac{\hbar(\omega_0 - n\Omega)}{E_{\text{det}}} J_n^2[(\omega_0 - n\Omega)\tau_g]$$

One can see that the term which is alternating in time in $P(t)$ arises due to interference of the different components of the photon wave packet (26).

Analysis of intensity correlations

(здесь вот такое прямо детектирование самое самое и пошло)

In the previous section, we showed that by appropriate filtering of the light signal, it is possible to extract oscillations associated with the past interaction with the GW.

While this previous proposal is promising and attractive, one can

take an alternative route for the detection of the GW signal: the measurement of intensity-intensity correlations of the photo-detection signal.

Consider the correlator $\langle\langle I(t_1) I(t_2) \rangle\rangle$: because of the time modulation due to the presence of the GW, it no longer depends on the variable $t_1 - t_2$ only.

Using the definition the intensity operator, one can see qualitatively that the Fourier components of Eq.

(43) will lead to a two particle interference effect.

The quantity to be measured on the Earth is thus the intensity correlator of the photo-detection signal which is the time ordered product of electrical field operators:

$$\langle\langle \hat{I}(t_1) \hat{I}(t_2) \rangle\rangle dt_1 dt_2 = \alpha^2 \frac{S^2 c^2 dt_1 dt_2}{E_{\text{det}}^2} \langle\langle \hat{\mathbf{E}}^{(-)}(t_1) \hat{\mathbf{E}}^{(-)}(t_2) \hat{\mathbf{E}}^{(+)}(t_2) \hat{\mathbf{E}}^{(+)}(t_1) \rangle\rangle$$

Substituting here the expression for the electrical field operators (36) and performing the quantum average one gets

$$\begin{aligned} \langle\langle : \hat{I}(t_1) \hat{I}(t_2) : \rangle\rangle dt_1 dt_2 &= 2\alpha^2 \frac{S^2 dt_1 dt_2}{r^4} \int d\omega_1 d\omega_2 e^{-i\omega_2[t_1-t_2+\tau_g(\sin\Omega t_1-\sin\Omega t_2)]} \\ &\times \frac{\hbar^2 \omega_1 \omega_2}{E_{det}^2} n(\omega_1) n(\omega_2) e^{i\omega_1[t_1-t_2+\tau_g(\sin\Omega t_1-\sin\Omega t_2)]} \end{aligned}$$

This correlator can be rewritten as

$$\langle\langle : \hat{I}(t_1) \hat{I}(t_2) : \rangle\rangle dt_1 dt_2 = 2\alpha^2 \frac{S^2 dt_1 dt_2}{r^4} |f[t_1 - t_2 + \tau_g(\sin\Omega t_1 - \sin\Omega t_2)]|^2$$

where

$$f(t) = \int \frac{\hbar\omega}{E_{det}} n(\omega) e^{-i\omega t} d\omega$$

The function $f(t)$ typically decays exponentially on a time scale given by the inverse of the line width $\Gamma_0 \gg \Omega$.

Thus one can safely linearize the dependence of the intensity-intensity correlator with respect to the relative time $\tau = t_1 - t_2$

$$\langle\langle : \hat{I}(t+\tau) \hat{I}(t) : \rangle\rangle = 2\alpha^2 \frac{S^2}{r^4} |f[\tau(1 + \Omega\tau_g \cos\Omega t)]|^2$$

where $\tau = t_1 - t_2$ is the relative time and t is the total time.

Let us now calculate the spectral power, $S(t)$ of the photons' noise defined as

$$S(t) = \int \langle\langle : \hat{I}(t+\tau) \hat{I}(t) : \rangle\rangle d\tau$$

Substituting here the expression for photon intensity correlator one can show that

$$S(t) = \frac{4\pi\alpha^2}{1 + \Omega\tau_g \cos\Omega t} \frac{S^2}{r^4} \int \frac{(\hbar\omega)^2}{E_{det}^2} n^2(\omega) d\omega$$

One can see from this equation that the intensity correlation contain periodic oscillations - although with a small amplitude - in the total time.

Note that contrary to the Fabry-Perot diagnosis, here a rather accurate measurement is implied, as the contribution to other noise sources (due to scattering...) needs to be minimized compared to this periodic signal.

In particular, it requires a large time acquisition window to filter out spurious fluctuations.

10.1 Другие методы детектирования волн

10.1.1 Методы на основе взаимодействия с электромагнитным полем

(выпишу их тут конкретно с формулами)

10.1.2 Методы на основе запаздывания времени вблизи пульсаров

(выпишу их тут конкретно с формулами, Лесовик вроде именно этот метод обсуждает)

10.2 Другие эксперименты

10.2.1 On ETrpathfinder

(так называется еще одна установка для экспериментов. но там всего 9 цитат за более чем год, так что мб проще забить на нее.)

10.2.2 Простейший эксперимент по детектированию волн

??

10.2.3 Исторически первый эксперимент

он был предложен в 1957 году Бонди и Фейнманом
это был стержень с бусинками, закрепленным на гироскопе.
(???? потом пропишу)

Part IV

Problems

11 Основные задачи о гравитационных волнах

(!! тут то, что понятно, как решается! разные вопросы, на которые мало кто ответит - дело предпоследней части про другое о грав волнах.)

11.0.1 Задачи на идейное понимание гравитационных волн

(особые вопросы теории, много что спросить можно. все это или отвечено или сказано все для ответа на это в теории! Многое можно спросить)

Изменение объема (?????)

(сформулирую потом, типа объем не меняется, напишу еще раз доказательство.)

11.0.2 Задачи на понимание формул гравитационных волн

(особые вопросы теории, много что спросить можно.)

11.0.3 Какие будут гравитационные волны от данной вращающейся системы?

(просто примеры и задачи на решаю когда-то)

12 Другие задачи о гравитационных волнах

12.1 Задачи на гравитационные волны в модифицированных теориях гравитации

(тоже наверняка многие есть, пока не до них вообще)

Part V

Other Models and Theories

(тут то, что решал и разбирал)

13 Различные волны и явления

13.1 Волны от разных источников

13.1.1 Волны от взрыва

(вот интересно, как сильно должен взорваться объект, чтобы от него были большие грав волны? подумаю потом. мб так и можно было бы их детектировать? хз.)

13.2 Разные объемные волны

(всем этим только по приколу занимаюсь, всерьез - никогда, ничего иначе не сделаешь!)

13.2.1 Объёмные волны Бонди - Пирани - Робинсона

Суть

Эти волны описываются метрикой вида

$$ds^2 = (dx^0)^2 - \left((dx^1)^2 + \alpha (dx^2)^2 + 2\beta (dx^2 dx^3) + \gamma (dx^3)^2 \right).$$

Если ввести переменную $u = x^0 - x^1$ и функцию $\lambda = \alpha\gamma - \beta^2$, то из уравнений ОТО получим уравнение

$$\frac{d^2\lambda}{du^2} - \frac{1}{2} \frac{d\lambda}{du} \frac{d(\ln \lambda)}{du} - \frac{d\alpha}{du} \frac{d\gamma}{du} - \left(\frac{d\beta}{du} \right)^2 = 0$$

Теория

(пока не смотрел)

13.2.2 Сильная гравитационная волна Бонди-Робинсона

(вроде Лашкевич знает это, так что мб у него спрошу просто потом.)

Рассмотрим неслабую гравитационную волну в плоском пространстве-времени.

Суть сильной гравитационной волны Бонди-Роберсона

(потом пропишу)

Обзор применений сильных гравитационных волн (??)

(где они нужны - хороший вопрос очень. пока хз)

Сильные волны нужны, например, для того, что при работе со вторым порядком малости, мы начинаем ловить некоторые эффекты, которые в первом отсутствуют.

(??? какие??)

Именно их следует изучить, прежде чем сказать, что чего-то по ОТО не существует.

(пока не четко понимаю, как это конкретно???)

теория

Пусть существует волна, в которой все компоненты метрического тензора оказываются при надлежащем выборе системы отсчета функциями лишь одной переменной, которую назовем x^0 . Поищем явный вид метрики такой волны.

(а дальше у меня слишком слабая теория про метрику, чтобы я повторил эти вещи. дополню её, потом вернусь сюда.)

Это условие допускает еще преобразования координат вида

$$\begin{aligned}x^\alpha &\rightarrow x^\alpha + \varphi^\alpha(x^0), \\x^0 &\rightarrow \varphi^0(x^0)\end{aligned}$$

где $\varphi^0, \varphi^\alpha$ - произвольные функции.

Характер решения существенно зависит от того, можно ли тремя преобразованиями с φ^i обратить все $g_{0\alpha}$ в нуль.

Это можно сделать, если определитель $|g_{\alpha\beta}| \neq 0$.

(?????) Действительно, при таком преобразовании $g_{0\alpha} \rightarrow g_{0\alpha} + g_{\alpha\beta}\dot{\varphi}^\beta$ (где точка означает дифференцирование по x^0); при $|g_{\alpha\beta}| \neq 0$ система уравнений

$$g_{0\alpha} + g_{\alpha\beta}\dot{\varphi}^\beta = 0$$

определяет $\varphi^\beta(x^0)$, осуществляющие требуемое преобразование.

Такой случай будет рассмотрен в разделе про плоскую анизотропную модель; здесь же нас будет интересовать решение, в котором

$$|g_{\alpha\beta}| = 0$$

В таком случае системы отсчета, в которой бы все $g_{0\alpha} = 0$, не существует.

Вместо этого, однако, четырьмя преобразованиями φ^i можно добиться того, чтобы было

$$g_{01} = 1, \quad g_{00} = g_{02} = g_{03} = 0 \quad (13.1)$$

Переменная x^0 имеет при этом «световой» характер: при $dx^\alpha = 0$ $dx^0 \neq 0$ интервал $ds = 0$; выбранную таким образом переменную x^0 будем обозначать ниже как $x^0 = \eta$.

Элемент интервала при условиях (13.1) можно представить в виде

$$ds^2 = 2dx^1d\eta + g_{ab}(dx^a + g^a dx^1)(dx^b + g^b dx^1)$$

Здесь и ниже в этом параграфе индексы $a, b, c, \dots$ пробегает значения 2, 3; $g_{ab}(\eta)$ можно рассматривать как двухмерный тензор, а две величины $g^a(\eta)$ - как компоненты двухмерного вектора.

Вычисление величин R_{ab} приводит к следующим уравнениям поля:

$$R_{ab} = -\frac{1}{2}g_{ac}\dot{g}^c g_{bd}\dot{g}^d = 0$$

Отсюда следует, что $g_{ac}\dot{g}^c = 0$ или $\dot{g}_c = 0$, т.е. $g^c = \text{const.}$

Преобразованием $x^a + g^a x^1 \rightarrow x^a$ можно поэтому привести рассматриваемую метрику к виду

$$ds^2 = 2dx^1d\eta + g_{ab}(\eta)dx^a dx^b$$

Определитель $-g$ этого метрического тензора совпадает с определителем $|g_{ab}|$, а из всех символов Кристоффеля отличны от нуля лишь следующие:

$$\Gamma_{b0}^a = \frac{1}{2}x_b^a, \quad \Gamma_{ab}^1 = -\frac{1}{2}x_{ab}$$

где мы ввели двухмерный тензор $x_{ab} = \dot{g}_{ab}$, $x_a^b = g^{bc} x_{ac}$.

Из всех компонент тензора Риччи не обращается тождественно в нуль лишь R_{00} , так что имеем уравнение

$$R_{00} = -\frac{1}{2}\dot{x}_a^a - \frac{1}{4}x_a^b x_b^a = 0$$

Таким образом, три функции $g_{22}(\eta)$, $g_{23}(\eta)$, $g_{33}(\eta)$ должны удовлетворять всего одному уравнению.

Поэтому две из них могут быть заданы произвольно.

Удобно представить уравнение $R_{00} = 0$ в другом виде, написав величины g_{ab} в виде

$$g_{ab} = -\chi^2 \gamma_{ab}, \quad |\gamma_{ab}| = 1$$

Тогда определитель $-g = |g_{ab}| = \chi^4$ и подстановка в $R_{00} = 0$ дает после простого преобразования

$$\ddot{\chi} + \frac{1}{8}(\dot{\gamma}_{ac}\gamma^{bc})(\dot{\gamma}_{bd}\gamma^{ad})\chi = 0$$

где γ^{ab} - двухмерный тензор, обратный тензору γ_{ab} .

Если задать произвольные функции $\gamma_{ab}(\eta)$ (связанные друг с другом соотношением $|\gamma_{ab}| = 1$), этим уравнением определится функция $\chi(\eta)$.

Мы приходим, таким образом, к решению, содержащему две произвольные функции. (???)

Сильная волна Бонди-Робинсона это действительно обобщение слабой (??)

(пока не вижу это)

Легко видеть, что решение уравнения выше представляет собой обобщение рассмотренной слабой плоской гравитационной волны (распространяющейся в одном направлении).

(??? как вообще оно выглядит то???)

Последняя получается, если произвести преобразование

$$\eta = \frac{t+x}{\sqrt{2}}, \quad x^1 = \frac{t-x}{\sqrt{2}}$$

и положить $\gamma_{ab} = \Delta_{ab} + h_{ab}(\eta)$ (где h_{ab} - малые величины, подчиненные условию $h_{22} + h_{33} = 0$) и $\chi = 1$; постоянное значение χ удовлетворяет уравнению $\ddot{\chi} + \frac{1}{8}(\dot{\gamma}_{ac}\gamma^{bc})(\dot{\gamma}_{bd}\gamma^{ad})\chi = 0$, если пренебречь в нем малыми членами второго порядка.

Об особенности в метрике сильной гравитационной волны

Пусть через какую-либо точку x пространства проходит слабая гравитационная волна конечной протяженности («волновой пакет»).

До начала прохождения имеет $h_{ab} = 0$, $\chi = 1$; после конца прохождения снова $h_{ab} = 0$, $\partial^2 \chi / \partial t^2 = 0$, но учет членов второго порядка в уравнении $\ddot{\chi} + \frac{1}{8}(\dot{\gamma}_{ac}\gamma^{bc})(\dot{\gamma}_{bd}\gamma^{ad})\chi = 0$ приведет к появлению отличного от нуля отрицательного значения $\partial \chi / \partial t$:

$$\frac{\partial \chi}{\partial t} \approx -\frac{1}{8} \int \left(\frac{\partial h_{ab}}{\partial t} \right)^2 dt < 0,$$

где интеграл берется по времени прохождения волны.

(???)

Поэтому после прохождения волны будет $\chi = 1 - \text{const} \cdot t$ и по истечении копеечного промежутка времени χ изменит знак.

Но обращение χ в нуль есть обращение в нуль метрического определителя g , т.е. особенность в метрике.

Эта особенность, однако, не имеет физического характера; она связана лишь с недостатками системы отсчета, «испорченной» проходящей гравитационной волной, и может быть устранена надлежащим ее преобразованием; после прохождения волны пространство-время оказывается в действительности снова плоским.

Действительно. Если отсчитывать переменную η от ее значения, соответствующего особой точке, то $\chi = \eta$, так что

$$ds^2 = 2d\eta dx^1 - \eta^2 \left[(dx^2)^2 + (dx^3)^2 \right]$$

После преобразования

$$\eta x^2 = y, \quad \eta x^3 = z, \quad x^1 = \xi - \frac{y^2 + z^2}{2\eta}$$

получаем:

$$ds^2 = 2d\eta d\xi - dy^2 - dz^2$$

и подстановка $\eta = (t+x)/\sqrt{2}$, $\xi = (t-x)/\sqrt{2}$ окончательно приводит метрику к галилеевой форме.

(?!!! проделаю это, пока не так понимаю.)

Это свойство гравитационной волны - возникновение фиктивной особенности - не связано, конечно, с ее слабостью и присуще также и общему решению уравнения Эйнштейна;

как и в рассмотренном примере, вблизи особенности $\chi \sim \eta$, т.е. $-g \sim \eta^4$.

Иллюстрация сингулярности за сильной волной

(пока актуален вопрос, как же выглядит эта особенность? пока не знаю, как изобразить её...с)

О метрике членом $f(t-x, y, z)(dt-dx)^2$

(зачем эта метрика?)

Определим функцию $f(t-x, y, z)$ в метрике

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2 + f(t-x, y, z)(dt-dx)^2,$$

при котором она является точным решением уравнений Эйнштейна для поля в пустоте

Тензор Риччи вычисляется проще всего в координатах $u = (t-x)/\sqrt{2}$, $v = (t+x)/\sqrt{2}$, y, z , в которых

$$ds^2 = -dy^2 - dz^2 + 2dudv + 2f(u, y, z)du^2$$

Помимо $g_{22} = g_{33} = -1$, отличны от нуля лишь следующие компоненты метрического тензора: $g_{uu} = 2f$, $g_{uv} = 1$; при этом $g^{vv} = -2f$, $g^{uv} = 1$, а определитель $g = -1$.

(?? дальше вычисления, которые проделать нужно ручками.)

Прямое вычисление дает для отличных от нуля компонент тензора кривизны:

$$\begin{aligned} R_{yuyu} &= -\frac{\partial^2 f}{\partial y^2}, \\ R_{zuzu} &= -\frac{\partial^2 f}{\partial z^2}, \\ R_{yuzu} &= -\frac{\partial^2 f}{\partial y \partial z}. \end{aligned}$$

Единственная отличная от нуля компонента тензора Риччи: $R_{uu} = \Delta f$, где Δ - оператор Лапласа по координатам y, z .

Таким образом, уравнение Эйнштейна имеют вид:

$$\Delta f = 0,$$

т.е. функция $f(t - x, y, z)$ должна быть гармонической по переменным y, z

Если функция f не зависит от y, z или линейна по этим переменным, поле отсутствует - пространство-время плоское (тензор кривизны обращается в нуль).

(?? дальше переход к ответу, который я не понял)

Здесь можно увидеть ответ, что именно квадратичная по y, z функция

$$f(u, y, z) = yz f_1(u) + \frac{1}{2} (y^2 - z^2) f_2(u)$$

отвечает плоской волне, распространяющейся в положительном направлении оси x ;

Действительно, тензор кривизны в такой волне зависят только от $t - x$:

$$R_{yuzu} = -f_1(u), \quad R_{yuyu} = -R_{zuzu} = -f_2(u)$$

В соответствии с двумя возможными поляризациями волны метрика содержит в этом случае две произвольные функции $f_1(u)$ и $f_2(u)$.

13.2.3 Метрика Такено

Суть

имеет вид

$$ds^2 = (P + S) (dx^0)^2 - 2S dx^0 dx^1 - (P - S) (dx^1)^2 - \alpha (dx^2)^2 - 2\beta (dx^2 dx^3) - \gamma (dx^3)^2$$

где P, S -функции, α, β, γ удовлетворяют тому же уравнению.

Теория

(пока не смотрел)

13.2.4 Метрика Розена

Суть

$$ds^2 = e^{2\mu} \left[(dx^0)^2 - (dx^1)^2 \right] - (u^2) \left[e^{2\nu} (dx^2)^2 - e^{-2\nu} (dx^3)^2 \right]$$

где μ, ν удовлетворяют уравнению

$$2 \frac{d\mu}{du} = u \left(\frac{d\nu}{du} \right)^2.$$

Теория

(пока не смотрел)

13.2.5 Метрика Переса

Суть

$$ds^2 = (dx^0)^2 - (dx^1)^2 - 2\varphi (dx^0 + dx^1)^2 - (dx^2)^2 - (dx^3)^2$$

при этом $\frac{\partial^2 \varphi}{\partial x^2 \partial x^2} + \frac{\partial^2 \varphi}{\partial x^3 \partial x^3} = 0$.

Теория

(пока не смотрел)

13.2.6 Цилиндрические волны Эйнштейна - Розена

Суть

В цилиндрических координатах такие волны имеют вид

$$ds^2 = \left[(dx^0)^2 - dr^2 \right] e^{2\gamma - 2\psi} - (dz^2) e^{2\psi} - r^2 (d\varphi^2) e^{-2\psi}$$

Суть

13.2.7 Другие гравитационные волны

(просто тут нарешаю самые разные задачи, которые придут в голову)

Гравитационные волны от махания кулаком

Гравитационные волны от махания кулаком, двигаясь с большой скоростью

Гравитационные волны от ядерного взрыва

13.2.8 Гравитационные волны в начале вселенной

и также в разных теориях посмотрим, как их там можно считать?

реликтовые гравитационные волны

????

хз, не думал, но инфа может нагуглится

13.3 Другие и редкие волны

(всем этим только по приколу занимаюсь, всерьез - никогда, ничего иначе не сделаешь!)

13.3.1 Ударные гравитационные волны

shock waves

(вроде этим тоже занимаются, статьи есть на архиве, мб посмотрю.)

Обзор

(хз)

Теория

(хз)

13.3.2 Турбулентность в гравитационных волнах

(статьи на архиве есть про это, мб потом посмотрю)

Обзор

(хз)

Теория

(хз)

13.3.3 Исследование коллапса гравитирующих систем

По мнению американского астрофизика Брайана Грина излучение гравитационных волн при движении различных космических объектов, действует подобно трению и в долгосрочной перспективе приводит к коллапсу космических систем, таких как планетарные системы звезд и галактики.

(тут всё, что я про это знаю, собираю. пока не знаю, как исследовать, но мб задачу про это найду решенную.)

13.3.4 Гравитационные волны в космологии (?)

(хз, Ландау писал, что чем-то на сильную волну похоже)

13.3.5 Многомерные гравитационные волны

Обсудим гравитационные волны в размерности пространства времени $d > 4$ (даже заниматься этим не хочу, звучит как тема кандидатской.)

14 Взаимодействия гравитационных волн

14.0.1 Взаимодействие гравитационных волн со сплошной средой

14.0.2 Взаимодействие гравитационных волн со сплошной средой по Лашкевичу

Теория

Вопросы

14.0.3 Взаимодействие гравитационных волн с электромагнитным полем

(у меня же целая статья про это выгружена!!!! где она????)

GW+EM fried по Лашкевичу

Может ли гравитационная волна порождаться электромагнитной? Рассмотрим тензор энергии-импульса электромагнитной волны, распространяющейся вдоль оси x^1 и поляризованной в плоскости (x^1, x^2) :

$$T^{\bullet\bullet} = \begin{pmatrix} S & S & 0 & 0 \\ S & S & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad S = E_0^2 \cos^2 \omega x^-.$$

Нас будут интересовать поперечные пространственные компоненты, поскольку продольные и временные компоненты не порождают гравитационную волну. Легко видеть, что они обращаются в нуль.

Включим внешнее постоянное однородное поперечное магнитное поле $\mathbf{H}_{\text{внеш}} = H_{\perp} \partial_2 + H_{\parallel} \partial_3$. Тогда мы найдем, что

$$T_{23} = -H_{\perp} E_0 \cos \omega x^- + \text{const},$$

$$\frac{1}{2} (T_{22} - T_{33}) = H_{\parallel} E_0 \cos \omega x^- + \text{const}.$$

Таким образом, электромагнитная волна в постоянном магнитном поле является когерентным источником гравитационной волны.

О деталях теории Лашкевича

14.1 Взаимодействие гравитационной волны с электромагнитными полем

14.1.1 Гравитационные волны в электромагнитных полях

(в дополнения вставляю минимальную теорию из той статьи, потому что именно она тут везде используется!!!)

Теория

Consider an electric field in the background of a GW travelling in the z direction. The general expression for Gauss' law (2.7), in vacuum, is now given by

$$[1 - f_+(z, t)]^{-1} \partial_x E_x + [1 + f_+(z, t)]^{-1} \partial_y E_y + \partial_z E_z - f_{\times}^{-1}(z, t) (\partial_y E_x + \partial_x E_y) + \left[\frac{1}{\sqrt{-g}} \partial_z (\sqrt{-g}) \right] E_z = 0$$

which clearly shows that physical (possibly observable) effects are induced by the propagation of GWs.

As for the Maxwell-Ampère law, Eq. (2.8) provides the following relations in vacuum:

$$\frac{1}{c^2} [f_{\times}^{-1} \partial_t E_y - (1 - f_+)^{-1} \partial_t E_x] + E_x \xi^{xx} + E_y \xi^{yx} - (1 - f_+)^{-1} [(1 + f_+)^{-1} \partial_y B^z - \partial_z B^y] - B^y \sigma^{zx} + B^x \sigma^{zy} - f_{\times}^{-1} (f_{\times}^{-1} \partial_y B^z + \partial_z B^x) = 0$$

$$\begin{aligned} \frac{1}{c^2} [f_{\times}^{-1} \partial_t E_x - (1 + f_+)^{-1} \partial_t E_y] + E_y \xi^{yy} + E_x \xi^{xy} \\ - (1 + f_+)^{-1} [(1 - f_+)^{-1} \partial_x B^z - \partial_z B^x] + B^x \sigma^{zyy} \\ - B^y \sigma^{zxy} + f_{\times}^{-1} (f_{\times}^{-1} \partial_x B^z + \partial_z B^y) = 0, \end{aligned}$$

$$\begin{aligned} \frac{1}{c^2} \partial_t E_z + E_z \xi^{zz} - f_{\times}^{-1} (\partial_y B^y - \partial_x B^x) \\ + [(1 - f_+)^{-1} \partial_x B^y - (1 + f_+)^{-1} \partial_y B^x] = 0, \end{aligned}$$

with the non-vanishing geometric coefficients given by

$$\begin{aligned} \xi^{xx} &= \frac{1}{c^2} \frac{f_{\times} (f_+ - 1) \partial_t f_{\times} - (f_{\times}^2 + f_+ - 1) \partial_t f_+}{(f_+ - 1)^2 (f_{\times}^2 + f_+^2 - 1)} \\ \xi^{yx} &= \xi^{xy} = \frac{1 - (f_+^2 - 1) \partial_t f_{\times} + f_{\times} f_+ \partial_t f_+}{c^2 f_{\times}^2 (f_{\times}^2 + f_+^2 - 1)}, \\ \xi^{yy} &= \frac{1 - f_{\times} (f_+ + 1) \partial_t f_{\times} + (f_{\times}^2 - f_+ - 1) \partial_t f_+}{c^2}, \\ \xi^{zz} &= -\frac{1}{c^2} \frac{f_{\times} (\partial_t f_{\times}) + f_+ (\partial_t f_+)}{f_{\times}^2 + f_+^2 - 1} \\ \sigma^{zxx} &= -\frac{f_{\times} (f_+ - 1) \partial_z f_{\times} - (f_{\times}^2 + f_+ - 1) \partial_z f_+}{(f_+ - 1)^2 (f_{\times}^2 + f_+^2 - 1)} \\ \sigma^{zyy} &= -\frac{-f_{\times} (f_+ + 1) \partial_z f_{\times} + (f_{\times}^2 - f_+ - 1) \partial_z f_+}{(f_+ + 1)^2 (f_{\times}^2 + f_+^2 - 1)} \\ \sigma^{zxy} &= \sigma^{zyx} = -\frac{-(f_+^2 - 1) \partial_z f_{\times} + f_{\times} f_+ \partial_z f_+}{f_{\times}^2 (f_{\times}^2 + f_+^2 - 1)} \end{aligned}$$

(тут вот задачка - сесть, проверить это!)

A natural consequence of these laws is the generation of electromagnetic waves induced by gravitational radiation. Initially static electric and magnetic fields become time dependent during the passage of GWs which might be detectable in this way.

In general, the system of coupled Eqs. (2.7) – (2.8) and the homogeneous equations in (2.3) have to be taken as a whole. As we will see from Eq. (3.5), in some specific situations the electric field can be solved directly from Gauss' law. This electric field can in turn act as a source for magnetism via the Maxwell-Ampère relations in Eqs. (3.6) – (3.8), where the presence of the GW induces extra terms proportional to the electric field. In this work, we will explore relatively simple situations in order to illustrate the effects of GWs in electric and magnetic fields. Let us start by considering the effects of GWs in electric fields.

14.1.2 Колебания электромагнитного поля за счет гравитационной волны (?)

(не знаю, насколько это нужно, мб даже и нужно)

We will consider electric fields in the following three scenarios.

Electric field aligned with the z axis

An electric field along the z axis can easily be achieved by charged plane plates constituting a capacitor. In the absence of GWs the electric field thus produced is approximately uniform (neglecting boundary effects) for static uniform charge distributions or time variable if there is an alternate current (as in the case of a RLC circuit with a variable voltage signal generator). With the passage of the GW, in general the electric field is perturbed by both the $(+)$ and (x) modes. To see this let us look at Gauss' law when the electric field is aligned with the direction of the GW propagation. From Eq. (3.5), we have

$$\partial_z E_z + E_z \left[\frac{1}{\sqrt{-g}} \partial_z (\sqrt{-g}) \right] = 0,$$

where $\frac{1}{\sqrt{-g}} \partial_z (\sqrt{-g})$ is given by the expression in Eq. (3.3). We can see that even if in the absence of any

GW the electric field was static and uniform, during the passage of the spacetime disturbance, the field will be time varying and non-uniform, oscillating with the same frequency of the passing GW. In fact, the general solution is

$$E_z(x, y, z, t) = \frac{E_0}{\sqrt{-g}} = \frac{E_0}{\sqrt{1 - f_+^2 - f_-^2}}$$

where in the most general case, $E_0 = E_0(x, y, t)$. To get the full description of the electric field one has to consider also both the Maxwell-Ampère relations in Eqs. (3.4)(3.8) and the Faraday law. Nevertheless it is already clear from Eq. (3.10) that GWs induce propagating electric oscillations.

We will consider the most simple case in which E_0 is a constant (without any dependence on x, y or t). Indeed, one can easily verify that the fields $\mathbf{E} = (0, 0, E_z)$, $\mathbf{B} = (0, 0, 0)$ constitute a (trivial) solution of the full Maxwell equations, namely Eqs. (3.5)-(3.8) together with the homogeneous equations in (2.3). Notice that for zero magnetic field the z Maxwell-Ampère equation (3.8) is

$$-\frac{1}{c^2} \partial_t E_z + E_z \xi^{zz} = 0$$

which is verified by the solution in (3.10) for a constant E_0 . This can easily be seen when one considers that

$$\xi^{zz} = -\frac{1}{c^2} \frac{1}{\sqrt{-g}} \partial_t (\sqrt{-g})$$

in accordance with the expressions previously shown for ξ^z and Eq. (3.4). In this case, the coupling between the electric field and the GW in the expression for the generalized Maxwell displacement current density, compensates the traditional term which depends on the time derivative of the electric field. In fact, by multiplying by c^2 , then Eq. (3.11) can be interpreted as the conservation of the total electric flux density. This situation is thus compatible with the experimental scenario where there are no currents producing any magnetic field and the electric field, although changing in time, due to the coupling with gravity does not give rise to any magnetic field, since the total electric flux is conserved. Naturally, this is not the general case. For example in the presence of currents along the z axis, $B^x, B^y \neq 0$ and due to the gravitational factors in the equations (3.6)-(3.8) the magnetic field is dynamical (time dependent). Therefore, this field necessarily affects the electric field via the Faraday law,

$$\partial_y E_z = \frac{\partial_y E_0}{\sqrt{-g}} = -\partial_t B^x, \quad \partial_x E_z = \frac{\partial_x E_0}{\sqrt{-g}} = \partial_t B^y,$$

which implies that in general $E_0 = E_0(x, y, t)$. Since E_0 is time dependent, in such a case the electric field contributes to the magnetic field via the (non-null) generalized Maxwell displacement current, in accordance with

Eq. (3.8), where now

$$-\frac{1}{c^2} \partial_t E_z + E_z \xi^{zz} \neq 0$$

As a practical application consider the following harmonic GW perturbation

$$\begin{aligned} f_+(z, t) &= a \cos(kz - wt), \\ f_\times(z, t) &= b \cos(kz - wt + \alpha). \end{aligned}$$

In this case, we get the following electric oscillations

$$\begin{aligned} E_z(z, t) &= \tilde{E}_0 [1 - a^2 \cos^2(kz - wt) \\ &\quad - b^2 \cos^2(kz - wt + \alpha)]^{-1/2}, \end{aligned}$$

for $a^2 + b^2 \leq 1$, which is obeyed by the extremely low amplitude GWs reaching the Solar System. Here $\tilde{E}_0$ is an arbitrary fixed constant and α is the phase difference. These electric oscillations will show distinct features sensitive to the (+) or ($\times$) GW modes. It provides a window for detecting and analysing GW signals directly converted into electromagnetic information.

Notice that the electric waves produced are longitudinal, since these are propagating along the same direction of the GW, even though the electric field is aligned with this direction. To grasp the physical interpretation behind this non-intuitive result, recall that the electric energy density depends quadratically on the field and therefore it is the energy density fluctuation induced by the GW which propagates along the direction of $\mathbf{k} = k^z \mathbf{e}_z$.

In order to have an approximate idea on the energy density u^{em} of the resulting electromagnetic wave we can use the usual expression (derived from Maxwell electrodynamics in Minkowski spacetime). We get

$$\begin{aligned} u^{em} &\sim \varepsilon_0 E_z^2(z, t) = \varepsilon_0 \tilde{E}_0^2 [1 - a^2 \cos^2(kz - wt) \\ &\quad - b^2 \cos^2(kz - wt + \alpha)]^{-1}, \end{aligned} \quad (3)$$

and the energy per unit area and unit time through any surface (with a normal making an angle ϑ with the z axis) is approximately expressed by

$$\begin{aligned} \|\vec{S}\| \cos \vartheta &= \varepsilon_0 c \tilde{E}_0^2 [1 - a^2 \cos^2(kz - wt) \\ &\quad - b^2 \cos^2(kz - wt + \alpha)]^{-1} \cos \vartheta \end{aligned}$$

where $\vec{S}$ is the Poynting vector, and $S \equiv u^{em} c$. If $\tilde{E}_0$ is the electric field in the absence of GWs, then the relevant dimensionless quantity to be measured is given by the following expression

$$\begin{aligned} \left| \frac{E_z(z, t) - \tilde{E}_0}{\tilde{E}_0} \right| &= | [1 - a^2 \cos^2(kz - wt) \\ &\quad - b^2 \cos^2(kz - wt + \alpha)]^{-1/2} - 1 | \end{aligned}$$

and in terms of the energy density, we get

$$\begin{aligned} \left| \frac{u^{em}(z, t) - u_0^{em}}{u_0^{em}} \right| &= | [1 - a^2 \cos^2(kz - wt) \\ &\quad - b^2 \cos^2(kz - wt + \alpha)]^{-1} - 1 | \end{aligned}$$

with $u_0^{em} = \varepsilon_0 \tilde{E}_0^2$. Substituting in these two expressions the typical amplitudes for GWs due to binaries ($10^{-25} - 10^{-21}$), the induced electric field and corresponding energy density

oscillations signal will be extremely small. Concerning GWs reaching the Solar System, the detectors which might have a response proportional to the electric field magnitude or rather to its energy (proportional to the square of the electric field magnitude), must be extremely sensitive. We emphasize the fact that, in principle, this electromagnetic wave can be confined in a cavity using very efficient reflectors for the frequency w . Then, under appropriate (resonant) geometric conditions, the signal can be amplified. This might have very important practical applications for future GW detectors.

Electric field in the xy plane

Suppose we have an electric field in the x direction. The electric field could initially be uniform and confined within a plane capacitor. In these conditions, the Gauss law in vacuum becomes

$$[1 - f_+(z, t)]^{-1} \frac{\partial E_x}{\partial x} - (f_\times)^{-1}(z, t) \frac{\partial E_x}{\partial y} = 0.$$

A similar expression is obtained if the electric field is aligned with the y axis. Assuming a separation of variables $E_x(x, y, z, t) = F_1(x, z, t)F_2(y, z, t)$, where z and t are seen as external parameters, substituting in the above equation and dividing it by E_x , we obtain

$$(1 - f_+)^{-1} \frac{\partial_x F_1}{F_1} = f_\times^{-1} \frac{\partial_y F_2}{F_2}$$

therefore, one arrives at the following expressions

$$F_1(x; z) = C_1(z, t)e^{-(1-f_+)x}, \quad F_2(x; z) = C_2(z, t)e^{-f_\times y}.$$

Since we can always add a constant to the solution obtained from $F_1(x, z, t)F_2(y, z, t)$, we can write

$$E_x(x, y, z, t) = E_{0x} \left[1 + \tilde{C}(z, t)e^{-[(1-f_+)x + f_\times y]} \right],$$

where in general $\tilde{C}(z, t) = \tilde{C}[f_+(z, t), f_\times(z, t)]$ can be obtained by taking into account the other Maxwell equations. The full solution should be compatible with the limit without gravity in which we recover the uniform field $E_x = E_{0x}$. Therefore

$$f_+ = f_\times = 0 \Rightarrow \tilde{C}(z, t) = 0.$$

A natural Ansatz is

$$\tilde{C}(z, t) = \eta f_+^{\alpha_1} + \beta f_\times^{\alpha_2} + \mu f_+^{\alpha_3} f_\times^{\alpha_4},$$

where η, β, μ and $\alpha_i (i = 1, 2, 3, 4)$ are constants. But as previously said the form of this function can be studied

by considering compatibility with the remaining Maxwell equations.

For the harmonic GW introduced before, the second term in the solution above, Eq. (3.25) is given by the following expression

$$E_{0x} \tilde{C}(z, t) \exp\{ - [(1 - a \cos(kz - wt))x + b \cos(kz - wt + \alpha)y] \}.$$

The solution obtained is also sensitive to the existence or not of two modes in the GW, to their amplitudes and phase difference. Although this solution obeys the Gauss law, it implies a non-zero dynamical magnetic field, according to Faraday's law. As mentioned, to get a full treatment one should then check the consistency with the other Maxwell equations, in order to derive restrictions on the mathematical form of $\tilde{C}(z, t)$.

Let us consider now the case where an electric field $\mathbf{E}_1 = (E_x, 0)$ is generated by a plane capacitor oriented along the x axis and a second electric field $\mathbf{E}_2 = (0, E_y)$ is generated by another similar capacitor oriented along the y axis. In this condition, the resulting electric field in the vacuum between the charged plates, $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 = (E_x, E_y)$, obeys the equation

$$(1 - f_+)^{-1} \partial_x E_x - (f_\times)^{-1} \partial_y E_x + (1 + f_+)^{-1} \partial_y E_y - (f_\times)^{-1} \partial_x E_y = 0$$

A possible solution to this equation is given by

$$E_x(x, y, z, t) = E_{0x} \left[1 + \tilde{C}_1(z, t) e^{-[(1-f_+)x + f_\times y]} \right],$$

$$E_y(x, y, z, t) = E_{0y} \left[1 + \tilde{C}_2(z, t) e^{-[f_\times x + (1+f_+)y]} \right],$$

where for $f_+ = f_\times = 0$ we get $\tilde{C}_1(z, t) = \tilde{C}_2(z, t) = 0$. The resulting electric oscillations propagate along the z axis as an electromagnetic wave with non-linear polarization. This wave results from a linear gravitational perturbation of Minkowski spacetime and therefore (in this first order approximation) it can be thought of as an electromagnetic disturbance propagating in Minkowski background with a dynamical polarization. In fact, the angle between the resulting electric field and the x axis is then $\Theta \simeq \arctan(E_y/E_x)$ i.e., for $E_{0x} = E_{0y}$

$$\Theta(x, y, z, t) \simeq \arctan \left\{ \frac{1 + \tilde{C}_2(z, t) e^{-[f_\times x + (1+f_+)y]}}{1 + \tilde{C}_1(z, t) e^{-[(1-f_+)x + f_\times y]}} \right\}$$

Even if we had $\tilde{C}_1 = \tilde{C}_2$, we still necessarily get a nonlinear, dynamical polarization. Such an oscillating polarization could in principle be another distinctive signature of the GW that is causing it. The solutions obtained already give sufficient information to conclude that it is

possible to obtain polarization fluctuations induced by GWs, where for $E_{0x} = E_{0y}$ the strength of the effect is given by $|\pi/2 - \Theta(x, y, z, t)|/(\pi/2)$. A dynamical spatial polarization pattern is therefore expected in our detector. This contrasts with the other cases where the resulting wave was linearly polarized. This effect is shown in Figs. 1 and 2.

Nevertheless, again, the Faraday law and the Maxwell-Ampère relations can provide constraints on the functions $\tilde{C}_1$ and $\tilde{C}_2$

Electric field in the background of a GW with zero ($\times$) mode

If we consider solely the (+) GW mode, the spacetime metric (3.2) becomes diagonal and the Gauss and Maxwell-Ampère equations simplify to the following expressions in vacuum

$$\partial_k \tilde{E}^k = 0, \quad \epsilon_{ijk} \partial_j \bar{B}^{ijjk} = \mu_0 j_D^i$$

respectively, where

$$\tilde{E}^j \equiv -g^{jj} g^{00} \sqrt{-g} E_j, \quad \bar{B}^{ijjk} \equiv g^{ii} g^{jj} \sqrt{-g} B^k$$

and the generalized Maxwell displacement current density is

$$J_D^i = \varepsilon_0 \partial_t \tilde{E}^i$$

in accordance with Eqs. (2.17)-(2.21). Let us search for a trivial electric field solution which is fully compatible with the complete system of Maxwell equations. If we consider the field

$$\tilde{\mathbf{E}} = \left(\tilde{E}_0^x(y, z, t), \tilde{E}_0^y(x, z, t), \tilde{E}_0^z(x, y, t) \right),$$

the Gauss law is trivially obeyed and the electric field is given by

$$\mathbf{E} = \left(\frac{1-f_+}{\sqrt{1-f_+^2}} \tilde{E}_0^x, \frac{1+f_+}{\sqrt{1-f_+^2}} \tilde{E}_0^y, \frac{\tilde{E}_0^z}{\sqrt{1-f_+^2}} \right)$$

Furthermore, if $\partial_t \tilde{E}_0^k = 0 (k = 1, 2, 3)$, the generalized Maxwell displacement current density j_D^i is zero, therefore effectively the electric field does not contribute to the Maxwell-Ampère equations. Consequently, in the absence of electric currents, such an electric field solution seems to be compatible with the condition $\mathbf{B} = 0$. Let us assume that this is the case. Regarding the remaining Maxwell equations, the Magnetic Gauss law $\partial_i B^i = 0$ is trivially obeyed but what about Faraday's law? In this case, one can show that the condition $\partial_t \mathbf{B} = -\text{curl } \mathbf{E} = 0$, leads to a field $\tilde{\mathbf{E}}$ which necessarily depends on time which contradicts the hypothesis of zero magnetic field according to the Maxwell-Ampère relations in (3.32) and the expression (3.34). In fact, one arrives at the field.

$$\tilde{\mathbf{E}} = \left(\tilde{E}_0^x(z, t), \tilde{E}_0^y(z, t), \tilde{E}_0^z \right),$$

where $\tilde{E}_0^z$ is a constant and $\tilde{E}_0^x(z, t), \tilde{E}_0^y(z, t)$ are given by

$$\begin{aligned} \tilde{E}_0^x &= \tilde{C}_0^x \exp \left[- \int \partial_z \left(\frac{1-f_+}{\sqrt{-g}} \right) \frac{\sqrt{-g}}{1-f_+} \right] \\ \tilde{E}_0^y &= \tilde{C}_0^y \exp \left[- \int \partial_z \left(\frac{1+f_+}{\sqrt{-g}} \right) \frac{\sqrt{-g}}{1+f_+} \right] \end{aligned}$$

where $\tilde{C}_0^x$ and $\tilde{C}_0^y$ are constants of integration. These functions clearly depend on time and therefore the generalized Maxwell displacement current cannot be zero leading to a non-vanishing magnetic field. When considering a generic electric field with three components as in (3.36), one cannot assume that $\partial_t \tilde{E}_0^k = 0$ neither a zero magnetic field.

Therefore in the general case one needs to consider the influence of the electric field on the magnetic field, through the generalized Maxwell displacement current. An exception to this is the special case first considered in III.B.1, where the electric field is aligned with the direction of the propagation of the GW.

Magnetic field oscillations induced by GWs

The passage of the GW can induce a non-vanishing time varying magnetic field, even for an initially static electric field. In general the full system of the Maxwell equations can be explored numerically to compute the resulting electric and magnetic oscillations. These magnetic fluctuations could be measured in principle using SQUIDS (Super Conducting Quantum Interference Devices) that are extremely sensitive to small magnetic field changes.

To get a glimpse of the gravitationally induced magnetic field fluctuations, we can consider for simplicity only the (+) GW mode and take the generalized Maxwell-Ampère law in the form of Eq. (2.17). We will be considering an electric field aligned with the z axis given by the following solution to the Gauss law

$$\mathbf{E} = \left(0, 0, \frac{\tilde{E}_0^z(x, y, t)}{\sqrt{1-f_+^2}} \right), \quad \partial_k \tilde{E}^k = 0$$

We can also consider an electric current I along the z axis such that in principle, by symmetry we expect a magnetic field in the xy plane, $\mathbf{B} = (B^x, B^y, 0)$. Then the Maxwell-Ampère equations (2.17) are

$$\nabla \times \bar{\mathbf{B}} = \mu_0 \left(\sqrt{-g} \mathbf{j} + \varepsilon_0 \partial_t \tilde{\mathbf{E}}_0 \right)$$

where $\bar{\mathbf{B}} \equiv (\bar{B}^{zzyy\mathbf{x}}, \bar{B}^{zzxx\mathbf{y}}, 0)$, while the Faraday law provides the equations

$$\partial_t B^x = -\frac{\partial_y \tilde{E}_0^z}{\sqrt{1-f_+^2}}, \quad \partial_t B^y = \frac{\partial_x \tilde{E}_0^z}{\sqrt{1-f_+^2}}$$

Then we can perform an integration over an "amperian" closed line coincident to a magnetic field line (in perfect analogy with the method taken in usual electromagnetism) to integrate the Maxwell-Ampère law, assuming axial symmetry, around the charge current distribution and electric flux (Maxwell displacement) current. We obtain the following solution to Eq. (3.41)

$$\bar{\mathbf{B}} = \frac{\mu_0 \tilde{I}_{tot}}{2\pi \sqrt{x^2 + y^2}} (\cos \phi \mathbf{e}_y - \sin \phi \mathbf{e}_x),$$

where $\tilde{I}_{tot}(x, y, z, t) = \sqrt{-g}I + I_D(x, y, t)$ and $I_D = \iint J_D^z dx dy$. I is the (constant) electric current and $j_D^z = \varepsilon_0 \partial_t \tilde{E}_0^z$ is the Maxwell displacement current density. We then get the magnetic field components

$$B^x = -\frac{1+f_+}{\sqrt{1-f_+^2}} \left[\frac{\mu_0 \tilde{I}_{tot}(x, y, z, t)}{2\pi (x^2 + y^2)} y \right],$$

and

$$B^y = \frac{1-f_+}{\sqrt{1-f_+^2}} \left[\frac{\mu_0 \tilde{I}_{tot}(x, y, z, t)}{2\pi (x^2 + y^2)} x \right]$$

respectively. Making the Ansatz

$$\partial_x \tilde{E}_0^z = -\partial_y \tilde{E}_0^z,$$

the Faraday equations then imply that $\partial_t B^x = \partial_t B^y$, from which one derives an equation for $\tilde{I}(x, y, z, t)$ with the general solution

$$\tilde{I} = C \exp \left\{ -\int \frac{\sqrt{1-f_+^2}}{(x-y)(1-f_+)} \times \left[x \partial_t \left(\frac{1-f_+}{\sqrt{1-f_+^2}} \right) + y \partial_t \left(\frac{1+f_+}{\sqrt{1-f_+^2}} \right) \right] dt \right\}$$

where C is an integration constant. In order to illustrate the general effect of the GW on the magnetic field, consider without great loss of generality that $\tilde{I} = I$ is a constant. Then the magnetic field in the background of the harmonic GW considered in (3.16) has the following fluctuations

$$B^x = -\frac{\mu_0 I y}{2\pi (x^2 + y^2)} [1 + a \cos(kz - wt)]$$

$$B^y = \frac{\mu_0 I x}{2\pi (x^2 + y^2)} [1 - a \cos(kz - wt)]$$

respectively. We can easily see that for any point (x, y) fixed on any magnetic field line, the x and y components of the magnetic field will oscillate in time out of phase, such that when one is at its maximum value, the other is at the minimum, and vice-versa. The overall result is that the magnetic field lines will oscillate with the passage of the GW, following the deformations of the spacetime geometry, perfectly mimicking the (+) mode deformations. Figure 3 illustrate this phenomenon and was obtained using the expressions in Eqs. (3.48) and (3.49). The strength of the effect as a function of time is independent from the current I and it depends on the position (x, y) as well as on the GW parameters (see Fig. 4). It can be easily shown that the strength of the fluctuations are much stronger in specific regions of the (x, y) plane.

Charge density fluctuations induced by GWs

In the previous analysis we considered the behaviour of electric and magnetic fields in vacuum regions and did not take into account the effect of the propagating GWs on charge distributions. The effect of spacetime geometry can be understood from the charge conservation equation in curved spacetime ($\nabla_\mu j^\mu = 0$). As a result of this equation even in the absence of (intrinsic) currents, a non-static spacetime will induce a time variability in the charge density according to $\partial_t \rho = -\partial_t(\log(\sqrt{-g}))\rho$, so we can write

$$\rho(t) = \rho_0 \sqrt{\frac{g_0}{g(t)}}$$

where ρ_0 is the initial charge density before the passage of the wave and g_0 is the determinant of the initially unperturbed background metric. For the simpler case of GWs travelling along the z direction, seen as disturbances of Minkowski spacetime, we have the simple result

$$\rho(t) = \rho_0 (1 - f_\times^2 - f_+^2)^{-\frac{1}{2}}$$

As an example, for the harmonic GW modes considered previously, we obtain

$$\rho(z, t) = \rho_0 [1 - b^2 \cos^2(kz - wt + \alpha) - a^2 \cos^2(kz - wt)]^{-\frac{1}{2}}$$

Consequently, one naturally predicts charge density fluctuations and, therefore, currents due to the passage of GWs. Such density oscillations propagate along the z direction following the GW penetrating a conducting material medium. This is analogous to Alfvén waves in plasmas [29], which are density waves induced by magnetic disturbances which propagate along the magnetic field lines. In this case, astrophysical sources of GWs such as Gamma Ray Bursts or generic coalescing binaries that happen to be surrounded by plasmas in accretion disks or in stellar atmospheres, might generate similar mass density waves and charge density waves induced by the GW propagation. A more realistic treatment would require the equations of Magneto-Hydrodynamics in the background of a GW (see [13]). An interesting study would be to consider the backreaction of the relativistic plasma and electromagnetic fields on the GW properties such as the frequency, amplitude and polarizations, so that the travelling wave after detection could, in principle, contain information about the physical properties of the medium through which it propagated.

The above expression can also indicate another window for GW detection. Conductors in perfect electrostatic equilibrium or superconducting materials at very low temperatures might reveal very dim electric oscillations with well-defined characteristics, induced by GWs.

GW effects on electromagnetic radiation

The vacuum equations for the 4-potential in the presence of a background GW can be derived from Eq. (2.23). In terms of the electric and magnetic components of the 4-potential, we have

$$\begin{aligned} \partial_\mu \partial^\mu \phi + \frac{f_\times (\partial_t f_\times) + f_+ (\partial_t f_+)}{f_\times^2 + f_+^2 - 1} \partial_t \phi \\ - \frac{f_\times (\partial_z f_\times) + f_+ (\partial_z f_+)}{f_\times^2 + f_+^2 - 1} \partial_z \phi = 0, \end{aligned}$$

and

$$\begin{aligned} \partial_\mu \partial^\mu A^k + \frac{f_\times (\partial_t f_\times) + f_+ (\partial_t f_+)}{f_\times^2 + f_+^2 - 1} \partial_t A^k \\ - \frac{f_\times (\partial_z f_\times) + f_+ (\partial_z f_+)}{f_\times^2 + f_+^2 - 1} \partial_z A^k = 0, \end{aligned}$$

respectively. In the absence of GWs we recover the usual wave equations.

The resulting expressions simplify significantly if one considers only one of the two possible GW modes. For example, for an electromagnetic wave travelling in the z direction and the harmonic GW in Eq. (3.15) with no (x) mode, we get the following wave equation for the electric potential

$$\partial_{tt}^2\phi - \partial_{zz}^2\phi - \frac{wa^2 \sin(wt - kz) \cos(wt - kz)}{a^2 \cos^2(wt - kz) - 1} \partial_t \phi - \frac{ka^2 \sin(wt - kz) \cos(wt - kz)}{a^2 \cos^2(wt - kz) - 1} \partial_z \phi = 0, \quad (3.55)$$

that can be studied by applying Fourier transformation methods.

In order to study in depth the physical (measurable) effects of the passage of the GW on electromagnetic wave dynamics, one needs to solve these equations and then compute the gauge invariant electric and magnetic fields. We see in the above wave equations, the presence of terms proportional to the first derivatives which are completely absent in the electromagnetic wave equations in flat spacetime (in cartesian coordinates). These terms are always induced by gravitational fields, but in this case the gravitational field is dynamical which represents a much richer electromagnetic wave signal with the signature of the GW (see also [18]). Such signals in the radio regime

might possibly be detectable through methods of Long Baseline Interferometry, in order to amplify it. Nevertheless, we can see from the expressions above that the extra terms on the electromagnetic wave equations, induced by GWs are proportional to the frequency. Such gravitational effects might become important for sufficiently high frequency GWs. Simulations are required to see the feasibility or not of such methods.

Обсуждения и замечания (?!!!!)

GW astronomy is an emerging field of science with the potential to revolutionize astrophysics and cosmology. The construction of GW observatories can also effectively boost major technological developments. Given the extremely low GW amplitudes reaching the Solar System, incredibly huge laser interferometers have been built and others are under development in order to reach the required sensitivities. In fact, the biggest of all, LISA is expected to be achieved in space possibly through an ESA-NASA collaboration. ESA's LISA-Pathfinder science mission officially started on March 2016 and during the following six months it conducted many experiments to pave the way for future space GW observatories, such as LISA. These huge observatories represent an amazing technological effort. A natural question is the following: Can we amplify the GW signal?

One fundamental prediction of the coupling between gravity and electromagnetism is the generation of electromagnetic waves due to gravitational radiation. Therefore, in principle under the appropriate resonant conditions, the electromagnetic signal thus produced can be amplified allowing us to measure GWs, not through the motion of test masses but rather by transferring the GW signal directly into electromagnetic information. This fact might represent an important change in perspective for future ground and space GW detectors.

The fact that GWs can generate electromagnetic waves is of course not evident if one restricts the analysis to the propagation of light rays (in the geometrical optics limit) in curved spacetime. On the other hand, the full Einstein-Maxwell system of equations have to take into account the curved spacetime within Maxwell's equations and also the contribution of the electromagnetic stressenergy tensor to the gravitational field. The first aspect of this coupling was considered in this work, and it is sufficient to show that GWs can be sources of electromagnetic waves. The full gravity-electromagnetic coupling also shows the reverse phenomenon.

In this work, we obtained electric and magnetic field oscillations fully induced by a GW travelling along the z axis. For simplicity we assumed harmonic GWs. We considered the Gauss law for the cases of an electric field along the z axis, along the x axis and in the (x, y) plane. In the first case, the solutions in Eqs. (3.10) and (3.17) allowed to make an estimation of the energy flux of the resulting radiation. It is important to emphasize the fact

that the electric fluctuation thus produced corresponds to a longitudinal wave. This means that a non-zero longitudinal mode in electromagnetic radiation can in general be induced by gravitational radiation. One should search for these GW signatures in the electromagnetic counterpart of GW sources. The solution we obtained shows the dependence on the amplitudes of the two GW modes, a and b , as well as on the frequency w and phase difference α . An important aspect of hypothetical electric-GW detectors is the fact that in general although the signal is very weak for any GW reaching the solar system, under appropriate resonant conditions it can be amplified. In fact, this can be used to improve the signal to noise ratio since a system analogous to optical resonators can act as a filter privileging the signal with a specific (resonant) frequency.

The changing electric field in Eq. (3.10) inside a capacitor, for instance, would also generate alternate currents in any conductor placed between the capacitor's charged plates. In particular, a diode placed in the appropriate orientation would allow a current signal in a single direction intermittently, following the rhythm of the GW fluctuations. In the (x, y) plane the electric field can be generated by two independent capacitors in perpendicular configuration. The approximate solutions obtained show electric field oscillations generated by the GW which propagate along the z axis with non-linear polarization. We can expect a spatial polarization pattern in our detector which changes with time. This contrasts with the other cases where the resulting wave was linearly polarized. This effect is shown in Figs. 1 and 2.

In all cases, the resulting electromagnetic signal has the signature of the GW that produces it, depending on a, b, w and α . In any of these examples, time varying electric fields are generated, which can contribute to the magnetic field via the Maxwell-Ampère law. In particular, they appear in the generalized Maxwell displacement current vector density, Eq. (2.18), induced by the GW. This in turn can generate a time varying magnetic field even in the absence of electric currents. Accordingly, GWs also induce magnetic field oscillations. We made an estimation of such an effect considering the case of a diagonal metric by setting the (x) GW mode to zero. We assumed a certain electric current I along the z axis and the electric field in Eq. (3.40) along the same direction. The magnetic field thus generated lies on the (x, y) plane and it is easy to see that for any point (x, y) fixed on the magnetic field lines, the x and y components of the magnetic field will oscillate in time out of phase, such that when one is at its maximum value, the other is at the minimum, and vice-versa.

The overall result is that the magnetic field lines will oscillate with the passage of the GW, following the deformations of spacetime geometry. Figure 3 illustrate this phenomenon and was obtained using the expressions in Eqs. (3.48) and (3.49). In Fig. 4 we see the strength of the effect as a function of time. The signal to be measured is independent from the current I and it depends on the position (x, y) as well as on the GW parameters. It can be easily shown that the strength of the fluctuations are much stronger in specific regions of the (x, y) plane. Such small magnetic field changes could in principle be measured with SQUIDS (Superconducting Quantum Interference Devices), which are sensitive to extremely small magnetic fields changes [30] – [33]. SQUIDS have amazing applications from biophysics (in particular to biomagnetism) and medical sciences but also to theoretical physics: studies of majorana fermions [34], dark matter [35], gravity wave resonant bar detectors [36], cosmological fluctuations [37, 38].

The calculations in this work point to electromagnetic effects induced by GWs such that

$$\frac{\delta \mathbf{E}}{\mathbf{E}} \sim h, \quad \frac{\delta \mathbf{B}}{\mathbf{B}} \sim h, \quad h \sim 10^{-21}$$

where h is the amplitude (strain) of the GW reaching the Solar System. SQUIDS have an incredible sensitivity [30] – [40] being able to measure magnetic fields of the order of $10^{-15}T$ or even $10^{-18}T$ for measurements performed over a sufficient period of time (the SQUIDS used in the GP-B experiment had this sensitivity). Using these values for the SQUIDS sensitivity, in order to be able to measure the tiny GW effects on magnetic fields we would require magnetic fields of the order of $\mathbf{B} \sim 10^6T$ or in the best case $\mathbf{B} \sim 10^3T$. Presently, the highest magnetic fields produced in the laboratory have values of $\mathbf{B} \sim 45T$ (continuous) and $\mathbf{B} \sim 100T - 10^3T$ (pulsed). therefore, although SQUIDS are extremely sensitive, there is a real limitation to perform these measurements coming from the huge magnetic fields required. Nevertheless, the science of SQUIDS and ultra-sensitive magnetometers is very active and evolving [39, 40] and it is natural to expect improvements in terms of sensitivities and noise reduction and modelling. For $\mathbf{B} \sim 10T$ laboratory magnetic fields we would require extremely higher sensitivities ($\delta\mathbf{B} \sim 10^{-20}$) which is not in the reach of present magnetometers. Besides these considerations, intrinsic and extrinsic noise should be extremely well modelled and if possible reduced by advanced cryogenics and filtering processes.

We may consider the use of electromagnetic cavity resonators to amplify the electromagnetic waves induced by the GWs. For magnetometers with $\delta\mathbf{B} \sim 10^{-18}T$ sensitivities and $10T$ reference magnetic fields, it means that the amplification of the signal would have to be about 2 orders of magnitude. Even if this cannot be achieved by present day electromagnetic resonators it might be in the near future. An important advantage of these cavities is that in practice they work has filters being able to amplify a signal centred around a specific frequency which corresponds to the fundamental frequency of the resonator. For cylindrical resonators with size L , the wavelength of the fundamental frequency is $\lambda \sim 2L$, meaning that different resonators of different sizes would be sensitive to the different parts of the GW spectrum. By effectively filtering and amplifying the signal around a certain frequency far from the noise peak, it is in principle possible to increase substantially the signal to noise ratio, which is essential for a good measurement/detection.

Let us consider the case where we se electric fields instead of magnetic fields in our electromagnetic detectors, for example the electric field inside a charged plane capacitor. By measuring the Voltage signal instead of electric field, we have the advantage of being able with the present technology to, on one hand, easily produce 10^3V or higher static fields and on the other hand, to reach sensitivities of $\delta V \sim 10^{-15}V$. This means that the signal should be amplified 2 to 3 orders of magnitude. The combination of electromagnetic cavity resonators and electronic amplifiers (for the Voltage signal) could make this a real possibility for GW detectors.

Moving now from human made laboratories on earth or in space to natural astrophysical observatories, we call the attention to the fact that the highest magnetic field values (indirectly) measured so far are those of neutron stars with values around $10^6T - 10^{11}T$. Radio and X ray astronomy is able to indirectly measure these astrophysical magnetic fields by considering the properties of Cyclotron radiation. A stochastic GW background signal due to innumerable sources in the galaxy and beyond is expected to leave a measurable imprint on the magnetic field of normal pulsars and magnetars. In fact, this method could be used in a complementary way to that of PTA (Pulsar Timing Arrays) to measure a stochastic GW signal. The huge magnetic fields in the surroundings of pulsars makes them natural laboratories to study the effects of GWs on electromagnetic fields. The use of arrays of Pulsars could be advantageous in order to distinguish the GW signal from intrinsic fluctuations of the magnetic field and to better deal with extrinsic noise. Pulsars are extremely precise clocks and if they behave as very stable dynamos, then it might be possible to generalize the methods and years of improvement in PTA by measuring the interaction of GWs with magnetic fiels. Is this another window for GW astronomy through VLBI (Very Long Baseline Interferometry)? We leave this as an open question that deserves more research from both theorists and observation experts.

We also obtained charge density oscillations induced by GWs. These can propagate as density waves in the case of charged fluids, through which a GW is propagating. This effect deserves to be taken in consideration within more complete magnetohydrodynamical computations, in order to have simulations of the effects of GWs in plasmas near the cores of highly energetic GW sources. These plasma environments might occur in different astrophysical sources such as Gamma Ray Bursts and some specific coalescing binaries.

Regarding electromagnetic waves in the presence of gravity, extra terms appear in the generalized wave equations which deserves further research to get a full analysis of the approximate solutions. Indeed, going beyond the geometrical optics limit, light deflection (null geodesics) and gravitational redshift are not the only effects arising from the coupling between light and gravity. More generally, all electromagnetic waves can experience gravitational effects on the amplitudes, frequencies and polarizations. Besides, as shown in [22], electric and magnetic wave dynamics can be coupled due to the non-stationary geometries, as is the case of GWs. Important studies have been made regarding the electromagnetic counterpart of GW sources (see for example [41] and [42]), but there is much to explore in the landscape of (multi-messenger) gravitational and electromagnetic astronomy

In general, one expects that GWs induce very rich electromagnetic wave dynamics. These effects might become more significant for very high frequency GWs as one can see from Eq. (3.55). Moreover, the terms proportional to the first derivatives of the 4 -potential have space and time varying coefficients. For the harmonic GWs considered in this work, these coefficients oscillate between positive and negative numbers, a fact that might imply a very distinctive wave modulation pattern of the resulting electromagnetic wave. This hypothesis and its implications requires further investigation as it might provide very rich GW information codified in the electromagnetic spectra of different astrophysical and even cosmological sources.

14.2 Прохождение света через гравитационную волну (!!!)

(тут именно про модели, где важно, что у нас именно свет, а не ЭМ волна)

14.2.1 Линзирование гравитационной волной

(статья про это есть, пока не читал)

Модель

О случаях применения

(хз)

14.2.2 Прохождение волн через сплошную среду

(пропишу с опытом, решать задачи разные нужно просто, вот что)

Теория Лашкевича

Слабое гравитационное поле изменяет расстояния между точками в соответствии с формулой

$$dl'^2 = dl^2 - h_{ij}dx^i dx^j$$

Согласно стандартным формулам теории сплошных сред, тензор деформации u_{ij} определяется как

$$dl'^2 = dl^2 + 2u_{ij}dx^i dx^j$$

Следовательно, слабое гравитационное поле вызывает деформацию

$$u_{ij} = -\frac{1}{2}h_{ij}$$

При этом, как мы видели, весь эффект гравитационной волны может быть сведен к такой деформации путем исключения временных компонент. Деформация, вызванная гравитационной волной полностью поперечна гравитационной волне.

Мы будем считать, что среда занимает все пространство, а скорость звука в среде много меньше скорости света в вакууме. Это приводит к тому, что релаксацией среды можно пренебречь. Если среда конечно, то рассуждения ниже верны, если частоты собственных колебаний тела много меньше частоты гравитационной волны. Частота собственных колебаний порядка $\sqrt{k/\rho l}$, где k — модуль упругости (сжатия или сдвига в зависимости от геометрии), ρ — плотность среды, l — характерный размер тела. Для изотропной среды упругие свойства среды определяются модулем сжатия K и модулем сдвига μ . Тензор напряжений согласно закону Гука равен

$$\sigma_{ij} = K u_{kk} \delta_{ij} + 2\mu \left(u_{ij} - \frac{1}{3} \delta_{ij} u_{kk} \right).$$

Учитывая, что в гравитационной волне $u_{kk} = 0$, она вообще не вызывает сжатия материала в системе отсчета, в которой среда остается покоящейся. Таким образом, она взаимодействует с веществом посредством сдвигов. Это приводит к тому, что тензор энергии-импульса вещества получает добавку

$$\Delta T_{ij} = -\Delta \sigma_{ij} = \mu h_{ij}$$

Взаимодействие с упругой средой приводит к тому, что гравитационная волна распространяется в среде со скоростью меньшей, чем скорость света. Действительно, уравнение для потенциалов принимает вид

$$\square \psi_{ij} + 16\pi G \mu \psi_{ij} = 0.$$

Отсюда для плоской волны получаем закон дисперсии

$$\omega^2 = \mathbf{k}^2 + 16\pi G \mu.$$

Групповая скорость равна

$$\mathbf{v}_{\text{гр}} = \frac{\partial \omega}{\partial \mathbf{k}} = \frac{\mathbf{k}}{\omega} < 1$$

Разумеется, для любых разумных твердых тел эффект ничтожно мал, эти формулы представляют интерес с точки зрения распространения гравитационного поля в межзвездном газе. Тело в том, что межзвездный газ можно считать бесстолкновительным, а в бесстолкновительной среде при деформации давление ведет себя анизотропно. Именно в направлении сжатия или растяжения $pL = \text{const}$. Если считать, что, тем не менее, до прохождения волны газ был равновесным с изотропным давлением p , то внутри волны

$$\Delta T_{ij} = -p u_{ij} = \frac{p}{2} h_{ij}.$$

Таким образом, межзвездный газ ведет себя как твердая среда с модулем сдвига $p/2$.

Для жидкости имеем

$$\sigma_{ij} = \zeta \dot{u}_{kk} \delta_{ij} + 2\eta \left(\dot{u}_{ij} - \frac{1}{3} \delta_{ij} \dot{u}_{kk} \right)$$

где η, ζ - коэффициенты вязкости. Вторая вязкость ζ , опять же, не играет роли. Имеем

$$\square\psi_{ij} + 16\pi G\eta\dot{\psi}_{ij} = 0$$

Закон дисперсии принимает вид

$$\omega^2 + 16\pi i G\eta\omega = k^2$$

Это значит, что частота имеет отрицательную мнимую часть, что отвечает поглощению. Более того, сколь бы ни был мал коэффициент взаимодействия, для больших длин волн поглощение становится большим.

Изменение плотности вещества при прохождении гравитационной волны

актуально, важный вопрос, пока хз.

Изменение показателя преломления при прохождении гравитационной волны

по идее нужно знать, как плотность меняется в ней.
и связь с показателем преломления.
хз.

Изменение температуры среды

оно вообще есть или нет?

Длина дуги прямой, проходящей, через гравитационную волну

(вроде уже почти понял, позже пропишу, мб даже скоро)

Уменьшение скорости Волны

хз просто, как нормально высчитать её.

О прохождении через межзвездный газ

(Лашкевич л9 на семинаре что-то говорил, теория упругости нужна кст)

14.3 Прохождение волн через волновод

(все работающие модели про волноводы и грав. волны тут соберу)

14.3.1 О разных факторах

Влияющие факторы в порядке приоритета

какой главный эффект?
изменение показателя преломления или размеров волновода?

О влиянии сейсмической активности

в лего говорят это было оценено.
важно!!!!

О зависимостях от геометрий волноводов

(в общем случае у нас вид сети, так что можно порешать задачи с разными параметрами. укажу на это, подробно - в отдельной записи буду решать задачу эту.)

14.3.2 The Response of Optical Fibres to Gravitational Waves by Thomas B. Mieling

Introduction and Summary

While free space laser interferometric detectors of gravitational waves (GW's) such as LIGO and VIRGO are mainly analysed in the low-frequency regime $\omega_g \ll \omega$ where the GW frequency is much smaller than the laser frequency (using either the geodesic deviation equation [1-3], geometrical optics [4-9] or Maxwell's equations directly [10 – 12]), GW detectors employing electromagnetic waveguides are typically discussed in the high-frequency regime [11; 13-16]. In this work, we bring together aspects of both of these kinds of analyses by studying an electromagnetic waveguide in the low-frequency regime.

One of our main motivations for such calculations is to study the effect of GW's on light in regimes which are beyond a description in terms of geometrical optics. While such descriptions suffice to describe light propagation in vacuum or in dielectrics of large extensions (generally, whenever the wavelength is small compared to all other relevant length scales), light propagating in waveguides is only consistently described by wave optics, since the diameter of optical fibres is comparable to the optical wavelength. As a consequence, polarisation-dependent effects are not necessarily as strongly suppressed as in geometrical optics, where the transport equation for the polarisation arises only at second order in the $1/\lambda$ -expansion, see e.g. the general discussion in Ref. [17, Chapter III].

Further motivation for such investigations is provided by recent analyses of the perturbation of fibre-modes by Earth's gravitational field [18] and by Earth's rotation [19]. These calculations assume time-independent metrics, and the study of the influence of GW's illustrates how optical fibres react to time-dependent perturbations of the space-time metric.

Of course, there are still many technical challenges still to be overcome before fibre-optic interferometers could be used in experimental gravitational wave detection, see e.g. the discussion in Ref. [20, Appendix B]. Apart from the limited power handling capacity of optical fibres due to intensity damage and the emergence of non-linear effects at high power levels, the main obstructions seem to be shot noise (which is much more pronounced than in free-space interferometers due to fibre attenuation, cf. Refs. [21, Sect. 5.1; 22, Sect. 2.2 and 4.2]), as well as thermal and mechanical noises inside the fibre (both of which are absent in free-space interferometers). As

these noise sources are more pronounced for long interferometer arms, we expect the detection of GW's in fibre-optic interferometers to be viable only for signals of higher frequencies than those measured in LIGO and Virgo (which are in the frequency range of roughly 30 to 250 Hz, see e.g. Refs. [23-26]). Consequently, it is to be expected that fibre-optic interferometers would be most suitable for searching for continuous signals (recent searches for continuous GW's in LIGO data can be found e.g. in Refs. [27-29]). Moreover, since such interferometers can be made much smaller than the hitherto constructed free-space interferometers, it is conceivable to construct fibre-optic interferometers of multiple arms, providing more detailed position information on the GW source than the standard two-arm Michelson interferometers. We plan to assess the experimental feasibility of such fibre-optic interferometers in the near future.

To summarise our findings: we consider a gravitational wave in transverse-traceless gauge

of the form

$$g_{\mu\nu} = \eta_{\mu\nu} + \varepsilon A_{\mu\nu} \cos(\omega_g(t - \hat{\kappa}_i x^i)),$$

2 propagating in an arbitrary direction $\hat{\kappa}$. (Here, we restrict the discussion to monochromatic plane waves. For arbitrary waveforms, one may use the Fourier transform and apply the analysis provided here to every Fourier component separately.) Within this GW background, we solve perturbatively the Maxwell equations for (single-mode) step-index optical fibres and obtain the following result. Choosing the z -axis to be the symmetry axis of the waveguide, we find the perturbed optical phase for fibres much shorter than the GW wavelength ($\omega_g z \ll 1$) to be

$$\psi = \omega t - \beta z + m\theta - \frac{1}{2}\varepsilon c_1 A_{zz} \omega z \cos(\omega_g t)$$

where ω is the angular frequency of the mode, β the propagation constant and m the azimuthal mode index. The Jones matrix, describing the perturbation of the polarisation (see e.g. Refs. [30, Chap. 3; 31, Sect. 1.4] for an introduction to the Jones formalism), is found to be

$$M = \mathbf{1} - i\varepsilon c_2 \Delta^2 \omega z \cos(\omega_g t) \begin{pmatrix} +\frac{1}{2}(A_{xx} - A_{yy}) & A_{xy} \\ A_{xy} & -\frac{1}{2}(A_{xx} - A_{yy}) \end{pmatrix}$$

where $\Delta = (n_1 - n_2)/n_1$ is the relative difference of the refractive indices in the core (n_1) and the cladding (n_2). The coefficients c_1 and c_2 arising here are computed numerically for typical single-mode fibres. We find that $c_1 \approx \bar{n}$, where $\bar{n}$ is the (frequency dependent) effective refractive index of the fibre, and $c_2 \approx 0.3$. The diagonal terms (of opposite sign) describe birefringence, while the off-diagonal terms (of equal sign) describe a deformation of linear polarisation to elliptic polarisation.

Fig. 1: Dependence of three terms A_{zz} , $\frac{1}{2}(A_{xx} - A_{yy})$ and A_{xy} (in absolute value) on the angle ϑ formed by the optical and the gravitational wave-vectors. The blue curve ("phase") describes the angular dependence of the overall phase shift, the dashed orange line ("birefringence") describes the difference in phase shift between the two polarisation states of light, and the dotted red curve ("deformation") quantifies the mixing of the polarisation states. 3

Figure 1 shows how these three effects (phase shift, birefringence and polarisation deformation) depend on the angle ϑ between the direction of light propagation and the direction of the gravitational wave. The precise dependence on this angle and on the polarisation of the gravitational wave is discussed in detail in Section 8.

It is seen that the phase shift is maximal when the gravitational wave propagates orthogonally to the fibre axis, and vanishes when it propagates collinearly with the light ray (i.e. parallel or antiparallel). For collinear propagation, the polarisation-dependent effects are maximal: the birefringence effect is then sensitive to the $+$ polarisation of the gravitational wave, while the deformation effect is sensitive to the $\times$ polarisation.

Conclusion

We have solved perturbatively the Maxwell equations for a step-index waveguide in the presence of a weak gravitational wave (gw) of low frequency. The correction of the central mode was found to describe a perturbation of the phase as expressed by (6.11). The parameter c_1 arising

here was found to be well approximated by the effective refractive index $\bar{n}$: the relative error is smaller than the relative index difference $\Delta = (n_1 - n_2)/n_1$.

Additionally, the "large sideband" (where the azimuthal mode index $m = \pm 1$ is flipped) was identified as a perturbation of the electromagnetic wave polarisation. This perturbation for linearly polarised fields is given explicitly in (7.14), and for circular polarisation in (7.15). Compared to the phase shift, this polarisation-dependent effect (describing birefringence and

coupling of polarisation states) is suppressed by a factor of Δ^2 . However, contrary to the phase shift, this effect is also present when the gravitational wave propagates along the symmetry axis of the waveguide, as it is sensitive to the components $\frac{1}{2}(A_{xx} - A_{yy})$ and A_{xy} of the gravitational wave amplitude. Similar results for birefringence induced by gravitational waves were obtained for plane waves in infinitely extended media [39; 40], and also in vacuum for exact plane gravitational wave backgrounds [41].

Further sidebands, where the azimuthal mode index is shifted by ± 1 or ± 2 , were also computed (equations (5.115) and (5.116)) but found to be negligibly small compared to the two main effects just mentioned: they are suppressed by the frequency ratio $\Omega = \omega_g/\omega$.

Comparing with light propagation in vacuum, we find that the waveguide dispersion enters in intermediate steps of the calculation, e.g. in (5.105) (which is structurally similar to equation (2.20) in Ref. [11]) via the term $\beta' = \partial\beta/\partial\omega$. However, having implemented physically plausible boundary conditions modelling incoming radiation, we find that the dispersion term β' cancelled in the final results for the phase correction (6.11) and for the perturbation of the polarisation (7.14).

Considering the calculations presented here in the context of previous calculations which determined the influence of Earth's gravitational field and its rotation on fibre-modes, we conclude that the phase shifts in all three problems are well described using semi-heuristic arguments based on the eikonal equation in a "fictitious optical metric" where the discontinuous function n in (3.4) is replaced by the constant value of the effective refractive index $\bar{n}$ (computed from the unperturbed problem). In all cases, the difference between such geometrical-optics estimates and the result obtained by solving the true field equations was essentially determined by the relative index difference $\Delta \ll 1$. In contrast, the perturbation of the polarisation found here goes beyond such simplistic models, and thus required a thorough analysis of Maxwell's equations.

Apart from neglecting terms of order ω_g/ω (low-frequency regime) and $\rho\omega_g$ (which is consistent since $p\omega \sim 1$), we have neglected the elastic response of the fibre to the gravitational wave, which influences light propagation therein due to the induced displacement of the core-cladding interface (where continuity conditions had to be imposed) and due to stress-induced changes of the optical properties of the materials. Effects of the second kind were previously studied for transparent solid bars (elasto-optical antennas), see e.g. Refs. [42-45], but such calculations have not been carried over to optical fibres yet.

Statement of the Problem

Consider the gravitational wave (GW) metric

$$g_{\mu\nu} = \eta_{\mu\nu} + \varepsilon A_{\mu\nu} \cos(\kappa \cdot x)$$

where $\kappa \cdot x = \kappa_\mu x^\mu$, $\kappa = \omega_g (dt - \hat{\kappa}_i dx^i)$ with $\hat{\kappa}_i = \hat{\kappa}^i$ being a "spatial" unit vector (normalised with respect to the unperturbed metric) determining the direction of wave propagation, ε is the amplitude of the gravitational wave (we assume $\varepsilon \ll 1$), and A is a symmetric matrix of constant entries which is transverse ($A_{\mu 0} = A_{\mu i} \hat{\kappa}^i = 0$) and traceless ($\eta^{\mu\nu} A_{\mu\nu} = 0$), so we are using TT coordinates.

We wish to describe the propagation of light in a cylindrical waveguide at rest in this coordinate system. More precisely, we consider a cylindrical step-index fibre consisting of a linear dielectric with constant refractive index n_1 in the core and n_2 in the cladding (with $n_1 > n_2$). Such waveguides are typically non-magnetic (i.e. of permeability $\mu = 1$) and hence the permittivity is $\epsilon = \sqrt{n}$.

We will neglect boundary effects from the ends of the waveguide and the outer boundary of the cladding (the waveguide is thus modelled as infinitely long and the cladding as infinitely

thick), and we shall neglect deformations of the fibre, so that in cylindrical coordinates r, θ, z (related to the above TT coordinates as in flat space) the core-cladding interface is located at $r = \rho$. A more accurate model would have to take into account the perturbation of the spatial metric when formulating the boundary conditions, but at such a level of accuracy we estimate that one would also have to consider elastic deformations of the waveguide due to the gravitational wave, which is beyond the scope of this article.

Being concerned with weak gravitational waves of low frequency, we work in a perturbative setting with two expansion parameters

$$\varepsilon \ll 1 \quad \Omega := \omega_g/\omega \ll 1$$

Decomposition of the Maxwell Equations

In this section, we write the Maxwell equations in the GW background (2.1) in a $3+1$ form, following Ref. [18] with $c = 1$. To this end, set

$$g_{ij} = g_{ij} = \delta_{ij} + \varepsilon A_{ij} \cos(\kappa \cdot x),$$

which is the time-dependent "spatial" metric induced on the hypersurfaces of constant time t , and denote by ∇ the associated (spatial) Levi-Civita derivative.

Since $g_{00} = -1 + O(\varepsilon^2)$ and $\partial_0(\det g) = O(\varepsilon^2)$, the space-time split of the Faraday two-form F and the displacement bivector $\bar{F}$ given in Ref. [18] yields

$$D^i = \bar{F}^{0i}, \quad H_i = \frac{1}{2} \epsilon_{ijk} \bar{F}^{jk}, \quad E_i = F_{i0}, \quad B^i = \frac{1}{2} \epsilon^{ijk} F_{jk},$$

where ϵ is the spatial volume form (not to be confused with the gravitational wave amplitude ε). Here and henceforward, we neglect terms of second and higher order in ε and, for brevity, we will not write the error term $O(\varepsilon^2)$ explicitly. The field equations then take the standard form

$$\dot{B}^i + \epsilon^{ijk} \nabla_j E_k = 0, \quad \dot{D}^i - \epsilon^{ijk} \nabla_j H_k = 0, \quad \nabla_i B^i = 0, \quad \nabla_i D^i = 0,$$

where an overset dot indicates a time derivative. Following Ref. [32], we define the optical metric

$$\gamma^{\mu\nu} = g^{\mu\nu} + (1 - n^2) u^\mu u^\nu$$

where u is the four-velocity of the dielectric and n its refractive index. This allows to write the constitutive equation in the form

$$\mu \bar{F}^{\alpha\beta} = \gamma^{\alpha\rho} \gamma^{\beta\sigma} F_{\rho\sigma}.$$

In the considered case, we have $u^\mu = \delta_0^\mu$ and

$$g^{\mu\nu} = \eta^{\mu\nu} - \varepsilon A^{\mu\nu} \cos(\kappa \cdot x),$$

where the indices of A were raised with the unperturbed metric η . The constitutive equations then reduce to

$$D^i = \epsilon g^{ij} E_j, \quad B^i = \mu g^{ij} H_j.$$

As the spatial metric is perturbed by the gravitational wave, this is the only part where the field equations differ from those of flat space.

To further simplify the equations, we define the Riemann-Silberstein vector

$$Z^i := \mu D^i + j n B^i,$$

where j is an imaginary unit ($j^2 = -1$), such that the Maxwell equations take the form

$$\begin{aligned} n\dot{Z}^i + j\epsilon^{ijk}\nabla_j(g_{kl}Z^l) &= 0 \\ \nabla_i Z^i &= 0 \end{aligned}$$

wherever n is locally constant. In the following calculations, we will use two independent (commuting) imaginary units i and j , where i is reserved for the usual complex description of waves (i.e. the physical fields are the i -real parts) and j is used for this compact formulation of the field equations.

As a side remark: the algebraic structure here is that of the complex vector space $\mathbf{C}^2$ with 1 identified with $(1, 0)$ and j identified with $(0, 1)$, and the product $(a, b) \times (c, d) := (ac - bd, ad + bc)$, where a, b, c, d are usual complex numbers. In particular, the product $ij = ji$ is represented by $(i, 0) \times (0, 1) = (0, i)$, which is not equal to -1 , here identified with $(-1, 0)$.

Using only a single complex unit, it would not be possible to use this complex description of waves while simultaneously encoding the entire electromagnetic field in a single vector field. Indeed, if D and B are complex, then a field of the form $Z' = \mu D + inB$ would only provide incomplete information about D and B , as it mixes the real part of D with the imaginary part of B , and vice versa. To recover those two fields entirely, one would have to use a second complex field $Z'' = \mu D - inB$. In contrast, the field Z defined in (3.8) allows obtaining the i -complex fields D and B simply by separating the j -real and j -imaginary parts.

The Unperturbed Modes

In this section, we briefly review the unperturbed solution to the problem in a notation which is convenient for the following calculations.

(скину в разделы по теории выше скорее всего)

4.1

Adapted Basis

To describe the modes in a cylindrical waveguide, it is useful to use the complex frame

$$\mathbf{c}_{\parallel} \equiv \mathbf{c}_z, \quad \mathbf{c}_+ = \frac{1}{\sqrt{2}}(\mathbf{c}_r - i\mathbf{c}_\theta), \quad \mathbf{c}_- = \frac{1}{\sqrt{2}}(\mathbf{c}_r + i\mathbf{c}_\theta),$$

where $(\mathbf{c}_r, \mathbf{c}_\theta, \mathbf{c}_z)$ is the standard orthonormal frame adapted to cylindrical coordinates (r, θ, z) . The associated coframe is given by

$$e^{\parallel} = dz, \quad e^+ = \frac{1}{\sqrt{2}}(dr + ir d\theta), \quad e^- = \frac{1}{\sqrt{2}}(dr - ir d\theta)$$

The metric of flat space can then be written as

$$\delta = e^+ \otimes e^- + e^- \otimes e^+ + e^{\parallel} \otimes e^{\parallel}$$

so raising and lowering of indices interchanges $+$ and $-$, e.g. for a vector v one has $v^{\pm} = v_{\mp}$ but $v^{\parallel} = v_1$. The only non-vanishing components of the spatial Levi-Civita connection one-form are

$$\omega^{++} = +id\theta = \frac{1}{\sqrt{2}r}(e^+ - e^-), \quad \omega^{--} = -id\theta = \frac{1}{\sqrt{2}r}(e^- - e^+)$$

We find the cross-products to be

$$\mathbf{c}_{\parallel} \times \mathbf{c}_{\pm} = \pm i\mathbf{c}_{\pm}, \quad \text{and} \quad \mathbf{c}_+ \times \mathbf{c}_- = i\mathbf{c}_{\parallel}$$

and consequently

$$\mathbf{c}_+ \cdot (\mathbf{c}_- \times \mathbf{c}_{\parallel}) = +i$$

so that for the Levi-Civita tensor (the volume form) we obtain

$$\epsilon_{+-1} := \epsilon(\mathbf{c}_+, \mathbf{c}_-, \mathbf{c}_{\parallel}) = +i, \quad \text{and} \quad \epsilon^{+-\parallel} = -i$$

Unperturbed Bessel Modes

The computation of the electromagnetic modes in a cylindrical step-index waveguide is a standard calculation which is discussed in many textbooks, e.g. Refs. [33, Chapter 8; 34, Section 3.1; 35, Section 16.10]. Here, we merely state the result, expressed in terms of the j -complex field Z and referred to the i -complex basis introduced above. The unperturbed Bessel modes take the concise form

$$Z^{(0)\parallel} = f_0(\alpha; r)e^{i(\omega t - \beta z + m\theta)}$$

$$i\zeta Z^{(0)\pm} = \frac{1}{\sqrt{2}}\rho^2(\beta \pm i j n \omega)c_m^\pm Z^{(0)\parallel}$$

where

$$\zeta = \rho^2 (n^2 \omega^2 - \beta^2)$$

and

$$c_m^\pm = \partial_r \mp \frac{m}{r}$$

Here, ω is the angular frequency of the wave, β the propagation constant, m is the azimuthal mode index, and ρ is the radius of the fibre core. The radial function f_0 is given by

$$f_0(\alpha_1, \alpha_2; r) = \begin{cases} \alpha_1 J_m(Ur/\rho), & r < \rho \\ \alpha_2 K_m(Wr/\rho), & r > \rho \end{cases}$$

where α_1, α_2 are j -complex constants (related by appropriate continuity conditions discussed below), and

$$U = \sqrt{\rho^2 (n_1^2 \omega^2 - \beta^2)}, \quad \text{and} \quad W = \sqrt{\rho^2 (\beta^2 - n_2^2 \omega^2)}$$

The function f_0 satisfies the Bessel equation

$$B_0 f_0(\alpha; r) \equiv \left[\frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} - \frac{m^2}{r^2} + \frac{\zeta}{\rho^2} \right] f_0(\alpha; r) = 0$$

both in the core and the cladding of the fibre. In passing, let us note that the $c_m^\pm$ act as "ladder operators", i.e. for any real number λ one has the identities

$$c_m^\pm J_m(\lambda r) = \mp \lambda J_{m\pm 1}(\lambda r), \quad \text{and} \quad c_m^\pm K_m(\lambda r) = -\lambda K_{m\pm 1}(\lambda r)$$

see e.g. Ref. [36, Eq. 9.1.27 on p. 361 and Eq. 9.6.26 on p. 376], so that we may write the $\pm$ -components of the field as

$$i\zeta Z^{(0)\pm} = \frac{1}{\sqrt{2}}\rho(\beta \pm i j n \omega)f_\pm(\alpha; r)e^{i(\omega t - \beta z + m\theta)}$$

where

$$f_\pm(\alpha_1, \alpha_2; r) = \rho c_m^\pm f_0(\alpha_1, \alpha_2; r) = \begin{cases} \mp \alpha_1 U J_{m\pm 1}(Ur/\rho), & r < \rho \\ -\alpha_2 W K_{m\pm 1}(Wr/\rho), & r > \rho \end{cases}$$

The dispersion relation is then obtained by imposing continuity conditions at the core-cladding interface. As we will discuss these conditions extensively for the perturbed problem, we merely outline the basic reasoning here.

1. Maxwell's equations imply that the field components $D^r, E_\theta, E_z, B^r, H_\theta, H_z$ are continuous at the core-cladding interface (at $r = \rho$).

2. Only four of these six continuity conditions are linearly independent.

3. Since the fields have two j -complex, and thus four j -real parameters, there is a j -real equation of the form (jumps) = Π_0 (parameters), where Π_0 is a 4×4 matrix.

4. The continuity conditions are thus only solvable if $\det \Pi_0 = 0$, which constitutes the dispersion relation for β and ω . This equation is transcendental (as it contains various Bessel functions) and must thus be solved numerically.

The Perturbed Modes

As we shall see in this section, the effect of weak Gw's on light in waveguides is twofold: (i) the phase is modulated by the GW, and (ii) due to the tensorial nature of the GW amplitude, optical sidebands with an angular dependence of the form $e^{i(m+k)\theta}$, with $k \neq 0$, arise. We find the dominant effects to be (a) the correction of the phase ($k = 0$) and (b) the perturbation of the polarisation, which arises when $m + k = -m$.

Outline Let us first give an outline of the calculation. We start by deriving and solving the first order correction to the wave equation for the longitudinal field component. From this, we derive the transverse components of the electromagnetic field, finding structures similar to the ones of the unperturbed case. Schematically, the correction to the central mode can be written As

$$Z_0^{(1)} = [f^+(\alpha) + g^+(\sigma^+)] e^{i(\omega t - \beta z + m\theta + \kappa x)} + [f^-(\alpha) + g^-(\sigma^-)] e^{i(\omega t - \beta z + m\theta - \kappa x)}$$

where the coefficients α describe the unperturbed field (they act as source terms in the wave equation for $Z^{(1)}$) and the coefficients $\sigma^\pm$ parametrise homogeneous solutions of the wave equation. The continuity condition then takes the schematic form

$$\Pi_0 \alpha + \varepsilon [P\alpha + \Sigma^+ \sigma^+] e^{+i\kappa \cdot x} + \varepsilon [P\alpha + \Sigma^- \sigma^-] e^{-i\kappa \cdot x} = 0$$

where $\alpha, \sigma^\pm$ are column vectors, each comprising four j -real numbers which parametrise α and $\sigma^\pm$, and $\Pi_0, P, \Sigma^\pm$ are j -real 4×4 -matrices.

The first term is the same as in the unperturbed problem, so the dispersion relation $\det \Pi_0 = 0$ and the coefficients α remain unchanged. The coefficients $\sigma^\pm$ are then determined from

$$\Sigma^\pm \sigma^\pm + P\alpha = 0$$

For the central mode, we find that $\Sigma^\pm$ is close to the singular matrix Π_0 and its inverse has the form $(\Sigma^\pm)^{-1} = \pm \Omega^{-1} \Sigma^y + O(\Omega^0)$, so that the parameters σ are given by

$$\sigma^\pm = \mp \Omega^{-1} \Sigma' P \alpha + O(\Omega^0)$$

In Section 6, we show that this correction of order $1/\Omega$ amounts to a perturbation to the phase of the electromagnetic wave.

For the sidebands, we obtain similar equations. In almost all cases ($m + k \neq -m$), the coefficient matrices are not close to being singular, and hence no terms of order $1/\Omega$ arise there. However, in the exceptional case $m + k = -m$ we have $\det \Pi_k = \det \Pi_0 = 0$, so that we again obtain terms of order $1/\Omega$ there. We show in Section 7 that these terms describe the perturbation of the polarisation of the electromagnetic wave.

The Polarisation of the Gravitational Wave

We start by writing the polarisation tensor A of the GW, as defined in (2.1), in the complex frame introduced above. Using

$$dx = \frac{1}{\sqrt{2}} [c^+ e^{+i\theta} + c^- e^{-i\theta}], \quad dy = \frac{-i}{\sqrt{2}} [e^+ e^{+i\theta} - c^- e^{-i\theta}]$$

one finds the components of

$$A = A_{\mu\nu} dx^\mu \otimes dx^\nu = A_{ij} c^i \otimes c^j$$

in the basis $c_+, c_-, c_{\parallel}$ to be

$$\begin{aligned} A_{I\parallel} &= A_{zz} \\ A_{+-} &= A_{-+} = \frac{1}{2}(A_{xx} + A_{yy}) \\ A_{\pm\parallel} &= A_{\parallel\pm} = \frac{1}{\sqrt{2}}(A_{xz} \mp iA_{yz})e^{\pm i\theta} \\ A_{++} &= \frac{1}{2}(A_{xx} - A_{yy} - 2iA_{xy})e^{+2i\theta} \\ A_{--} &= \frac{1}{2}(A_{xx} - A_{yy} + 2iA_{xy})e^{-2i\theta} \end{aligned}$$

The condition that A is traceless now translates to

$$2A_{+-} + A_{\parallel\parallel} = 0$$

while the transversality condition takes the form

$$A_{i+}\hat{\kappa}_- + A_{i-}\hat{\kappa}_+ + A_{i\parallel}\hat{\kappa}_{\parallel} = 0$$

where the frame components of $\hat{\kappa}$ are

$$\hat{\kappa}_{\pm} = \frac{1}{\sqrt{2}}(\hat{\kappa}_x \mp i\hat{\kappa}_y)e^{\pm i\theta}, \quad \text{and} \quad \hat{\kappa}_{\parallel} = \hat{\kappa}_z.$$

To separate the amplitudes of A_{ij} from their θ -dependence, we write

$$A_{\parallel\parallel} = a_0, \quad A_{\parallel\pm} = a_{\pm 1}e^{\pm i\theta}, \quad A_{\pm\pm} = a_{\pm 2}e^{\pm 2i\theta}$$

and due to the trace condition we have $A_{+-} = A_{-+} = -\frac{1}{2}a_0$.

The Wave Equation

As shown in Appendix A, in regions where n is constant, the field equations (3.9) and (3.10) imply the wave equation

$$n^2\ddot{Z}^i - (\Delta Z)^i + R_j^i Z^j + jn\epsilon^{ijk}\nabla_j(\dot{g}_{kl}Z^l) = 0$$

where Δ is the perturbed spatial vector Laplacian (i.e. associated to the perturbed spatial metric g) and R_j^i is the spatial Ricci tensor. As the exterior derivative is independent of the connection (provided it is torsion-free), it suffices to use the unperturbed Levi-Civita connection here: $\nabla \equiv \nabla^{(0)}$.

Let us now expand the electromagnetic field $Z = Z^{(0)} + \varepsilon Z^{(1)}$ as well as the Laplacian $\Delta = \Delta^{(0)} + \varepsilon\Delta^{(1)}$. Since $Z^{(0)}$ satisfies the unperturbed equation $n^2\ddot{Z}^{(0)} - \Delta^{(0)}Z^{(0)} = 0$, only terms of order ε remain, so we obtain at first order

$$\square Z^{(1)i} - (\Delta^{(1)}Z^{(0)})^i + R^{(1)i}_j Z^{(0)j} + jn\epsilon^{ijk}\nabla_j(\dot{g}_{kl}^{(1)}Z^{(0)l}) = 0$$

where we have set

$$\square Z^i := n^2\ddot{Z}^i - (\Delta^{(0)}Z)^i$$

The derivatives of the unperturbed field are of order w , which is much larger than the terms arising from derivatives of the metric or the spatial Ricci tensor, so we neglect the last two

terms in (5.17), which are suppressed by $\Omega = \omega_g/\omega$ or more. At this level of accuracy, we also have

$$\Delta^{(1)} Z^{(0)i} \approx g^{(1)jk} \nabla_j \nabla_k Z^{(0)i} = -\cos(\kappa \cdot x) A^{jk} \nabla_j \nabla_k Z^{(0)i}$$

With these approximations, we arrive at the wave equation

$$\square Z^{(1)i} + \cos(\kappa \cdot x) A^{jk} \nabla_j \nabla_k Z^{(0)i} = 0$$

We shall use one further natural approximation:

$$\cos(\kappa \cdot x) = \cos(\omega_g(t - \hat{\kappa}_z z - r(\hat{\kappa}_x \cos \theta + \hat{\kappa}_y \sin \theta))) \approx \cos(\omega_g(t - \hat{\kappa}_{\parallel} z))$$

which is admissible since the diameter of the waveguide is negligible compared to the Gw wavelength ($\rho\omega_g \ll 1$). As an illustration: for typical optical fibres, ρ measures a couple of micrometers and GW wavelengths detected by LIGO measure hundreds of kilometres: $\rho\omega_g \sim 10^{-10}$, so the error terms here will be much larger than those arising from quadratic terms in the GW amplitude $\varepsilon \sim 10^{-21}$. At this level of accuracy, we may thus replace $\kappa \cdot x$ by $\kappa \cdot x_{\parallel} := \omega_g(t - \hat{\kappa}_{\parallel} z)$ to obtain

$$\square Z^{(1)i} + \cos(\kappa \cdot x_{\parallel}) A^{jk} \nabla_j \nabla_k Z^{(0)i} = 0$$

Using the fact that the connection one-form satisfies $\omega_{ij}|_{ij} = 0$, the second covariant derivative in this equation evaluates to

$$\nabla_j \nabla_k Z^{\parallel} = \mathbf{c}_j(c_k(Z^{\parallel})) - \omega_{kj}^l c_l(Z^{\parallel})$$

Note that the unperturbed field $Z^{(0)}$ depends on the angle θ only as $e^{im\theta}$. Since A has components which vary with θ as $e^{\pm i\theta}$ and $e^{\pm 2i\theta}$, the first order correction will have sidebands with m shifted by one and two in either direction. Moreover, we find it convenient to separate terms which oscillate like the metric perturbation (i.e. trigonometric functions of $\kappa \cdot x_{\parallel}$) into terms with complex exponentials and thus decompose the perturbation of the electromagnetic field (and other quantities if needed) as

$$Z^{(1)a} = \sum_{k=-2}^2 \sum_{l=\pm 1} \mathcal{Z}_{kl}^a(r) e^{i(\omega t - \beta z + (m+k)\theta + 1\kappa \cdot x_{\parallel})}$$

Thus, in $\mathcal{Z}_{kl}^a$ the first index refers to the vector component (typically taken with respect to the complex frame $\mathbf{c}_{\parallel}, \mathbf{c}_{\pm}$), the first lower index determines the angular mode number, and the For the wave equation, we discuss the central mode ($k = 0$) and the sidebands ($k \neq 0$) separately.

Central Mode and Sidebands

For the central mode ($k = 0$), the only contributions come from A III and $A^{+-} = A^{-+}$, so we have

$$(A^{jk} \nabla_j \nabla_k Z^{(0)\parallel})_0 = A^{\parallel\parallel} \partial_z \partial_z Z^{(0)\parallel} + A^{+-} \left[c_+ c_- + c_- \mathbf{c}_+ + \frac{1}{\sqrt{2}r} (\mathbf{c}_+ + \mathbf{c}_-) \right] Z^{(0)\parallel}$$

where the subscript indicates $k = 0$. Now, $\mathbf{c}_{\pm}$ acts on $Z^{(0)\parallel}$ as $\frac{1}{\sqrt{2}} c_m^{\pm}$, as defined in (4.11), so

$$\left[\mathbf{c}_+ \mathbf{c}_- + \mathbf{c}_- \mathbf{c}_+ + \frac{1}{\sqrt{2}r} (\mathbf{c}_+ + \mathbf{c}_-) \right] Z^{(0)1} = (\partial_r^2 + r^{-1} \partial_r - m^2 r^{-2}) Z^{(0)\parallel}$$

Using the unperturbed radial equation (4.14) as well as the trace condition (5.12), we thus obtain

$$(A^{jk}\nabla_j\nabla_k Z^{(0)\parallel})_0 = \frac{1}{2}a_0(\zeta/\rho^2 - 2\beta^2)Z^{(0)\parallel}$$

For the sidebands with m shifted by ± 1 , the only relevant terms are those involving $A^{\mp\parallel} = A_{\pm\parallel} = a_{\pm 1}e^{\pm i\theta}$, so we find

$$(A^{jk}\nabla_j\nabla_k Z^{(0)\parallel})_{\pm 1} = -\sqrt{2}i\beta a_{\pm 1}c_m^{\pm}Z^{(0)\parallel}$$

and for the sidebands with m shifted by ± 2 , the relevant terms come from $A^{\mp\mp} = A_{\pm\pm} = a_{\pm 2}e^{\pm 2i\theta}$, so

$$(A^{jk}\nabla_j\nabla_k Z^{(0)\parallel})_{\pm 2} = \frac{1}{2}a_{\pm 2}\left(c_m^{\pm}c_m^{\pm} - \frac{1}{r}c_m^{\pm}\right)Z^{(0)\parallel} = \frac{1}{2}a_{\pm 2}c_{m\pm 1}^{\pm}c_m^{\pm}Z^{(0)\parallel}$$

The Radial Equations Central Mode

Using separation of variables, the wave equations just derived are reduced to radial equations. We give explicit solutions Bessel equations arising this way.

Using the decomposition (5.24), the inhomogeneous wave equation (5.22) yields

$$B_0^{\pm}Z_{0\pm}^{\parallel 0} = \frac{1}{4}a_0(\zeta/\rho^2 - 2\beta^2)f_0(\alpha, r)$$

where

$$\begin{aligned} B_0^{\pm} &= \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{\zeta_{\pm}}{\rho^2} - \frac{m^2}{r^2}, \\ \zeta_{\pm} &= \rho^2(n^2\omega_{\pm}^2 - \beta_{\pm}^2), \\ \omega_{\pm} &= \omega \pm \omega_g, \\ \beta_{\pm} &= \beta \pm \hat{\kappa}\omega_g. \end{aligned}$$

The homogeneous solutions to this equation which are regular at the origin and vanish at infinity are evidently

$$f_0^{\pm}(\sigma, r) = \begin{cases} \sigma_1 J_m(U_{\pm}r/\rho), & r < \rho \\ \sigma_2 K_m(W_{\pm}r/\rho), & r > \rho \end{cases}$$

where σ_1, σ_2 are constants (continuity conditions relating them are discussed in a later section), and where $U_{\pm}, W_{\pm}$ are defined in analogy to U, W in (4.13):

$$U_{\pm} = \sqrt{\rho^2(n_1^2\omega_{\pm}^2 - \beta_{\pm}^2)}, \quad \text{and} \quad W_{\pm} = \sqrt{\rho^2(\beta_{\pm}^2 - n_2^2\omega_{\pm}^2)}$$

To find particular solutions, we consider the more general problem

$$\rho^2 B_{\nu}^{\pm}p_{\nu}^{\pm} = \begin{cases} \alpha_1 J_{m+\nu}(Ur/\rho), & r < \rho \\ \alpha_2 K_{m+\nu}(Wr/\rho), & r > \rho \end{cases}$$

where α_1, α_2 are arbitrary constants and $B_{\nu}^{\pm}$ is a Bessel operator of arbitrary order ν :

$$B_{\nu}^{\pm} = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{\zeta_{\pm}}{\rho^2} - \frac{(m+\nu)^2}{r^2}.$$

Particular solutions to this equation are given by

$$p_{\nu}^{\pm}(\alpha, r) = \begin{cases} \alpha_1 \left[Y_{m+\nu}(U_{\pm}r/\rho) \Gamma_{m+\nu}^{\pm}(r/\rho) + J_{m+\nu}(U_{\pm}r/\rho) \tilde{\Gamma}_{m+\nu}^{\pm}(r/\rho) \right], & r < \rho, \\ \alpha_2 \left[I_{m+\nu}(W_{\pm}r/\rho) \Delta_{m+\nu}^{\pm}(r/\rho) + K_{m+\nu}(W_{\pm}r/\rho) \tilde{\Delta}_{m+\nu}^{\pm}(r/\rho) \right], & r > \rho, \end{cases}$$

where

$$\begin{aligned}\Gamma_{\nu}^{\pm}(z) &= \frac{\pi}{2} \int_0^z J_{\nu}(U_{\pm} z') J_{\nu}(U z') z' dz', \\ \tilde{\Gamma}_{\nu}^{\pm}(z) &= \frac{\pi}{2} \int_z^1 Y_{\nu}(U_{\pm} z') J_{\nu}(U z') z' dz', \\ \Delta_{\nu}^{\pm}(z) &= - \int_z^{\infty} K_{\nu}(W_{\pm} z') K_{\nu}(W z') z' dz', \\ \tilde{\Delta}_{\nu}^{\pm}(z) &= - \int_1^z I_{\nu}(W_{\pm} z') K_{\nu}(W z') z' dz'\end{aligned}$$

Equation (5.37) is readily verified using the Wronskians

$$\begin{vmatrix} J_m(r) & J'_m(r) \\ Y_m(r) & Y'_m(r) \end{vmatrix} = \frac{2}{\pi r}, \quad \text{and} \quad \begin{vmatrix} K_m(r) & K'_m(r) \\ I_m(r) & I'_m(r) \end{vmatrix} = \frac{1}{r},$$

cf. e.g. Ref. [36, eqns. 9.1.15 on p. 360 and 9.6.15 on p. 375]. The integral ranges were chosen as follows. The function $\Gamma_{\nu}^{\pm}$ multiplies Y_{ν} which diverges at the origin, so we chose the region of integration such that $\Gamma_{\nu}^{\pm}$ vanishes there. The region of integration for $\Delta_{\nu}^{\pm}$ was chosen for a similar reason, as I_{ν} diverges at large distances. Finally, the ranges for $\tilde{\Gamma}_{\nu}^{\pm}$ and $\tilde{\Delta}_{\nu}^{\pm}$ were chosen such that these functions vanish at the core-cladding interface $r = \rho$, where certain components of the electromagnetic field are required to be continuous.

Using the function p_0 , one obtains the general solution to (5.30) which is regular at the origin and vanishes at infinity:

$$\mathcal{Z}_{0\pm}^{\parallel 0\pm} = a_0 \left[\frac{1}{4} (\zeta - 2\rho^2 \beta) p_0^{\pm}(\alpha) + f_0^{\pm}(\sigma_{0,\pm}) \right],$$

where a_0 was factored out for convenience.

The Radial Equations Sidebands

Sidebands Using the decomposition (5.24) for the sidebands also, the inhomogeneous wave equation (5.22) reduces to

$$\begin{aligned}B_{\pm 1}^+ \mathcal{Z}_{\pm 1+}^V &= B_{\pm 1}^- \mathcal{Z}_{\pm 1-}^{\parallel} = \frac{i}{\sqrt{2}} a_{\pm 1} \beta^{-1} f_{\pm 1}(\alpha) \\ B_{\pm 2}^+ \mathcal{Z}_{\pm 2+}^V &= B_{\pm 2}^- \mathcal{Z}_{\pm 2-}^{\parallel} = \frac{1}{4} a_{\pm 2} \rho^{-2} f_{\pm 2}(\alpha)\end{aligned}$$

where $B_{\nu}^{\pm}$ is as in (5.38). Proceeding in the same way as for the central mode, one obtains

$$\begin{aligned}\mathcal{Z}_{\pm 1+}^{\parallel} &= a_{\pm 1} \left(\frac{i}{\sqrt{2}} \rho \beta p_{\pm 1}^+ (\mp U \alpha_1, -W \alpha_2, r) + f_{\pm 1}^+(\sigma_{\pm 1,+}, r) \right), \\ \mathcal{Z}_{\pm 1-}^{\parallel} &= a_{\pm 1} \left(\frac{i}{\sqrt{2}} \rho \beta p_{\pm 1}^- (\mp U \alpha_1, -W \alpha_2, r) + f_{\pm 1}^-(\sigma_{\pm 1,-}, r) \right), \\ \mathcal{Z}_{\pm 2+}^{\parallel} &= a_{\pm 2} \left(\frac{1}{4} p_{\pm 2}^+ (U^2 \alpha_1, W^2 \alpha_2, r) + f_{\pm 2}^+(\sigma_{\pm 2,+}, r) \right) \\ \mathcal{Z}_{\pm 2-}^{\parallel} &= a_{\pm 2} \left(\frac{1}{4} p_{\pm 2}^- (U^2 \alpha_1, W^2 \alpha_2, r) + f_{\pm 2}^-(\sigma_{\pm 2,-}, r) \right)\end{aligned}$$

where the various σ 's are parameters to be determined from continuity conditions, and the functions $f_k^\pm$ are derived from $f_0^\pm$, defined in (5.35), by successive application of the ladder operators c_m defined in (4.11):

$$\begin{aligned} f_{\pm 1}^+ &= \rho c_m^\pm f_0^+, & f_{\pm 2}^+ &= \rho^2 c_{m\pm 1}^\pm c_m^\pm f_0^+, \\ f_{\pm 1}^- &= \rho c_m^\pm f_0^-, & f_{\pm 2}^- &= \rho^2 c_{m\pm 1}^\pm c_m^\pm f_0^- \end{aligned}$$

Having found the longitudinal components of the electromagnetic field, one can now compute the remaining (transverse) components. 5.2 The Transverse Components To obtain the transverse components of Z , consider the $\pm$ components of (3.9). Expanding $g = g^{(0)} + gg^{(1)}$, $Z = Z^{(0)} + \varepsilon Z^{(1)}$, and using the unperturbed equation, one obtains at first order

$$n\partial_0 Z^{(1)i} + j\epsilon^{ijk}\nabla_j \left(g_{kl}^{(0)} Z^{(1)l} \right) + j\epsilon^{ijk}\nabla_j \left(g_{kl}^{(1)} Z^{(0)l} \right) = 0$$

In the last term, we may neglect derivatives of the metric perturbation (of order ω_g) compared to derivatives of the unperturbed electromagnetic field (of order ω) to arrive at

$$(n\partial_0 \pm ij\partial_z) Z^{(1)\pm} = \pm ij [c_\mp (Z^{(1)\parallel}) + \chi^\pm \cos(\kappa \cdot x_1)]$$

where we have used (4.7) and have set

$$\chi^\pm = A_{\parallel l} \nabla_\mp Z^{(0)l} + i\beta A_{\mp l} Z^{(0)l}$$

Using the notation of (5.24), a direct calculation shows that

$$\begin{aligned} \chi_{-2}^+ &= i\beta a_{-2} Z^{(0)-}, & \chi_{+2}^- &= i\beta a_{+2} Z^{(0)+} \\ \chi_{-1}^+ &= \frac{i}{2} a_{-1} (3\beta - i j n \omega) Z^{(0)}, & \chi_{+1}^- &= \frac{i}{2} a_{+1} (3\beta + i j n \omega) Z^{(0)1} \\ \chi_0^+ &= -i a_0 \left(\frac{3}{2} \beta - i j n \omega \right) Z^{(0)+}, & \chi_0^- &= -i a_0 \left(\frac{3}{2} \beta + i j n \omega \right) Z^{(0)-} \\ \chi_{+1}^+ &= \frac{1}{\sqrt{2}} a + 1 c_{m+1}^+ Z^{(0)+}, & \chi_{-1}^- &= \frac{1}{\sqrt{2}} a - 1 c_{m-1}^- Z^{(0)-} \\ \chi_{+2}^+ &= 0, & \chi_{-2}^- &= 0 \end{aligned}$$

To arrive at the stated formula for $\chi_0^\pm$, we have used $e_\mp (Z^{(0)\parallel}) = -i(\beta \mp i j n \omega) Z^{(0)\pm}$ which follows from (4.9), and the trace condition (5.12). For χ_{-1}^+ and χ_{+1}^- we have used

$$\rho^2 c_{m+1}^- c_m^+ f_0 = \rho^2 c_{m-1}^+ c_m^- f_0 = -\zeta f_0$$

which can be shown either using the recursion relations (4.15) or by noting that $c_{m+1}^- c_m^+ = c_{m-1}^+ c_m^- = B_0 - \zeta/\rho^2$ and using the Bessel equation $B_0 f_0 = 0$. These formulae have the general structure that the radial dependence of χ_b^a is proportional to $a_b f_{a+b}$.

Using the notation (5.24), as well as the abbreviations $\omega_\pm$ and $\beta_\pm$ as defined in (5.33) and (5.34), equation (5.55) reduces to

$$\begin{aligned} i(n\omega + \mp i j \beta_+) \mathcal{Z}_{k+}^\pm &= \pm i j \left[\frac{1}{\sqrt{2}} c_{m+k}^\pm \mathcal{Z}_{k+}^\parallel + \frac{1}{2} \chi_k^\pm \right] \\ i(n\omega - \mp i j \beta_-) \mathcal{Z}_{k-}^\pm &= \pm i j \left[\frac{1}{\sqrt{2}} c_{m+k}^\pm \mathcal{Z}_{k-}^\parallel + \frac{1}{2} \chi_k^\pm \right] \end{aligned}$$

Multiplying the first equation by $n\omega_+ \pm i j \beta_+$ and the second one by $n\omega_- \pm i j \beta_-$, and using the definition of $\zeta_\pm$ given in (5.32), one obtains

$$\begin{aligned} i\zeta_+ \mathcal{Z}_{k+}^\pm &= \frac{1}{\sqrt{2}} \rho^2 [\beta_+ \pm i j n \omega_+] \left(c_{m+k}^\pm \mathcal{Z}_{k+}^\parallel + \frac{1}{\sqrt{2}} \chi_k^\pm \right) \\ i\zeta_- \mathcal{Z}_{k-}^\pm &= \frac{1}{\sqrt{2}} \rho^2 [\beta_- \pm i j n \omega_-] \left(c_{m+k}^\pm \mathcal{Z}_{k-}^\parallel + \frac{1}{\sqrt{2}} \chi_k^\pm \right) \end{aligned}$$

which is structurally similar to the unperturbed equation (4.9). Now that the field equations are solved in the core and cladding separately, we consider how these solutions match at the core-cladding interface of the waveguide. 5.3 Continuity Conditions From (3.3), one finds that the field components $D^r, E_\theta, E_z, B^r, H_\theta, H_z$ must be continuous at the core-cladding interface $r = \rho$. As is well-known, these six continuity conditions are linearly dependent, and it suffices to impose continuity of the four components

$$D^r = \Re Z^r, \quad B^r = n^{-1} \Im_\xi^r, \quad H_\parallel = n^{-1} g_{\parallel i} \Im^i, \quad E_\parallel = n^{-2} g_{\parallel i} \Re Z^i$$

where $\Re$ and $\Im$ refer to the j -real and j -imaginary parts, respectively. To assess the jumps of the fields, we introduce the notation

$$[[f]] := \lim_{r \nearrow \rho} f(r) - \lim_{r \leq \rho} f(r),$$

and write symbolically

$$[[Z]] = \begin{pmatrix} [[D^r]] & [[E_\parallel]] [[B^r]] & [[H_\parallel]] \end{pmatrix}^\top,$$

which conflates the jumps of all relevant fields in one column vector. Since the jumps arising in (5.67) are determined by the jumps of the two components

$$\begin{aligned} Z^r &= \frac{1}{\sqrt{2}} (Z^+ + Z^-) \\ g_{\parallel i} Z^i &= Z^{(0)\parallel} + \varepsilon Z^{(1)\parallel} + \varepsilon A_{\parallel i} Z^{(0)i} \cos(\kappa \cdot x_\parallel) \end{aligned}$$

for the perturbed problem it suffices to consider the jumps of

$$\begin{aligned} \mathcal{Z}_{kl}^r &= \frac{1}{\sqrt{2}} (\mathcal{Z}_{kl}^+ + \mathcal{Z}_{kl}^-), \\ \mathcal{Z}_{\parallel kl} &= \mathcal{Z}_{kl}^\parallel + \frac{1}{2} \chi_k^d, \end{aligned}$$

where the $\mathcal{Z}_{kl}^a$ are the components from the decomposition (5.24), and where

$$\chi_0^\parallel = a_0 Z^{(0)\parallel}, \quad \chi_{\pm 1}^1 = a_{\pm 1} Z^{(0)\pm}, \quad \chi_{\pm 2}^\parallel = 0.$$

Similar to the decomposition (5.24), we write

$$[[Z]] \equiv \left([[Z]]^{(0)} + \varepsilon \sum_{k=-2}^{+2} \sum_{l=\pm} [[Z]]_{kl}^{(1)} e^{i(k\theta + l\kappa \cdot x_1)} \right) e^{i(\omega t - \beta z + m\theta)} = 0$$

where the first term does not contain any ε corrections because all components of $Z^{(1)}$ oscillate with $\exp(\pm i\kappa \cdot x_1)$.

The four j -real components of $[[Z]]_{kl}^{(1)}$ are parametrised by α (coming from the particular solutions) and σ_{kl} (homogeneous solutions), each having four j -real parameters. Hence, one can write

$$[[Z]]_{kl}^{(1)} = a_k (M_{kl} \alpha + \Sigma_{kl} \sigma_{kl}) \quad (\text{no summation implied})$$

where α and σ_{kl} are j -real vectors:

$$\alpha = (\Re \alpha_1, \Re \alpha_2, \Im \alpha_1, \Im \alpha_2)^\top, \quad \sigma_{kl} \text{ similar,}$$

and M_{kl} and Σ_{kl} are j -real 4×4 matrices. Note that the jumps are not simply linear in the complex coefficients α, σ_{kl} since (5.67) discriminates between the real and complex parts.

In Appendix B.1, it is shown that because the electric and magnetic fields are derived from a single complex field, these matrices are fully determined by their first two rows, which motivates the following abbreviating notation:

$$\begin{vmatrix} a & b & c & d \\ e & f & g & h \end{vmatrix} := \begin{pmatrix} a & b & c & d \\ e & f & g & h \\ -c/n_1 & -d/n_2 & a/n_1 & b/n_2 \\ -gn_1 & -hn_2 & en_1 & fn_2 \end{pmatrix}$$

As many matrices considered here have similar structures, we define the following abbreviation

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ \beta & \omega & U & W \end{bmatrix}_\nu :=$$

where the parameters $a_1, \dots, a_4$ and β, ω determine the numerical coefficients in the first row of the matrix, and the last two parameters U, W determine the arguments of the Bessel functions, whose order is given by the integer ν which is given as a general subscript.

Within this class of matrices, we identify the following families, which are useful for the problem considered here:

$$\Pi_k = \begin{bmatrix} 1/U & 1/W & 1/U^2 & 1/W^2 \\ \beta & \omega & U & W \end{bmatrix}_{m+k}$$

$$\Sigma_k^\pm = \begin{bmatrix} 1/U_\pm & 1/W_\pm & 1/U_\pm^2 & 1/W_\pm^2 \\ \beta_\pm & \omega_\pm & U_\pm & W_\pm \end{bmatrix}_{m+k}$$

where $\beta_\pm, \omega_\pm$ are defined in (5.33) and (5.34), and $U_\pm, W_\pm$ are defined in (5.36). The Π family is used to describe the central mode and to formulate the dispersion relation, and the Σ matrices arise when considering the jumps of the homogeneous solutions to the radial equations. Similar to the family (5.79), we define

$$\left\{ \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ \beta & \omega & U & W \end{bmatrix}_\nu \mid \begin{vmatrix} -ia_1\rho\beta Y'_\nu(U) & -ia_2\rho\beta I'_\nu(W) & a_3n_1\nu\rho\omega Y_2(U) & a_4n_2\nu\rho\omega I_2(W) \\ \frac{1}{n_1^2}Y_\nu(U) & -\frac{1}{n_2^2}I_\nu(W) & 0 & 0 \end{vmatrix} \right\}$$

where the Bessel functions J_ν and K_ν are replaced by Y_ν and I_ν . Using this notation, let

$$P_k^\pm = \left\{ \begin{bmatrix} U/U_\pm^2 & W/W_\pm^2 & 1/U_\pm^2 & 1/W_\pm^2 \\ \beta_\pm & \omega_\pm & U & W \end{bmatrix}_{m+k} \Lambda_{m+k}^\pm \right\}$$

where

$$\Lambda_\nu^\pm = \text{diag}(\Gamma_\nu^\pm(1), \Delta_\nu^\pm(1), \Gamma_\nu^\pm(1), \Delta_\nu^\pm(1))$$

This is used to describe the jumps of the particular solutions $p_\nu^\pm$ defined in (5.39). The functions $\Gamma_\nu^\pm, \Delta_\nu^\pm$ appearing here are defined in (5.40) and (5.42), while the functions $\tilde{\Gamma}_\nu^\pm, \tilde{\Delta}_\nu^\pm$ defined in (5.41) and (5.43) do not contribute, since they vanish at the core-cladding interface $r = \rho$. Having prepared the notation, we now analyse (5.75). By linear independence of the exponential functions, all terms $\llbracket Z \rrbracket^{(0)}$ and $\llbracket Z \rrbracket_{kl}^{(1)}$ must vanish separately.

Dispersion Relation

Because all corrections coming from the gravitational wave oscillate as $\exp(\pm i\kappa.x_1)$, the first term in (5.75) is the same as for the unperturbed problem. As shown explicitly in Appendix B.2, the continuity condition for this part is given by

$$\Pi_0\alpha = 0$$

where Π_0 is given by (5.80). To obtain a non-trivial solution, the coefficient matrix must be singular: $\det \Pi_0 = 0$, which is equivalent to

$$(\mathcal{Z}_m + \mathcal{K}_m) (n_1^2 \mathcal{Z}_m + n_2^2 \mathcal{K}_m) = (m\beta/\omega)^2 (U^{-2} + W^{-2})^2$$

where we have used the abbreviations

$$\mathcal{F}'_m = \frac{J'_m(U)}{U J_m(U)}, \quad \mathcal{K}_m = \frac{K'_m(W)}{W K_m(W)},$$

cf. e.g. Ref. [34, Eq. 3.27 on p. 124] and Ref. [35, Eq. 16.166 on p. 511] (where ϵ_2/ϵ_2 should read ϵ_1/ϵ_2). Because β and ω enter this equation through U and W as arguments of Bessel functions, this is a transcendental equation which must be solved numerically. More precisely, by solving $\det \Pi_0 = 0$ for a prescribed vacuum wavelength $\lambda = 2\pi/\omega$ and given waveguide parameters (i.e. core radius ρ and refractive indices n_1, n_2), one can determine the effective refractive index $\bar{n} := \beta/\omega$ as a function of λ : $\bar{n} = \bar{n}(\lambda)$, which constitutes the waveguide dispersion relation. Subsequently, the coefficients α can be determined from (5.85) up to an overall factor which describes the amplitude of the electromagnetic wave.

In general, there are multiple solutions to this equation for given waveguide parameters and given values of m and w . Here, we focus on single-mode fibres (i.e. fibres with sufficiently thin cores, operated at suitable frequencies) where there is only one single solution for $m = +1$ and there are no solutions for higher values of m .

Central Mode

For the central mode ($k = 0$), the continuity condition can be written as

$$\Sigma_0^\pm \sigma_{0,\pm} + \mathbf{P}_0^\pm \Xi \boldsymbol{\alpha} + \frac{1}{2} \Psi_0 \boldsymbol{\alpha} = 0$$

The first term describes the jumps arising from the homogeneous solution in (5.45). In the second term, which comes from the particular solution, the matrix Ξ describes the discontinuity of the factor $\frac{1}{4}(\zeta - 2\rho^2\beta^2)$ in (5.45):

$$\Xi = -\frac{1}{4} \text{diag} (2\rho^2\beta^2 - U^2, 2\rho^2\beta^2 + W^2, 2\rho^2\beta^2 - U^2, 2\rho^2\beta^2 + W^2)$$

Дальше многое я не копировал в надежде, что не нужно

(лишь скопировано, чем заканчивается этот анализ)

Since the range of Σ'' is the kernel of Π_{-2} , the two vectors $C_{-2}\sigma_{-2,\pm}$ are proportional to α^* , and we find, similar to the central mode

$$C_{-2}\sigma_{-2,\pm} = \pm a_{-2} \frac{\Delta^2}{2\Omega} \frac{c_2}{\beta' - \hat{\kappa}_\parallel} \alpha^*$$

where β' is the same constant as in the phase correction (5.105), and the coefficient c_2 is a new parameter, which we determine numerically. Here, the relative index difference Δ is defined as

$$\Delta = 1 - \frac{n_2}{n_1}$$

and we have factored out $\Delta^2/2$ such that c_2 is typically of order unity (see the numerical examples below).

As shown explicitly in Section 7, this formula essentially describes the perturbation of the polarisation by the gravitational wave.

Perturbation of the Phase

In this section, we implement boundary conditions which model a setup where monochromatic laser light of definite polarisation is injected into the fibre. As we will see, contrary to the unperturbed case, this will require adding further terms to the field computed so far. This leads to an expression for the phase correction which vanishes at the point where the light enters the fibre and grows with increasing distance. A similar construction for the eikonal in a non-dispersive regime can be found in Ref. [4, Sect. 2].

Neglecting boundary effects from the ends of the waveguide, the light sent into the waveguide (or coming out of it) is simply obtained by restricting the field to a definite value of z , where we consider $z = 0$ for simplicity. Also, we restrict the discussion to the "large corrections", i.e. to $k = 0$ and $k = -2m$ where the perturbations were found to be of order $1/\Omega$. The remaining "small components" are expected to be experimentally irrelevant. Recall that the coefficients of the central mode ($k = 0$) were given by

$$\sigma_{0,\pm} = \mp \frac{1}{4\Omega} \frac{c1}{\beta' - \hat{\kappa}_{\parallel}} \alpha$$

Since this is proportional to the unperturbed coefficients α , we may write the central mode ($k = 0$) of the overall field in the form

$$Z_0^{(0)a} + \varepsilon Z_0^{(1)a} = 3^a \bar{\Phi}$$

where

$$\begin{aligned} 3^{\parallel} &= f_0(\alpha; r), \\ i\zeta_{\Xi\pm} &= \frac{1}{\sqrt{2}} \rho(\beta \pm i j n \omega) f_{\pm}(\alpha; r), \\ \Phi &= e^{i\psi^{(0)}} \left[1 - \frac{\varepsilon}{4\Omega} \frac{a_0 \beta'}{\beta' - \hat{\kappa}_{\parallel}} (e^{+i\kappa \cdot x_1} - e^{-i\kappa \cdot x_{\parallel}}) \right], \\ \psi^{(0)} &= \omega t - \beta z + m\theta. \end{aligned}$$

Evaluating the perturbation terms at $z = 0$, we find that this describes a situation where the incoming phase is modulated in a highly specific way by the gravitational wave. This is not the situation which we want to model so that correction terms are needed: instead of the phase modulation behaviour given by Φ , we are concerned with the scenario where a polarised monochromatic plane wave (as emitted by a laser source) is injected into the fibre at $z = 0$. Hence, we require the injected field to have a definite phase of $\psi = \omega t + m\theta$ (and definite polarisation, see the next section) at $z = 0$.

Such boundary conditions can be implemented by adding appropriate solutions of the unperturbed equations of slightly shifted frequencies. Specifically, we add terms of the form

$$3_{\pm}^a = 3^a \exp(i(\omega_{\pm} t - (\beta \pm \omega_g \beta^t) z + m\theta))$$

where the term $\beta' = \partial\beta/\partial\omega$ ensures that the unperturbed dispersion relation is satisfied. (The shift in the frequency also causes a slight perturbation in the amplitude, but this is of order Ω and thus negligible here.) The corrected overall field is thus

$$Z_0^a = 3^a \tilde{\Psi} + \frac{\varepsilon}{4\Omega} \frac{a_0 c_1}{\beta^s - \hat{\kappa}_{\parallel}} (\dot{j}_+^a - \dot{j}_-^a) = 3^a \Phi$$

where

$$\Phi = e^{i\psi^{(0)}} \left[1 - \frac{\varepsilon}{4\Omega} \frac{a_0 c_1}{\beta' - \hat{\kappa}_{\parallel}} \left(e^{+i\omega_g t} \left(e^{-i\omega_g \hat{\kappa}_{\parallel} z} - e^{-i\omega_g \beta^t z} \right) - e^{-i\omega_g t} \left(e^{+i\omega_g \hat{\kappa}_{\parallel} z} - e^{+i\omega_g \beta^t z} \right) \right) \right]$$

Assuming the fibre to be much shorter than the GW wavelength, i.e. $\omega_g z \ll 1$, one finally obtains

$$\Phi \approx e^{i\psi^{(0)}} \left[1 - \frac{i}{2} \varepsilon a_0 c_1 \cos(\omega_g t) \omega z \right] \approx e^{i\psi^{(0)}} \exp \left(-\frac{i}{2} \varepsilon a_0 c_1 \cos(\omega_g t) \omega z \right)$$

which can be written as a perturbation of the phase:

$$\psi = \omega t - \beta z + m\theta - \frac{1}{2} \varepsilon a_0 c_1 \omega z \cos(\omega_g t)$$

This shows that the phase perturbation grows linearly with the distance from the emission point and oscillates in time exactly as the gravitational wave does.

Perturbation of the Polarisation

Having identified the perturbation of the central mode $k = 0$ as a correction to the phase, we now show that the other "large term" with $m + k = -m$ corresponds to a perturbation of the polarisation.

To see this, we follow Ref. [34], where it is shown that modes with real field patterns, i.e. an angular dependence of the form $\cos \theta$ or $\sin \theta$, correspond to linear polarisation. To describe such fields, one must take linear combinations of the solutions with $m = +1$ and $m = -1$, and we shall make the standard assumption of weak guidance $\Delta \equiv (n_1 - n_2)/n_1 \ll 1$. It should be noted that non-zero values of Δ cause the modes not to be perfectly linearly polarised, but

(это тоже не копировал в надежде, что не нужно)

Numerical Examples and Angular Dependence (????)

A Derivation of the Wave Equation

Here, we derive the wave equation for the electromagnetic field in the Gw metric considered above, neglecting terms of order ε^2 . Throughout this section, we raise and lower indices with the full spatial metric g_{ij} . Equation (3.9) can then be written as

$$n\dot{Z}^i + j\epsilon^{ijk}\nabla_j Z_k = 0$$

where ∇ denotes the spatial covariant derivative. Lowering the index, keeping in mind that the spatial metric is time dependent, one obtains the equivalent form

$$n\dot{Z}_i - n\dot{g}_{ij}Z^j + j\epsilon_{ijk}\nabla^j Z^k = 0$$

Applying ∂_0 to the first equation and using this last equation, we find

$$n^2\ddot{Z}^i + j\epsilon^{ijk}\nabla_j (n\dot{g}_{kl}Z^l - j\epsilon_{klm}\nabla^l Z^m) = 0$$

Here, no time derivatives of the volume form occur as the trace of the metric perturbation is time-independent, and no derivatives of the Christoffel symbols occur as the exterior derivative is independent of the connection. Using $\nabla_j Z^j = 0$ and the Ricci identity, we have

$$\epsilon^{ijk}\epsilon_{klm}\nabla_j \nabla^l Z^m = \nabla_j \nabla^i Z^j - \nabla_j \nabla^j Z^i = R_j^i Z^j - \Delta Z^i$$

where R_j^i is the spatial Ricci tensor. Inserting this into the previous equation then yields

$$n^2\ddot{Z}^i - \Delta Z^i + R_j^i Z^j + j\epsilon^{ijk}\nabla_j (n\dot{g}_{kl}Z^l) = 0.$$

B Continuity Conditions

The jumps of the various field components are linear combinations of the various parameters α and σ . For example, if w denotes any field component

$$[w] = a^1\alpha_1 + a^2\alpha_2 + s^1\sigma_1 + s^2\sigma_2$$

with some coefficients a^1, a^2, s^1, s^2 , for which we shall simply write

$$[w] = (a^1, a^2)_\alpha + (s^1, s^2)_\sigma$$

To analyse the jumps of the real fields E, D, B, H , we separate the j -real and j -imaginary parts. Accordingly, we write

$$(a + jb, c + jd)_\alpha = [a + jb, c + jd, -b + ja, -d + jc]_\alpha$$

(далее не копировал, потому что все равно очень плохо понимаю, что происходит.)

15 Другие темы

15.1 Cosmology with Gravitational Wave Observations by Sathyaprakash

Gravitational wave observations may inform us about cosmology in at least two ways: by studies of individual sources at cosmological distances that give information about cosmography (the structure and kinematics of the universe) and about early structure formation, and by direct observation of a stochastic background of gravitational waves of cosmological origin. In turn, a stochastic background could either be astrophysical in origin (generated by any of a myriad of astrophysical systems that have arisen since cosmological structure formation began, as described in Section 8.2.2), or it could come from the Big Bang itself (generated by quantum processes associated with inflation or with spontaneous symmetry breaking in the extremely early universe, as described in Section 8.2.1). The observation of a cosmic gravitational wave background (CGWB) is probably the most fundamentally important observation that gravitational wave detectors can make. But the astrophysical gravitational wave background (AGWB) also contains important information and may mask the CGWB over much of the accessible spectrum.

The detection of discrete sources at cosmological distances will require high sensitivity. Advanced ground-based detectors should be able to see a few individual sources (mainly stellar-mass black hole binaries) at redshifts approaching 1, with which they may be able to make a good determination of the Hubble constant. But LISA's observations of the coalescences of massive black hole binaries at all redshifts should make LISA a significant tool for cosmography. We examine cosmography measurements in Section 8.3. These high- z observations may also contain interesting information about early structure formation, such as the relationship between SMBH formation and galaxy formation. We have mentioned this already in Section 7.2.4.

Both kinds of detectors will search for a stochastic background in their own wave band. As we have seen earlier, LISA will almost certainly detect an AGWB from binary systems in our galaxy, and both LISA and advanced ground-based detectors may see a CGWB, if the more optimistic estimates of its strength are correct. But scientists are already sketching designs for a mission to follow LISA with much higher sensitivity, dedicated to observing the CGWB from inflation. Stochastic searches are described in Section 8.1.2.

Other detection methods are also being used to probe the spectrum of the background radiation at longer wavelengths. Pulsar timing observations (Section 8.1.3) are already being used to set limits on the background at periods of a few years, and they will reach much greater sensitivity when coherent antenna arrays (like the Square Kilometer Array [224, 107]) are available. And observations of the temperature fluctuations of the cosmic microwave background (Section 8.1.4) have the potential to reveal the gravitational wave content of the universe at the redshift of decoupling, which means at wavelength scales comparable to the size of the universe [300, 214].

Before examining the details of detection, we begin by examining the statistics of a random gravitational wave background. A good introduction to the theory of the CGWB is the set of lectures by Bruce Allen at the 1996 Les Houches summer school [27]. The first paper of the LSC on searches for a stochastic background [1] also contains a brief introduction.

15.1.1 Detecting a stochastic gravitational wave background

Describing a random gravitational wave field

By definition, a stochastic background of gravitational waves is a superposition of waves arriving at random times and from random directions, overlapping so much that individual waves are not identifiable. We assume that there are so many sources (either astrophysical sources or the quantum fluctuations that create the CGWB) that individual ones are not distinguishable. Such a gravitational wave field will appear in detectors as a time-series noise, which by the central limit theorem should have a Gaussian-normal distribution function if there are enough overlapping sources. This kind of background will compete with instrumental noise. It will be detectable by a single detector, if it is stronger than instrumental noise, but a weaker background could still be detected by using a pair of detectors and looking for a correlated component of their "noise" output, on the assumption that their instrumental noise is not correlated.

As a random phenomenon, the gravitational wave fields at two different locations are uncorrelated, because gravitational waves arrive from all directions and at all frequencies. It might, therefore, be thought that two detectors' responses would be correlated only if they were located at the same position. But if one considers one component of the wave field with a single frequency, then it is clear that there will be strong correlations between points if they are separated along the wave's propagation direction by much less than a wavelength. We shall see that these frequencydependent correlations allow one to detect a background by cross-correlating the output of two separated detectors, albeit with less sensitivity than if they were co-located. We shall consider cross-correlation as a detection method in Section 8.1.2.

Random gravitational waves are conventionally described in terms of their energy density spectrum $\rho_{\text{gw}}(f)$, rather than their mean amplitude. It is convenient to normalize this energy density to the critical density ρ_c required to close the universe, which is given in terms of the Hubble constant H_0 as

$$\rho_c = 3H_0^2/8\pi G.$$

We then define

$$\Omega_{\text{gw}} := \frac{d\rho_{\text{gw}}/\rho_c}{d\ln f}$$

This can be interpreted as the fraction of the closure energy density that is in random gravitational waves between the frequency f and $e \times f$. If the source of radiation is scale-free (which means that there is no preferred length or time scale in the process), then it will

produce a power-law spectrum, i.e., one in which $\Omega_{\text{gw}}(f)$ depends on a power of f . Inflation, as we describe below, predicts a flat energy spectrum, one in which Ω_{gw} is essentially independent of frequency [27].

The energy in the cosmological background is, of course, related to the spectral density of the noise that the background would produce in a gravitational wave detector. Since we describe the gravitational wave noise in terms of amplitude rather than energy, there are scaling factors involving the frequency between the two. An isotropic gravitational wave background incident on an interferometric detector will induce a strain spectral noise density equal to [359, 27]

$$S_{\text{gw}}(f) = \frac{3H_0^2}{10\pi^2} f^{-3} \Omega_{\text{gw}}(f)$$

Note that the explicit dependence on frequency is f^{-3} : two factors come from the relation of energy and squared-strain, and one factor from the fact that Ω_{gw} is an energy distribution per unit logarithmic frequency. Note also that there are no explicit factors of c or G needed in this formula if one wants to work in nongeometrized units.

If we scale H_0 by $h_{100} = H_0 / (100 \text{ km s}^{-1} \text{ Mpc}^{-1})$, and we note that $100 \text{ km s}^{-1} \text{ Mpc}^{-1} = 3.24 \times 10^{-18} \text{ s}^{-1}$, then this equation implies that the strain noise is

$$S_{\text{gw}}^{1/2} = 5.6 \times 10^{-22} \text{ Hz}^{-1/2} \Omega_{\text{gw}}^{1/2} \left(\frac{f}{100 \text{ Hz}} \right)^{-3/2} h_{100}$$

Observations with gravitational wave detectors

To be observed by a single gravitational wave detector, the gravitational wave noise must be larger than the instrumental noise. This is a bolometric method of detection of the background, and it requires great confidence in the understanding of the detector, in order to believe that the observed noise is external. This is how the cosmic microwave background was originally discovered in a radio telescope by Penzias and Wilson.

If there are two detectors, then one may be able to get better sensitivity by cross-correlating their output, as mentioned in Section 4.7.3 above. This works best when the two detectors are close enough together to respond to the same random wave field. Even when they are separated, however, they are correlated well at lower frequencies.

From Equation 135 and the discussion in Section 4.7.3 it is straightforward to deduce that two co-located detectors, each with spectral noise density S_h and fully uncorrelated instrumental noise, observing over a bandwidth f at frequency f for a time T , can detect a stochastic background with energy density

$$\Omega_{\text{gw}}^{1/2} h_{100} = \left(\frac{S_h^{1/2}}{3.1 \times 10^{-18} \text{ Hz}^{-1/2}} \right) \left(\frac{f}{10 \text{ Hz}} \right)^{5/4} \left(\frac{T}{3 \text{ yrs}} \right)^{-1/4}$$

The two LIGO detectors (separated by about 10 ms in light-travel time) are reasonably well placed for performing such correlations, particularly when upgrades push their lower frequency limit to 20 Hz or less. Two co-located first-generation LIGO instruments operating at 100 Hz could, in a one-year correlation, reach a sensitivity of $\Omega_{\text{gw}} \sim 1.7 \times 10^{-8}$. But the separation of the actual detectors takes its toll at this frequency, so that they can actually only reach $\Omega_{\text{gw}} \sim 10^{-6}$. Advanced LIGO may improved this by two or three orders of magnitude, going well below the nucleosynthesis bound. The third-generation instrument ET, with instrumental noise as shown in Figure 5, can go even deeper. Two co-located ETs, observing at 10 Hz for three years, could reach $\Omega_{\text{gw}} \sim 10^{-12}$. At this frequency the detectors could be as far apart as 5000 km without a substantial loss in correlation sensitivity. The numbers given here are reflected in the curves in Figure 5.

Correlation searches are also possible between resonant detectors or between one resonant and one interferometric detector [53]. This has been implemented with bar detectors [52] and between LIGO and the ALLEGRO bar detector [391].

LISA does not gain by a simple correlation between any two of its independent interferometers, since they share a common arm, which contributes common noise that competes with that of the background. A gravitational wave background of $\Omega_{\text{gw}} \sim 10^{-10}$ would compete with LISA's expected instrumental noise. However, using all three interferometers together can improve things for LISA at low frequencies, assuming that the LISA instrumental noise is well behaved [194]. This might enable LISA to go below 10^{-11} .

Observations with pulsar timing

Other less-direct methods are also being used to search for primordial gravitational waves. As we saw in Section 4.4.2, pulsar timing can, in principle, detect gravitational-wave-induced fluctuations in the arrival times of pulses. Millisecond pulsars are such stable clocks when averaged over years of observations that they are being used to search for gravitational waves with periods longer than one year. A single pulsar can set limits on a stochastic background by removing the slow spindown and looking for random timing residuals. Although one would never have enough confidence in the stability of a single pulsar to claim a detection, this sets upper limits in the important frequency range below that accessible to man-made instruments. The best such limits are on pulsar PSR B1855+09, with an upper limit (at 90% confidence) of $\Omega_{\text{gw}} < 4.8 \times 10^{-9}$ at $f = 4.4\text{ nHz}$ [212].

Arrays of pulsars offer the possibility of cross-correlating their fluctuations, which makes it possible to distinguish between intrinsic variability and gravitational-wave-induced variability. Pulsars are physically separated by much more than a wavelength of the gravitational waves even with periods of 10yrs, so that the correlated fluctuations come from the wave amplitudes at Earth. It will soon be possible to monitor many pulsars simultaneously with multibeam instruments, as mentioned in Section 4.4.2. This method could push the limits on $h_c \equiv (f S_{\text{gw}})^{1/2}$ [cf. Equation (134)] down to 10^{-16} at 10nHz [223], which translates into a limit on Ω_{gw} of around 10^{-12} .

Observations using the cosmic microwave background

Observations of the cosmic microwave background (CMB) may in fact make the first detections of stochastic (or any other!) gravitational waves. The temperature fluctuations first detected by the Cosmic Background Explorer (COBE) [342] and more recently measured with great precision by the Wilkinson Microwave Anisotropy Probe (WMAP) [72] are produced by both density perturbations and long-wavelength gravitational waves in the early universe (see the next Section 8.2.1). Inflation suggests that the gravitational wave component may be almost as large as the density component, but it can only be separated from the density perturbations by looking at the polarization of the cosmic microwave background [215]. WMAP made the first measurements of polarization [280], but it did not have the sensitivity to see the weak imprint of gravitational waves, which appears in the B -component of the polarization, the part that is divergence-free on the whole sky. The best limits on the B -component so far (early 2008) have been made by the QUaD¹³ detector [300], a cryogenic detector that operated for three seasons in Antarctica. These have not yet shown any evidence for gravitational waves. Results are expected soon from the Background Imaging of Cosmic Extragalactic Polarization (BICEP) detector, also in Antarctica [214]. The next satellite to study the microwave background will be Planck, due for launch by the European Space Agency in 2009 [154].

The gravitational waves detectable in the CMB have wavelengths a good fraction of the horizon size at the time of decoupling, and today they have been redshifted to much longer

wavelengths. They are, therefore, much lower frequency than the radiation that would be observed directly by LISA or ground-based detectors, or even by pulsar timing.

15.1.2 Origin of a random background of gravitational waves

Gravitational waves from the Big Bang

Gravitational waves have traveled almost unimpeded through the universe since they were generated. The cosmic microwave background [72] is a picture of the universe at a time 3×10^5 yrs after the Big Bang, and studies of nucleosynthesis [346] (how the primordial hydrogen, helium, deuterium, and lithium were created) reveal conditions in the universe a few minutes after the Big Bang. Gravitational waves, on the other hand, were produced at times earlier than 10^{-24} s after the Big Bang. Observing this background would undoubtedly be one of the most important measurements that gravitational wave astronomy could make. It would provide a test of inflation, and it would have the potential to give information about the fundamental interactions of physics at energies far higher than we can reach with accelerators.

The most well-defined predictions about the energy in the cosmological gravitational wave background come from inflationary models. Inflation is an attractive scenario for the early universe because, among other things, it provides a natural mechanism for producing the initial density perturbations that evolved into galaxies and galaxy clusters as the universe expands. These perturbations start out as quantum fluctuations in the (hypothetical) scalar inflaton field that is responsible for the inflationary expansion of the universe. The fluctuations are parametrically amplified by the expansion [182, 262, 27] and lead to fluctuations in the density of normal matter after inflation ends.

Several strands of evidence - among them the statistical distribution of density perturbations seen in the cosmic microwave background (most recently by WMAP [343]), the present distribution of galaxies [229], and numerical simulations of structure formation in the early universe [286] - are fully consistent with the now-standard model of a universe dominated by dark energy and whose matter density is dominated by some kind of cold (i.e., massive) dark matter particles [350] with density perturbations consistent with those that inflation could have produced.

The scalar inflaton fluctuations are accompanied by tensor quantum fluctuations in the gravitational field that similarly get amplified by inflation and form a random background [26, 27]. Different models of inflation make different predictions about the relative strength of the scalar and tensor components.

Although inflation is in excellent agreement with observation, other mechanisms in the early universe may have led to the additional production of gravitational waves. Defects that arise from symmetry breaking as the presumed early unified interactions separate from one another can lead to cosmic strings [379], which can produce both a continuous observable gravitational wave background [102] and characteristic isolated bursts of gravitational waves [140, 141, 142]. String theory [93, 99] and brane theory [192, 248] may also provide mechanisms for generating observable radiation.

The various models usually predict significantly different spectra for background radiation. Standard inflationary models predict that the spectrum of Ω_{gw} should be nearly flat, independent of frequency, but variants exist that allow a spectrum that rises with frequency (positive spectral index) or falls. Symmetry-breaking and brane model cosmologies can make very different predictions, even leading to narrow spectral features. It is, therefore, important to measure the spectrum at as many frequencies as possible. Limits on power at one frequency (such as at

¹³QUaD stands for QUEST (Q and U Extragalactic Survey Telescope) at DASI (Degree Angular Scale Interferometer).

the very low-frequency end in the cosmic microwave background) do not necessarily predict the power at other frequencies (such as at ground-based frequencies, a factor 10^{20} times higher).

It is even possible that there will be a feature in the spectrum in the observing band of ground-based or space-based detectors. In standard cosmologies, the radiation observable by LISA (1 mHz) had a wavelength comparable to the (then) horizon size at around the time when the temperature of the universe was equal to the electroweak symmetry-breaking energy. If electroweak symmetry breaking led to a first-order phase transition, where density fluctuations occurred on the length scale of the typical symmetry domain size, then it is likely that these density fluctuations produced gravitational waves with wavelengths of the size of the horizon, which would be in the LISA band today [253]. Detection of this radiation would have deep implications for fundamental physics.

The other expected phase transition is the GUTs (Grand Unified Theory) transition, whose energy might have been 10^{13} times higher. Any gravitational radiation from this transition today would then be at a frequency 10^{13} times that from the electroweak transition, i.e., at centimeter wavelengths. This is one motivation for building microwave-based table-top detectors aimed at high frequencies [126]. For this radiation to be observable by standard interferometers, the GUTs transition would have to have an energy 10^7 times smaller than expected, i.e., around 10^9 GeV. We shall have to wait for observations at these frequencies to tell us if it is there!

In addressing the possibility of new physics, observation of gravitational waves in the cosmic microwave background would play a unique role. These waves originated long after nucleosynthesis, at energies where physics is presumably well understood. They would, therefore, normalize the amount of power in the initial tensor perturbations. Then observations at higher frequencies can use this normalization to measure the excess energy due to any exotic effects due to string theory, phase transitions, or other unknown physics [183, 99].

Pulsar timing arrays (see Section 8.1.3) will also be used to search for a CGWB at frequencies of a few nanoHertz. As for the microwave background, the physics of the universe when gravitational waves at these frequencies originated is well understood, so they could be used to normalize the spectrum. If the power at pulsar frequencies and that in the microwave background are not consistent, then this could indicate something about the conditions in the universe before inflation began.

The predicted spectrum from inflation, strings, and symmetry breakings is highly nonthermal. Any thermal radiation produced in the Big Bang (for example, if, hypothetically, there was some kind of equipartition between gravitational degrees of freedom and other fields in the initial data at the singularity, whatever that might mean!) would have been redshifted away to unobservability by the subsequent inflationary expansion. If inflation did not in fact occur, then this radiation today would have a temperature only a little below that of the cosmological microwave background. So far no instrument has been proposed that would be sensitive to this radiation, but its detection would presumably be inconsistent with inflation.

Astrophysical sources of a stochastic background

After galaxy formation, it is possible that many systems arose that have been radiating gravitational waves in the bands observable by pulsar timing, LISA, and ground-based detectors. There are likely to be strong extra-galactic backgrounds in the LISA band from compact binary systems, which would limit searches for a CGWB [325] by LISA, even if the sensitivity were better. At lower frequencies, even down to pulsar timing frequencies, black hole binaries may make the strongest background, while at frequencies above the LISA band (i.e., above 0.1 Hz) the universe should be relatively free of serious backgrounds [158, 365].

In the LISA band our galaxy is a strong source of backgrounds [190]. This presents a serious confusion noise in searching for other sources at frequencies below 1 mHz. It should be possible

to distinguish this from a CGWB by its intrinsic anisotropy [366].

15.1.3 Cosmography: gravitational wave measurements of cosmological parameters

Since inspiral signals are standard candles [329], as described in Section 6, observations of massive black hole coalescences at cosmological distances by space-based detectors can facilitate an accurate determination of the distance to the source. Our earlier expressions for the chirp waveform can be generalized to the cosmological case (a source at redshift z) by multiplying all masses by $1 + z$ and by replacing the physical distance D by the cosmological luminosity distance D_L [226]. If the wave amplitude, frequency, and chirp rate of the binary can be measured, then its luminosity distance can be inferred. It is not, however, possible to infer the redshift z from the observed signal: the scale-invariance of black hole solutions means that a signal with a redshift of two and a chirp mass $\mathcal{M}$ looks identical to a signal with no redshift and a chirp mass of $\mathcal{M}/3$. To use these distance measures for cosmography, one has to obtain redshifts of the host galaxies.

Before considering how this might be done, we should ask about the accuracy with which the distance can be measured. The relative error in the distance is dominated by the relative error in the measurement of the intrinsic amplitude of the gravitational wave, because the masses will normally be much more accurately measured (by fitting the evolving phase of the signal) than the amplitude. Several factors contribute to the amplitude uncertainty:

- **Signal-to-noise ratio.** The intrinsic measurement uncertainty in the amplitude of the detector's response is simply the inverse of the SNR. Since LISA can have an SNR of several thousand when it observes an SMBH coalescence at high redshift, LISA has great potential for cosmography.
- **Position error.** From the detector response one must infer the intrinsic amplitude of the wave, which means projecting it on the antenna pattern of the detectors. This requires a knowledge of the source position, and this will be potentially a bigger source of uncertainty because the sensitivity of LISA depends on the location of the source in its antenna pattern. Recent work [232, 58] has shown that LISA may be able to achieve position accuracies between one and ten arcminutes. At, say, three arcminutes error, the amplitude uncertainty will be of order 0.1%. This error can be reduced to the SNR-limited error if the source can be identified. Although the coalescence of two SMBHs itself may not have an immediate effect on the visible light from a galaxy, the host galaxy might be identifiable either because it shows great irregularity (mergers of black holes follow from mergers of galaxies) or because some years after the merger an X-ray source turns on (accretion will be disrupted by the tidal forces of the orbiting black holes, but will start again after they merge) [257]. Other effects that might lead to an identification include evidence that stars have been expelled from the core of a galaxy, fossil radio jets going in more than two directions from a common center, and evidence for accretion having stopped in the recent past.
- **Microlensing.** If the source is at a redshift larger than one, as we can expect for LISA, then random microlensing can produce a magnification or demagnification on the order of a few percent [195, 131]. The measured intrinsic amplitude then does not match the amplitude that the signal would have in an ideal smooth cosmology.

The relatively small error boxes within which the LISA coalescences can be localized are promising for identifications, especially if the X-ray indicators mentioned above pick out the

host in the error box. These factors and their impact on cosmography measurements have been examined in detail by Holz and Hughes [195], who coined the term "standard siren" for the chirp sources whose distance can be determined by gravitational wave measurements. The potential for cosmographic measurements by advanced ground-based detectors have been considered in a further paper by the same authors and collaborators [131]. Nearby coalescences and IMRIs should provide an accurate determination of the Hubble Constant [204, 250]. Perhaps the most interesting measurement will be to characterize the evolution of the dark energy, which is usually characterized by inserting a parameter w in the equation of state of dark energy, $p = wp$. If $w = -1$, then the dark energy is equivalent to a cosmological constant [108] and the energy density will be the same at all epochs. If $w > -1$, the dark energy is an evolving field whose energy density diminishes in time. According to [131], gravitational wave measurements have the potential to measure w to an accuracy better than 10% (for advanced ground-based detectors) and around 4% (for LISA). The accuracy with which parameters can be measured improves greatly when one includes in the computation of the covariance matrix the harmonics of the binary inspiral signal that is normally neglected [374]. Arun et al. [48] have shown that the source location in the sky can be greatly improved when the signal harmonics (up to fifth harmonic) are included, which further helps in measuring the parameter w even better.

15.2 О других шумах и факторах

15.2.1 Шумы от вакуумных эффектов (???)

(использую теорию из записи про силы и вакуум, там это подробно обсуждается, тут про грав волны напишу следствия. теорию отсюда удалил, потому что она хоть и крутая, она не про грав волны.)

15.3 Более редкие случаи

15.3.1 Волны в искривленном 4-пространстве

Обсудим подробно волны в искривленном пространстве-времени. В этом разделе все тензорные операции производятся с помощью негалилеевой метрики $g_{ik}^{(0)}$.

(интересный вопрос, но пока не актуальный, так что пусть повесит, пока не сделаю я важные темы.)

Основные формулы и свойства волн в искривленном пространстве

Формулы (??)

(выпишу запросто из лл2, пока не до этого просто)

В итоге от поля Земли какое влияние на плоские волны?

(важное указание)

Поправки к тензору кривизны в случае негалилеевой метрики

Рассмотрим распространение малых возмущений h_{ik} по отношению к негалилеевой «невозмущенной» метрике $g_{ik}^{(0)}$:

$$g_{ik} = g_{ik}^{(0)} + h_{ik}.$$

Поправка первого порядка к символам Кристоффеля выражается через поправки h_{ik} согласно

$$\Gamma_{kl}^{i(1)} = \frac{1}{2} (h_{k;l}^i + h_{l;k}^i - h_{kl}^{;i}),$$

где “;” - ковариантная производная от негалилеевой метрики.

(в этом можно убедиться прямым вычислением (???) потом просчитаю.)

Для поправок к тензору кривизны получается: (??)

$$R_{klm}^{i(1)} = \frac{1}{2} (h_{k;m;l}^i + h_{m;k;l}^i - h_{km;l}^{;i} - h_{k;l;m}^i - h_{l;k;m}^i + h_{kl;m}^{;i}).$$

Отсюда поправки к тензору Риччи:

$$R_{ik}^{(1)} = R_{ilk}^{l(1)} = \frac{1}{2} (h_{i;k;l}^l + h_{ki;l}^l - h_{ik;l}^{;l} - h_{;i;k}^l)$$

Поправки же к смешанным компонентам тензора Риччи получаются из соотношения

$$R_i^{k(0)} + R_i^{k(1)} = (R_{il}^{(0)} + R_{il}^{(1)}) (g^{kl(0)} - h^{kl}),$$

откуда

$$R_i^{k(1)} = g^{kl(0)} R_{il}^{(1)} - h^{ki} R_{il}^{(0)}.$$

Уравнение гравитационной волны в искривленном пространстве-времени

Точная метрика в пустоте должна удовлетворять точным уравнениям Эйнштейна $R_{ik} = 0$. Поскольку невозмущенная метрика $g_{ik}^{(0)}$ удовлетворяет уравнениям $R_{ik}^{(0)} = 0$, то для возмущения получается уравнение $R_{ik}^{(1)} = 0$, т.е.

$$h_{i;k;l}^l + h_{k;i;l}^l - h_{ik;l}^{;l} - h_{;i;k}^l = 0$$

В общем случае произвольных гравитационных волн упрощение этого уравнения до формы, подобной $\square h_i^k = 0$, невозможно. Это можно, сделать в случае волн большой частоты, т.е. когда длина волны $\lambda \ll L$ и период колебаний $\lambda/c \ll L/c$, то есть малы по сравнению с характерными расстояниями L и характерными временами L/c , на которых меняется «фоновое поле».

Каждое дифференцирование компонент h_{ik} увеличивает тогда порядок величины в отношении L/λ по сравнению с производными от невозмущенной метрики $g_{ik}^{(0)}$.

Если ограничиться точностью лишь до членов двух наибольших порядков $((L/\lambda)^2$ и (L/λ)), то в $R_i^{k(1)}$ можно менять порядок дифференцирований, ведь разность

$$h_{i;k;l}^l - h_{i;l;k}^l \approx h_m^l R_{ikl}^{m(0)} - h_i^m R_{mkl}^{l(0)}$$

имеет порядок $(L/\lambda)^0$, между тем как каждое из выражений $h_{i;k}^l$ и $h_{i,l;k}^l$ содержит члены обоих больших порядков.

Наложив теперь на h_{ik} дополнительные условия с помощью функции $\psi_i^k = h_i^k - \frac{1}{2}\delta_i^k h$:

$$\psi_{i;k}^k = 0,$$

аналогичные плоскому случаю $\frac{\partial \psi_i^k}{\partial x^k} = 0$, получим уравнение

$$h_{ik;l}^{;l} = 0$$

обобщающее уравнение $\square h_i^k = 0$.

Условие $h_{ik};_{;l}^l$ также не фиксирует однозначный выбор координат.

(? потом поясню, почему)

Последние можно еще подвергнуть преобразованию $x'^i = x^i + \xi^i$, где малые величины ξ^i удовлетворяют уравнению $\xi_{;k}^{i;k} = 0$.

Этими преобразованиями можно, в частности, воспользоваться для того, чтобы наложить на h_{ik} также и условие $h \equiv h_i^i = 0$.

Тогда $\psi_i^k = h_i^k$, так что h_i^k подчинены условиям

$$h_{i;k}^k = 0, \quad h = 0$$

Круг все еще допустимых преобразований суживается после этого требованием $\xi^i_{;i} = 0$.

Псевдотензор Ландау-Лифшица волны в искривленном пространстве

Псевдотензор t^{ik} находится просто с помощью усреднения.

(???)

Псевдотензор t^{ik} содержит, вообще говоря, наряду с невозмущенной частью $t^{ik(0)}$ также и члены различных порядков по h_{ik} .

Мы придем к выражению, аналогичному $t^{ik} = \frac{c^4}{32\pi k} h_q^{n,i} h_n^{q,k}$, если рассмотрим величины t^{ik} , усредненные по участкам 4-пространства с размерами, большими по сравнению с λ , но малыми по сравнению с L .

Такое усреднение (обозначаемое ниже угловыми скобками $\langle \dots \rangle$) не затрагивает $g_{ik}^{(0)}$ и обращает в нуль все члены, линейные по быстро осциллирующим величинам h_{ik} .

Из квадратичных же членов сохраним лишь члены наиболее высокого (второго) порядка по $1/\lambda$; это - члены, квадратичные по производным $h_{ik,l} \equiv \partial h_{ik}/\partial x^l$.

При такой степени точности все члены в t^{ik} , представляющие собой 4-дивергенции, могут быть опущены.

Действительно, интегралы от таких выражений по области 4-пространства (области усреднения) преобразуются согласно теореме Гаусса, в результате чего их порядок величины по $1/\lambda$ уменьшается на единицу.

Кроме того, выпадают члены, обращающиеся в нуль в силу (108.7) и (108.8) после интегрирования по частям.

Так, интегрируя по частям и опуская интеграл от 4-дивергенции, находим:

$$\begin{aligned} \langle h^{ln} h_{l,n}^p \rangle &= - \langle h^{ln} h_{l,p,n}^p \rangle = 0 \\ \langle h^{il}_{,n} h_l^{k,n} \rangle &= - \langle h^{il} h_{l,n}^{k,n} \rangle = 0 \end{aligned}$$

В результате из всех членов второго порядка остается лишь

$$\langle t^{ik(2)} \rangle = \frac{c^4}{32\pi k} \langle h_q^{n,i} h_n^{q,k} \rangle$$

Отметим, что при этом, с той же точностью, $\langle t_i^{i(2)} \rangle = 0$.

Гравитационное поле гравитационной волны (!)

Обладая определенной энергией, гравитационная волна сама является источником некоторого дополнительного гравитационного поля.

Вместе с создающей его энергией это поле - эффект второго порядка по величинам h_{ik} .

Но в случае высокочастотных гравитационных волн эффект существенно усиливается: тот факт, что псевдотензор t^{ik} квадратичен по производным от h_{ik} , привносит в его порядок величины большой множитель λ^{-2} .

В таком случае можно сказать, что сами волны создают фоновое поле, на котором они распространяются.

Это поле целесообразно рассматривать, проводя описанное выше усреднение по участкам 4-пространства с размерами, большими по сравнению с λ .

Такое усреднение сглаживает коротковолновую «рябь» и оставляет медленно меняющуюся фоновую метрику.

Для вывода уравнения, определяющего эту метрику, надо учесть в разложении тензора R_{ik} члены не только линейные, но и квадратичные по h_{ik} : $R_{ik} = R_{ik}^{(0)} + R_{ik}^{(1)} + R_{ik}^{(2)}$.

Как уже указывалось, усреднение не затрагивает членов нулевого порядка.

Таким образом, усредненные уравнения поля $\langle R_{ik} \rangle = 0$ принимают вид

$$R_{ik}^{(0)} = - \langle R_{ik}^{(2)} \rangle$$

причем в $R_{ik}^{(2)}$ надо сохранить лишь члены второго порядка по $1/\lambda$.

Их легко найти из тождества (96.7).

Квадратичные по h_{ik} члены, возникающие из правой части этого тождества, имеющей вид 4-дивергенции, исчезают (с рассматриваемой точностью) при усреднении и, таким образом, остается

$$\left\langle \left(R^{ik} - \frac{1}{2} g^{ik} R \right)^{(2)} \right\rangle = - \frac{8\pi k}{c^4} \langle t^{ik(2)} \rangle$$

или, поскольку $\langle t_i^{(2)} \rangle = 0$, с той же точностью:

$$\langle R_{ik}^{(2)} \rangle = - \frac{8\pi k}{c^4} \langle t_{ik}^{(2)} \rangle$$

Наконец, используя (108.9), получим окончательно уравнение (108.10) в виде

$$R_{ik}^{(0)} = \frac{1}{4} \langle h_{q,i}^n h_{n,k}^q \rangle$$

Если «фон» создается целиком самими волнами, уравнения (108.11) и (108.7) должны решаться совместно.

Оценка выражений в обеих сторонах уравнения (108.11) показывает, что в этом случае радиус кривизны фоновой метрики по своему порядку величины L связан с длиной волны λ и порядком величины ее поля h согласно $L^{-2} \sim h^2 \lambda^2$, т.е. $\lambda/L \sim h$.

О другом про волны в искривленном пространстве

Обзор применений волн в искривленном пространстве

(приведу критерии, когда это нужно, а когда нет? в поле Земли вроде бы слишком слабое искривление, так что не нужно. тут четко укажу критерии. Мб еще когда-то есть польза от этой теории, пока не знаю.)

(пока считаю, что это никогда не нужно, так что и не занимаюсь этим.)

Количественная оценка влияния искривленности пространства на гравитационные волны

хз, какая она.

к вопросу о влиянии гравитационного поля земли.

хз вообще.

Гравитационные волны в гравитационном поле Земли

(решу позже эту модель)

Вид гравитационной волны вблизи источников

(как там они выглядят. какая их формула? они ведь не малые там? пока представляю картинку как для ЭМ волн.)

Отличие волн в плоском пространстве от волн в искривленном пространстве

(в дополнение к теме укажу явное отличие(??))

15.4 Другие модели и вопросы

(то, что просто есть, но чего еще не скоро я коснусь.)

15.4.1 Другие вопросы для иллюстрации грав. волн

можно придумать разные модели, которые интересно было бы количественно посчитать.

некоторые задачи, которые я придумал, заготовлю тут.

Гравитационные волны от махания кулаком

(порешаю Лейтамана Прайса, давно про эту задачу думал уже)

время, через которое гравитационные волны унесут всю энергию двойной системы?

хз

15.4.2 Другие идейные вопросы о грав. волнах

можно придумать разные модели, которые интересно было бы количественно посчитать.

некоторые задачи, которые я придумал, заготовлю тут.

откуда эта уносимая энергия?

типа масса что ли меняется этих объектов?

гравитационные волны от тройной системы

я вообще хз, как там тройная будет излучать.

15.4.3 Гравитационные антенны

(хз, что с этим делать, просто слышал, что такое есть)

Основные формулы

(хз)

Когда встречаются?

хз

15.4.4 Рассеяние гравитационных волн

(мб когда-то подумаю про это, пока не думал, но вообще, такое ведь может быть, да?)

15.4.5 Дифракция гравитационной волной

(такая идея есть, хз, потом пропишу мб)

Модель

О случаях применения

(мб как детектирование грав волн, но пока не уверен, что это нужно)

Принцип Гюйгенса-Френеля при наличии гравитационной волны

(мб в другое место этот п-ф перенесу)

(вроде тоже почти понимаю)

Part VI

Adds

A Введение

A.1 Еще мотивация

Обсудим всю мотивацию, которая позволит нам познать предмет без проблем.

A.1.1 Множество приложений

Укажем приложения
(потом раскрою)

A.1.2 О пользе в плане понимании природы и саморазвитии

(потом раскрою)

A.1.3 Удивительные факты

(потом раскрою)

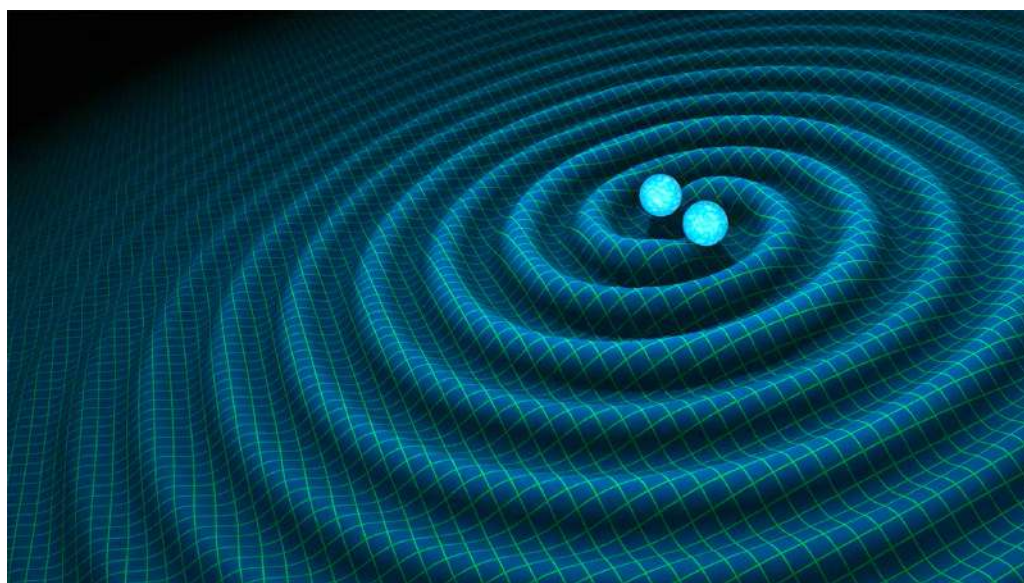
A.2 Обзор гравитационных волнах

Приведем общий обзор гравитационных волн для кругозора.

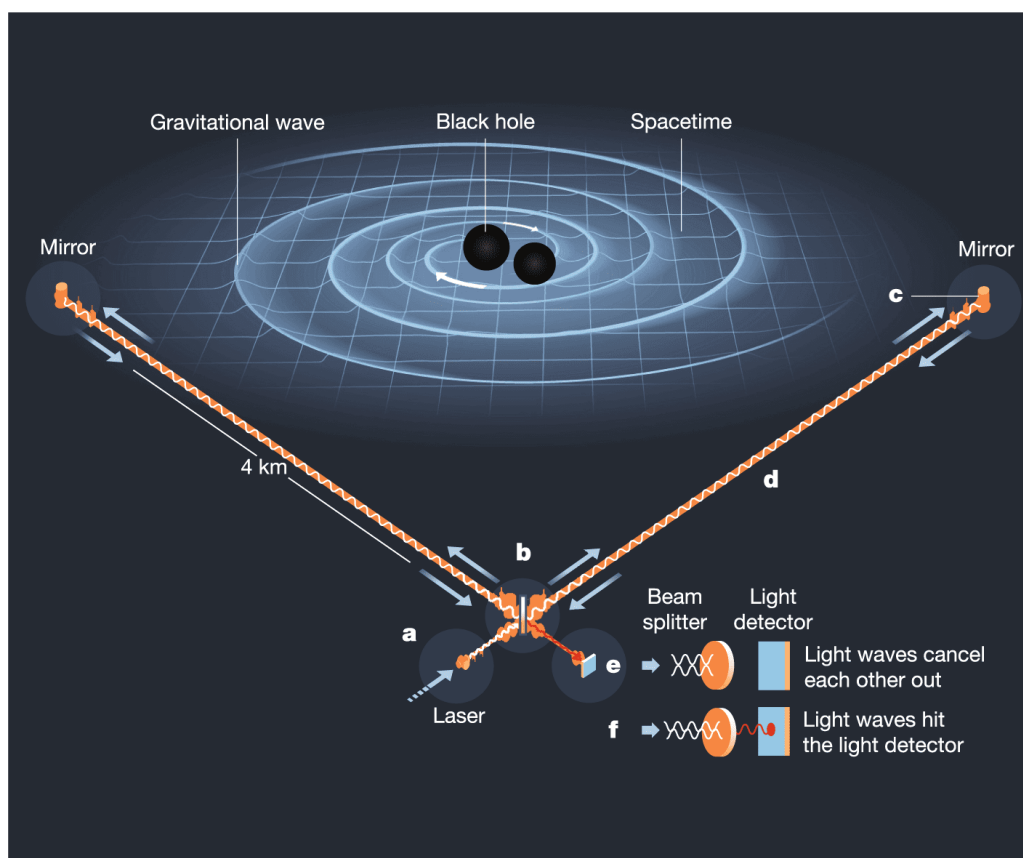
A.2.1 Популярно о гравитационных волнах

Объяснение гравитационных волн на пальцах (???)

(не знаю, потом напишу, не актуально.)



Популярно о детектировании гравитационных волн



А.3 Мышление профессионала в гравитационных волнах

А.3.1 Отношение к гравитационным волнам

Так все-таки есть ли в волнах перспективы? (???)

(ответу потом. первые несколько лет могут пройти в духе "давай просто изучай и все, другие люди над смыслом этого будут думать")

Стоит ли изучать малополезные способы детектирования волн?

(я думаю, что нет, потому что слишком маловероятно, что они будут нужны. при этом наверняка сотни людей именно сейчас (на 2023 год) это и изучают. Не знаю, потом отвечу.)

А.3.2 Особенности изучения гравитационных волн

Обсудим особенности эффективного изучения свойств гравитационных волн и физики в контексте их. В целом, все как и всегда, просто неплохо бы понимать небольшие указания.

(сам не занимаюсь ими, потому что нет места, где бы мне платили за это)

1/2) Убедись, что в этом есть у тебя будущее, а что тебя не окружают неавторитеты и по сути тебе просто сказали фигней заниматься умными словами.

Если это второй случай, что крайне вероятно в 2020х годах и скорее всего долгое время так и было и будет, то забей и найди другое место, если есть возможность. Seriously. Самому это не стоит того тянуть, особенно учитывая, что есть в других вузах (скорее всего в других странах) группы умных людей, которые все вдоль и поперек проходят, которым платят много. Конечно, можно взять цель на переход в эту новую тусовку, но это отдельное направление жизни. Так вот случайно такого направления не будет.

Всякие не очень умные люди (статьи не считаются) что пробуют их понять - окей, круто, пусть пробуют, но идти за ними нечего!

Такой во идейный совет: в некоторых случаях просто забей на гравитационные волны, это может не стоить того абсолютно! Слишком много заморочек, слишком каждый вопрос много подготовки требует. Сдвинуть это можно, но одному тянуть - ох сколько несчастья принесет, а это очень очень плохо.

Но это совет для людей до аспирантуры, после успешной аспирантуры это будет очевидно.

2/2) Не занимайся разными вопросами, в частности отделяй теорию от экспериментов и моделей для экспериментов!

Не нужно стремиться к полноте, пока нет результатов исследований, в этом предмете это особенно важно понимать. Фокусироваться всегда нужно, но здесь тоже очень полезно это напомнить. Слишком всего много! Каждый вопрос можно как минимум неделю изучать и только по уже написанным статьям, а вообще - можно и по месяцам исследовать.

Так что разово создаешь разделы для возникших вопросов и не думаешь про них, пока не нужно будет! Общая согласованная теория всего, что уже (на 2023 год) развито, мб страниц 2000 займет, по ощущениям так пишу. Много статей. Так что вот, не думаем о лишнем. В прочем тому, у кого есть опыт исследования это будет очевидно сразу, но зато тому, у кого нет опыта исследования, один этот малый параграф сэкономит прямо, может быть, полгода.

Так что по итогу многие годы активной работы в гравитационных волнах скорее всего многие разделы будут очень не сделанными и близко. Зато будут результаты.

Что еще полезно делать для понимания гравитационных волн?

Есть очень много популярных и обзорных статей, с хорошее их усвоение очень полезно включить в первый этап изучения

Очень много где все написано четко - вот по сути и основа всей структуры, там и собраны основные формулы. Так что просто прочитываешь - и выписываешь, вот и все, в чем вообще проблема? И все, ты готов к изучению.

О предметах, очень полезных для изучения гравитационных волн

О важности понимания основ механики сплошных сред и теории упругости

(про то, что это часто может использоваться для приложений, так что важно понимать суть тех методов. потом напишу, очень очень часто эти концепции встречаются!)

А.3.3 Особенности изучения детектирования волн

Обсудим особенности эффективного изучения методов детектирования гравитационных волн.

Параллельно двигаем необходимые предметы (МСС, теория упругости, т.п.), которые по умолчанию не связаны с волнами, но тут они очень нужны, но ни в коем случае в них не останавливаемся отдельно много!

Потому что без их понимания - не понять методы детектирования, но в то же время если усесться - то минус месяцы на то, что не особо нужно.

А.3.4 Как догадаться до всех главных идей самому?

(потом раскрою, еще я не профессионал, а вопрос этот самый профессиональный)

А.4 Литература

А.4.1 Основная

Основная обзорная

Еще раз напоминаю, что обзорная - круто, но ее по ходу дела дочитывать лучше всего. Отдельно в обзоры погружаться - трата времени.

[1] Sathyaprakash, B. S. etc. Physics, Astrophysics and Cosmology with Gravitational Waves

Хороший обзор гравитационных волн. Много именно из него выгружено, многое что нужно, там и написано.

[2] Peter R. Saulson Gravitational wave detection: Principles and practice

Хороший простой обзор, много хороших формул и картинок.

[3] Medvedev, Alexander S. and Yiğit, Erdal Gravity Waves in Planetary Atmospheres

статья про влияние гравитационных волн на атмосферу, говорят, типичный эффект такой есть. мб почитаю, пока абсолютно это не важно. . вообще, аналогичных статей тоже много, пока я не знаю, что сказать и мне это не актуально

Основная о свойствах волн и их детектировании (??)

(пока не выбрал)

А.4.2 Профессиональная и углубленная литература

Основная профессиональная литература

[4] Maggiore, Michele Gravitational Waves. Vol. 1: Theory and Experiments

800 страниц чего только не написано про гравитационные волны, можно просто 1 год их изучать по ней.

[5] Maggiore, Michele Gravitational Waves. Vol. 2: Astrophysics and Cosmology

Еще 800 страниц большой теории про космологию и гравитационные волны, только на нее 1 год уйдет изучать.

[6] Sanjay Kumar Sahay Data Analysis of Gravitational Waves

120 страниц тоже содержательной, большой теории, так что тоже минимум 2 месяца на эту книгу ушло бы. Нужно будет анализировать данные - открою ее.

А.4.3 Статьи о детектировании гравитационных волн и шумах

Статьи о шумах

Gregory M. Harry et. al. Thermal noise in interferometric gravitational wave detectors due to dielectric optical coatings

Вроде очень известная статья про шумы, очень крутой выглядит.

[7] Reid, Stuart and Martin, Iain W. Development of Mirror Coatings for Gravitational Wave Detectors

Просто обзор покрытий, мб она полезная и говорит, что такие-то есть шумы и такие-то есть способы покрыть зеркало, чтобы их уменьшить, но я не особо вчитывался, просто там ссылки есть, в самой ней не особо много полезного увидел.

[8] Utina, A et. al. ETPATHFINDER: a cryogenic testbed for interferometric gravitational-wave detectors

Обзор шумов, ссылки на их описание, полезная, но мб есть статьи полезнее, я не смотрел просто особо много про это еще.

[9] Harry, Gregory M et. al. Thermal noise in interferometric gravitational wave detectors due to dielectric optical coatings

Конкретно решена задача про спектр от термо шума, можно подтянуть статфиз и теорию упругости, разбираясь в этом. Еще про материалы какие-то заключения есть.

[10] Abbott, B P, et. al. A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals

Статья с более 1000 авторов про шумы в гравитационных волнах. Вроде именно по этой статье с шумами и работают, мб она объясняет итоговые методы работы. Не особо вникал.

[11] Gravitational-wave sensitivity curves Moore, C J and Cole, R H and Berry, C P L

Еще одна цитируемая статья о шумах, может, там наконец-то объяснено качественно, почему такие графики шумов? Вроде формул не мало.

[12] Caves, Carlton M. Quantum-mechanical noise in an interferometer

Первая статья про квантовый шум, там глубокие есть идеи про его природу, я их пока вообще не понимаю.

О различных способах детектирования

(потом мб отдельный раздел создам, пока не так много читал статей, так что пишу тут.)

Thomas B. Mieling The Response of Optical Fibres to Gravitational Waves

Та самая статья, где решена моя научка, посмотрел, но тут отдельно нужно разбираться, минимум неделю, ибо очень особый подход для задачи. Так что пока что оставил, будет нужно - выучу, пропишу в структуру. Abstract The response of optical fibre modes to plane gravitational waves of low frequency is computed. By solving perturbatively the Maxwell equations for step-index optical fibres in a gravitational wave background and implementing appropriate boundary conditions to describe single-mode fibres, explicit formulae for the perturbations of the phase and the polarisation of the fibre modes are obtained.

Reginald T. Cahill Optical-Fiber Gravitational Wave Detector: Dynamical 3-Space Turbulence Detected

Максимально стремная статья, что оказывается можно так вот детектировать их. Не могу разобраться в этом пока что, потом мб посмотрю еще раз, подумаю про это. Abstract: Preliminary results from an optical-fiber gravitational wave interferometric detector are reported. The detector is very small, cheap and simple to build and operate. It is assembled

from readily available opto-electronic components. A parts list is given. The detector can operate in two modes: one in which only instrument noise is detected, and data from a 24 hour period is reported for this mode, and in a 2nd mode in which the gravitational waves are detected as well, and data from a 24 hour period is analysed. Comparison shows that the instrument has a high S/N ratio. The frequency spectrum of the gravitational waves shows a pink noise spectrum, from 0 to 0.1Hz.

А.4.4 Статьи о разных эффектах и теоретических моделях

О другой теории

Dynamical 3-Space: Gravitational Wave Detection and the Shnoll Effect David P. Rothall and Reginald T. Cahill

(??? я хз, не до этой статьи, пока более проверенными статьями заниматься буду)

Аннотация:

Shnoll has investigated the non-Poisson scatter of rate measurements in various phenomena such as biological and chemical reactions, radioactive decay, photodiode current leakage and germanium semiconductor noise, and attributed the scatter to cosmophysical factors.

While Shnoll didn't pinpoint the nature of the cosmophysical factors the Process Physics model of reality leads to a description of space, which is dynamic and fractal and exhibits reverberation effects, and which offers an explanation for the scattering anomaly. The work presented here shows a new way of generating the effects Shnoll discovered, through studying the phase difference of RF EM waves travelling through a dual coaxial cable Gravitational Wave Detector experiment

Об эффектах электродинамики в гравитации

Статьи про разные другие эффекты

(тут то, что пусть и прикольно, но я не планирую это учить, пока мне не будут платить.)

[13] G. B. Lesovik, A. V. Lebedev, V. Mounutcharyan, T. Martin Detection of gravity waves by phase modulation of the light from a distant star

Модель отклонения света и погрешности детектирования. 1 неделя ушла бы ее изучить, но абсолютно всегда не до нее, не на таком я уровне, чтобы развлекаться за чтением ее.

Статьи о гравитационных волнах и космологии

Статьи о теоретических методах

А.4.5 Дополнительная литература

Другая обзорная

[?] Colpi, Monica etc. Gravitational Wave Sources in the Era of Multi-Band Gravitational Wave Astronomy

Где-то в статье была ссылка на эту статью (книгу?). Что-то я не могу найти эту книгу, есть только платная и уверен, что она не стоит того.

Parivesh Choudhary Role of Modern Optical Techniques in Gravitational Wave Detection

небольшой обзор с оптикой и красивыми картинками, мб это полезное, но ноль цитат и не авторитетный автор.

Статьи о результатах детектирований волн и анализе астрономических объектов из этих данных

[14] Karl Wette Searches for continuous gravitational waves from neutron stars: A twenty-year retrospective

Возможно, хорошая статья, там таблица есть, типа от такого-то тела такие-то волны. Может, нейтронные звезды обсуждены? Все равно не актуально про это думать.

[15] Abbott, B. P. et. al. Observation of Gravitational Waves from a Binary Black Hole Merger

Первая статья про детектирование гравитационных волн. Таких статей навалом, из нее схожие графики где только ни увидеть, но эта была первой.

А.4.6 Литература в помощь

По физическим основам

(тут из ОТО потом впишу книги, Тонг, Ландау, Вайнберг, мб другие. вообще, многие могут быть очень полезны тут.)

А.5 Описание записи

А.5.1 Общие особенности записи

А.5.2 Обозначения и константы

А.6 Головоломки на разные случаи жизни

(потом напишу, пока не актуально, итак тут 1000 вопросов можно разных с ходу придумать.)

А.6.1 Бытовые головоломки

А.6.2 Принципиальные головоломки

А.6.3 Головоломки о деталях

А.7 Короткий исторический обзор

Обсудим вкратце историю развития...

А.7.1 О первых открытиях

(то, что мало кто знает и особо это не нужно.)

А.7.2 О результатах конца 20го и начала 21го веков

(то, что все знают.)

В Некоторые концепции математики

(тоже скорее всего разное может становиться нужно.)

В.0.1 О методах диффема для грав волн (!!!)

(очень важные указания, чтобы не делать много всего, а делать ровно то, что нужно для волн! потренируюсь еще на этом!)

С Некоторые концепции физики

(тут в основном о гравитации, мб чуть о теории поля)

С.1 О теориях в помощь и других теоретических моделях

С.1.1 Об электродинамике в искривленном пространстве для грав волн

(чисто в этом контексте и укажу это!)

Bibliography

- [1] Sathyaprakash, B. S. and Schutz, Bernard F.: *Physics, astrophysics and cosmology with gravitational waves*. Living Reviews in Relativity, 12(1), Mar 2009, ISSN 1433-8351. <http://dx.doi.org/10.12942/lrr-2009-2>.
- [2] Saulson, Peter R.: *Gravitational wave detection: Principles and practice*. Comptes Rendus Physique, 14(4):288–305, 2013, ISSN 1631-0705. <https://www.sciencedirect.com/science/article/pii/S163107051300008X>, Gravitational waves / Ondes gravitationnelles.
- [3] Medvedev, Alexander S. and Yiğit, Erdal: *Gravity waves in planetary atmospheres: Their effects and parameterization in global circulation models*. Atmosphere, 10(9), 2019, ISSN 2073-4433. <https://www.mdpi.com/2073-4433/10/9/531>.
- [4] Maggiore, Michele: *Gravitational Waves. Vol. 1: Theory and Experiments*. Oxford University Press, 2007, ISBN 978-0-19-171766-6, 978-0-19-852074-0.
- [5] Maggiore, Michele: *Gravitational Waves. Vol. 2: Astrophysics and Cosmology*. Oxford University Press, March 2018, ISBN 978-0-19-857089-9.
- [6] Sahay, Sanjay Kumar: *Data Analysis of Gravitational Waves*. Svenska fysikarkivet, 2008, ISBN 978-91-85917-05-1.
- [7] Reid, Stuart and Martin, Iain W.: *Development of mirror coatings for gravitational wave detectors*. Coatings, 6(4), 2016, ISSN 2079-6412. <https://www.mdpi.com/2079-6412/6/4/61>.
- [8] Utina A, et. al: *Etpathfinder: a cryogenic testbed for interferometric gravitational-wave detectors*. Classical and Quantum Gravity, 39(21):215008, September 2022, ISSN 1361-6382. <http://dx.doi.org/10.1088/1361-6382/ac8fdb>.
- [9] Harry, Gregory M, Gretarsson, Andri M, Saulson, Peter R, Kittelberger, Scott E, Penn, Steven D, et al.: *Thermal noise in interferometric gravitational wave detectors due to dielectric optical coatings*. Classical and Quantum Gravity, 19(5):897–917, February 2002, ISSN 1361-6382. <http://dx.doi.org/10.1088/0264-9381/19/5/305>.
- [10] Abbott, B P, Abbott, R, Abbott, T D, Abraham, S, Acernese, F, Ackley, K, and Adams, C: *A guide to ligo–virgo detector noise and extraction of transient gravitational-wave signals*. Classical and Quantum Gravity, 37(5):055002, February 2020, ISSN 1361-6382. <http://dx.doi.org/10.1088/1361-6382/ab685e>.
- [11] Moore, C J, Cole, R H, and Berry, C P L: *Gravitational-wave sensitivity curves*. Classical and Quantum Gravity, 32(1):015014, December 2014, ISSN 1361-6382. <http://dx.doi.org/10.1088/0264-9381/32/1/015014>.
- [12] Caves, Carlton M.: *Quantum-mechanical noise in an interferometer*. Phys. Rev. D, 23:1693–1708, Apr 1981. <https://link.aps.org/doi/10.1103/PhysRevD.23.1693>.
- [13] Lesovik, G. B., Lebedev, A. V., Mounutcharyan, V., and Martin, T.: *Detection of gravity waves by phase modulation of the light from a distant star*. Physical Review D, 71(12), Jun 2005, ISSN 1550-2368. <http://dx.doi.org/10.1103/PhysRevD.71.122001>.

- [14] Wette, Karl: *Searches for continuous gravitational waves from neutron stars: A twenty-year retrospective*. Astroparticle Physics, 153:102880, 2023, ISSN 0927-6505. <https://www.sciencedirect.com/science/article/pii/S092765052300066X>.
- [15] Abbott, B. P. et. al.: *Observation of gravitational waves from a binary black hole merger*. Phys. Rev. Lett., 116:061102, Feb 2016. <https://link.aps.org/doi/10.1103/PhysRevLett.116.061102>.
- [16] Брагинский, В. Б. and Халили, Ф. Я.: *Взаимодействие электромагнитных и механических колебаний на уровне основного квантового состояния*. Усп. физ. наук, 163(6):107–109, 1993. <https://ufn.ru/ru/articles/1993/6/e/>.
- [17] Coleman, Sidney and De Luccia, Frank: *Gravitational effects on and of vacuum decay*. Phys. Rev. D, 21:3305–3315, Jun 1980. <https://link.aps.org/doi/10.1103/PhysRevD.21.3305>.
- [18] Briet, Jop and Hobill, David: *Determining the dimensionality of spacetime by gravitational lensing*, 2008.
- [19] Nacyan, Shahan: *Effects of gravitational waves on the polarization of pulsars*. International Journal of Modern Physics A, 31(02n03):1641023, Jan 2016, ISSN 1793-656X. <http://dx.doi.org/10.1142/S0217751X16410232>.
- [20] Лайтман А., Пресс В., Прайс Р. Тюкольски С.: *Сборник задач по теории относительности и гравитации*. " Мир".
- [21] Б., Хрипович И.: *Общая теория относительности*. " Регулярная и хаотическая динамика".
- [22] Б., Хрипович И.: *Общая теория относительности*.
- [23] Tong, David: *General relativity*.
- [24] Мизнер, Ч., Торн, К., and Уилер, Дж.: *Гравитация ч. 1*. Рипол Классик.
- [25] Мизнер, Ч., Торн, К., and Уилер, Дж.: *Гравитация ч. 2*. Рипол Классик.
- [26] Мизнер, Ч., Торн, К., and Уилер, Дж.: *Гравитация ч. 3*. Рипол Классик.
- [27] Cabral, Francisco and Lobo, Francisco S. N.: *Electrodynamics and spacetime geometry: Foundations*. Foundations of Physics, 47(2):208–228, Nov 2016, ISSN 1572-9516. <http://dx.doi.org/10.1007/s10701-016-0051-6>.
- [28] Cabral, Francisco and Lobo, Francisco S. N.: *Electrodynamics and spacetime geometry: Astrophysical applications*. The European Physical Journal Plus, 132(7), Jul 2017, ISSN 2190-5444. <http://dx.doi.org/10.1140/epjp/i2017-11618-2>.
- [29] М, Лашкевич Ю: *Лекции по ОТО для студентов 3 курса*.
- [30] Степаньянц, К: *Классическая теория поля*. Litres, 2018.
- [31] Лашкевич: *Лекции по общей теории относительности для 3 курса*. 2018.
- [32] Лифшиц Е. М., Ландау Л. Д.: *Теория поля*. Наука, 2004.