Cosmology and Fundamental Physics in the Era of Gravitational-Wave Astronomy

Jenkins, Alex

Awarding institution:
King's College London

Cosmology and Fundamental Physics in the Era of Gravitational-Wave Astronomy

by
Alexander C. Jenkins

A thesis presented for the degree of
Doctor of Philosophy
at
KING'S COLLEGE LONDON
Department of Physics
27 January 2022

Abstract

The advent of gravitational-wave (GW) astronomy has presented us with a completely new means for observing the Universe, allowing us to probe its structure and evolution like never before. In this thesis, we explore three distinct but complementary avenues for using GW observations to gain new insights into cosmology and fundamental physics.

In chapter 1, we study the astrophysical GW background (AGWB): the cumulative GW signal arising from a large number of compact binary coalescences (CBCs) throughout the Universe. Since these compact binaries reside in galaxies, the AGWB contains *anisotropies* (i.e., intensity fluctuations on the sky) that trace out the large-scale structure of the cosmic matter distribution. Despite their intrinsic interest as a source of novel cosmological information, these anisotropies have been neglected in most studies of the AGWB until quite recently. Applying tools and concepts from other cosmological observables such as galaxy surveys and the cosmic microwave background (CMB), we investigate the angular power spectrum of the AGWB, with the goal of developing predictions that can be confronted with current and future directional AGWB searches. Our key result is a simulated full-sky map of the AGWB anisotropies that we construct using data from the Millennium N-body simulation. We find that these anisotropies are much larger in amplitude than in early-Universe observables such as the CMB, and that the lowest few multipoles of the angular power spectrum are likely to be observed by third-generation GW observatories. We also highlight the issue of *shot noise* due to the relatively low rate of CBCs in our frequency band of interest, and develop an optimal data-analysis strategy for estimating the true angular power spectrum in the presence of this shot noise.

In chapter 2, we investigate the *nonlinear GW memory effect*, a fascinating prediction of general relativity in the dynamical, nonlinear regime, in which essentially all GW emission is accompanied by a hereditary, monotonically-increasing GW strain sourced by the energy of the escaping gravitons. Essentially all of the literature on this effect has focused on the memory signals associated with CBCs; we broaden this scope by calculating, for the first time, the nonlinear memory emitted by cusps and kinks on cosmic string loops, which are among the most promising *cosmological* sources of GWs. Working in the Nambu-Goto approximation, we obtain simple analytical waveforms for the memory emitted by cusps and kinks, as well as the 'memory of the memory' and other higher-order effects. Summing over all of these contributions, we show that, surprisingly, the combined cusp memory signal *diverges* for sufficiently large loops, indicating a breakdown in the validity of the weak-field description of the cusp. We trace this divergence back to the high-frequency behaviour of the original cusp waveform, which gives rise to a trans-Planckian energy flux in the direction of the cusp's motion. We then present one tentative possible solution to this divergence, in which the portion of the string surrounding the cusp collapses to form a primordial black hole (PBH). We investigate the observational predictions of this scenario, and show that these PBHs could act as a 'smoking gun' signature of cosmic strings.

Finally, in chapter 3 we develop a powerful new method for GW detection based on precision measurements of the orbits of binary systems. In the presence of a stochastic GW background (GWB) the trajectories of the binary's components are perturbed, giving rise to a random walk in the system's orbital parameters over time. By searching for this stochastic orbital evolution, we can infer the presence or absence of a GWB, turning the binary into a dynamical GW detector. We develop here a novel Fokker-

Planck formalism for calculating the expected evolution in all six orbital elements. We then apply this formalism to two observational probes: timing of binary millisecond pulsars, and laser ranging of the Moon and artificial satellites. We use a Fisher-forecasting approach to estimate the sensitivity of each of these probes to the GWB, and show that present data are already sensitive enough to place the strongest constraints to date in the μHz frequency band. This band lies between the frequencies probed by pulsar timing arrays and by future space-based interferometers such as LISA, and is therefore an extremely attractive observational target, which could contain numerous cosmological GW signals. As an example, we consider the GWB sourced by a cosmological first-order phase transition (FOPT), and show that the binary resonance searches we propose (in particular, with lunar laser ranging) will be sensitive to a region of the FOPT parameter space that no other current or near-future GW experiment can reach.

The contents of this thesis are the result of my own work, except where specific reference is made to the work of others. This work was originally presented in the publications listed below.

Chapter 1 is based on:

- **ACJ** and M. Sakellariadou, *Anisotropies in the stochastic gravitational-wave background: Formalism and the cosmic string case*, Phys. Rev. D **98** (2018), 063509, arXiv:1802.06046 [astro-ph.CO], [345]

- **ACJ**, M. Sakellariadou, T. Regimbau, and E. Slezak, *Anisotropies in the astrophysical gravitational-wave background: Predictions for the detection of compact binaries by LIGO and Virgo*, Phys. Rev. D **98** (2018), 063501, arXiv:1806.01718 [astro-ph.CO], [349]

- **ACJ**, R. O'Shaughnessy, M. Sakellariadou, and D. Wysocki, *Anisotropies in the astrophysical gravitational-wave background: The impact of black hole distributions*, Phys. Rev. Lett. **122** (2019), 111101, arXiv:1810.13435 [astro-ph.CO], [342]

- **ACJ** and M. Sakellariadou, *Shot noise in the astrophysical gravitational-wave background*, Phys. Rev. D **100** (2019), 063508, arXiv:1902.07719 [astro-ph.CO], [346]

- **ACJ**, J. D. Romano, and M. Sakellariadou, *Estimating the angular power spectrum of the gravitational-wave background in the presence of shot noise*, Phys. Rev. D **100** (2019), 083501, arXiv: 1907.06642 [astro-ph.CO], [344]

Chapter 2 is based on:

- **ACJ** and M. Sakellariadou, *Primordial black holes from cusp collapse on cosmic strings* (2020), arXiv:2006.16249 [astro-ph.CO], [347]

- **ACJ** and M. Sakellariadou, *Nonlinear gravitational-wave memory from cusps and kinks on cosmic strings*, Class. Quant. Grav. **38** (2021), 165004, arXiv:2102.12487 [gr-qc], [348]

Chapter 3 is based on:

- D. Blas and **ACJ**, *Detecting stochastic gravitational waves with binary resonance*, Phys. Rev. D **105** (2022), 064021, arXiv: 2107.04063 [gr-qc], [138]

- D. Blas and **ACJ**, *Bridging the μHz gap in the gravitational-wave landscape with binary resonance*, Phys. Rev. Lett. **128** (2022), 101103, arXiv:2107.04601 [astro-ph.CO], [137]

The following publications, of which I am also an author, are beyond the scope of this thesis:

- **ACJ**, A. G. A. Pithis, and M. Sakellariadou, *Can we detect quantum gravity with compact binary inspirals?*, Phys. Rev. D **98** (2018), 104032, arXiv:1809.06275 [gr-qc], [343]

- P. Auclair, J.-J. Blanco-Pillado, D. G. Figueroa, **ACJ**, M. Lewicki, M. Sakellariadou, S. Sanidas, L. Sousa, D. A. Steer, J. M. Wachter, and S. Kuroyanagi, *Probing the gravitational wave background from cosmic strings with LISA*, JCAP **04** (2020), 034, arXiv:1909.00819 [astro-ph.CO], [87]

- D. Bertacca, A. Ricciardone, N. Bellomo, **ACJ**, S. Matarrese, A. Raccanelli, T. Regimbau, and M. Sakellariadou, *Projection effects on the observed angular spectrum of the astrophysical stochastic gravitational wave background*, Phys. Rev. D **101** (2020), 103513, arXiv:1909.11627 [astro-ph.CO], [117]

- G. Boileau, **ACJ**, M. Sakellariadou, R. Meyer, and N. Christensen, *Ability of LISA to detect a gravitational-wave background of cosmological origin: the cosmic string case*, Phys. Rev. D **105** (2022), 023510, arXiv:2109.06552 [gr-qc], [145]

- N. Bellomo, D. Bertacca, **ACJ**, S. Matarrese, A. Raccanelli, T. Regimbau, A. Ricciardone, and M. Sakellariadou, *CLASS_GWB: robust modeling of the astrophysical gravitational wave background anisotropies*, arXiv:2110.15059 [gr-qc], [111]

- N. Bartolo, D. Bertacca, R. Caldwell, C. R. Contaldi, G. Cusin, V. De Luca, E. Dimastrogiovanni, M. Fasiello, D. G. Figueroa, G. Franciolini, **ACJ**, M. Peloso, M. Pieroni, A. Renzini, A. Ricciardone, A. Riotto, M. Sakellariadou, L. Sorbo, G. Tasinato, J. Torrado, S. Clesse, and S. Kuroyanagi, *Probing Anisotropies of the Stochastic Gravitational Wave Background with LISA*, arXiv:2201.08782 [astro-ph.CO], [100]

- A. Renzini, B. Goncharov, **ACJ**, and P. M. Meyers, *Stochastic Gravitational-Wave Backgrounds: Current Detection Efforts and Future Prospects*, Galaxies **10** (2022), 34, arXiv:2202.00178 [gr-qc], [477]

- We generally use ~ to mean equality to within an order of magnitude, $\approx$ to mean approximate equality, $\simeq$ to mean asymptotic equality (e.g., $e^x \simeq 1 + x$), and $\propto$ to mean proportionality. The symbol $\equiv$ denotes equality by definition.

- We use units such that the speed of light in vacuum and the Boltzmann constant are both equal to unity, $c = k_{\mathrm{B}} = 1$, but keep Newton's constant G and the reduced Planck constant $\hbar$ explicit throughout. Unless otherwise specified, we set the Hubble constant to the *Planck* 2018 value of $H_0 \approx 67.7 \, \mathrm{km \, s^{-1} \, Mpc^{-1}} \approx 2.19 \times 10^{-18} \, \mathrm{Hz}$ [55].

- Spacetime indices running over $(0, 1, 2, 3)$ are indicated by lowercase Greek letters, while Latin letters indicate either spatial indices running over $(1, 2, 3)$ or, where specified, some other set of positive integers $(1, 2, \dots)$. We also use A, A' to denote GW polarisation indices running over $(+, \times)$. Round brackets around pairs of indices indicate a symmetrisation operation, e.g.,

$$X_{(\mu\nu)} \equiv \frac{1}{2}(X_{\mu\nu} + X_{\nu\mu}).$$

- Our conventions for general relativity match those of Misner, Thorne, and Wheeler [429] and Wald [574]. The spacetime metric $g_{\mu\nu}$, which is used to raise and lower indices, has signature $(-+++)$, and defines the Christoffel symbols

$$\Gamma^\mu_{\alpha\beta} \equiv \frac{1}{2}g^{\mu\nu}(\partial_\alpha g_{\beta\nu} + \partial_\beta g_{\alpha\nu} - \partial_\nu g_{\alpha\beta}).$$

The Riemann curvature is given by

$$R^\mu{}_{\nu\alpha\beta} \equiv \partial_\alpha \Gamma^\mu_{\nu\beta} - \partial_\beta \Gamma^\mu_{\nu\alpha} + \Gamma^\mu_{\rho\alpha}\Gamma^\rho_{\nu\beta} - \Gamma^\mu_{\rho\beta}\Gamma^\rho_{\nu\alpha},$$

which defines the Ricci curvature, $R_{\mu\nu} \equiv R^\alpha{}_{\mu\alpha\nu}$, and Ricci scalar, $R \equiv R^\mu{}_\mu$. We *do not* use commas and semicolons to denote partial and covariant derivatives.

- Spatial 3-vectors (and Euclidean vectors more generally) are written in boldface, $\boldsymbol{x}$, with the corresponding letter used to denote their length, $x = |\boldsymbol{x}|$. Unit-length 3-vectors are written with a hat, $\hat{\boldsymbol{x}} = \boldsymbol{x}/x$.

- Complex conjugates are denoted with an asterisk, z^*.

- Matrices are denoted by uppercase sans-serif Latin letters, M, with their transpose and Hermitian conjugate written as M^T and $\mathsf{M}^\dagger$, respectively.

- We denote integrals over the real line $\mathbb{R} = (-\infty, +\infty)$ and over 3D Euclidean space $\mathbb{R}^3$ by

$$\int_\mathbb{R} \mathrm{d}x \equiv \int_{-\infty}^{+\infty} \mathrm{d}x \,, \qquad \int_{\mathbb{R}^3} \mathrm{d}^3\boldsymbol{x} \equiv \int_{-\infty}^{+\infty} \mathrm{d}x_1 \int_{-\infty}^{+\infty} \mathrm{d}x_2 \int_{-\infty}^{+\infty} \mathrm{d}x_3 \,,$$

while integrals over the 2-sphere S^2 are written as

$$\int_{S^2} \mathrm{d}^2\hat{\boldsymbol{r}} \equiv \int_{-1}^{+1} \mathrm{d}(\cos\theta) \int_0^{2\pi} \mathrm{d}\phi = \int_0^{\pi} \mathrm{d}\theta \sin\theta \int_0^{2\pi} \mathrm{d}\phi \,,$$

where (θ, ϕ) are the usual polar coordinates.

- We define the Fourier transform of a function $h(t)$ by

$$\tilde{h}(f) = \mathcal{F}[h](f) \equiv \int_{\mathbb{R}} \mathrm{d}t \, e^{-2\pi i f t} h(t), \qquad h(t) = \int_{\mathbb{R}} \mathrm{d}f \, e^{2\pi i f t} \tilde{h}(f).$$

Contents

Contents

0   Introduction: the dawn of gravitational-wave astronomy

'When I ask myself "What are the great things we got from the Renaissance?", it's the great art, the great music, the science insights of Leonardo da Vinci. Two hundred years from now when you ask "What are the great things that came from this era?", I think it's going to be an understanding of the Universe around us.'

Kip Thorne

The era of gravitational-wave (GW) astronomy has begun. On 14 September 2015, almost a century after Einstein published his General Theory of Relativity (GR) [250, 251], one of the most important and controversial predictions of his theory was verified in spectacular fashion by the Advanced LIGO interferometers [2, 302] with the detection of GWs from two merging black holes (BHs) [5]. This event—dubbed GW150914—helped dispel decades of scepticism around the physical existence of GWs due to theoretical misconceptions (including by Einstein himself) [186, 368] and spurious observational claims (most notably by Weber and his collaborators) [208, 583], and was immediately heralded as a major breakthrough by the scientific community. In the six years since this first detection, LIGO and its European counterpart Virgo [44] have discovered dozens more GW signals [22, 33, 40, 41], firmly establishing GW astronomy as a new observational science for the 21$^{\text{st}}$ century.

While testing the predictions of GR is a laudable goal in itself, it is far from the only motivation for GW science. Much of the appeal of GWs lies in their rôle as an astronomical *messenger*, akin to the photon or the neutrino: a new and fundamentally different tool with which to observe the Universe. The power of combining this new messenger with more traditional astronomical observations has already been forcefully demonstrated by the neutron star (NS) merger event GW170817 [12–14], whose GW signal, combined with observations across the electromagnetic (EM) spectrum, has eliminated a large class of modified gravity theories [93, 219, 261, 495], probed the properties of nuclear matter under extreme conditions [20, 413, 466], confirmed NS mergers as key sites of r-process nucleosynthesis [198, 244], and provided a novel and independent measurement of the local expansion rate of the Universe [15, 502].

Even more exciting, in my view, is the opportunity to probe objects and phenomena that *only* GWs can access: 'dark' parts of the Universe that are invisible to EM astronomy. This is important because we know that luminous matter only constitutes a small fraction of the cosmic matter inventory, with

the majority consisting of non-baryonic 'dark matter' (DM) whose nature is still mysterious [121]. Black holes (which may or may not account for most or all of the DM, see section 0.3.4) are a quintessential example of a dark astronomical object, made not of hot radiating gas like the Sun and other stars, but formed instead from pure spacetime geometry. Until September 2015, all observational evidence for BHs was *indirect*, mostly due to their gravitational influence on surrounding luminous matter (e.g., in quasars [516] or X-ray binaries [473]). With the advent of GW astronomy, we can now observe BHs *directly*, and LIGO/Virgo are already giving us fascinating new insights into their properties and origins as a result [23, 31, 35, 77, 110, 427, 485, 547].

Going back further, beyond the myriad dark objects in our cosmic neighbourhood, we reach an entire epoch in the Universe's history that is inaccessible to EM observations. For the first 370,000 years or so after the Big Bang, the Universe was so hot that the electrons were unbound from their atomic nuclei, forming a dense ionised plasma that was impenetrable to EM radiation [584]. The light emitted at the end of this epoch, which we observe today as the Cosmic Microwave Background (CMB) [238, 453], represents the earliest point in the Universe's history that can be probed with EM observations. Since the CMB has been mapped out in exquisite detail by the Planck satellite [54] and other missions [48, 112, 116, 297, 521], we have already, in a sense, hit the limit of EM astronomy's reach into the early Universe. GWs, on the other hand, face no such restriction; since they couple so weakly to matter, they pass through the hot pre-recombination plasma unimpeded, carrying signals from the Universe's earliest moments. Thus, while EM observations can tell us about conditions in the Universe at energies below $\sim 0.3\,\mathrm{eV}$ (i.e. $\sim 3000\,\mathrm{K}$, the temperature at which photons decoupled from baryons), GWs can reach far further, far beyond the $\sim 10^{13}\,\mathrm{eV}$ energies probed by terrestrial experiments such as the Large Hadron Collider (LHC) [260], all the way up to the Planck scale $m_\mathrm{Pl} \sim 10^{28}\,\mathrm{eV}$ and into the realm of quantum gravity.

This ability to probe the hidden side of the Universe, from BHs to the Big Bang, makes GWs the theoretical physicist's most powerful new tool, opening up countless new avenues for studying the nature of the cosmos and the fundamental laws that govern it. The goal of this thesis is to explore a few of these avenues. In chapter 1, we look at ways in which GW observations can be used to probe the large-scale structure (LSS) of the Universe by studying *anisotropies* in the cosmic GW background, analogous to those in the CMB. In chapter 2, we study novel GW signatures and phenomenology of *cosmic strings*, which are one of the most the promising ways in which GWs can probe the pre-recombination era. Finally, in chapter 3, we explore the possibility of using binary systems as *dynamical detectors* of GWs, and investigate the constraints that such detectors can place on cosmological first-order phase transitions.

In the remainder of this chapter, we set the stage by giving a self-contained introduction to GW astronomy. We begin in section 0.1 with a discussion of GWs themselves, in which we derive their basic physical properties in the framework of linearised GR. Then, in section 0.2, we introduce a GW signal which is of particular importance from the point of view of cosmology: the stochastic GW background (GWB). In section 0.3 we give an overview of some of the various GW sources which contribute to the GWB, focusing on those which are the most promising as probes of cosmology and fundamental physics. Finally, in section 0.4 we describe the basic principles and current status of observational searches for GWs.
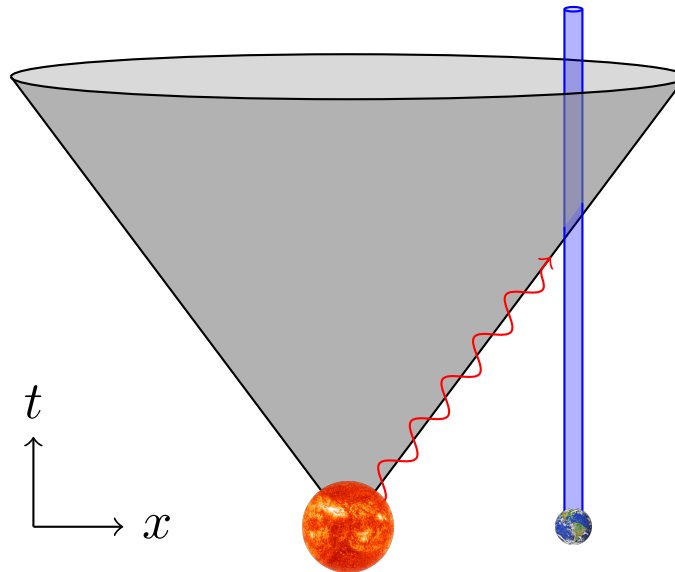
Figure 0.1: In Einsteinian gravity, disturbances in the Sun's gravitational field (shown in red) travel along the lightcone (grey), reaching the Earth's worldline (blue) some finite time after they originated. We call these propagating disturbances *gravitational waves*.

0.1 WHAT ARE GRAVITATIONAL WAVES?

The existence of GWs arises very naturally from the basic premise of GR, which is to formulate a *relativistic* theory of the gravitational field; i.e., one in which all observers agree on the speed of light in vacuum, $c \approx 3 \times 10^8 \, \mathrm{m \, s^{-1}}$. This requirement imposes a particular kind of causal structure on the theory, in which no physical degree of freedom can propagate faster than light, and the consequences of any event can therefore only be felt within its 'lightcone': the region of spacetime that can be reached by particles travelling at speeds $\leq c$ (see figure 0.1). Classical electrodynamics is a key example of such a theory.

This causal structure is in marked contrast with that of Newtonian gravity, in which changes in the gravitational field are felt *simultaneously* everywhere, no matter how far from the source. If, for example, the Sun were to spontaneously disappear, then according to Newton the Earth would immediately deviate from its orbit. Meanwhile, the last photons emitted by the Sun, obeying the relativistic laws of electrodynamics, would travel along the lightcone, and it would take ~ 8.3 minutes for the Sun to cease shining in the sky (as measured by an observer comoving with the Earth). This instantaneous Newtonian propagation is nonsensical from the perspective of a relativistic theory, as different observers cannot agree on what is meant by 'simultaneous'—some would even see the Earth's orbit change *before* the Sun disappeared, making the causal relationship between the two events ambiguous. However, all observers would agree that the sunlight ceased to reach the Earth *after* the Sun disappeared, as the speed-of-light propagation of this signal has a special status in a relativistic theory.

Clearly this situation would make much more sense if changes to the gravitational field also propagated at the speed of light. As we discuss below, this is indeed the situation in GR.[1] This simple example illustrates the most general and straightforward way to understand what GWs are: they are degrees of

---

[1]In fact, this idea was discussed decades before Einstein's formulation of GR, most notably by Heaviside [309] and Poincaré [458].

freedom of the gravitational field which propagate at the speed of light, thereby endowing gravitational interactions with a relativistic causal structure akin to that of electrodynamics.[2]

In the remainder of this section, we explore the physical properties of these light-like excitations of the gravitational field. After a very brief recap of GR, we introduce the linearised form of the theory and use this to show that there are in fact *two* degrees of freedom, before discussing how these interact with test masses, how they are sourced by the motion of matter, and how we can assign energy and momentum to them. We end with a discussion of the *GW memory effect*, a fascinating prediction of GR which is further explored in chapter 2 of this thesis. Throughout we use units in which $c = 1$.

0.1.1 A LIGHTNING REVIEW OF GENERAL RELATIVITY

We have referred several times above to 'the gravitational field'. In GR, this means the *spacetime metric* $g_{\mu\nu}(x)$: a symmetric rank-$(0, 2)$ tensor field, whose components in any coordinate chart $\{x^\mu\}$ determine the spacetime interval $\mathrm{d}s$ between two infinitesimally separated events,

$$\mathrm{d}s^2 = g_{\mu\nu}\, \mathrm{d}x^\mu\, \mathrm{d}x^\nu. \tag{0.1}$$

The resulting spacetime geometry is generically *curved*, and this curvature influences the trajectories of particles through spacetime. In the absence of non-gravitational interactions, these trajectories are solutions of the geodesic equation

$$\frac{\mathrm{d}U^\mu}{\mathrm{d}\lambda} + \Gamma^\mu_{\alpha\beta} U^\alpha U^\beta = 0, \tag{0.2}$$

where $U^\mu(\lambda)$ is the 4-velocity of the particle along a trajectory with affine parameter λ, and $\Gamma^\mu_{\alpha\beta}$ are the Christoffel symbols associated with the metric $g_{\mu\nu}$. The metric, in turn, evolves dynamically according to the Einstein field equation (EFE),[3]

$$R_{\mu\nu} - \frac{1}{2}R g_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{0.3}$$

where $R_{\mu\nu}$ is the Ricci tensor corresponding to $g_{\mu\nu}$ and R is the Ricci scalar, while the *energy-momentum tensor* $T_{\mu\nu}$ describes how matter (i.e., all other fields that exist in spacetime) acts as a source of spacetime curvature. This beautiful reciprocal interplay between the geometry of spacetime and the matter fields that inhabit it is succinctly captured by Wheeler [588]: 'Spacetime tells matter how to move; matter tells spacetime how to curve.'

The EFE can also be derived from an action principle, writing

$$S = S_{\mathrm{EH}} + S_{\mathrm{mat}}, \qquad S_{\mathrm{EH}} \equiv \frac{1}{16\pi G} \int \mathrm{d}^4 x \sqrt{-g} R, \qquad T^{\mu\nu} \equiv \frac{2}{\sqrt{-g}} \frac{\delta S_{\mathrm{mat}}}{\delta g_{\mu\nu}}, \tag{0.4}$$

---

[2]Note however that GWs do not *have* to propagate at the speed of light in a generic relativistic theory [175, 480]; this is merely a particularly appealing and well-motivated option. In practice, joint GW and EM observations of GW170817 have demonstrated that any differences between the speed of light and the speed of GWs must be incredibly small: less than one part in 10^{15} [93, 219, 261, 495].

[3]This is sometimes written with an additional cosmological constant term, $R_{\mu\nu} - \frac{1}{2}R g_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}$. Here we choose instead to absorb Λ into $T_{\mu\nu}$, treating it as a homogeneous matter field with negative pressure, $p = -\rho$.

where $g \equiv \det g_{\mu\nu}$ is the metric determinant, S_{EH} is the *Einstein-Hilbert action*, and S_{mat} is the action of the matter fields. The metric which extremises the total action S then obeys equation (0.3).

One of the most important features of the EFE (0.3) is that it holds in any coordinate chart; we are free to arbitrarily redefine our coordinates $x \to x'(x)$ (subject to differentiability conditions) by appropriately redefining the metric,

$$g_{\mu\nu}(x) \to g'_{\mu\nu}(x') = \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} g_{\alpha\beta}(x), \qquad (0.5)$$

and similarly redefining the energy-momentum tensor. Such a transformation is called a *diffeomorphism*, and the indifference of GR to diffeomorphisms is called the principle of *general covariance*. This principle reflects the notion that spacetime coordinates are merely artificial constructs that we introduce to facilitate calculations, and should play no fundamental rôle in the laws of physics. The resulting freedom to choose coordinates in GR can thus be thought of as a gauge freedom, analogous to the choice of potential in electrodynamics. Practically, this huge redundancy in description means that we must be careful to separate the physical degrees of freedom from gauge degrees of freedom; historically it was this difficulty that caused much of the early confusion over the existence of GWs in GR [186].

0.1.2 LINEARISED GENERAL RELATIVITY

The EFE (0.3), while famous for its elegance and conceptual simplicity, is also infamously difficult to solve. While many exact solutions do exist [526], only a handful represent physically relevant spacetimes, and those that do often assume a high degree of symmetry that is all but guaranteed to be broken in reality. Chief among these are the flat Minkowski spacetime, the Friedmann-Lemaître-Roberson-Walker (FLRW) solutions used in cosmology, and the Kerr-Newman family of BH spacetimes.

Despite the relative simplicity of these solutions, there are many physical situations where differences between the true spacetime metric $g_{\mu\nu}$ and that of some exact solution $g_{\mu\nu}^{(0)}$ are 'small' in some sense. We can then make progress by using *perturbation theory*, defining the small quantity $h_{\mu\nu} \equiv g_{\mu\nu} - g_{\mu\nu}^{(0)}$, and writing the relevant equations as series expansions in h. While these methods are necessarily approximate, due to the need to truncate the expansion at some order in h, they are often incredibly powerful: for example, cosmological perturbation theory (where $g_{\mu\nu}^{(0)}$ belongs to the FLRW family) provides an excellent description of the anisotropies in the CMB [584] as well as many other cosmological observables at lower redshift [115], while BH perturbation theory (where $g_{\mu\nu}^{(0)}$ belongs to the Kerr-Newman family) allows us to compute the quasinormal 'ringing' of BHs formed in mergers [119, 377] and the orbits of small bodies around supermassive BHs through the 'self-force' formalism [96].

The case where $g_{\mu\nu}^{(0)}$ is the Minkowski metric $\eta_{\mu\nu} = \mathrm{diag}(-1, +1, +1, +1)$ is by far the simplest, describing a situation where there are no strongly-gravitating objects (such as black holes and neutron stars) and where cosmological expansion is negligible. Despite its simplicity, however, this flat-space perturbation theory is sufficient to capture many of the most salient features of GWs. In fact, for much of the discussion we can simplify things even further by considering only the first-order (linear) term in the series expansion. The resulting theory is called *linearised* GR.[4]

---

[4]It is important to note that linearised GR is not the only tool for studying GWs, and that it is not necessary to assume that GWs are perturbatively small. For example, one can use the Bondi-Sachs formalism [147, 240, 492] to study the large-distance

Inserting the perturbed flat-space metric

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad |h_{\mu\nu}| \ll 1, \tag{0.6}$$

into the EFE (0.3) and keeping only the linear-order terms, we obtain [406]

$$\Box \bar{h}_{\mu\nu} + \eta_{\mu\nu} \partial^\alpha \partial^\beta \bar{h}_{\alpha\beta} - 2\partial_{(\mu} \partial^\alpha \bar{h}_{\nu)\alpha} = -16\pi G T_{\mu\nu}, \tag{0.7}$$

which we have written in terms of the trace-reversed GW perturbation,

$$\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2} h^\alpha{}_\alpha \eta_{\mu\nu}. \tag{0.8}$$

The resulting equation (0.7) is a hyperbolic partial differential equation (PDE), i.e., one which has wavelike solutions which propagate at a characteristic finite velocity; in this case, the speed of light, as suggested by the presence of the (flat-space) d'Alembertian operator $\Box \equiv \partial^\mu \partial_\mu = -\partial_0^2 + \partial^i \partial_i$, and the absence of any mass term $\sim m^2 \bar{h}_{\mu\nu}$. The other derivative terms in equation (0.7) do not spoil this behaviour, and in fact we see below that we can safely set them to zero with an appropriate gauge choice.

0.1.3 Fixing the gauge

Since $\bar{h}_{\mu\nu}$ has two symmetric spacetime indices, the linearised EFE (0.7) has, in principle, ten independent degrees of freedom. However, due to general covariance, we must be extremely cautious in identifying which of these degrees of freedom are physical (if any), and which can be removed by an appropriate gauge choice.

Writing a general diffeomorphism as $x^\mu \to x^\mu + \xi^\mu$, and using equation (0.5), we find that the metric perturbation transforms as

$$h_{\mu\nu} \to h_{\mu\nu} - 2\partial_{(\mu} \xi_{\nu)}. \tag{0.9}$$

We thus see that specifying an almost-flat metric (0.6) has already ruled out a large class of diffeomorphisms, as the only remaining gauge transformations consistent with this ansatz are those in which $|\partial \xi| \ll 1$. Nonetheless, the $\partial \xi$ terms are still permitted to be of the same size as h, so they must still be carefully separated from the physical degrees of freedom.

We saw above that the interpretation of the linearised EFE (0.7) as a wave equation was obscured somewhat by the presence of terms like $\partial^\mu \bar{h}_{\mu\nu}$, and mentioned that these can be removed by a gauge transformation. Indeed, from equations (0.8) and (0.9) we find that $\partial^\mu \bar{h}_{\mu\nu} \to \partial^\mu \bar{h}_{\mu\nu} - \Box \xi_\nu$, which shows that it is always possible to set

$$\partial^\mu \bar{h}_{\mu\nu} = 0 \tag{0.10}$$

---

behaviour of GWs of arbitrary amplitude emitted by an isolated source, and there are also numerous exact solutions describing arbitrarily strong gravitational plane waves [240, 526].

with an appropriate choice of ξ, as the latter can always be constructed using the Green's function for the d'Alembertian. This is the *harmonic* gauge condition,[5] in which the EFE becomes

$$\Box \bar{h}_{\mu\nu} = -16\pi G T_{\mu\nu}, \qquad (0.11)$$

making its interpretation as a massless relativistic wave equation much clearer.

Setting $\partial^{\mu}\bar{h}_{\mu\nu} = 0$ removes four gauge degrees of freedom from the ten that we started with: one for each value of the free spacetime index ν. This does not fully specify ξ however, as it is still possible to perform a smaller set of gauge transformations that obey $\Box \xi_{\mu} = 0$; i.e., we are free to choose four harmonic functions. If we are in a region of spacetime with no matter sources, $T_{\mu\nu} = 0$, then this allows us to set certain components of $\bar{h}_{\mu\nu}$ to zero.[6] A particularly appealing choice is to set

$$\bar{h}_{0i} = 0, \qquad (0.12)$$

which is called the *transverse* condition. To see why, consider a plane-wave solution to the vacuum EFE, $\Box \bar{h}_{\mu\nu} = 0$, given by

$$\bar{h}_{\mu\nu}(t, \boldsymbol{x}) = e^{2\pi i f(t - \hat{\boldsymbol{r}} \cdot \boldsymbol{x})} \bar{H}_{\mu\nu}, \qquad (0.13)$$

where f is the frequency of the wave, $\hat{\boldsymbol{r}}$ is a unit vector pointing in the propagation direction, and the components of $\bar{H}_{\mu\nu}$ are constants describing the magnitude of the trace-reversed metric perturbation. Applying the transverse condition (0.12) in the harmonic gauge (0.10) gives us

$$\partial^{\mu}\bar{h}_{\mu i} = \partial^{j}\bar{h}_{ji} = 0 \qquad \Longrightarrow \qquad \hat{r}^{j}\bar{H}_{ji} = 0, \qquad (0.14)$$

such that the projection of the plane wave onto the propagation direction is zero; i.e., there is no perturbation along the longitudinal direction, only in the transverse directions.

The transverse condition (0.12) uses up three of our four free harmonic functions (one for each of the three values of the spatial index i). We can use the final one to additionally enforce the *traceless* condition,

$$\bar{h}^{\mu}{}_{\mu} = 0, \qquad (0.15)$$

which erases the distinction between the metric perturbation and its trace-reversed form, $h_{\mu\nu} = \bar{h}_{\mu\nu}$.[7] Since the coordinate volume element in our perturbed flat spacetime is proportional to $\sqrt{-g} = \sqrt{1 + h^{\mu}{}_{\mu}}$,

---

[5] This name comes from the fact that the spacetime coordinates are then solutions of the curved-space wave equation, $\nabla^{\mu}\nabla_{\mu}x^{\nu} = 0$. Other names include the de Donder gauge, the Hilbert gauge, and in the context of linearised GR specifically, the Lorenz or (erroneously) Lorentz gauge [334], by analogy with electrodynamics.

[6] We cannot set components of $\bar{h}_{\mu\nu}$ to zero inside a matter source by subtracting only harmonic functions obeying $\Box \xi_{\mu} = 0$, because the flat-space d'Alembertian commutes with partial derivatives to give $\Box \bar{h}_{\mu\nu} \rightarrow \Box(\bar{h}_{\mu\nu} - 2\partial_{(\mu}\xi_{\nu)} + \eta_{\mu\nu}\partial^{\alpha}\xi_{\alpha}) = \Box \bar{h}_{\mu\nu} \neq 0$.

[7] Note that while we have set $h_{0i} = 0$, we have kept h_{00} thus far. This component must be constant since $\partial^{0}h_{00} = \partial^{\mu}h_{\mu 0} = 0$ by the harmonic condition, such that the vacuum EFE reduces to the Laplace equation, $\partial^{i}\partial_{i}h_{00} = 0$, and thus does not exhibit wavelike behaviour. In fact, in the appropriate weak-field, slow-motion limit of GR, this component is related to the Newtonian gravitational potential, $\phi(\boldsymbol{x}) = -\frac{1}{2}h_{00}$ [429]. We can thus neglect it if we are sufficiently far from any massive objects, leaving just the spatial components h_{ij} as the nonzero parts of the metric perturbation. The traceless condition is then just $h^{i}{}_{i} = 0$.

equation (0.15) can be interpreted as choosing the gauge such that the coordinate volume of a spacetime region is not affected by a passing GW.

Combining the three conditions in equations (0.10), (0.12), and (0.15), we have exhausted our possibilities for specifying ξ, and have thus removed all gauge freedom from the vacuum linearised EFE. We call this the *transverse-traceless* (TT) gauge.[8] Of our initial ten degrees of freedom, we have removed four by choosing the harmonic gauge, and a further four with the transverse and traceless conditions, leaving *two* physical degrees of freedom. We explore these two GW modes below.

0.1.4 THE TWO POLARISATION MODES

Let us now return to the plane wave (0.13) and apply the TT gauge. Without loss of generality, we can choose our spatial coordinates (x, y, z) such that the GW is propagating in the $\hat{z}$ direction. The transverse condition then becomes $h_{zi} = 0$, leaving just four nonzero components: two on-diagonal (h_{xx} and h_{yy}) and two off-diagonal (h_{xy} and h_{yx}). The former must be the same up to a minus sign to ensure that the trace vanishes, and the latter must be equal to each other to ensure h_{ij} is symmetric, so we see that only two components can be freely chosen; these are the two radiative degrees of freedom of the vacuum gravitational field in GR, which we call the *plus* and *cross polarisations*, h_+ and $h_\times$, for reasons that will become apparent in section 0.1.5.

We can thus write the $\hat{z}$-pointing plane wave as

$$h_{ij}(t, z) = e^{2\pi i f(t-z)} \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix} = \sum_{A=+,\times} e^{2\pi i f(t-z)} e_{ij}^A h_A, \qquad (0.16)$$

where in the second equality we have defined the TT *polarisation tensors*,

$$e_{ij}^+ = \hat{x}_i \hat{x}_j - \hat{y}_i \hat{y}_j, \qquad e_{ij}^\times = 2\hat{x}_{(i} \hat{y}_{j)}, \qquad (0.17)$$

which encapsulate the geometric pattern of each of the two polarisations. We can also generalise to an arbitrary propagation direction $\hat{r}$, which we specify using the standard spherical polar angles $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$; we then have

$$h_{ij}(t, \boldsymbol{x}) = \sum_{A=+,\times} e^{2\pi i f(t-\hat{r}\cdot\boldsymbol{x})} e_{ij}^A(\hat{r}) h_A, \qquad (0.18)$$

with the polarisation tensors given by

$$e_{ij}^+(\hat{r}) = \hat{\theta}_i \hat{\theta}_j - \hat{\phi}_i \hat{\phi}_j, \qquad e_{ij}^\times(\hat{r}) = 2\hat{\theta}_{(i} \hat{\phi}_{j)}, \qquad (0.19)$$

---

[8]In constructing the TT gauge, we have chosen our coordinates such that the metric perturbation is transverse and traceless. However, it is useful to note that this is not strictly necessary in practice. Given a generic coordinate system in linearised GR, it is always possible to extract the *TT part* of the metric perturbation with an appropriate projection operation. So long as the coordinate system is in the harmonic gauge (so that equation (0.11) holds), one can show that this TT part is equal to the value of the metric perturbation in the TT gauge, allowing us to identify this part as representing GWs even when we are not in the TT gauge.

where the propagation direction $\hat{\boldsymbol{r}}$ and two transverse directions $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$ are given by

$$
\begin{aligned}
\hat{\boldsymbol{r}} &= \sin\theta\cos\phi\,\hat{\boldsymbol{x}} + \sin\theta\sin\phi\,\hat{\boldsymbol{y}} + \cos\theta\,\hat{\boldsymbol{z}}, \\
\hat{\boldsymbol{\theta}} &= \cos\theta\cos\phi\,\hat{\boldsymbol{x}} + \cos\theta\sin\phi\,\hat{\boldsymbol{y}} - \sin\theta\,\hat{\boldsymbol{z}}, \\
\hat{\boldsymbol{\phi}} &= -\sin\phi\,\hat{\boldsymbol{x}} + \cos\phi\,\hat{\boldsymbol{y}}.
\end{aligned}
\tag{0.20}
$$

Note that these are normalised such that

$$
e^A_{ij}(\hat{\boldsymbol{r}})e^{A',ij}(\hat{\boldsymbol{r}}) = 2\delta_{AA'},
\tag{0.21}
$$

where A, A' are polarisation indices running over $+, \times$, and such that

$$
a_{ij} = \frac{1}{2}e^A_{ij}e^{A,kl}a_{kl}
\tag{0.22}
$$

for any symmetric TT rank-$(0,2)$ tensor $a_{\mu\nu}$, with implicit summation over the polarisation index. We are free to redefine the two polarisation modes by rotating $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$ around the $\hat{\boldsymbol{r}}$-axis by an arbitrary angle ψ; this gives

$$
\begin{pmatrix} e^+_{ij} \\ e^\times_{ij} \end{pmatrix} \rightarrow \begin{pmatrix} \cos 2\psi & -\sin 2\psi \\ \sin 2\psi & \cos 2\psi \end{pmatrix}\begin{pmatrix} e^+_{ij} \\ e^\times_{ij} \end{pmatrix} = \begin{pmatrix} \cos 2\psi\, e^+_{ij} - \sin 2\psi\, e^\times_{ij} \\ \sin 2\psi\, e^+_{ij} + \cos 2\psi\, e^\times_{ij} \end{pmatrix},
\tag{0.23}
$$

mixing the amplitudes of the two polarisations. We refer to ψ as the 'polarisation angle'.

For a generic plane wave propagating in the $\hat{\boldsymbol{r}}$-direction, it is often convenient to define

$$
h(t,\boldsymbol{x}) \equiv \frac{1}{2}\left[e^{+,ij}(\hat{\boldsymbol{r}}) - \mathrm{i}e^{\times,ij}(\hat{\boldsymbol{r}}) \right]h_{ij}(t,\boldsymbol{x}) = h_+(t,\boldsymbol{x}) - \mathrm{i}h_\times(t,\boldsymbol{x}),
\tag{0.24}
$$

thereby encoding the two real TT degrees of freedom as a single complex variable, whose amplitude is

$$
|h|^2 = hh^* = h_+^2 + h_\times^2 = \frac{1}{2}h^{ij}h_{ij}.
\tag{0.25}
$$

Under rotations of the polarisation angle, this complex strain transforms like

$$
h \rightarrow \mathrm{e}^{-2\mathrm{i}\psi}h,
\tag{0.26}
$$

which is the transformation law for a massless spin-2 particle.

Since we are working in linearised GR, any interactions between different plane waves (which correspond to terms of order $\sim h^2$) are neglected. This allows us to trivially superpose as many plane-wave solutions as we like, each with arbitrary frequency, propagation direction, and plus- and cross-mode amplitudes (so long as the latter are both much smaller than unity). In fact, the set of all plane waves

serves as a basis in the space of solutions to the linearised vacuum EFE. We can thus write an arbitrary vacuum GW solution as a superposition of plane waves,

$$h_{ij}(t, \boldsymbol{x}) = \int_{\mathbb{R}} \mathrm{d}f \int_{S^2} \mathrm{d}^2 \hat{\boldsymbol{r}} \, \mathrm{e}^{2\pi \mathrm{i} f(t - \hat{\boldsymbol{r}} \cdot \boldsymbol{x})} e_{ij}^A(\hat{\boldsymbol{r}}) \tilde{h}_A(f, \hat{\boldsymbol{r}}), \tag{0.27}$$

with the contribution from each polarisation, frequency, and propagation direction described by the Fourier components $\tilde{h}_A(f, \hat{\boldsymbol{r}})$, and with implicit summation over the polarisation index. We can also invert this to write

$$\tilde{h}_A(f, \hat{\boldsymbol{r}}) = \frac{1}{2} f^2 \int_{\mathbb{R}^3} \mathrm{d}^3 \boldsymbol{x} \, \mathrm{e}^{-2\pi \mathrm{i} f(t - \hat{\boldsymbol{r}} \cdot \boldsymbol{x})} e^{A, ij}(\hat{\boldsymbol{r}}) h_{ij}(t, \boldsymbol{x}). \tag{0.28}$$

This plane-wave expansion will prove useful when we discuss stochastic GW backgrounds in section 0.2.

0.1.5 Interaction of gravitational waves with test masses

Since GWs are distortions in the curvature of spacetime, they affect the solutions to the geodesic equation (0.2), and can thus influence the trajectories of particles. (Of course, there is no absolute reference frame in GR, so we are really interested in how GWs affect the pairwise separation between different geodesics.) These effects are obscured if we work in TT coordinates, as the relevant Christoffel symbols for a particle initially at rest ($U^i = 0$) are

$$\Gamma_{00}^i = \partial_0 h_0{}^i - \frac{1}{2} \partial^i h_{00}, \tag{0.29}$$

which vanish by construction in the TT frame. As a result, the *coordinate* distance between two initially stationary freely-falling particles is unaffected by GWs in this frame at linear order. However, the *proper* distance between them (i.e., as measured using the spacetime metric) is affected, due to the time-varying metric perturbation h_{ij}. Since proper distances are diffeomorphism-invariant, we see that this must be a physical effect and not a coordinate artefact.

In order to analyse the response of a system of test masses to an impinging GW, it is therefore preferable to choose an alternative coordinate frame in which coordinate distances more closely follow the behaviour of proper distances. We can do this by constructing the *proper detector frame* (also called *Fermi coordinates*), in which the metric is locally flat with vanishing first derivatives along the trajectory of the system, with no rotation of the coordinates along this trajectory (in the sense that the coordinate spin axes of freely-spinning gyroscopes do not change over time). Since the metric is locally flat, coordinate distances are then equal to proper distances if we work sufficiently close to the origin. It is straightforward to show that, in this frame, the separation vector ξ^i between two nearby geodesics obeys $\ddot{\xi}^i = R^i{}_{00j} \xi^j$ [406], where dots denote derivatives with respect to the coordinate time (which is locally equivalent to the system's proper time).

At this point, we could evaluate the Riemann tensor as a function of the metric perturbation in the proper detector frame. However, it turns out that the components of the Riemann tensor are *invariant* under gauge transformations in linearised GR (note this is a stronger statement than the tensor obeying
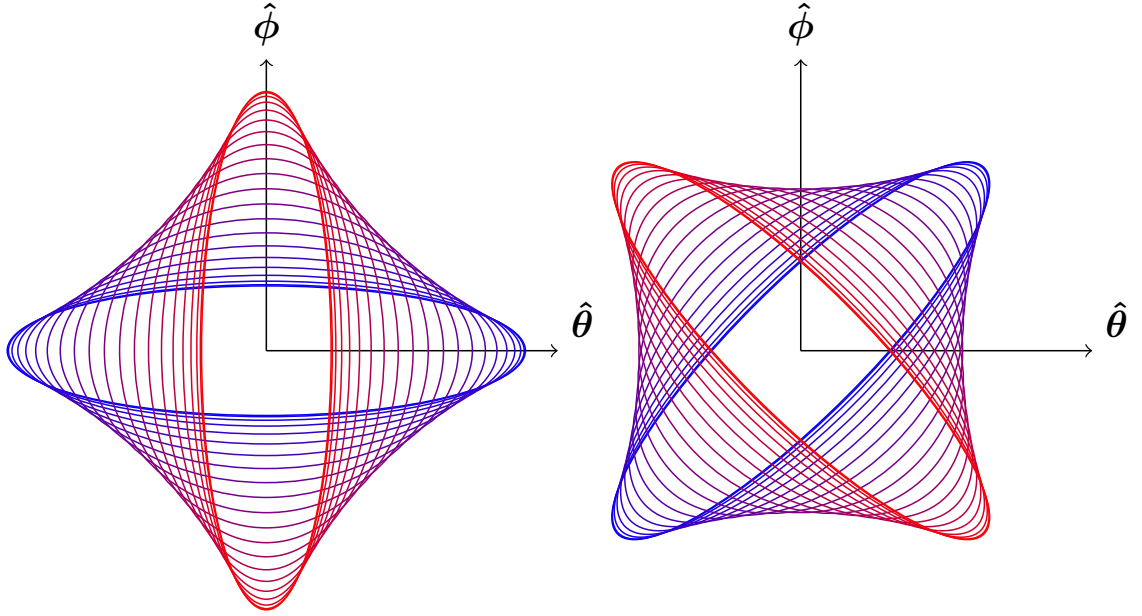
Figure 0.2: Deformation of a circular ring of test particles by a plane GW propagating out of the page. The left panel shows the pure +-polarised case, while the right panel shows the ×-polarised case. The colour gradient (from blue to red) shows how the deformation evolves over a time interval equal to half the GW oscillation period. (Note that the amplitude of the deformation has been greatly exaggerated here for clarity.)

general covariance), so we are free to evaluate it in the TT frame, making the link to the physical GW perturbation clearer. We therefore find that

$$\ddot{\xi}^i = \frac{1}{2}\ddot{h}^{ij}\xi_j, \tag{0.30}$$

where ξ^i is the coordinate separation in the proper detector frame as before, and h_{ij} is the metric perturbation in the TT frame.

Substituting in the $\hat{\boldsymbol{z}}$-pointing plane wave solution from equation (0.16), we find that equation (0.30) has solutions

$$\boldsymbol{\xi} = \frac{1}{2}e^{2\pi i f (t-z)}\left[h_+(x_0\hat{\boldsymbol{x}} - y_0\hat{\boldsymbol{y}}) + h_\times(y_0\hat{\boldsymbol{x}} + x_0\hat{\boldsymbol{y}})\right], \tag{0.31}$$

where (x_0, y_0) are the initial coordinates of the separation vector (we can treat these as constant on the right-hand side here, as their variation over time only gives corrections of order $\sim h^2$). The resulting deformations of a circular ring of test particles are shown in figure 0.2, clearly demonstrating where the plus and cross polarisations get their names from. We see that $\boldsymbol{\xi} \cdot \hat{\boldsymbol{z}} = 0$, so the GW is transverse as expected.

When we say that equation (0.30) holds for pairs of 'nearby' geodesics, we mean that the size of the separation vector ξ is much less than the lengthscale over which the spacetime curvature varies—in this case, the wavelength of the GW, $\lambda \sim 1/f$. This situation, in which $\xi \ll \lambda$, is called the *small-antenna limit* in the context of GW detectors. As we discuss in section 0.4, this limit holds for some GW experiments (e.g., ground-based interferometers such as LIGO and Virgo), but not for others (e.g., pulsar timing arrays). In the latter case, the right-hand side of equation (0.30) picks up higher-order terms in ξ, as well as a nontrivial dependence on the frequency of the GW.

0.1.6 Generation of gravitational waves

We have focused thus far on the propagation of GWs in vacuum. However, in order to use GWs as *messengers* which tell us about objects in the cosmos, we need to understand how said objects generate GWs in the first place. To do so, we abandon the TT gauge momentarily (since our freedom to set certain metric components to zero is inhibited by the presence of a source term), and return to the linearised EFE in the harmonic gauge, equation (0.11). We can invert the flat-space d'Alembertian using the retarded Green's function,

$$G(x, x') = -\frac{\delta(t - t' - |\boldsymbol{x} - \boldsymbol{x}'|)}{4\pi |\boldsymbol{x} - \boldsymbol{x}'|}, \qquad \Box G(x, x') = \delta^{(4)}(x - x'). \tag{0.32}$$

The corresponding solution for the trace-reversed metric perturbation is then given by

$$\bar{h}_{ij}(t, \boldsymbol{x}) = -16\pi G \int \mathrm{d}^4 x' \, G(x, x') T_{ij}(x') = 4G \int \frac{\mathrm{d}^3 \boldsymbol{x}'}{|\boldsymbol{x} - \boldsymbol{x}'|} T_{ij}(t - |\boldsymbol{x} - \boldsymbol{x}'|, \boldsymbol{x}'), \tag{0.33}$$

where the integral runs over a constant-time hyperslice. Outside of the source we are once again free to impose the TT gauge, writing

$$h_{ij}(t, \boldsymbol{x}) = 4G \int \frac{\mathrm{d}^3 \boldsymbol{x}'}{|\boldsymbol{x} - \boldsymbol{x}'|} \left[T_{ij}(t - |\boldsymbol{x} - \boldsymbol{x}'|, \boldsymbol{x}') \right]^{\mathrm{TT}}, \tag{0.34}$$

where the 'TT' superscript indicates that we have extracted the transverse-traceless part of the energy-momentum tensor (with the transverse directions defined relative to the propagation from the source to $\boldsymbol{x}$), so that the quantity on the left-hand side corresponds to a gauge-invariant GW propagating in vacuum (see footnote 8).

In many physical situations, the motion of the source is more straightforwardly described in the Fourier domain. We thus rewrite the energy-momentum tensor in terms of its spacetime Fourier transform,

$$T_{ij}(t, \boldsymbol{x}) = \int \mathrm{d}f \int \mathrm{d}^3 \boldsymbol{k} \, \mathrm{e}^{2\pi \mathrm{i}(ft - \boldsymbol{k} \cdot \boldsymbol{x})} \tilde{T}_{ij}(f, \boldsymbol{k}). \tag{0.35}$$

Inserting this into equation (0.34), we find that the GW solution at distances much greater than the size of the source is given by

$$h_{ij}(t, \boldsymbol{x}) = \frac{4G}{r} \int \mathrm{d}f \, \mathrm{e}^{2\pi \mathrm{i} f(t - r)} \left[\tilde{T}_{ij}(f, f\hat{\boldsymbol{r}}) \right]^{\mathrm{TT}}, \qquad r \gg \max |\boldsymbol{x}'|, \tag{0.36}$$

where $\boldsymbol{r}$ is a 3-vector pointing from the centre of the source to $\boldsymbol{x}$. Equivalently, we can write this in terms of the plane wave components $\tilde{h}_A$ using equation (0.28),

$$\tilde{h}_A(f, \hat{\boldsymbol{n}}) = \frac{2G}{r} \mathrm{e}^{-2\pi \mathrm{i} f r} e^{A, ij}(\hat{\boldsymbol{r}}) \tilde{T}_{ij}(f, f\hat{\boldsymbol{r}}) \delta^{(2)}(\hat{\boldsymbol{r}}, \hat{\boldsymbol{n}}). \tag{0.37}$$

Here the exponential phase factor reflects the time taken for the signal to reach distance r (this can be removed by translating the time coordinate used in the Fourier transforms), while the Dirac delta tells us

that the only nonzero plane wave component is that which points in the direction of the source.[9] (Note that the 'TT' subscript is no longer needed, as we have projected onto the TT polarisation tensors.)

Thus far we have made no assumptions about the source other than requiring it to be weakly-gravitating, so as to be consistent with linearised GR. However, we can greatly simplify our description of GW generation if we consider sources with internal velocities much smaller than the speed of light,[10] $v \ll 1$. In this limit, the resulting GW signal is dominated by the mass quadrupole of the source,

$$h_{ij}(t, \boldsymbol{x}) \simeq \frac{2G}{r} \big[\ddot{Q}_{ij}(t - r) \big]^{\mathrm{TT}}, \qquad Q_{ij}(t) \equiv \int \mathrm{d}^3 \boldsymbol{x}' \, \rho(t, \boldsymbol{x}') x_i' x_j', \qquad (0.38)$$

where $\rho \equiv T_{00}$ is the mass density. As a simple example, consider a system of two slowly-moving point masses. In a frame where the centre of mass of the system is fixed at the origin, the quadrupole moment is given by

$$Q^{ij} = m_1 x_1^i x_1^j + m_2 x_2^i x_2^j = \mu R^i R^j \qquad (0.39)$$

(ignoring terms of order v^2), where $\boldsymbol{R} \equiv \boldsymbol{x}_1 - \boldsymbol{x}_2$ is a 3-vector pointing from the position of one mass to the other, and $\mu \equiv m_1 m_2 / (m_1 + m_2)$ is the reduced mass of the system. The emitted GW signal is thus

$$h_{ij} \simeq \frac{4G\mu}{r} \big[\dot{R}_i \dot{R}_j + R_{(i} \ddot{R}_{j)} \big]^{\mathrm{TT}}, \qquad (0.40)$$

where the right-hand side is evaluated at the retarded time $t - r$.

0.1.7 ENERGY AND MOMENTUM OF GRAVITATIONAL WAVES

The notions of energy and momentum are extremely intuitive and useful in understanding the physics of GWs. However, they are also notoriously thorny issues in GR [535], as general covariance makes it impossible to find a local measure (i.e., one defined at each point in spacetime) of gravitational energy-momentum that all observers can agree on. To see this, note that the energy density of the gravitational field in Newtonian gravity goes like $\sim (\nabla \phi)^2$ (where ϕ is the Newtonian potential), so to ensure consistency in the Newtonian limit we would expect any analogous quantity in GR to be quadratic in the first derivatives of the metric, $\sim (\partial g)^2$. The problem is that it is always possible to transform to a local inertial frame in which these first derivatives vanish, and we can therefore always set any such quantity to zero at any point in spacetime. (This problem arises in both linearised GR and the full nonlinear theory, and is just as much of an obstacle for defining, e.g., the mass density of a BH as it is for defining the energy density of GWs.)

Nonetheless, we can still define a sensible *quasi-local* energy-momentum tensor for the gravitational field by averaging over a region of spacetime, rather than focusing on a single point. The intuition here is that, while ∂g can be set to zero at any individual point, it cannot be simultaneously set to zero at every point in a sufficiently 'large' region of spacetime in any one coordinate frame.

---

[9] Here $\delta^{(2)}(\hat{\boldsymbol{r}}, \hat{\boldsymbol{r}}') \equiv \delta(\cos\theta - \cos\theta')\delta(\phi - \phi')$ is the Dirac delta function on the 2-sphere, defined such that we have $\int_{S^2} \mathrm{d}^2 \hat{\boldsymbol{r}}' \, f(\hat{\boldsymbol{r}}') \delta^{(2)}(\hat{\boldsymbol{r}}, \hat{\boldsymbol{r}}') = f(\hat{\boldsymbol{r}})$ for test functions $f : S^2 \to \mathbb{R}$.

[10] Note that in virialised self-gravitating systems, v^2 is of the same order of magnitude as the gravitational potential GM/R, where M and R are the mass and radius of the system; any self-gravitating system which is consistent with the linearised GR requirement that $GM/R \ll 1$ must therefore also be slowly-moving, $v \ll 1$.

There are two key approaches for constructing a quasi-local energy-momentum tensor for GWs.[11] One is to write down a Lagrangian for linearised GR and use Noether's theorem to identify the conserved currents associated with the invariance of this Lagrangian under spacetime translations (this is the usual procedure for defining energy and momentum in other field theories). The other is to work directly with the EFE and study how GWs *backreact* on spacetime, acting as a source for further perturbations to the metric.[12] Both approaches yield the same answer:

$$T_{\text{gw}}^{\mu\nu} = \frac{1}{32\pi G} \langle \partial^\mu h^{\alpha\beta} \partial^\nu h_{\alpha\beta} \rangle, \tag{0.41}$$

where the angle brackets denote a quasi-local spatial average over a region much larger than the wavelength of the GW (or equivalently, a temporal average over many periods). In particular, the quasi-local GW energy density is

$$\rho_{\text{gw}} = \frac{1}{32\pi G} \langle \dot{h}^{\alpha\beta} \dot{h}_{\alpha\beta} \rangle. \tag{0.42}$$

These expressions apply so long as we are in the harmonic gauge, and are invariant under the remaining gauge transformations by virtue of the averaging process.

Using equation (0.41), we can thus calculate the total GW energy flux radiated by a generic isolated source in a given direction $\hat{\boldsymbol{r}}$ over all time,

$$\frac{\mathrm{d}E_{\text{gw}}}{\mathrm{d}^2\hat{\boldsymbol{r}}} = \frac{r^2}{32\pi G} \int_{\mathbb{R}} \mathrm{d}t \, \dot{h}^{ij}\dot{h}_{ij} = \frac{r^2}{16\pi G} \int_{\mathbb{R}} \mathrm{d}t \, |\dot{h}|^2 = \frac{\pi r^2}{2G} \sum_{A=+,\times} \int_0^\infty \mathrm{d}f \, f^2 |\tilde{h}_A|^2. \tag{0.43}$$

Here we assume that the flux is measured at a radius r much larger than both the size of the source and the wavelengths of interest,[13] yet small enough that cosmological expansion can be neglected. We have assumed we are in vacuum, such that we are free to select the TT gauge and replace α and β by purely spatial indices. We have also removed the angle brackets, since the time integral averages over many GW periods for us. The final equality uses Parseval's theorem to convert the time integral into a frequency integral.

0.1.8 Gravitational-wave memory

Much of our basic intuition for the behaviour of GWs comes from examples of wavelike systems we are familiar with in our everyday lives, such as the propagation of ripples on the surface of a pond. These analogies encourage us to visualise GWs as shown in the left panel of figure 0.3: an incoming wave packet

---

[11] A detailed derivation of equation (0.41) using both approaches described here can be found in Maggiore [406].

[12] This approach requires us to go beyond the flat-background perturbation theory we have adopted so far, as we must allow the background spacetime to evolve dynamically to study how it responds to the energy-momentum carried by the GWs. In order for the distinction between the background and the GWs to be unambiguous, we require a separation of scales (either spatial or temporal) between the low-frequency/long-wavelength variations in the background metric and the high-frequency/short-wavelength perturbations due to the GWs. One can then develop an effective theory of the metric on large scales by 'integrating out' the GW perturbations, with equation (0.41) representing a renormalisation correction to the total energy-momentum tensor due to this coarse-graining.

[13] The leading-order piece of the GW strain we are interested in goes as $\sim 1/r$ at large distances, and is thus compensated by the factor of r^2 in equation (0.43). To get an invariant expression for the energy flux we must therefore choose r large enough that the next-to-leading order $\sim 1/r^2$ piece of the strain is negligible. Taking r much larger than the GW wavelength and the size of the source achieves this.

1. Burst without memory

2. Burst with memory
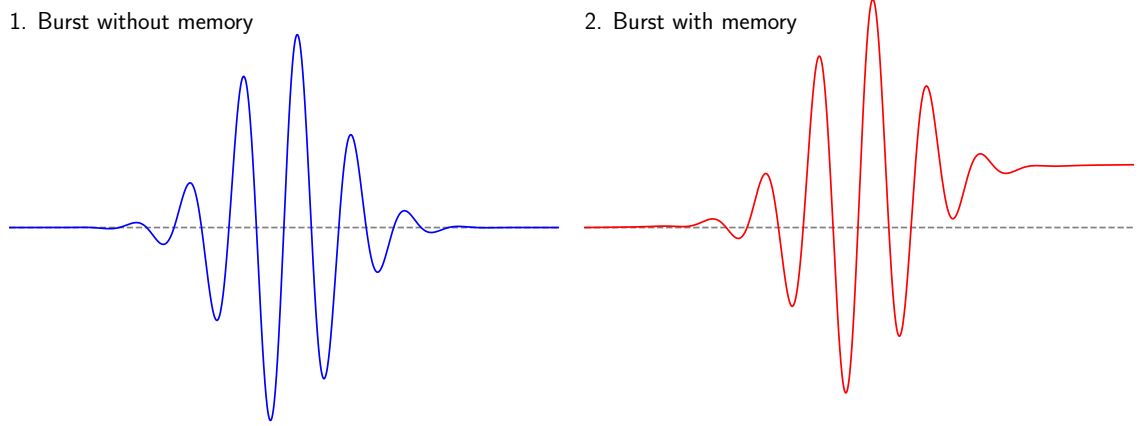
Figure 0.3: A cartoon illustration of the memory effect. In the left panel, a gravitational-wave burst passes and the strain returns to its initial value. In the right panel, a gravitational-wave burst passes and leaves a nonzero memory offset.

causes some initially quiescent quantity (e.g., the vertical displacement of a leaf floating on the pond, or the proper distance between two freely-falling test masses) to oscillate until it eventually returns to its starting value. Remarkably, this intuitive picture is often wrong in the case of GWs, as there are many situations in which the GW strain does *not* return to its initial value, but instead settles down with some fixed offset from its starting point (as illustrated in the right panel of figure 0.3), i.e.,

$$\Delta h_{ij} \equiv \lim_{t \to +\infty} h_{ij}(t) - \lim_{t \to -\infty} h_{ij}(t) \neq 0. \tag{0.44}$$

This corresponds to a permanent change in the GW measurement apparatus, and is therefore called the *GW memory effect* [151, 152, 265], as the apparatus 'remembers' the passage of the wave packet.

Though surprising at first, it is straightforward to derive this effect by solving the linearised EFE for sources whose energy-momentum tensor undergoes a permanent change between early and late times. For example, consider the GW signal (0.40) that we derived above for two slowly-moving test masses. We see immediately that if the relative velocity or acceleration of the two bodies is different at early and late times, then the resulting GW signal will exhibit a memory effect like that shown in figure 0.3. Such situations typically occur when the system is gravitationally unbound,[14] either in 'scattering-like' events where the masses follow a hyperbolic orbit, undergoing a permanent change in their velocities due to their gravitational interaction, or in 'explosion-like' events where the masses are initially bound, but become unbound due to some injection of energy and escape to infinity with nonzero velocities. These examples are both simplified descriptions of real astrophysical GW sources, such as stellar scattering [149, 379, 517, 550, 551, 595], core-collapse supernovae [162, 257, 378, 440, 552], and gamma-ray bursts [60, 127, 494, 503],

---

[14]Gravitationally-bound systems such as compact binaries do generate *some* linear memory if the merger product has a non-zero recoil velocity, but this linear memory signal is likely far too weak to be detected [262], even for systems with maximal recoil velocities (the so-called 'super-kick' configuration) [167].

as well as more idealised theoretical setups such as high-energy particle decay [68, 548, 549]. In both cases the GW memory associated with the signal (0.40) is given by

$$\Delta h_{ij} \simeq \Delta \frac{4G}{r} \mu (v_i v_j)^{\mathrm{TT}}, \tag{0.45}$$

where $\boldsymbol{v} \ll 1$ is the relative velocity, and we have neglected the acceleration term, which vanishes at $t \to \pm\infty$ in both examples. This expression is readily generalised to relativistic velocities and more than two particles, giving [542]

$$\Delta h_{ij} = \Delta \frac{4G}{r} \sum_a m_a \gamma_a \frac{(v_{a,i} v_{a,j})^{\mathrm{TT}}}{1 - \hat{\boldsymbol{r}} \cdot \boldsymbol{v}_a}, \tag{0.46}$$

where the index a labels the particles, $\gamma_a \equiv (1 - v_a^2)^{-1/2}$ is the Lorentz factor of particle a, and $\hat{\boldsymbol{r}}$ is the GW propagation direction.

   While GW memory effects associated with unbound systems have been studied since the 1970s [149, 257, 379, 517, 550–552, 595], it was not until 1991 that Christodoulou [201] pointed out a much more generic memory effect which accompanies essentially *any* source of GWs,[15] called the *nonlinear* memory (or sometimes, 'Christodoulou memory'),

$$\Delta h_{ij} = \frac{4G}{r} \int_{S^2} \mathrm{d}^2\hat{\boldsymbol{r}}' \, \frac{\mathrm{d}E_{\mathrm{gw}}}{\mathrm{d}^2\hat{\boldsymbol{r}}'} \frac{(\hat{r}'_i \hat{r}'_j)^{\mathrm{TT}}}{1 - \hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}'}. \tag{0.47}$$

(In contrast, the effect encapsulated by equation (0.46) is called the *linear* memory.) The reason the nonlinear memory effect went unnoticed for so long is that it is absent in the linearised theory, and was only discovered by Christodoulou using rigorous asymptotic methods [201] developed in his proof (with Klainerman) of the global nonlinear stability of Minkowski space [202]. (The same effect was also discovered soon after by Blanchet and Damour [129] in the very different context of post-Minkowskian theory.) We see that the nonlinear memory is sourced not by the momenta of unbound particles escaping to infinity as in equation (0.46), but instead by the GW energy radiated by the source. In fact, as Thorne [542] pointed out soon after Christodoulou's result, we can straightforwardly interpret equation (0.47) as the *linear* memory (0.46) associated with the gravitons radiated by the source. To see this, we need only make three small modifications to equation (0.46):

1. replace $m_a \gamma_a$ by E_a, the energy of each graviton (as measured in the observer's rest frame);

2. let $\boldsymbol{v}_a$ have unit magnitude (since gravitons propagate at the speed of light);

3. take the limit of infinitely many particles (i.e., the classical limit), so that the sum $\sum_a$ is replaced by an integral over a smooth angular distribution.

(The Δ symbol is also unnecessary on the right-hand side of equation (0.47), since we assume the GW flux from the source is zero in the distant past, $t \to -\infty$.) This interpretation makes clear why such an effect is only present in the full, nonlinear theory: it is due to the backreaction of gravitational perturbations on spacetime, a phenomenon which is neglected when working with a fixed, non-dynamical background.

---

[15] The only exception is when the integrand of equation (0.47) has vanishing TT component; e.g., when the radiated GW flux is exactly isotropic.
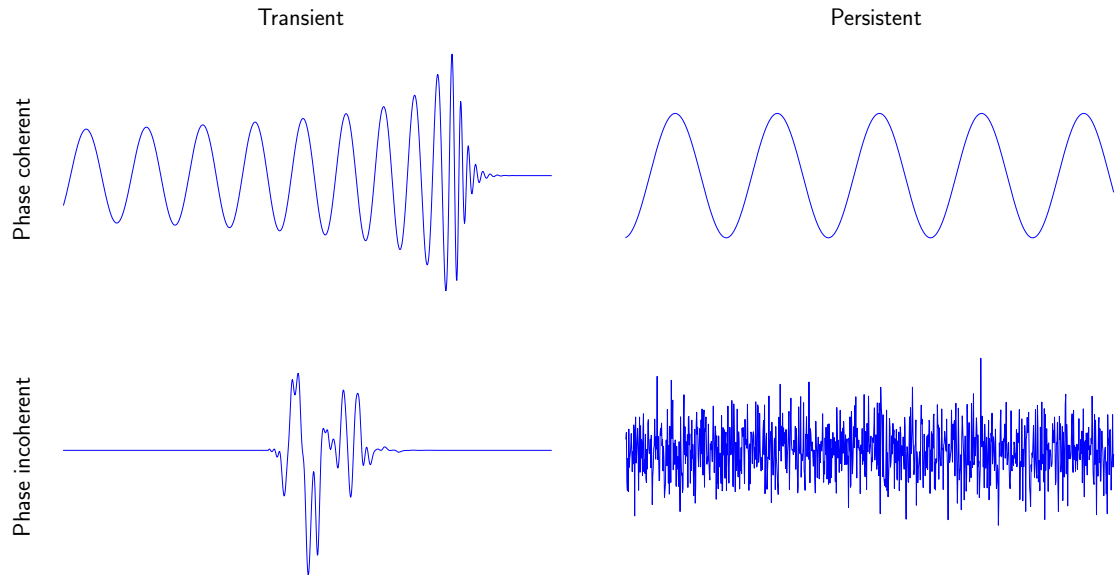
Figure 0.4: A taxonomy of four key gravitational-wave signal morphologies: phase-coherent transients such as compact binaries; persistent phase-coherent signals such as the 'continuous waves' emitted by spinning neutron stars; incoherent transients (i.e., bursts), such as those emitted by core-collapse supernovae; and persistent, incoherent signals—namely, the stochastic background.

While the memory effect was first discovered in the context of classical GR, it has been recognised in recent years that analogous memory effects are a generic feature of field theories with massless degrees of freedom [124, 448, 534, 548, 549]. Explicit examples that have been studied include 'electromagnetic memory' in electrodynamics [124, 168, 360, 534] and 'colour memory' in Yang-Mills theory [94, 168, 351, 352, 448], with the photon and the gluon playing the rôle of the graviton, respectively. Related work (principally conducted by Strominger and his collaborators) has shown that these memory effects are intimately related to both the asymptotic symmetries of the fields in each theory and to 'soft theorems' which govern the production of low-energy massless particles in the corresponding quantum theory [271, 308, 360, 446, 447, 530, 531]. These deep theoretical links add to the appeal of GW memory as a tool for understanding GR and probing the true nature of the gravitational field.

0.2 The gravitational-wave background

We have seen in section 0.1 that GR predicts the existence of two massless propagating degrees of freedom associated with TT perturbations to the gravitational field. These are generated by the motion of matter fields, meaning that the GW signals we receive on Earth can be used to infer the properties of distant objects in the Universe. In this section we turn our attention to a GW signal that is of particular interest to cosmologists: the stochastic GW background (SGWB) [67, 170, 200, 405, 470, 487]. We begin by discussing what distinguishes this signal from the compact binary signals detected thus far by LIGO/Virgo, before covering its statistical properties, how to model its expected amplitude, and how best to search for it in noisy data.

0.2.1 WHAT IS THE STOCHASTIC BACKGROUND?

Imagine we have constructed a GW detector (a few examples of which are described in section 0.4), and we switch it on. What kinds of signals might we expect to see? This simple question is incredibly important for GW astronomy, due to the fact that GWs are typically extremely faint and therefore difficult to distinguish from instrumental noise. Unlike in EM astronomy, where the signal is often very easily distinguishable from the noise (even instruments with very high noise levels, such as the naked eye, can easily detect many EM signals in the night sky), signal detection in GW astronomy relies on sophisticated statistical methods, with the best approach depending strongly on the morphology of the signal.

Of all the ways we might characterise a GW signal, there are two questions that are perhaps the most important:

1. Is the signal *transient*, or is it *persistent*? i.e., does the signal only appear in the detector for a relatively short time, or is it 'always on'?[16]

2. Is the signal *coherent*, or is it *incoherent*? i.e., are we able to deterministically model the phase of the signal, or does our lack of knowledge about the source force us to treat the phase as random?[17]

These two pairs of categories define four different kinds of GW signal (as illustrated in figure 0.4), each of which has its own set of specialised search methods.[18]

Phase-coherent signals are by far the easiest to search for, since one can use the technique of *matched filtering* (convolving the data with a template which encapsulates our knowledge of the signal); assuming a perfect template and Gaussian noise, this returns the maximum signal-to-noise ratio (SNR) of any possible search [573]. In the case of incoherent signals it is, by definition, impossible to construct a template, so this option is not available to us. Transient signals also present a key search advantage compared to persistent signals, as their arrival represents a change in the state of the detector; persistent signals, on the other hand, are harder to disentangle from the instrumental noise, which is also 'always on'.[19] It is unsurprising, then, that all of the GWs detected by LIGO/Virgo thus far originated from compact binaries; as we show in section 0.3.1, these are prime examples of transient, phase-coherent signals.

Consider, however, the situation depicted in figure 0.5, in which we have a cosmic population of sources distributed throughout the Universe; these could be compact binaries, or any other population of GW sources.[20] Since the strain signal from each of these sources falls off like $\sim 1/r$, any instrument will

---

[16]Note that 'relatively short' means relative to a typical observational timescale. A signal lasting for, say, several centuries, while certainly transient on astronomical timescales, would be treated as persistent for our purposes here.

[17]It is interesting to note that in EM astronomy, essentially *all* signals are incoherent; one rarely attempts to measure the phase of electromagnetic waves, only their intensity. This is because astronomical EM radiation is typically associated with the random, thermal motion of charges on microscopic scales, which is impossible to model in a phase-coherent way. Viewed through this lens, we see that the use of matched filtering in GW astronomy relies on the large-scale, coherent motion of macroscopic massive bodies (such as the components of a compact binary) to generate signals whose phase can be modelled accurately. One can see this as a consequence of the universally-attractive nature of gravity; an analogous system of coherently oscillating macroscopic charges would be very difficult to realise in the Universe due to the cancellation of positive and negative charges on smaller scales.

[18]As a result of this, these four categories correspond exactly to the four main data analysis working groups in the LIGO/Virgo/KAGRA Collaboration. For an introduction to some of the methods used by each group, see Creighton and Anderson [218] or Maggiore [406].

[19]On the other hand, of course, there are many advantages to searching for a signal one knows is always present. With transient searches, there is always a risk of missing rare and exciting signals due to changes in the state of the detector.

[20]In some contexts it makes sense to replace the notion of 'individual sources' here with 'individual Hubble patches' in which some cosmological GW production process takes place; e.g., the GWs generated by a first-order phase transition, which we
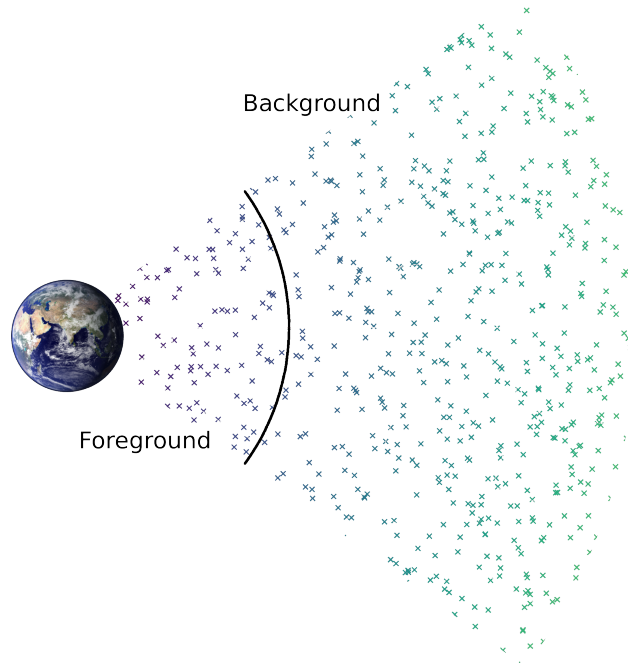
Figure 0.5: A cosmic population of gravitational-wave sources gives typically gives rise to a small number of signals that are individually resolvable (the 'foreground'), and a much larger number that are not (the 'background'). The combined emission from the latter gives rise to a persistent, incoherent signal that we call the stochastic background.

inevitably have a 'detection horizon' beyond which the signals are too faint to be distinguished from noise. We can also see in figure 0.5 that, so long as the number density of sources doesn't fall off too quickly, the total number of sources grows rapidly as a function of distance (since the size of a spherical shell of radius r grows like $\sim r^2$). As a result, we might have a situation where it is impossible to distinguish individual sources from each other past a certain distance due to the 'confusion noise' from many overlapping signals, even if these are all inside the detection horizon.

In both situations, matched filtering is no longer effective at detecting individual sources. But the GWs emitted by these sources are still present in the detector, and can still be searched for using other methods. Instead of focusing on individual sources, it becomes more useful to consider the entire population of sources together, treating their combined GW emission as one aggregate signal. Since these sources are typically very widely separated from each other, we expect them to be causally disconnected, meaning that there is no discernible pattern in the arrival times of each individual signal. This makes it impossible to deterministically model the phase of the aggregate signal, even if we have a perfect template for each individual signal, as we cannot predict how each of their independent phases will combine. It also means that we should expect the signal to be time-translation invariant (at least in a statistical sense), since there is no reason to expect that all of these distant, independent sources should conspire to emit their GWs at the same time. We therefore have a *persistent, incoherent* signal associated with essentially any cosmic

---

discuss below in section 0.3.2. The rest of the discussion here still holds in cases like this, with the caveat that typically *none* of the 'sources' are individually resolvable.

population of GW sources. We call this signal a *stochastic gravitational-wave background* (SGWB);[21] 'stochastic' because of its random, nondeterministic phase evolution, and 'background' because of its association with a large number of distant sources. In contrast, the small number of nearby, individually resolvable sources form what we might call a 'foreground'.[22]

As discussed above, persistent and incoherent signals like the SGWB are the most difficult to search for, since our ignorance of the phase evolution means we are unable to use matched filtering, and the lack of a distinct arrival time makes it hard to differentiate between the signal and the detector noise. So why should we bother? The answer can be seen in figure 0.5: being associated with distant sources, the SGWB can give us a glimpse of the Universe at *high redshift* and on *large scales*, revealing cosmological information that is inaccessible with individual nearby detections alone. We therefore see that the SGWB is an indispensable tool in our effort to probe cosmology and fundamental physics with GW observations.

In practice, the distinction between 'phase-coherent' and 'phase-incoherent' signals is somewhat blurry, particularly when the incoherence comes from a finite number of overlapping signals (e.g., in the compact binary case) [216]. If the number of such signals is sufficiently small, and if the phase evolution of each individual signal is well-known, then it is possible to use semi-coherent search methods that leverage this knowledge and outperform traditional stochastic searches—see, in particular, Smith and Thrane [518] for a Bayesian methodology that relies on the known phase evolution of BBH signals, as well as the relatively low expected rate of stellar-mass BBHs in the Universe. In recognition of this ambiguity, we will often drop the word 'stochastic' below, and simply refer to the GW signal we have described above as the *gravitational-wave background* (GWB). Our philosophy is that the words 'deterministic/stochastic', 'coherent/incoherent', etc., are best applied to GW *search methods*, and that different search methods with varying degrees of phase coherence may be suited to different GWB signals from different populations of sources. Notwithstanding this distinction, the remainder of this section (and indeed, of this thesis) is primarily concerned with phase-incoherent GWB searches, as these are simpler conceptually, and have thus far been applied much more successfully in practice due to computational challenges associated with semi-coherent Bayesian searches.

0.2.2 STATISTICAL PROPERTIES OF THE GRAVITATIONAL-WAVE BACKGROUND

Assuming we are unable to deterministically model the phase evolution of the GWB, the best we can hope to do is specify the *statistical* properties of the strain. Performing the plane wave expansion as in equation (0.27), we therefore treat the components $\tilde{h}_A(f, \hat{r})$ as random variables. In this section, we 'derive' a list of standard assumptions that are made about the probability distribution governing these random strain components.

---

[21] We will often refer to 'a' stochastic background, meaning the persistent, incoherent signal associated with one particular population of sources (e.g., compact binaries), as opposed to 'the' stochastic background, which refers to the superposition of all such stochastic signals, which is what we see in a detector. The question of how best to distinguish between multiple stochastic signals which might be simultaneously measured by a detector is an important open problem in GW astronomy [144, 146, 171, 272, 300, 415, 445, 471, 491, 513, 532], which we do not discuss further in this thesis.

[22] In some contexts the term 'foreground' has negative connotations, and implies that the signal in question is a nuisance factor which one must mitigate in order to observe the underlying target signal (e.g., polarised dust emission from the Milky Way acts as a nuisance foreground for CMB observatories [54]). The situation is different here, as both foreground and background are extremely scientifically valuable in GW astronomy.

The fact that the GWB is, by definition, composed of a large number of statistically independent sources allows us to invoke the central limit theorem and treat $\bar{h}_A(f, \hat{\boldsymbol{r}})$ as Gaussian. This reduces the problem of specifying an entire probability density function down to just specifying the first two moments,

$$\langle \bar{h}_A(f, \hat{\boldsymbol{r}}) \rangle, \qquad \langle \bar{h}_A(f, \hat{\boldsymbol{r}}) \bar{h}_{A'}^*(f', \hat{\boldsymbol{r}}') \rangle, \tag{0.48}$$

where $\langle \cdots \rangle$ here denotes an expectation value. This follows from Isserlis' theorem, which states that all higher-order moments are trivial for a set of Gaussian random variables.

Since $\bar{h}_A(f, \hat{\boldsymbol{r}})$ is a complex number, we can write it in terms of an amplitude and a complex phase (both of which are functions of frequency, propagation direction, and polarisation mode),

$$\bar{h}_A(f, \hat{\boldsymbol{r}}) = \mathcal{A}_A(f, \hat{\boldsymbol{r}}) e^{i\varphi_A(f, \hat{\boldsymbol{r}})}. \tag{0.49}$$

It is not difficult to convince oneself that the phase $\varphi_A(f, \hat{\boldsymbol{r}})$ should be uniformly distributed on $[0, 2\pi]$, and should be statistically independent of the amplitude $\mathcal{A}_A(f, \hat{\boldsymbol{r}})$. Indeed, this follows immediately from the assumption that the signal is statistically invariant under time translations $t \rightarrow t - \tau$ (i.e., the signal is *stationary*), since the definition of the Fourier transform implies that such a translation gives $\varphi_A(f, \hat{\boldsymbol{r}}) \rightarrow \varphi_A(f, \hat{\boldsymbol{r}}) - 2\pi f \tau$ while leaving the amplitude constant. We see that the probability distribution of $\varphi_A(f, \hat{\boldsymbol{r}})$ must be identical to that of $\varphi_A(f, \hat{\boldsymbol{r}}) + c$ (modulo 2π) for an arbitrary constant shift c, which can only be satisfied if the distribution is uniform. We also see that the amplitude is invariant even though the phase is shifted, meaning that the two must be statistically independent of each other. As a result, we see that the first moment of the GWB strain must vanish,

$$\langle \bar{h}_A(f, \hat{\boldsymbol{r}}) \rangle = \left\langle \mathcal{A}_A(f, \hat{\boldsymbol{r}}) e^{i\varphi_A(f, \hat{\boldsymbol{r}})} \right\rangle = \langle \mathcal{A}_A(f, \hat{\boldsymbol{r}}) \rangle \left\langle e^{i\varphi_A(f, \hat{\boldsymbol{r}})} \right\rangle = 0, \tag{0.50}$$

since $\langle e^{i\varphi} \rangle$ vanishes if φ is uniform.

In order to evaluate the second moment, we need to know about the two-point statistics of $\mathcal{A}_A(f, \hat{\boldsymbol{r}})$ and $\varphi_A(f, \hat{\boldsymbol{r}})$. Here we can use the distinguishing feature of the stochastic background, which is that the arrival times of GWs from each individual source are statistically independent due to each source being causally disconnected from the others. (Recall that this is the fundamental reason why we are unable to use matched filtering for stochastic signals.) This implies that the phases of any two plane-wave components are statistically independent from each other, with the only exception being when they have the same frequency, propagation direction, and polarisation state (since they are then just two copies of the same random variable),[23]

$$\mathrm{Cov}\left[e^{i\varphi_A(f, \hat{\boldsymbol{r}})}, e^{i\varphi_{A'}(f', \hat{\boldsymbol{r}}')} \right] = \left\langle e^{i[\varphi_A(f, \hat{\boldsymbol{r}}) - \varphi_{A'}(f', \hat{\boldsymbol{r}}')]} \right\rangle \propto \delta_{AA'} \delta(f - f') \delta^{(2)}(\hat{\boldsymbol{r}}, \hat{\boldsymbol{r}}'). \tag{0.51}$$

---

[23] One might object that cosmological sources of GWs *can* be causally connected if the Universe underwent a period of inflation at early times which shrank the comoving size of the cosmological horizon; this is the leading explanation for why different Hubble patches in the CMB have such similar temperatures. (Indeed, this 'horizon problem' was one of the original theoretical motivations for inflation [584].) However, Margalit *et al.* [412] have shown that even if phase-coherent correlations do exist in the GW signal at early times, this coherence is lost as the GWs propagate across the Universe, as each GW experiences a large phase shift $\delta\varphi \gg 2\pi$ due to inhomogeneities in the gravitational potential along its line of sight. This argument holds for GW frequencies $f \gtrsim 10^{-12}$ Hz, which covers the entirety of the frequency range we are interested in here, meaning that we can safely assume zero phase correlation between different lines of sight.

As a result, we have

$$\left\langle \bar{h}_A(f,\hat{\boldsymbol{r}})\bar{h}^*_{A'}(f',\hat{\boldsymbol{r}}')\right\rangle = \left\langle \mathcal{A}_A(f,\hat{\boldsymbol{r}})\mathcal{A}_{A'}(f',\hat{\boldsymbol{r}}')\right\rangle\left\langle e^{i[\varphi_A(f,\hat{\boldsymbol{r}})-\varphi_{A'}(f',\hat{\boldsymbol{r}}')]}\right\rangle$$
$$\propto \left\langle [\mathcal{A}_A(f,\hat{\boldsymbol{r}})]^2\right\rangle\delta_{AA'}\delta(f-f')\delta^{(2)}(\hat{\boldsymbol{r}},\hat{\boldsymbol{r}}').$$
(0.52)

(Note that we have made no assumptions about whether the amplitudes of different sources are statistically independent or not.)

The two Dirac delta functions in equation (0.52) are somewhat surprising at first, as they seem to imply that the variance of each individual plane wave component, $\left\langle|\bar{h}_A(f,\hat{\boldsymbol{r}})|^2\right\rangle$, is infinite. In fact, this is only true in the idealised situation where we have infinitely fine frequency resolution δf and angular resolution $\delta\hat{\boldsymbol{r}}$. In practice, we are only able to resolve a finite set of discrete frequency bins f_i and sky pixels $\hat{\boldsymbol{r}}_i$, so that the Dirac deltas are replaced by Kronecker deltas,

$$\delta(f_i-f_j)\to\frac{\delta_{ij}}{\delta f}, \qquad \delta^{(2)}(\hat{\boldsymbol{r}}_i,\hat{\boldsymbol{r}}_j)\to\frac{\delta_{ij}}{\delta\hat{\boldsymbol{r}}}.$$
(0.53)

However, since we are usually only interested in integrals over the second moment (0.52) (e.g., in equation (0.57) below) we can safely keep the Dirac deltas in equation (0.52), with the understanding that they represent an idealised continuum limit.

We have thus determined the second moment up to some unknown deterministic function of frequency, sky direction, and polarisation,

$$\left\langle \bar{h}_A(f,\hat{\boldsymbol{r}})\bar{h}^*_{A'}(f',\hat{\boldsymbol{r}}')\right\rangle = S_A(f,\hat{\boldsymbol{r}})\delta_{AA'}\delta(f-f')\delta^{(2)}(\hat{\boldsymbol{r}},\hat{\boldsymbol{r}}').$$
(0.54)

At this point it is useful to recall from section 0.1.4 that the amplitudes of the two polarisation modes are not fixed for a given plane wave, but undergo mixing when ones rotates around the line of sight, as in equation (0.23). If we assume the polarisation angle of each source, $\psi(f,\hat{\boldsymbol{r}})$, is uniformly distributed, then by a very similar argument to that above we conclude that this unknown function must be equal for each polarisation mode,

$$S(f,\hat{\boldsymbol{r}})\equiv S_+(f,\hat{\boldsymbol{r}})=S_\times(f,\hat{\boldsymbol{r}}),$$
(0.55)

since we can arbitrarily rotate +-modes into ×-modes and vice versa. This assumption of a uniform distribution in $\psi(f,\hat{\boldsymbol{r}})$ is fairly robust, particularly for astrophysical sources, since there is generally no reason to expect sources to be preferentially aligned along any particular direction around the line of sight. Note, however, that there are several models of the early Universe that give rise to a GWB with nontrivial polarisation content, for example through couplings to parity-violating matter fields [51, 52, 64, 98, 239, 401], the addition of parity-violating terms to the gravitational action [63, 105, 211, 499, 536, 576], or the generation of helical magnetohydrodynamic (MHD) turbulence in the primordial plasma [173, 252, 354, 374]. Specialised search methods have been developed for polarised backgrounds such as these [242, 253, 414, 507, 508, 510, 511, 519]; however, we will not discuss these further here, since the unpolarised assumption (0.55) is expected to be valid for all of the GW sources we investigate in this thesis.
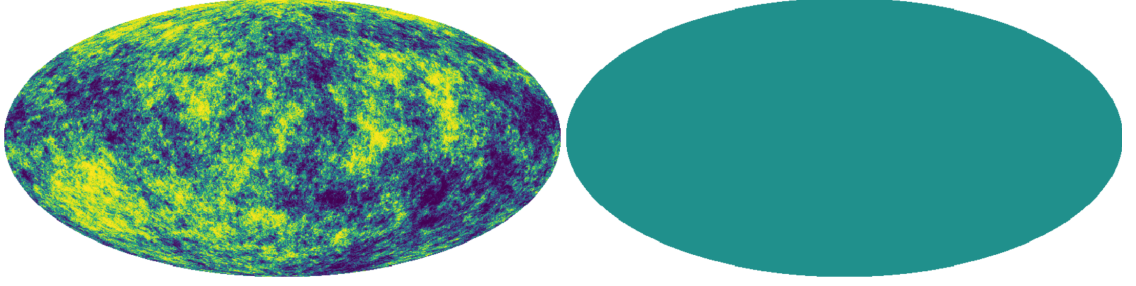
Figure 0.6: Left panel: a statistically isotropic field on the sphere. Right panel: an *exactly* isotropic field on the sphere. Most stochastic background searches assume the latter.

The final assumption that is usually made in describing the statistics of the GWB is that its intensity is *isotropic*, i.e., the same in all directions on the sky. This implies that the function $S(f, \hat{\boldsymbol{r}})$ characterising the second moment of the GWB strain is independent of sky direction, and therefore depends only on frequency. The motivation for this assumption comes from the statistical isotropy of many other cosmological observables; the CMB, for instance, varies in intensity by only a few parts in 10^5 across the sky. It is important to point out, however, that *statistical* isotropy (i.e., invariance of expectation values of the observed field under sky rotations) is not the same as *exact* isotropy (i.e., invariance of *the field itself* under rotations), which is being assumed here; see figure 0.6. Indeed, modern cosmology has forcefully demonstrated that, even if departures from exact isotropy and homogeneity are small, they are incredibly important tools for understanding the structure and evolution of the Universe. In this sense, isotropy is the least well-justified of all the assumptions we have made thus far. In chapter 1 of this thesis we revisit this assumption, and investigate what can be learned by considering *anisotropies* in the GWB.

0.2.3 The gravitational-wave density parameter

To summarise, we have shown that by assuming the GWB to be Gaussian, stationary, unpolarised, and isotropic, with no nontrivial phase correlations, we can write

$$\left\langle \bar{h}_A(f, \hat{\boldsymbol{r}}) \right\rangle = 0, \qquad \left\langle \bar{h}_A(f, \hat{\boldsymbol{r}}) \bar{h}_{A'}^*(f', \hat{\boldsymbol{r}}') \right\rangle = S(f) \delta_{AA'} \delta(f - f') \delta^{(2)}(\hat{\boldsymbol{r}}, \hat{\boldsymbol{r}}'), \qquad (0.56)$$

such that the strain statistics are fully characterised by a single function of frequency, $S(f)$. The fact that $h_{ij}(t, \boldsymbol{x})$ is real implies that $S(f)$ is real and even-valued, $S(f) = S^*(f) = S(-f)$.

This function is not arbitrary; it is directly related to the energy spectrum of the GWB. To see this, we insert the plane-wave expansion (0.27) and the GWB moments (0.56) into the expression for the

quasi-local GW energy density we obtained previously, equation (0.42). Placing our detector at the origin $\boldsymbol{x} = \boldsymbol{0}$ (which we assume to be effectively in vacuum, such that we can work in the TT gauge), this gives

$$
\begin{aligned}
\rho_{\mathrm{gw}} &= \frac{1}{32\pi G}\langle \dot{h}^{ij}\dot{h}_{ij}\rangle \\
&= \frac{\pi}{8G}\int_{\mathbb{R}}\mathrm{d}f\int_{\mathbb{R}}\mathrm{d}f'\int_{S^2}\mathrm{d}^2\hat{\boldsymbol{r}}\int_{S^2}\mathrm{d}^2\hat{\boldsymbol{r}}'\,e^{2\pi i t(f-f')}e^{A,ij}(\hat{\boldsymbol{r}})e_{ij}^{A'}(\hat{\boldsymbol{r}}')ff'\langle \bar{h}_A(f,\hat{\boldsymbol{r}})\bar{h}_{A'}^*(f',\hat{\boldsymbol{r}}')\rangle \\
&= \frac{\pi}{8G}\int_{\mathbb{R}}\mathrm{d}f\int_{S^2}\mathrm{d}^2\hat{\boldsymbol{r}}\,e^{A,ij}(\hat{\boldsymbol{r}})e_{ij}^{A}(\hat{\boldsymbol{r}})f^2 S(f) = \frac{4\pi^2}{G}\int_{\mathbb{R}}\mathrm{d}(\ln f)\,f^3 S(f),
\end{aligned}
\tag{0.57}
$$

where in the last line we have used equation (0.21), and have restricted ourselves to positive frequencies (without loss of generality), so that $\ln f$ is well-defined.[24] We therefore find

$$
S(f) = \frac{G}{4\pi^2 f^3}\frac{\mathrm{d}\rho_{\mathrm{gw}}}{\mathrm{d}(\ln f)},
\tag{0.58}
$$

so that the strain statistics of the GWB are uniquely determined by the frequency spectrum of the GW energy density.

Since the GWB is a cosmological observable, it is often convenient to normalise its energy density with respect to the present-day critical energy density of the Universe (i.e., the energy density required for the Universe to be spatially flat). We therefore define the dimensionless *GWB density parameter* (or simply *GWB spectrum*),

$$
\Omega_{\mathrm{gw}}(f) \equiv \frac{1}{\rho_{\mathrm{c}}}\frac{\mathrm{d}\rho_{\mathrm{gw}}}{\mathrm{d}(\ln f)}, \qquad \rho_{\mathrm{c}} \equiv \frac{3H_0^2}{8\pi G},
\tag{0.59}
$$

with the total normalised energy density in GWs being given by $\int_{\mathbb{R}}\mathrm{d}(\ln f)\,\Omega_{\mathrm{gw}}(f)$. This brings the GWB in line with how cosmologists describe other sources of energy density in the Universe; see, e.g., table 0.1. (When there is no chance of confusion with other density parameters we will omit the 'gw' subscript, and simply write $\Omega(f)$.) Equation (0.56) thus becomes

$$
\langle \bar{h}_A(f,\hat{\boldsymbol{r}})\rangle = 0, \qquad \langle \bar{h}_A(f,\hat{\boldsymbol{r}})\bar{h}_{A'}^*(f',\hat{\boldsymbol{r}}')\rangle = \frac{3H_0^2\Omega_{\mathrm{gw}}(f)}{32\pi^3|f|^3}\delta_{AA'}\delta(f-f')\delta^{(2)}(\hat{\boldsymbol{r}},\hat{\boldsymbol{r}}').
\tag{0.60}
$$

As we will see below, the values of $\Omega_{\mathrm{gw}}(f)$ predicted by most models of the GWB (and indeed, the values allowed by current observations) are much smaller than any of those in table 0.1, even those of photons and neutrinos. This shows that GW cosmology is 'doubly hard': not only are GWs difficult to detect due to how weakly they couple to our detectors, but they are also not very abundant in the Universe, making up a minuscule fraction of the cosmic energy budget. However, as we will see in subsequent

---

[24]Note that we have indulged in a slight abuse of notation here: we have treated the angle brackets $\langle \cdots \rangle$ in equation (0.42) identically to those in equation (0.56), even though the former represent a spatial or temporal average of a single GWB realisation over a region of spacetime, while the latter represent an ensemble average over many random realisations of the GWB at a single point in spacetime. We can interchange these two averaging processes if we assume that the GWB is *ergodic* such that, given sufficient time or a large enough spatial volume, the number of occurrences of a given value of the GWB strain is proportional to the probability assigned to that value by the distribution in equation (0.56). Ergodic processes are ubiquitous in physics whenever we have a configuration space of finite size—as is the case here, since the GWB strain cannot become arbitrarily large.

Source of energy density, X	Density parameter, $\Omega_X \equiv \rho_X/\rho_c$
Dark energy (Λ)	0.689
Cold dark matter (CDM)	0.261
Baryonic matter	0.0490
Photons (γ)	0.0000543
Neutrinos (ν)	0.0000375
All matter (baryons + CDM)	0.311
All radiation ($\gamma + \nu$)	0.0000918
Total	1.00

Table 0.1: Present-day cosmological density parameters of the main components of the ΛCDM model, as determined by *Planck* 2018, TTTEEE+lowE+lensing+BAO [55]. We assume the Standard Model value of the effective number of neutrino species, $N_{\text{eff}} = 3.045$ [496].

chapters, the challenges involved in detecting the GWB are more than compensated by the immense scientific value such a detection would have.

It is worth noting briefly that the energy density we have chosen to normalise against, ρ_c, depends on a quantity whose value has historically been somewhat poorly determined: the Hubble constant, H_0. For this reason, some authors characterise the GWB in terms of $h^2 \Omega_{\text{gw}}$ rather than just Ω_{gw}, where h is defined by $H_0 = h \times 100\,\text{km\,s}^{-1}\,\text{Mpc}^{-1}$; this then has the advantage of being insensitive to uncertainties in H_0. We choose not to do this here. Instead, we assume the *Planck* 2018 value $h = 0.677$ whenever quoting values for Ω_{gw}. While this is discrepant with the value $h = 0.732$ inferred from local distance ladder measurements [482] (the infamous 'Hubble tension'—see, e.g., Verde *et al.* [563]), the resulting change in Ω_{gw} is only a factor of ≈ 1.17, which is negligible in most situations.

0.2.4 MODELLING GRAVITATIONAL-WAVE BACKGROUNDS

Given a model for a particular cosmic population of GW sources, we can calculate the corresponding density parameter $\Omega(f)$ as follows. First, consider the GW energy density we observe today due to a single source, which is equal to the rate of GW energy passing through a spatial 2-surface surrounding us per unit surface area. In a Euclidean Universe, we can write this as

$$\rho = \frac{1}{4\pi r^2} \frac{\mathrm{d}E}{\mathrm{d}t}, \tag{0.61}$$

where r is the distance to the source, and $\mathrm{d}E/\mathrm{d}t$ is the rate at which it emits energy in GWs (which we have implicitly averaged over inclination). In a FLRW Universe, this expression still holds so long as we replace the Euclidean distance r by the luminosity distance d_L, and replace the global time t by the time in the source's rest frame, t_s. Both d_L and t_s are functions of redshift, z.

For an individual signal, $\mathrm{d}E/\mathrm{d}t_s$ is an arbitrary function of time. However, when we combine a large number of signals, we expect this energy emission rate to reduce to its time-averaged value (due to

ergodicity, see footnote 24), which is just the product of the mean rate of signals[25] per unit source time, R, and the total energy emitted by each signal, E, so that $\mathrm{d}E/\mathrm{d}t_\mathrm{s} \to RE$. If we focus on just the GWs emitted in a particular logarithmic frequency bin, then (0.61) becomes

$$\frac{\mathrm{d}\rho}{\mathrm{d}(\ln f)} = \frac{R}{4\pi d_L^2} \frac{\mathrm{d}E}{\mathrm{d}(\ln f_\mathrm{s})}, \tag{0.62}$$

where we distinguish between the observed frequency f and the source-frame frequency $f_\mathrm{s} = (1 + z) \times f$. (Note that we would pick up a factor of $1 + z$ here if we were using linear frequency bins, but redshifting does not change the size of a logarithmic bin.)

Now we go from the time-averaged contribution from a single source to the total contribution from all sources at all redshifts. Since the signal rate R will, in general, vary over cosmological timescales, it is more useful to think in terms of the *comoving rate density* $\mathcal{R}(z)$, which is the rate of signals per unit comoving volume. We can then encapsulate the contributions from all sources by computing a volume integral,

$$\frac{\mathrm{d}\rho}{\mathrm{d}(\ln f)} = \int \mathrm{d}\mathcal{V} \, \frac{\mathcal{R}(z)}{4\pi d_L^2(z)} \frac{\mathrm{d}E}{\mathrm{d}(\ln f_\mathrm{s})}, \tag{0.63}$$

where $\mathrm{d}\mathcal{V}$ is the comoving volume element, which can also be written as $\mathrm{d}\mathcal{V} = 4\pi r^2 \, \mathrm{d}z \, / H(z)$ with $r(z)$ the comoving distance. Inserting this into equation (0.63), and using the fact that $d_L/r = 1 + z$, we therefore find

$$\Omega(f) = \frac{1}{\rho_\mathrm{c}} \int_0^\infty \frac{\mathrm{d}z}{(1 + z)^2} \frac{\mathcal{R}(z)}{H(z)} \frac{\mathrm{d}E}{\mathrm{d}(\ln f_\mathrm{s})}. \tag{0.64}$$

This result (sometimes called the 'Phinney formula' [456]) allows us to calculate the expected GWB spectrum from a cosmic population of sources. All we need to know is the energy spectrum radiated by each source, the comoving rate density of sources, and the expansion rate of the cosmological background. Often the first two of these three quantities will depend on some set of parameters $\boldsymbol{\zeta}$ describing the sources (e.g., the masses of compact binaries), in which case the formula becomes

$$\Omega(f) = \frac{1}{\rho_\mathrm{c}} \int_0^\infty \frac{\mathrm{d}z}{(1 + z)^2} \int \mathrm{d}\boldsymbol{\zeta} \, \frac{\mathcal{R}(z, \boldsymbol{\zeta})}{H(z)} \frac{\mathrm{d}E(\boldsymbol{\zeta})}{\mathrm{d}(\ln f_\mathrm{s})}. \tag{0.65}$$

0.2.5 Stochastic search methods

As mentioned above, persistent and incoherent signals such as the GWB are arguably the most challenging to search for, since we cannot use matched filtering, and there is no distinct arrival time to distinguish the signal from instrumental noise. Faced with these limitations, how does one go about searching for the GWB?

Cross-correlation searches

The most powerful approach that has been developed thus far is to conduct *cross-correlation* searches between two or more detectors [69, 199, 270, 487]. The key insight here is that even if the GWB is

---

[25] In using this language, we appear to be restricting ourselves to transient signals. However, we can easily accommodate individual sources that are persistent (e.g., spinning neutron stars) by letting E be the total energy radiated by a persistent source in some fixed time interval τ, and setting R equal to the number of such sources divided by τ.

indistinguishable from noise in any individual detector, two different detectors will always 'see' the same GWB signal, while ideally having statistically independent sources of noise. We can therefore use the output of one detector as a (noisy) template for the signal we expect to see in the other detector.

More quantitatively, let us write the data stream $d(t)$ in each detector as a linear sum of the GWB strain $h(t)$ and some random noise process $n(t)$. The expectation value of the cross-correlation statistic $d_1 d_2$ can then be written as a sum of four terms,

$$\langle d_1 d_2 \rangle = \langle (h_1 + n_1)(h_2 + n_2) \rangle = \langle h_1 h_2 \rangle + \langle h_1 n_2 \rangle + \langle n_1 h_2 \rangle + \langle n_1 n_2 \rangle, \qquad (0.66)$$

with the subscripts labelling the two detectors. We expect the second and third of these terms to vanish, since there is no reason for the detector noise to have any coherence with the GW signal. Similarly, the fourth term should vanish if each detector is subject to statistically-independent sources of noise.[26] This leaves just the first term, which, heuristically, is proportional to the GWB density parameter, $\langle d_1 d_2 \rangle = \langle h_1 h_2 \rangle \propto \Omega$.

Of course, we don't have observational access to ensemble-averaged quantities, only to individual random realisations. However, since both the signal and noise are ergodic (see footnote 24), we can approximate the ensemble average by summing over many successive measurements of the cross-correlation statistic made at different times. As the number of measurements, N_t, goes to infinity, we recover exactly the ensemble average,

$$\lim_{N_t \to \infty} \frac{1}{N_t} \sum_{i=1}^{N_t} d_1(t_i) d_2(t_i) = \langle d_1 d_2 \rangle = \langle h_1 h_2 \rangle. \qquad (0.67)$$

Since N_t is finite in practice, the cross-correlation statistic will have some random scatter around its expectation value; this is the fundamental limitation on our ability to measure the GWB. Assuming that the signal and the noise are both Gaussian, and that measurements at different times are independent and identically distributed (i.i.d.), this scatter is given by

$$\begin{aligned}
\mathrm{Var}\left[\frac{1}{N_t} \sum_{i=1}^{N_t} d_1(t_i) d_2(t_i) \right] &= -\langle d_1 d_2 \rangle^2 + \frac{1}{N_t^2} \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} \langle d_1(t_i) d_2(t_i) d_1(t_j) d_2(t_j) \rangle \\
&= -\langle h_1 h_2 \rangle^2 + \frac{1}{N_t^2} \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} \Bigg[\langle d_1(t_i) d_2(t_i) \rangle \langle d_1(t_j) d_2(t_j) \rangle + \langle d_1(t_i) d_1(t_j) \rangle \langle d_2(t_i) d_2(t_j) \rangle \\
&\qquad\qquad\qquad\qquad\qquad + \langle d_1(t_i) d_2(t_j) \rangle \langle d_1(t_j) d_2(t_i) \rangle \Bigg] \\
&= -\langle h_1 h_2 \rangle^2 + \frac{1}{N_t^2} \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} \left[\langle h_1 h_2 \rangle^2 + \delta_{ij} \left(\langle n_1^2 \rangle + \langle h_1^2 \rangle \right) \left(\langle n_2^2 \rangle + \langle h_2^2 \rangle \right) + \delta_{ij} \langle h_1 h_2 \rangle^2 \right] \\
&= \frac{1}{N_t} \left[\left(\langle n_1^2 \rangle + \langle h_1^2 \rangle \right) \left(\langle n_2^2 \rangle + \langle h_2^2 \rangle \right) + \langle h_1 h_2 \rangle^2 \right] \simeq \frac{1}{N_t} \langle n_1^2 \rangle \langle n_2^2 \rangle,
\end{aligned}$$

$$(0.68)$$

---

[26]Note that this is not always the case. In particular, any pair of ground-based interferometers, no matter how widely separated, are expected to possess some level of correlated noise due to *Schumann resonances* [199, 424, 544]: coherent electromagnetic fields in the cavity between the Earth's surface and the ionosphere that are continuously excited by lightning strikes.

where we have used Isserlis' theorem to rewrite the fourth moment in terms of products of second moments. In the final line we have assumed we are in the 'weak-signal' regime such that $|h| \ll |n|$, as is the case for essentially all GW experiments. We can then calculate a signal-to-noise ratio (SNR) for the cross-correlation search by dividing the GWB term by the square-root of equation (0.68),

$$\mathrm{SNR} \simeq \sqrt{N_t} \, \frac{\langle h_1 h_2 \rangle}{\sqrt{\langle n_1^2 \rangle \langle n_2^2 \rangle}}. \tag{0.69}$$

There are two comments to make about equation (0.69). First, we see that the cross-correlated GWB power is divided by the geometric mean of the auto-correlated noise power in each detector. In the weak-signal regime, this factor is very small. However, since N_t is proportional to the total observing time T_{obs}, we see that $\mathrm{SNR} \propto \sqrt{T_{\mathrm{obs}}}$. This is the key reason that cross-correlation searches are so useful: in principle, *any* pair of detectors can eventually measure the GWB with arbitrarily large SNR, so long as we observe for a long enough time.[27]

THE OVERLAP REDUCTION FUNCTION

There are many complications that we have swept under the carpet here (see Romano and Cornish [487] for an authoritative treatment), but one which is important to highlight is the precise relationship between the GW strain cross-correlation $\langle h_1 h_2 \rangle$ and the GW density parameter $\Omega(f)$. If the two detectors are co-located, then these quantities really are proportional to each other,[28]

$$\langle h_1(t) h_2(t) \rangle = \frac{3 H_0^2}{4\pi^2} \int_0^\infty \frac{\mathrm{d}f}{f^3} \Gamma_{12} \Omega(f). \tag{0.70}$$

Here Γ_{12} is a proportionality constant which encodes the coupling of each detector's strain readout $h(t)$ to the actual GW strain $h_{ij}(t)$. Defining the *detector response functions* $\mathcal{D}^A(\hat{r})$ through the response of a detector to each polarisation mode of a GW plane wave propagating in the $\hat{r}$-direction (and assuming we are in the small-antenna limit),

$$h(t) \equiv \mathcal{D}^A(\hat{r}) h_A(t, \hat{r}), \tag{0.71}$$

we find that the proportionality constant is

$$\Gamma_{12} = \frac{1}{8\pi} \int_{S^2} \mathrm{d}^2 \hat{r} \, \mathcal{D}_1^A(\hat{r}) \mathcal{D}_2^A(\hat{r}). \tag{0.72}$$

For the simplest case of two identical, equal-arm, co-located and co-aligned interferometers, this is given by $\Gamma_{12} = \sin^2 \beta / 5$, where β is the opening angle between the arms of each interferometer.

---

[27] In practice, if the noise power is too large compared to the GWB, then the requisite time might be far too long for such a measurement to be feasible. For this reason, any time spent upgrading the sensitivity of the detector is often more than worth the consequent loss of observing time.

[28] This quantity isn't really what is measured in cross-correlation searches (which usually work with frequency-domain data $\tilde{d}(f)$ instead), and even if it was, we're missing a factor in the integrand to account for the loss of sensitivity of the instrument outside of the small-antenna regime, where equation (0.71) breaks down. However, these details are unimportant for our purpose here, which is to illustrate the origin of the overlap reduction function.
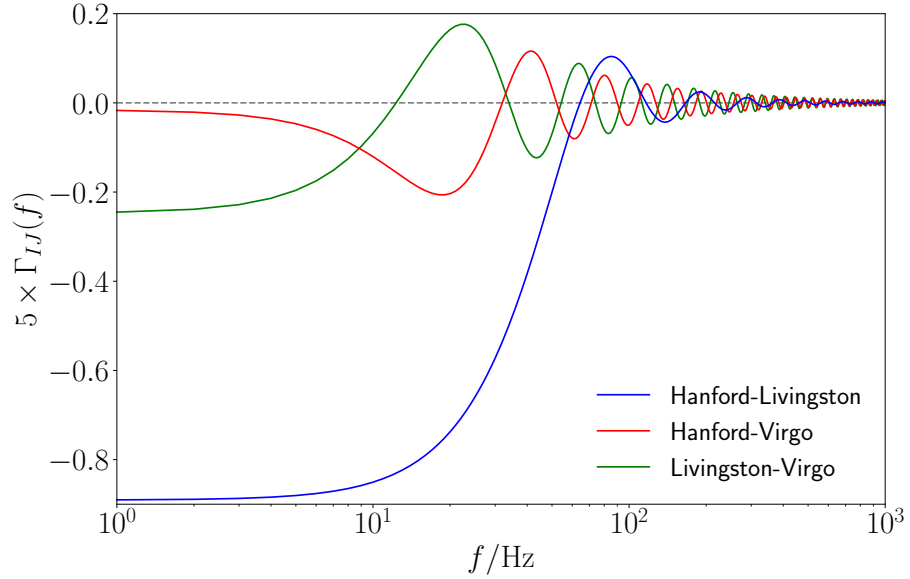
Figure 0.7: Overlap reduction functions for the three detectors that make up the LIGO/Virgo network. The factor of 5 ensures that $5 \times \Gamma_{IJ}(f) = 1$ for two co-located, co-aligned detectors in the small-antenna limit. At low frequencies the Hanford-Livingston ORF is approximately $5 \times \Gamma_{\mathrm{HL}} \approx -0.89$; this is close to -1, due to the two LIGO interferometers being nearly co-aligned up to a 90° rotation (there is a slight misalignment due to the curvature of the Earth between the two detector sites, which causes $|5 \times \Gamma_{\mathrm{HL}}|$ to drop below unity). Each pair of detectors has a series of 'blind spots' set by their separation, with longer baselines corresponding to lower-frequency blind spots.

The issue with cross-correlating between two co-located detectors is that they both experience the same sources of environmental noise, making it extremely difficult to ensure that $\langle n_1 n_2 \rangle$ vanishes.[29] We are therefore usually interested in cross-correlating between widely-separated pairs of interferometers such as LIGO Hanford and LIGO Livingston, which are situated about 3000 km apart. In this case, however, Γ_{12} becomes a frequency-dependent transfer function,

$$\Gamma_{12}(f) = \frac{1}{8\pi} \int_{S^2} \mathrm{d}^2\hat{\boldsymbol{r}}\, \mathcal{D}_1^A(\hat{\boldsymbol{r}}) \mathcal{D}_2^A(\hat{\boldsymbol{r}}) \cos[2\pi f \hat{\boldsymbol{r}} \cdot (\boldsymbol{x}_1 - \boldsymbol{x}_2)], \qquad (0.73)$$

called the *overlap reduction function* (ORF). The cosine in the integrand here means that this function is generally smaller than the constant Γ_{12} we found for the co-located case. This is because the two detectors, being at different locations, are generally out of phase in their response to a given plane wave, and this phase difference causes a loss of coherence between their strain readouts. In fact, $\Gamma_{12}(f)$ usually oscillates between positive and negative values as a function of f, meaning that there are a series of frequencies f_* set by the inverse separation $|\boldsymbol{x}_1 - \boldsymbol{x}_2|^{-1}$ where there is, on average, *no* coherence between the two detectors, $\Gamma_{12}(f_*) = 0$ (see figure 0.7). Since we can only measure the combination $\Gamma_{12}(f)\Omega(f)$, these frequencies represent 'blind spots' in the search, at which it is impossible to measure $\Omega(f)$. (Note however that this can be mitigated by combining multiple pairs of detectors, each with different pairwise separations and therefore different blind spots.)

---

[29]Nonetheless, GWB searches have been conducted with co-located detectors in the past; in particular, Aasi *et al.* [1] carried out such a search using two interferometers at the LIGO Hanford site, H1 and H2. (H2 has since been decommissioned.)

Excess-power searches

What if we only have access to a single detector? In principle, we can still carry out a GWB search if we have some prior information about the amplitude or spectral shape of the detector noise. If the measured amplitude is greater than expected, then we can try to attribute the excess power to a GWB signal. This is the exact method that Penzias and Wilson [453] used in their groundbreaking first detection of the CMB, which relied on ruling out all other possible sources of noise in their radio antenna. This method is very difficult to apply in GWB searches, however, since a detection would require us to determine the noise spectrum of our detector with an uncertainty much smaller than $\langle h^2 \rangle$. Since $|h| \ll |n|$, this is practically impossible in most cases. One important exception is the future space-based interferometer LISA, for which it will be possible to construct a *null channel* which is (approximately) insensitive to GWs, giving a direct measurement of the noise power [46].

Sensitivity curves

Whether one is conducting a cross-correlation search in a network of detectors, or an excess-power search in a single detector, the ultimate goal is to reconstruct the GWB spectrum $\Omega(f)$ within some frequency band set by the sensitivity of the experiment. In both cases, however, this is often a very difficult task; for cross-correlation searches, we have seen that the ORF generally causes 'blind spots' which make it difficult to reconstruct certain frequencies, while for excess-power searches a lack of knowledge about the spectral shape of the GWB makes it even harder to distinguish signal power from noise power. Fortunately, in many situations the GWB spectrum is sufficiently slowly-varying across the frequency range of the search that we can model it with a simple power law,

$$\Omega(f) = \Omega_\alpha (f/f_{\text{ref}})^\alpha, \tag{0.74}$$

where α is the spectral index and Ω_α is the amplitude at some reference frequency f_{ref}. The case $\alpha = 0$ corresponds to a scale-invariant spectrum, while $\alpha = 3$ corresponds to white noise (since $\langle h^2(t) \rangle$ then receives equal contributions from each frequency bin; c.f. equation (0.70)). As we will see in section 0.3 below, the GWB from inspiralling compact binaries is described by $\alpha = 2/3$, while cosmic strings have $\alpha = 0$ at high frequencies, and first-order phase transitions are typically described by a *broken* power law that interpolates between $\alpha = 3$ and $\alpha = -4$.

Using equation (0.74) massively reduces the complexity of the search, as we are now attempting to infer just two parameters for the whole spectrum, rather than one free parameter for every frequency bin. We can even reduce this further to just one parameter by assuming a particular value of α and attempting to infer the corresponding value of Ω_α; this approach forms the basis of the stochastic searches conducted by the LIGO/Virgo/KAGRA Collaboration, which typically focus on $\alpha = 0$, $2/3$, and 3. Since the GWB has yet to be detected, this yields upper limits on Ω_α for each of the spectral indices that are searched for.
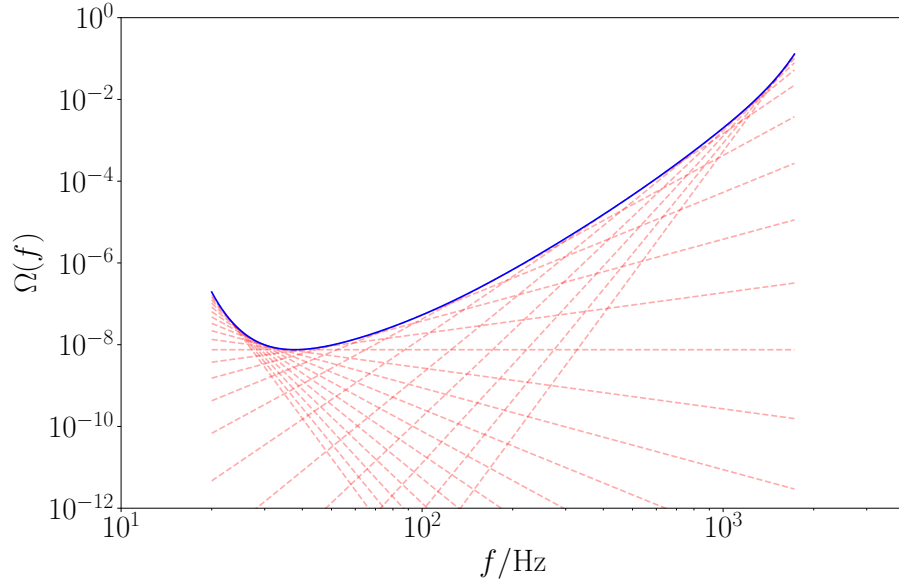
Figure 0.8: The power-law integrated sensitivity curve for LIGO/Virgo's first three observing runs [32, 36] (solid blue curve), along with the individual power-law upper limits for spectral indices $\alpha = -10, -9, \ldots, +10$ (dashed red lines).

Taking this power-law search method to its logical conclusion, we can construct a *power-law integrated sensitivity curve* (PI curve) [545] by scanning over a large set of spectral indices $A = \{\alpha_0, \alpha_1, \ldots\}$ and combining the individual power-law constraints to give

$$\Omega_{\mathrm{PI}}(f) \equiv \max_{\alpha \in A} \Omega_\alpha (f/f_{\mathrm{ref}})^\alpha. \tag{0.75}$$

Under the assumption that the signal can be described by a power law with index $\alpha \in A$, this PI curve has the property that any signal which intersects the curve, or lies tangent to it at any point, will be detected by the search. In principle, we could imagine having a very sharply peaked signal (e.g., a delta function), which could intersect the PI curve without being detected. In practice, however, there are few models that predict such sharply peaked signals, and the PI curve therefore represents an extremely useful summary of the sensitivity of a given GW experiment to the GWB.

0.3 Sources of gravitational waves

As mentioned at the beginning of this chapter, our primary goal is to use GWs (and in particular, the GWB) as *messengers*, allowing us to probe exotic sources in the early and late Universe—particularly those 'dark' sources that are invisible to EM astronomy. In this section we give a brief introduction to a few of these, focusing primarily on *cosmological* sources, whose GW emission can teach us about the structure of the Universe on the largest scales and the fundamental laws that govern it.

We begin, however, with a quintessentially *astrophysical* source: compact binaries. Despite not being cosmological in origin, these sources are still extremely important for GW cosmology, as they give rise to a strong stochastic background; in fact, this signal is likely to be the dominant component of the GWB

across a very broad frequency range. As we will see in chapter 1 below, compact binaries can also be used as tracers of the cosmic matter distribution, allowing us to use anisotropies in the GWB as a novel probe of the large-scale structure of the Universe.

There is one key class of cosmological GWs that we will not discuss here in any detail, which is the primordial tensor spectrum that is generated during inflation through the amplification of quantum fluctuations [293, 525]. Though these primordial GWs are of fundamental importance in modern cosmology, and are doubtless a fascinating observational target, their spectrum is so weak at the frequencies we are interested in ($f \gtrsim 10^{-10}$ Hz) that they are not likely to be directly detected by GW observatories in the foreseeable future.[30] Instead, they are much more likely to be probed indirectly by future CMB missions, through the 'B-mode' polarisation patterns that they imprint on the CMB [48, 358, 504].

0.3.1 COMPACT BINARIES

We saw in section 0.1.6 that in the weak-field, slow-motion limit, GW generation is driven by the second time derivative of the mass quadrupole of the source. In order to maximise the strength of the GW signal, we thus require a large, high-mass system, with rapid (but non-spherically symmetric) internal velocities. Perhaps the most efficient mechanism nature has invented for achieving this is the self-gravitating motion of *binary stars*. Binaries are ubiquitous in astronomy; roughly half of solar-type stars, and a majority of more massive stars, are in binaries or systems of higher multiplicity (triples, etc.) [245]. However, as we show below, for the purposes of GW detection our attention is restricted to *compact* binaries, whose components are not stars but dense stellar remnants such as white dwarfs (WDs), neutron stars (NSs), and black holes (BHs).

CIRCULAR NEWTONIAN BINARIES

As a first approximation, we can model a binary system as two point masses interacting via Newtonian gravity, whose separation vector $\boldsymbol{R}$ thus obeys

$$\ddot{\boldsymbol{R}} + \frac{GM}{R^3}\boldsymbol{R} = \boldsymbol{0}, \tag{0.76}$$

where $M \equiv m_1 + m_2$ is the total mass of the system. This is a good approximation when $GM/R \ll 1$, or equivalently, when the orbital velocity of the binary is small, $v \ll 1$. Assuming for now that the binary is on a circular orbit, equation (0.76) fixes the orbital period P of the binary such that

$$\frac{GM}{R^3} = \left(\frac{2\pi}{P}\right)^2. \tag{0.77}$$

(The same equation holds for more general elliptical orbits if one replaces R—which is no longer constant— with a, the semi-major axis of the ellipse; this equation is then called Kepler's third law.) It is straightfor-

---

[30]At least, this is the case in the standard inflationary scenario of a single slowly-rolling inflaton field. For a summary of some alternative scenarios in which the inflationary GWB is enhanced at high frequencies, see Bartolo *et al.* [104], as well as Iacconi *et al.* [325, 326] for more recent work on this.

Object	m	d	Gm/d
Solar-type star	$\sim m_\odot$	$\sim R_\odot$	$\sim 10^{-6}$
White dwarf	$\sim 0.6\,m_\odot$	$\sim 10^{-2}\,R_\odot$	$\sim 10^{-4}$
Neutron star	$\sim 1.4\,m_\odot$	$\sim 10^{-5}\,R_\odot$	~ 0.2
Black hole	$\ldots$	$2Gm$	$1/2$

Table 0.2: Typical mass, radius, and compactness of various astronomical objects. The compactness of a (nonspinning) BH is always $1/2$ regardless of its mass, as the Schwarzschild radius is always $2Gm$.

ward to substitute equations (0.76) and (0.77) into the GW signal we derived for a system of two point masses (0.40), giving[31]

$$h_{ij} = -\frac{4G\mu}{r}\frac{GM}{R}\begin{pmatrix} \cos(4\pi t/P) & \sin(4\pi t/P) & 0 \\ \sin(4\pi t/P) & -\cos(4\pi t/P) & 0 \\ 0 & 0 & 0 \end{pmatrix}^{\mathrm{TT}}, \qquad (0.78)$$

where we have, without loss of generality, chosen our coordinates such that the binary is confined to the $z = 0$ plane, and such that $\hat{R} = \hat{x}$ at time $t = 0$.

We see that the GWs emitted by a Newtonian binary are monochromatic, with a frequency which is twice that of the binary's orbital frequency,

$$f = \frac{2}{P} = \frac{1}{\pi}\sqrt{\frac{GM}{R^3}}. \qquad (0.79)$$

One way of understanding this intuitively is to notice that after half an orbital period the angular positions of the two binary components are interchanged, resulting in a configuration which is equivalent from the point of view of GW production; this halving of the period corresponds to a doubling of the frequency.

We also see that equation (0.78) contains a factor of GM/R as a result of Kepler's third law. This dimensionless factor is called the *compactness* of the binary, and is a measure of how strongly-gravitating the system is. For a binary composed of objects of mass m and size d, the compactness of the system is bounded from above by the compactness of the objects, Gm/d, since $R \gtrsim d$. We therefore see that the maximum amplitude of the signal (and thus the prospect of detection) is greatly enhanced if the binary is composed of highly compact objects such as WDs, NSs, or BHs.[32] In particular, we see from table 0.2 that the maximum strain from binary black hole (BBH) and binary neutron star (BNS) systems is about six orders of magnitude greater than that of a binary composed of solar-type stars. This is perhaps the

---

[31] Here the time t on the right-hand side should properly be interpreted as the retarded time $t - r$ at the observer's position. However, if the observer is at a fixed distance from the binary (as is usually a good approximation), this distinction is unimportant, as the two are equivalent up to a time translation.

[32] Of course, the weak-field, slow-motion approach we have adopted here breaks down when $GM/R \sim 1$, so it is not strictly legitimate to apply equation (0.78) here. In principle, one must account for relativistic effects to say anything about this regime, either by going to much higher order in perturbation theory (e.g. in a post-Newtonian framework), or by using numerical simulations of full general relativity. However, in practice the simple Newtonian description of the signal is sufficient to illustrate its most important features, once one accounts for radiation reaction as discussed below.

simplest way of understanding why compact binaries, particularly those composed of BHs and NSs, are the primary observational target for current and planned GW observatories.

Eliminating the binary separation R in favour of the GW frequency f using equation (0.79), and projecting onto the polarisation tensors to obtain the complex strain (0.24), we can rewrite the GW signal emitted by a Newtonian binary as

$$h = -\frac{2G\mathcal{M}}{r}(\pi G\mathcal{M}f)^{2/3}\left[(1 + \cos^2\theta)\cos(2\pi ft - 2\phi) + 2\mathrm{i}\cos\theta\sin(2\pi ft - 2\phi)\right]. \qquad (0.80)$$

This is determined by a peculiar combination of m_1 and m_2 called the *chirp mass*,

$$\mathcal{M} \equiv \mu^{3/5}M^{2/5} = \eta^{3/5}M = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}. \qquad (0.81)$$

Here $\eta \equiv \mu/M$ is a dimensionless quantity called the *symmetric mass ratio*, which varies between $\eta = 1/4$ for equal-mass binaries, and $\eta \to 0$ for 'extreme mass-ratio' binaries in which one object is much more massive than the other. Since $\eta \leq 1/4$, we see that the chirp mass is always smaller than the total mass by a factor of $\leq (1/4)^{3/5} \approx 0.44$; for example, GW150914 had a total mass of $M \approx 66\,m_\odot$, but a chirp mass of only $\mathcal{M} \approx 29\,m_\odot$. The angles (θ, ϕ) in equation (0.80) specify the observer's line of sight with respect to the binary reference frame used in equation (0.78); from the observer's point of view θ is thus interpreted as the inclination of the binary with respect to this line of sight, and we will therefore denote it by I. The factor $\pi G\mathcal{M}f$ that appears in equation (0.80)—which will also crop up in many of the expressions below—is another measure of how strongly-gravitating the system is, and is related to the compactness GM/r and orbital velocity v of the binary by $(\pi G\mathcal{M}f)^{2/3} = \eta^{2/5}GM/R = \eta^{2/5}v^2$.

Radiation reaction

If we believe the Newtonian result (0.78), then a binary system will continuously emit GWs at the same fixed frequency. However, we know from section 0.1.7 that these GWs carry away energy, and there must therefore be a corresponding loss of energy from the dynamics of the system. Inserting the complex strain (0.80) into equation (0.43), we find that the binary loses energy at a rate

$$\frac{\mathrm{d}E_{\mathrm{gw}}}{\mathrm{d}t} = \int_{S^2} \mathrm{d}^2\hat{r}\,\frac{\langle|r\dot{h}|^2\rangle}{16\pi G} = \frac{32}{5G}(\pi G\mathcal{M}f)^{10/3}. \qquad (0.82)$$

Assuming a circular orbit, the binary's orbital energy (kinetic plus potential) is

$$E_{\mathrm{orb}} = -\frac{GM\mu}{2R} = -\frac{1}{2}\mathcal{M}(\pi G\mathcal{M}f)^{2/3}. \qquad (0.83)$$

Taking the time derivative of this and setting $\dot{E}_{\mathrm{orb}} + \dot{E}_{\mathrm{gw}} = 0$, we obtain

$$\dot{f} = \frac{96\pi}{5}f^2(\pi G\mathcal{M}f)^{5/3}, \qquad (0.84)$$
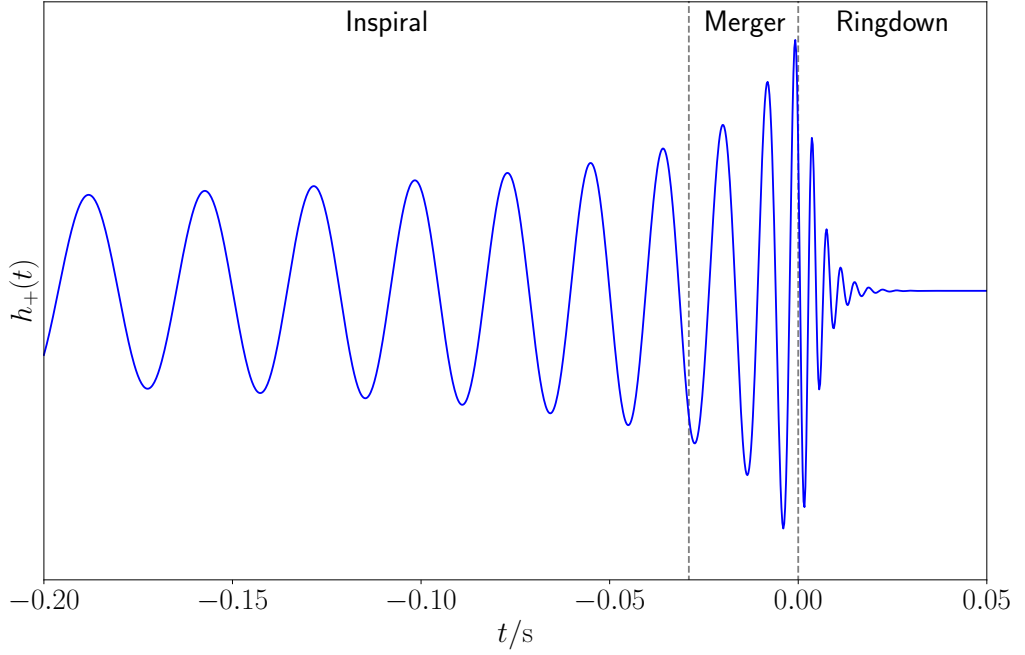
Figure 0.9: Gravitational-wave strain signal from the final stages of a compact binary coalescence. We show the +-polarisation (in arbitrary units) for an observer at inclination $I = 0$. This waveform was generated using the `SEOBNRv4_opt` approximant [143] (as implemented in the Python package PyCBC [438, 556]) for a GW150914-like system with $m_1 = 35.6\,m_\odot$ and $m_2 = 30.6\,m_\odot$ [22].

or equivalently, in terms of the orbital period $P = 2/f$,

$$\dot{P} = -\frac{192\pi}{5}\left(\frac{2\pi G\mathcal{M}}{P}\right)^{5/3}. \tag{0.85}$$

We thus see that energy loss due to GW emission causes a circular binary to spiral inwards, orbiting at an ever-increasing rate.[33] It was this effect which provided the first indirect evidence for the existence of GWs, thanks to observations of the binary pulsar B1913+16 by Hulse and Taylor [324]. We call this process 'radiation reaction', in analogy with the damping force experienced by radiating charges in classical electromagnetism [333].

Note that, while our focus here is on circular orbits, equation (0.85) is readily extended to elliptical orbits, giving [455]

$$\dot{P} = -\frac{192\pi}{5}\left(\frac{2\pi G\mathcal{M}}{P}\right)^{5/3}\frac{1 + \frac{73}{24}e^2 + \frac{37}{96}e^4}{(1-e^2)^{7/2}}, \tag{0.86}$$

where $e \in [0, 1)$ is the eccentricity of the ellipse. In this case, there is also a radiation reaction effect on the eccentricity itself due to the radiation of angular momentum [455],

$$\dot{e} = -\frac{608\pi}{15P}\left(\frac{2\pi G\mathcal{M}}{P}\right)^{5/3}\frac{e + \frac{121}{304}e^3}{(1-e^2)^{5/2}}. \tag{0.87}$$

---

[33]Strictly speaking the orbit is no longer circular, since the separation R shrinks as the binary period decays. However, this process happens on much longer timescales than that of the orbit so long as $|\dot{P}| \ll 1$, which is equivalent to our earlier requirement that $GM/R \ll 1$. We can thus think of these orbits as 'quasi-circular', since the separation is effectively constant over the course of a single period.

Since the right-hand side above is always negative, we see that GW emission causes eccentric binaries to circularise over time, at a rate which increases drastically as the binary's period decays. This justifies somewhat our focus on circular binaries here; indeed all of the GW signals detected thus far by LIGO/Virgo are consistent with zero eccentricity [33] (see however Gayathri *et al.* [284]).

Returning to the circular case, we see from equation (0.80) that as the binary's period decays the amplitude of the GW strain grows, and as a result, so does the rate at which the system loses energy (0.82). The resulting 'chirp' signal, whose amplitude and frequency grow at an accelerating rate over time, is highly distinctive (see figure 0.9), and reflects the inherent gravitational instability of binary systems.[34] The runaway inspiral associated with this instability only stops once the binary's components are so close to each other that they merge to become a single object; to a first approximation, we expect this to happen once the separation reaches $R = 6GM$, which is the radius of the innermost stable circular orbit (ISCO) around a Schwarzschild BH of mass M. Inserting this radius into equation (0.79), we see that the merger frequency is roughly

$$f_{\text{ISCO}} = \frac{1}{6^{3/2}\pi GM} \approx 63\,\text{Hz} \times \left(\frac{M}{70\,m_\odot}\right)^{-1}. \tag{0.88}$$

We can find the time taken for the binary starting at some initial frequency $f_0 \ll f_{\text{ISCO}}$ to reach merger by integrating equation (0.84),

$$t_{\text{merge}} \simeq \frac{5GM}{256}(\pi GM f_0)^{-8/3} \approx 4.8\,\text{s} \times \left(\frac{M}{30\,m_\odot}\right)^{-5/3}\left(\frac{f_0}{10\,\text{Hz}}\right)^{-8/3}. \tag{0.89}$$

The resulting merger product is initially highly excited, and undergoes a period of further GW emission before settling down to a quiescent state. If the final object is a BH (as is always the case for BBHs, and also occurs for sufficiently massive BNSs [92]), this post-merger emission is described by a superposition of damped sinusoids at a fixed set of quasi-normal mode (QNM) frequencies determined by the structure of the BH spacetime; the resulting signal is called the 'ringdown'. The overall inspiral-merger-ringdown process is called a *compact binary coalescence* (CBC).

Frequency spectrum and the stochastic background

So far our discussion has been purely in the time domain, where the evolution of the binary is clearer. However, it is often more convenient to work in the frequency domain, particularly when studying stochastic backgrounds. In general, deriving the frequency-domain GW signal from a binary is rather involved, as we must account for both the rapid phase oscillations in the signal and the more gradual increase in the orbital frequency when taking the Fourier transform. Fortunately, for quasicircular Newtonian binaries in the quadrupole approximation there is a simple trick we can use to obtain the frequency spectrum of the radiated GW energy [486]. Taking advantage of the fact that the GW signal in

---

[34] We can understand this instability from a thermodynamical point of view. Equation (0.83) implies that a decrease in the energy of the system corresponds to an *increase* in its 'temperature' (whether this is defined in terms of the energy radiated by the system or in terms of the velocities of its components). The system therefore has a *negative heat capacity*, meaning that it is unable to reach equilibrium, and is thus inherently unstable. This behaviour is not restricted to binaries, but is a generic feature of finite-size self-gravitating systems, including galaxies [125] and black holes [574].

this approximation is emitted at a single frequency which can be written as a function of time, $f(t)$, by integrating equation (0.84), we write the GW energy spectrum as

$$\frac{\mathrm{d}E_{\mathrm{gw}}}{\mathrm{d}(\ln f)} = \frac{f}{\dot{f}}\frac{\mathrm{d}E_{\mathrm{gw}}}{\mathrm{d}t} = \frac{1}{3}\mathcal{M}(\pi G\mathcal{M}f)^{2/3} = \frac{1}{3}\eta M v^2. \tag{0.90}$$

This agrees exactly with what one finds from a more careful analysis, using the stationary-phase approximation to evaluate the Fourier transform of the strain [406]. Note that this expression is identical to the orbital energy (0.83) up to an $\mathcal{O}(1)$ constant; this tells us that GW emission is highly efficient in extracting energy from the orbit. Indeed, if we integrate equation (0.90) up to the ISCO frequency, we find that the total energy radiated is $E_{\mathrm{gw}} = \mu/12$, or just over 2% of the total rest-mass of the system for equal-mass binaries. (In comparison, the equivalent fraction for Hydrogen burning in stars is roughly 0.7%.) In reality, the true fraction is even higher due to radiation emitted in the merger-ringdown phase.

Inserting equation (0.90) into the Phinney formula (0.64), we see that the GWB signal from a cosmic population of CBCs has a characteristic $\sim f^{2/3}$ spectral slope, the amplitude of which is determined by the CBC rate density and mass distribution. Note however that this $\sim f^{2/3}$ power law does not continue indefinitely; since the spectrum in equation (0.90) is truncated when the two objects merge around the ISCO frequency (0.88), the GWB spectrum will gradually reach a peak and then drop off at higher frequencies, as fewer and fewer CBCs contribute.

Ignoring this for now, we can combine equations (0.90) and (0.64) to find

$$\Omega(f) = \frac{8f^{2/3}}{9H_0^2} \int_0^\infty \frac{\mathrm{d}z}{(1+z)^{4/3}} \frac{\mathcal{R}(z)}{H(z)} \int \mathrm{d}\mathcal{M}\, p(\mathcal{M}|z)(\pi G\mathcal{M})^{5/3}, \tag{0.91}$$

where $p(\mathcal{M}|z)$ is the chirp mass distribution as a function of redshift. We can obtain a crude, order-of-magnitude estimate for the GWB from CBCs by setting the redshift integral equal to the value of its integrand at redshift zero. For the LIGO/Virgo frequency band, this gives

$$\Omega(f) \sim 10^{-9} \times \left(\frac{f}{100\,\mathrm{Hz}}\right)^{2/3} \left(\frac{\langle\mathcal{M}\rangle}{30\,m_\odot}\right)^{5/3} \frac{\mathcal{R}_0}{10\,\mathrm{Gpc}^{-3}\,\mathrm{yr}^{-1}}, \tag{0.92}$$

where $\langle\mathcal{M}\rangle$ is the mean chirp mass (which we have normalised to the chirp mass of GW150914) and $\mathcal{R}_0$ is the BBH merger rate at $z = 0$ (which we have normalised to the value reported in Abbott *et al.* [22]). This simple estimate is actually remarkably close to the answer we find from a much more detailed calculation in chapter 1.

0.3.2 First-order phase transitions

We now turn to look at GWs of cosmological origin. Although we have yet to detect GWs from any source other than CBCs, the rise of GW astronomy has prompted many cosmologists and particle theorists to investigate mechanisms for producing GWs in the early Universe. Few such possibilities have excited the interest of the community as much as *cosmological first-order phase transitions*, which present an extremely promising way to search for new physics beyond the Standard Model [169, 170, 172, 359].
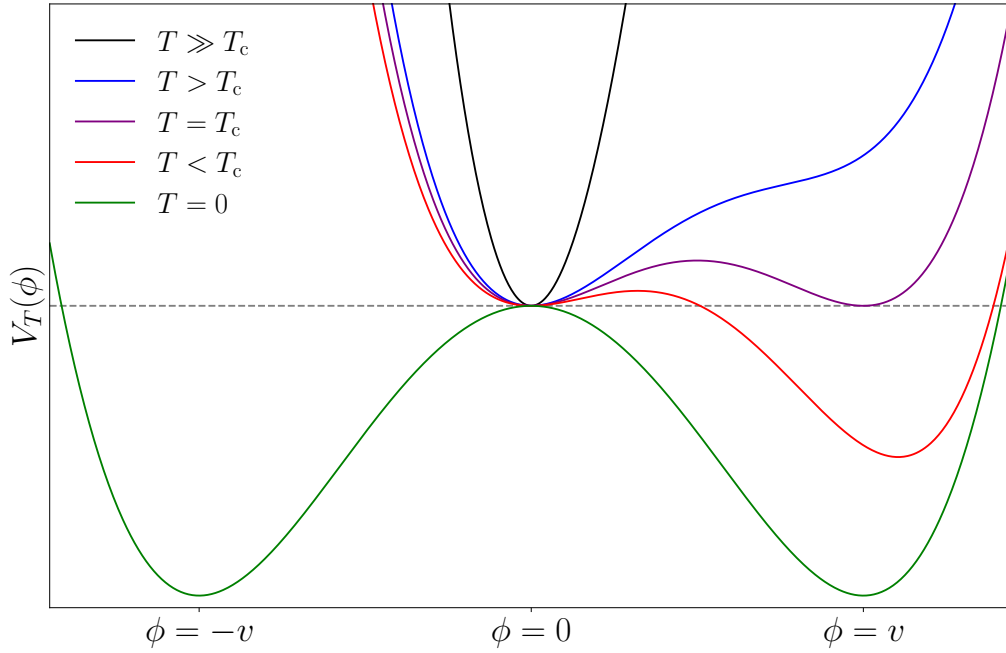
Figure 0.10: The thermal effective potential (0.93) for a scalar field ϕ that undergoes a first-order phase transition. The field is initially localised around $\phi = 0$, which is a stable vacuum state at high temperatures. As the Universe cools below a critical temperature T_c due to cosmic expansion, $\phi = 0$ becomes a metastable false vacuum, separated from the true vacuum by a potential barrier.

The basic picture is as follows: at some high-temperature epoch in the early Universe, we have a scalar field ϕ whose dynamics are described by a thermal effective potential of the form [317]

$$\hbar^3 V_T(\phi) \simeq A(T^2 - T_0^2)\phi^2 - BT\phi^3 + C\phi^4, \tag{0.93}$$

where A, B, and C are dimensionless constants, T is the temperature of the Universe, and T_0 is some fixed energy scale. (Recall that we use units with $k_B = 1$ but $\hbar \neq 1$. The factor of $\hbar^3$ here ensures that the potential has dimensions of mass per unit volume, while T and ϕ both have dimensions of mass.) For example, ϕ could be the Higgs field, in which case T_0 is related to the Higgs mass, and the values of A, B, C are set by the Higgs' couplings to other particles. As the Universe expands and cools, T decreases, changing the shape of the potential and therefore the dynamics of the scalar field. At zero temperature and at very high temperatures, there is a discrete $\mathbb{Z}_2$ symmetry under $\phi \rightarrow -\phi$.

For appropriate values of A, B, C, the potential exhibits the behaviour shown in figure 0.10 where, as well as possessing a local minimum at $\phi = 0$ at high temperatures, a second minimum develops at some $\phi > 0$ once the temperature falls below a certain threshold. As the Universe continues to cool, it passes a critical temperature T_c where this point becomes a *global* minimum, and it becomes energetically favourable for the field to fall into this state and acquire a nonzero vacuum expectation value (VEV), $v \equiv \langle \phi \rangle > 0$. We call this state the *broken phase*, as it breaks the $\phi \rightarrow -\phi$ symmetry, in contrast with the *symmetric phase* at $\phi = 0$.

For a broad class of values of A, B, C, the potential has a barrier between $\phi = 0$ and $\phi = v$ at temperatures $T \lesssim T_c$, which prevents the field from straightforwardly rolling down into the broken phase.

The symmetric phase thus becomes a metastable *false vacuum* (as opposed to the *true vacuum* at the bottom of the potential well) which can persist for some time after the critical temperature is reached. Eventually, however, the field escapes this metastable state, either through thermal fluctuations 'over the barrier', or through quantum tunnelling 'under the barrier'. These random decay events occur at localised points in spacetime, resulting in the nucleation of 'bubbles' of true vacuum which then expand outwards at relativistic velocities and collide with each other until the entire Universe has transitioned to the broken phase. This process is called a *first-order phase transition* (FOPT).

While we have motivated this discussion in terms of a Higgs-like scalar field ϕ, it turns out that the masses and couplings of the heaviest particles in the Standard Model are such that the Higgs potential does *not* have a barrier, and instead of a FOPT there is a simple cross-over transition which occurs homogeneously throughout the Universe in thermal equilibrium [356]. However, in alternative particle physics models which couple new fields to the Higgs, or which have new Higgs-like scalar fields, the potential generically acquires a barrier between the symmetric and broken phases, meaning that FOPTs are a very generic consequence of such modifications. As such, any detection of a FOPT in the early Universe would be incredibly exciting, as it would provide compelling evidence for physics beyond the Standard Model.

GW astronomy provides the most promising means of making such a detection. This is because false vacuum decay represents a violent and highly energetic departure from equilibrium in the early-Universe plasma, with the resulting dynamics of the matter fields leading to copious production of GWs. Modelling the resulting GWB signal is a considerable theoretical and numerical challenge, and is currently a very active area of research. However, the general picture is that there are three key processes that contribute to the GWB [172]:

1. the dynamics of the scalar field during collisions of true-vacuum bubbles;

2. the bulk motion of the plasma due to sound waves;

3. the chaotic motion of the plasma due to MHD turbulence.

While all three of these mechanisms depend in a complicated way on the details of the underlying particle physics model, the sound-wave contribution dominates the combined GWB spectrum in most cases. This contribution is usually modelled as a broken power law,

$$\Omega(f) = \Omega(f_*) \times (f/f_*)^3 \left[\frac{7}{4 + 3(f/f_*)^2} \right]^{7/2}, \tag{0.94}$$

which grows like $\sim f^3$ at low frequencies and falls off as $\sim f^{-4}$ at high frequencies. The changeover between these two regimes occurs at a peak frequency [172, 501]

$$f_* \approx 19\,\mu\mathrm{Hz} \times \frac{T_*}{100\,\mathrm{GeV}} \frac{\beta/H_*}{v_{\mathrm{w}}} \left(\frac{g_*}{106.75} \right)^{1/6}, \tag{0.95}$$

which depends on a few key parameters describing the phase transition: T_*, the temperature at which the transition takes place (which we have normalised here relative to the electroweak scale); β, the inverse duration of the transition (normalised to the Hubble rate at the transition epoch, H_*); and v_{w}, the velocity

of the bubble walls. (Here g_* is the number of relativistic degrees of freedom in the plasma, which we have normalised to the value it takes in the Standard Model at early times.) The peak amplitude of the spectrum is approximated by

$$\Omega(f_*) \approx 5.7 \times 10^{-6} \times \frac{\nu_\mathrm{w}}{\beta/H_*} \left(\frac{\kappa\alpha}{1+\alpha}\right)^2 \left(\frac{g_*}{106.75}\right)^{-1/3} \left[1 - (1 + 2\tau_\mathrm{sw}H_*)^{-1/2}\right], \qquad (0.96)$$

where α is the energy density released by the transition, in units of the energy density of the plasma at that epoch; κ is an efficiency parameter determined by α and ν_w; and τ_sw is the lifetime of the sound-wave source, which is a function of α, β, and ν_w.

We can understand the strongly-peaked nature of this spectrum as being a consequence of the transient nature of the FOPT source: the GW production is centred around a single epoch in the Universe's history (set by the transition temperature T_*), and as a result this imprints a characteristic frequency scale f_* on the spectrum. This makes it very challenging to probe the entire FOPT parameter space $(T_*, \alpha, \beta/H_*, \nu_\mathrm{w})$, as it is possible for the peak of the spectrum to lie between the sensitive frequency bands of various GW experiments. We will revisit this problem in chapter 3, and show how the novel GW detection method we develop there can be used to probe a unique region of the FOPT parameter space.

0.3.3 Cosmic strings

Consider now what happens if we replace the real scalar field ϕ of the previous section with a *complex* scalar field, whose effective potential at low temperatures $T \ll T_\mathrm{c}$ is given by

$$\hbar^3 V(\phi) \simeq C(|\phi|^2 - \nu^2)^2, \qquad (0.97)$$

which is minimised by setting $|\phi| = \nu$. (This is the same potential as equation (0.93) at $T \ll T_0$, up to a constant shift.) Rather than having two disconnected vacua at $\phi = \pm\nu$ related by a broken $\mathbb{Z}_2$ symmetry, we now have a connected circle of vacua at $\phi = e^{i\vartheta}\nu$ that are related by a broken $U(1)$ symmetry, as illustrated in figure 0.11; we label each of these vacua by a phase angle ϑ.

As the Universe cools, different spatial regions decay from the symmetric phase $\phi = 0$ to a different, randomly-selected point on the vacuum circle, resulting in a phase ϑ that is a smoothly-varying field on spacetime. This process generically gives rise to configurations such as the one shown in figure 0.12, where ϑ executes a nonzero winding around the vacuum circle. As we approach the centre of such a configuration, we see that it is impossible for the scalar field ϕ to reside in any of the vacua without causing a discontinuity; ϕ is therefore forced into the higher-energy symmetric phase at this point. Topologically, it is impossible to have $\phi = 0$ at just this single isolated point. Instead, we necessarily obtain a linelike object of nonzero energy density by stacking many of these points together, as illustrated in figure 0.12. We call this object a *cosmic string* [316, 371, 565, 567]. As with FOPTs, cosmic strings do not occur in the Standard Model, but are a generic consequence of many high-energy modifications to the Standard Model [341], and are therefore an extremely well-motivated means of searching for new physics in the early Universe.

$$T \gg T_{\mathrm{c}} \qquad\qquad T \ll T_{\mathrm{c}}$$
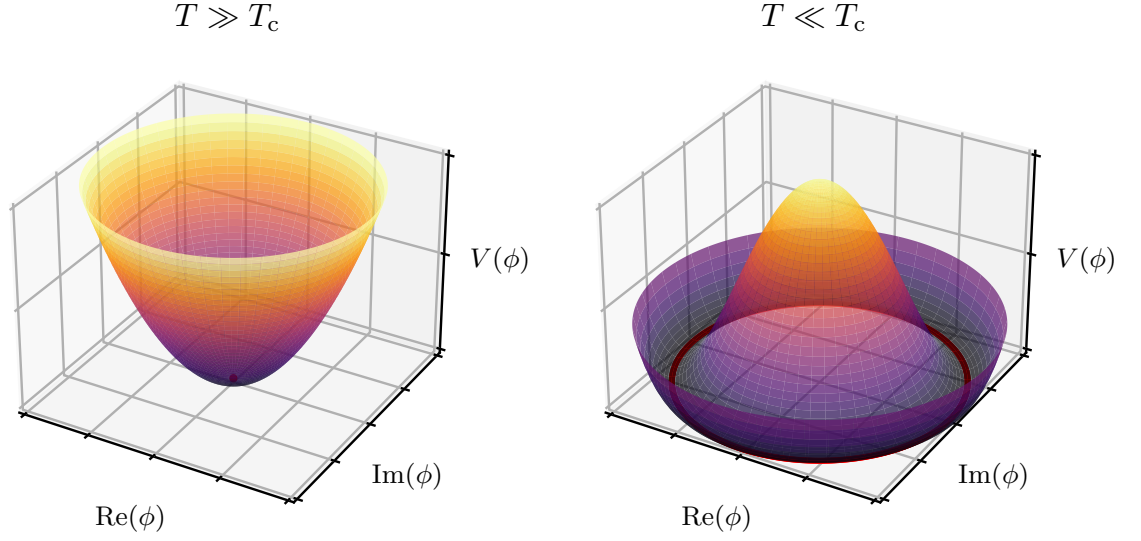
Figure 0.11: Effective potential for the complex scalar field ϕ. At high temperatures (left panel), this is simply $\sim |\phi|^2 T^2$, and the symmetric phase $\phi = 0$ is the unique vacuum state. At low temperatures (right panel), the potential transitions to the form in equation (0.97); the symmetric phase then becomes unstable, and the field decays into one of a continuous family of vacua given by $|\phi| = v$.

Once formed, cosmic strings are protected from decay by the topological stability of the underlying field configuration. This stability is much more robust than the metastability of the false vacuum state we encountered in the previous section, which was vulnerable to decay via any single point in space tunnelling to the true vacuum. Due to the continuity of the phase angle ϑ throughout space, it is impossible to 'unwind' the scalar field at any single point along the string: one would have to perform this unwinding simultaneously *everywhere* along the entire length of the string. The probability of the field tunnelling to such a globally unwound configuration is vanishingly small. As a result, cosmic strings can survive over cosmological timescales, acting as *relics* of their early formation epoch that we can, in principle, observe at late times.

We can understand the importance of cosmic strings as GW sources by estimating the amount of mass/energy they carry per unit length. To calculate this, we need to know (*i*) the energy density ρ of the scalar field near the core of the string, and (*ii*) the characteristic width δ of the string, i.e., the lengthscale over which ϕ decays from the symmetric phase to one of the vacuum states as we travel away from the core of the string. We can then approximate the linear mass density, μ, as the product of the cross-sectional area of the string and the energy density in its core,

$$\mu \approx \pi \delta^2 \times \rho. \tag{0.98}$$

It is natural to interpret ρ as the difference in the potential (0.97) between the symmetric phase and the broken phase, which gives us $\rho = V(0) - V(v) \sim \hbar^{-3} v^4$. For the string width δ, we notice that since the scalar field VEV, v, is the only dimensionful quantity in equation (0.97), the only lengthscale that we can

construct is the associated Compton wavelength, $\delta \sim \hbar/v$. Neglecting $\mathcal{O}(1)$ constants, we therefore find that the linear density must be

$$G\mu \sim (v/m_{\mathrm{Pl}})^2, \tag{0.99}$$

which we have made dimensionless by normalising relative to the Planckian linear density $m_{\mathrm{Pl}}/\ell_{\mathrm{Pl}} = m_{\mathrm{Pl}}^2/\hbar = 1/G$. Equivalently, this means we can write the string width δ as

$$\delta \sim \sqrt{\hbar/\mu} = \ell_{\mathrm{Pl}}/\sqrt{G\mu}. \tag{0.100}$$

As with any string, the linear density μ is related to the string tension, $\mathcal{T}$, and the propagation speed of waves along the string, u, by the equation $\mu = \mathcal{T}/u^2$. As we will see in chapter 2, the motion of cosmic strings can be written as a *relativistic* wave equation with $u = 1$, so this equation tells us we are free to interchange their tension and linear density. We therefore refer to $G\mu$ as the *(dimensionless) string tension* of a cosmic string.

Equation (0.99) shows that strings formed at early epochs in the Universe's history will have enormous string tensions compared to objects from our everyday experience. For example, a string formed at the grand unified theory (GUT) scale, $v \sim 10^{13}$ TeV $\sim 10^{-3} m_{\mathrm{Pl}}$, would have a dimensionless string tension of $G\mu \sim 10^{-6}$. (Compare this with, e.g., a steel guitar string, which has $G\mu \approx 5 \times 10^{-30}$.) This means that a GUT-scale string spanning the size of the Earth's orbit around the Sun would have a mass of $\mu \times \mathrm{AU} \sim 100\, m_\odot$, far greater than the total mass of the Solar System. Such objects would clearly exert considerable gravitational influence over their surroundings. The fact that they also oscillate at relativistic velocities tells us that cosmic strings must be incredibly powerful sources of GWs.

Since cosmic strings are topologically forbidden from having endpoints, they are forced to form either closed loops [372, 554] or open-ended 'long' strings (also called 'infinite' strings) which stretch across an entire Hubble volume. Any given Hubble patch will typically have $\mathcal{O}(1)$ long strings passing through it [567], which are formed at the epoch of the symmetry-breaking phase transition. These long strings are free to oscillate on sub-horizon scales, and generically intersect themselves to chop off closed loops, which produce copious GWs and decay to smaller sizes by trading their length for radiated energy (since their linear density μ is a constant fixed by the underlying field theory, any loss of energy necessarily translates into a decay in the loop's size). This process (combined with loop-loop intersections which also chop off smaller loops) causes a downwards cascade, eventually filling the Hubble patch with a network of loops with a continuous spectrum of sizes, through which the energy in the long strings is gradually dissipated into GWs [113].

Since loop production persists over cosmological timescales, the GW emission from the loop network accumulates over time to give a strong GWB spectrum [87]. Unlike the peaked spectrum (0.94) we saw in the FOPT case, which encoded the single epoch of GW production, this collective emission from cosmic strings of all sizes and at all epochs leads to a GWB which is flat over a very broad frequency band (although this plateau is usually also accompanied by a peak at lower frequencies). As a result, cosmic strings allow us to probe much higher energy scales than those accessible with FOPT signals. For example, the GWB from a FOPT occurring at the GUT scale would peak at $f \sim 10^{10}$ Hz, far beyond the sensitive
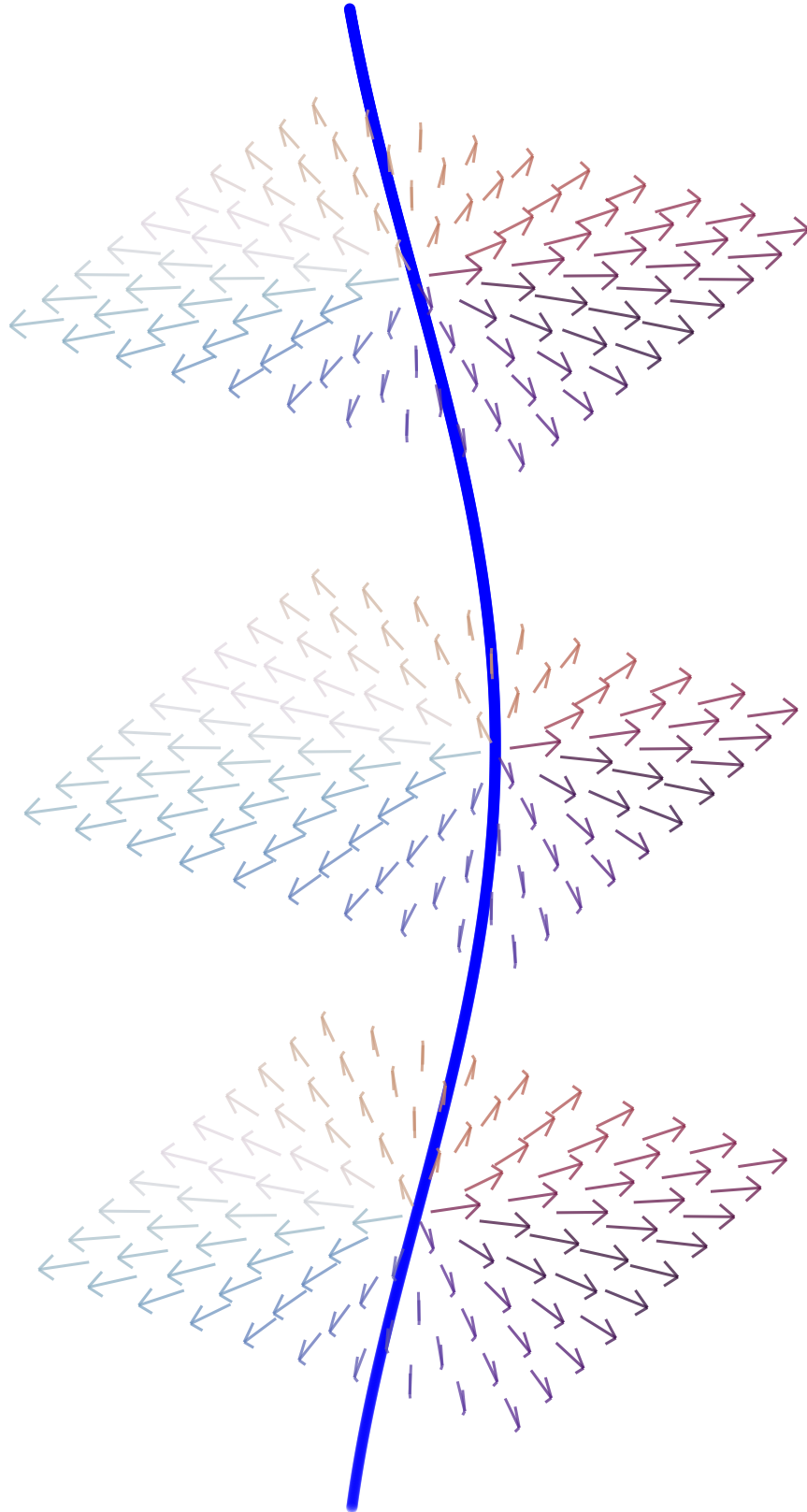
Figure 0.12: Configuration of the complex scalar field ϕ on three spatial slices intersecting a cosmic string (shown in blue). The colour and direction of the arrows at each point on these slices indicate the phase angle ϑ of the vacuum state at that point in space. If we travel around a closed spatial loop that encloses the string, we find that this phase has a nonzero winding number n around the vacuum circle, $\vartheta \to \vartheta + 2\pi n$.

frequencies of current or future GW experiments, while the signal from a cosmic string network formed at that epoch would be observable at much lower frequencies.

GW searches for cosmic strings with LIGO/Virgo [4, 18, 24, 36, 37] and pulsar timing arrays [131, 388, 593] have so far returned only null results, allowing us to place a conservative upper limit on the string tension of $G\mu \lesssim 10^{-11}$ (there is some uncertainty due to the modelling of the cosmic string loop network—the most stringent constraints are at the level of $G\mu \lesssim 10^{-15}$ [37]). These constraints are orders of magnitude better than those derived from CMB observations, which are of the order $G\mu \lesssim 10^{-7}$ [47, 355, 419].

In chapter 2 we will describe the GW emission from loops in more detail, focusing on three key emission mechanisms called *cusps*, *kinks*, and *kink-kink collisions*. We will derive, for the first time, the nonlinear GW memory (0.47) associated with these mechanisms, and will show that this has surprising implications for the gravitational dynamics of cusps.

0.3.4 Primordial black holes

We saw in section 0.3.1 that compact binaries, and binary black holes in particular, are vitally important sources of GWs. Previously, we characterised these sources as *astrophysical* rather than cosmological, as the main formation channel for BHs and NSs is thought to be through the gravitational collapse of stars that have exhausted their nuclear fuel. However, it is also possible that BHs may have formed through *cosmological* processes in the early Universe, meaning that the BBH signals observed by LIGO/Virgo could be a hint of something much more exotic. Efforts to understand the formation and observational signatures of these *primordial* black holes (PBHs) [180, 181, 303, 596] have formed a major strand of research in cosmology for the past fifty years.

While there are various ways of producing PBHs in the early Universe that have been studied in the literature, the most well-established mechanism is the gravitational collapse of horizon-sized density perturbations. Since a given Hubble patch must be significantly more massive than average to undergo such a collapse, this mechanism requires the production of very large overdensities during inflation, e.g., due to a phase transition between two distinct inflationary phases [279], or the presence of a feature in the inflationary potential [332]. These large perturbations are 'frozen out' by the inflationary expansion, and only become dynamical again and collapse once they re-enter the horizon during the radiation-dominated era. We can estimate the masses of the resulting PBHs by noticing that, since they correspond to the collapse of an entire Hubble patch, their Schwarzschild radius $2Gm_{\mathrm{pbh}}$ should be of the same order of magnitude as the particle horizon at the time of horizon re-entry. In a radiation-dominated Universe this is given by $r_{\mathrm{hor}} = 2t_{\mathrm{age}}$, where t_{age} is the age of the Universe at that epoch. We therefore have the rough estimate

$$m_{\mathrm{pbh}} \sim \frac{t_{\mathrm{age}}}{G} \sim m_{\mathrm{Pl}} \times \left(\frac{t_{\mathrm{age}}}{t_{\mathrm{Pl}}}\right) \sim 10^6 \, m_\odot \times \left(\frac{t_{\mathrm{age}}}{10\,\mathrm{s}}\right). \tag{0.101}$$

This shows that the masses of horizon-collapse PBHs can span an enormous range of physical scales, from tiny Planckian objects formed in the Universe's earliest moments, all the way up to supermassive BHs comparable to those residing at the centres of galaxies, formed shortly before the epoch of big-bang nucleosynthesis. PBHs that are sufficiently massive can produce detectable GW signals at late times by
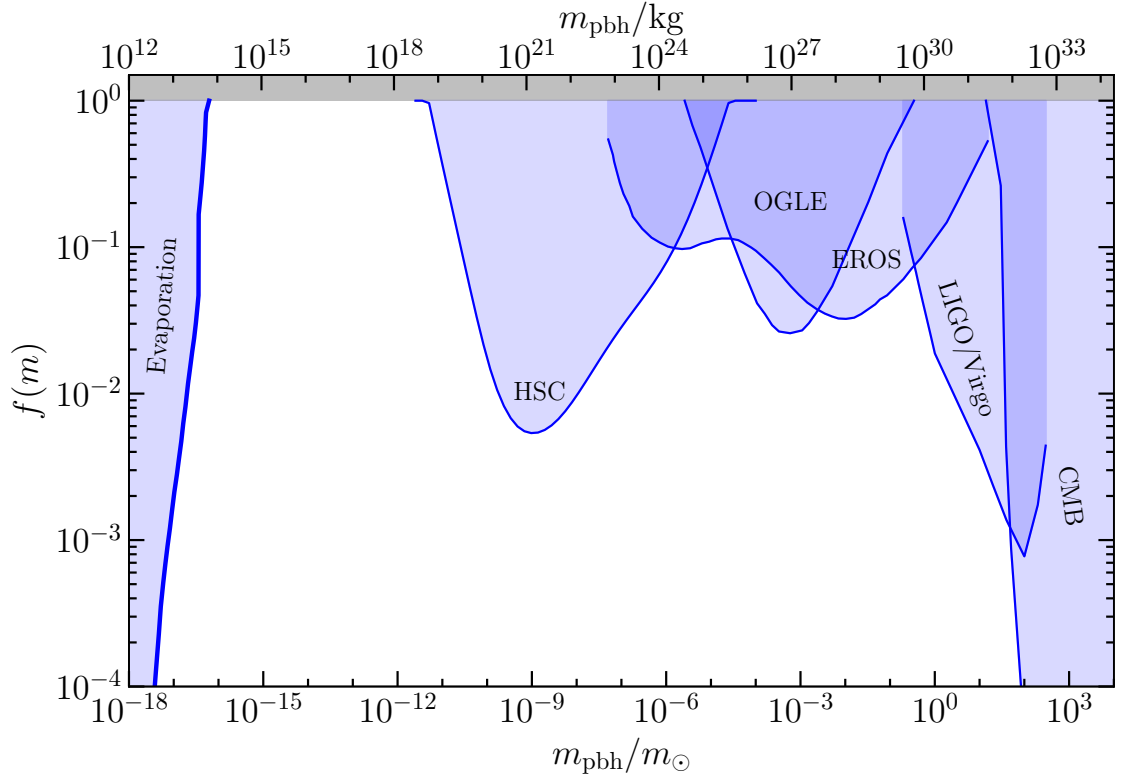
Figure 0.13: Current observational bounds on the PBH mass spectrum. Very light PBHs ($m_{\text{pbh}} \lesssim 10^{-16} \, m_{\odot}$) are constrained by the non-observation of high-energy particles (e.g. gamma rays) emitted during evaporation [150, 182, 204, 205, 235, 373, 386, 463], while very massive PBHs ($m_{\text{pbh}} \gtrsim 10 \, m_{\odot}$) are constrained by the sensitivity of the CMB anisotropies to PBH accretion at the epoch of recombination [505]. PBHs at intermediate masses ($10^{-11} \, m_{\odot} \lesssim m_{\text{pbh}} \lesssim 100 \, m_{\odot}$) are constrained by stellar microlensing experiments such as OGLE [437], EROS [546], and Subaru-HSC [220], and by the BBH event rate observed by LIGO/Virgo [26, 365]. PBHs in the 'sublunar' mass range ($10^{-16} \, m_{\odot} \lesssim m_{\text{pbh}} \lesssim 10^{-12} \, m_{\odot}$) could account for the entirety of the Universe's CDM budget. This figure was produced using the Python package `PBH-bounds` [291, 364].

forming compact binaries [126, 206, 231, 467, 468, 497, 498], as described in section 0.3.1. They can also produce a powerful GWB signal during the collapse process itself, with the dynamics of the large scalar overdensities leading to the generation of sizeable tensor perturbations [74].

One of the most exciting features of equation (0.101) is that it provides a mechanism for producing BHs with masses less than that of the Sun, $m_{\text{pbh}} < m_{\odot}$, which is impossible to achieve through the standard astrophysical process of stellar collapse. This means that any detection of a subsolar-mass BH, whether through GWs or otherwise, would provide powerful evidence for PBH formation in the early Universe, and could substantially advance our understanding of inflation and other areas of fundamental physics. Aside from their masses, PBHs may also possess other features which could help to distinguish them from astrophysical BHs. For example, the radiation-era formation process described above is expected to lead to the PBHs having negligibly small spins, due to the angular momentum of the collapsing region being dissipated by radiation pressure during the collapse; this is in contrast with astrophysical BHs, which can have quite sizeable spins due to angular momentum conservation during their collapse.

As we go towards lower end of the mass range predicted by equation (0.101), it becomes increasingly important to treat the PBHs as quantum-mechanical objects. By considering the behaviour of quantum fields on BH spacetimes, Hawking [305] famously showed that BHs emit thermal blackbody radiation with a temperature[35] $T_{\mathrm{BH}} = \hbar/(8\pi G m)$, causing them to radiate away their mass and evaporate on a timescale [304, 442, 443]

$$t_{\mathrm{evap}} \sim t_{\mathrm{Pl}} \times \left(\frac{m}{m_{\mathrm{Pl}}} \right)^3. \tag{0.102}$$

As a result, PBHs with masses less than

$$m_* \approx 3 \times 10^{19} \, m_{\mathrm{Pl}} \approx 5 \times 10^{11} \, \mathrm{kg} \approx 3 \times 10^{-19} \, m_\odot \tag{0.103}$$

are expected to have evaporated completely in the time since their formation. Equation (0.103) therefore provides a lower bound on the masses of PBHs surviving to the present day.

As we mentioned near the beginning of this chapter, non-evaporating PBHs are interesting not just as GW sources, but also as a very well-motivated candidate for dark matter [178, 188]. Since they are massive, nonbaryonic, have nonrelativistic velocities, and interact only through gravity, PBHs fulfil all of the properties that we require for a DM particle. For this reason, constraints on their cosmological abundance are usually phrased in terms of their *CDM mass fraction*,

$$f(m) \equiv \frac{1}{\rho_{\mathrm{cdm}}} \frac{\mathrm{d}\rho_{\mathrm{pbh}}}{\mathrm{d}(\ln m)} = \frac{m^2 n_{\mathrm{pbh}}(m)}{\rho_{\mathrm{cdm}}}, \tag{0.104}$$

which tells us how much of the observed CDM mass density in the Universe is made up of PBHs of mass m. (We will also refer to this quantity as the *PBH mass spectrum*.) Here ρ_{pbh} and n_{pbh} are the PBH mass density and number density respectively, while $\rho_{\mathrm{cdm}} \approx 0.261 \, \rho_{\mathrm{c}} \approx 30 \, m_\odot \, \mathrm{kpc}^{-3}$ (c.f. table 0.1). To successfully account for the entirety of the DM, a given PBH formation scenario must give $\int \mathrm{d}(\ln m) \, f(m) = 1$; however, consistency with various observational probes places upper bounds on $f(m)$ over a broad range of masses, as shown in figure 0.13.

We will encounter PBHs again in chapter 2, in which we investigate the possibility of them being formed through the gravitational collapse of cusps on cosmic string loops.

0.4 GRAVITATIONAL-WAVE DETECTION: BASIC PRINCIPLES AND CURRENT RESULTS

Having familiarised ourselves with a few important sources of gravitational waves in section 0.3, the obvious question now is: how do we go about detecting these signals?

---

[35] Strictly speaking, this expression is only correct for Schwarzschild BHs. For Kerr BHs one must multiply this by a factor of $2\sqrt{1-\chi^2}/(1+\sqrt{1-\chi^2})$, where $\chi \equiv J/(Gm^2) \in [0,1]$ is the BH's dimensionless spin.

The key idea is to use the results of section 0.1.5, in which we found that the passage of a GW causes an oscillation in the proper distance between two freely-falling test masses. In particular, this results in an oscillation in the light-travel time T between the two masses [259],

$$\delta T = \frac{1}{2} \int_0^T dt \, \hat{u}^i \hat{u}^j h_{ij}(t, \boldsymbol{x}(t)), \tag{0.105}$$

with $\hat{\boldsymbol{u}}$ the propagation direction of the light pulse. If one knows the expected arrival time of the pulse with sufficient accuracy, then searching for offsets from this expected arrival time allows us to probe the GW strain h_{ij}. The two leading methods for GW detection, laser interferometry and pulsar timing, are both fundamentally based on this principle, each using different pairs of objects as test masses and different means of calibrating the expected arrival time. We give a brief overview of each of these methods and their current observational results in sections 0.4.1 and 0.4.2 respectively. We then describe a completely different class of indirect cosmological bounds in section 0.4.3, before briefly touching on a variety of other GW searches in section 0.4.4.

0.4.1 LASER INTERFEROMETERS

One approach to extracting the perturbations (0.105) in the round-trip time of a light pulse is to use an *interferometer* (IFO), sending two laser pulses of the same wavelength along different paths through space before recombining them and measuring their relative phase. The power of this method is that it removes the need to calibrate the light travel time along each path; instead, we can perform a *differential* measurement between the two paths. If each path is affected differently by the passage of a GW, then this results in a perturbation to the relative phase. For example, suppose we use a beam-splitter to send half of our laser photons along the $\hat{\boldsymbol{x}}$-axis and the other half along the $\hat{\boldsymbol{y}}$-axis, each being reflected by a test mass and returning to the origin. The resulting differential signal is then

$$T_x(t) - T_y(t) = \frac{1}{2} \int_0^T dt \, (\hat{x}^i \hat{x}^j - \hat{y}^i \hat{y}^j) h_{ij}(t, \boldsymbol{x}(t)) \simeq \frac{T}{2} (\hat{x}^i \hat{x}^j - \hat{y}^i \hat{y}^j) h_{ij}(t, \boldsymbol{0}), \tag{0.106}$$

where in the second equality we have taken the *small-antenna limit* in which the arms of the IFO are much smaller than the wavelength of the GW.

In order to measure this signal, we require the test masses at the ends of each arm of the IFO to be free to move along the axis of the corresponding arm, so that their response to an incoming GW corresponds to that of a freely-falling object along that axis. However, this freedom means that there are countless sources of instrumental noise that can shake the test masses and therefore generate a phase difference (0.106), limiting the sensitivity to GW signals. The two key sources of noise for ground-based IFOs are seismic noise at low frequencies and photon shot noise in the laser at high frequencies [406, 500]; as a result, the best sensitivity comes at a 'sweet spot' between these two noise floors at about 10–1000 Hz.

Several ground-based IFOs have been used to conduct GW searches since this idea was first proposed by Weiss [587] in 1972. Of these, by far the most successful to date are the Laser Interferometer Gravitational-Wave Observatory [2, 302] (LIGO, which consists of two instruments in Hanford, Washington and Livingston, Louisiana, USA), and its European counterpart Virgo [44] (which is a single instrument in

Cascina, Italy). These three IFOs, operating together as a global GW observatory network, have jointly detected dozens of CBC signals [22, 33, 40], primarily from BBHs, but also including several BNSs and, more recently, two systems which have been interpreted as black hole-neutron star binaries (BHNS) [39]. These detections have revealed fascinating insights into the population properties of BHs and NSs in the local Universe [35], and have been used to perform stringent new tests of GR [6, 34, 594]. They have also included a number of 'exceptional events' with unexpected properties [28–31] that challenge our understanding of the astrophysical formation channels for these systems.

Aside from their success in detecting CBCs, LIGO/Virgo have also conducted some of the most sensitive GWB searches of any experiment to date [8, 24, 36]. In figure 0.14 we show the PI curve from LIGO/Virgo's third observing run, which reaches a maximum sensitivity of $\Omega(f) \approx 7.5 \times 10^{-9}$ at $f \approx 38$ Hz for an isotropic background [36]. LIGO/Virgo have also carried out searches for anisotropic backgrounds [9, 25, 38], backgrounds with non-GR polarisation modes [19, 166], and modelled searches for the GWB from cosmic strings [18, 37].

Looking to the future, the recent addition of the Japanese IFO KAGRA [27, 61] and the future addition of LIGO India [555] will extend the capabilities of the GW observatory network, while planned upgrades to the existing instruments will further enhance their sensitivity. This network is expected to detect the GWB signal from compact binaries once LIGO and Virgo reach their design sensitivities [17]. By that time, we can expect to have detected many more individual CBC signals, which in turn will give us a much firmer understanding of the expected amplitude and spectral shape of the stochastic CBC signal.

By the late 2030s, these experiments will likely be replaced by so-called 'third generation' IFOs such as Einstein Telescope (ET) [464] and Cosmic Explorer (CE) [295, 472]. These observatories will be so sensitive that their BBH detection horizons will reach all the way out to the earliest epochs of star formation at $z \approx 20$ and beyond [407] (the BNS and BHNS detection horizons are at later epochs, due to the lower masses of these systems), allowing them to individually resolve more than 99.9% of all BBH signals in the Universe [471] (assuming that they are not primordial in origin), as well as enormous numbers of BNS and BHNS signals. As a result, it might not even make sense to talk about a stochastic CBC signal for ET and CE; instead, the goal will likely be to coherently model and subtract the majority of individual CBCs, and to use stochastic methods to search for cosmological GWs beneath this astrophysical foreground [471, 491, 513].

By this time, we also expect to have data from the first *space-based* GW observatory, the Laser Interferometer Space Antenna (LISA) [73]. Freed from the seismic noise that limits the sensitivity of ground-based IFOs below ~ 10 Hz, LISA will be able to search for GWs at much lower frequencies, with a sensitivity peaking in the mHz regime. This will allow LISA probe a completely different population of CBCs, with masses much greater than those of the stellar-mass binaries detected by LIGO/Virgo. (Recall that CBC signals are truncated around the ISCO frequency (0.88), so that binaries with masses greater than $\sim 200\, m_\odot$ merge below the LIGO/Virgo frequency band.) In particular, these more massive sources are expected to include extreme mass-ratio inspirals (EMRIs) [72, 89] in which a stellar-mass object orbits a supermassive black hole (SMBH) with mass $\gtrsim 10^6\, m_\odot$, as well as supermassive binary black holes (SMBBHs) [376] that are expected to form when two galaxies merge. LISA will also prove incredibly useful in searching for cosmological signals, probing cosmic strings with tensions of $G\mu \gtrsim 10^{-17}$ [87],
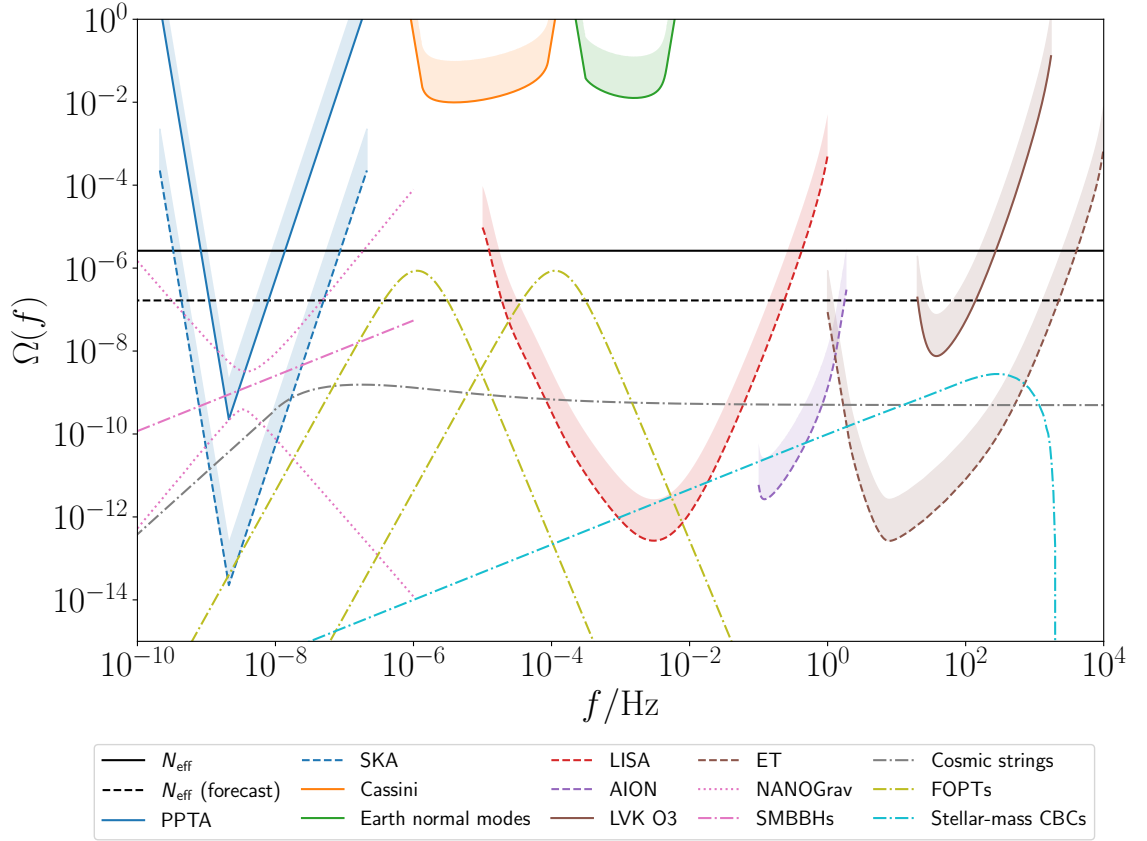
Figure 0.14: An overview of current and future observational bounds on the stochastic gravitational-wave background. Solid curves indicate existing results from the LIGO/Virgo/KAGRA Collaboration (LVK) [36], pulsar timing by the Parkes PTA [388], CMB+BBN constraints from N_{eff} [441], gravimeter monitoring of the Earth's normal modes [217], and Doppler tracking of the Cassini spacecraft [82]. Dashed curves indicate forecast bounds from Einstein Telescope (ET) [464], LISA [73], the Square Kilometre Array (SKA) [339], and the proposed km-scale atom interferometer AION [91], as well as improved N_{eff} constraints [441]. Dot-dashed curves show various potential GWB signals. The light blue curve shows the expected spectrum from stellar-mass CBCs (based on our current knowledge of the cosmic BBH, BNS, and BHNS populations), which we discuss in chapter 1. The grey curve shows an example spectrum from Nambu-Goto cosmic string loops (using 'model 2' of the loop network, with string tension $G\mu = 10^{-11}$), which we discuss in chapter 2. The yellow curves show two first-order phase transition (FOPT) spectra at temperatures $T_* = 2$ GeV and 200 GeV, peaking at $f \approx 1\,\mu$Hz and $\approx 100\,\mu$Hz respectively. The pink dotted curves indicate a range of possible signals associated with the common process (CP) detected by NANOGrav [86], while the overlaid dot-dashed curve shows the median inferred amplitude for the NANOGrav CP when assuming a $\Omega \sim f^{2/3}$ spectrum, as expected for inspiralling supermassive binary black holes (SMBBHs).

and first-order phase transitions at temperatures $T_* \sim 1$ GeV–10^3 TeV [169, 172] (note that this includes transitions at the electroweak scale ~ 200 GeV, which are particularly well-motivated from the point of view of particle physics models).

0.4.2 Pulsar timing

The other main GW detection technique that has been explored to date relies on *millisecond pulsars* [90, 397]: highly-spinning neutron stars that emit strong beams of EM radiation (particularly radio emission) from their magnetic poles. Since these beams are usually misaligned with the pulsar's rotation axis, the EM emission acts like the beam of a lighthouse, only reaching the Earth once per spin period. Pulsar
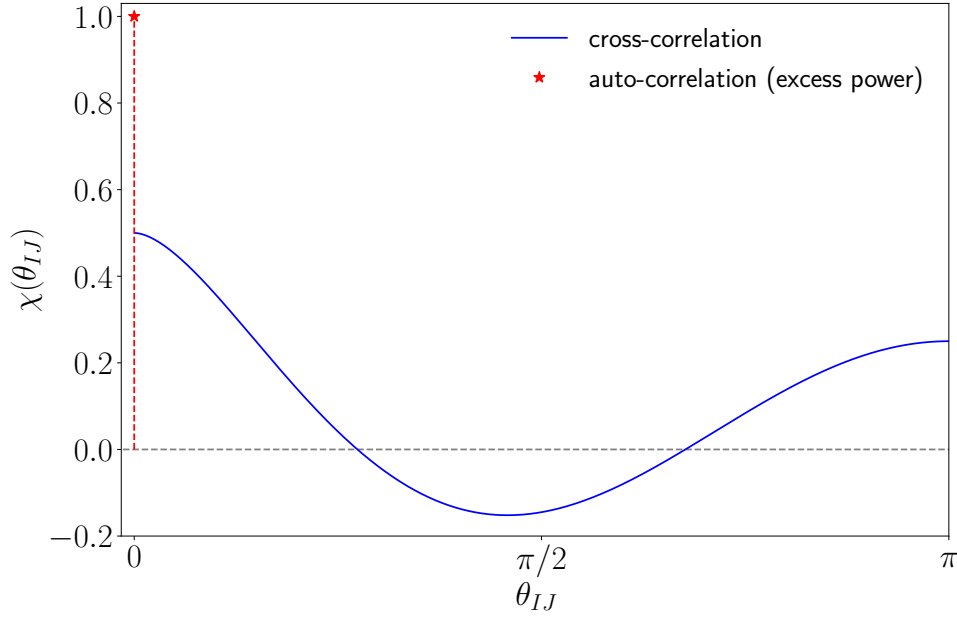
Figure 0.15: The Hellings-Downs function (blue curve) for cross-correlation GWB searches as a function of the angle θ_{IJ} between the two pulsars. The red point emphasises that the signal power is doubled for auto-correlation (i.e., excess power) searches with individual pulsars.

spin periods are incredibly stable over time, and in many cases have been measured to within one part per trillion, allowing us to predict with very high precision the times of arrival (ToAs) of subsequent pulses. Subtracting this 'timing formula' from the observed ToAs then gives us a set of *timing residuals* that directly measure the GW-induced perturbations to the light-travel time between the pulsar and the Earth. These residuals are subject to noise in the radio antenna (which causes a certain level of uncertainty in the reconstruction of each pulse, and therefore in determining the ToA), as well as some level of intrinsic noise in the pulse profiles and emission times from each pulsar, the nature of which is not yet fully understood.

We saw in section 0.2.5 that the strongest GWB searches typically come from cross-correlating data between multiple detectors. In the context of pulsar timing, this corresponds to measuring the ToAs of multiple pulsars and cross-correlating the timing residuals between them [237]. Such an experiment is called a *pulsar timing array* (PTA). The ORF in this case has a simple analytical expression, and is proportional to the *Hellings-Downs curve* [311],

$$\chi(\theta_{IJ}) \equiv \frac{1 + \delta_{IJ}}{2} - \frac{1 - \cos\theta_{IJ}}{2}\left[\frac{1}{4} - \frac{3}{2}\ln\left(\frac{1 - \cos\theta_{IJ}}{2}\right)\right], \tag{0.107}$$

with I and J labelling different pulsars in the array. This depends only on the angle θ_{IJ} between the two pulsars on the sky, and is maximised for pulsars in the same sky direction, $\theta_{IJ} = 0$ (see figure 0.15). It is also possible to perform excess-power searches for each individual pulsar in the array by looking at the auto-correlation of their timing residuals; we see from the Kronecker delta term in equation (0.107) that this auto-correlation signal is a factor of two stronger than the cross-correlation signal for two different pulsars at the same sky location, $\theta_{IJ} = 0$, as a single set of timing residuals is guaranteed to be perfectly in phase with itself.

In equation (0.106), we used the small-antenna approximation to simplify the phase difference measured by an IFO. One important difference between IFOs and PTAs is that the latter are *never* in the small-antenna regime. Since the distances to known pulsars are on the order of $d \sim 0.1$–100 kpc, this regime would correspond to frequencies much lower than $1/d \sim 10^{-10}$–10^{-13} Hz; but GWs with these frequencies would cause much less than a single oscillation in the timing residuals over observational timescales, making them extremely challenging to measure. Essentially, the issue is that the light-travel time across the 'detector' is much longer than the observation time. PTAs are therefore forced to measure GWs with frequencies $f \gg 1/d$, which undergo many oscillations during the light-travel time from the pulsar to the Earth, causing the total GW perturbation (0.105) to average to a smaller value. As a result, PTAs are most sensitive at the lower end of their frequency band, giving a sharp PI curve which grows like $\sim f^5$ [545] (see figure 0.14). This curve is truncated at a minimum frequency corresponding to the inverse of the total observation time, $f_{\min} = 1/T_{\mathrm{obs}}$, which for decade-long pulsar timing campaigns can reach below a few nHz. The key observational target in this frequency band is the GWB from inspiralling SMBBHs [336], which has a characteristic $\sim f^{2/3}$ spectrum, just like the GWB from stellar-mass CBCs in the LIGO/Virgo band. However, PTAs have also proven extremely useful in probing cosmological GW sources—particularly cosmic strings [131], whose GWB spectrum typically contains a peak around this frequency range.

There are three key PTAs currently operating, all of which have been taking data for more than a decade: the Parkes Pulsar Timing Array (PPTA) [318], the European Pulsar Timing Array (EPTA) [380], and the North American Nanohertz Observatory for Gravitational Waves (NANOGrav) [420]. As well as conducting independent GW searches, these three collaborations also perform joint analyses as the International Pulsar Timing Array (IPTA) [409]. The strongest reported upper limit on the nHz GWB comes from a 2015 PPTA analysis, setting $\Omega(f) < 2.3 \times 10^{-10}$ at a reference frequency of $f = \mathrm{yr}^{-1} \approx 32$ nHz [388, 512] (see figure 0.14). However, NANOGrav have recently claimed a detection of *excess power* in the auto-correlation spectra of their timing residuals, with an amplitude and spectral shape that is consistent across all of the pulsars in the array [86]. This 'common process' signal is completely consistent with what one would expect from a GWB, leading several authors to explore how a spectrum with this amplitude and spectral tilt could be generated by various exotic cosmological GW sources [123, 135, 141, 158, 232, 254, 384, 435, 469, 558, 561] (though the spectrum is also perfectly consistent with the GWB from SMBBHs [425]). Consistent findings have very recently been reported by PPTA and EPTA in a re-analysis of their data [190, 288], as well as in a joint analysis by the IPTA [76].

Confirming the NANOGrav common process signal as being due to the GWB will require a detection of the corresponding *cross-correlation* signal; if the measured cross-correlation spectra are consistent with the distinctive shape of the Hellings-Downs curve as a function of the inter-pulsar angle θ_{IJ}, then this will provide extremely strong evidence for GWs. We know from equation (0.107) that the cross-correlation signal is a factor of two smaller than the auto-correlation signal detected by NANOGrav, meaning that the present situation is unsurprising, and also that the cross-correlation signal will likely be within reach once NANOGrav have collected a few more years of data. We have therefore reached a very exciting point in the development of GW astronomy, where NANOGrav and other PTAs appear to be on the brink of making the first-ever detection of the GWB. If this signal is genuine, then future PTA observations such

as those planned for the Square Kilometre Array (SKA) [339] will detect it with extremely large SNR, allowing highly detailed reconstruction of the GWB spectrum in the nHz band.

0.4.3 Cosmological bounds

We saw in section 0.2.3 that the energy density of GWs in the Universe can be thought of in the same terms as the other key components of the cosmic energy inventory: matter, radiation, and dark energy. In fact, since GWs redshift in exactly the same way as photons and other relativistic particles, the GW density parameter $\Omega_{\mathrm{gw}}(f)$ should really be treated as just one part of the total radiation density, in much the same way that CDM and baryons each contribute to the total matter density. This radiation density parameter is crucially important for early-Universe physics, as it determines the expansion rate of the Universe at redshifts $z \gtrsim 3400$. As a result, if the GW energy density $\Omega_{\mathrm{gw}}(f)$ is large enough at early times, it can contribute significantly to the Hubble rate, and thereby leave an imprint on early-Universe observables such as the light element abundances predicted by big-bang nucleosynthesis (BBN), and the temperature and polarisation anisotropies of the CMB [405]. Of course, this only holds for primordial GWs that were already present at these early epochs, such as those from FOPTs; astrophysical sources have no impact on these probes, and even some cosmological sources do not contribute (e.g. the GWs emitted by cosmic string loops or PBH binaries at later times).

The resulting constraints are usually phrased in terms of the effective number of neutrino species, N_{eff}, which is a key model parameter for both BBN and the CMB. Since neutrinos have extremely small masses, they are relativistic at these early epochs, and therefore redshift in the same way as photons and GWs. BBN and CMB analyses typically do not account for the GW energy density, meaning that a given GWB spectrum $\Omega_{\mathrm{gw}}(f)$ is effectively absorbed into the measured value of N_{eff}, giving an extra contribution [405] relative to the Standard Model prediction of $N_{\mathrm{eff}}^{(\mathrm{SM})} \approx 3.045$ [496],

$$N_{\mathrm{eff}} - N_{\mathrm{eff}}^{(\mathrm{SM})} = \frac{16}{7\Omega_\gamma} \left[\frac{g_S(T_*)}{g_S(T_0)} \right]^{4/3} \int_{\ln f_*}^{\infty} \mathrm{d}(\ln f)\, \Omega_{\mathrm{gw}}(f) \approx 2.50 \times 10^5 \int_{\ln f_*}^{\infty} \mathrm{d}(\ln f)\, \Omega_{\mathrm{gw}}(f). \quad (0.108)$$

Here $\Omega_\gamma \approx 5.43 \times 10^{-5}$ is the present-day photon density, and $g_S(T)$ is the number of entropic degrees of freedom in the Universe as a function of photon temperature T, with T_* and T_0 the temperatures at the relevant early-Universe epoch and at the present day, respectively.

We see from equation (0.108) that, rather than constraining the GWB spectrum in a certain frequency band like the IFOs and PTAs we have discussed above, N_{eff} measurements constrain the *integrated* GW energy density across a very broad frequency range. (This range is cut off at a minimum frequency $f_* \approx 10^{-15}$ Hz corresponding to the horizon size at the epoch of BBN, since super-horizon GWs are non-dynamical and thus do not contribute to the GW energy density.) As a result, if we were to detect the GWB through its contribution to N_{eff}, we would have no information about its frequency content, making it extremely difficult to infer the source of the signal. In fact, it would be very difficult to attribute such a detection to the GWB in the first place, as there are numerous other mechanisms for changing the value of N_{eff} (e.g., by adding further relativistic particle species to the Standard Model).

Despite these caveats, N_{eff} measurements still provide very useful constraints on the GWB. The most up-to-date analysis, which combines the Planck 2015 temperature and polarisation angular power spectra

with information from BBN Deuterium abundances, CMB lensing, and baryon acoustic oscillation (BAO) data, gives [441]

$$\int_{\ln f_*}^{\infty} \mathrm{d}(\ln f) \, \Omega_{\mathrm{gw}}(f) < 2.6 \times 10^{-6}, \qquad (0.109)$$

Forecasts by the same authors predict this bound will improve to 1.7×10^{-7} with future CMB and BAO measurements by COrE and EUCLID. Both bounds are shown in figure 0.14, along with the PI curves from various IFOs and PTAs, and the predicted GWB spectra from a variety of astrophysical and cosmological sources.

There are several other cosmological constraints on the GWB that we do not mention here, primarily coming from CMB observations (e.g. through searches for *B*-mode polarisation patterns [48, 358, 504], or GW-induced spectral distortions [375]). However, N_{eff} measurements provide the only cosmological constraint on GWs at the frequencies $f \gtrsim 10^{-10}$ Hz that we are interested in, which is why we have focused on them here.

0.4.4 OTHER SEARCHES

There are many GW experiments that we have omitted from the discussion above. In this section we very briefly highlight a few that are of particular interest, all of which are shown in figure 0.14.

The oldest method of GW detection, pioneered by Weber [582, 583] in the 1960s with his 'resonant bar' experiments, is to monitor the resonant frequencies of some macroscopic test object. When acted upon by a GW with a frequency matching one of the object's normal modes, the stretching and squeezing action they induce excites vibrations in these modes which can be amplified and detected. Numerous such resonant-mass experiments have been operated since the 1960s [56], but none has had sufficient sensitivity to place meaningful bounds on the GWB below the $\Omega(f) < 1$ level required to prevent GWs from over-closing the Universe. However, the same idea has been applied with success on a much larger scale, using the Earth itself as a resonant mass. By monitoring the Earth's normal mode frequencies using gravimeter data, Coughlin and Harms [217] placed an upper limit on the GWB in the mHz frequency band. The corresponding PI curve has a minimum of $\Omega(f) \approx 1.3 \times 10^{-2}$ at $f \approx 1.6$ mHz. Recently, a trio of similar experiments have been proposed in which gravimeters installed on the Lunar surface could be used to monitor the Moon's normal modes, providing a cleaner probe of GWs in the mHz frequency band [301, 338, 363].

Another GW experiment which has successfully constrained the GWB below the $\Omega(f) < 1$ level is the work of Armstrong *et al.* [82], who carried out Doppler-tracking observations of the Cassini spacecraft. The principle here is more similar to IFOs and PTAs in that it tracks the trajectories of photons in the presence of GWs; however, rather than measuring changes in the light-travel time, one attempts to measure changes in the photon frequency due to GW-induced Doppler shifts between the two test masses (this is essentially the time derivative of the perturbation to the light-travel time). This analysis probed GWs in the 1–100 μHz band, with the corresponding PI curve reaching a minimum of $\Omega(f) \approx 9.9 \times 10^{-3}$ at $f \approx 3.9$ μHz.

Finally, a GW detection technique which is subject to growing interest in the community is *atom interferometry*, in which, rather than using interference between photons to measure the phase difference

between two spatial paths, one exploits the wavelike behaviour of matter on the quantum scale to measure interference between atoms. While first-generation atom interferometry experiments such as AION [91] and MAGIS [43, 290] are in their early stages and are not expected to have significant sensitivity to GWs, proposals for future km-scale experiments are forecast to provide extremely useful GWB constraints in the 0.1–1 Hz band. In particular, the AION-km proposal has a forecast sensitivity of $\Omega(f) \approx 2.7 \times 10^{-12}$ at $f \approx 0.12$ Hz.

1 ANISOTROPIES IN THE GRAVITATIONAL-WAVE BACKGROUND

In section 0.2 we introduced the *gravitational-wave background*, and argued that, by giving us access to physics at high redshift, this provides one of our most powerful tools for probing cosmology with GW observations. We also discussed the standard set of statistical assumptions that are usually made about the GWB: that it has no phase correlations between different sky directions, and that it is Gaussian, stationary, unpolarised, and isotropic. This last assumption, however, is particularly restrictive and unrealistic in the context of the inhomogeneous Universe we inhabit, with its beautiful and fascinating array of structures on a vast range of scales. In order to put the GWB on an equal footing with other cosmological observables, we must treat it as *anisotropic* (though for extragalactic sources it is still *statistically* isotropic—see section 1.1 below).

Aside from providing a more faithful description of reality, one of the key motivations for treating the GWB as anisotropic is the *information content* of the anisotropies, which is generally complementary to that of the isotropic component. As we saw with the 'Phinney formula' (0.64) in section 0.2.4, the isotropic GWB spectrum encodes the average rate, number density, and energy spectra of GW sources in the Universe, as well as the isotropic expansion rate of the FLRW background. Meanwhile, the anisotropies, as well as depending on all of these ingredients, also encode the *spatial clustering* of GW sources, and the inhomogeneities in the spacetime metric that the GWs encounter as they propagate towards us. This situation is entirely analogous to other cosmological observables such as the CMB: while the isotropic component (i.e., the mean CMB temperature $T_0 \approx 2.73\,\mathrm{K}$) is sensitive to the expansion history of the homogeneous background, the anisotropies reveal a treasure trove of information about the origins and dynamical evolution of perturbations in the Universe's matter content and spacetime metric.

Of course, given the historical headstart electromagnetic astronomy has had relative to GW astronomy, the observational status of EM probes such as the CMB is vastly more advanced than that of the GWB. As emphasised by Romano [486], it took half a century for CMB observations to progress from Penzias and Wilson's initial discovery of the isotropic component to the exquisitely detailed maps of the anisotropies provided by Planck, whereas in 2021 we're yet to even detect the isotropic component of the GWB. It therefore seems unlikely that studies of the GWB anisotropies will be able to compete with EM probes any time soon in terms of, say, constraining the values of cosmological parameters like the Hubble constant. So why study GWB anisotropies?

The key motivation for GWB searches, in my view, lies in their *discovery potential*. While EM observables each probe a known set of sources described by well-understood physics, the GWB has the potential to unveil previously-unknown sources across an enormous redshift range, thereby giving us the opportunity to probe exotic new physics. The sources we discussed in section 0.3—first-order phase transitions,

cosmic strings, and primordial black holes—are just a few of the most prominent possibilities studied in the literature, any of which would revolutionise our understanding of fundamental physics if detected. Studying the anisotropies associated with each of these GWB signals will likely be crucial in distinguishing and characterising the underlying sources, enabling us to extract as much cosmological information as possible from our GW data. Another key motivation is that searches for GWB anisotropies are already underway, with LIGO/Virgo in particular providing impressively strong directional upper limits on the GWB intensity [9, 25, 38]. As these searches become increasingly sensitive, it is important for us to understand what they should expect to see, so that the results can be properly interpreted.

For these reasons, there has been a recent surge of interest in modelling the expected spectrum of anisotropies in the GWB [62, 100–103, 117, 174, 209, 210, 221–225, 285, 342, 345, 346, 349, 382, 392, 457, 559, 578], and their cross-correlations with EM observables [50, 153, 221, 224, 312, 408, 481], as well as in developing associated data analysis techniques [57, 65, 66, 70, 71, 95, 212, 320, 344, 423, 444, 474–476, 478, 532, 533, 592]. While this body of work has included studies of the anisotropies associated with inflation [62, 102, 103], FOPTs [285, 382], cosmic strings [345, 383, 439], and PBHs [101, 578], the primary focus has been on the *astrophysical* GW background (AGWB) from compact binary coalescences. As I see it, there are three key reasons why this signal is interesting:

1. Since the black holes and neutron stars that give rise to the AGWB are created via stellar evolution, we expect them to reside in galaxies and other star-forming environments. As a result, these CBCs act as *tracers* of the inhomogeneous distribution of galaxies, and therefore of the cosmic matter distribution. By studying the statistical clustering of this background signal on the sky, we can therefore probe the large-scale structure of the Universe at late times, and gain insights into the nonlinear gravitational dynamics that governs these structures. While this LSS is already being measured by late-Universe EM observations such as galaxy surveys, there are several features of GWB searches which might allow us to access new and complementary information compared to these existing probes: namely that they automatically include the full sky and extend to arbitrarily high redshifts (in contrast with galaxy surveys, which typically cover only a fraction of the sky, and are inevitably limited in their depth), as well as using a different population of tracers of the matter distribution (i.e., CBC-hosting galaxies), whose clustering properties will, in general, differ from those used by existing cosmological probes.

2. Aside from providing novel cosmological information, the GWB from CBCs also has the potential to reveal interesting *astrophysical* information, principally about the rates and masses of CBCs and how these evolve throughout cosmic history. This could help us answer some of the many open questions about these systems; for example, what fraction of them originate from isolated binary evolution, and what fraction are dynamically assembled in dense stellar environments [485]? Are there other important formation channels, e.g., in AGN discs [106]? What are the mass distributions of BHs and NSs, and how do these depend on the stellar physics of their progenitors [287]? What are the properties of the host galaxies of these CBCs [49], and how do these impact the CBCs themselves? While individually-resolved CBCs can help us try to pin down these questions

at low redshifts, the properties of the AGWB and its anisotropies could help us paint a much more complete picture, covering the full history of star formation in the Universe.

3. Perhaps most importantly of all, the AGWB is likely to be the loudest stochastic signal across a broad range of frequencies, and is therefore likely to the first component of the GWB that we detect. In order to dig out the cosmological signals which may be lurking beneath it, and thereby exploit the GWB to its full potential, we are thus obliged to understand the AGWB as thoroughly as possible, including a complete characterisation of its anisotropies.

With these motivations in mind, the goal of this chapter is to investigate the anisotropies in the AGWB, focusing on the background from stellar-mass CBCs that is the target of ground-based interferometers such as LIGO/Virgo. We begin in section 1.1 by introducing some of the necessary tools for studying the GWB as a random field on the sphere; this is almost entirely a review of pre-existing results, though the presentation is somewhat different to that which one usually encounters in the GWB literature. In section 1.2 we calculate predictions for the AGWB spectrum, starting with the isotropic component, before constructing a simulated map of the AGWB's anisotropies using data from the Millennium *N*-body simulation. In section 1.3 we examine the issue of *shot noise* as an obstacle for AGWB searches, deriving the expected amplitude of this noise, before developing an optimal statistical methodology for mitigating its impact on measurements of the angular power spectrum. We summarise our results in section 1.4, and discuss some of the ways forward for future studies of GWB anisotropies.

1.1 CHARACTERISING THE ANISOTROPIC BACKGROUND

In section 0.2.2, we saw that the plane wave components of the GWB strain, $\bar{h}_A(f, \hat{r})$, must be treated as random variables due to our ignorance of their incoherent phases, and that under the standard set of assumptions the statistics of these components are fully characterised by a single function of frequency. In section 0.2.3, we identified this function with the GW density parameter $\Omega(f)$. Motivated by the discussion above, we now relax the assumption of perfect isotropy, so that the GWB strain moments become

$$\left\langle \bar{h}_A(f, \hat{r}) \right\rangle = 0, \qquad \left\langle \bar{h}_A(f, \hat{r})\bar{h}_{A'}^*(f', \hat{r}') \right\rangle = \frac{3\pi H_0^2}{(2\pi|f|)^3}\Omega(f, \hat{r})\delta_{AA'}\delta(f - f')\delta^{(2)}(\hat{r}, \hat{r}'), \quad (1.1)$$

where we have promoted the GWB intensity to a field on the 2-sphere,

$$\Omega(f, \hat{r}) \equiv \frac{1}{\rho_c}\frac{d\rho_{gw}}{d(\ln f)\, d^2\hat{r}}. \quad (1.2)$$

This has units of inverse steradians, and when integrated gives the isotropic GWB spectrum from before,

$$\Omega(f) = \int_{S^2} d^2\hat{r}\, \Omega(f, \hat{r}). \quad (1.3)$$

57

1.1.1 THE GRAVITATIONAL-WAVE BACKGROUND AS A RANDOM FIELD

Previously, we treated $\Omega(f)$ as a deterministic (i.e., non-random) quantity, fixed by the average properties of the cosmic population of GW sources. However, when studying the distribution of GW intensity on the sky, $\Omega(f, \hat{\boldsymbol{r}})$, we are faced once again with a fundamental degree of uncertainty, similar to our ignorance about the phases of the strain components $\tilde{h}_A(f, \hat{\boldsymbol{r}})$. In this case, the uncertainty is due to our lack of knowledge about the angular positions of individual GW sources on the sky. For example, while we can attempt to deterministically model the average rate of CBC signals that contribute to the GWB, there is no hope of modelling where on the sky these signals should come from without a complete knowledge of the positions of all the BBHs and BNSs in the Universe. We are therefore forced to treat $\Omega(f, \hat{\boldsymbol{r}})$ itself as essentially random.[1]

Of course, the above argument doesn't hold for GW sources which inhabit a preferred set of directions on the sky. For example, the stochastic signal from white dwarf binaries that will be observed by LISA will come primarily from the galactic plane, and it should therefore be possible to model the angular distribution of this signal using the known distribution of stars within the galaxy. However, we are interested here in *extragalactic* sources, for which there are no preferred directions. (Or at least, if there are preferred directions, these depend on information we do not have access to, such as the locations of CBCs throughout the Universe.) We therefore assume that the GWB is *statistically isotropic*, such that all ensemble-averaged functions of $\Omega(f, \hat{\boldsymbol{r}})$ are invariant under sky rotations. In particular, this implies that the mean of the GWB intensity field is independent of direction, and that the covariance between any two field points depends only on the invariant angle between those points,

$$\langle \Omega(f, \hat{\boldsymbol{r}}) \rangle \equiv \bar{\Omega}(f), \qquad \mathrm{Cov}[\Omega(f, \hat{\boldsymbol{r}}), \Omega(f, \hat{\boldsymbol{r}}')] \equiv C(f, \hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}'). \tag{1.4}$$

If the GWB is a Gaussian random field, then these two moments are sufficient to fully describe its statistics.

Note that we haven't written down the most general covariance function here, which would specify what happens when we cross-correlate two different frequencies, f and f', as well as two different sky directions. In fact, this is not necessary, since we generally expect that it is possible to factorise the GWB into a frequency spectrum (which is deterministic) and an angular intensity map (which is a random field),

$$\Omega(f, \hat{\boldsymbol{r}}) = \mathcal{H}(f) \Omega(\hat{\boldsymbol{r}}). \tag{1.5}$$

To see why this might be a reasonable assumption, it is useful to focus for a moment on the AGWB in the LIGO/Virgo frequency band. In this case, the frequency spectrum encodes the characteristic $\sim f^{2/3}$ emission of each of the individual CBCs as they 'chirp' up and approach merger, while the angular intensity encodes the distribution of CBCs throughout the Universe. Since each CBC spans the entire frequency range of interest while it is in-band, there is no distinction between the angular distribution of 'high-frequency CBCs' and 'low-frequency CBCs'—they are just the same sources, with the same distribution on the sky, and we can therefore capture this distribution by factorising out the frequency

---

[1] Note that we now have two distinct levels of randomness in the GWB strain: the randomness of the incoherent phases, and the randomness of the intensity as a function of sky direction. It is important to stress that these are two logically distinct sources of randomness, with the former being tied to the random GW emission times of different sources in the Universe, and the latter tied to the spatial clustering of those sources due to LSS. We will return to this point in section 1.3.5.

spectrum as in equation (1.5).[2] We therefore suppress the frequency arguments in equation (1.4), and focus on the angular intensity $\Omega(\hat{\boldsymbol{r}})$ at some reference frequency.

1.1.2 SPHERICAL HARMONICS

The statistical invariance of the GWB under rotations means that it's not actually very useful to think about Ω as a function of sky direction, since its value at any given direction $\hat{\boldsymbol{r}}$ is entirely dependent on our arbitrary choice of spherical coordinates. This is exactly the same reason why it's not useful to describe the GWB strain in the time domain, as statistical time-translation invariance (i.e., stationarity) means that specifying the value of the strain at any given time t is not very informative. Instead, we treat the strain in the frequency domain, projecting $h_A(t, \hat{\boldsymbol{r}})$ onto the Fourier modes $e^{2\pi i f t}$, which provide a complete and orthogonal set of basis functions on the real line $\mathbb{R}$. Stationarity then means that the phases of the Fourier components $\tilde{h}_A(f, \hat{\boldsymbol{r}})$ are uninformative, but their amplitudes *are* useful, allowing us to describe the statistics of the GWB strain over different timescales (i.e., different frequencies) in terms of the power spectrum (which is proportional to $\Omega(f, \hat{\boldsymbol{r}})$). Ideally we would like to do something analogous here, projecting $\Omega(\hat{\boldsymbol{r}})$ onto some set of basis functions on the sphere S^2, and thereby describing the statistics of the field on different *angular* scales.

The obvious candidate for these basis functions is the set of *spherical harmonics* $Y_{\ell m}$ (see figure 1.1), which are defined as solutions to the equation $[\nabla^2 + \ell(\ell + 1)]g(\hat{\boldsymbol{r}}) = 0$, where ℓ is any non-negative integer. Note that this is analogous to the equation $(\partial_t^2 + \omega^2)g(t) = 0$ obeyed by the Fourier modes $e^{i\omega t}$, with ℓ here playing the rôle of 'angular frequency'. One important difference that comes with the higher dimensionality is an increase in the number of independent solutions of the same frequency: for any given ℓ there are $2\ell + 1$ linearly independent harmonics $Y_{\ell m}(\hat{\boldsymbol{r}})$, which we label with the index $m \in \{-\ell, -\ell + 1, \ldots, +\ell\}$ such that $|m| \leq \ell$. These form a complete basis on the 2-sphere S^2, and are orthonormal,

$$\int_{S^2} \mathrm{d}^2\hat{\boldsymbol{r}} \, Y_{\ell m}(\hat{\boldsymbol{r}}) Y_{\ell' m'}^*(\hat{\boldsymbol{r}}) = \delta_{\ell \ell'} \delta_{m m'}. \tag{1.6}$$

The first few spherical harmonics are

$$Y_{00}(\hat{\boldsymbol{r}}) = \sqrt{\frac{1}{4\pi}},$$

$$Y_{10}(\hat{\boldsymbol{r}}) = \sqrt{\frac{3}{4\pi}} \cos\theta, \qquad Y_{11}(\hat{\boldsymbol{r}}) = -\sqrt{\frac{3}{8\pi}} \sin\theta \, e^{i\phi},$$

$$Y_{20}(\hat{\boldsymbol{r}}) = \sqrt{\frac{5}{16\pi}} (3\cos^2 - 1), \qquad Y_{21}(\hat{\boldsymbol{r}}) = -\sqrt{\frac{15}{8\pi}} \sin\theta \cos\theta \, e^{i\phi}, \qquad Y_{22}(\hat{\boldsymbol{r}}) = \sqrt{\frac{15}{32\pi}} \sin^2\theta \, e^{2i\phi},$$

$$\tag{1.7}$$

with the expressions for negative values of m given by the identity

$$Y_{\ell m}^*(\hat{\boldsymbol{r}}) = (-)^m Y_{\ell, -m}(\hat{\boldsymbol{r}}). \tag{1.8}$$

---

[2]This argument does not hold in all settings. For example, many of the CBCs that will be observed by LISA will undergo negligible frequency evolution over observational timescales (c.f. equation (0.89)), meaning that there *is* then a distinction between the CBC populations emitting at different frequencies, and this could in principle lead to interesting frequency-dependent effects in the two-point statistics of the GWB. We do not investigate this further here.
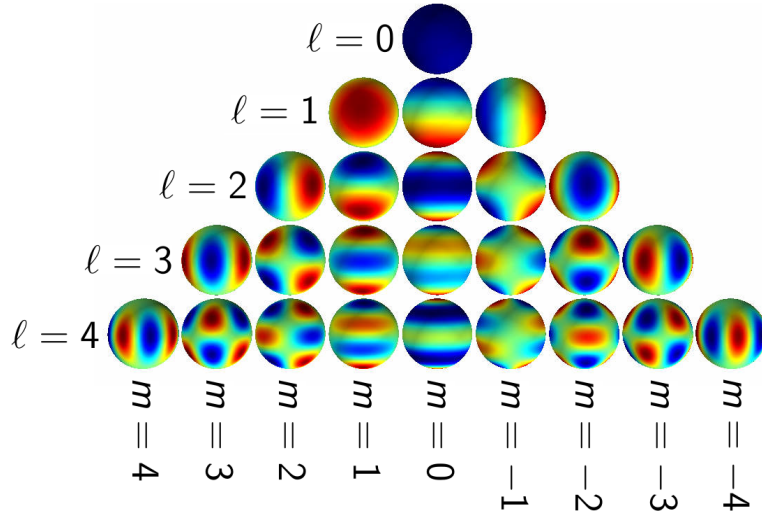
Figure 1.1: Visualisations of the real parts of the first few spherical harmonics, $\mathrm{Re}\{Y_{\ell m}(\hat{\boldsymbol{r}})\}$, for multipoles $\ell \le 4$. The polar axis ($\theta = 0, \pi$) is oriented vertically.

We therefore decompose the GWB intensity into spherical harmonics,

$$\Omega_{\ell m} \equiv \int_{S^2} \mathrm{d}^2\hat{\boldsymbol{r}}\, Y_{\ell m}^*(\hat{\boldsymbol{r}})\Omega(\hat{\boldsymbol{r}}), \tag{1.9}$$

allowing us to replace a real random field $\Omega : S^2 \to \mathbb{R}$ with an infinite set of complex random numbers $\Omega_{\ell m} \in \mathbb{C}$, which we call the spherical harmonic components (SHCs) of the field. Note that the fact that $\Omega(\hat{\boldsymbol{r}})$ is real, combined with equation (1.8), implies that $\Omega_{\ell m}^* = (-)^m \Omega_{\ell,-m}$. We can invert equation (1.9) to write

$$\Omega(\hat{\boldsymbol{r}}) = \sum_{\ell m} Y_{\ell m}(\hat{\boldsymbol{r}})\Omega_{\ell m}, \tag{1.10}$$

where we use the shorthand notation $\sum_{\ell m} \equiv \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{+\ell}$ for brevity. Equations (1.9) and (1.10) are the spherical analogues of the Fourier transform and its inverse. One important difference with frequency spectra on $\mathbb{R}$ is that the spherical harmonic spectrum (1.9) is discrete rather than continuous, which we can understand as a consequence of S^2 being compact (i.e., having finite volume).

1.1.3 THE ANGULAR POWER SPECTRUM

Now that we have decomposed the GWB into spherical harmonics, we would like to replace the field moments in equation (1.4) with the first and second moments of the SHCs, writing these in terms of the constant $\bar{\Omega}$ and the function $C(\hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}')$. We can obtain the first moment pretty much immediately by writing

$$\langle \Omega_{\ell m} \rangle = \bar{\Omega}\int_{S^2} \mathrm{d}^2\hat{\boldsymbol{r}}\, Y_{\ell m}^*(\hat{\boldsymbol{r}}) = \sqrt{4\pi}\,\bar{\Omega}\,\delta_{\ell 0}\delta_{m0}, \tag{1.11}$$

where in the second equality we have used the orthonormality condition (1.6) with the $\ell = m = 0$ harmonic from equation (1.7). This tells us that all of the SHCs have zero mean, except for the monopole Ω_{00}.

For the second moment, it is useful to note that since the dot product $\hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}'$ always lies in the interval $[-1, 1]$, the function $C(\hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}')$ can be decomposed into Legendre polynomials $P_\ell(\hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}')$, which form a complete basis for functions on this interval. We therefore write

$$C(\hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}') \equiv \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} C_\ell P_\ell(\hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}'), \tag{1.12}$$

where the coefficients C_ℓ are a set of real numbers describing the decomposition of the covariance function in this basis, and the normalisation factor $(2\ell+1)/(4\pi)$ is chosen such that we can write the inverse expression as

$$C_\ell = \int_{S^2} \mathrm{d}^2\hat{\boldsymbol{r}}' \, P_\ell(\hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}') C(\hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}'), \tag{1.13}$$

using the orthogonality relation for the Legendre polynomials, $\int_{-1}^{+1} \mathrm{d}x \, P_\ell(x) P_{\ell'}(x) = 2\delta_{\ell\ell'}/(2\ell+1)$.

The Legendre polynomials are related to the spherical harmonics of the same angular scale ℓ by the *addition theorem* [247]

$$P_\ell(\hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}') = \sum_{m=-\ell}^{+\ell} \frac{4\pi}{2\ell+1} Y_{\ell m}(\hat{\boldsymbol{r}}) Y_{\ell m}^*(\hat{\boldsymbol{r}}'), \tag{1.14}$$

which allows us to write equation (1.12) as

$$C(\hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}') = \sum_{\ell m} C_\ell Y_{\ell m}(\hat{\boldsymbol{r}}) Y_{\ell m}(\hat{\boldsymbol{r}}'). \tag{1.15}$$

We can now derive the second moment of the SHCs,[3]

$$\begin{aligned}
\mathrm{Cov}[\Omega_{\ell m}, \Omega_{\ell' m'}] &= \int_{S^2} \mathrm{d}^2\hat{\boldsymbol{r}} \int_{S^2} \mathrm{d}^2\hat{\boldsymbol{r}}' \, Y_{\ell m}(\hat{\boldsymbol{r}}) Y_{\ell' m'}^*(\hat{\boldsymbol{r}}') C(\hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}') \\
&= \sum_{LM} C_L \left[\int_{S^2} \mathrm{d}^2\hat{\boldsymbol{r}} \, Y_{\ell m}(\hat{\boldsymbol{r}}) Y_{LM}^*(\hat{\boldsymbol{r}}) \right] \left[\int_{S^2} \mathrm{d}^2\hat{\boldsymbol{r}}' \, Y_{\ell' m'}^*(\hat{\boldsymbol{r}}') Y_{LM}(\hat{\boldsymbol{r}}') \right] \\
&= \sum_{LM} C_L \delta_{\ell L} \delta_{mM} \delta_{\ell' L} \delta_{m'M} = C_\ell \delta_{\ell\ell'} \delta_{mm'},
\end{aligned} \tag{1.16}$$

which shows us that the SHCs $\Omega_{\ell m}$ are all uncorrelated from each other (which, if they are Gaussian, further implies that they are statistically independent). This illustrates the usefulness of going to spherical harmonic space, since having a diagonal covariance matrix like equation (1.16) is extremely convenient for both theory and data analysis. Compare with the covariance matrix in field space (also called 'pixel space', since in practice the sky is usually partitioned into a finite number of pixels), $\mathsf{C}_{ij} \equiv C(\hat{\boldsymbol{r}}_i \cdot \hat{\boldsymbol{r}}_j)$, in which the off-diagonal terms are generically nonzero (otherwise the field would have no spatial correlations at all), making it much more cumbersome to work with. Again, the analogy with Fourier analysis is clear: one of the key strengths of working in Fourier space is that modes of different frequency are usually uncorrelated, giving a diagonal covariance matrix.

---

[3]Note that here, and throughout, we define the covariance of two complex random variables with a complex conjugate on the second argument, i.e., $\mathrm{Cov}[X, Y] \equiv \langle XY^* \rangle - \langle X \rangle \langle Y^* \rangle = \mathrm{Cov}[Y, X]^*$. This ensures that the variance is always real, since $\mathrm{Var}[X] \equiv \mathrm{Cov}[X, X] = \langle |X|^2 \rangle - |\langle X \rangle|^2$.

We have thus shown that we can replace equation (1.4) with[4]

$$\langle \Omega_{\ell m} \rangle = \sqrt{4\pi}\,\bar{\Omega}\,\delta_{\ell 0}, \qquad \mathrm{Cov}[\Omega_{\ell m}, \Omega_{\ell' m'}] = C_\ell\,\delta_{\ell\ell'}\,\delta_{mm'}, \qquad (1.17)$$

which is analogous to equation (1.1) for the Fourier components of the GWB strain. In particular, the set of numbers C_ℓ is analogous to the Fourier power spectrum; we call it the *angular power spectrum*. One important difference is that the monopole Ω_{00} has a nonzero mean value, as opposed to the Fourier components, which are all zero-mean. This is because, while we can set the mean value of the strain $\langle h_{ij}(t) \rangle$ to zero by a coordinate transformation, the mean GWB intensity $\bar{\Omega}$ must always be positive— otherwise there would be no GWB to observe.

Intuitively, the angular power spectrum C_ℓ tells us about the amplitude of statistical fluctuations in the intensity field on angular scales $\theta \sim \pi/\ell$. We can make this notion more precise by using equation (1.12) to write the variance of a single field point $\hat{\boldsymbol{r}}$ as[5]

$$\mathrm{Var}[\Omega(\hat{\boldsymbol{r}})] \equiv C(\hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}) = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} C_\ell \approx \int \mathrm{d}(\ln\ell)\,\frac{\ell(\ell+1)}{2\pi} C_\ell, \qquad (1.18)$$

where we have used the fact that $P_\ell(1) = 1$. The quantity $\ell(\ell+1)C_\ell/(2\pi)$ thus approximately describes the field variance associated with a given logarithmic bin in ℓ, and is therefore the quantity that is usually presented in, e.g., CMB analyses, rather than the bare angular power spectrum C_ℓ. Comparing with the GWB strain once again, $\ell(\ell+1)C_\ell/(2\pi)$ is analogous to the density parameter $\Omega(f)$, while C_ℓ is analogous to the strain power spectrum $S(f) \sim f^{-3}\Omega(f)$. Having $\ell(\ell+1)C_\ell = $ constant corresponds to a scale-invariant angular power spectrum, while having $C_\ell = $ constant corresponds to a white spectrum.

1.1.4 Estimating the angular power spectrum

Suppose we detect the GWB and measure some set of SHCs, $\Omega_{\ell m}$. What can we say about the angular power spectrum based on these measurements? (We ignore here the problem of actually inferring the SHCs from noisy strain data, and the associated uncertainties in their values—see, e.g., Thrane *et al.* [543] or Romano and Cornish [487] for thorough treatments of these topics.)

The SHCs that we observe represent just a single random realisation from the distribution parameterised by the angular power spectrum. As such, we cannot measure these distribution parameters directly; the best we can do is construct some function of the observed SHCs whose ensemble-averaged value is related to the parameters we're interested in. We call such a function an *estimator*. If the ensemble average of the estimator is equal to the parameter we're trying to infer, then we call it *unbiased*.

---

[4] Here we replace $\delta_{\ell 0}\delta_{m0}$ with just $\delta_{\ell 0}$. We can always do this, since having $\ell = 0$ necessarily implies that $m = 0$, in order to satisfy $|m| \leq \ell$.

[5] At first glance, it seems like there should be a factor of $\ell(\ell+1/2)$ here rather than $\ell(\ell+1)$. However, one can show that $(\ell+1/2)/[\ln(\ell+1) - \ln\ell] \simeq \ell(\ell+1) + O(1)$ for $\ell \gg 1$, so this is actually the consistent choice when approximating the sum as an integral. In any case, the different is negligible in the $\ell \gg 1$ regime where this approximation holds.

From equation (1.17), we see immediately that we can define the naive estimator $\hat{C}_\ell^{(m)} \equiv |\Omega_{\ell m}|^2$, which is unbiased for all $\ell > 0$,

$$\left\langle \hat{C}_\ell^{(m)} \right\rangle = \mathrm{Var}[\Omega_{\ell m}] + |\langle \Omega_{\ell m} \rangle|^2 = C_\ell + 4\pi \bar{\Omega}^2 \delta_{\ell 0}. \tag{1.19}$$

If we knew the value of $\bar{\Omega}$, we could make this unbiased for all ℓ by defining $\hat{C}_\ell^{(m)} \equiv |\Omega_{\ell m} - \sqrt{4\pi}\,\bar{\Omega}\,\delta_{\ell 0}|^2$; however, $\bar{\Omega}$ is not a known quantity, but another parameter to be estimated.

For any given set of measured SHCs, the estimator $\hat{C}_\ell^{(m)}$ will not be exactly equal to its average value, but will be subject to some random scatter. We can quantify this by calculating the variance of the estimator,

$$\mathrm{Var}[\hat{C}_\ell^{(m)}] = \left\langle |\Omega_{\ell m}|^4 \right\rangle - \left\langle |\Omega_{\ell m}|^2 \right\rangle^2 = \left[2\left\langle |\Omega_{\ell m}|^2 \right\rangle^2 + \left\langle (\Omega_{\ell m})^2 \right\rangle \left\langle (\Omega_{\ell m}^*)^2 \right\rangle \right] - \left\langle |\Omega_{\ell m}|^2 \right\rangle^2$$
$$= \left\langle |\Omega_{\ell m}|^2 \right\rangle^2 + (-)^{2m} \left\langle \Omega_{\ell m} \Omega_{\ell,-m}^* \right\rangle \left\langle \Omega_{\ell,-m} \Omega_{\ell m}^* \right\rangle = (1 + \delta_{m0}) C_\ell^2, \tag{1.20}$$

where we have assumed Gaussianity in order to write the fourth moment in terms of second moments. We see that this is a rather poor estimator, as its standard deviation is equal to (or greater than, for $m = 0$) the mean value we want to extract, meaning that we can typically expect our estimate to be off by 100%.

Normally with a problem like this we would try to make as many independent measurements as possible, and combine all of these to create an estimator with lower variance. More explicitly, if we have N independent and identically distributed (i.i.d.) unbiased estimators $\hat{X}_i$, $i = 1, 2, \ldots, N$ for some parameter X, each with some variance σ^2, then we can define an unbiased mean estimator $\hat{X} \equiv (1/N) \sum_{i=1}^N \hat{X}_i$ whose variance is reduced by a factor of $1/N$,

$$\mathrm{Var}[\hat{X}] = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \mathrm{Cov}[\hat{X}_i, \hat{X}_j] = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \sigma^2 \delta_{ij} = \frac{\sigma^2}{N}. \tag{1.21}$$

In cosmological contexts, taking more measurements is not an option, as nature presents us with a single random realisation of LSS. This results in a certain irreducible uncertainty when trying to estimate ensemble-averaged quantities like the angular power spectrum from this single realisation. We call this uncertainty *cosmic variance*.

Despite this fundamental limitation, however, we can still do significantly better than the naive estimator (1.19). The key here is to note that, while we only have one random realisation of each SHC $\Omega_{\ell m}$, the $(2\ell + 1)$ different m indices for a given ℓ index are i.i.d., and therefore act as independent measurements of the same C_ℓ multipole in the angular power spectrum. We can therefore average over these to define an improved estimator,

$$\hat{C}_\ell \equiv \frac{1}{2\ell + 1} \sum_m |\Omega_{\ell m}|^2, \tag{1.22}$$

whose variance is

$$
\begin{aligned}
\mathrm{Var}[\hat{C}_\ell] &= \frac{1}{(2\ell+1)^2} \sum_m \sum_{m'} \mathrm{Cov}\big[|\Omega_{\ell m}|^2, |\Omega_{\ell m'}|^2\big] \\
&= \frac{1}{(2\ell+1)^2} \sum_m \sum_{m'} \big[\langle \Omega_{\ell m} \Omega_{\ell m'}^* \rangle \langle \Omega_{\ell m}^* \Omega_{\ell m'} \rangle + \langle \Omega_{\ell m} \Omega_{\ell m'} \rangle \langle \Omega_{\ell m}^* \Omega_{\ell m'}^* \rangle \big] \quad (1.23) \\
&= \frac{1}{(2\ell+1)^2} \sum_m \sum_{m'} C_\ell^2 (\delta_{mm'} + \delta_{m,-m'}) = \frac{2}{2\ell+1} C_\ell^2.
\end{aligned}
$$

This is significantly better than the variance in equation (1.20), particularly at higher multipoles $\ell \gg 1$. In fact, if we assume the intensity field to be Gaussian, we can prove that the improved estimator (1.22) is the best that we can construct, in the sense that it is the minimum-variance unbiased estimator (MVUE) for the angular power spectrum. To show this, we use the fact that the variance of any unbiased estimator $\hat{\theta}(x)$ for a parameter θ of a probability distribution $p(x|\theta)$ obeys the *Cramér-Rao bound* [366],

$$
\mathrm{Var}[\hat{\theta}] \geq -\left\langle \frac{\partial^2 \mathcal{L}}{\partial \theta^2} \right\rangle^{-1}, \quad (1.24)
$$

where $\mathcal{L}(x|\theta) \equiv \ln p(x|\theta)$ is the log-likelihood for the distribution. In our case, the Gaussian log-likelihood for the SHCs is

$$
\mathcal{L}(\Omega_{\ell m}|\bar{\Omega}, C_\ell) = -\frac{1}{2} \sum_{\ell m} \left[\ln(2\pi C_\ell) + \frac{|\Omega_{\ell m} - \sqrt{4\pi}\,\bar{\Omega}\,\delta_{\ell 0}|^2}{C_\ell} \right], \quad (1.25)
$$

so equation (1.24) becomes

$$
\mathrm{Var}[\hat{C}_\ell] \geq \left\langle \sum_m \left[\frac{|\Omega_{\ell m} - \sqrt{4\pi}\,\bar{\Omega}\,\delta_{\ell 0}|^2}{C_\ell^3} - \frac{1}{2C_\ell^2} \right] \right\rangle^{-1} = \left(\sum_m \frac{1}{2C_\ell^2} \right)^{-1} = \frac{2}{2\ell+1} C_\ell^2, \quad (1.26)
$$

which is identical to what we found in equation (1.23). This corresponds to the irreducible uncertainty on the angular power spectrum due to cosmic variance. We thus see that the estimator (1.22) saturates this bound, meaning that it must be the MVUE for all $\ell > 0$.

For the monopole $\ell = 0$ we still have the issue of the estimator being biased, due to our inability to subtract off the unknown mean intensity $\bar{\Omega}$. All that we can do with the monopole is construct an estimator for the mean intensity, $\hat{\Omega} \equiv \Omega_{00}/\sqrt{4\pi}$; this is unbiased, $\langle \hat{\Omega} \rangle = \bar{\Omega}$, and has variance $\mathrm{Var}[\hat{\Omega}] = C_0/(4\pi)$. The problem here is that we are attempting to infer both the mean and the variance of a population (the ensemble of all possible realisations of the GWB monopole) with only a single sample (the GWB monopole that we observe). This is impossible to do, since we have no way of knowing whether the single realisation of the monopole that we observe is near to the cosmic mean value, or whether it is out in the tails of the distribution (which could be the case if, e.g., we live near the centre of a cosmological under-/over-density of GW sources). We therefore neglect C_0 in what follows, focusing on the higher multipoles $\ell > 0$.

1.1.5 The gravitational-wave background in an inhomogeneous Universe

We have now discussed in some detail how to treat the anisotropic GWB intensity as a random field on the sphere, but so far we are lacking an anisotropic version of the Phinney formula (0.64) to predict what this field should look like for a given population of GW sources.

To write down such a formula, we consider three sources of anisotropies in the GWB. First, we treat the comoving number density of sources, n, as inhomogeneous, writing $n(t, \boldsymbol{x}) = \bar{n}(t)[1 + \delta_n(t, \boldsymbol{x})]$, where δ_n is the number density contrast, and $\bar{n}$ is the homogeneous mean number density. (Since the comoving rate density $\mathcal{R}$ is proportional to n, we therefore also have $\mathcal{R} = \bar{\mathcal{R}}(1 + \delta_n)$, where $\bar{\mathcal{R}}$ is the homogeneous mean rate density.) Second, we assume the emitted GWs propagate through a FLRW spacetime with scalar perturbations, whose line element in the Newtonian gauge is

$$\mathrm{d}s^2 = a^2(\eta)\big[-(1 + 2\psi(\eta, \boldsymbol{x}))\,\mathrm{d}\eta^2 + (1 - 2\phi(\eta, \boldsymbol{x}))\,\mathrm{d}\boldsymbol{x} \cdot \mathrm{d}\boldsymbol{x}\big], \tag{1.27}$$

where ψ and ϕ are the Bardeen potentials [97], and η is conformal time, $\mathrm{d}\eta \equiv \mathrm{d}t/a(t)$. Finally, we allow the sources and the observer to have peculiar velocities with respect to the cosmic rest frame, which follow some cosmological velocity field $\boldsymbol{v}(\eta, \boldsymbol{x})$. The GWB intensity is then given to first order in the perturbations by[6] [210, 224, 345]

$$\Omega(f, \hat{\boldsymbol{r}}) = \frac{2G}{3H_0^2} \int_0^{\eta_0} \mathrm{d}\eta\, a^2(\eta) \int \mathrm{d}\boldsymbol{\zeta}\, \bar{\mathcal{R}}(\eta, \boldsymbol{\zeta}) \frac{\mathrm{d}E}{\mathrm{d}(\ln f_{\mathrm{s}})}$$
$$\times \left[1 + \delta_n - 4\psi_0 + 5\psi + \hat{\boldsymbol{r}} \cdot (4\boldsymbol{v}_0 - 3\boldsymbol{v}) + 2\int_\eta^{\eta_0} \mathrm{d}\eta'\, \partial_\eta(\psi + \phi)\right], \tag{1.28}$$

where '0' subscripts indicate quantities evaluated at the observer's position $\boldsymbol{x}_0$, while all other perturbations are evaluated along the line of sight, $\boldsymbol{x}(\eta, \hat{\boldsymbol{r}}) = \boldsymbol{x}_0 + (\eta_0 - \eta)\hat{\boldsymbol{r}}$. We recall that $\boldsymbol{\zeta}$ denotes the set of parameters characterising each CBC signal (masses, spins, etc.), and that $\mathrm{d}E/\mathrm{d}(\ln f_{\mathrm{s}})$ is the dimensionless GW energy spectrum as a function of the source-frame frequency f_{s}, which in this context is related to the observed frequency f by

$$f_{\mathrm{s}} = \frac{f}{a}\left[1 + \psi_0 - \psi - \hat{\boldsymbol{r}} \cdot (\boldsymbol{v}_0 - \boldsymbol{v}) + \int_\eta^{\eta_0} \mathrm{d}\eta'\, \partial_\eta(\psi + \phi)\right]. \tag{1.29}$$

Equation (1.28) is analogous to the standard approach for calculating the scalar perturbations in the CMB temperature field, with the same three key contributions [247, 493]: a Sachs-Wolfe (SW) term due to gravitational redshifting of the photons/gravitons, expressed in terms of the values of the Bardeen potential ψ at the source and at the observer; a Doppler term written in terms of the peculiar velocities of the source and the observer projected along the line of sight; and an integrated Sachs-Wolfe (ISW) term encoding the time-evolution of both Bardeen potentials, ψ and ϕ. An important difference arises, however, in the density contrast term δ_n. While the CMB is emitted from a last-scattering surface which fills the entire sky, the GWB is in many cases (including the astrophysical background that is the focus of

---

[6]See Bertacca *et al.* [117] for a thorough derivation of a more complete version of this formula, as well as Bellomo *et al.* [111] for a numerical implementation of this.

this chapter) emitted from a discrete population of sources, whose number density fluctuations are far and away the most important contribution to the angular power spectrum. As we will see in section 1.2.2 below, the resulting anisotropies are significantly larger than those in the CMB from the SW, Doppler, and ISW effects. (Indeed, Bertacca *et al.* [117] found that these other terms contribute at most $\sim 10\%$ of the total angular power spectrum at a reference frequency of $f = 50$ Hz.) In the remainder of this chapter, we therefore neglect the metric perturbations ψ, ϕ, and consider an inhomogeneous population of discrete sources in a homogeneous FLRW background spacetime.

One other effect which is important to include, aside from the source density contrast, is the effect of the observer's peculiar velocity $\boldsymbol{v}_0$. Using equations (1.28) and (1.29), we see that this induces a *kinematic dipole* along the $\boldsymbol{v}_0$ direction,

$$\Omega(\hat{\boldsymbol{r}}) = \bar{\Omega} + \mathcal{D}\,\hat{\boldsymbol{v}}_0 \cdot \hat{\boldsymbol{r}} + \cdots, \qquad \mathcal{D} \equiv v_0 \left(4\bar{\Omega} - \frac{\partial \bar{\Omega}}{\partial (\ln f_s)} \right)\bigg|_{f_s = f/a}, \qquad (1.30)$$

or equivalently, in spherical harmonics,

$$\Omega_{10} = \sqrt{\frac{4\pi}{3}} \cos \theta_v \mathcal{D}, \qquad \Omega_{11} = -\sqrt{\frac{2\pi}{3}} \sin \theta_v \mathrm{e}^{-\mathrm{i}\phi_v} \mathcal{D}, \qquad C_1 = \frac{|\Omega_{10}|^2 + 2|\Omega_{11}|^2}{3} = \frac{4\pi}{9} \mathcal{D}^2, \tag{1.31}$$

where $\hat{\boldsymbol{v}}_0 = (\theta_v, \phi_v)$ and $v_0 \equiv |\boldsymbol{v}_0|$ are the direction and magnitude of the observer's peculiar velocity, with the latter being measured by Planck as $v_0 \approx 1.23 \times 10^{-3}$ for the Solar System barycentre [54]. We see that the coefficient $\mathcal{D}$ of the dipole term depends on the tilts of the GW energy spectra of the sources; the intuition here is that for a fixed observer-frame frequency f, the corresponding source-frame frequency f_s is lower for sources in the Doppler-boosted direction $\boldsymbol{v}_0$, meaning that sources with red-tilted spectra give rise to a larger coefficient, and vice versa. (We have been slightly lazy with our notation for this term—the derivative should in fact be taken inside of both integrals in equation (1.28), so that the spectral tilt is averaged over the source population.) For a population of inspiralling CBCs this tilt is simply $\sim f_s^{2/3}$, so we have $\mathcal{D} = (10/3)v_0\bar{\Omega}$. This provides an extremely useful consistency relationship between the GWB monopole and kinematic dipole; if the dipole were observed to deviate from this value, this could provide interesting evidence for some other population of sources contributing to the GWB.

Before we move on, it is important to point out that the statistics of the density contrast δ_n that determine the AGWB angular power spectrum are usually *non-Gaussian*, particularly for late-Universe sources such as CBCs. While the primordial perturbations that act as seeds for these inhomogeneities are extremely well-described by Gaussian statistics (with CMB observations and other probes providing strong bounds on primordial non-Gaussianity), the nonlinear dynamics of gravitational clustering are known to cause significant non-Gaussianity in the cosmic matter distribution at late times [115, 451]. This implies that our treatment in terms of the angular power spectrum does not capture the full statistical information present in the AGWB anisotropies, and that it may be interesting to consider higher-order statistics such as the bispectrum, trispectrum, etc. We leave this for future work.

1.2 MODELLING THE ASTROPHYSICAL BACKGROUND AND ITS ANISOTROPIES

The goal of this section is to compute the angular power spectrum of the AGWB, encapsulating the clustering statistics of CBC host galaxies. To accomplish this, we use an all-sky mock lightcone galaxy catalogue constructed from the Millennium *N*-body simulation. There are two key advantages of this approach, compared to other calculations based on CMB-inspired linear Boltzmann codes [221]: first, we are able to capture the full nonlinear gravitational dynamics of the cosmological matter distribution at late times; and second, we are able to extract detailed astrophysical information about the galaxies that form at the peaks of this distribution, which is necessary to calculate the GW emission from CBCs hosted in each galaxy.

Before we come to this calculation, it is first necessary to construct a model for the isotropic component of the AGWB, $\bar{\Omega}$, as this depends on the CBC rate density and population properties that will determine the contribution of each galaxy in the mock catalogue to our AGWB sky map.

1.2.1 THE ISOTROPIC COMPONENT

Averaging all cosmological perturbations to zero, we can rewrite equation (1.28) as

$$\bar{\Omega}(f) = \sum_i \frac{2G}{3H_0^2} \int_0^\infty \frac{\mathrm{d}z}{(1+z)^2 H(z)} \int \mathrm{d}\boldsymbol{\zeta}\, \bar{\mathcal{R}}_i(z, \boldsymbol{\zeta}) \frac{\mathrm{d}E_i}{\mathrm{d}(\ln f_\mathrm{s})}, \tag{1.32}$$

which we have written as a sum over the three key CBC populations: BBHs, BNSs, and BHNSs. (In principle we could include other astrophysical sources here too, e.g. supernovae. However, the contributions from these other sources are expected to be negligible compared to CBCs [470].) There are three ingredients we must specify in order to compute equation (1.32):

- the Hubble rate $H(z)$ of the FLRW background;

- the isotropic GW energy spectra $\mathrm{d}E_i/\mathrm{d}(\ln f_\mathrm{s})$ of the CBCs;

- the mean comoving rate density $\bar{\mathcal{R}}_i(z, \boldsymbol{\zeta})$ of each population, as a function of both the galaxy parameters and the CBC parameters.

The first ingredient is the easiest; we assume a standard ΛCDM Universe with Planck 2018 parameters,

$$H(z) \simeq H_0 \sqrt{\Omega_\Lambda + \Omega_m (1+z)^3}, \qquad H_0 \approx 67.7\,\mathrm{km\,s^{-1}\,Mpc^{-1}}, \qquad \Omega_\mathrm{m} \approx 0.311, \qquad \Omega_\Lambda = 1 - \Omega_\mathrm{m}. \tag{1.33}$$

For the second ingredient, we could use the $\sim f^{2/3}$ energy spectrum we computed in section 0.3.1 if we were interested purely in the inspiral regime, and were happy to neglect all post-Newtonian effects. While this might be an adequate approximation for BNSs in the LIGO/Virgo frequency band, many of the more massive BBHs in our population have their merger and ringdown in this frequency band, meaning that we must adopt a more complete description of the energy spectrum. To this end, we use the hybrid waveform models developed by Ajith *et al.* [58, 59] (as shown in figure 1.2), which combine
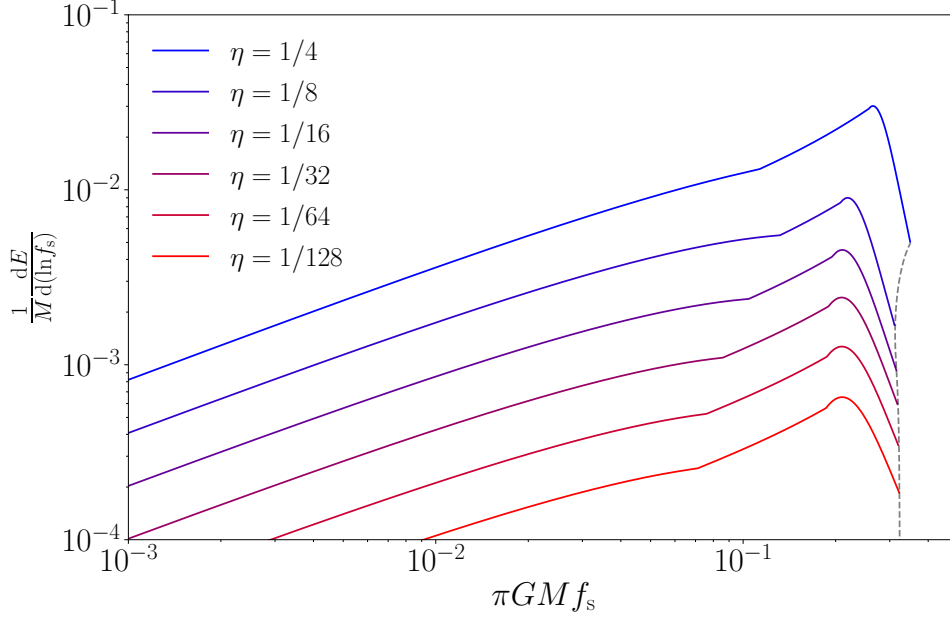
Figure 1.2: The hybrid energy spectrum (1.34) for non-spinning BBHs from Ajith *et al.* [58, 59], shown here as a fraction of the total mass M of the system, and as a function of the dimensionless frequency $\pi G M f_s = v_s^3$. Written in this form, the only remaining variable is the dimensionless mass ratio $\eta \equiv m_1 m_2 / (m_1 + m_2)^2 \leq 1/4$. The black dashed line shows the high-frequency cutoff at $\pi G M f_4$ as a function of η.

post-Newtonian results in the inspiral regime with data from numerical relativity simulations to fit a parameterised function covering the full inspiral-merger-ringdown spectrum. The hybrid BBH energy spectrum is written as a piecewise function,

$$
\frac{\mathrm{d}E}{\mathrm{d}(\ln f_s)} \approx \frac{1}{3} \eta M v_s^2 \times
\begin{cases}
\left[1 - \left(\frac{323}{224} - \frac{451}{168}\eta\right)v_s^2\right]^2, & f_s < f_1 \\
c_1 (f_s/f_1)\left(1 - 1.8897\, v_s + 1.6557\, v_s^2\right)^2, & f_1 \leq f_s < f_2 , \\
c_2 (f_s/f_2)^{7/3} \mathcal{L}^2(f_s|f_2, f_3), & f_2 \leq f_s < f_4
\end{cases}
\tag{1.34}
$$

with the three pieces corresponding to the inspiral, merger, and ringdown respectively. Here $v_s \equiv (\pi G M f_s)^{1/3}$ is the binary's relative velocity, c_1 and c_2 are numerical constants enforcing continuity of the spectrum, and $\mathcal{L}(x|x_0, \sigma) \equiv (\sigma/\pi)/[\sigma^2 + (x - x_0)^2]$ is the Lorentzian function. (Note that this reduces to the simple Newtonian result (0.90) in the limit $v_s \ll 1$, as expected.) The frequencies $\{f_1, f_2, f_3, f_4\}$ are given by

$$
\begin{aligned}
\pi G M f_1 &\approx 0.0660 + 0.6437\, \eta - 0.05822\, \eta^2 - 7.092\, \eta^3, \\
\pi G M f_2 &\approx 0.185 + 0.1469\, \eta - 0.0249\, \eta^2 + 2.325\, \eta^3, \\
\pi G M f_3 &\approx 0.0925 - 0.4098\, \eta + 1.829\, \eta^2 - 2.87\, \eta^3, \\
\pi G M f_4 &\approx 0.3236 - 0.1331\, \eta - 0.2714\, \eta^2 + 4.922\, \eta^3,
\end{aligned}
\tag{1.35}
$$

and correspond to the ISCO frequency, QNM frequency, inverse QNM damping time, and high-frequency cutoff, respectively. We assume here that the constituents of the binary are non-spinning; while
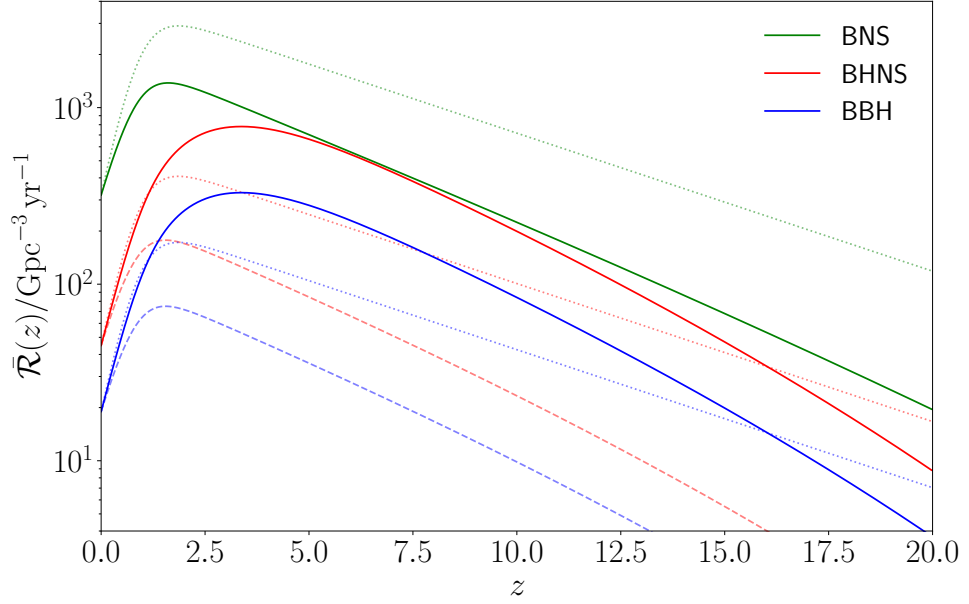
Figure 1.3: Comoving merger rate densities as functions of redshift for each of the three CBC populations we consider: BBHs, BHNSs, and BNSs. The dotted curves, given by equation (1.36), are directly proportional to the cosmic mean SFR. The dashed curves are proportional to the *delayed* SFR (1.38), with delay time distributions as described in the text. The solid curves additionally account for the suppression in BH formation from high-metallicity stellar progenitors using equation (1.39). (For BNSs, the latter two curves are identical.) In each case, the overall amplitude is fixed by the redshift-zero merger rates (1.48) inferred from the direct detections made by LIGO/Virgo; this has the slightly counter-intuitive effect of boosting the merger rate at higher redshifts when the metallicity suppression is taken into account, since this suppression is strongest at redshift zero.

there is evidence for some large BH spins amongst the CBCs detected by LIGO/Virgo [33], modelling the distribution of these spins greatly complicates our calculations by significantly increasing the dimensionality of the parameter space, and has negligible effect on the final GWB spectrum [36, 342], so we choose to ignore them here.

Equation (1.34) is not valid for BNSs and BHNSs, as in these cases the NS matter leaves imprints on the spectrum at frequencies $f_s \gtrsim f_1$. However, the lower masses of these systems mean that they typically merge outside of the LIGO/Virgo frequency band. We can therefore conservatively use equation (1.34) for BNSs and BHNSs too, but truncate the spectrum at the ISCO frequency f_1.

For the final ingredient, the comoving rate density, we need to capture the evolution in the number density of CBCs of different masses over cosmological timescales. Since all CBCs are products of stellar evolution, one simple choice is to tie the CBC rate density to the star formation rate (SFR) density,

$$\bar{\mathcal{R}}_i(z, \zeta) \propto \bar{\psi}(z) p_i(\zeta), \tag{1.36}$$

where $\bar{\psi}$ is the mean SFR density as a function of redshift, and $p_i(\zeta)$ is the probability density function (PDF) of the parameters of each population of CBCs (in particular, their masses), with i running over the different CBC populations as before. The unknown proportionality factor here encodes the efficiency with which newly-formed stars are converted into merging compact objects. This factor is challenging to model directly; but fortunately, we don't need to model it. Instead, we can fix the proportionality

using the local (i.e., $z = 0$) CBC rate density, $\bar{\mathcal{R}}_i^{(0)}$, inferred from the individual detections made by LIGO/Virgo,

$$\bar{\mathcal{R}}_i(z, \boldsymbol{\zeta}) = \bar{\mathcal{R}}_i^{(0)} p_i(\boldsymbol{\zeta}) \frac{\bar{\psi}(z)}{\bar{\psi}(0)}. \tag{1.37}$$

(Here we assume that the PDFs are normalised, $\int \mathrm{d}\boldsymbol{\zeta}\, p_i(\boldsymbol{\zeta}) = 1$.)

One major problem with this model is that it implicitly assumes that newly-formed stars are *immediately* converted into merging CBCs. In reality, the stars must first go through their main-sequence evolution before eventually becoming compact objects, and then must gradually inspiral from large initial separations until they reach the LIGO/Virgo band and merge. This entire process gives rise to a significant *delay time* between star formation and the associated CBC mergers, typically several Myr. We therefore convolve the SFR density at each redshift with some distribution of delay times,

$$\bar{\psi}_{\mathrm{d},i}(z) = \int \mathrm{d}t_{\mathrm{d}}\, p_i(t_{\mathrm{d}})\bar{\psi}(z_{\mathrm{f}}(z, t_{\mathrm{d}})), \tag{1.38}$$

where z_{f} is the redshift at which the stars were formed, which occurs at a lookback time t_{d} before the redshift z at which the binary merges. By replacing $\bar{\psi}$ with $\bar{\psi}_{\mathrm{d},i}$ in equation (1.36), we obtain an appropriately delayed estimate for the merger rate density.

So far we have assumed that the proportionality in equation (1.36) holds regardless of the properties of the binary and of the host galaxy. One exception to this which is important to capture here is that BH formation is suppressed for stellar progenitors with large metallicities. Following Abbott *et al.* [36], we model this metallicity dependence with a sharp cutoff on all BH formation for metallicities $Z > 0.1\,Z_\odot$, where $Z_\odot \approx 0.02$ is the solar metallicity. For BBHs and BHNSs, we therefore re-weight the rate density by the fraction of SFR occurring at metallicities below this threshold,

$$\bar{\psi}_{\mathrm{d},i} \to \bar{\psi}_{\mathrm{d},i}^{(Z)} \equiv \int \mathrm{d}t_{\mathrm{d}}\, p_i(t_{\mathrm{d}})f_i(z_{\mathrm{f}})\bar{\psi}(z_{\mathrm{f}}), \qquad f_i(z) \equiv \begin{cases} 1, & i = \text{BNS}, \\ \int_0^{0.1\,Z_\odot} \mathrm{d}Z\, p(Z|z), & i = \text{BBH, BHNS}. \end{cases} \tag{1.39}$$

This is our model for the CBC rate density. See figure 1.3 for a comparison between this and the more simplistic models described above.

We are now ready to flesh out each of the individual parts that make up this model; many of the details here agree with the 'fiducial' model adopted by the LVK in, e.g., Abbott *et al.* [36]. For the cosmic mean SFR density, we use the fitting function from Vangioni *et al.* [560],

$$\bar{\psi}(z) = \bar{\psi}_{\text{peak}} \frac{\alpha \exp\left[\beta(z - z_{\text{peak}})\right]}{\alpha - \beta + \beta \exp\left[\alpha(z - z_{\text{peak}})\right]}, \tag{1.40}$$

which has a maximum value of $\bar{\psi}_{\text{peak}} \approx 0.145\, m_\odot\, \mathrm{yr}^{-1}\,\mathrm{Mpc}^{-3}$ at redshift $z_{\text{peak}} \approx 1.86$, with redshift scaling either side of this peak set by the dimensionless constants $\alpha = 2.80$ and $\beta = 2.62$. This SFR density determines the rate at which stars produce metals and return these to the interstellar medium.
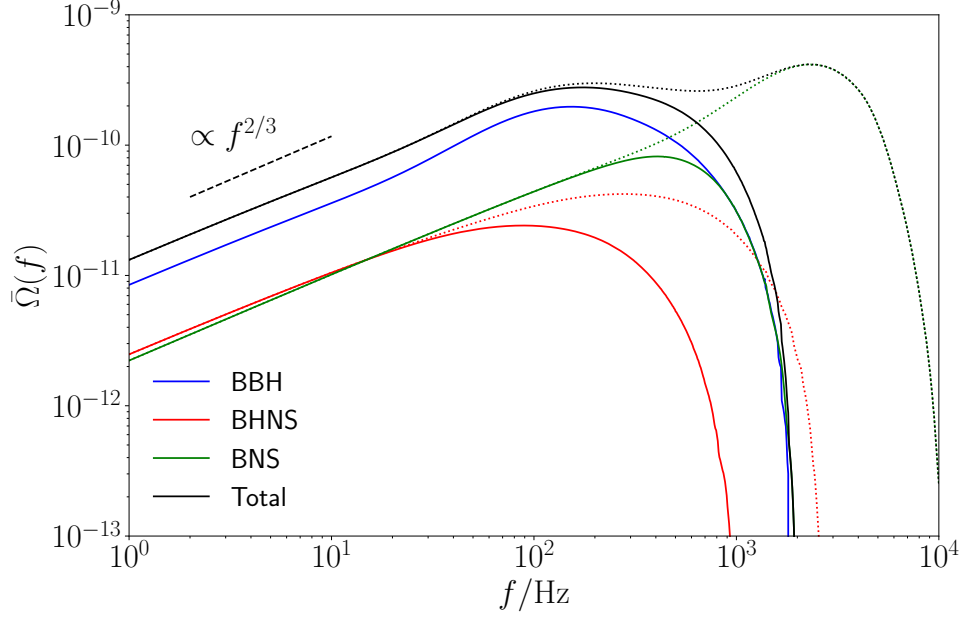
Figure 1.4: The mean GWB intensity $\bar{\Omega}(f)$ from stellar-mass CBCs in the LIGO/Virgo band, as predicted by our model. (Note that this has units of inverse steradians, and is therefore smaller than the isotropic GWB spectrum $\Omega(f)$ discussed in chapter 0 by a factor of 4π.) Solid curves show our default model, in which the GW emission from BNSs and BHNSs is truncated at the ISCO frequency, while dotted curves show the case where all binaries use the BBH waveform (1.34). The true GWB spectrum will likely lie somewhere between these two extremes; however, note that LIGO/Virgo stochastic searches are insensitive to this frequency range.

Equation (1.40) therefore sets the cosmic mean metallicity, using an expression which we take from Belczynski *et al.* [109],

$$\bar{Z}(z) = \frac{H_0}{\bar{\psi}_{\text{norm}}} \int_z^{20} \mathrm{d}z' \, \frac{\bar{\psi}(z')}{(1+z')H(z')}, \qquad \bar{\psi}_{\text{norm}} \approx 8.90 \, m_\odot \, \text{yr}^{-1} \, \text{Mpc}^{-3}. \qquad (1.41)$$

We assume that $\log_{10} Z$ follows a Gaussian distribution around this mean with standard deviation $\sigma = 1/2$. This distribution is cut off at $\log_{10} Z = 0$ (since $Z \le 1$ by definition), such that its PDF is

$$p(\log_{10} Z | z) = \sqrt{8/\pi} \, \frac{\exp\left[-2\left(\log_{10} Z - \log_{10} \bar{Z}(z)\right)^2\right]}{\text{erfc}\left[\sqrt{2} \log_{10} \bar{Z}(z)\right]}, \qquad (1.42)$$

and the corresponding metallicity correction factor is

$$f_i(z) = \begin{cases} 1, & i = \text{BNS}, \\[2ex] \dfrac{\text{erfc}\left[\sqrt{2}\left(\log_{10} \bar{Z}(z) - \log_{10} 0.1 Z_\odot\right)\right]}{\text{erfc}\left[\sqrt{2} \log_{10} \bar{Z}(z)\right]}, & i = \text{BBH, BHNS}, \end{cases} \qquad (1.43)$$

where $\text{erfc}(x) \equiv 1 - \int_0^x \mathrm{d}t \, e^{-t^2}$ is the complementary error function.

For the delay time distribution, we take $p_i(t_d) \propto 1/t_d$ between a minimum value t_{min} of 20 Myr for BNSs and 50 Myr for BBHs and BHNSs, and a maximum value equal to the age of the Universe at that redshift, which for our cosmological model is

$$t_{age}(z) = \frac{2}{3} \Omega_\Lambda^{-1/2} H_0^{-1} \sinh^{-1}\left(\sqrt{\frac{\Omega_\Lambda}{\Omega_m(1+z)^3}} \right).$$ (1.44)

The delayed, metallicity-corrected SFR density can thus be written as

$$\bar{\psi}_{d,i}^{(Z)} = \frac{1}{\ln\left(t_{age}(z)/t_{min,i} \right)} \int_{\ln t_{min,i}}^{\ln t_{age}(z)} d(\ln t_d)\, f_i(z_f(z,t_d))\, \bar{\psi}(z_f(z,t_d)),$$ (1.45)

where the formation redshift z_f is given by

$$\frac{1+z_f}{1+z} = \left[\cosh\left(\frac{3}{2} \Omega_\Lambda^{1/2} H_0 t_d \right) - \frac{H(z)}{\Omega_\Lambda^{1/2} H_0} \sinh\left(\frac{3}{2} \Omega_\Lambda^{1/2} H_0 t_d \right) \right]^{-2/3}.$$ (1.46)

Finally, we need to specify the local rates and mass distributions of each CBC population. For BBHs, following Abbott *et al.* [36], we assume the primary mass m_1 follows a broken power law distribution (as inferred from the second GW transient catalogue, GWTC-2),

$$p(m_1) \propto \begin{cases} m_1^{-1.58}, & 5.875\, m_\odot < m_1 < 40.82\, m_\odot, \\ m_1^{-5.59}, & 40.82\, m_\odot < m_1 < 87.14\, m_\odot, \end{cases}$$ (1.47)

with the secondary mass $m_2 \leq m_1$ given by a power-law distribution in the mass ratio $q \equiv m_2/m_1$, $p(q|m_1) \propto q^{1.4}$. For BNSs, we assume that both NS masses are uniformly distributed between $1\, m_\odot$ and $2.5\, m_\odot$, while for BHNSs we assume the NS mass follows this same distribution, while the BH mass follows the same distribution as the primary mass m_1 of the BBHs. We set the local BBH and BNS rates equal to those inferred for these models [36], while for BHNSs we use the recent value inferred in Abbott *et al.* [39],

$$\bar{\mathcal{R}}_{BBH}^{(0)} \approx 19\,\mathrm{Gpc}^{-3}\,\mathrm{yr}^{-1}, \qquad \bar{\mathcal{R}}_{BHNS}^{(0)} \approx 45\,\mathrm{Gpc}^{-3}\,\mathrm{yr}^{-1}, \qquad \bar{\mathcal{R}}_{BNS}^{(0)} \approx 320\,\mathrm{Gpc}^{-3}\,\mathrm{yr}^{-1}.$$ (1.48)

Putting all of these ingredients together, we find the AGWB spectrum shown in figure 1.4. We see that, as expected, the spectrum follows a $\sim f^{2/3}$ power law at low frequencies, before peaking and falling off rapidly above $\sim 1\,\mathrm{kHz}$ due to the post-merger truncation of the CBC energy spectrum. This fall-off occurs at higher frequencies for BBHs and BNSs than for BHNSs, since the BBH spectrum includes the merger-ringdown emission from equation (1.34), while the BNS spectrum is dominated by much lower total masses, and therefore much higher ISCO frequencies. The dotted curves show the hypothetical case in which BNSs and BHNSs have the same merger-ringdown emission as BBHs; we see that this only changes the overall spectrum at frequencies $f \gtrsim 1\,\mathrm{kHz}$, and is therefore beyond the reach of LIGO/Virgo stochastic searches. As such, we can safely neglect post-merger emission from these systems.
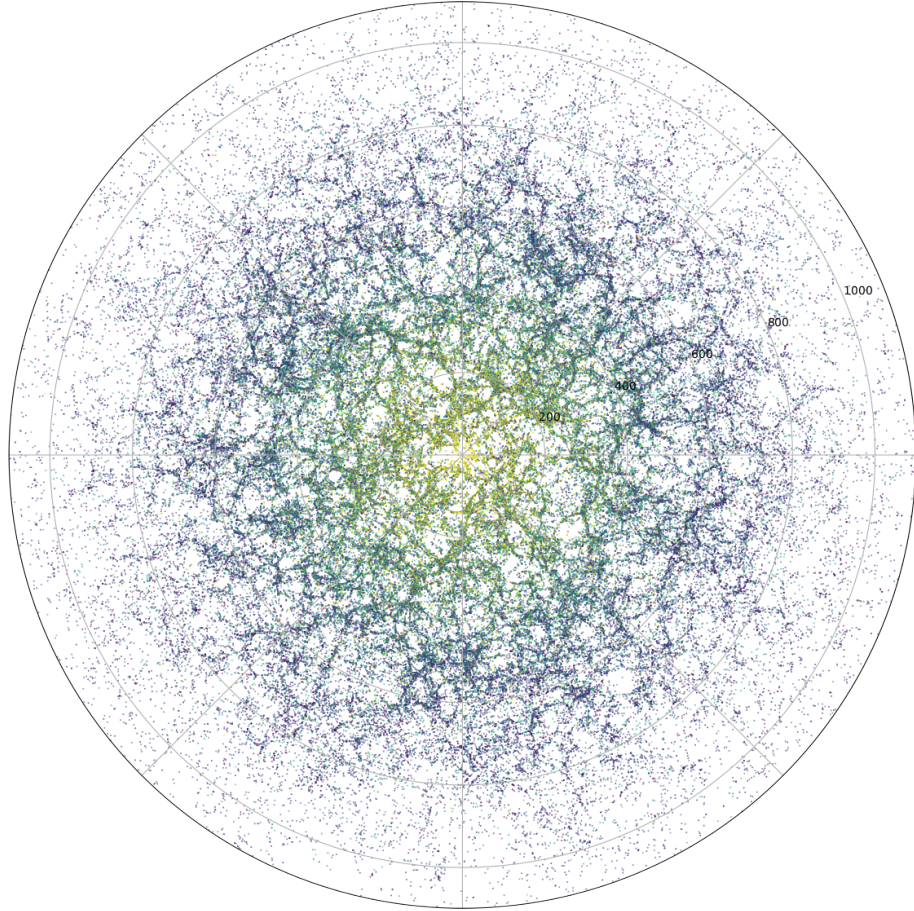
Figure 1.5: Equatorial slice through the three-dimensional distribution of galaxies in the Millennium mock galaxy catalogue. Each galaxy is coloured according to its contribution to the AGWB intensity, with nearby galaxies dominating due to the $\sim 1/r^2$ fall-off in the GW flux. The numbered radial rings indicate comoving distances in Mpc.

1.2.2 THE FULL AGWB SKY

We now construct a full AGWB sky map by applying the recipes described above to a mock light cone galaxy catalogue [128]. This catalogue consists of 5,715,694 galaxies, and was built using a random tiling technique applied to 64 post-processed snapshots saved during the Millennium N-body simulation [523], with a time step of roughly 100 Myr and a box size of 500 h^{-1} Mpc. These snapshots of the CDM distribution were analysed using the 'L-galaxies' semi-analytic model [233] to populate the simulation box with galaxies and capture the baryonic physics of each of these galaxies (star formation, metallicity, etc.), which is crucial for our purposes. The resulting large-scale galaxy distribution is illustrated in figure 1.5.

In order to compute the contribution of each galaxy to the AGWB, we query the catalogue database [390] to extract its sky location, redshift, star formation history, metallicity history, and peculiar velocity (the last of which has a small effect on the galaxy's total GW energy emission and source frame frequency through equations (1.28) and (1.29).) Computing the galaxy's *delayed* SFR, as well as the metallicity at the time of star formation, requires us to extract information that is not on the lightcone, so for each galaxy we query the full Millennium simulation to extract its SFR and metallicity at each of
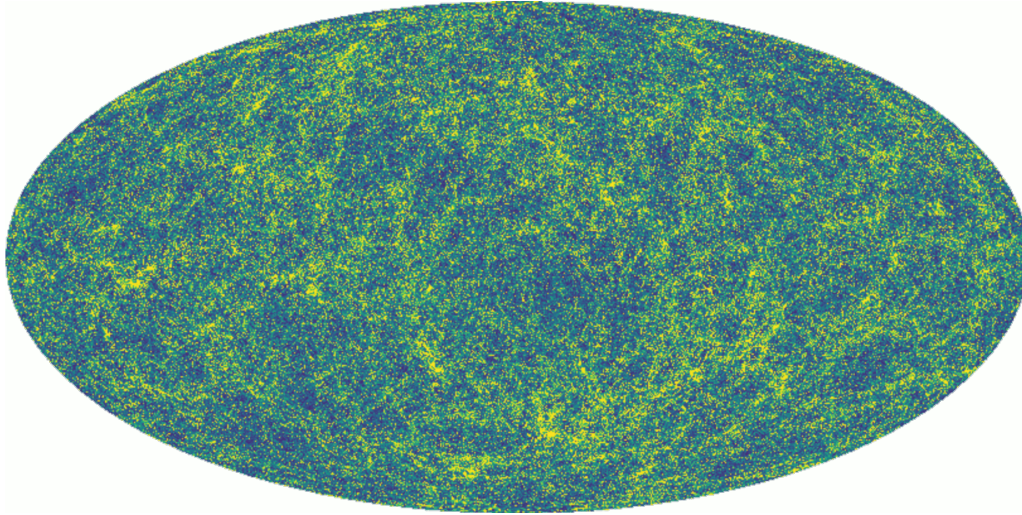
Figure 1.6: Mollweide projection of our full-sky mock HEALPix map of the AGWB, based on data from the Millennium simulation as described in the main text. Bright/dark pixels indicate over-/under-densities in the AGWB intensity field at a reference frequency of 25 Hz.

the previous snapshots. This process is made more complicated by the fact that the lightcone galaxies are the result of a sequence of mergers of smaller galaxies, each with its own independent star formation history. Tracing all of the branches of these merger trees for all 5,715,694 lightcone galaxies results in a set of 973,224,532 measurements of SFR, metallicity, and redshift, which we process to calculate the CBC merger rate of each galaxy on the lightcone.

This results in a list of sky locations and GW intensities for each of the galaxies, which we combine to produce the full-sky AGWB map shown in figure 1.6. This map uses the HEALPix[7] pixelisation scheme to partition the sky into equal-area curvilinear quadrilaterals, whose configuration is chosen to optimise the computation of spherical harmonics [289]. We estimate the angular power spectrum C_ℓ of this map using HEALPix's '`anafast`' routine; the result is shown in figure 1.7. Note that this spectrum corresponds to a single random realisation of LSS,[8] such that each of the C_ℓ multipoles shown here should be interpreted as random draws from a distribution centred on the true angular power spectrum, with width set by the cosmic variance we calculated in equation (1.23), $\sigma_\ell \equiv C_\ell / \sqrt{\ell + 1/2}$. We see in figure 1.7 that the anisotropies are very large compared to those in early-Universe observables such as the CMB, with fluctuations on the order of a few percent at the largest angular scales $\ell \sim 1$, rising to fluctuations of order unity at smaller angular scales $\ell \sim 100$. The resulting dipole component C_1 is significantly larger than the kinematic dipole due to our peculiar motion, which is denoted with a red star in figure 1.7. These large fluctuations reflect the nonlinear gravitational clustering of matter at late times, and retroactively justify our decision to focus on the density contrast term in equation (1.28), neglecting the much smaller contributions from the SW and ISW terms.

---

[7] http://healpix.sourceforge.net

[8] In principle, we could reduce the cosmic variance here by averaging the angular power spectra from multiple statistically-independent mock catalogues; at the time of writing, however, the catalogue we use is the only full-sky lightcone which is publicly available and is sufficiently large for our purposes. Furthermore, it is not possible to re-sample from this one catalogue without introducing correlations that would interfere with the goal of reducing cosmic variance.
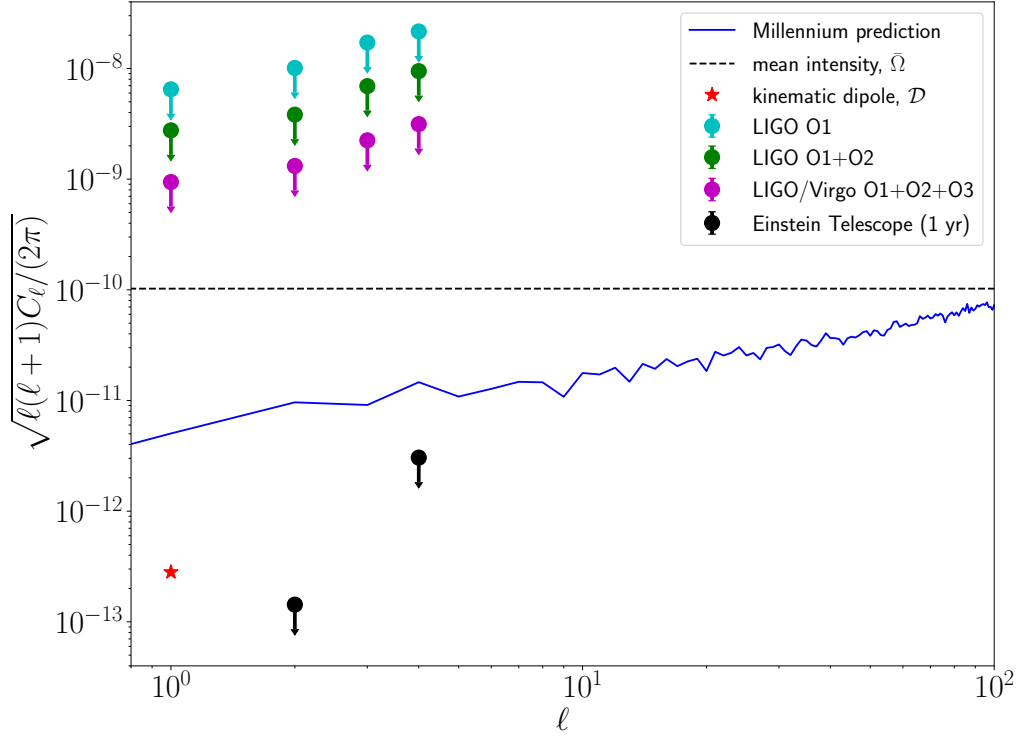
Figure 1.7: The AGWB angular power spectrum computed from the map shown in figure 1.6. The quantity plotted here, $\sqrt{\ell(\ell+1)C_\ell/(2\pi)}$, is roughly the contribution of each angular multipole ℓ to the standard deviation in the intensity field fluctuations. This has units of inverse steradians, and can be directly compared with the mean intensity $\bar{\Omega}$, which is shown as a horizontal black dashed line. The shaded blue region shows the 1σ uncertainty in the angular power spectrum due to cosmic variance. The red point shows the kinematic dipole, $\sqrt{\ell(\ell+1)C_\ell/(2\pi)} = (2/3)\mathcal{D} \simeq (20/9)v_0\bar{\Omega}$. The other coloured points with arrows show current upper limits (95 % confidence) on the AGWB angular power spectrum from LIGO/Virgo, as well as forecast upper limits from the Einstein Telescope (as computed in appendix A). All of the values shown here assume a reference frequency of 25 Hz.

The impact of this nonlinearity can also be seen in the one-point statistics of the AGWB map. In figure 1.8 we show a histogram of the GW intensity $\Omega(\hat{\boldsymbol{r}}_i)$ in each of the pixels $\hat{\boldsymbol{r}}_i$ in our map, normalised to the mean value $\bar{\Omega}$. The distribution is strongly skewed towards large, nonlinear overdensities, and is clearly in very poor agreement with a Gaussian distribution (as expected based on our discussion in section 1.1.5).

In figure 1.7 we also compare our predicted angular power spectrum with existing upper limits from LIGO/Virgo searches, as well as sensitivity forecasts for Einstein Telescope [464]. These upper limits focus on the lowest few multipoles (i.e., the largest angular scales), as these are the ones that the current detector network is most sensitive to. We see that, despite steady improvement from O1 onward, LIGO/Virgo are still several orders of magnitude away from detecting the AGWB anisotropies—which is unsurprising, given that the isotropic component has still not yet been detected, and this is louder by a factor of ~ 100. However, looking forward to the late 2030s, we see that ET is likely to make a confident detection of the AGWB anisotropies within just one year of observing time. This forecast assumes a cross-correlation search between the three co-located interferometers that make up the ET proposal (one at each corner
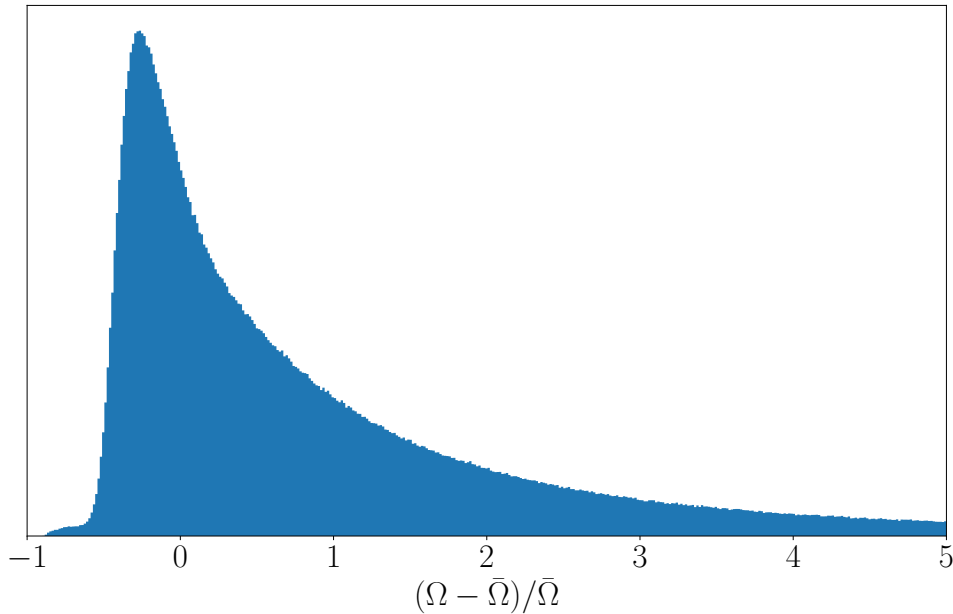
Figure 1.8: Histogram showing the distribution in GW intensity for each of the pixels in the map shown in figure 1.6, normalised to the mean intensity $\bar{\Omega}$. We see that the distribution is strongly non-Gaussian, with a heavy tail due to numerous nonlinear overdensities.

of an equilateral triangle). As we show in appendix A, the geometry of this setup is such that ET on its own is only sensitive to the $\ell = 2$ and $\ell = 4$ multipoles (as well as the monopole). However, adding further third-generation ground-based interferometers to the network, such as the US-based Cosmic Explorer proposal [295, 472], would break this degeneracy and provide sensitivity for a broader range of multipoles, as well as improving the sensitivity to the AGWB overall. It is important to note that these third-generation interferometers will be able to individually resolve the vast majority of BBH events in the Universe (though only a smaller fraction of the BNS and BHNS events) [407, 471, 491, 513]. As such, it is not clear how stochastic searches for the AGWB will change in the third-generation era, or indeed whether it will still be useful to think about this signal as being inherently stochastic, rather than attempting some kind of semi-coherent search along the lines of Smith and Thrane [518]. We leave a deeper investigation of these questions, and how they impact upon cosmological studies using CBCs as tracers of the cosmic matter distribution, for future work.

1.2.3 ESTIMATING THE IMPACT OF POPULATION UNCERTAINTIES

There are clearly many modelling assumptions which go into our model for the AGWB and its angular power spectrum, of which perhaps the least well-determined at present are the rates and mass distributions we assume for the CBC populations. In order to investigate the impact of these assumptions on our results, we can re-calculate the mean AGWB intensity and the angular power spectrum for a range of alternative rates and mass distributions, and see how much variation this causes in the final results.

In figures 1.9 and 1.10 we perform a scan over $\sim 10^4$ different models for the BBH distribution, computing the isotropic component and angular power spectrum for each model. Each set of model parameters
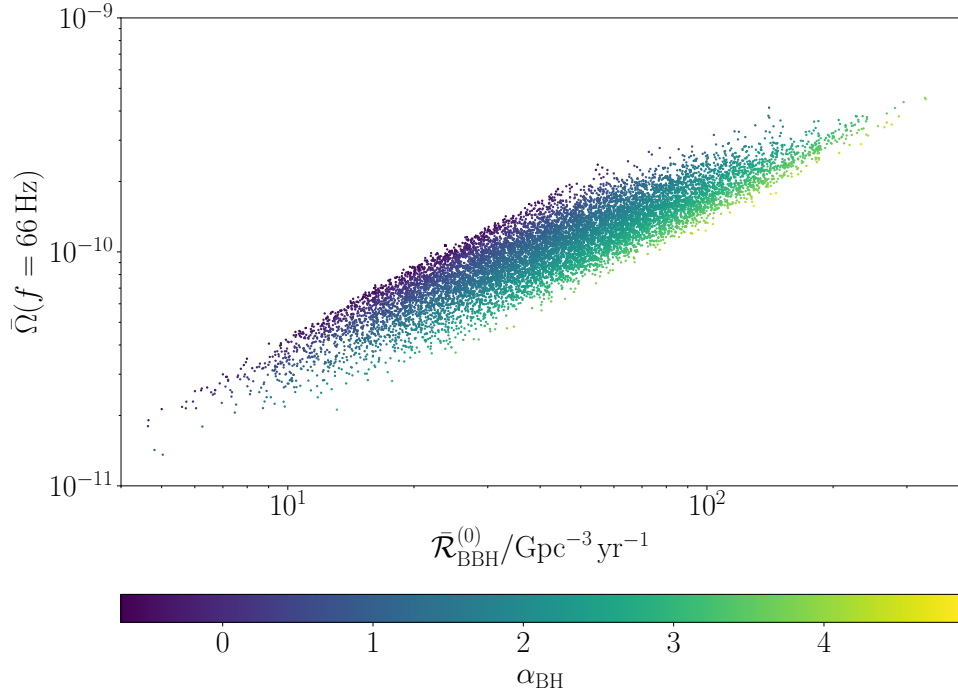
Figure 1.9: Scatter plot of the mean AGWB intensity (at a reference frequency of 66 Hz) and local BBH rate density $\bar{\mathcal{R}}_{\mathrm{BBH}}^{(0)}$ in $\sim 10^4$ samples from the distribution of possible BBH populations inferred from the first GW transient catalogue. Each sample is coloured according to its value of the BH mass power law index α_{BH}, with larger values corresponding to steeper fall-offs at large BH masses.

is drawn from a posterior hyperparameter distribution inferred from the first five confident BBH detections reported by LIGO/Virgo (GW150914 [5], GW151226 [7], GW170104 [10], GW170608 [16], and GW170814 [11]),[9] using the methodology developed by Wysocki *et al.* [590]. For simplicity we assume a straightforward power-law distribution for the primary BH mass, $p(m_1) \propto m_1^{-\alpha_{\mathrm{BH}}}$, between a minimum mass of 5 $m_\odot$ and a maximum mass m_{max} that is allowed to vary. The secondary BH mass is assumed to be uniformly distributed between 5 $m_\odot$ and m_1. The parameters α_{BH} and m_{max}, along with the local BBH rate density $\bar{\mathcal{R}}_{\mathrm{BBH}}^{(0)}$, therefore fully specify the BBH distribution needed for our calculations, with all other details (star formation rate, metallicity threshold, etc.) held fixed.

In figure 1.9, we see that this variation in the BBH population causes the mean AGWB intensity $\bar{\Omega}$ to vary by roughly an order of magnitude. While there is a clear (and intuitively sensible) trend of $\bar{\Omega}$ increasing in line with the local rate density $\bar{\mathcal{R}}_{\mathrm{BBH}}^{(0)}$, there is still a significant amount of scatter around this trend. Much of this scatter is captured by changes in the BH mass power law index α_{BH}, with shallower indices giving rise to stronger AGWB spectra. Again, this makes intuitive sense, as decreasing α_{BH} while holding the local rate density fixed effectively increases the relative importance of more massive BBHs, and these provide a much larger contribution in terms of GW energy radiated. This interesting structure highlights the fact that, once detected, the isotropic component of the AGWB could provide interesting and useful astrophysical information about the CBC populations from which it is emitted.

---

[9]This was the most complete study possible in Autumn 2018, when this work was first performed. Many more BBHs have been detected since that time [22, 33, 40].
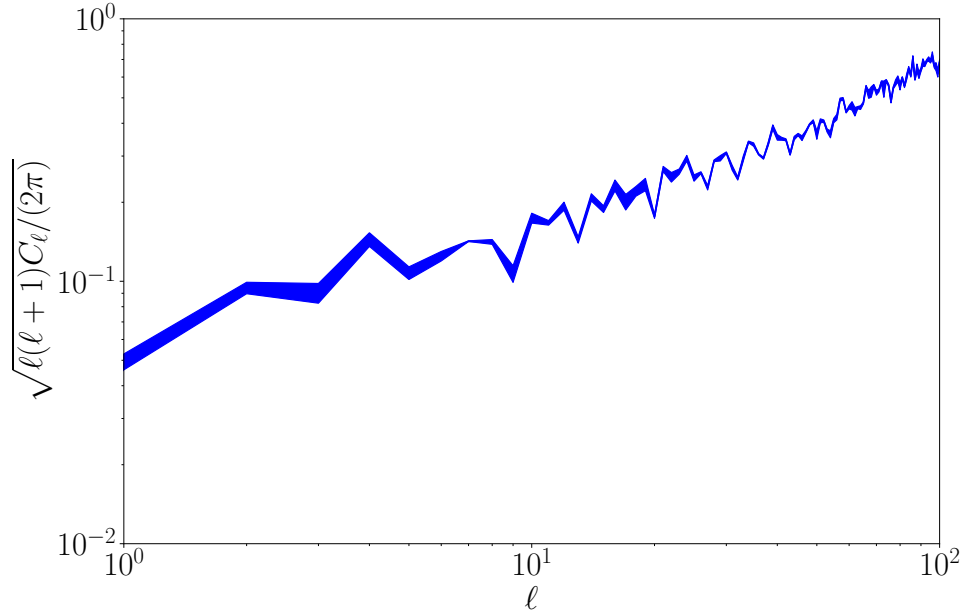
Figure 1.10: The AGWB angular power spectrum, normalised to the mean intensity $\bar{\Omega}$. While the light shaded region shows 1σ cosmic variance as before, the dark shaded region now shows the 3σ uncertainty region corresponding to the variation in the BBH population described in the main text. We see that this population uncertainty is always much smaller than cosmic variance (at least for the range of BBH population models considered here).

In contrast with this large variation encoded in the isotropic component, we see in figure 1.10 that, provided we normalise with respect to the isotropic amplitude, the AGWB angular power spectrum varies surprisingly little over the range of BBH population models we consider. This shows us that slightly re-weighting different galaxies in terms of their BBH populations does little to change the overall clustering statistics of the signal. As a result, our prediction for the angular power spectrum appears reasonably robust to population uncertainties, once one factors out the variation in the overall isotropic amplitude.

In the time since this work was first performed, the number of detected BBH events has grown significantly, and the GW astronomy community has begun to use ever more sophisticated models to capture their mass spectrum [33]. As such, it would be interesting to repeat this exercise again in future work, to see how our evolving knowledge of BBHs (and other CBCs) affects our expectations for the AGWB and its anisotropies. It would also be interesting to explore the impact of varying other model assumptions, particularly in ways which could have more influence on the relative contributions of different galaxy populations with different clustering properties. For example, Adhikari *et al.* [49] showed that if CBCs have longer delay times, then the GW events we observe today are concentrated in galaxies with larger stellar mass (since these dominate the star formation at higher redshifts), which live in more massive dark matter haloes and are therefore more tightly clustered.

1.3 Shot noise in the astrophysical background

In sections 1.1 and 1.2, we characterised the anisotropies in the AGWB in terms of their angular power spectrum components C_ℓ. These are statistical quantities, describing the expected angular correlation of the AGWB after averaging the signal in at least three distinct ways:

1. averaging over random realisations of LSS (i.e., over an 'ensemble of Universes');

2. averaging over the discrete positions of galaxies within the matter distribution to give a continuous galaxy number density field;

3. averaging over the merger times of CBCs within each galaxy to give a mean merger rate.

However, the AGWB we observe corresponds to a single Universe, with some discrete number of galaxies, each with some discrete number of CBCs. In practice, therefore, we are unable to perform the averaging process described above, leading to random fluctuations in the observed C_ℓ's, which must be accounted for when comparing to theoretical predictions. What's more, our theoretical predictions in the previous section are themselves based on a simulated galaxy catalogue with a single realisation of LSS and a finite number of galaxies, so it is doubly important to understand these effects.

We have already accounted for the first point above: the uncertainty due to our observation of a single realisation of LSS, i.e., cosmic variance. In section 1.1.4 we saw that, for a Gaussian AGWB, this uncertainty is given by $\mathrm{Var}[C_\ell] = C_\ell^2 / (\ell + {}^1\!/{}_2)$. Besides cosmic variance, we must account for the fact that the AGWB is emitted from a finite number of galaxies (sampled from the underlying density field), each hosting a finite number of CBCs (sampled from the mean merger rate). These two sampling processes, corresponding to the second and third points in the list above, follow Poissonian statistics and introduce *shot noise* to the observed angular power spectrum. This is a very important effect, which has been studied for decades in contexts such as galaxy redshift surveys [268, 296] and the cosmic infrared background [361]. In the context of the AGWB, Meacher *et al.* [422] used numerical simulations to study the effects of shot noise on the isotropic component, but until now the effects of shot noise on the AGWB anisotropies have been ignored.

In this section, we derive expressions for the AGWB angular power spectrum in the presence of shot noise, and calculate the size of these shot-noise effects for our model of the AGWB in the LIGO/Virgo frequency band. We then develop an optimal data-analysis method for inferring the *true*, cosmological angular power spectrum in the presence of shot noise.

1.3.1 The angular power spectrum, with and without shot noise

We now show how the inclusion of generic shot-noise effects in the underlying statistics of the AGWB leads to an additional term in the angular power spectrum. For later convenience, we write the intensity field as an integral over comoving distance,

$$\Omega(f, \hat{\boldsymbol{r}}) = \int \mathrm{d}r \, r^2 \omega(f, \boldsymbol{r}),$$ (1.49)

where ω is an 'AGWB density' with dimensions $[\text{length}]^{-3}$. Comparing with equation (1.28), we see that

$$\omega(f, \boldsymbol{r}) = \frac{2G}{3(1+z)^2} \left(\frac{r_H}{r}\right)^2 \int d\boldsymbol{\zeta}\, \mathcal{R}(\boldsymbol{r}, \boldsymbol{\zeta}) \frac{dE}{d(\ln f_s)}, \tag{1.50}$$

where $r_H \equiv 1/H_0$ is the Hubble radius. Note that the rate density $\mathcal{R}$ is now a random field in three dimensions, with mean equal to the isotropic but redshift-dependent value $\bar{\mathcal{R}}$ that we discussed in the previous section. The SHCs are given in terms of the density ω by

$$\Omega_{\ell m} \equiv \int_{S^2} d^2\hat{\boldsymbol{r}}\, Y^*_{\ell m}(\hat{\boldsymbol{r}}) \Omega(\hat{\boldsymbol{r}}) = \int d^3\boldsymbol{r}\, Y^*_{\ell m}(\hat{\boldsymbol{r}}) \omega(\boldsymbol{r}). \tag{1.51}$$

We calculate the angular power spectrum by specifying the second moment of the ω field. Neglecting shot noise, this is simply

$$\text{Cov}[\omega(\boldsymbol{r}), \omega(\boldsymbol{r}')]_{\text{LSS}} = \bar{\omega}(r)\bar{\omega}(r')\xi(r, r', \theta), \tag{1.52}$$

where ξ is the two-point correlation function of ω, describing the probability in excess of random of similar values of the field being clustered together. Due to statistical isotropy, ξ only depends on the angular positions $\hat{\boldsymbol{r}}, \hat{\boldsymbol{r}}'$ of the two points through their separation $\theta \equiv \cos^{-1}(\hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}')$. (It does, however, depend on their radial distances, as these influence the intensity of the GW flux.) Using equations (1.49), (1.13), and (1.52), we find the C_ℓ's in the absence of shot noise,

$$C_\ell^{\text{LSS}} = 2\pi \int_{-1}^{+1} d\cos\theta\, P_\ell(\cos\theta) \int dr\, r^2 \int dr'\, r'^2 \bar{\omega}(r)\bar{\omega}(r')\xi(r, r', \theta). \tag{1.53}$$

We now introduce a shot-noise term that encompasses both the galaxy sampling and the CBC rate sampling. Assuming these effects can jointly be treated as a local Poisson process that is independent of LSS, we have

$$\text{Cov}[\omega(\boldsymbol{r}), \omega(\boldsymbol{r}')] = \bar{\omega}(r)\bar{\omega}(r')\xi(r, r', \theta) + \mathcal{V}(r)\delta^3(\boldsymbol{r} - \boldsymbol{r}'), \tag{1.54}$$

where $\mathcal{V}$ is some function describing the variance due to the finite sample, which is independent of direction due to statistical isotropy. The form of this new shot-noise term (in particular, the fact that it is proportional to $\delta^3(\boldsymbol{r} - \boldsymbol{r}')$), is motivated by the equivalent expression for galaxy surveys (see, e.g., appendix A of Feldman *et al.* [268], or section 2.4 of Hamilton [296]). It is quite simple to convince oneself that the modification due to shot noise should take this form: there should be an extra term added to the covariance, as shot-noise fluctuations increase the variations in the measured values of ω throughout space, above the intrinsic variance due to the clustering of GW sources. However, since the shot-noise fluctuations at one point in space are causally disconnected from those at any other point, the fluctuations at any two points are statistically independent, and the extra term in the covariance should vanish except when the two points are coincident, leading to the delta function. Equation (1.54) thus arises naturally from the superposition of independent Poisson processes at each point in space. We flesh this argument out more quantitatively in section 1.3.2 below.

Using equations (1.49), (1.13), (1.53), and (1.54), as well as the fact that $P_\ell(1) = 1$ for all ℓ, the full angular power spectrum is then

$$C_\ell = C_\ell^{\text{LSS}} + \mathcal{W}, \qquad \mathcal{W} \equiv \int \mathrm{d}r\, r^2 \mathcal{V}(r). \tag{1.55}$$

We therefore see that the shot noise generates a spectrally white (i.e., independent of ℓ) contribution to the C_ℓ's.

1.3.2 Calculating the shot-noise power

In order to evaluate equation (1.55), we must derive an expression for the local Poisson variance function $\mathcal{V}$, accounting for the random sampling of both the galaxy number density $n(\mathbf{r})$ and the CBC event rate per galaxy R, whose product gives the CBC rate density $\mathcal{R} = nR$.

Consider a volume element δV at position $\mathbf{r}$. We treat the number of galaxies in this region as a Poisson random variable,[10] $N \sim \text{Pois}[\delta V\, n(\mathbf{r})]$. Assuming for now that these galaxies and the CBCs in them all have the same parameter values $\boldsymbol{\zeta}$, the CBC event counts for each galaxy in a given source-frame time interval T_s are i.i.d. Poisson random variables, $\lambda_i \sim \text{Pois}[RT_s]$. The total CBC event count from δV is then $\Lambda = \sum_{i=1}^N \lambda_i$. This quantity—a sum over Poisson random variables, with a number of terms that is itself a Poisson random variable—follows a *compound* Poisson distribution. In appendix B, we show that the variance of this distribution is

$$\text{Var}[\Lambda] = \delta V\, \bar{n}\big[RT_s + (RT_s)^2\big]. \tag{1.56}$$

Taking the sampling as independent for different spatial volumes and for different parameter values, we therefore find

$$\text{Cov}[\Lambda(\mathbf{r}_i, \boldsymbol{\zeta}), \Lambda(\mathbf{r}_j, \boldsymbol{\zeta}')] = \delta V\, \bar{n}\big[RT_s + (RT_s)^2\big]\delta_{ij}\delta(\boldsymbol{\zeta}, \boldsymbol{\zeta}'), \tag{1.57}$$

where we have introduced a Kronecker delta over the volume cells and a Dirac delta over parameter space to enforce this independence, and have averaged over LSS so that $n(\mathbf{r}) \to \bar{n}(\mathbf{r})$. Replacing Λ with the comoving rate density, $\mathcal{R} = nR \simeq \Lambda/(\delta V T_s)$, and taking $\delta V \to 0$, we are left with

$$\text{Cov}[\mathcal{R}, \mathcal{R}']_{\text{shot}} = \left(\frac{\bar{\mathcal{R}}}{T_s} + \frac{\bar{\mathcal{R}}^2}{\bar{n}}\right)\delta^{(3)}(\mathbf{r} - \mathbf{r}')\delta(\boldsymbol{\zeta}, \boldsymbol{\zeta}'), \tag{1.58}$$

where $\mathcal{R}'$ is shorthand for $\mathcal{R}(\mathbf{r}', \boldsymbol{\zeta}')$, and where we have set $\delta_{ij}/\delta V \to \delta^{(3)}(\mathbf{r} - \mathbf{r}')$. The relationship between the source-frame time T_s and the observer-frame time T will generally depend on cosmological metric perturbations and peculiar velocities at the location of the source (c.f. equation (1.29) for the source-frame GW frequency), but to leading order it is simply $T_s = T/(1 + z)$.

---

[10] We stress that this is only an approximation. A more sophisticated approach would use the halo model of LSS [213], accounting for the statistical properties of dark matter haloes and of different populations of galaxies within them. However, this approximation is sufficiently accurate for our purposes (particularly as the galaxy-number contribution to the shot noise is much less than the CBC rate contribution—see below).
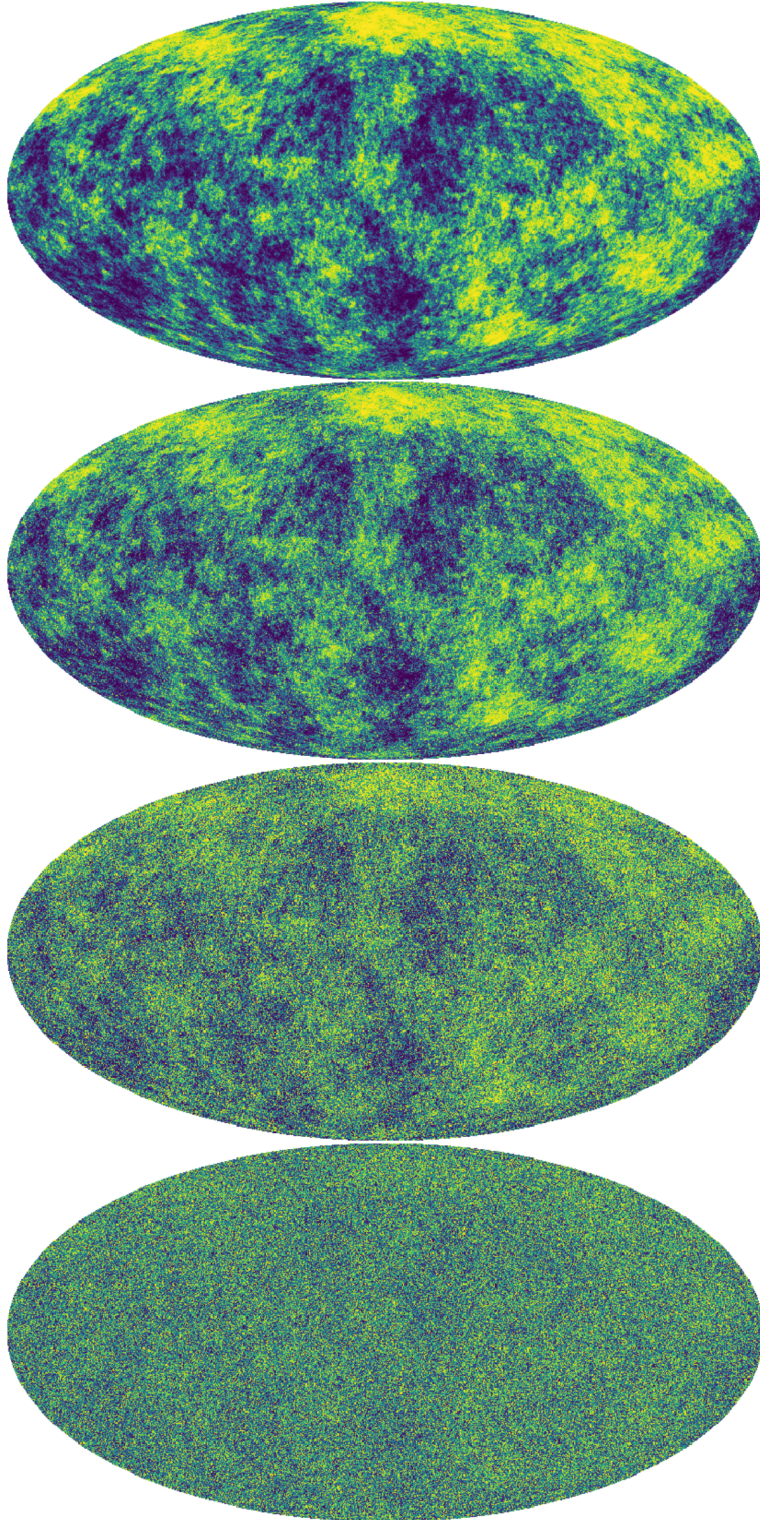
Figure 1.11: A toy-model depiction of shot noise. All four images are HEALPix maps with scale-invariant angular power spectra $\ell(\ell+1)C_\ell = $ constant, plus varying degrees of shot noise power. From top to bottom, the shot noise power is equal to $\mathcal{W} = 0$, $\mathcal{W} = 10^{-5}\bar{\Omega}^2$, $\mathcal{W} = 10^{-4}\bar{\Omega}^2$, and $\mathcal{W} = 10^{-3}\bar{\Omega}^2$. Physically, these represent different observations of the AGWB, with different observation time intervals, leading to different levels of shot noise power. All four maps have the same underlying random realisation of LSS, which is why the same large-scale features can be recognised in each of them. However, increasing the amount of shot noise leads to much stronger anisotropies on small scales, making it harder to discern the relatively subtle large-scale features.

Using equations (1.50) and (1.54), the shot-noise power is therefore

$$\mathcal{W} = \frac{4G^2}{9H_0^2} \int \frac{\mathrm{d}r}{(1+z)^4} \left(\frac{r_H}{r}\right)^2 \int \mathrm{d}\zeta \left(\frac{\bar{\mathcal{R}}}{T_s} + \frac{\bar{\mathcal{R}}^2}{\bar{n}}\right) \left[\frac{\mathrm{d}E}{\mathrm{d}(\ln f_s)}\right]^2 \tag{1.59}$$

Equation (1.59) is our main result, and has two distinct applications: (*i*) interpreting future observations of the AGWB, and (*ii*) improving our theoretical models. For the first application, we can use equation (1.59) to calculate the expected shot noise in the observed C_ℓ spectrum as a function of observing time, given a model for the galaxy number density and CBC merger rate. The $\bar{\mathcal{R}}/T_s$ term dominates over the $\bar{\mathcal{R}}^2/\bar{n}$ term, since the latter is suppressed by a factor of $\bar{\mathcal{R}}T_s/\bar{n} = RT_s$, which is $\sim 10^{-6}$ for a typical galaxy (i.e., the CBC rate per galaxy is on the order of $\sim 1\,\mathrm{Myr}^{-1}$). For the second application, we can use equation (1.59) to calculate the error inherent to our theoretically-predicted spectrum from the previous section due to the finite galaxy sampling in the simulated galaxy catalogue. Since these predictions do not involve simulating the time of arrival of discrete GW signals, but average over the CBC rate, they exclude the shot noise due to sampling of this rate. This is identical to taking the limit $T \to \infty$, meaning that the catalogue predictions contain shot noise due to the $\bar{\mathcal{R}}^2/\bar{n}$ term only. We therefore distinguish between the 'observational shot noise' and the 'catalogue shot noise',

$$\begin{aligned}
\mathcal{W}_{\mathrm{obs}} &= \frac{4G^2}{9H_0^2 T} \int \frac{\mathrm{d}r}{(1+z)^3} \left(\frac{r_H}{r}\right)^2 \int \mathrm{d}\zeta\, \bar{\mathcal{R}} \left[\frac{\mathrm{d}E}{\mathrm{d}(\ln f_s)}\right]^2, \\
\mathcal{W}_{\mathrm{cat}} &= \frac{4G^2}{9H_0^2} \int \frac{\mathrm{d}r}{(1+z)^4} \left(\frac{r_H}{r}\right)^2 \int \mathrm{d}\zeta\, \frac{\bar{\mathcal{R}}^2}{\bar{n}_{\mathrm{cat}}} \left[\frac{\mathrm{d}E}{\mathrm{d}(\ln f_s)}\right]^2.
\end{aligned} \tag{1.60}$$

Here, $\bar{n}_{\mathrm{cat}}$ represents the galaxy number density in the catalogue, which is significantly less than the true galaxy number density $\bar{n}$ (this is because only galaxies brighter than a certain magnitude are included).

1.3.3 Removing the foreground

Inspecting equation (1.59), we see that the integrand diverges as $r \to 0$. This is to be expected, for two reasons. First, the Poisson statistics become progressively worse at small distances, as we are looking at smaller spherical shells that contain fewer galaxies, so the notion of a smooth number density $\bar{n}$ breaks down as $r \to 0$. Second, the contribution of a single CBC to the total AGWB flux becomes much larger at smaller distances, so a CBC that is arbitrarily close can bias the power spectrum by an arbitrarily large amount.

In order to regulate this divergence, it is necessary to introduce a cutoff distance r_*, below which we remove any CBC signals and do not consider them part of the AGWB. (This is similar to what is done with, e.g., the cosmic infrared background [361].) We are free to choose the value of r_*, with larger values helping to reduce the shot noise as much as possible. However, the choice of r_* will also affect C_ℓ^{LSS}, and this must be accounted for when making theoretical predictions.

In principle, one can implement this foreground cut by removing from the GW strain time series $h(t)$ any intervals in which an individual CBC with a comoving distance $r < r_*$ is identified. The CBC chirp signal encodes the luminosity distance d_L, which can easily be converted to the comoving distance r by assuming a fiducial cosmology. However, d_L cannot be measured with arbitrary precision, particularly
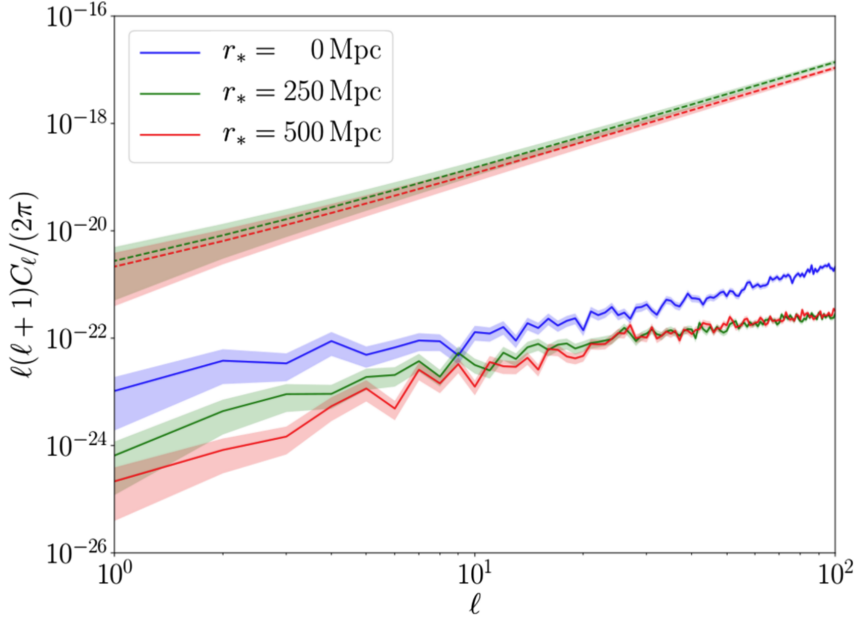
Figure 1.12: The AGWB angular power spectrum, plotted for different values of the cutoff distance r_*. The solid curves are calculated using a mock galaxy catalogue from the Millennium Simulation (with the catalogue error $\mathcal{W}_{cat}$ removed, though this has negligible effect), and represent the cosmological power spectrum C_ℓ^{LSS}. The dashed curves include the observational shot noise after $T = 1\,\mathrm{yr}$, giving the full power spectrum $C_\ell = C_\ell^{LSS} + \mathcal{W}_{obs}$. The shaded regions indicate cosmic variance. $\mathcal{W}_{obs}$ is divergent in the case $r_* = 0$, and so is not plotted.

as it is degenerate with the inclination of the binary. What's more, the detectability of CBCs at a given distance is also a function of sky position, due to the anisotropic beam pattern of the detectors. Removing nearby sources based on a signal-to-noise threshold (as in Meacher *et al.* [422]) thus risks biasing the power spectrum by favouring particular kinds of CBC events and particular regions of the sky. The alternative is to set r_* small enough that *all* CBCs at distances $r < r_*$ are detectable, so the foreground removal can be implemented in an unbiased and isotropic manner, without selection effects. This would mean including some number of CBCs at $r > r_*$ in the stochastic analysis despite the fact that they can be individually resolved. This proposal presents a serious data analysis challenge, and it is unclear whether it will be possible in practice to remove the foreground in a way which does not introduce directional biases. Looking forward to third generation interferometers such as Einstein Telescope, we expect that essentially all BBHs in the observable Universe will be individually resolvable [407, 471, 491, 513], meaning that our framing here in terms of a resolvable foreground and a stochastic target signal will likely have to be modified. Indeed, it will likely become preferable to study anisotropies in the distribution of resolved events, although some care will have to be taken in combining this with the large number of BNS and BHNS events that will remain unresolvable. We leave a detailed examination of these issues for future work.

1.3.4 Results and discussion

Using the AGWB model described above, we can calculate the size of the observational and catalogue shot-noise terms from equation (1.60) for different values of the foreground cutoff distance r_*. We find that the catalogue shot noise is at most $\mathcal{W}_{\mathrm{cat}} \approx 3 \times 10^{-29}$, and is typically several orders of magnitude smaller than this for larger values of r_*. Since the monopole is $\bar{\Omega} \approx 10^{-10}$, this represents shot-noise fluctuations of $\lesssim 0.01\%$. This is negligible compared to the true power spectrum C_ℓ^{LSS}, so our earlier predictions based on the Millennium simulation are completely unaffected.

On the other hand, we find that the observational shot noise $\mathcal{W}_{\mathrm{obs}}$ is generically several orders of magnitude larger than the true angular power spectrum C_ℓ^{LSS} for any reasonable values of r_* and T. (In fact, $\sqrt{\mathcal{W}_{\mathrm{obs}}} \approx \bar{\Omega}$, so the shot-noise fluctuations are typically as large as the monopole itself.) This is illustrated in figure 1.12 for $r_* = 250\,\mathrm{Mpc}$ and $r_* = 500\,\mathrm{Mpc}$. Note that C_ℓ^{LSS} also changes with r_*, partly because the total GW emission is reduced, causing an overall reduction in the spectrum, and partly because the nearest galaxies contribute most strongly to the AGWB, so that their removal changes the shape of the spectrum. Calculations of C_ℓ^{LSS} using the catalogue are not reliable for values of r_* significantly larger than $\sim 500\,\mathrm{Mpc}$, due to the incompleteness of the catalogue at high redshift. However, our calculations indicate that even increasing r_* to $2\,\mathrm{Gpc}$ can only reduce $\mathcal{W}_{\mathrm{obs}}$ by less than an order of magnitude, so this seems unlikely to solve the problem.

The numerical values given in figure 1.12 depend on the details of the AGWB model, and include a multitude of random and systematic uncertainties in, e.g., the populations of astrophysical sources that contribute, their emission rates, and the nature of their clustering. However, we stress that the main result, equation (1.55), is generic, and is grounded in simple and realistic physical principles. Any finite population of sources will have random Poissonian fluctuations. If these fluctuations are statistically independent at different spatial locations (i.e., if the shot-noise fluctuations are causally disconnected), then the angular power spectrum generically gains an extra white-noise component, $\mathcal{W}$. If these sources are finite in time, then basic Poisson statistics dictates that this noise decays as the inverse of the observation time, $\mathcal{W} \propto 1/T$.

1.3.5 Temporal shot noise and hierarchical averaging

Now that we have diagnosed the issue of shot noise, we turn our attention to possible ways in which it can be mitigated when carrying out searches for AGWB anisotropies. In order to do this, it is helpful to first think more carefully about the statistics of the observed AGWB intensity under shot noise fluctuations, and how these statistics are distinct from those associated with LSS.

We have shown above that, for the AGWB, the *temporal* shot noise associated with finite CBCs per observation time is significantly larger than the true C_ℓ spectrum, while the *spatial* shot noise associated with a finite galaxy number density is significantly smaller; we therefore focus exclusively on the former in this analysis. This allows us to exploit the fact that we have observational access to multiple realisations of the temporal shot noise: for each successive observation interval τ, we can obtain a set of SHCs with shot noise power $\mathcal{W}_\tau \propto 1/\tau$. The shot-noise fluctuations in each set of SHCs are associated with a different set of CBCs in a different set of galaxies, so it is immediately clear that each successive shot noise realisation is statistically independent of the others. This is in contrast with the spatial shot noise, for which we can
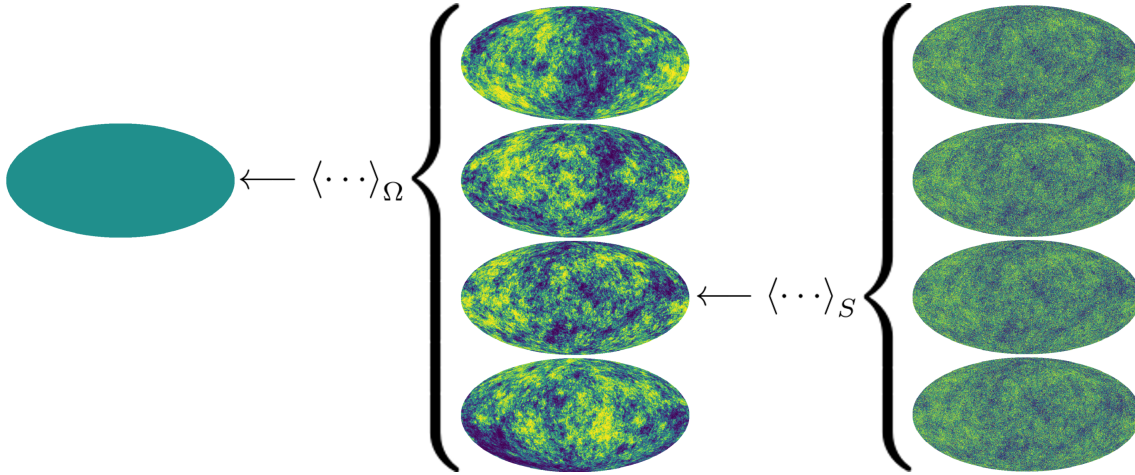
Figure 1.13: An illustration of the two averaging procedures introduced in Sec. 1.3.5. The right column shows four independent realisations of shot noise (at a level of $\mathcal{W} = 10^{-4}\bar{\Omega}^2$) for a single given realisation of LSS. Averaging over many such realisations corresponds to the $\langle\cdots\rangle_S$ operation. Similarly, the central column shows four independent realisations of LSS, with zero shot noise. Averaging over many such realisations corresponds to the $\langle\cdots\rangle_\Omega$ operation, and results in a perfectly uniform field (i.e. all SHCs equal to zero, except the monopole).

only observe a single realisation (in other words, the random positions of galaxies are a *persistent* random field, while the random sky locations of CBCs form a *transient* random field).

There are two logically distinct random processes that govern the observed SHCs: the distribution of matter on large scales, and the emission of a finite number of GW signals from this matter distribution in a given observation period. We model these processes, and their corresponding ensemble averages, in a hierarchical manner.

1. The true $\Omega_{\ell m}$ are drawn from Gaussian distributions with variance C_ℓ. This process is associated with the cosmological averaging operation from section 1.1,

$$\langle\cdots\rangle_\Omega \equiv \text{cosmological average,} \tag{1.61}$$

   which can be thought of as an average over an 'ensemble of Universes', with each Universe having a distinct random realisation of LSS. (Of course, we only have access to a single such realisation, and this gives rise to cosmic variance.)

2. The true $\Omega_{\ell m}$ are modulated by shot noise, so that a set of 'noisy' SHCs $\Omega_{\ell m}^i$ is drawn from a distribution with the true components $\Omega_{\ell m}$ as its mean. This draw is independent for each time interval, with different intervals being labelled by the index i. We write the associated average over shot noise realisations as

$$\langle\cdots\rangle_S \equiv \text{shot noise average.} \tag{1.62}$$

   Physically, this reflects the fact that a given cosmological distribution of galaxies can correspond to many different GWB realisations, as the number and times-of-arrival of transient GW signals from each galaxy are essentially random.

Figure 1.13 is a schematic representation of the two averaging procedures.

More quantitatively, the first two moments of the noisy SHCs under the shot noise average for a fixed realisation of LSS are

$$\langle \Omega^i_{\ell m} \rangle_S = \Omega_{\ell m}, \qquad \text{Cov}[\Omega^i_{\ell m}, \Omega^j_{\ell' m'}]_S = \delta_{\ell \ell'} \delta_{mm'} \delta_{ij} \mathcal{W}_\tau. \qquad (1.63)$$

The first equality holds by definition, while the second states that each SHC in each time interval has equal shot noise power, and is uncorrelated with all the others.[11] Averaging also over realisations of LSS, we find

$$\langle \Omega^i_{\ell m} \rangle_{S,\Omega} = 0, \qquad \text{Cov}[\Omega^i_{\ell m}, \Omega^j_{\ell' m'}]_{S,\Omega} = \delta_{\ell \ell'} \delta_{mm'} (C_\ell + \delta_{ij} \mathcal{W}_\tau), \qquad (1.64)$$

where we have used equation (1.17), and have introduced the shorthand $\langle \cdots \rangle_{S,\Omega} \equiv \langle \langle \cdots \rangle_S \rangle_\Omega$.[12] (Recall that we are focusing on $\ell > 0$, so the SHCs all have zero mean.)

1.3.6 MITIGATING THE SHOT NOISE

In this section, we use (1.64) to define a function of the noisy SHCs $\Omega^i_{\ell m}$ that is an unbiased estimator of the true angular power spectrum C_ℓ in the presence of shot noise. We then show that (in the appropriate limit) this is the minimum-variance unbiased estimator (MVUE) for the true angular power spectrum.

We start by modifying the standard autocorrelation estimator (1.22), forming a set of cross-correlations between different time intervals,

$$\hat{C}^{ij}_\ell \equiv \frac{1}{2\ell + 1} \sum_{m=-\ell}^{+\ell} \Omega^i_{\ell m} \Omega^{j*}_{\ell m}. \qquad (1.65)$$

These are unbiased if and only if $i \neq j$,

$$\left\langle \hat{C}^{ij}_\ell \right\rangle_{S,\Omega} = C_\ell + \delta_{ij} \mathcal{W}_\tau. \qquad (1.66)$$

We can combine these estimators in much the same way that we combine multiple 'naive' estimators (1.19) to form the standard estimator (1.22). Suppose that our total observing time is T, so that there are $N_\tau \equiv T/\tau$ time segments. Then we have $N_\tau(N_\tau - 1)/2$ pairs (i, j) for which $i \neq j$. Summing over these, we define the combined estimator

$$\hat{C}_\ell \equiv \frac{2}{N_\tau(N_\tau - 1)} \sum_{ij} \hat{C}^{ij}_\ell, \qquad (1.67)$$

where we introduce the shorthand $\sum_{ij} \equiv \sum_{i=1}^{N_\tau} \sum_{j=i+1}^{N_\tau}$. This is unbiased, $\langle \hat{C}_\ell \rangle_{S,\Omega} = C_\ell$, with variance given by

$$\text{Var}[\hat{C}_\ell]_{S,\Omega} = \left[\frac{2}{N_\tau(N_\tau - 1)(2\ell + 1)} \right]^2 \sum_m \sum_{m'} \sum_{ij} \sum_{i'j'} \text{Cov} \left[\Omega^i_{\ell m} \Omega^{j*}_{\ell m}, \Omega^{i'}_{\ell m'} \Omega^{j'*}_{\ell m'} \right]_{S,\Omega}. \qquad (1.68)$$

---

[11] The lack of correlation between time intervals is due to the acausal relationship between distant GW sources, while the lack of correlation between different ℓm is due to statistical isotropy.

[12] Note that $\text{Cov}[X, Y]_{S,\Omega} \equiv \langle XY^* \rangle_{S,\Omega} - \langle X \rangle_{S,\Omega} \langle Y^* \rangle_{S,\Omega}$, which is *not* equal to $\langle \text{Cov}[X, Y]_S \rangle_\Omega$.

Evaluating (1.68) requires us to evaluate the fourth moment of the noisy SHCs. This would be trivial if the SHCs were all Gaussian, but we must account for the Poisson-like nature of the shot noise. In appendix B we calculate the fourth moment using the same statistical model for the CBC rate density as in section 1.3.2; this results in

$$\text{Var}[\hat{C}_\ell]_{S,\Omega} = \frac{2}{2\ell+1}\left[C_\ell^2 + \frac{2\mathcal{W}_\tau C_\ell}{N_\tau} + \frac{\mathcal{W}_\tau^2}{N_\tau(N_\tau-1)}\right]. \tag{1.69}$$

This expression is tied to the fact that we have excluded the on-diagonal terms $i = j$ when constructing (1.67); otherwise, there would be additional contributions to the variance (the estimator would also no longer be unbiased). Note that since $\mathcal{W}_\tau \propto 1/\tau$, we have $\mathcal{W}_\tau/N_\tau \propto 1/T$. This means that we can't 'win' by decreasing the length of the data segments τ, only by increasing the total observing time T. In fact, writing $\mathcal{W}_T = \mathcal{W}_\tau(\tau/T) = \mathcal{W}_\tau/N_\tau$, we see that in the limit where $N_\tau \gg 1$ (i.e., the limit in which we minimise the contribution of auto-power from each data segment), (1.69) becomes

$$\text{Var}[\hat{C}_\ell]_{S,\Omega} \simeq \frac{2}{2\ell+1}(C_\ell + \mathcal{W}_T)^2. \tag{1.70}$$

This is exactly the standard cosmic variance expression from (1.23), but with C_ℓ replaced by $C_\ell + \mathcal{W}_T$. In section 1.3.7 below, we show that this is in fact the minimum possible variance of any unbiased estimator for the C_ℓ's in the presence of shot noise, saturating the Cramér-Rao bound. The estimator (1.67) is therefore the MVUE in the limit $N_\tau \gg 1$.

At the opposite extreme, for the minimum number of segments, $N_\tau = 2$, the variance is nearly twice as large (taking $\mathcal{W}_T \gg C_\ell$),

$$\text{Var}[\hat{C}_\ell]_{S,\Omega} \approx \frac{4}{2\ell+1}\mathcal{W}_T^2. \tag{1.71}$$

An illustrative example for $N_\tau = 10$ is shown in figure 1.14.

The term $\mathcal{W}_T$ in (1.70) is the same as that appearing in the mean of the standard estimator, $\langle C_\ell^{(\text{std})}\rangle_{S,\Omega} = C_\ell + \mathcal{W}_T$, so this new optimal estimator is still affected by the presence of shot noise; the crucial improvement is that the shot noise only adds to the variance of the estimator, and does not bias the spectrum as in the standard case.

1.3.7 Minimum-variance estimation of the angular power spectrum

Recall that in section 1.1.4 we used the Gaussian likelihood (1.25) to show that the standard cosmic variance expression (1.23) saturates the Cramér-Rao bound, making it the MVUE for the angular power spectrum in the absence of shot noise. Now we include shot noise, and consider the noisy SHCs, $\Omega_{\ell m}^i$. Though we know these are not Gaussian (see appendix B), the Gaussian case is by far the most tractable, so we consider it first. The joint Gaussian log-likelihood is fully specified by (1.64),

$$\mathcal{L} = -\frac{1}{2}\sum_{\ell m}\left[\ln(\det(2\pi\mathsf{C}_\ell)) + \boldsymbol{\Omega}_{\ell m}^\dagger \mathsf{C}_\ell^{-1}\boldsymbol{\Omega}_{\ell m}\right], \tag{1.72}$$

where $\boldsymbol{\Omega}_{\ell m} = (\Omega_{\ell m}^1, \ldots, \Omega_{\ell m}^{N_\tau})$ is a vector of the noisy SHCs for a given ℓm, and C_ℓ is the corresponding $N_\tau \times N_\tau$ covariance matrix,

$$
\mathsf{C}_\ell = \begin{pmatrix}
C_\ell + \mathcal{W}_\tau & C_\ell & \cdots & C_\ell \\
C_\ell & C_\ell + \mathcal{W}_\tau & \cdots & C_\ell \\
\vdots & \vdots & \ddots & \vdots \\
C_\ell & C_\ell & \cdots & C_\ell + \mathcal{W}_\tau
\end{pmatrix}.
\tag{1.73}
$$

One can show that this has determinant $\det \mathsf{C}_\ell = (N_\tau C_\ell + \mathcal{W}_\tau)\mathcal{W}_\tau^{N_\tau - 1}$, and inverse

$$
\mathsf{C}_\ell^{-1} = \frac{1}{(N_\tau C_\ell + \mathcal{W}_\tau)\mathcal{W}_\tau} \begin{pmatrix}
(N_\tau - 1)C_\ell + \mathcal{W}_\tau & -C_\ell & \cdots & -C_\ell \\
-C_\ell & (N_\tau - 1)C_\ell + \mathcal{W}_\tau & \cdots & -C_\ell \\
\vdots & \vdots & \ddots & \vdots \\
-C_\ell & -C_\ell & \cdots & (N_\tau - 1)C_\ell + \mathcal{W}_\tau
\end{pmatrix},
\tag{1.74}
$$

using the matrix determinant lemma and the Sherman-Morrison formula, respectively. Taking the second derivative with respect to C_ℓ, we therefore find

$$
\begin{aligned}
\text{Var}[\hat{C}_\ell]_{S,\Omega} &\geq -\left\langle \frac{\partial^2 \mathcal{L}}{\partial C_\ell^2} \right\rangle_{S,\Omega}^{-1} = -\left\langle \sum_{m=-\ell}^{+\ell} \frac{1}{2(C_\ell + \mathcal{W}_T)^2} - \frac{1}{N_\tau^2(C_\ell + \mathcal{W}_T)^3} \sum_{i=1}^{N_\tau} \sum_{j=1}^{N_\tau} \Omega_{\ell m}^{i*} \Omega_{\ell m}^{j} \right\rangle_{S,\Omega}^{-1} \\
&= \left[\sum_{m=-\ell}^{+\ell} -\frac{1}{2(C_\ell + \mathcal{W}_T)^2} + \frac{1}{N_\tau^2(C_\ell + \mathcal{W}_T)^3} \sum_{i=1}^{N_\tau} \sum_{j=1}^{N_\tau} \left(C_\ell + \delta_{ij}\mathcal{W}_\tau \right) \right]^{-1} \\
&= \left[\sum_{m=-\ell}^{+\ell} \frac{1}{2(C_\ell + \mathcal{W}_T)^2} \right]^{-1} = \frac{2}{2\ell + 1}(C_\ell + \mathcal{W}_T)^2,
\end{aligned}
\tag{1.75}
$$

so the Cramér-Rao bound is the same as before, but with $C_\ell \to C_\ell + \mathcal{W}_T$. This is exactly the variance we derived for our estimator in (1.70).

Equation (1.75) was derived using the Gaussian log-likelihood (1.72), and one may worry about whether it holds in the case we are interested in, given that the noisy SHCs do not follow a Gaussian distribution. However, we have seen that our estimator saturates this bound in the limit $N_\tau \gg 1$. As shown in, e.g., Jaynes [340], the only probability distribution that saturates the Cramér-Rao bound under a given set of constraints (e.g., the constraints on the first two moments in (1.64)) is that which maximises the entropy under those constraints. The maximum-entropy distribution with fixed mean and variance is Gaussian [340], so this shows that our estimator must be Gaussian[13] in the limit $N_\tau \gg 1$, and that the calculations above are valid in that limit. (The approach to Gaussianity for $N_\tau \gg 1$ can also be shown using the central limit theorem.)

---

[13]At least, under the shot noise distribution—the LSS distribution is non-Gaussian as discussed above, but we ignore this complication for now.
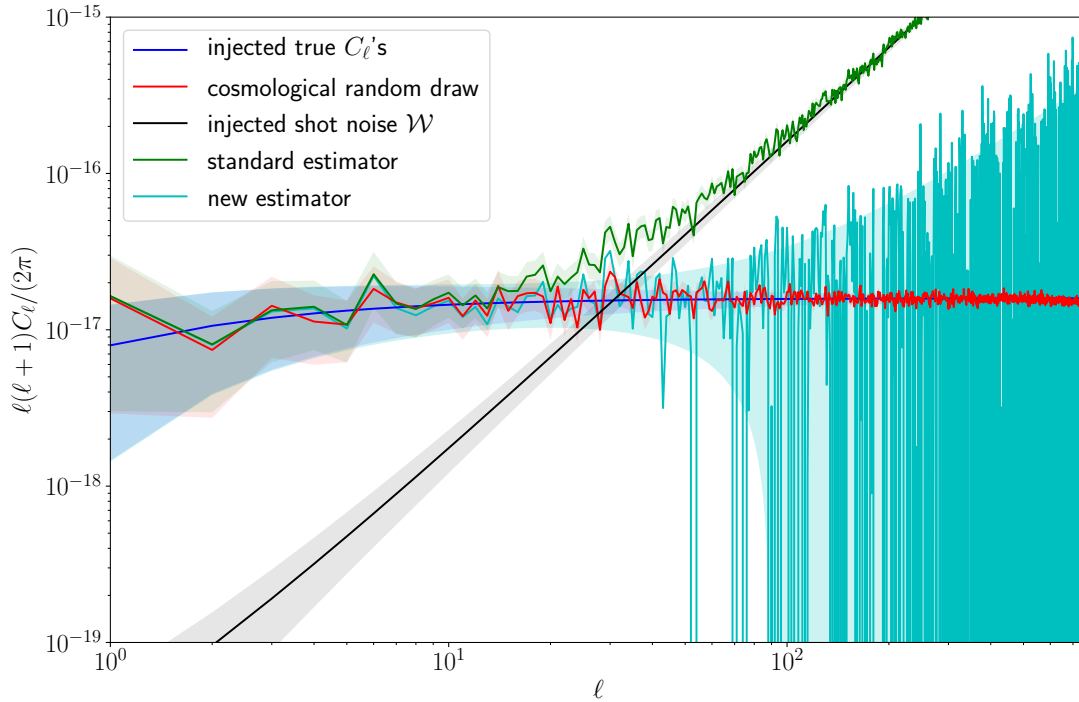
Figure 1.14: Simulated angular power spectra using the standard estimator (1.22) and the new estimator (1.67). The dark blue line is the chosen 'true' spectrum to be estimated, which is here taken as scale-invariant for simplicity, $\ell(\ell+1)C_\ell \approx$ constant. The red line is the spectrum of a single random cosmological realisation of the AGWB (i.e. a single Universe), distributed around the dark blue line according to cosmic variance. The black line is the shot-noise power $\mathcal{W}$, here set to 10^{-3} times the monopole. The green line is the spectrum resulting from the standard estimator (1.22) for a single random realisation from the shot noise ensemble, which follows the sum of the true spectrum and the shot-noise power, $C_\ell + \mathcal{W}$. The cyan line is the spectrum resulting from the new estimator (1.67), for the same shot noise realisation, subdivided into $N_\tau = 10$ independent segments. The shaded regions in all cases show the 1σ uncertainty, which for the cyan line is given by (1.69).

1.4 SUMMARY AND OUTLOOK

In this chapter, we have investigated the angular power spectrum of the astrophysical gravitational-wave background from stellar-mass compact binary coalescences in the LIGO/Virgo frequency band, with the goal of using this signal as a novel probe of late-Universe cosmology. Our key result is the predicted spectrum shown in figure 1.7, which is constructed using a full-sky mock lightcone galaxy catalogue based on the Millennium simulation. Applying simple astrophysical recipes, we are able to model the GW emission from CBCs in all $\sim 5.7 \times 10^6$ galaxies in this catalogue, allowing us to construct a full-sky simulated map of the AGWB (shown in figure 1.6), from which we calculate the corresponding angular power spectrum (shown in figure 1.7). We find that the predicted anisotropies are much larger than in early-Universe observables such as the CMB, and that the one-point statistics of the AGWB map are strongly non-Gaussian, both of which can be understood in terms of the nonlinear gravitational dynamics that generate the clustering signal. We have also calculated the kinematic dipole anisotropy associated with the peculiar motion of our detectors with respect to the cosmic rest frame; however, this is an order of magnitude smaller than the 'intrinsic' clustering signal. While our predicted C_ℓ spectrum is several

orders of magnitude below the current sensitivity of LIGO/Virgo, we saw in figure 1.7 that it should be detected by third-generation ground-based interferometers such as Einstein Telescope.

In section 1.2.3, we considered the impact of CBC population uncertainties on our results. This is an important factor to account for, as the rates and mass distributions of each class of CBC (BBH, BNS, and BHNS) are still rather poorly-determined from present observations. We have therefore calculated the isotropic intensity and angular power spectrum of the AGWB using $\sim 10^4$ different possible sets of BBH population hyperparameters, as inferred from the first five BBH events detected by LIGO/Virgo. We found that the isotropic component of the AGWB is very sensitive to the underlying BBH distribution, varying by more than an order of magnitude, and encoding some interesting information about the shape of the BH mass function. The angular power spectrum, on the other hand, is fairly insensitive to these changes in the BBH population once we normalise with respect to the isotropic amplitude, meaning that our predicted spectrum is robust to these uncertainties. In future work, it would be interesting to carry out a more thorough and up-to-date study of the dependence of the AGWB and its anisotropies on the underlying astrophysical modelling, both in terms of the rates and mass distributions of CBCs, and also in terms of other modelled quantities like the delay-time distribution or the metallicity threshold for BH formation, which could have more of an impact on the angular power spectrum.

In section 1.3, we have pointed out a serious obstacle to using AGWB anisotropies as a cosmological probe: *shot noise*. This is an additional white-noise component of the observed angular power spectrum caused by the finite sampling of the CBC event rate in each galaxy. Since a typical galaxy has a CBC rate of $\sim \mathrm{Myr}^{-1}$, one has to observe for an extremely long time to fully sample the large-scale clustering of galaxies using only GW observations. (There is also an analogous contribution to the shot noise from the finite number of galaxies in a given spatial volume, but this is many orders of magnitude smaller than both the temporal shot noise associated with the CBC event rate and the 'true' angular power spectrum, so we can safely neglect it.) We have developed a novel data-analysis method for mitigating the impact of shot noise on measurements of the C_ℓ spectrum, constructing an estimator which provides an unbiased measurement of the true spectrum, and which attains the minimum possible variance of any such estimator. However, while this method optimally removes the shot noise offset in our C_ℓ measurements, the presence of shot noise still greatly increases the *uncertainty* in these measurements, far beyond the level of cosmic variance.

One promising avenue for future work is to focus on cross-correlating the AGWB anisotropies with other probes of LSS, such as galaxy surveys, rather than simply auto-correlating the AGWB with itself. There are two key advantages to this: first, since the galaxy clustering signal has already been measured with very high SNR in galaxy surveys, the cross-correlation search can act as a 'template' for the AGWB anisotropies, making a detection much easier; second, since galaxy surveys have no temporal shot noise, the total shot noise associated with the cross-correlation spectrum is drastically reduced compared to that of the auto-correlation spectrum, making it much easier (in principle) to extract interesting new information about LSS.

2 NONLINEAR GRAVITATIONAL-WAVE MEMORY FROM COSMIC STRINGS

In section 0.1.8, we introduced the nonlinear gravitational-wave memory effect—a fascinating prediction of general relativity in which oscillatory GW signals are generically accompanied by a permanent strain offset [201], as illustrated in figure 0.3. This effect is sourced by the energy-momentum of the radiated gravitons [542], thus allowing us to directly probe the 'ability of gravity to gravitate' (in the words of Favata [264]). The nonlinear memory effect also has surprising and illuminating links to other fundamental aspects of GR in the infrared regime [530], as well as analogous memory effects in other field theories [124, 448], and therefore presents a powerful tool for advancing our theoretical understanding of gravitational physics.

With these motivations in mind, there has been a significant effort in recent years to calculate [263–266, 370, 436, 460, 537, 589] and search for [53, 85, 142, 160, 215, 241, 249, 294, 321, 330, 350, 367, 389, 402, 509, 577, 591] the nonlinear memory associated with compact binary coalescences. This focus on CBCs is unsurprising, as we have seen in section 0.4 that these are the primary observational target of both interferometer experiments and pulsar timing arrays (PTAs), and are predicted to be abundant sources of nonlinear memory. However, CBCs are far from the only source of interest in GW astronomy, and in particular, it would be interesting to expand our scope to study the nonlinear memory emitted by *cosmological* sources of GWs. On the one hand, these sources might lead to memory observables that are distinct from those associated with CBCs, giving us novel signals to search for. On the other hand, understanding the memory radiated by cosmological sources of GWs might help us to sharpen our theoretical understanding of these sources.

In this chapter, we calculate the nonlinear GW memory emitted by *cosmic strings*, which, as we saw in section 0.3.3, are one of the most important and well-motivated cosmological sources of GWs. We focus in particular on the memory associated with *cusps* and *kinks*, which are the two main mechanisms for GW emission from cosmic string loops. There are at least two reasons to expect *a priori* that cusps and kinks could be important sources of nonlinear memory.[1]

1. GW memory waveforms tend to be associated with lower frequencies than the primary GW emission that sources them [265]. This is because the memory grows monotonically over a period of order the duration of the primary signal, which is typically much longer than the oscillation period

---

[1]We note that cosmic string loops can also source significant amounts of *linear* memory by emitting radiation in the underlying matter fields they are made from. These effects are ignored by the Nambu-Goto approximation we adopt here, and can only be resolved by field-theory simulations. In reality, we expect loops to generate a combination of linear and nonlinear memory, by radiating both matter and GWs. This has recently been demonstrated for collapsing circular cosmic string loops, using numerical-relativity/field-theory simulations [88].

of the primary signal. This means that, e.g., massive stellar binary black holes which merge near the bottom end of the LIGO/Virgo frequency band produce memory signals which are shifted to lower frequencies, and are thus challenging to detect with LIGO/Virgo. Cosmic string signals, on the other hand, have durations which are comparable to their oscillation period (i.e. they have a 'burst-like' morphology, which is very different to an inspiralling compact binary), meaning that we should expect the resulting memory signals to have power at similar frequencies to the original signal. What's more, the cosmic string signals we consider also have significant power at very high frequencies, and it is possible that the hereditary nature of the memory effect could transfer some of this power, enhancing the amplitude of the signal at observable frequencies. (A similar idea of 'orphan' signals—where the memory emission is detectable even though the primary signal is not—was studied by McNeill *et al.* [421].)

2. The angular pattern of the GW memory signal on the sky is typically different to that of the primary GW emission that sources it. (We can see this from the angular integral in equation (0.47).) As we will see in section 2.1.3, cosmic string cusps and kinks emit GWs in narrow beams, meaning that only a very small fraction of all cusps and kinks are oriented such that their GWs can be observed. However, if the associated GW memory signal is more broadly distributed on the sphere, this might allow us to observe some of the many cusps and kinks whose beams are not oriented towards us.

We find that both expectations are borne out by our calculations below: the GW memory from cusps and kinks is indeed emitted in a much broader range of directions than the initial beam, and the memory signal does indeed have a similar frequency profile to the primary signal, with the high-frequency behaviour of the primary GWs playing an important rôle in determining the strength of the memory effect. In fact, we show that these two ingredients lead to a *divergence* in the memory signal from cusps on sufficiently large loops, potentially signifying a breakdown of the standard weak-field approach for calculating the GW signal from cusps. We attempt to clarify the root cause of this breakdown, and suggest one tentative possible resolution in which the cusp collapses to form a primordial black hole, 'trapping' the high-frequency GW emission behind a horizon and thereby preventing the memory from diverging. Other resolutions are possible however, and ultimately a fully general-relativistic treatment will be required to understand the true behaviour of cusps.

The remainder of this chapter is structured as follows. We begin in section 2.1 by introducing the dynamics of cosmic string loops in the Nambu-Goto approximation in the context of linearised GR, deriving the standard GW waveforms for cusps and kinks, and briefly discussing how these are used to calculate the GW background emitted by the cosmic string loop network. This first section is a review of existing work, mainly following the exposition in chapter 6 of Vilenkin and Shellard [567], as well as the derivation of the cusp and kink waveforms by Damour and Vilenkin [227]. The remainder of the chapter consists of original work. In section 2.2 we derive some useful formulae for nonlinear memory, building on equation (0.47) to find expressions for the late-time memory and Fourier-domain memory waveform in terms of the primary strain signal. In section 2.3, we calculate the nonlinear GW memory signal associated with cusps, obtaining the simple frequency-domain waveform (2.64); we show that the total GW energy radiated by this memory signal diverges for Nambu-Goto strings, and regularise this divergence

by imposing a cutoff at the scale of the string width δ; we then go on to consider higher-order memory effects (the nonlinear GW memory sourced by the memory GWs themselves) and show that accounting for all such contributions leads to a divergence for large loops, which persists even after applying the string-width regularisation; finally, we sketch how this divergence is resolved in the cusp-collapse scenario. In section 2.4 we repeat the memory calculation for kinks, obtaining the leading-order waveform (2.97); unlike in the cusp case, the memory signal is strongly suppressed at high frequencies due to interference effects, and no divergence occurs. In section 2.5 we flesh out the cusp-collapse proposal, showing how it arises naturally from arguments based on the hoop conjecture, and derive some of the properties of the resulting PBHs. In section 2.6 we study the observable consequences of our results. We find that, in the scenario where the memory divergence is cured by cusp collapse, the GW emission associated with the memory is strongly suppressed, and is beyond the reach of current or planned GW searches. However, we show that the PBHs formed in this scenario form a unique population compared to other astrophysical and primordial formation mechanisms, making them a 'smoking gun' signature of cosmic strings. Finally, we summarise our results in section 2.7. In appendix C we argue that the nonlinear memory divergence is associated with a trans-Planckian GW flux, as well as giving some technical details of the angular distribution of the memory radiation from cusps and kinks.

2.1 C O S M I C   S T R I N G   L O O P S :   D Y N A M I C S   A N D   G R A V I T A T I O N A L - W A V E   E M I S S I O N

The formation of cosmic strings in the early Universe naturally leads to the production of cosmic string *loops* through the self-intersection of long strings. The lengths of these loops span an enormous range of physical scales; initially created with lengths comparable to the Hubble scale, the emission of GWs causes loops to gradually shrink to microscopic sizes, eventually unwinding and dispersing once their length is comparable to the string width scale. In order to understand the GWs emitted through this process, it is vital to understand the dynamics of individual loops. In principle, we should do this by solving the equations of motion for the matter fields that constitute the loop, which usually involves running highly expensive large-scale lattice simulations. However, in practice we can make significant progress analytically by effacing the microscopic dynamics of the underlying fields, and focusing on the effective, macroscopic equations of motion for the loops themselves. We do this using the *Nambu-Goto approximation*, which we introduce below, following sections 6.1 and 6.2 of Vilenkin and Shellard [567].

2.1.1 T H E   N A M B U - G O T O   A P P R O X I M A T I O N

We saw in section 0.3.3 that the width of a cosmic string is roughly $\delta \sim \sqrt{\hbar/\mu}$, where μ is the string tension. Since cosmic strings are formed at such high energies, this is a *tiny* lengthscale: some seventeen orders of magnitude smaller than the radius of a proton for GUT-scale strings. By comparison, the lengths ℓ of the loops we are interested in are astronomically large, $\ell \sim$ pc–Gpc. It is therefore natural for us to treat these loops as purely one-dimensional objects, setting their width δ to zero. (Note that this is equivalent to taking $\hbar \rightarrow 0$, so we can think of this as a classical limit in which we neglect the

microphysical degrees of freedom.) This allows us to describe their motion in terms of a two-dimensional surface in spacetime called the *worldsheet*.

We cover this surface with coordinates ζ^I, $(I = 0, 1)$, and describe the location of a given point on the worldsheet in four-dimensional spacetime in terms of four functions, $X^\alpha(\zeta)$. The metric that determines the spacetime interval between points on the worldsheet, γ_{IJ}, is given by restricting the metric of the full spacetime, $g_{\alpha\beta}$, down to this surface,

$$\gamma_{IJ} \equiv g_{\alpha\beta} \frac{\partial X^\alpha}{\partial \zeta^I} \frac{\partial X^\beta}{\partial \zeta^J}. \tag{2.1}$$

If the strings are formed through the breaking of a *local* symmetry, then we should be able to encapsulate their equations of motion with a local worldsheet action[2] of the form $S = \int \mathrm{d}^2\zeta \sqrt{-\gamma}\mathcal{L}$, where $\gamma \equiv \det \gamma_{IJ}$. Since S has dimensions of (mass × length), the Lagrangian density $\mathcal{L}$ has dimensions of (mass/length), just like the string tension μ. One obvious choice is therefore to take $\mathcal{L} = A\mu$, where A is some unknown dimensionless constant. We can fix the appropriate value of A by evaluating the energy-momentum tensor of a string with this action, using the procedure in equation (0.4) to obtain

$$T^{\alpha\beta}(x) = \frac{2}{\sqrt{-g}} \frac{\delta S}{\delta g_{\alpha\beta}} = \frac{A\mu}{\sqrt{-g}} \int \mathrm{d}^2\zeta \sqrt{-\gamma}\gamma^{IJ} \frac{\partial X^\alpha}{\partial \zeta^I} \frac{\partial X^\beta}{\partial \zeta^J} \delta^{(4)}(x - X), \tag{2.2}$$

where we have used the standard identity

$$\delta\sqrt{-\gamma} = \frac{1}{2}\sqrt{-\gamma}\gamma^{IJ}\delta\gamma_{IJ}. \tag{2.3}$$

Evaluating this for the simplest possible case of an infinite straight string (aligned with the x^1 axis without loss of generality) on a flat background spacetime, and choosing $\zeta^0 = x^0$, $\zeta^1 = x^1$, we find the energy density

$$T_{00}(x) = -A\mu\delta(x^2 - X^2)\delta(x^3 - X^3). \tag{2.4}$$

We see that if we set $A = -1$, then this agrees exactly with the energy density of an infinitely thin string of linear density μ. We thus obtain the *Nambu-Goto action*

$$S_{\mathrm{NG}} = -\mu \int \mathrm{d}^2\zeta \sqrt{-\gamma}. \tag{2.5}$$

Since $\sqrt{-\gamma}$ is the area element on the worldsheet, we can interpret this as the action that minimises the total worldsheet area.

Note that μ is not the only quantity we can use to construct the Lagrangian. On dimensional grounds, we could write down the more general expression

$$\mathcal{L} = -\mu + B\hbar\kappa + C\frac{\hbar^2\kappa^2}{\mu} + \cdots, \tag{2.6}$$

---

[2]In contrast, strings associated with *global* symmetries are characterised by long-range non-gravitational interactions, and have a very distinct phenomenology compared to local strings as a result.

where the κ's represent worldsheet curvature invariants, with indices suppressed. However, since a loop of length ℓ will generally have curvature of order $\kappa \sim \ell^{-2}$, we see that $\hbar\kappa/\mu \sim (\delta/\ell)^2 \ll 1$, so we can safely discard all but the leading term, leaving the Nambu-Goto action (2.5).

This approximation breaks down if we consider small enough loops. In particular, a more detailed analysis shows that field-theoretic effects become important for loops smaller than

$$\ell_{\min} \equiv \frac{\delta}{G\mu} \approx \frac{\ell_{\rm Pl}}{(G\mu)^{3/2}} \approx 5.1 \times 10^{-19} \, {\rm m} \times \left(\frac{G\mu}{10^{-11}}\right)^{-3/2}. \tag{2.7}$$

(Recall from section 0.3.3 that GWB searches by LIGO/Virgo [4, 18, 24, 36, 37] and PTAs [131, 388, 593] constrain the string tension to be $G\mu \lesssim 10^{-11}$. We therefore use $G\mu = 10^{-11}$ as a representative value throughout this chapter, and use this to give numerical values of key physical quantities where relevant.) These loops are expected to rapidly lose their energy through radiation in the underlying matter fields [316, 418, 524] or through topological unwinding and dispersion [88, 310], and are therefore uninteresting for our purposes.[3]

2.1.2 THE FLAT-SPACE EQUATIONS OF MOTION

We can find the equations of motion for a Nambu-Goto string by extremising the action (2.5) with respect to the spacetime coordinates $X^\alpha(\zeta)$ which, from the worldsheet perspective, are just a set of four scalar fields. Since our goal is ultimately to use the EoMs to calculate GW signals emitted by loops in linearised GR, we can specialise to a flat background spacetime, such that

$$\gamma_{IJ} = \eta_{\alpha\beta} \frac{\partial X^\alpha}{\partial \zeta^I} \frac{\partial X^\beta}{\partial \zeta^J}. \tag{2.8}$$

Varying the action then gives

$$\begin{aligned}
\delta S &= -\frac{\mu}{2} \int {\rm d}^2\zeta \, \sqrt{-\gamma}\, \gamma^{IJ} \delta\gamma_{IJ} = -\mu\eta_{\alpha\beta} \int {\rm d}^2\zeta \, \sqrt{-\gamma}\, \gamma^{IJ} \frac{\partial X^\alpha}{\partial \zeta^I} \frac{\partial}{\partial \zeta^J} \delta X^\beta \\
&= \mu \int {\rm d}^2\zeta \, \delta X_\alpha \frac{\partial}{\partial \zeta^J}\left(\sqrt{-\gamma}\, \gamma^{IJ} \frac{\partial X^\alpha}{\partial \zeta^I}\right),
\end{aligned} \tag{2.9}$$

---

[3]There is a long-standing debate within the cosmic string community as to whether matter radiation can be safely neglected for loops much larger than $\ell_{\min}$. Field theory simulations of Abelian-Higgs loop networks appear to show that matter radiation dominates over GW emission on large scales [314, 315, 568], causing the network to decay rapidly into the matter fields, thus significantly suppressing the expected GWB signal. These results are in tension with theoretical arguments and with simulations of individual loops, which support the idea that loops much larger than $\ell_{\min}$ can safely be treated in the Nambu-Goto approximation. It has been suggested that the anomalous matter radiation rates observed in the field-theory network simulations may be a numerical artefact, tied to the limited dynamical range of these simulations compared to the enormous separation of scales in the cosmological scenarios of interest. However, this remains an open question. Throughout this chapter, we assume that field-theory effects are negligible for loops much larger than $\ell_{\min}$ (though we will return to this point in section 2.3.9). We refer the interested reader to, e.g., section 3.4 of Auclair *et al.* [87] or section 9.2 of Caprini and Figueroa [170] for further discussion of this issue.

where we have used equation (2.3) again, and have ignored the boundary term when integrating by parts.[4] In order for this to vanish, the coordinates X^α must each obey a wave equation,

$$^{(2)}\Box X^\alpha \equiv \frac{1}{\sqrt{-\gamma}} \frac{\partial}{\partial \zeta^I} \left(\sqrt{-\gamma} \gamma^{IJ} \frac{\partial X^\alpha}{\partial \zeta^J} \right) = 0, \tag{2.10}$$

where $^{(2)}\Box$ is the two-dimensional covariant d'Alembertian operator on the worldsheet. The lack of a mass term here shows us that the loop's degrees of freedom propagate at the speed of light.

Thus far, we have kept our worldsheet coordinates completely general, but as with the gauge freedom in the four-dimensional spacetime coordinates that we dealt with in section 0.1, there is a great deal of redundancy here. We can simplifying things considerably with a judicious choice of gauge. One particularly convenient choice is to select ζ such that the worldsheet metric is conformally flat, $\gamma_{IJ} = \sqrt{-\gamma}\eta_{IJ}$; this is called the 'conformal gauge'. Equations (2.8) and (2.10) then become

$$\ddot{X}^\alpha - X^{\alpha\prime\prime} = 0, \qquad \dot{X}^\alpha \dot{X}_\alpha + X^{\alpha\prime} X'_\alpha = 0, \qquad \dot{X}^\alpha X'_\alpha = 0, \tag{2.11}$$

where dots and primes indicate derivatives with respect to ζ^0 and ζ^1, respectively. These equations are simplified even further if we now choose the timelike coordinate such that $\zeta^0 = X^0 \equiv t$. The loop's trajectory is then fully described by a 3-vector $\boldsymbol{X}(t, \sigma)$, where $\sigma \equiv \zeta^1$ is the spacelike coordinate in this gauge, and the EoMs become

$$\ddot{\boldsymbol{X}} - \boldsymbol{X}'' = \boldsymbol{0}, \tag{2.12}$$

$$|\dot{\boldsymbol{X}}|^2 + |\boldsymbol{X}'|^2 = 1, \tag{2.13}$$

$$\dot{\boldsymbol{X}} \cdot \boldsymbol{X}' = 0. \tag{2.14}$$

Each of these three equations has an intuitive physical interpretation. Since $\boldsymbol{X}'$ is the tangent vector pointing along the string, equation (2.14) tells us that $\dot{\boldsymbol{X}}$ should be interpreted as the *transverse* velocity of the string; we cannot observe velocities along the tangential direction. Rearranging equation (2.13) shows us that intervals in the worldsheet coordinate σ are related to intervals in spatial distance $|\boldsymbol{X}|$ by

$$d\sigma = \frac{|d\boldsymbol{X}|}{|\boldsymbol{X}'|} = \frac{|d\boldsymbol{X}|}{\sqrt{1 - |\dot{\boldsymbol{X}}|^2}} = \gamma|d\boldsymbol{X}| = \frac{dE}{\mu}, \tag{2.15}$$

where $\gamma = 1/\sqrt{1 - |\dot{\boldsymbol{X}}|^2}$ is the Lorentz factor of the string segment (not to be confused with the determinant of the worldsheet metric γ_{IJ}), and $dE = \gamma\mu|d\boldsymbol{X}|$ is the energy (kinetic plus rest mass) of the string segment.[5] This shows us that by choosing this gauge, we have scaled σ such that it is proportional to the energy at every point along the string, $\mu \, d\sigma = dE$. Finally, equation (2.12) tells us that the degrees of

---

[4]If we had kept a general spacetime metric $g_{\alpha\beta}$, then we would have picked up an additional term here containing the four-dimensional Christoffel symbols $\Gamma^\nu_{\alpha\beta}$. This correction vanishes on a flat spacetime.

[5]This expression for the energy accords with our intuition from point particles in flat space, but we also could have derived it from the energy momentum tensor (2.2).

freedom propagating tangentially along the string obey a wave equation and propagate at the speed of light. We can make this more explicit by decomposing $\boldsymbol{X}$ into left- and right-moving modes,

$$\boldsymbol{X}(t, \sigma) = \frac{1}{2}[\boldsymbol{X}_+(\sigma_+) + \boldsymbol{X}_-(\sigma_-)], \qquad \sigma_\pm \equiv t \pm \sigma. \tag{2.16}$$

This decomposition is always possible for any solution to equation (2.12), and also obeys equations (2.13) and (2.14) if we set

$$|\dot{\boldsymbol{X}}_+| = |\dot{\boldsymbol{X}}_-| = 1, \tag{2.17}$$

showing that both modes do indeed propagate at the speed of light. (Note that differentiating $\boldsymbol{X}_\pm$ with respect to time gives the same result as differentiating with respect to their argument, $\sigma_\pm$.)

Since we are interested in cosmic string *loops*, $\boldsymbol{X}$ must be periodic in σ, with some finite periodicity length ℓ. From equation (2.15), we see that this length is set by the total energy of the loop, since $E = \mu \int_0^\ell \mathrm{d}\sigma = \mu\ell$. This implies that ℓ must be constant over time (neglecting for now energy loss from GW radiation), so we call it the *invariant length*. Note that this differs from the coordinate length of the loop,

$$L = \int_0^\ell \mathrm{d}\sigma\, |\boldsymbol{X}'| = \int_0^\ell \mathrm{d}\sigma\, \sqrt{1 - |\dot{\boldsymbol{X}}^2|} \le \ell, \tag{2.18}$$

which fluctuates over time, undergoing relativistic length contraction due to the motion of the loop.

The left-right decomposition (2.16) makes it clear that $\boldsymbol{X}_\pm$ must also be periodic in $\sigma_\pm \in [0, \ell)$, which in turn implies that $\boldsymbol{X}$ is periodic in *time* as well as space. In fact, the period of the loop's motion over time is $\ell/2$ rather than ℓ, since equation (2.16) gives $\boldsymbol{X}(t + \ell/2, \sigma + \ell/2) = \boldsymbol{X}(t, \sigma)$. This shows us that it's not just modes moving along the loop that are relativistic, but that the loop itself must oscillate at relativistic speeds. Indeed, using equations (2.12) and (2.13) we see that the loop's rms velocity, averaged over its length and over an oscillation period, is given by

$$v_{\mathrm{rms}}^2 \equiv \left\langle |\dot{\boldsymbol{X}}|^2 \right\rangle_\ell = \left\langle \boldsymbol{X} \cdot \ddot{\boldsymbol{X}} \right\rangle_\ell = \left\langle \boldsymbol{X} \cdot \boldsymbol{X}'' \right\rangle_\ell = \left\langle |\boldsymbol{X}'|^2 \right\rangle_\ell = \left\langle 1 - |\dot{\boldsymbol{X}}|^2 \right\rangle_\ell = 1 - v_{\mathrm{rms}}^2,$$

$$\langle \cdots \rangle_\ell \equiv \int_0^\ell \frac{\mathrm{d}\sigma}{\ell} \int_0^{\ell/2} \frac{\mathrm{d}t}{\ell/2} (\cdots), \tag{2.19}$$

where periodicity allows us to drop the boundary terms when integrating by parts. We therefore have $v_{\mathrm{rms}} = 1/\sqrt{2}$; i.e., the loop moves at $\approx 71\%$ the speed of light on average.

Cusps and kinks

Another way of understanding equation (2.19) is to think about the net velocity of the loop as the average of the velocity vectors of the left- and right-moving modes, $\dot{\boldsymbol{X}} = (\dot{\boldsymbol{X}}_+ + \dot{\boldsymbol{X}}_-)/2$. Since we saw in equation (2.17) that these each have unit magnitude at all times and at all points on the loop, the net velocity depends only on the *relative* velocity of the two modes. If we define $\theta_\pm \equiv \cos^{-1}(\dot{\boldsymbol{X}}_+ \cdot \dot{\boldsymbol{X}}_-)$ as the angle between the two velocity vectors, we see that the net velocity is given by

$$|\dot{\boldsymbol{X}}|^2 = \frac{1}{4}\left(|\dot{\boldsymbol{X}}_+|^2 + 2\cos\theta_\pm + |\dot{\boldsymbol{X}}_-|^2\right) = \frac{1 + \cos\theta_\pm}{2}. \tag{2.20}$$
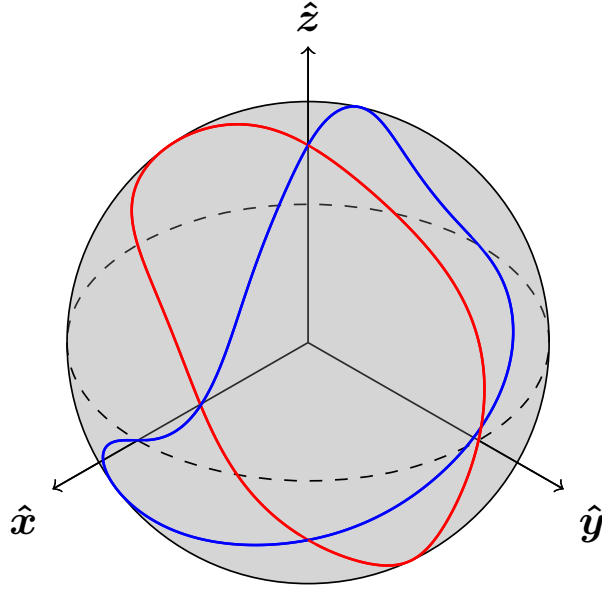
Figure 2.1: The left- and right-moving mode velocities $\dot{\boldsymbol{X}}_\pm$ for a given loop configuration can be drawn as two curves on the Kibble-Turok sphere. These curves generically intersect each other, giving rise to cusps. Note that the curves drawn here are continuous, and therefore correspond to loops which have no kinks.

We thus reproduce equation (2.19) if $\cos\theta_\pm$ averages to zero over an oscillation period (i.e., the two mode velocities are just as likely to be aligned as anti-aligned).

From equation (2.20), we see that it is possible for the loop itself to move at the speed of light if we can arrange for both the left- and right-moving modes to be moving in the same direction, $\dot{\boldsymbol{X}}_+ = \dot{\boldsymbol{X}}_-$ (since then $\theta_\pm = 0$). A point on the worldsheet where such a situation occurs is called a *cusp* [316, 554, 565, 567]. At first glance, cusps seem finely-tuned and unlikely to occur for generic loop trajectories. However, while it is true that a randomly-selected point on the worldsheet is unlikely to be a cusp, the occurrence of a cusp *somewhere* on the loop during the course of an oscillation period is actually very generic. To see this, consider the possible forms that the functions $\dot{\boldsymbol{X}}_\pm$ can take. These are both periodic 3-vector functions with unit magnitude, so we can draw them as closed curves on the unit 2-sphere (as illustrated in figure 2.1; this is often referred to as the Kibble-Turok sphere). Since $\int_0^\ell \mathrm{d}\sigma_\pm \, \dot{\boldsymbol{X}}_\pm = \boldsymbol{X}_\pm(\ell) - \boldsymbol{X}_\pm(0) = \boldsymbol{0}$, neither of these curves can reside entirely in one hemisphere, as the integral $\int_0^\ell \mathrm{d}\sigma_\pm \, \dot{\boldsymbol{X}}_\pm$ would then be nonzero. As a result, it is quite hard for the two curves to avoid each other completely, and there is generically at least one point at which they intersect, with each intersection corresponding to a cusp.

One way in which the two curves can more easily avoid intersecting is if they possess discontinuities; i.e., if there are values of $\sigma_\pm$ at which $\dot{\boldsymbol{X}}_\pm$ jumps between two disconnected points on the unit sphere. Physically this corresponds to a discontinuity in the loop's tangent vector $\boldsymbol{X}'$, at which two smooth string segments are joined at a sharp angle.[6] These features are called *kinks*, and occur very naturally due to the string self-intersections from which loops are created. Each kink is associated with a particular fixed value

---

[6]Of course, such a sharp feature is only possible in our idealised Nambu-Goto setup, where the string has zero width. In the full field-theoretic description, kinks correspond to points where the loop's tangent vector changes significantly over an interval of order the string width δ, but in such a way that the matter fields are everywhere continuous.

of σ_+ or σ_-, meaning that the kink propagates around the loop at the speed of light, with the direction of travel depending on whether the discontinuity is associated with a left- or a right-moving mode.

In the following section, we show that cusps and kinks are the two main mechanisms through which loops radiate GWs.

2.1.3 GRAVITATIONAL-WAVE EMISSION FROM LOOPS

We now turn to the loop's GW emission, following the derivation of the cusp and kink waveforms first given by Damour and Vilenkin [227]. Using the results from section 0.1.6, we can write the Fourier transform of the complex strain waveform $h = h_+ - ih_\times$ emitted by the loop in direction $\hat{\boldsymbol{r}}$ as

$$\tilde{h}(f, \hat{\boldsymbol{r}}) = \frac{2G}{r} \left[e^{+,ij}(\hat{\boldsymbol{r}}) - ie^{\times,ij}(\hat{\boldsymbol{r}}) \right] \tilde{T}_{ij}(f, f\hat{\boldsymbol{r}}), \tag{2.21}$$

where $\tilde{T}_{ij}$ is the spacetime Fourier transform of the energy-momentum tensor,

$$\tilde{T}_{ij}(f, \boldsymbol{k}) = \int \mathrm{d}t \int \mathrm{d}^3\boldsymbol{x} \, e^{-2\pi i(ft - \boldsymbol{k}\cdot\boldsymbol{x})} T_{ij}(t, \boldsymbol{x}). \tag{2.22}$$

In our case, this is given by evaluating equation (2.2) in the conformal gauge,

$$T_{ij}(t, \boldsymbol{x}) = \mu \int_0^\ell \mathrm{d}\sigma \, \delta^{(3)}(\boldsymbol{x} - \boldsymbol{X}) \left(\dot{X}_i \dot{X}_j - X_i' X_j' \right),$$

$$\tilde{T}_{ij}(f, \boldsymbol{k}) = \mu \int_0^{\ell/2} \mathrm{d}t \int_0^\ell \mathrm{d}\sigma \, e^{-2\pi i(ft - \boldsymbol{k}\cdot\boldsymbol{X})} \left(\dot{X}_i \dot{X}_j - X_i' X_j' \right), \tag{2.23}$$

where we have restricted the time integral to $t \in [0, \ell/2]$, as we are interested in the GW emission from a single oscillation period. We can simplify this by writing everything in terms of the left- and right-moving modes defined in equation (2.16). Using $\boldsymbol{X}_\pm' = \pm\dot{\boldsymbol{X}}_\pm$, and changing the area element to $\mathrm{d}\sigma_+ \, \mathrm{d}\sigma_- = 2 \, \mathrm{d}t \, \mathrm{d}\sigma$, we see that the energy-momentum tensor can be factorised into two separate integrals corresponding to the left- and right-moving parts, such that the waveform is given by

$$\tilde{h}(f, \hat{\boldsymbol{r}}) = \frac{G\mu}{r} (e_{ij}^+ - ie_{ij}^\times) \mathcal{I}_+^i \mathcal{I}_-^j, \qquad \mathcal{I}_\pm^i(f, \hat{\boldsymbol{r}}) \equiv \int_{-\ell/2}^{+\ell/2} \mathrm{d}\sigma_\pm \, e^{-\pi i f(\sigma_\pm - \hat{\boldsymbol{r}}\cdot\boldsymbol{X}_\pm)} \dot{X}_\pm^i. \tag{2.24}$$

Since the loop's motion is periodic, the GW emission described by equation (2.24) is not a continuous spectrum, but is concentrated at the discrete set of frequencies $2n/\ell$, where n is a nonzero integer, and $2/\ell$ is the fundamental mode of the loop. This distinction is unimportant in most cases however, as the fundamental mode corresponds to a very low frequency for the loop sizes $\ell \sim$ pc–Gpc we are interested in, $2/\ell \approx 2 \times 10^{-8}$ Hz $\times (\ell/\mathrm{pc})^{-1}$, meaning that we are typically interested in very high harmonics $|n| \gg 1$, particularly for interferometers like LIGO/Virgo. In this high-frequency regime, the emission modes of the loop are so closely spaced that they effectively form a continuous spectrum. The downside of this is that the integrals $\mathcal{I}_\pm^i$ are highly oscillatory in this regime, meaning that the GW emission typically falls off exponentially with frequency.

There are two key ways in which this exponential suppression can be avoided, giving a strong high-frequency GW signal [227]:

1. If there is a *stationary point* in the phase, such that $\partial/\partial\sigma_\pm\left(\sigma_\pm - \hat{\boldsymbol{r}} \cdot \boldsymbol{X}_\pm\right) = 0$ for some value of $\sigma_\pm$. This corresponds to having $\dot{\boldsymbol{X}}_\pm = \hat{\boldsymbol{r}}$, which occurs whenever the left-/right-moving mode in question is propagating directly towards the observer.

2. If there is a *discontinuity* in the integrand or one of its derivatives. Since $\boldsymbol{X}_\pm$ itself must be continuous (recall that cosmic strings are topologically forbidden from having open ends), the strongest discontinuity that we can allow is a jump in the first derivative $\dot{\boldsymbol{X}}_\pm$, i.e., a kink.

In order to prevent the GW strain (2.24) from decaying exponentially, we need one or the other of these conditions to be satisfied for *both* the left- and right-moving modes. There are three possible situations that achieve this:

Cusp: Both integrals $\mathcal{I}_\pm^i$ possess a stationary point along the same propagation direction $\hat{\boldsymbol{r}}_{\mathrm{c}}$. This corresponds to having $\hat{\boldsymbol{r}}_{\mathrm{c}} = \dot{\boldsymbol{X}}_+ = \dot{\boldsymbol{X}}_-$; in other words, a cusp pointing in the $\hat{\boldsymbol{r}}_{\mathrm{c}}$ direction.

Kink: One integral has a stationary point along the GW propagation direction $\hat{\boldsymbol{r}}$, and the other has a kink. This corresponds to a point in the loop's oscillation at which the kink is travelling in the $\hat{\boldsymbol{r}}$ direction.

Kink-kink collision: Both integrals have kinks. We can then interpret the emitted GWs as being due to a *kink-kink collision*. Since neither kink is associated with a particular direction $\hat{\boldsymbol{r}}$ (as there is no stationary-phase requirement), the resulting GW emission must be isotropic.

We calculate the corresponding waveform for each of these cases below.

Cusps

Let's look at the cusp case first. Here we can use the stationary-phase approximation to evaluate the integrals $\mathcal{I}_\pm^i$, focusing on a small interval in $\sigma_\pm$ near the cusp. This is useful because the gauge conditions (2.17) constrain the form of the functions $\boldsymbol{X}_\pm(\sigma_\pm)$ in a small region near the cusp, allowing us to derive results that are generic to *all* cusps. Choosing our coordinates so that the cusp occurs at $\sigma_\pm = 0$, we can write the position and velocity of the left- and right-moving modes as Taylor expansions around this point,

$$\boldsymbol{X}_\pm = \hat{\boldsymbol{r}}_{\mathrm{c}}\sigma_\pm + \frac{1}{2}\ddot{\boldsymbol{X}}_\pm\sigma_\pm^2 + \frac{1}{6}\dddot{\boldsymbol{X}}_\pm\sigma_\pm^3 + \cdots, \qquad \dot{\boldsymbol{X}}_\pm = \hat{\boldsymbol{r}}_{\mathrm{c}} + \ddot{\boldsymbol{X}}_\pm\sigma_\pm + \frac{1}{2}\dddot{\boldsymbol{X}}_\pm\sigma_\pm^2 + \cdots, \qquad (2.25)$$

with all the derivatives evaluated at $\sigma_\pm = 0$. We can then apply the gauge conditions (2.17) order-by-order to find

$$\hat{\boldsymbol{r}}_{\mathrm{c}} \cdot \ddot{\boldsymbol{X}}_\pm = 0, \qquad \hat{\boldsymbol{r}}_{\mathrm{c}} \cdot \dddot{\boldsymbol{X}}_\pm = -|\ddot{\boldsymbol{X}}_\pm|^2. \qquad (2.26)$$

As an example of how this determines the properties of the loop near the cusp, consider the loop's position at time $t = 0$,

$$\boldsymbol{X}(0, \sigma) = \frac{1}{2}\left[\boldsymbol{X}_+(\sigma) + \boldsymbol{X}_-(-\sigma)\right] = \frac{1}{4}(\ddot{\boldsymbol{X}}_+ + \ddot{\boldsymbol{X}}_-)\sigma^2 + \frac{1}{12}(\dddot{\boldsymbol{X}}_+ - \dddot{\boldsymbol{X}}_-)\sigma^3 + \cdots. \qquad (2.27)$$
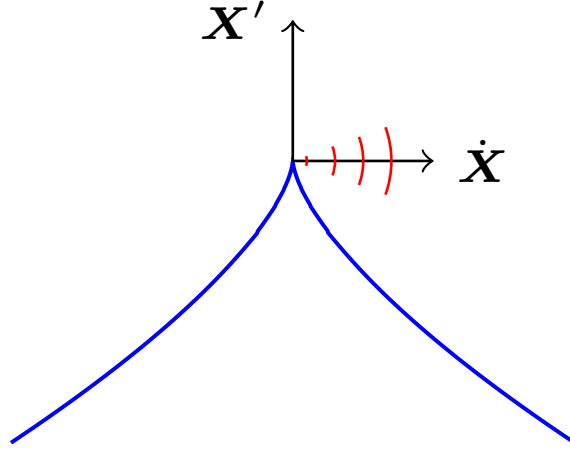
Figure 2.2: Diagram of a string segment (blue) near a cusp, showing the tangent vector $\boldsymbol{X}'$ and velocity $\dot{\boldsymbol{X}}$ at the cusp. Note the $y \propto x^{2/3}$ shape of this curve (where x and y are the distances of the loop from the cusp in the directions parallel to and transverse to the cusp velocity, respectively), which is characteristic of all cusps. As we show below, the emitted GWs (red) are beamed along the direction of the cusp velocity.

Using the constraints in equation (2.26), we see that the loop's distance from the cusp in the direction parallel to $\dot{\boldsymbol{X}}$ (call this x) is related to the distance in the transverse direction (call this y) by $y \propto x^{2/3}$. This gives rise to a universal shape for the string segment near the cusp, which is illustrated in figure 2.2. The proportionality constant relating x and y depends on the values of the derivatives $\ddot{\boldsymbol{X}}_\pm, \dddot{\boldsymbol{X}}_\pm$ at the cusp, but we can generically expect it to include a factor of $\ell^{1/3}$ on dimensional grounds, as the loop length ℓ is the only relevant lengthscale in the problem (here it is hidden in the values of the derivatives, as we discuss below when estimating $|\dddot{\boldsymbol{X}}_\pm|$). In other words, this universal cusp shape is scaled up and down according to the total length of the loop; a slightly counterintuitive idea, given that the cusp is a small, localised feature which one might not expect to 'know' about the rest of the loop, but one which is important later in this chapter.

Returning now to the GW signal emitted by a cusp, we see that equation (2.26) implies that the leading contribution to the phase factor in equation (2.24) is of order $\sigma_\pm^3$,

$$\sigma_\pm - \hat{\boldsymbol{r}}_{\mathrm{c}} \cdot \boldsymbol{X}_\pm = -\frac{1}{2}\hat{\boldsymbol{r}}_{\mathrm{c}} \cdot \ddot{\boldsymbol{X}}_\pm \sigma_\pm^2 - \frac{1}{6}\hat{\boldsymbol{r}}_{\mathrm{c}} \cdot \dddot{\boldsymbol{X}}_\pm \sigma_\pm^3 + \cdots = \frac{1}{6}|\ddot{\boldsymbol{X}}_\pm|^2 \sigma_\pm^3 + \cdots . \tag{2.28}$$

We therefore write

$$\mathcal{I}_\pm^i(f, \hat{\boldsymbol{r}}_{\mathrm{c}}) \simeq \int_{-\ell/2}^{+\ell/2} \mathrm{d}\sigma_\pm \left(\hat{r}_*^i + \ddot{X}_\pm^i \sigma_\pm \right) \exp\left(-\frac{\pi \mathrm{i}}{6} f |\ddot{\boldsymbol{X}}_\pm|^2 \sigma_\pm^3 \right). \tag{2.29}$$

Notice that the leading-order term here is directed along the line of sight, $\mathcal{I}_\pm^i \propto \hat{r}_*^i$, and therefore vanishes when contracted with the TT polarisation tensors e_{ij}^A. This means that actual leading-order contribution to the GW signal is given by

$$\mathcal{I}_\pm^i(f, \hat{\boldsymbol{r}}_{\mathrm{c}}) \simeq \ddot{X}_\pm^i \int_{-\ell/2}^{+\ell/2} \mathrm{d}\sigma_\pm \, \sigma_\pm \exp\left(-\frac{\pi \mathrm{i}}{6} f |\ddot{\boldsymbol{X}}_\pm|^2 \sigma_\pm^3 \right) \simeq \frac{\ddot{X}_\pm^i}{|\ddot{\boldsymbol{X}}_\pm|^{4/3}} (\pi f/6)^{-2/3} \int \mathrm{d}u \, u \, \mathrm{e}^{-\mathrm{i}u^3}, \tag{2.30}$$

where we have rewritten the integral in terms of the dimensionless variable $u \equiv \sigma_\pm (\pi f |\ddot{\boldsymbol{X}}_\pm|^2/6)^{1/3}$. We have assumed here that the GW frequency is positive; for negative frequencies, we can replace f with $|f|$ and take the complex conjugate of the integral.

We can estimate $|\ddot{\boldsymbol{X}}_\pm|$ by noting that generic solutions for $\boldsymbol{X}_\pm$ can be written as a sum of Fourier modes, $\boldsymbol{X}_\pm = \sum_{n=1}^\infty \boldsymbol{X}_\pm^{(n)} \exp(2\pi \mathrm{i} n \sigma_\pm/\ell)$. Consider first the unrealistic case of a solution with a single mode, n. Since $|\dot{\boldsymbol{X}}_\pm| = 1$, we would then have $|\ddot{\boldsymbol{X}}_\pm| = 2\pi n/\ell$. For a more realistic solution, there are cross-terms from various different modes, but in general we can write $|\ddot{\boldsymbol{X}}_\pm| = 2\pi \bar{n}_\pm/\ell$, where the 'effective mode number' $\bar{n}_\pm$ is of order unity for smooth strings, and becomes larger for very wiggly strings. (One generally expects gravitational backreaction to dampen higher-order modes, which would dynamically drive $\bar{n}_\pm$ towards smaller values over time.) We therefore write $u = \sigma_\pm \pi (\bar{n}_\pm/\ell)^{2/3} (2f/3)^{1/3}$.

Note that we neglected the limits of the u integral in equation (2.30). These are given by $|u| = \pi \bar{n}_\pm^{2/3} (f\ell/12)^{1/3}$, and are thus much greater than unity, since we are interested in the high-frequency regime $f\ell \gg 1$. As such, we can safely extend the limits of the integral to infinity, since the final value is dominated by the region $|u| \lesssim 1$ anyway. This allows us to evaluate the integral analytically, giving

$$\int_{-\infty}^{+\infty} \mathrm{d}u\, u\, \mathrm{e}^{-\mathrm{i}u^3} = -\frac{2\pi \mathrm{i}}{3\Gamma(^1/_3)}, \tag{2.31}$$

where $\Gamma(z)$ is the Euler Gamma function.

Going back to equation (2.24), we therefore find that the cusp waveform is given at high frequencies $|f|\ell \gg 1$ by[7]

$$\bar{h}_\mathrm{c}(f, \hat{\boldsymbol{r}}_\mathrm{c}) \simeq -A_\mathrm{c} \frac{G\mu \ell^{2/3}}{r|f|^{4/3}} (e_{ij}^+ - \mathrm{i}e_{ij}^\times) \hat{t}_+^i \hat{t}_-^j, \qquad A_\mathrm{c} \equiv \frac{4 \times (2/3)^{2/3}}{(\bar{n}_+ \bar{n}_-)^{1/3} \Gamma^2(^1/_3)} \lesssim 0.4253, \tag{2.32}$$

where $\hat{\boldsymbol{\tau}}_\pm \equiv \ddot{\boldsymbol{X}}_\pm/|\ddot{\boldsymbol{X}}_\pm|$ are two unit vectors transverse to the GW propagation direction. Each of these is defined by some angle $\theta_\pm$ in the transverse plane, and it is straightforward to show that

$$(e_{ij}^+ - \mathrm{i}e_{ij}^\times) \hat{t}_+^i \hat{t}_-^j = \mathrm{e}^{-2\mathrm{i}(\theta_+ + \theta_-)}, \tag{2.33}$$

so that this factor can be absorbed into our choice of coordinates by performing a polarisation rotation, c.f. equation (0.26). The fact that we can do this indicates that the cusp's GW emission is *linearly polarised*.

Now that we know the GW signal emitted along the direction of the cusp, $\hat{\boldsymbol{r}}_\mathrm{c}$, we can ask how this emission is modified if there is some nonzero angle θ between this direction and the observer's line of sight $\hat{\boldsymbol{r}}$. If θ is too large, then the phase in $\mathcal{I}_\pm^i$ will no longer be stationary, and the GW emission will be exponentially suppressed at high frequencies. The question then is, how small does this angle have to be for the observer to still see the much slower $f^{-4/3}$ power law decay? We can obtain a rough answer to this question by rewriting the phase (2.28) with a small inclination angle $\theta \equiv \cos^{-1}(\hat{\boldsymbol{r}}_\mathrm{c} \cdot \hat{\boldsymbol{r}})$,

$$\sigma_\pm - \hat{\boldsymbol{r}} \cdot \boldsymbol{X}_\pm = \frac{1}{2}\theta^2 \sigma_\pm - \frac{1}{2}\boldsymbol{\delta} \cdot \ddot{\boldsymbol{X}}_\pm \sigma_\pm^2 + \frac{1}{6}|\ddot{\boldsymbol{X}}_\pm|^2 \sigma_\pm^3 + \cdots, \tag{2.34}$$

---

[7]Note that the value of A_c given here is a factor of two smaller than the value of 0.8507 quoted by Damour and Vilenkin [227]. This is because we work in terms of the complex strain h instead of the metric perturbation h_{ij}, with the missing factor of two being carried by the normalisation of the polarisation tensors e_{ij}^A.

where $\boldsymbol{\delta} \equiv \hat{\boldsymbol{r}} - \hat{\boldsymbol{r}}_c$ is a 3-vector pointing from the cusp direction to the line of sight, with magnitude $|\boldsymbol{\delta}| = \sqrt{2 - 2\cos\theta} \simeq \theta$. The first two terms here are the ones that cause the rapid oscillations that spoil the power-law frequency scaling. We can therefore estimate the angle θ at which this power-law scaling fails by setting the first and third terms equal to each other, and setting $\sigma_{\pm} = (\pi f |\ddot{\boldsymbol{X}}_{\pm}|^2/6)^{-1/3}$ such that the dummy variable u is equal to unity (since we saw that the integral is dominated by $|u| \lesssim 1$). This defines the cusp's *beaming angle*,

$$\theta_b(f) \simeq \frac{2^{2/3}}{3^{1/6}}(|f|\ell)^{-1/3}, \tag{2.35}$$

which is necessarily very small, since we have $|f|\ell \gg 1$. (Note that we have set the effective mode numbers $\bar{n}_{\pm}$ to unity here—these can increase the size of the beam like $\theta_b \sim \bar{n}_{\pm}^{1/3}$ for very wiggly strings.)

Our final expression for the cusp waveform is therefore

$$\tilde{h}_c(f, \hat{\boldsymbol{r}}) \simeq A_c \frac{G\mu\ell^{2/3}}{r|f|^{4/3}}\Theta(\hat{\boldsymbol{r}}_c \cdot \hat{\boldsymbol{r}} - \cos\theta_b)\Theta(|f| - 2/\ell), \tag{2.36}$$

where $\Theta(x)$ is the Heaviside step function. We approximate the beam with a sharp, step-function cutoff, as the gradual onset of the exponential frequency damping is somewhat difficult to calculate precisely. We also include a second step function, to remind us that this expression should not be applied below the fundamental frequency $2/\ell$. Strictly speaking we should only use equation (2.36) in the high-frequency regime, but in practice we expect that this *under*estimates the low-frequency GW emission, so we can safely use it as a conservative proxy for the GW emission across the whole frequency range (particularly as our results for the corresponding GW memory are dominated by the high-frequency emission).

Using equation (2.36), we can calculate the total GW energy radiated by the cusp. It is useful to normalise this against the total energy of the loop, $\mu\ell$, to give the dimensionless energy spectrum

$$\epsilon(f, \hat{\boldsymbol{r}}) \equiv \frac{1}{\mu\ell}\frac{\mathrm{d}E_{\mathrm{gw}}}{\mathrm{d}(\ln f)\,\mathrm{d}^2\hat{\boldsymbol{r}}} = \frac{\pi r^2 f^3}{4G\mu\ell}\left(|\tilde{h}(f, \hat{\boldsymbol{r}})|^2 + |\tilde{h}(-f, \hat{\boldsymbol{r}})|^2\right). \tag{2.37}$$

For cusps, this is given by

$$\epsilon_c(f, \hat{\boldsymbol{r}}) \simeq \frac{\pi}{2}A_c^2 G\mu(f\ell)^{1/3}\Theta(\hat{\boldsymbol{r}}_c \cdot \hat{\boldsymbol{r}} - \cos\theta_b)\Theta(f - 2/\ell), \tag{2.38}$$

which diverges at high frequencies for observers in the beaming direction, $\hat{\boldsymbol{r}} = \hat{\boldsymbol{r}}_c$. This divergence is 'hidden' by the beaming angle θ_b, which shrinks fast enough with frequency to ensure that the total radiated energy is finite. We can see this by defining the isotropically-averaged energy spectrum,

$$\bar{\epsilon}(f) \equiv \int_{S^2} \mathrm{d}^2\hat{\boldsymbol{r}}\, \epsilon(f, \hat{\boldsymbol{r}}), \tag{2.39}$$

which gives us an extra factor of $\pi\theta_b^2$, so that

$$\bar{\epsilon}_c(f) \simeq (2/3)^{1/3}(\pi A_c)^2 G\mu(f\ell)^{-1/3}\Theta(f - 2/\ell). \tag{2.40}$$

Integrating over frequency, we define the total fraction of the loop's energy that is radiated as

$$\mathcal{E} \equiv \int_0^\infty \frac{\mathrm{d}f}{f} \bar{\epsilon}(f), \tag{2.41}$$

which ends up being of order $G\mu$ for cusps,

$$\mathcal{E}_\mathrm{c} \simeq 3^{2/3}(\pi A_\mathrm{c})^2 G\mu \ll 1. \tag{2.42}$$

As we will see below, however, the high-frequency divergence in equation (2.38) can still cause problems elsewhere due to the nonlinear nature of GR. Indeed, this divergence lies at the heart of the divergent behaviour we will encounter in the nonlinear memory waveforms.

Kinks

Now let us consider the kink case. Suppose one of the left/right mode velocities $\dot{\boldsymbol{X}}_\pm$ jumps discontinuously from $\hat{\boldsymbol{n}}_1$ to $\hat{\boldsymbol{n}}_2$ at some value of $\sigma_\pm$. The oscillatory integral is then dominated by this jump, giving [227]

$$\mathcal{I}_\pm^i \simeq \frac{\mathrm{i}}{\pi f}\left(\frac{\hat{n}_2^i}{1 - \hat{\boldsymbol{n}}_2 \cdot \hat{\boldsymbol{r}}} - \frac{\hat{n}_1^i}{1 - \hat{\boldsymbol{n}}_1 \cdot \hat{\boldsymbol{r}}}\right). \tag{2.43}$$

While this expression diverges if either $\hat{\boldsymbol{n}}_1$ or $\hat{\boldsymbol{n}}_2$ points along the line of sight $\hat{\boldsymbol{r}}$, we always project onto the transverse plane by contracting with the polarisation tensors e_{ij}^A, meaning that the GW strain is always finite. Replacing one copy of the stationary-phase integral (2.30) with this kink integral, and neglecting a factor of order unity which depends on the 3-vectors $\hat{\boldsymbol{n}}_1$ and $\hat{\boldsymbol{n}}_2$, we therefore obtain the kink waveform

$$\bar{h}_\mathrm{k}(f, \hat{\boldsymbol{r}}) \simeq A_\mathrm{k} \frac{G\mu\ell^{1/3}}{r|f|^{5/3}} \Theta(\hat{\boldsymbol{r}}_\mathrm{k} \cdot \hat{\boldsymbol{r}} - \cos\theta_\mathrm{b})\Theta(|f| - 2/\ell), \tag{2.44}$$

where $A_\mathrm{k} \equiv \sqrt{A_\mathrm{c}}/\pi \lesssim 0.2076$. The corresponding energy spectrum is

$$\epsilon_\mathrm{k}(f, \hat{\boldsymbol{r}}) \simeq \frac{\pi}{2} A_\mathrm{k}^2 G\mu(f\ell)^{-1/3} \Theta(\hat{\boldsymbol{r}}_\mathrm{k} \cdot \hat{\boldsymbol{r}} - \cos\theta_\mathrm{b})\Theta(f - 2/\ell), \tag{2.45}$$

which, unlike the cusp spectrum, converges at high frequencies, even when the beam points exactly along the line of sight, $\hat{\boldsymbol{r}} = \hat{\boldsymbol{r}}_\mathrm{k}$.

We see that the kink waveform is remarkably similar to the cusp waveform, in that it is also linearly polarised and is also characterised by a power-law falloff in frequency, although the $\sim f^{-5/3}$ falloff here is slightly steeper than the $\sim f^{-4/3}$ we found for cusps. The beaming angle θ_b is the same as before, as it is still set by the range of directions $\hat{\boldsymbol{r}}$ in which the stationary-phase approximation holds.

One very important difference is that kinks are *persistent* features of cosmic string loops, as opposed to cusps, which are *transient*; this is because cusps are associated with a single value of both σ_+ and σ_- at which $\boldsymbol{X}_+ = \boldsymbol{X}_-$, which determines a fixed time $t = (\sigma_+ + \sigma_-)/2$, while kinks have only one of the two coordinates $\sigma_\pm$ fixed, with the corresponding locus in (t, σ) describing the propagation of the kink around the loop over time. As a result, the cusp waveform is associated with a single beaming direction

$\hat{r}_c$, while the kink waveform is associated with a one-dimensional 'fan' of directions defined by the kink's propagation direction as it travels around the loop. The unit vector $\hat{r}_k$ that appears in equation (2.44) is therefore defined as the direction in this 'fan' which passes closest to the observer's line of sight $\hat{r}$. Here we assume that the length of the curve traced out by this fan on the unit sphere is comparable to the sphere's circumference, $\sim 2\pi$. As a result, integrating equation (2.45) over the sphere picks up a factor of $4\pi\theta_b$ rather than $\pi\theta_b^2$, giving

$$\bar{\epsilon}_k(f) \simeq \frac{2^{5/3}}{3^{1/6}}(\pi A_k)^2 G\mu(f\ell)^{-2/3}\Theta(f-2/\ell), \qquad \mathcal{E}_k \simeq 3^{5/6}(\pi A_k)^2 G\mu. \tag{2.46}$$

Note that we could repeat this calculation to find the GWs radiated by milder discontinuities in the loop, due to jumps in higher-order derivatives of $X_\pm$. However, for each order of the derivative we expect the waveform to be reduced by a factor of $\sim 1/(|f|\ell) \ll 1$, so that these features are less important than kinks in the high-frequency regime we are interested in.

KINK-KINK COLLISIONS

Finally, the waveform for a kink-kink collision is given by setting both the integrals $\mathcal{I}_\pm^i$ equal to equation (2.43). This gives

$$\tilde{h}_{kk}(f,\hat{r}) \simeq \frac{G\mu}{r\pi^2 f^2}\Theta(|f|-2/\ell). \tag{2.47}$$

Since the left-moving kink is associated with a fixed value of σ_+, and the right-moving kink with a fixed value of σ_-, their collision occurs at a single point (σ_+, σ_-) and is therefore *transient* like a cusp, in contrast with the persistent GW emission from each kink individually. As mentioned previously, the resulting emission is perfectly isotropic, as there is no stationary-phase requirement and therefore no beaming direction. As a result, kink-kink collisions produce no nonlinear memory, since the angular integral in equation (0.47) has no TT part for this isotropic emission. For this reason, kink-kink collisions are unimportant for much of the rest of this chapter.

For completeness, we include the energy spectrum associated with the waveform (2.47),

$$\epsilon_{kk}(f,\hat{r}) = \frac{1}{4\pi}\bar{\epsilon}_{kk}(f) \simeq \frac{G\mu}{2\pi^3 f\ell}, \qquad \mathcal{E}_{kk} \simeq \frac{G\mu}{\pi^2}. \tag{2.48}$$

THE GRAVITATIONAL-WAVE BACKGROUND FROM LOOPS

We can calculate the GWB spectrum generated by cusps, kinks, and kink-kink collisions on a cosmic population of loops using the 'Phinney formula' from section 0.2.4. Of the two key ingredients that go into this formula, we have already calculated one: the energy spectra of each GW emission mechanism. The other ingredient is the comoving rate density $\mathcal{R}$, which we can factorise into the comoving number density of loops, n, and the mean GW burst rate from each loop. (Note that these are all functions of the loop length ℓ.) Since we know that each loop's oscillation period is given by $\ell/2$, we can write the latter in terms of the mean number of cusps, kinks, and kink-kink collisions per oscillation, which we denote by N_i, $i \in \{c, k, kk\}$. Determining these numbers for a realistic loop network is somewhat challenging; the number of cusps per oscillation N_c is thought to be of order unity, while the number of kinks N_k could

in principle be much larger. The number of kink-kink collisions is always $N_{kk} = N_k^2/4$, since each of the $N_k/2$ left-moving kinks will collide with all of the $N_k/2$ right-moving kinks once per oscillation period. Putting all this together, the Phinney formula (0.65) gives

$$\Omega(f) = \frac{16\pi G\mu}{3H_0^2} \sum_{i \in \{c,k,kk\}} N_i \int dt\, a(t) \int d\ell\, n(\ell, t) \bar{\epsilon}_i(f/a(t)), \qquad (2.49)$$

where we have replaced the redshift integral with an integral over cosmic time t, with $a(t)$ the FLRW scale factor.

The loop network usually evolves toward a scaling solution [113], such that the number density is given by

$$n(\ell, t) = \frac{a^3(t)}{t^4} \mathcal{F}(\gamma), \qquad (2.50)$$

where the dimensionless function $\mathcal{F}$ depends on the loop length and on cosmic time only through their dimensionless ratio $\gamma \equiv \ell/t$. There are three widely-used models for the distribution function:

1. The original 'one-scale' model of Vilenkin and Shellard [567] (see also Siemens *et al.* [514]);

2. The model of Blanco-Pillado *et al.* [130], which is calibrated to numerical simulations;

3. The model of Ringeval *et al.* [483] (see also Lorenz *et al.* [396]), which is calibrated to a different set of simulations, and includes additional modelling of the effects of backreaction on the loops.

Following Auclair *et al.* [87], we refer to these as 'model 1', 'model 2', and 'model 3' respectively. Model 1 is widely considered obsolete, as it is incompatible with both of the main sets of Nambu-Goto network simulations [130, 483]; however, we include it here for completeness.

2.2 Nonlinear memory waveforms

Now that we have derived the standard cusp and kink waveforms, equations (2.36) and (2.44), we are almost ready to calculate the nonlinear memory associated with each of these. Before we start, we develop here some useful expressions—based on the nonlinear memory formula in equation (0.47)—which will allow us to carry out these calculations entirely in the frequency domain.

We work in terms of the complex GW strain $h(t, \boldsymbol{r}) = h_+ - i h_\times$ throughout. We distinguish between the primary, oscillatory strain signal (sourced at linear order) and the additional strain due to the nonlinear memory effect by writing these as $h^{(0)}$ and $h^{(1)}$ respectively, reserving $h^{(n)}$ with $n \geq 2$ for the 'memory of the memory' and other higher-order memory contributions, which we discuss in sections 2.3.7 and 2.4.6. The leading nonlinear memory correction term can be written in gauge-invariant form as [265, 542]

$$h^{(1)}(t, \boldsymbol{r}) = \frac{2G}{r} \int_{-\infty}^{t} dt' \int_{S^2} d^2\hat{\boldsymbol{r}}' \frac{(e^{+,ij} - i e^{\times,ij})\hat{r}_i'\hat{r}_j'}{1 - \hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}'} \frac{dE_{gw}^{(0)}}{dt'\, d^2\hat{\boldsymbol{r}}'}, \qquad (2.51)$$

which reduces to equation (0.47) in the limit $t \to +\infty$. Since the integral is over the (non-negative) GW energy flux from the source, $dE_{gw}^{(0)}/dt\, d^2\hat{\boldsymbol{r}}$, we generically obtain a nonzero memory correction which grows monotonically with time while the source is 'on'. By projecting onto the GW polarisation tensors

we ensure that only the TT part of the angular integral contributes; this TT part vanishes if the emitted flux is exactly isotropic, but is generically non-vanishing for anisotropic emission.

Note that equation (2.51) corresponds to just one of many quadratic terms that appear on the right-hand side of the relaxed Einstein field equation [263, 449]. While this term has received particular attention in the literature due to its pleasing intuitive interpretation as the nonlinear gravitational counterpart to the linear GW memory effect, the other quadratic terms will also give nonlinear corrections to the GWs emitted by cosmic strings. Our results here therefore represent just one particular nonlinear contribution to these GW signals. Since we argue below that nonlinear effects are important near cusps on large cosmic string loops, it would be interesting to explore these additional contributions further in future work. However, we note that there are two features of the nonlinear memory term that make it particularly important here, and justify our focus on this one term. The first is that it is *hereditary*, i.e., it depends on the entire history of the source; as we will see, this has the effect of transferring power from very high frequencies down to the frequencies we can observe. The second is that it is *radiative*, i.e., it determines the GW signal measured by distant observers, rather than just affecting the spacetime near to the source (indeed, Christodoulou's original expression for the nonlinear memory effect [201] was derived entirely in terms of gauge-invariant quantities at null infinity, and can be very cleanly identified with equation (2.51) above).

Returning to equation (2.51), we can insert the expressions for the polarisation tensors from equation (0.19) to write

$$(e^{+,ij} - \mathrm{i}e^{\times,ij})\hat{r}'_i\hat{r}'_j = [(\hat{\boldsymbol{\theta}} - \mathrm{i}\hat{\boldsymbol{\phi}}) \cdot \hat{\boldsymbol{r}}']^2. \tag{2.52}$$

We can also rewrite the energy flux (2.51) in terms of the linear GW strain of the source using equation (0.43). The leading GW memory term then becomes

$$h^{(1)}(t, \boldsymbol{r}) = \int_{-\infty}^{t} \frac{\mathrm{d}t'}{2r} \int_{\hat{\boldsymbol{r}}'} |r\dot{h}^{(0)}(t', \boldsymbol{r}')|^2, \tag{2.53}$$

where we have written the angular integral as

$$\int_{\hat{\boldsymbol{r}}'} [\cdots] \equiv \int_{S^2} \frac{\mathrm{d}^2\hat{\boldsymbol{r}}'}{4\pi} \frac{[(\hat{\boldsymbol{\theta}} - \mathrm{i}\hat{\boldsymbol{\phi}}) \cdot \hat{\boldsymbol{r}}']^2}{1 - \hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}'} [\cdots], \tag{2.54}$$

for brevity. Throughout we refer to the oscillatory strain $h^{(0)}$ on the right-hand side that sources the memory as the 'primary' GW emission.

2.2.1 Late-time memory

One drawback of equation (2.53) is that it is written in terms of the time-domain primary strain, whereas the cosmic string waveforms that we want to consider are much more naturally expressed in the frequency domain. This problem disappears when we consider the *late-time memory*, i.e. the total strain offset caused by the cusp,

$$\Delta h^{(1)} \equiv \lim_{t \to \infty} h^{(1)}(t), \tag{2.55}$$

as the time integral in equation (2.53) is then over the entire real line, and we can thus use Parseval's theorem to write

$$\Delta h^{(1)}(\boldsymbol{r}) = \frac{2\pi^2}{r} \int_{\mathbb{R}} \mathrm{d}f \int_{\hat{\boldsymbol{r}}'} |rf\bar{h}^{(0)}(f, \boldsymbol{r}')|^2, \qquad (2.56)$$

where $\bar{h}^{(0)}$ is the Fourier transform of the primary strain signal, and the extra factor of frequency comes from the time derivative on the strain.

2.2.2 FREQUENCY-DOMAIN MEMORY WAVEFORMS

The late-time memory (2.56) is useful for giving a sense of the total size of the memory effect, but in many cases is not directly observable. For example, the test masses in ground-based GW interferometers like LIGO/Virgo are not freely-falling in the plane of the interferometer arms; they are acted upon by feedback control systems at low frequencies to mitigate seismic noise [500]. These low-frequency forces mean that the test masses cannot sustain a permanent displacement after the GW has passed. However, the 'ramping up' of the memory signal from zero at early times to $\Delta h^{(1)}$ at late times can be measured if it contains power in the sensitive frequency band of the interferometer.

We are therefore interested in calculating the full memory signal in the frequency domain. The simplest way of doing this is to use the result[8]

$$\mathcal{F}\left[\int_{-\infty}^{t} \mathrm{d}t' \, g(t')\right] = \frac{1}{2}\tilde{g}(0)\delta(f) - \frac{\mathrm{i}}{2\pi f}\tilde{g}(f) \qquad (2.57)$$

for a general function $g(t)$, where $\mathcal{F}[\cdots]$ denotes the Fourier transform, and $\tilde{g} = \mathcal{F}[g]$. Applying this to equation (2.53), we obtain

$$\bar{h}^{(1)}(f) = \frac{1}{2}\Delta h^{(1)}\delta(f) - \frac{\mathrm{i}}{2\pi f} \int_{\mathbb{R}} \frac{\mathrm{d}t}{2r} \int_{\hat{\boldsymbol{r}}'} \mathrm{e}^{-2\pi\mathrm{i}ft} |r\dot{h}^{(0)}|^2. \qquad (2.58)$$

We can neglect the term proportional to $\delta(f)$, as this only contributes at $f = 0$, and we are interested here in frequencies which are accessible to GW experiments (this zero-frequency term is still captured in equation (2.56)). By replacing each factor of $\dot{h}^{(0)}$ with its Fourier transform and massaging the resulting expression, we find

$$\bar{h}^{(1)}(f) = -\frac{\mathrm{i}\pi}{rf} \int_{\mathbb{R}} \mathrm{d}f' \int_{\hat{\boldsymbol{r}}'} f'(f' - f)r^2\bar{h}^{(0)}(f', \hat{\boldsymbol{r}}')\bar{h}^{(0)*}(f' - f, \hat{\boldsymbol{r}}'). \qquad (2.59)$$

This is the simplest way of calculating the frequency-domain memory using only the primary frequency-domain signal $\bar{h}^{(0)}$.

2.3 MEMORY FROM CUSPS

We are now ready to calculate the nonlinear memory from cusps, inserting the primary waveform (2.36) into equations (2.56) and (2.59) to obtain the late-time memory and frequency-domain waveform,

---

[8]One can show this by noting that $\int_{-\infty}^{t} \mathrm{d}t' \, g(t')$ is just the convolution of $g(t)$ with the Heaviside step function $\Theta(t)$. Since $\mathcal{F}[\Theta] = (1/2)\delta(f) - \mathrm{i}/(2\pi f)$, the result follows from the convolution theorem, $\mathcal{F}[\Theta * g] = \mathcal{F}[\Theta]\mathcal{F}[g]$.

and then iterating this process to obtain higher-order memory corrections. (This iteration process is conceptually similar to the procedure described by Khera *et al.* [370], Talbot *et al.* [537] to incorporate higher-order memory effects in BBH waveform models.)

2.3.1 BEAMING EFFECTS

The anisotropic beaming of the GWs from cusps is what gives rise to a nonzero memory effect (due to its nonzero TT projection), and is captured in the angular integral $\int_{\hat{r}'} \Theta(\hat{r}_c \cdot \hat{r}' - \cos\theta_b)$, where we are using the shorthand (2.54). To compute this integral, it is convenient to define polar coordinates $\hat{r}' = (\theta', \phi')$ such that the North pole $\theta' = 0$ coincides with the centre of the beam $\hat{r}_c$. The integrand then only has support for $\theta' \in [0, \theta_b]$, so that we obtain

$$
\int_{\hat{r}'} \Theta(\hat{r}_c \cdot \hat{r}' - \cos\theta_b) = \int_0^{\theta_b} d\theta' \frac{\sin\theta'}{2} \int_0^{2\pi} \frac{d\phi'}{2\pi} \frac{[(\hat{\theta} - i\hat{\phi}) \cdot \hat{r}']^2}{1 - \hat{r} \cdot \hat{r}'}
$$
$$
= \frac{1 + \cos I}{1 - \cos I} \left[\frac{1}{2}(1 - \cos\theta_b)\cos\theta_b - \frac{1}{4}\cos I \sin^2\theta_b \right],
$$
(2.60)

where $I \equiv \cos^{-1}\hat{r}_c \cdot \hat{r}$ is the inclination of the beam to the observer's line of sight. In the high-frequency regime where the primary cusp and kink waveforms are valid, the beam angle is very small, so we expand equation (2.60) to leading order in θ_b to obtain

$$
\int_{\hat{r}'} \Theta(\hat{r}_c \cdot \hat{r}' - \cos\theta_b) \simeq \frac{\theta_b^2}{4}(1 + \cos I).
$$
(2.61)

There are a few remarks worth making about equation (2.61). First, we note that it assumes $I \gg \theta_b$; when instead the inclination is much smaller than the beaming angle, $I \ll \theta_b$, the integral (2.60) drops to zero, as the TT part of the angular emission vanishes when the beam is aligned with the line of sight. The fact that the result (2.61) is purely real, despite the integrand being complex, shows that the memory strain is linearly polarised, much like the primary cusp and kink waveforms. Geometrically, the $\theta_b^2/4$ factor represents the fraction of the sphere taken up by the beam, while the $(1 + \cos I)$ factor shows how the strength of the memory effect varies with inclination. In particular, we notice that the memory strain is nonzero when the observer lies outside of the beam, $I > \theta_b$. In fact, the observed memory strain vanishes only when the beam is face-on ($I \ll \theta_b$) or face-off ($I = \pi$), as in both cases the angular pattern of the primary GW emission is isotropic around the line of sight.

It is interesting to note that this angular pattern—a broad $\sim (1 + \cos I)$ distribution, except at very small inclinations where the memory signal drops to zero—is exactly the same as that of the *linear* GW memory generated by the ejection of an ultrarelativistic 'blob' of matter from a massive object along a fixed axis [503]. Intuitively, this makes complete sense: the setup here is essentially the same, except that the 'blob' is replaced by a burst of GWs.
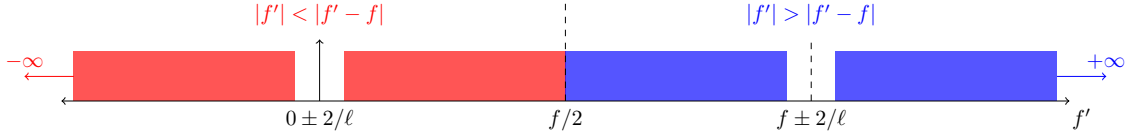
Figure 2.3: Schematic illustration of the different contributions to $\bar{h}_c^{(1)}(f)$ from the integral over f' in equation (2.59). By introducing a dimensionless dummy variable $u \equiv f'/|f|$ we obtain the two integrals shown in equation (2.63), one corresponding to the finite interior region $2/\ell < f' < f - 2/\ell$, and the other corresponding to the two semi-infinite exterior regions $f' < -2/\ell$ and $f' > f + 2/\ell$.

2.3.2 Late-time memory

Now that we have computed the angular integral (2.60), it is straightforward to obtain the late-time memory from the cusp. Inserting equations (2.36) and (2.35) into equation (2.56) and integrating over frequency, we find

$$\Delta h_c^{(1)} = 2 \times 3^{2/3}(\pi A_c G \mu)^2 (1 + \cos I)\ell/r. \tag{2.62}$$

Inserting the numerical factors, this has a maximum value of $\Delta h_c^{(1)} \approx 14.86 \times (G\mu)^2 \ell/r$ for nearly-face-on cusps $I \gtrsim 0$, and smoothly tapers to zero for face-off cusps $I = \pi$. We note that equation (2.62) should be taken with a pinch of salt, as it is sensitive to the low-frequency regime where the primary waveform is less accurate; nonetheless, we expect this to give a reasonable estimate of the magnitude of the memory effect.

2.3.3 Full waveform

We now calculate the full frequency-domain cusp memory waveform by inserting equations (2.36) and (2.35) into equation (2.59). In doing so, we must be careful to correctly account for the behaviour of the two frequency arguments f' and $f' - f$; in particular, the integrand is only nonzero when both of these arguments have magnitude greater than $2/\ell$, and the size of the beam angle θ_b must always be set by whichever of the two arguments has the greater magnitude (as this corresponds to a smaller, more restrictive beam). These different contributions are illustrated in figure 2.3 for the case where f is positive. Assuming that $|f| > 4/\ell$, we find

$$\bar{h}_c^{(1)}(f) = -\frac{\mathrm{i}\Delta h_c^{(1)}}{2^{2/3}3\pi\ell^{1/3}f|f|^{1/3}}\left[\int_{2/|f|\ell}^{\infty} \frac{\mathrm{d}u}{u^{1/3}(1+u)} - \int_{2/|f|\ell}^{1/2} \frac{\mathrm{d}u}{u^{1/3}(1-u)}\right], \tag{2.63}$$

where u is a dimensionless dummy variable. We can simplify this by taking $|f| \gg 4/\ell$, as this is the regime where the primary waveforms are valid; in this high-frequency limit, both integrals can be evaluated analytically. Inserting equation (2.62), the final result is

$$\bar{h}_c^{(1)}(f) \simeq -\mathrm{i}B_c \frac{(G\mu)^2\ell^{2/3}}{rf|f|^{1/3}}(1 + \cos I)\Theta(|f| - 2/\ell), \tag{2.64}$$

with a numerical prefactor,

$$B_c \equiv \pi A_c^2 (3/2)^{2/3}\left[\frac{4\pi}{3\sqrt{3}} - {}_2F_1\left(\tfrac{2}{3}, \tfrac{2}{3}; \tfrac{5}{3}; -1\right)\right] \approx 1.191, \tag{2.65}$$
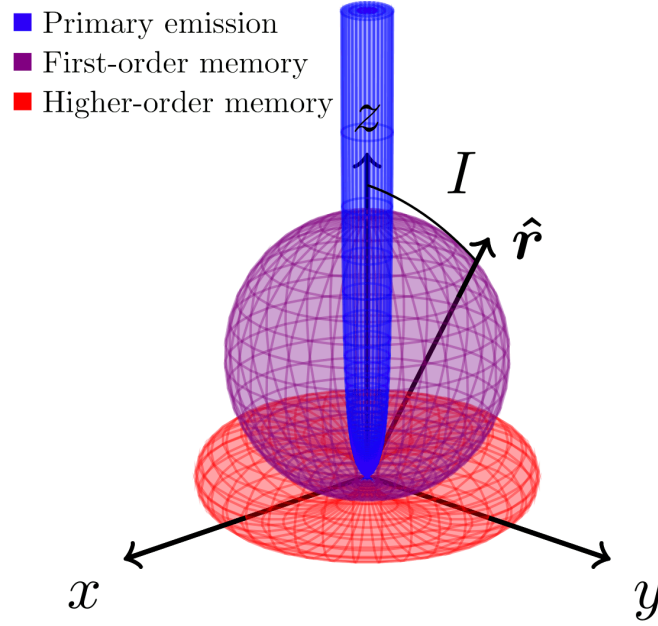
Figure 2.4: A cartoon illustration of the angular distribution of the energy radiated by a cusp. The primary emission (blue) is concentrated in a narrow beam of width $\theta_b \sim (f\ell)^{-1/3}$, while the first-order memory emission (violet) is proportional to $(1 + \cos I)^2$, and the second-order memory emission (red) is proportional to $\sin^4 I$.

where $_2F_1$ is a hypergeometric function. While we assumed $|f| > 4/\ell$ in order to obtain equation (2.63), we have checked that our final expression (2.64) underestimates the true memory signal in the region $2/\ell < |f| < 4/\ell$ (assuming the primary signal is accurate at these frequencies, which we expect to be the case to within an order of magnitude if one ignores the low-frequency motion not associated with the cusp), so we can safely leave the low-frequency cutoff at $2/\ell$ as in the primary waveform. The simple expression (2.64) thus gives a conservative but generally accurate model of the time-varying part of the cusp memory waveform at all frequencies greater that the fundamental mode of the loop, while the zero-frequency offset is described by equation (2.62).

Comparing equation (2.64) with the primary cusp waveform (2.36), we see that they are remarkably similar to each other, with both being given by the same simple frequency power law $\sim f^{-4/3}$ and the same dependence on the loop length ℓ (the latter being required by dimensional arguments). There are, however, some important differences:

1. The memory GW emission has broad support on the sphere, while the primary waveform only has support inside a narrow beam (see figure 2.4).

2. The memory waveform is suppressed by an additional power of $G\mu$ (this makes intuitive sense, given that it is a nonlinear effect sourced by the primary GW emission).

3. The numerical constant in front of the memory waveform, B_c, is an order of magnitude larger than that in front of the primary waveform, A_c.

The first point is particularly crucial, as it lies at the heart of the divergent behaviour that we investigate in sections 2.3.6 and 2.3.7.

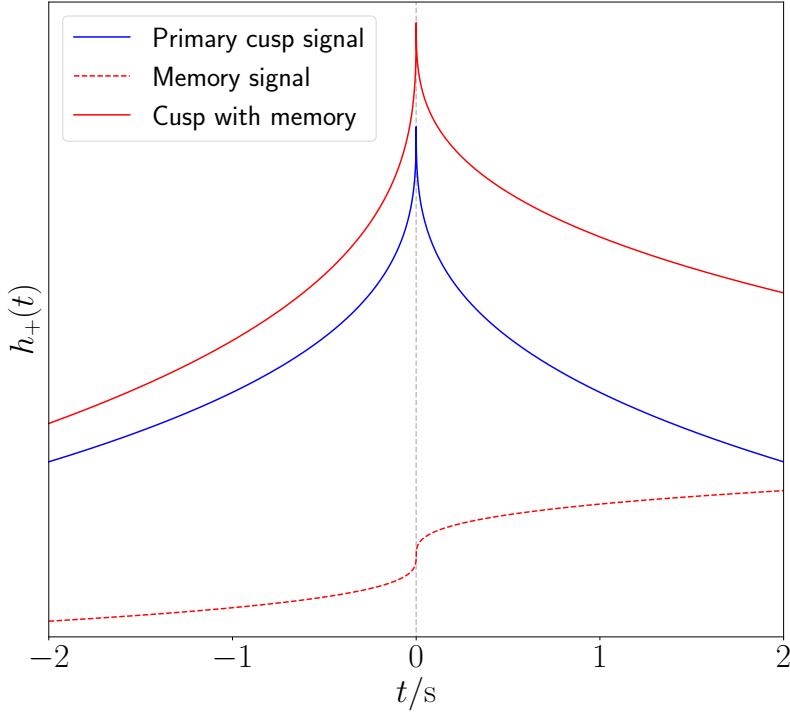2 Nonlinear gravitational-wave memory from cosmic strings

Figure 2.5: The time-domain GW strain $h_c(t)$ from a cusp, with and without the leading-order memory contribution (2.66). The memory is exaggerated by a factor of $\sim 1/G\mu$ here to make it visible. For small inclinations $I < \theta_b$ the observer lies within the cusp's beam, and sees the memory superimposed on the primary cusp signal (solid red line). If the inclination is very small, $I \ll \theta_b$, then the memory vanishes and only the primary cusp signal is observable (solid blue line). In most cases however, the observer lies outside of the beam, and only the memory is observable (red dashed line). Note that the higher-order memory contributions (order $n \geq 2$) are *not* shown here; these would look like step functions in the time domain, with height that either diverges rapidly with n ('large' loops, $\ell \gtrsim \delta/(G\mu)^3$) or converges so rapidly that the contribution to the total signal is negligible ('small' loops, $\ell \lesssim \delta/(G\mu)^3$).

2.3.4 TIME-DOMAIN WAVEFORM NEAR THE ARRIVAL TIME

While we have focused on deriving the memory signal in the frequency domain, we can inverse-Fourier-transform equation (2.64) to obtain a simple closed-form expression in the time domain,

$$ h_c^{(1)}(t) - h_c^{(1)}(t_0) \simeq -2^{5/3} 3\pi B_c \frac{(G\mu)^2}{r}(1 + \cos I)(t - t_0)\left[1 - \frac{\Gamma(2/3)}{(4\pi|t - t_0|/\ell)^{2/3}}\right]. \qquad (2.66) $$

(Here we have re-introduced the time of arrival of the primary cusp signal, t_0, rather than setting it to zero.) This time-domain waveform is shown in figure 2.5. Note that this is real, which means that the signal is linearly polarised, just like the primary waveform (2.36).

It is important to note that since equation (2.64) is only valid for high frequencies $f \gg 2/\ell$, equation (2.66) must only be valid for a short duration $|t - t_0| \ll \ell$ around the arrival time. However, for typical loop sizes this 'short duration' is actually much longer than the relevant observational timescale (typically a few seconds), so the approximation in equation (2.66) may be a useful one.

2.3.5 Radiated energy

Inserting the memory waveform (2.64) into equation (2.37), we find the radiated energy spectrum

$$\epsilon_c^{(1)}(f, \hat{\boldsymbol{r}}) \simeq \frac{\pi}{2} B_c^2 (G\mu)^3 (f\ell)^{1/3} (1 + \cos I)^2 \Theta(f - 2/\ell). \tag{2.67}$$

The fact that $\tilde{h}_c^{(0)}$ is purely real while $\tilde{h}_c^{(1)}$ is purely imaginary means that there is no coherent cross-energy between the two contributions, so the total energy is just $\epsilon_c^{(0)} + \epsilon_c^{(1)}$. For observers in the beaming direction the energy spectra of the primary and memory signals scale in the exact same way with frequency, but with a much smaller coefficient for the memory signal. However, the picture changes drastically when integrating the spectra over the sphere to compute the total emission, as the primary emission is then suppressed by a factor of $\theta_b^2 \sim (f\ell)^{-2/3}$, while the memory waveform gives

$$\bar{\epsilon}_c^{(1)}(f) \simeq \frac{8}{3} (\pi B_c)^2 (G\mu)^3 (f\ell)^{1/3} \Theta(f - 2/\ell). \tag{2.68}$$

We see that the isotropic energy spectrum due to the memory emission is blue-tilted (i.e. grows with frequency), while the primary spectrum (2.40) is red-tilted. This means that the memory emission dominates at very high frequencies,

$$f > \frac{3}{16\ell} \left(\frac{A_c}{B_c G\mu} \right)^3 \approx 8 \times 10^{22}\,\mathrm{Hz} \times \left(\frac{\ell}{\mathrm{pc}} \right)^{-1} \left(\frac{G\mu}{10^{-11}} \right)^{-3}. \tag{2.69}$$

2.3.6 Ultraviolet divergence of the radiated energy

The total fraction of the loop's energy radiated by the cusp memory can be found by integrating over frequency, as in equation (2.41). For the primary signal, we saw in section 2.1.3 that this gives a small total energy fraction of order $G\mu$. For the memory signal, however, the integral diverges. This is clearly unphysical, and shows a breakdown in the validity of equation (2.64). Note however that this breakdown is *not* in the low-frequency regime where we know that equation (2.64) is inaccurate; rather, the integral has an ultraviolet divergence that goes like $\sim f^{1/3}$ at high frequencies $f \to \infty$. Instead, we can understand this divergence as a breakdown of the Nambu-Goto approximation for the loop dynamics.

In section 2.1.1, we justified the use of the Nambu-Goto action by arguing that possible curvature corrections $\hbar\kappa/\mu$ should generically be very small for macroscopic loops, typically of order $(\delta/\ell)^2$. However, as we saw in section 2.1.2, the Nambu-Goto action generically predicts the formation of cusps, where the GW strain looks locally like $h(t) \sim |t - t_0|^{1/3}$. As was already recognised by Damour and Vilenkin [227], the curvature associated with this scales as $\kappa \propto \ddot{h} \propto |t - t_0|^{-5/3}$, which diverges at the peak of the cusp, clearly ruining the validity of the Nambu-Goto approximation. This problem does not manifest itself in the primary cusp waveform, since the beaming angle θ_b decreases fast enough with frequency to ensure the total radiated energy is finite, cf. equation (2.42). However, the energy flux in the centre of the beam is still divergent, and as we have shown here, this sources further divergences due to the nonlinear nature of gravity. Since the memory effect is not beamed, there is nothing suppressing it at high frequencies.
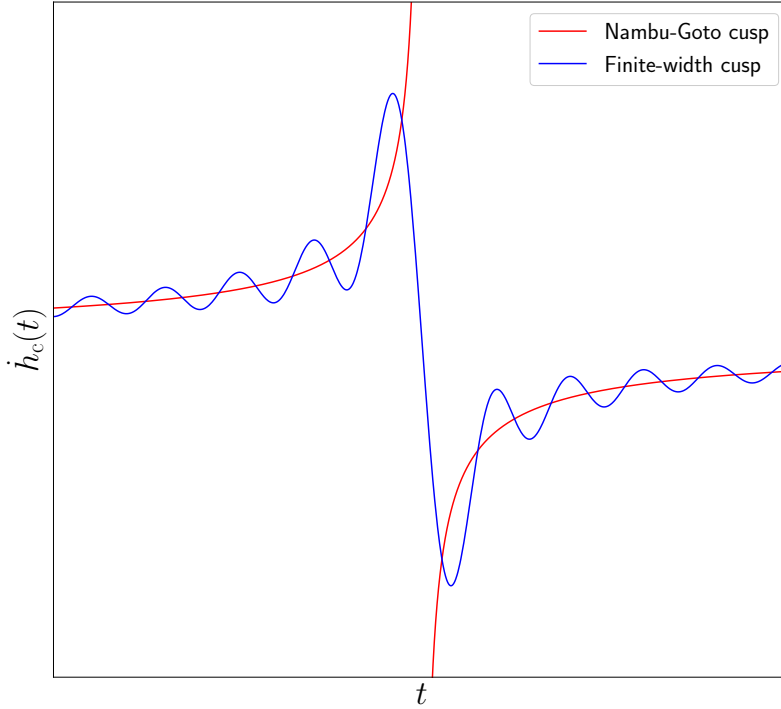
Figure 2.6: The time derivative of the cusp strain signal close to the peak, $|t - t_0| \ll \ell$, for an observer in the beaming direction, $I = 0$. This diverges for the standard Nambu-Goto waveform (2.36), ruining the validity of the Nambu-Goto approximation, and causing a divergence in the energy radiated by the first-order memory. By introducing a frequency cutoff due to the finite string width, $f < 1/\delta$, we see that the derivative becomes finite and continuous, and the first-order memory divergence is regularised.

The simplest way to regularise this divergence is to impose an ultraviolet cutoff at some high energy scale, for which the only obvious candidate is the string width scale. By truncating the frequency-domain cusp waveform at $f \sim 1/\delta$, the cusp is smoothed out on timescales $|t - t_0| \sim \delta$, and the curvature reaches a finite maximum value that scales like $\kappa \propto \delta^{-5/3}$. This smoothing is shown explicitly in figure 2.6, where we see that the time derivative of the strain diverges in the Nambu-Goto case, but is finite and continuous if a finite width is introduced. (In this heuristic setup, it is not immediately clear whether or not the smoothing afforded by a finite string width is strong enough to prevent higher-curvature terms in the action from becoming important near the cusp; we return to this point later.)

We therefore consider only GWs with frequency $f < 1/\delta$. This has no impact on the primary waveform, since $1/\delta \approx 10^{38}$ Hz $\times \left(G\mu/10^{-11}\right)^{1/2}$, which is beyond the reach of any current or planned GW experiments. However, the cutoff *does* impact the GW memory. For example, setting an upper limit of $f = 1/\delta$ in the integral (2.41), we find a finite value for the energy radiated by the cusp memory,

$$\mathcal{E}_c^{(1)} \simeq 8(\pi B_c)^2 (G\mu)^3 (\ell/\delta)^{1/3} \approx 2 \times 10^{-16} \times \left(\frac{G\mu}{10^{-11}}\right)^{19/6} \left(\frac{\ell}{\text{pc}}\right)^{1/3}, \qquad (2.70)$$

where we have used $\delta \approx \ell_{\text{Pl}}/\sqrt{G\mu}$ in the second expression. This shows that the energy radiated due to the first-order memory effect is smaller than that from the primary emission for observationally-allowed values of the string tension, $G\mu \lesssim 10^{-11}$. However, the factor of $(\ell/\delta)^{1/3} \gg 1$ in equation (2.70) is

concerning. Based on equation (2.70), cusps on GUT-scale strings ($G\mu = 10^{-6}$) would radiate far more energy through the memory effect than through the primary GWs; for $\ell \gtrsim 0.4\,\mathrm{pc}$ they would radiate more than the entire energy of the loop. Such a situation would be unphysical, and would indicate the breakdown of the validity of the primary cusp waveform. One might argue that this is not an issue, since GUT-scale Nambu-Goto strings are already ruled out by observations; however, we show below that accounting for higher-order memory effects exacerbates the problem, leading to similar unphysical results even for much lower string tensions.

We should note that the cutoff imposed here is rather ad-hoc, and is done in a way that is agnostic to the underlying microphysics of the string. In reality, the divergence will be resolved in a way which may depend on the microphysics, and the GW memory observables may be sensitive to this; indeed, this is implied by the fact that equation (2.70) depends directly on the string width δ.

The fact that equation (2.70) depends on the loop length ℓ distinguishes it from the equivalent expression for the primary signal, equation (2.42), in which we found that a fixed fraction of the loop's energy (roughly $\sim G\mu$) was radiated by the cusp regardless of the loop length. The reason for the difference here is that we have introduced a new lengthscale to the problem, the string width δ, so on dimensional grounds we can expect powers of ℓ/δ to appear. The reason we have a *positive* power of ℓ/δ here is that, as we can see from equation (2.68), the energy radiated by the cusp memory at any fixed frequency f is larger for larger loops, scaling like $\propto \ell^{1/3}$. In equation (2.70) we are essentially calculating the energy radiated at the string-width frequency $1/\delta$, and thus end up with a factor of $(\ell/\delta)^{1/3}$. This is a foretaste of one of our key findings in this chapter, which is that the cusp memory diverges only for loops greater than some critical length.

2.3.7 SECOND-ORDER MEMORY

The memory waveform (2.64) describes the spacetime curvature generated by the energy-momentum of the primary GWs from the cusp. However, these memory GWs themselves carry energy-momentum, and will in turn act as a source of their own GW memory (see reference [537] for a discussion of this effect in the context of binary black hole coalescences). We refer to this 'memory of the memory' here as the *second-order memory effect*.

The calculation of this second-order memory contribution is straightforward for cusps, simply substituting equation (2.64) into the right-hand side of equation (2.59) and following all of the same steps as before. The key difference is in the angular integral, $\int_{\hat{r}'} (1 + \cos I')^2 = \frac{1}{6} \sin^2 I$, which, due to the broader emission of the first-order memory signal, is not suppressed by a factor of θ_b^2; cf. equation (2.61). The resulting expressions for the late-time memory and frequency-domain waveform are

$$\Delta h_c^{(2)} = \frac{2(\pi B_c)^2}{3r}(G\mu)^4 \ell^{4/3} \sin^2 I \int_{2/\ell}^{\infty} \frac{\mathrm{d}f}{f^{2/3}},$$

$$\tilde{h}_c^{(2)}(f) = -\frac{\mathrm{i}\pi B_c^2}{3rf}(G\mu)^4 \ell^{4/3} \sin^2 I \left[\int_{2/\ell}^{\infty} \frac{\mathrm{d}f'}{f'^{1/3}(f' + |f|)^{1/3}} + \int_{2/\ell}^{|f|/2} \frac{\mathrm{d}f'}{f'^{1/3}(|f| - f')^{1/3}} \right], \tag{2.71}$$

both of which contain integrals which diverge due to high-frequency contributions from the first-order memory. Introducing the $f < 1/\delta$ cutoff from section 2.3.6 once again, we find to leading order in δ,

$$\Delta h_{\rm c}^{(2)} \simeq \frac{2\pi^2 \ell}{r} B_{\rm c}^2 (G\mu)^4 (\ell/\delta)^{1/3} \sin^2 I, \qquad \bar{h}_{\rm c}^{(2)}(f) \simeq -\frac{i\Delta h_{\rm c}^{(2)}}{2\pi f} \Theta(1/\delta - |f|)\Theta(|f| - 2/\ell). \quad (2.72)$$

This is interesting in that it departs from the the $\sim f^{-4/3}$ scaling of both the primary waveform (2.36) and the first-order memory waveform (2.64); the second-order memory instead has a slower decay with frequency, due to the lack of beaming in the first-order memory. In fact, for intermediate frequencies $2/\ell < |f| < 1/\delta$ the second-order memory is identical to the Fourier transform of a Heaviside step function of height $\Delta h_{\rm c}^{(2)}$. This means that on timescales much shorter than the loop oscillation period $\ell/2$ and much longer than the light-crossing time of the loop width δ, the second-order memory signal looks like a step function in the time domain; this is because the signal is dominated by high frequencies $f \lesssim 1/\delta$, and therefore 'switches on' in a very short time interval[9], $|t - t_0| \lesssim \delta$.

The waveform (2.72) can be used to calculate a second-order correction to the energy radiated by the cusp,

$$\epsilon_{\rm c}^{(2)} = \frac{\pi r^2 f^3}{2G\mu\ell}\left(|\bar{h}_{\rm c}^{(0)} + \bar{h}_{\rm c}^{(1)} + \bar{h}_{\rm c}^{(2)}|^2 - |\bar{h}_{\rm c}^{(0)} + \bar{h}_{\rm c}^{(1)}|^2\right) = \frac{\pi r^2 f^3}{2G\mu\ell}\left(|\bar{h}_{\rm c}^{(2)}|^2 + 2\bar{h}_{\rm c}^{(1)}\bar{h}_{\rm c}^{(2)*}\right)$$

$$= \left[\frac{\pi^3}{2} B_{\rm c}^4 (G\mu)^7 (\ell/\delta)^{2/3} f\ell \sin^4 I + \pi^2 B_{\rm c}^3 (G\mu)^5 (\ell/\delta)^{1/3} (f\ell)^{2/3} \sin^2 I (1 + \cos I)\right] \quad (2.73)$$

$$\times \Theta(1/\delta - |f|)\Theta(|f| - 2/\ell).$$

We see that, unlike for the first-order correction, there are now nonzero cross-terms due to $\bar{h}_{\rm c}^{(1)}$ and $\bar{h}_{\rm c}^{(2)}$ being exactly in phase with each other, resulting in two new contributions to the energy. Integrating over frequency and emission direction, the total energy is given by

$$\mathcal{E}_{\rm c}^{(2)} = \frac{16}{15}(\pi B_{\rm c})^4 (G\mu)^7 (\ell/\delta)^{5/3} + 4(\pi B_{\rm c})^3 (G\mu)^5 (\ell/\delta). \quad (2.74)$$

Comparing with the corresponding first-order result (2.70) we see a clear pattern emerging, where the second term in equation (2.74) is multiplied by a factor of $\sim \pi B_{\rm c}(G\mu)^2(\ell/\delta)^{2/3}$ compared to the first-order energy, and multiplying by the same factor again gives the first term in equation (2.74). This factor is greater than unity so long as the loop length ℓ is larger than

$$\ell_* \equiv \frac{\delta}{(\pi B_{\rm c})^{3/2}(G\mu)^3} \approx 0.7\,{\rm km} \times \left(\frac{G\mu}{10^{-11}}\right)^{-7/2}, \quad (2.75)$$

which applies to all macroscopically-large loops.

One would naively expect successive memory corrections for a generic GW source to become less and less important at higher order; the fact that they become *more* important for such a large class of cosmologically-relevant cosmic string loops is surprising, and suggests that something unphysical is

---

[9]This picture is closely related to the zero-frequency limit (ZFL) method for calculating radiation from high-energy gravitational scattering [149, 517, 552, 572], which leverages the fact that the scattering is effectively instantaneous compared to the GW frequencies of interest.

happening. Indeed, if we plug numerical values into equation (2.74), we find that loops larger than about $0.1\,\mathrm{pc} \times (G\mu/10^{-11})^{-47/10}$ have $\mathcal{E}_{\mathrm{c}}^{(2)}$ greater than unity, meaning that they radiate away more than the total energy of the loop. This represents a significant worsening of the issue we identified at the end of section 2.3.6. The problem only gets worse when we go beyond second-order corrections, as we demonstrate below.

2.3.8 HIGHER-ORDER MEMORY, AND ANOTHER DIVERGENCE

Iterating the procedure described above to calculate the third-order GW memory, it is straightforward to show that it obeys the same step-function-like relation that we found at second order,

$$\bar{h}_{\mathrm{c}}^{(3)}(f) = -\frac{\mathrm{i}\Delta h_{\mathrm{c}}^{(3)}}{2\pi f}\Theta(1/\delta - |f|)\Theta(|f| - 2/\ell), \tag{2.76}$$

with the late-time memory given by

$$\Delta h_{\mathrm{c}}^{(3)} = \frac{2\pi^2}{r}\int_{\mathbb{R}}\mathrm{d}f\int_{\hat{\mathbf{r}}'} r^2 f^2\left(|\bar{h}_{\mathrm{c}}^{(2)}|^2 + 2\bar{h}_{\mathrm{c}}^{(1)}\bar{h}_{\mathrm{c}}^{(2)*}\right), \tag{2.77}$$

where we have made sure to include both of the second-order energy contributions as source terms for the third-order memory. Upon integration, this becomes

$$\Delta h_{\mathrm{c}}^{(3)} = -\frac{4\delta}{5r}(\ell/\ell_*)^{8/3}\sin^2 I\left(1 - \frac{1}{3}\cos^2 I\right) - \frac{2\delta}{r}(\ell/\ell_*)^2\sin^2 I\left(1 + \frac{3}{5}\cos I\right), \tag{2.78}$$

where ℓ_* is the $G\mu$-dependent lengthscale defined in equation (2.75).

The third-order memory clearly scales differently for loops with $\ell \gg \ell_*$ (which we call 'large loops') compared to those with $\ell \ll \ell_*$ (which we call 'small loops'). When going to fourth order and beyond, we obtain an increasing number of cross-terms in the energy at each order (starting with $|\bar{h}_{\mathrm{c}}^{(n)}|^2$, then $2\bar{h}_{\mathrm{c}}^{(n-1)}\bar{h}_{\mathrm{c}}^{(n)*}$, $2\bar{h}_{\mathrm{c}}^{(n-2)}\bar{h}_{\mathrm{c}}^{(n)*}$, and so on, down to $2\bar{h}_{\mathrm{c}}^{(1)}\bar{h}_{\mathrm{c}}^{(n)*}$), resulting in a proliferation of terms in the resulting memory expressions, each with a different power of ℓ/ℓ_*. It is therefore much simpler to treat small and large loops separately, and focus on the leading power of ℓ/ℓ_* in each case. This results in a very economical formula for iterating the memory calculation, valid for all $n \geq 2$,

$$\Delta h_{\mathrm{c}}^{(n)} = \begin{cases} \dfrac{r}{\delta}\displaystyle\int_{\hat{\mathbf{r}}'}|\Delta h_{\mathrm{c}}^{(n-1)}|^2, & \text{for } \ell \gg \ell_* \\[2ex] \dfrac{6\pi r}{\delta^2}\displaystyle\int_{\hat{\mathbf{r}}'}\Delta h_{\mathrm{c}}^{(n-1)}|\bar{h}_{\mathrm{c}}^{(1)}(1/\delta)|, & \text{for } \ell \ll \ell_* \end{cases} \tag{2.79}$$

where we have evaluated the frequency integral in each case, leaving just the angular integral. The frequency-domain waveform is then given by the same quasi-step-function form as before,

$$\bar{h}_{\mathrm{c}}^{(n)} = -\frac{\mathrm{i}\Delta h_{\mathrm{c}}^{(n)}}{2\pi f}\Theta(1/\delta - |f|)\Theta(|f| - 2/\ell), \tag{2.80}$$

and the corresponding energy spectra are given by

$$\epsilon_{\mathrm{c}}^{(n)} = \begin{cases} \dfrac{r^2 f}{8\pi G\mu\ell}|\Delta h_{\mathrm{c}}^{(n)}|^2, & \text{for } \ell \gg \ell_* \\[2mm] \dfrac{r^2 f^2}{2G\mu\ell}\Delta h_{\mathrm{c}}^{(n)}|\bar{h}_{\mathrm{c}}^{(1)}|, & \text{for } \ell \ll \ell_* \end{cases} \tag{2.81}$$

Solving equation (2.79) iteratively with equation (2.72) as an input, we find for all $n \geq 2$,

$$\Delta h_{\mathrm{c}}^{(n)} = \begin{cases} \dfrac{\delta}{r}(\ell/\ell_*)^{2^n/3}L_n(I), & \text{for } \ell \gg \ell_* \\[2mm] \dfrac{\delta}{r}(\ell/\ell_*)^{2n/3}S_n(I), & \text{for } \ell \ll \ell_* \end{cases} \tag{2.82}$$

where $L_n(I)$ and $S_n(I)$ are polynomials in $\cos I$ which describe the angular pattern of the memory for large and small loops, respectively—these are described in detail in appendix C.2. For small loops, all memory effects are subdominant compared to the primary emission, and equation (2.82) gives a convergent geometric series. For large loops, on the other hand, the memory emission becomes stronger at each order, and equation (2.82) gives a lacunary series which diverges extremely quickly. One might hope that the polynomials $L_n(I)$ decrease in magnitude fast enough to counteract the divergence, but we find empirically in appendix C.2 that $|L_n(I)| \approx 5(2/5)^{2^{n-2}}\sin^2 I$, so the series diverges as long as $\ell \gtrsim (5/2)^3\ell_* \approx 1\,\mathrm{km} \times (G\mu/10^{-11})^{-7/2}$, as shown in figure 2.7. We discuss the cause of this divergence in appendix C.1, and argue that it is caused by a trans-Planckian GW energy flux from the cusp.

2.3.9 MEMORY FROM CUSP COLLAPSE

The results of the previous section show that, even with an ultraviolet cutoff in place at the scale of the string width, the standard cusp waveform (2.36) leads to a divergence for all 'large' loops with length $\ell \gtrsim \delta/(G\mu)^3$. The fact that the divergence appears in observable, gauge-independent quantities (the memory strain and, therefore, the radiated energy) means that something unphysical must be going on. Since the only inputs to our calculation are the cusp waveform (2.36) and the GW memory formula (2.51), at least one of these two ingredients must break down for cusps on loops of this size.

In order to track down the cause of the divergence, let us list the assumptions that go into equations (2.36) and (2.51):

1. The GW frequency is assumed to be much greater than the fundamental mode of the loop, $f \gg 2/\ell$. This is because the waveform (2.36) is derived using the universal behaviour of the loop on scales $\ll \ell$ near the cusp.

2. The loop's dynamics are assumed to follow the Nambu-Goto action (2.5) on lengthscales larger than the loop width δ. (We have imposed a cutoff that effaces scales below this, in order to regularise the first divergence we encountered in section 2.3.6.)

3. The loop is assumed to evolve according to the flat-space equations of motion (2.12)–(2.14); i.e. gravitational backreaction is assumed to be negligible.

4. The GWs generated by the loop are assumed to be well-described by linear perturbations on a flat background.

As mentioned earlier, the first assumption cannot be the source of the problem, as the divergence is associated with very high frequencies near the string width scale.

The second assumption is robust so long as (i) the higher-order curvature terms in the worldsheet Lagrangian (2.6) are negligible, and (ii) the strings are created through the breaking of a local gauge symmetry, so that the underlying field theory does not give rise to long-range interactions (i.e. we are not considering global strings, which would instead be described by the Kalb-Ramond action [567]). As mentioned in section 2.3.6, while introducing a finite string width prevents the curvature from diverging, it does not necessarily guarantee that the higher-order curvature terms are negligible. In principle the memory divergence identified here could be cured by departures from the Nambu-Goto action near the cusp. However, these departures would have to take place on scales much greater than the string width, which seems very difficult to achieve.

By elimination, it seems that the problem is mostly likely due to assumptions 3 and 4: i.e., that the flat-space description of the cusp's dynamics and GW generation is inconsistent. Indeed, we can trace the divergence back to the fact that the integrated GW energy flux diverges in the centre of the cusp's beam, $\int d(\ln f)\, \epsilon_c^{(0)}(f, \hat{\boldsymbol{r}}_c) \to \infty$, which already suggests that a flat-space description is insufficient. This divergent flux is hidden somewhat by the narrowness of the beam, $\theta_b \sim (f\ell)^{-1/3}$, which ensures that $\mathcal{E}_c^{(0)}$ is finite, but we have shown that the GW memory inherits and amplifies the divergence.

One possible resolution is that the gravitational backreaction of the loop on itself could smooth out the cusp on scales much larger than δ. However, there is large body of literature on cosmic string backreaction [107, 133, 134, 159, 183, 191, 214, 465, 539, 570, 571] which indicates that the backreaction timescale is $\sim \ell/(G\mu)$, much longer than the loop's oscillation period. While it is true that these studies all assume linearised gravity, and are thus likely to underestimate the magnitude of the effect near cusps, it is still hard to see how the loop could backreact fast enough to smooth out the cusp on relatively large scales before the peak of the signal.

All of this suggests that we need some strong-gravity mechanism which acts on a very short timescale while the cusp is forming, and suppresses the cusp's GW emission at frequencies far below the cutoff, $f \ll 1/\delta$. In section 2.5, we propose exactly such a mechanism: we argue that when cusps form on sufficiently large cosmic string loops, they source such extreme spacetime curvature that a small portion of the loop could collapse to form a black hole at a time $\sim G\mu\ell$ before the peak of the cusp emission. We will see that, remarkably, the loops for which this 'cusp collapse' process is predicted to occur are those with length $\ell \gtrsim \delta/(G\mu)^3$—exactly the same loops for which the higher-order memory divergence occurs.

It is difficult to calculate the precise GW signal associated with cusp collapse, but as we will show later in section 2.6.1, there are two main qualitative differences it introduces compared to the standard cusp waveform: (i) the Fourier transform of the primary strain signal, $\tilde{h}_c^{(0)}$, is reduced by a factor $\approx 1/2$, as it only receives contributions from the half of the signal at times before the peak; (ii) more importantly, there is a loss of power at frequencies $f \gtrsim 1/(G\mu\ell)$, due to the truncation immediately before the peak. Both of these effects influence the corresponding GW memory signal. We can account for (i) by multiplying the strain at each order in the memory expansion by the appropriate power of 1/2, and can approximate the
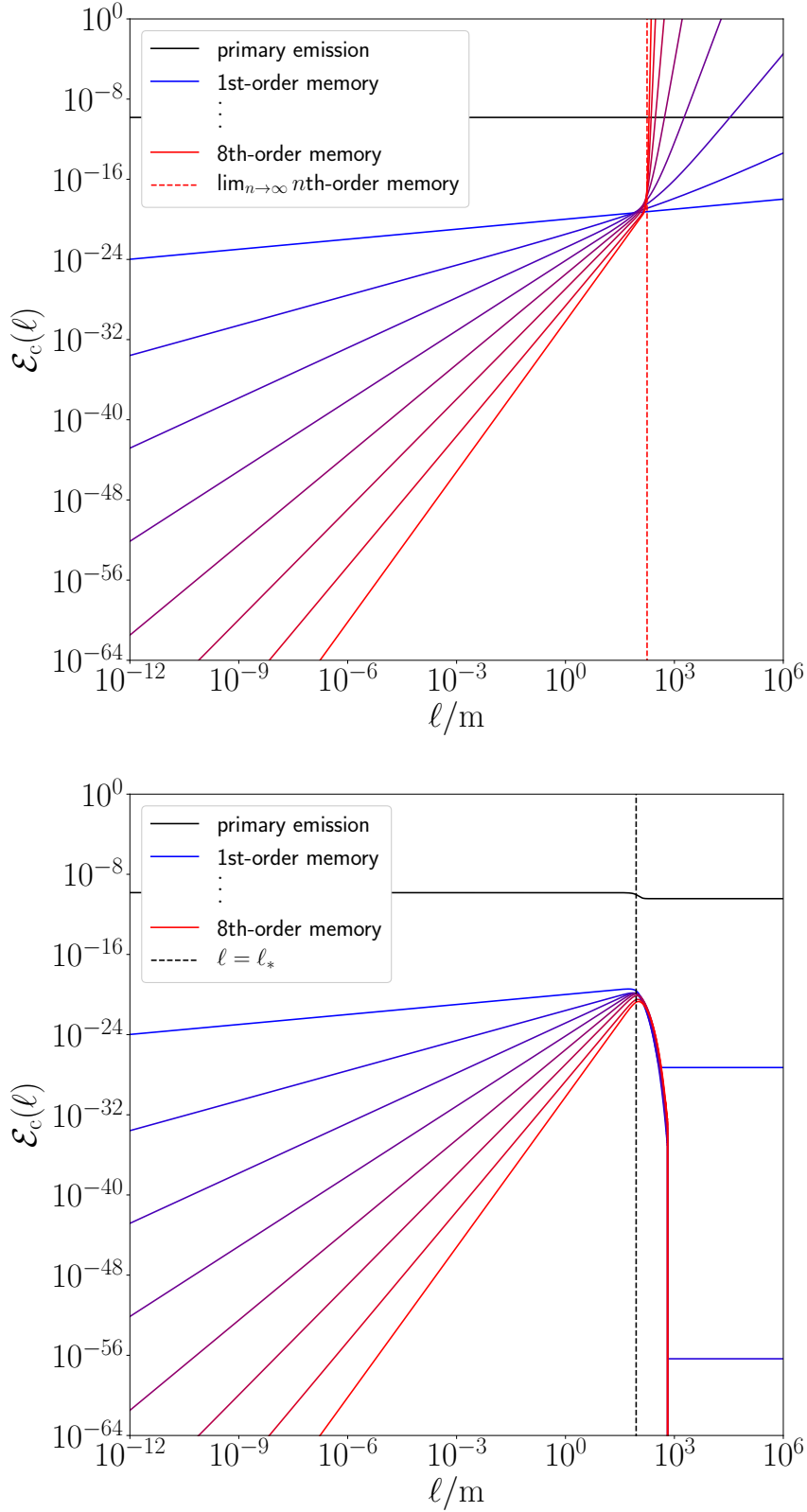
Figure 2.7: The fractional energy radiated by a cusp at different orders in the memory expansion as a function of loop length, with $G\mu = 10^{-11}$. The top panel shows the standard cusp case, clearly illustrating the divergence at $\ell \gtrsim \ell_* \approx 90\,\mathrm{m}$. The bottom panel shows the cusp collapse case, for which the radiated energy at each order drops to a small, ℓ-independent value for $\ell \gtrsim \ell_*$, curing the divergence.

effect of (ii) by introducing a sharp cutoff at frequency $f = 1/(G\mu\ell)$, which is equivalent to replacing $\delta \to G\mu\ell$ in all previous expressions. The critical lengthscale (2.75) which previously marked the onset of the divergence then becomes

$$\ell_* \to \frac{\ell}{(\pi B_c)^{3/2}(G\mu)^2} \gg \ell, \tag{2.83}$$

which means that we are always in the 'small loop' regime, $\ell \ll \ell_*$; higher-order memory corrections are suppressed by powers of $G\mu$, and the divergence is avoided completely, as shown in figure 2.7. Note that the cusp collapse process is only predicted to take place for loops with $\ell \gtrsim \delta/(G\mu)^3$, and that the GW memory from smaller loops is still described by the results given above, with ℓ_* given by equation (2.75).

More explicitly, if the divergence is indeed cured by invoking cusp collapse, then the GW observables from the cusp are as follows: the primary waveform is

$$\tilde{h}_c^{(0)} \simeq A_c \frac{G\mu\ell^{2/3}}{r|f|^{4/3}}\Theta(\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{r}}_c - \cos\theta_b)\Theta(|f| - 2/\ell)\begin{cases}(1/2)\Theta(1/G\mu\ell - |f|), & \text{for } \ell \gg \ell_* \\ \Theta(1/\delta - |f|), & \text{for } \ell \ll \ell_*\end{cases} \tag{2.84}$$

with ℓ_* given by (2.75); the first-order memory waveform is

$$\tilde{h}_c^{(1)} \simeq -\mathrm{i}B_c \frac{(G\mu)^2\ell^{2/3}}{rf|f|^{1/3}}(1 + \cos I)\Theta(|f| - 2/\ell)\begin{cases}(1/4)\Theta(1/G\mu\ell - |f|), & \text{for } \ell \gg \ell_* \\ \Theta(1/\delta - |f|), & \text{for } \ell \ll \ell_*\end{cases} \tag{2.85}$$

with the corresponding late-time memory given by

$$\Delta h_c^{(1)} \approx 2 \times 3^{2/3}(\pi A_c G\mu)^2(1 + \cos I)\frac{\ell}{r}\begin{cases}1/4, & \text{for } \ell \gg \ell_* \\ 1, & \text{for } \ell \ll \ell_*\end{cases} \tag{2.86}$$

and the nth-order memory waveforms for $n \geq 2$ are all step-function-like,

$$\tilde{h}_c^{(n)} \approx -\frac{\mathrm{i}\Delta h_c^{(n)}}{2\pi f}\Theta(|f| - 2/\ell)\begin{cases}\Theta(1/G\mu\ell - |f|), & \text{for } \ell \gg \ell_* \\ \Theta(1/\delta - |f|), & \text{for } \ell \ll \ell_*\end{cases} \tag{2.87}$$

with height given by

$$\Delta h_c^{(n)} \approx \begin{cases}\dfrac{\ell}{r}\left(\dfrac{\pi B_c}{4}\right)^{2^{n-1}}(G\mu)^{1 + \frac{2^{n+1}}{3}}L_n(I), & \text{for } \ell \gg \ell_* \\ \dfrac{\delta}{r}(\ell/\ell_*)^{2n/3}S_n(I), & \text{for } \ell \ll \ell_*\end{cases} \tag{2.88}$$

It is interesting to note that loops with length just below the cusp-collapse threshold, $\ell \lesssim \ell_*$, emit more energy through GW memory than those above the threshold. We can see this already in the primary emission, where the total energy emission is

$$\mathcal{E}_c^{(0)} \approx 3^{2/3}(\pi A_c)^2 G\mu \times \begin{cases}1/4, & \text{for } \ell \gg \ell_* \\ 1, & \text{for } \ell \ll \ell_*\end{cases} \tag{2.89}$$

123

with only cusps above the collapse threshold being subject to the 1/4 reduction in GW power due to the truncation of the signal. However, the difference becomes more significant in the memory emission,

$$\mathcal{E}_{\mathrm{c}}^{(1)} \approx \begin{cases} \dfrac{1}{2}(\pi B_{\mathrm{c}})^2 (G\mu)^{8/3}, & \text{for } \ell \gg \ell_* \\[2mm] 8(\pi B_{\mathrm{c}})^{3/2}(G\mu)^2 (\ell/\ell_*), & \text{for } \ell \ll \ell_* \end{cases} \tag{2.90}$$

where there is an extra power of $(G\mu)^{2/3}$ for cusps above the collapse threshold, compared to those just below it. This pattern continues for higher-order memory, where the total radiated energy for all $n \geq 2$ is given by

$$\mathcal{E}_{\mathrm{c}}^{(n)} \approx \begin{cases} \dfrac{1}{2}\left(\dfrac{\pi B_{\mathrm{c}}}{4}\right)^{2^n} (G\mu)^{2^{n+2}/3} \displaystyle\int_{S^2} \dfrac{\mathrm{d}^2\hat{\boldsymbol{r}}}{4\pi} |L_n(I)|^2, & \text{for } \ell \gg \ell_* \\[3mm] \dfrac{1}{2}(\pi B_{\mathrm{c}})^{3/2}(G\mu)^2 (\ell/\ell_*)^{(2n-1)/3} \displaystyle\int_{S^2} \dfrac{\mathrm{d}^2\hat{\boldsymbol{r}}}{4\pi} 6(1+\cos I) S_n(I), & \text{for } \ell \ll \ell_* \end{cases} \tag{2.91}$$

In section 2.6.1, we will also discuss a further GW signature associated with cusp collapse: the high-frequency quasi-normal ringing of the newly-formed black hole after the collapse. We neglect this effect here however, as it is hard to say anything concrete about the phase evolution and angular pattern of this additional GW emission, and this prevents us from calculating the associated memory signal. It would be interesting to revisit this contribution to the memory if and when more detailed phase-coherent cusp collapse waveform models become available.

2.4 Memory from kinks

As we saw in section 2.1.3, kinks are persistent features of cosmic string loops (unlike cusps, which are transient): they propagate around the loop at the speed of light, continuously emitting GWs in a beam which traces out a one-dimensional 'fan' of directions, like a lighthouse. This fact is usually unimportant when computing the primary GW emission from kinks, since the beam only overlaps with the observer's line of sight for a small fraction of the loop oscillation time, meaning that kinks are effectively transient sources for any given observer. However, in order to calculate a GW memory signal, we need to know the primary GW flux in all directions over the entire history of the source, which for kinks means specifying the beaming direction as a function of time.

We consider here the simplest case, where the beam of the kink traces out a great circle on the sphere at a constant rate. We choose our polar coordinates such that this circle lies in the equatorial plane, $\theta_{\mathrm{k}} = \pi/2$. The azimuthal direction of the kink's beam is then given by $\phi_{\mathrm{k}} = 4\pi\varsigma t/\ell$ with $\varsigma = \pm 1$, where the two different signs correspond to left- and right-moving kinks respectively.[10] From equation (2.44) we then have

$$\tilde{h}_{\mathrm{k}}^{(0)}(f, \hat{\boldsymbol{r}}) \simeq A_{\mathrm{k}} \frac{G\mu\ell^{1/3}}{r|f|^{5/3}} e^{-\mathrm{i}\varsigma\phi f\ell/2} \Theta(\theta_{\mathrm{b}} - |I|)\Theta(|f| - 2/\ell), \tag{2.92}$$

---

[10] The definition of whether a given kink is left- or right-moving is somewhat arbitrary here. For concreteness, we call kinks with $\varsigma = +1$ 'left-moving'; these move anti-clockwise around the loop when viewed from the North pole $\theta = 0$. Conversely, kinks with $\varsigma = -1$ are called 'right-moving'; these move clockwise when viewed from $\theta = 0$.

where the inclination $I \equiv \pi/2 - \theta$ describes the angle between the observer's line of sight and the closest point on the equatorial plane, and takes values $I \in [-\pi/2, \pi/2]$ (with positive/negative values corresponding to the observer being above/below the plane). Note that we have picked up a phase factor $e^{-i\varsigma\phi f\ell/2}$ to account for the time at which the kink passes closest to the line of sight. As we show below, this direction-dependent phase ultimately leads to a strong suppression of the kink memory signal compared to the cusp case.

2.4.1 Beaming effects

As in the cusp case, the first step in calculating the GW memory signal is to compute the angular integral that captures the beaming effects of the kink, $\int_{\hat{r}'} e^{-i\varsigma\phi' f\ell/2}\Theta(\theta_b - |I|)$. Here the small circle of radius θ_b around the North pole that we considered for cusps has been replaced with a narrow band of half-width θ_b around the equator, and we have included the ϕ-dependent phase factor from equation (2.92). It is straightforward to integrate out the zenith angle θ' if we keep only the leading-order term in θ_b; this gives

$$\int_{\hat{r}'} e^{-i\varsigma\phi' f\ell/2}\Theta(\theta_b - |I|) \simeq \theta_b K_{f\ell/2}(-\varsigma I), \tag{2.93}$$

where we have defined a family of azimuthal integrals,

$$K_n(I) \equiv \int_0^{2\pi} \frac{\mathrm{d}\phi}{2\pi} e^{in\phi} \frac{(\sin I \cos\phi - i\sin\phi)^2}{1 - \cos I \cos\phi}. \tag{2.94}$$

Notice that the argument n is always an integer, as the GW frequency is always an integer multiple of the loop's fundamental mode $2/\ell$ (although we often ignore this when taking the continuum limit at high frequencies). Computing the integral equation (2.94) for general n therefore corresponds to finding the Fourier spectrum of a complicated nonlinear function of ϕ; we perform this calculation explicitly in appendix C.3.

2.4.2 Late-time memory

Using equation (2.56), along with the expression for the angular integral $K_0(I)$ from equation (C.19), we find that the late-time memory from the kink is given by

$$\Delta h_k^{(1)} = \frac{3^{5/6}\ell}{r}(\pi A_k G\mu)^2 \frac{4\sin|I| + \cos 2I - 3}{\cos^2 I}. \tag{2.95}$$

Inserting numerical values, the strongest effect is $\Delta h_k^{(1)} \approx -2.125 \times (G\mu)^2 \ell/r$ when the observer lies in the plane of the kink $I = 0$, an order of magnitude smaller than the maximum cusp memory. The late-time kink memory decreases smoothly to zero as one approaches the poles $I \to \pm\pi/2$.
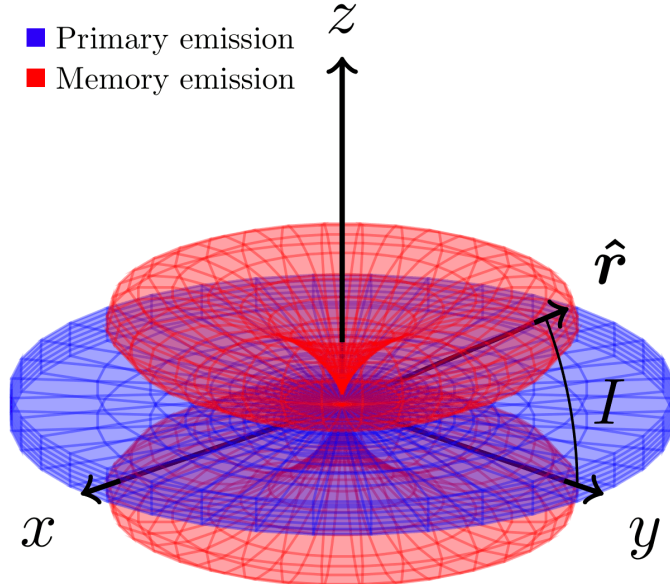
Figure 2.8: A cartoon illustration of the angular distribution of the energy radiated by a kink. The primary emission (blue) is concentrated in a narrow fan of half-width $\theta_b \sim (f\ell)^{-1/3}$ around the plane of the kink, while the memory emission (red) is concentrated in two lobes either side of this plane, which are exponentially suppressed as one approaches either of the directions normal to the plane.

2.4.3 FULL WAVEFORM

The calculation here is very similar to the cusp case in section 2.3.3. Inserting equation (2.92) into equation (2.59), and taking care to enforce the frequency cutoffs and to account for the two competing beam angles $\theta_b(f')$ and $\theta_b(f'-f)$, we obtain

$$\bar{h}_k^{(1)}(f) \simeq -\frac{\mathrm{i}2^{5/3}\pi(A_kG\mu)^2\ell^{1/3}}{3^{1/6}rf|f|^{2/3}}K_{f\ell/2}(-\varsigma I)\left[\int_{2/|f|\ell}^{\infty}\frac{\mathrm{d}u}{u^{2/3}(1+u)} - \int_{2/|f|\ell}^{1/2}\frac{\mathrm{d}u}{u^{2/3}(1-u)}\right] \quad (2.96)$$

for $|f| > 4/\ell$. (This is analogous to equation (2.63) from the cusp case.) Taking the limit $|f| \gg 2/\ell$, we can evaluate the integrals analytically to find

$$\bar{h}_k^{(1)}(f) \simeq -\mathrm{i}B_k\frac{(G\mu)^2\ell^{1/3}}{rf|f|^{2/3}}K_{f\ell/2}(-\varsigma I)\Theta(|f|-4/\ell),$$

$$B_k \equiv 12^{5/6}\pi A_k^2\left[\frac{2\pi}{3\sqrt{3}} - {}_2F_1(\tfrac{1}{3},\tfrac{1}{3};\tfrac{4}{3};-1)\right] \approx 0.2915, \quad (2.97)$$

where we have set the frequency cutoff at double the fundamental mode $2/\ell$ due to the different behaviour of the angular integral $K_n(I)$ for $n = 1$ compared to $n \geq 2$.

This waveform (2.97) shares many features with both the cusp memory waveform (2.64) and the primary kink waveform (2.92). The most important difference from both of those waveforms is the de-

pendence on inclination, which here is a function of frequency. Using the results derived in appendix C.3, we can rewrite equation (2.97) as

$$
\tilde{h}_{\mathrm{k}}^{(1)}(f) \simeq -\mathrm{i}B_{\mathrm{k}} \frac{(G\mu)^2 \ell^{1/3}}{r f |f|^{2/3}} \frac{4 \sin |I|}{\cos^2 I}
\begin{cases}
\left(\dfrac{\cos I}{1 + \sin |I|} \right)^{f\ell/2} \Theta(f - 4/\ell), & \text{for } \varsigma I < 0 \\[3mm]
\left(\dfrac{\cos I}{1 + \sin |I|} \right)^{-f\ell/2} \Theta(-f - 4/\ell), & \text{for } \varsigma I > 0
\end{cases}
\tag{2.98}
$$

meaning that the memory signal from left-moving kinks contains only negative frequencies above the equatorial plane and only positive frequencies below the plane, and vice versa for right-moving kinks.

For high frequencies $|n| \gg 1$ the angular integral $K_n(I)$ has a maximum value of $\simeq 4/(\mathrm{e}|n|)$ at inclination $I \simeq 1/n$. This means that the kink memory signal is only observable very close to the plane of the kink (but *not* in the plane, where it vanishes; see figure 2.8), and is suppressed by an extra power of frequency compared to the primary signal, $\tilde{h}_{\mathrm{k}}^{(1)} \sim f^{-8/3}$.

2.4.4 Time-domain waveform near the arrival time

As with the cusp case, we can reverse-Fourier-transform equation (2.98) to find the time-domain memory strain around time of arrival of the primary kink signal, $|t - t_0| \ll \ell$. Unlike the cusp case, we obtain a signal which is not linearly polarised, but contains both + and × polarisation content,

$$
\begin{aligned}
h_{\mathrm{k},+}^{(1)}(t) - h_{\mathrm{k},+}^{(1)}(t_0) &\simeq \frac{16\pi B_{\mathrm{k}}(G\mu)^2}{2^{1/3} r} (t - t_0) \frac{\sin |I|}{\cos^2 I} E_{2/3}\left(2\ln \frac{1 + \sin |I|}{\cos I} \right), \\
h_{\mathrm{k},\times}^{(1)}(t) - h_{\mathrm{k},\times}^{(1)}(t_0) &\simeq \frac{64\pi^2 B_{\mathrm{k}}(G\mu)^2}{2^{1/3} \ell r} (t - t_0)^2 \frac{\sin \varsigma I}{\cos^2 I} E_{-1/3}\left(2\ln \frac{1 + \sin |I|}{\cos I} \right),
\end{aligned}
\tag{2.99}
$$

where we have used the generalised exponential integral function, $E_n(z) \equiv \int_1^\infty \mathrm{d}x\, \mathrm{e}^{-zx} x^{-n}$. The ×-polarised component is suppressed by an additional factor of $(t - t_0)/\ell \ll 1$, meaning that the signal is still approximately +-polarised.

An important difference with respect to the cusp case is that since kinks are long-lived rather than transient, their memory signal is not concentrated around a particular arrival time, but can in principle be observed at *all* times. One can easily obtain expressions analogous to equation (2.99) at any point in the kink's periodic motion by substituting in the appropriate time when evaluating the reverse Fourier transform; in general this leads to a mixing between the +- and ×-polarisation modes.

2.4.5 Radiated energy

As in the cusp case, we are interested in the total energy radiated by the kink, and how the memory emission adds to this. Using equation (2.37), we find that the dimensionless energy spectrum for the first-order memory is

$$
\epsilon_{\mathrm{k}}^{(1)}(f, \hat{\boldsymbol{r}}) \simeq 4\pi B_{\mathrm{k}}^2 \frac{(G\mu)^3}{(f\ell)^{1/3}} \frac{\sin^2 I \cos^{f\ell-4} I}{(1 + \sin |I|)^{f\ell}} \Theta(f - 4/\ell).
\tag{2.100}
$$

Recall that when integrating over the sphere, the primary spectrum (2.45) is suppressed by a factor of the beaming angle $\theta_b \sim (f\ell)^{-1/3}$. The spherical integral for the memory contribution can be evaluated explicitly to give a much stronger suppression,

$$\int_{S^2} \mathrm{d}^2\hat{\boldsymbol{r}}\, \frac{\sin^2 I \cos^{f\ell-4} I}{(1 + \sin|I|)^{f\ell}} = \frac{8\pi}{f\ell(f\ell - 2)(f\ell + 2)} \simeq 8\pi/(f\ell)^3, \tag{2.101}$$

so that at high frequencies, the isotropic energy spectrum is approximately

$$\bar{\epsilon}_k^{(1)}(f) \simeq 32(\pi B_k)^2 \frac{(G\mu)^3}{(f\ell)^{10/3}} \Theta(f - 4/\ell). \tag{2.102}$$

We see that the angular pattern of the kink memory strongly suppresses the isotropic spectrum at high frequencies, meaning that unlike in the cusp case, there is no frequency range where the memory contribution dominates, and the total radiated energy converges without imposing an ultraviolet cutoff,

$$\mathcal{E}_k^{(1)} \simeq \frac{3(\pi B_k)^2}{2^{8/3} \times 5} (G\mu)^3. \tag{2.103}$$

2.4.6 HIGHER-ORDER MEMORY

As with the cusp case, we can iterate the memory calculation with our 1st-order memory waveform (2.97) as an input to calculate the 2nd-order memory effect (the 'memory of the memory'). This involves calculating the angular integral

$$\int_{\hat{\boldsymbol{r}}'} \mathrm{e}^{-\mathrm{i}\varsigma\phi' f\ell/2} K_{f'\ell/2}(-\varsigma I') K_{(f'-f)\ell/2}(-\varsigma I'), \tag{2.104}$$

which in the high-frequency regime $|f'|, |f' - f| \gg 2/\ell$ is well-approximated by

$$\simeq \frac{16 K_{f\ell/2}(-\varsigma I)}{(f'\ell - f\ell/2)^3} [\Theta(f' - 4/\ell)\Theta(f' - f - 4/\ell) - \Theta(-f' - 4/\ell)\Theta(-f' + f - 4/\ell)]. \tag{2.105}$$

With this result to hand, the remaining steps are very similar to the calculations for the 1st-order memory described above, yielding

$$\begin{aligned} \bar{h}_k^{(2)}(f) &\simeq -\mathrm{i}\frac{512\pi B_k^2 (G\mu)^4}{r f |f|^{10/3}\ell^{7/3}} K_{f\ell/2}(-\varsigma I)\Theta(|f| - 4/\ell) \int_{4/|f|\ell}^{\infty} \frac{\mathrm{d}u}{u^{2/3}(u+1)^{2/3}(u+1/2)^3} \\ &\approx -\mathrm{i}\frac{1666 \times (G\mu)^4}{r f |f|^{10/3}\ell^{7/3}} K_{f\ell/2}(-\varsigma I)\Theta(|f| - 4/\ell). \end{aligned} \tag{2.106}$$

The fact that this has the exact same dependence on the inclination I as the 1st-order memory makes it straightforward to iterate the process to higher orders. Doing this, we find that for all $n \geq 2$, the kink GW memory is given schematically by

$$\bar{h}_k^{(n)}(f) \sim \frac{-\mathrm{i}(G\mu)^{2n}f\ell^3}{r(|f|\ell)^{8n/3}} K_{f\ell/2}(-\varsigma I)\Theta(|f| - 4/\ell), \tag{2.107}$$

multiplied by some numerical constant (for which there does not seem to be a simple expression for all n).

The situation here is drastically different from the cusp case. For cusps we saw that each successive order in the memory was suppressed by larger powers of $G\mu \ll 1$, but also enhanced by larger powers of $\ell/\delta \gg 1$, and that in certain situations the latter would dominate, causing a divergence. For kinks, we see instead that each order in the memory is not only suppressed by powers of $G\mu$, but is further suppressed by a factor of $(|f|\ell)^{-8/3} \ll 1$ each time. The signal falls off quickly enough with frequency that higher-order memory contributions are not sensitive to the string-width scale, and no factors of ℓ/δ appear. This means that there is no situation where the kink memory diverges, and that the higher-order contributions are negligible in any observational scenario. This lack of divergence is in agreement with the cusp-collapse mechanism we have invoked as a possible resolution for the cusp divergence, as we will see below that kinks are not predicted to form PBHs.

2.4.7 CAVEATS OF OUR APPROACH

We have only considered the simplest case where kink's beam traverses a fixed plane at a constant rate, in order to make detailed analytical calculations feasible. This situation is highly idealised, and is not representative of the loops one would find in a cosmological loop network, which would typically contain structure on scales smaller than the loop length ℓ, causing the path of the beam to vary on those scales. We expect that such structure is only likely to make a significant *qualitative* difference to our results if it is on a scale corresponding to the GW frequency of interest. We note that small-scale structure on loops is expected to be damped over time through gravitational backreaction [465], so our simple treatment here is not likely to be too unreasonable. In any case, we do not expect such considerations to change the main conclusion of this section: that memory from kinks is highly suppressed compared to the cusp case.

We have also neglected the fact that kinks always appear in pairs on loops, with one left-mover for every right-mover. A realistic loop is likely to have several pairs of kinks, with each of these sourcing a GW memory signal like the one calculated here. Since the kinks travel around the loop at the same average rate, it is possible that the superposition of their memory signals could give rise to interesting coherent effects. However, regardless of whether the kink memory signals are coherent or not, the GW energy flux will still be of the same order of magnitude, so our main conclusions are unaffected.

2.5 PRIMORDIAL BLACK HOLES FROM CUSP COLLAPSE

We saw in section 2.3 that the standard Nambu-Goto cusp waveform (2.36) leads to a divergence in the total GW strain when accounting for all higher-order memory terms. In section 2.3.9, we argued that this is due to a breakdown in the weak-field assumption used to derive the waveform in the first place. In order to resolve the divergence, there must be some form of strong-gravity mechanism which truncates the high-frequency GW emission.

In this section, we explore one tentative candidate for such a mechanism: that the GW emission is suppressed by the formation of a black hole event horizon enclosing the cusp. Aside from curing the cusp memory divergence, this presents a novel formation channel for primordial black holes. We present two heuristic arguments for why this might happen, both based on the *hoop conjecture* (which
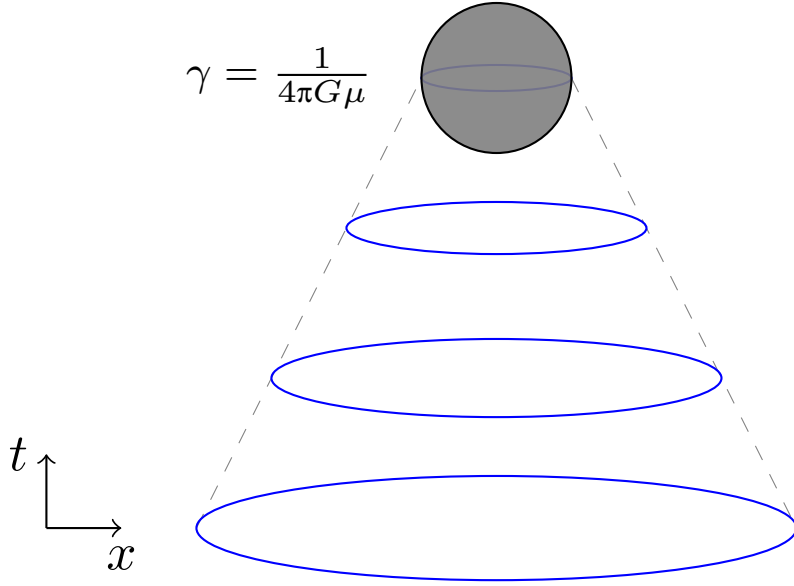
Figure 2.9: A collapsing circular cosmic string loop (in blue) forms a PBH (in grey) once its Lorentz factor satisfies equation (2.109).

we will define below). These arguments are not mathematically rigorous, and are subject to some open conceptual issues around how exactly the hoop conjecture applies in highly dynamical and relativistic situations such as cusps [132]. Ultimately we will require fully general-relativistic calculations to gain a complete understanding of strong-gravity effects near cusps, probably in the form of numerical relativity simulations. With these caveats in mind, we explore this 'cusp-collapse' scenario as one possible resolution of the nonlinear memory divergence identified above, with particular emphasis on potential observational signatures that could allow us to test this proposal.

2.5.1 The hoop conjecture

The hoop conjecture, first formulated by Thorne [540], is a powerful diagnostic for the formation of BH horizons in GR. The conjecture states that 'horizons form when, and only when, a mass M gets compacted into a region whose circumference in every direction is $C \leq 4\pi G M$' [429]. In other words, if a sphere containing mass M fits inside its own Schwarzschild radius $r_S \equiv 2GM$, it forms a black hole. This conjecture is intentionally somewhat vaguely defined in several ways; for example, there is no unambiguous way to assign a mass to the gravitational field inside the sphere.[11] For the present situation, we include only the mass due to the matter fields, $M_{\text{sphere}} \equiv \int_{\mathcal{B}_r} d^3 x\, T_{00}(t, x)$, where $\mathcal{B}_r$ is a ball of radius r, and the mass is a function of time and of the centre of the ball. The hoop conjecture then predicts BH formation if

$$\frac{2GM_{\text{sphere}}}{r} \geq 1. \tag{2.108}$$

which we refer to as the 'hoop condition'.

---

[11]Doing so would require a *quasi-local* measure of gravitational mass in GR, which is challenging for reasons we touched on in section 0.1.7.

We are interested in cosmic string loops which lead to PBH formation, i.e., solutions to equations (2.12)–(2.14) which satisfy the hoop condition (2.108) at some point in their evolution. The simplest example is a circular loop, which contracts at an accelerating rate until the entire loop is compact enough to form a PBH, as illustrated in figure 2.9. This occurs within a single loop oscillation period, and results in a PBH of mass $M_{\mathrm{pbh}} \sim M_{\mathrm{loop}} = \mu\ell$, which is smaller than the original loop by a factor of $G\mu$. One can show that the horizon forms only once the loop's Lorentz factor satisfies

$$\gamma = \frac{1}{4\pi G\mu}. \tag{2.109}$$

Since we know that $G\mu \ll 1$, this corresponds to an ultrarelativistic contraction velocity $v \simeq 1 - 8\pi^2(G\mu)^2$, and we can understand the PBH formation as being due to relativistic length contraction. This mechanism for PBH formation from (quasi)circular loops has been studied extensively in the literature [88, 164, 165, 282, 283, 306, 307, 310, 337, 400, 461]. However, circular collapse is only possible if all three components of the loop's angular momentum are smaller than those of a typical loop by a factor of $\sim G\mu$ [567]. Circular collapse is thus finely-tuned, and only a very small fraction of the cosmic loop population is expected to collapse in this way.

This naturally leads one to ask whether generic (i.e. noncircular) loops can form PBHs. It is easy to see intuitively that a Lorentz factor of order $\gamma \sim (G\mu)^{-1}$ like that in equation (2.109) is a necessary condition for PBH formation, even for noncircular loops. Suppose we want to form a PBH which contains some fraction f of the loop's mass, $M_{\mathrm{pbh}} = f\mu\ell$, corresponding to a σ interval of $\Delta\sigma = f\ell$. This length of string must be compacted into a region of diameter $\lesssim 2r_{\mathrm{S}} = 4GM_{\mathrm{pbh}}$. The ratio between this lengthscale and the corresponding σ interval is related to the loop's tangent vector,

$$|\boldsymbol{X}'| \approx \frac{|\Delta\boldsymbol{X}|}{\Delta\sigma} \lesssim \frac{4GM_{\mathrm{pbh}}}{f\ell} = 4G\mu. \tag{2.110}$$

We can relate the tangent vector to the loop's dynamics by rearranging equation (2.13) to get

$$|\boldsymbol{X}'| = \sqrt{1 - |\dot{\boldsymbol{X}}|^2} = \frac{1}{\gamma}, \tag{2.111}$$

which shows that the hoop condition is generically satisfied if part of the loop has a large enough Lorentz factor,

$$\gamma \gtrsim \frac{1}{4G\mu}. \tag{2.112}$$

This is not a sharp bound, just an order-of-magnitude estimate. The corresponding (exact) inequality (2.109) for circular loops agrees to within a factor of π. We expect equation (2.112) to have a similar level of accuracy for generic loop configurations.

Note that in the above argument we have *not* assumed that the entire loop must be moving at such high velocities, only some fraction f of it. This is in contrast with the literature on (quasi)circular loop collapse, which has looked exclusively at cases where *all* of the loop's mass ends up behind the PBH horizon.[12] The

---

[12] In fact, Polnarev and Zembowicz [461] briefly mention the possibility of forming a PBH from just part of the loop, but do not discuss this idea in any detail.
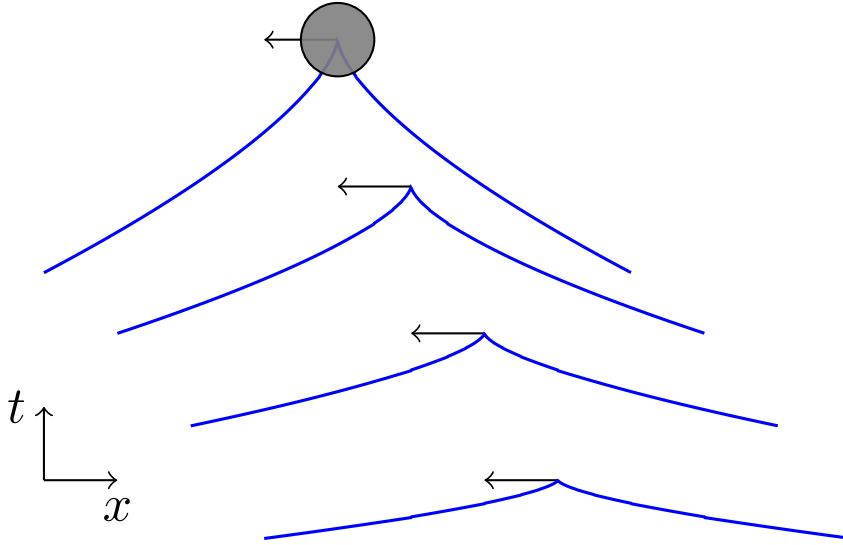
Figure 2.10: A segment of a cosmic string loop (in blue) becomes more compact as it develops a cusp. Once it satisfies the hoop condition (2.108) it collapses to form a PBH (in grey).

argument sketched above therefore suggests a change in focus: rather than looking at PBH formation *from* loops, we should be concerned with PBH formation *on* loops.

2.5.2 THE CUSP-COLLAPSE MECHANISM

Now consider what happens for cusps. Since these correspond to a point on the loop moving at the speed of light, the corresponding Lorentz factor diverges, instantaneously compacting a finite fraction of the loop's mass into an infinitesimally small region. Equation (2.112) then suggests that cusps should therefore lead to some fraction of the loop's mass being enclosed behind a horizon, as illustrated in figure 2.10.

We can look at this idea in more detail by considering the behaviour of solutions to the flat-space EoMs (2.12)–(2.14) near a cusp, as we did in section 2.1.3. Using equation (2.25), we see that the position and velocity of the loop near the cusp at time $t = 0$ are given by

$$
\begin{aligned}
\boldsymbol{X}_0(\sigma) &\equiv \frac{1}{2}[\boldsymbol{X}_+(\sigma) + \boldsymbol{X}_-(-\sigma)] = \frac{1}{2}\ddot{\boldsymbol{X}}\sigma^2 + O(\sigma^3), \\
\dot{\boldsymbol{X}}_0(\sigma) &\equiv \frac{1}{2}[\dot{\boldsymbol{X}}_+(\sigma) + \dot{\boldsymbol{X}}_-(-\sigma)] = \hat{\boldsymbol{r}}_c + \frac{1}{2}\dddot{\boldsymbol{X}}\sigma^2 + O(\sigma^3),
\end{aligned}
\tag{2.113}
$$

so that the distance from the cusp as a function of σ is given by

$$
r_0(\sigma) = \sqrt{\boldsymbol{X}_0 \cdot \boldsymbol{X}_0} = \frac{1}{2}|\ddot{\boldsymbol{X}}|\sigma^2 + O(\sigma^3).
\tag{2.114}
$$

We see that the fact that $\dot{\boldsymbol{X}}_+ = \dot{\boldsymbol{X}}_- = \hat{\boldsymbol{r}}_c$ at the cusp means that there is no term of order σ in equation (2.114), and the distance grows much more slowly for small σ than it would on a non-cuspy part of the loop; this is the crucial ingredient for fulfilling the hoop condition.

Consider now a sphere of radius $r \ll \ell$, centred at the cusp. We see from equation (2.114) that the portion of the loop contained in the sphere is given by $-\sigma_* \leq \sigma \leq \sigma_*$, where $\sigma_* \ll \ell$ is defined by $r = r_0(\sigma_*)$, such that $\sigma_* = \sqrt{2r/|\ddot{\boldsymbol{X}}|}$. We thus see that the mass contained in the sphere is

$$M_{\mathrm{sphere}} = \mu \int_{-\sigma_*}^{+\sigma_*} \mathrm{d}\sigma = 2\mu\sigma_* = \left(\frac{8\mu^2 r}{|\ddot{\boldsymbol{X}}|}\right)^{1/2}. \tag{2.115}$$

The hoop condition (2.108) is therefore satisfied if $r|\ddot{\boldsymbol{X}}| \leq 32(G\mu)^2$, with the limiting PBH mass being $M_{\mathrm{pbh}} = 16G\mu^2/|\ddot{\boldsymbol{X}}|$. The fact that this depends on the cusp's acceleration $|\ddot{\boldsymbol{X}}|$ rather than its velocity may seem surprising at first, but we can understand this intuitively by using equations (2.12) and (2.111) to write $|\ddot{\boldsymbol{X}}| = |\boldsymbol{X}''| = \mathrm{d}(1/\gamma)/\mathrm{d}\sigma$. The acceleration therefore tells us about the rate of change of the Lorentz factor along the loop, and thus controls the size of the region that satisfies the hoop condition, which sets the PBH mass.

We therefore find that, assuming the hoop conjecture can be applied in this simple way, cusps are predicted to form PBHs with mass

$$M_{\mathrm{pbh}} = \frac{8}{\pi\bar{n}}G\mu^2\ell \approx G\mu M_{\mathrm{loop}}, \tag{2.116}$$

which are a factor of $G\mu$ smaller than those formed from circular collapse. (Here we have used the effective mode number $\bar{n} \equiv \ell|\ddot{\boldsymbol{X}}|/(2\pi)$ from section 2.1.3.)

2.5.3 Properties of the PBHs

We can estimate the properties of the PBHs formed through cusp collapse by assuming that all of the energy-momentum inside the sphere of radius $2GM$ at time $t = 0$ is trapped behind the horizon. Using equation (2.23), the PBH's linear and angular momenta are then given by

$$\begin{aligned} P^i &= \int_{\mathcal{B}_r} \mathrm{d}^3\boldsymbol{x}\, T^{0i}(0, \boldsymbol{x}) = \mu \int_{-\sigma_*}^{+\sigma_*} \mathrm{d}\sigma\, \dot{X}_0^i, \\ J^i &= \int_{\mathcal{B}_r} \mathrm{d}^3\boldsymbol{x}\, \varepsilon^i{}_{jk} x^j T^{0k}(0, \boldsymbol{x}) = \mu \int_{-\sigma_*}^{+\sigma_*} \mathrm{d}\sigma\, \varepsilon^i{}_{jk} X_0^j \dot{X}_0^k, \end{aligned} \tag{2.117}$$

where ε_{ijk} is the Levi-Civita symbol. Inserting the leading-order terms from equation (2.113), and using $M = 2\mu\sigma_* = 16G\mu^2/|\ddot{\boldsymbol{X}}|$, we find

$$\boldsymbol{P} = M\left[\hat{\boldsymbol{r}}_{\mathrm{c}} + \frac{32(G\mu)^2}{3}\frac{\ddot{\boldsymbol{X}}}{|\ddot{\boldsymbol{X}}|^2}\right], \qquad \boldsymbol{J} = \frac{2GM^2}{3}\frac{\ddot{\boldsymbol{X}}}{|\ddot{\boldsymbol{X}}|} \times \hat{\boldsymbol{r}}_{\mathrm{c}}. \tag{2.118}$$

Thus we see that immediately after formation, the PBH is moving in the cusp direction $\hat{\boldsymbol{r}}_{\mathrm{c}}$ with an ultrarelativistic velocity $v = |\boldsymbol{P}|/M \approx 1$. In fact, the PBH's Lorentz factor is of the same order of magnitude as our estimate (2.112),

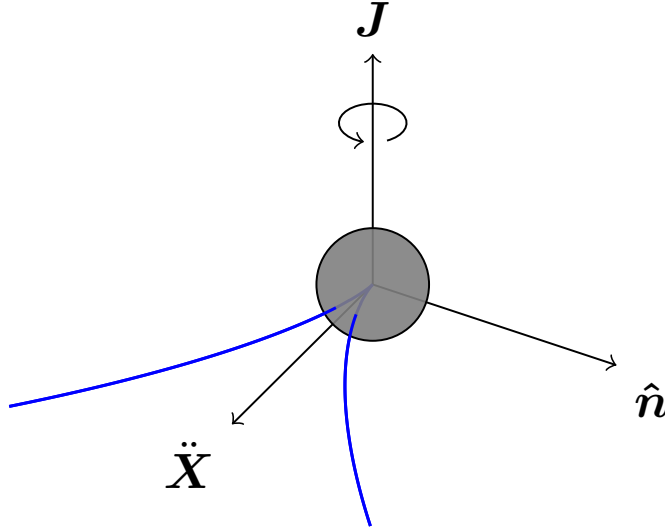$$\gamma \lesssim \sqrt{\frac{3}{128}}(G\mu)^{-1}, \tag{2.119}$$

Figure 2.11: An illustration of the PBH (in grey) immediately after formation. The cusp acceleration $\ddot{\boldsymbol{X}}$, cusp velocity $\hat{\boldsymbol{r}}_{\mathrm{c}}$, and PBH angular momentum $\boldsymbol{J}$ are all orthogonal to each other. The cosmic string (in blue) punctures the horizon at two points separated by a small angle $\sim G\mu$, with its cusp hidden behind the horizon.

where we have used equation (2.26) and the triangle inequality $|\ddot{\boldsymbol{X}}_+|^2 + |\ddot{\boldsymbol{X}}_-|^2 \geq |\ddot{\boldsymbol{X}}_+ + \ddot{\boldsymbol{X}}_-|^2$. We also see that the PBH is spinning around an axis orthogonal to both the cusp's velocity $\hat{\boldsymbol{r}}_{\mathrm{c}}$ and its acceleration $\ddot{\boldsymbol{X}}$, as illustrated in figure 2.11, with a dimensionless spin parameter,

$$\chi \equiv \frac{|\boldsymbol{J}|}{GM^2} = 2/3, \tag{2.120}$$

that is two-thirds of the extremal Kerr value $\chi = 1$. This large spin is a direct consequence of the orthogonality of the cusp's velocity and acceleration, as enforced by equation (2.26). It is perhaps surprising at first that we obtain such a specific prediction for the spin, but we can understand this as a consequence of the universality of the string solution near the cusp, as discussed in section 2.1.3. We should point out, however, that this prediction does not account for the mass and angular momentum radiated during the collapse; we return to this point in section 2.5.6 below.

Note that the mass M includes the kinetic energy of the PBH, which is large due to its ultrarelativistic velocity. The PBH's rest mass is given by

$$m \equiv \sqrt{M^2 - P^2} = \frac{M}{\gamma} \approx (G\mu)^2 \mu\ell, \tag{2.121}$$

which is smaller than the total mass of the loop by a factor of $\sim (G\mu)^2$. This means that the PBH radius is

$$r \sim (G\mu)^3 \ell, \tag{2.122}$$

i.e., the size of the PBH scales linearly with the size of the loop, as expected from the fact that the PBH's mass is a fixed fraction of that of the loop (c.f. equation (2.116)). By using the Nambu-Goto approximation, we have assumed throughout that the cosmic strings have zero width, effacing any physics which occurs

on lengthscales smaller than the string width $\delta \sim \sqrt{\hbar/\mu}$. We therefore expect our results to hold only if the PBH radius is larger than δ, which implies

$$\ell \gtrsim \frac{\delta}{(G\mu)^3} = \frac{\ell_*}{(G\mu)^2} \approx 5.1 \, \text{km} \times \left(\frac{G\mu}{10^{-11}}\right)^{-7/2}, \tag{2.123}$$

and corresponds to a minimum PBH rest mass of

$$m_{\min} \approx \frac{\delta}{G} \approx \frac{m_{\text{Pl}}}{\sqrt{G\mu}} \approx 6.9 \, \text{g} \times \left(\frac{G\mu}{10^{-11}}\right)^{-1/2}, \tag{2.124}$$

where we recall that $m_{\text{Pl}}/\sqrt{G\mu}$ is the energy scale at which the cosmic strings are formed. We thus see that there are three different classes of loops, corresponding to three broad ranges of loop lengths: loops smaller than $\delta/(G\mu)$ are driven by field-theoretic effects (as we mentioned in section 2.1.1), loops larger than $\delta/(G\mu)^3$ are able to form PBHs from cusps, and loops inbetween are unchanged compared to the standard treatment in the Nambu-Goto literature. (These different regimes are summarised in figure 2.12.) Remarkably, this cusp-collapse regime $\ell \gtrsim \delta/(G\mu)^3$ corresponds exactly to the 'large-loop' regime in which the nonlinear memory from the cusp diverges, as we showed in section 2.3. This suggests that we are along the right lines in searching for a strong-gravity resolution to this divergence.

The loop punctures the PBH horizon at two points, corresponding to $(t, \sigma) = (0, \pm\sigma_*)$. Using the expansion (2.113) for $\boldsymbol{X}_0(\pm\sigma_*)$, we see that both points lie very near to the $\ddot{\boldsymbol{X}}$ axis, which is in the PBH's equatorial plane. By continuing the expansion to at least $O(\sigma^3)$, one can show that

$$\cos\theta_* \equiv \frac{\boldsymbol{X}_0(\sigma_*) \cdot \boldsymbol{X}_0(-\sigma_*)}{|\boldsymbol{X}_0(\sigma_*)||\boldsymbol{X}_0(-\sigma_*)|} = 1 + O(\sigma_*^2), \tag{2.125}$$

which implies that the angle between the two puncture points is $\theta_* \sim \sigma_*/\ell \sim G\mu$, and thus that both points are very close to the equatorial plane. (This is important for our discussion of the subsequent dynamics of the loop near the PBH in Section 2.5.7.) Accounting for the finite string width δ, we see that these two puncture points are so close that it is possible for the loop to self-intersect at the PBH horizon. Using simple trigonometry, the separation between the puncture points on the horizon is roughly

$$2G m_{\text{pbh}} \tan\frac{\theta_*}{2} \sim (G\mu)^4 \ell, \tag{2.126}$$

so the loop self-intersects at the horizon if this separation is smaller than the string width δ, which occurs if

$$\ell \lesssim \frac{\delta}{(G\mu)^4} = \frac{\ell_*}{(G\mu)^3} \approx 1.7 \times 10^{-2} \, \text{pc} \times \left(\frac{G\mu}{10^{-11}}\right)^{-9/2}. \tag{2.127}$$

In this case, one would expect the string to intercommute near the horizon, meaning that the PBH would be immediately chopped off from the loop at formation.

It is worth pointing out that we could have found the exact same PBH radius (2.122) in a much quicker way if we had applied the hoop condition to the GW energy density radiated by the cusp, rather than the

energy density of the loop itself. To see this, write the total GW flux through a sphere of radius r around the cusp as

$$M_{\mathrm{gw}}(r) = \mu\ell \int_{2\pi/r}^{\infty} \frac{\mathrm{d}f}{f} \bar{\epsilon}_{\mathrm{c}}(f) \sim G\mu^2 \ell^{2/3} r^{1/3}, \tag{2.128}$$

where we have made sure to only include GWs with wavelengths smaller than the sphere, so that their energy density (0.43) is well-defined. This expression then satisfies the hoop condition $2GM_{\mathrm{gw}}/r = 1$ for a sphere of size $r \sim (G\mu)^3\ell$, perfectly matching what we found in equation (2.122). Indeed, recalling our finding from section 2.1.3 that GWs of frequency f are generated by a region $|\sigma| \sim \ell^{2/3} f^{-1/3}$ surrounding the cusp, we see that substituting in $f \sim 2\pi/r \sim (G\mu)^{-3}/\ell$ gives $|\sigma| \sim G\mu\ell \sim \sigma_*$, so that the string segment radiating these GWs corresponds exactly to the region that collapses according to our arguments above.

2.5.4 OTHER POINTS ON THE LOOP: PSEUDOCUSPS AND KINKS

Consider now a generic loop segment at some time $t = 0$, whose configuration is locally described by

$$X_{\pm}(\sigma_{\pm}) = \hat{r}_{\pm}\sigma_{\pm} + \frac{1}{2}\ddot{X}_{\pm}\sigma_{\pm}^2 + O(\sigma_{\pm}^3). \tag{2.129}$$

If $\hat{r}_+ = \hat{r}_-$, the point $\sigma = 0$ is a cusp with a divergent Lorentz factor $\gamma \to \infty$, which forms a PBH as described above. However, our simple argument in equation (2.112) suggests that a divergent Lorentz factor is not necessary for PBH formation; all we need is $\gamma = O(1/G\mu)$. We therefore expect PBH formation in situations where $\theta \equiv \cos^{-1}\hat{r}_+ \cdot \hat{r}_- \simeq |\hat{r}_+ - \hat{r}_-|$ is nonzero, so long as θ is small enough (note that this is very reminiscent of our discussion around the beaming angle for the cusp GW signal in section 2.1.3). We refer to points on the loop where θ is small but nonzero as 'pseudocusps'.

In order to estimate how small θ must be, we generalise equation (2.114) to give the distance from the pseudocusp for small σ at time $t = 0$,

$$r_0(\sigma) \approx \frac{1}{2}\theta|\sigma| + \frac{1}{2}|\ddot{X}|\sigma^2, \tag{2.130}$$

where we have used the constraints (2.26), and have approximated $(\hat{r}_+ - \hat{r}_-) \cdot \ddot{X} \approx \theta|\ddot{X}|$. Repeating the arguments of section 2.5.2, we find that a PBH forms so long as $\theta < 8G\mu$, with the corresponding mass given by $M = 2\mu(8G\mu - \theta)/|\ddot{X}|$. The loop velocity at $\sigma = 0$ is $v \simeq 1 - \theta^2/8$, so we can translate the bound $\theta < 8G\mu$ into a bound on the pseudocusp velocity. Doing so, we see that pseudocusps collapse to form PBHs so long as their Lorentz factor obeys

$$\gamma \geq \frac{1}{4G\mu}, \tag{2.131}$$

in agreement with our simple estimate (2.112). Since pseudocusps occur even more generically on loops than cusps do [528], this result further enhances the PBH formation rate.

Note that in writing the Taylor expansion (2.129) we have assumed that the loop is smooth in the neighbourhood of $\sigma = 0$, which precludes any discontinuities in the loop's tangent vector, i.e., kinks. However, it is easy to convince oneself that kinks do not contribute to PBH formation. A kink at some σ_{k}

near $\sigma = 0$ would make equation (2.130) a piecewise smooth function, with different coefficients for each order in σ on either side of the kink. Generically these coefficients are of the same order of magnitude on both sides of the kink; there is nothing about the kink which forces the $O(\sigma)$ term in equation (2.130) to be small, which is what we require for PBH formation. As we have shown above, the smallness of this term is uniquely associated with a large Lorentz factor, and therefore with (pseudo)cusps.

Of course, this last argument depends strongly on the Nambu-Goto approximation; in a full field-theoretic setting one would expect kinks to carry gradient energy, which may be sufficiently concentrated to satisfy the hoop condition. However, one would only expect the gradient energy to be large in a region of size comparable to the string width δ, meaning the resulting PBH masses would be near the minimal mass $m_{\min} \sim \delta/G$ from equation (2.124). For kinks, as for cusps, we can trust the Nambu-Goto approximation so long as we consider PBHs with mass $m \gg m_{\min}$.

2.5.5 Timescale of cusp collapse

One of the main assumptions we have made so far is that the loop is described by the flat-space equations of motion (2.12)–(2.14) right up to the instant of PBH formation. A more complete analysis would account for the gravitational backreaction of the loop on its own dynamics, which one would expect to suppress the cusp. One might worry whether this suppression is strong enough to prevent the PBH from forming.

Significant evidence against this worry comes from the extensive literature on cosmic-string backre-action [107, 133, 134, 159, 183, 191, 214, 465, 539, 570, 571], in which numerous different approaches (both analytical and numerical) have repeatedly shown that backreaction does not prevent cusps from forming. There is general agreement that cusps are suppressed to some degree by backreaction, but that this suppression occurs gradually over many loop oscillation periods, on a timescale of order the loop decay time

$$t_{\text{decay}} \sim \frac{\ell}{G\mu}. \tag{2.132}$$

A serious problem with this argument is that essentially all of the existing work on string backreaction has been done in the weak-field limit, treating the string's gravity as a linear perturbation on the background spacetime. This linearised approach is clearly unable to answer questions about strong-gravity effects, such as whether or not PBH formation takes place.

One piece of evidence we can consider is the timescale on which the PBH formation occurs in the scenario described above. The velocity of the string point $\sigma = 0$ at times near to the cusp, $|t| \ll \ell$, can be written as

$$\dot{X}_{\text{c}}(t) = \hat{r}_{\text{c}} + t\ddot{X} + \frac{1}{2}t^2\dddot{X} + O(t^3), \tag{2.133}$$

with the corresponding Lorentz factor given by

$$\gamma_{\text{c}}(t) \simeq \frac{2}{|t|}|\ddot{X}_+ - \ddot{X}_-|^{-1} \approx \frac{\ell}{\pi\bar{n}|t|}, \tag{2.134}$$

where we have used the constraints (2.26), and the last equality generally holds to within an order of magnitude. Since PBH formation is associated with the Lorentz factor growing above a certain
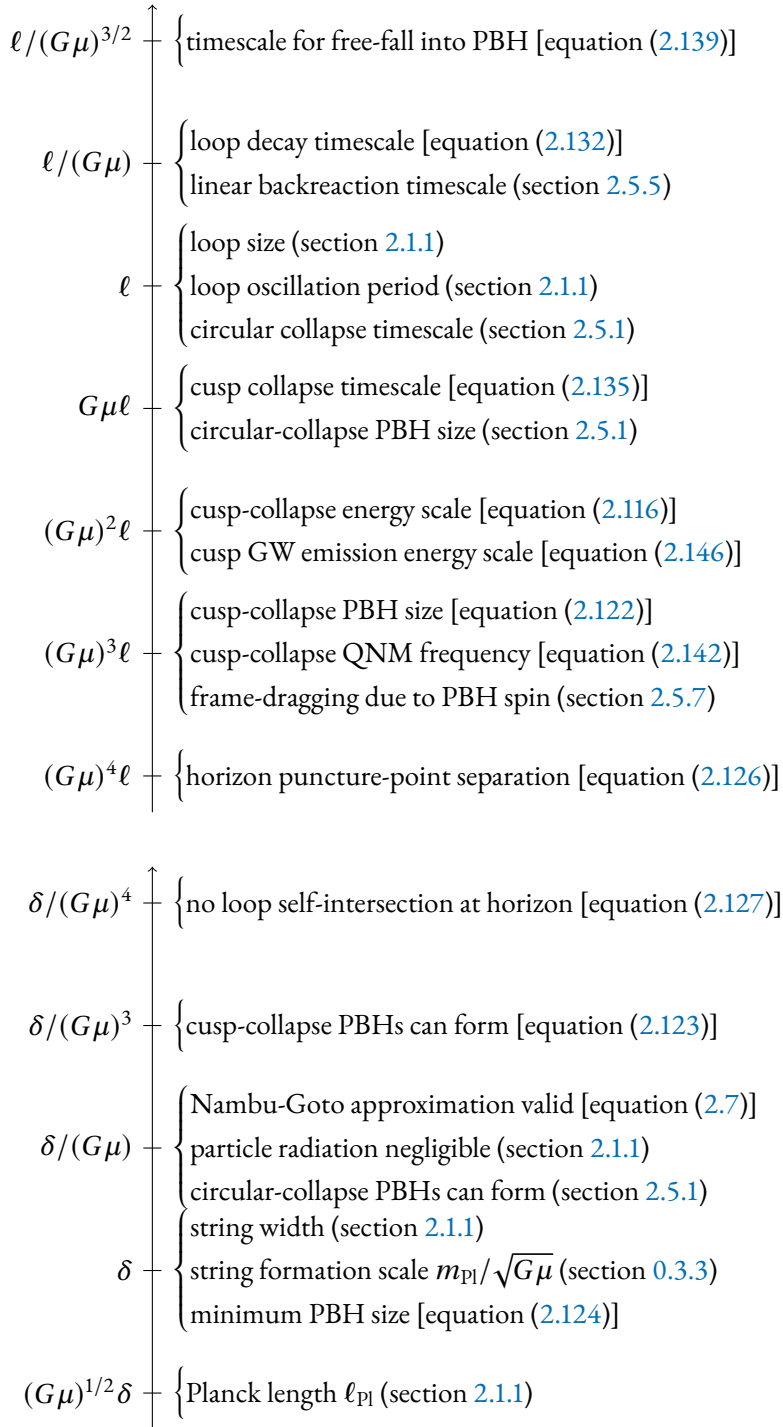
$\ell/(G\mu)^{3/2}$ — {timescale for free-fall into PBH [equation (2.139)]}

$\ell/(G\mu)$ — {loop decay timescale [equation (2.132)]
linear backreaction timescale (section 2.5.5)}

ℓ — {loop size (section 2.1.1)
loop oscillation period (section 2.1.1)
circular collapse timescale (section 2.5.1)}

$G\mu\ell$ — {cusp collapse timescale [equation (2.135)]
circular-collapse PBH size (section 2.5.1)}

$(G\mu)^2\ell$ — {cusp-collapse energy scale [equation (2.116)]
cusp GW emission energy scale [equation (2.146)]}

$(G\mu)^3\ell$ — {cusp-collapse PBH size [equation (2.122)]
cusp-collapse QNM frequency [equation (2.142)]
frame-dragging due to PBH spin (section 2.5.7)}

$(G\mu)^4\ell$ — {horizon puncture-point separation [equation (2.126)]}

$\delta/(G\mu)^4$ — {no loop self-intersection at horizon [equation (2.127)]}

$\delta/(G\mu)^3$ — {cusp-collapse PBHs can form [equation (2.123)]}

$\delta/(G\mu)$ — {Nambu-Goto approximation valid [equation (2.7)]
particle radiation negligible (section 2.1.1)
circular-collapse PBHs can form (section 2.5.1)}

δ — {string width (section 2.1.1)
string formation scale $m_{\text{Pl}}/\sqrt{G\mu}$ (section 0.3.3)
minimum PBH size [equation (2.124)]}

$(G\mu)^{1/2}\delta$ — {Planck length ℓ_{Pl} (section 2.1.1)}

Figure 2.12: A summary of the different scales in the cusp-collapse problem, including references to the equation or section where they first appear. In the Nambu-Goto approximation there are only two dimensionful quantities: the string tension μ and the loop length ℓ. Since the string tension usually appears in the dimensionless combination $G\mu$, all of the system's time- and length-scales can be written as $(G\mu)^p \times \ell$ for some power p. The fact that $G\mu \ll 1$ means that there is a strong hierarchy between these scales. Going beyond Nambu-Goto introduces another dimensionful quantity with its own hierarchy of scales: the string width $\delta \approx (\mu/\hbar)^{-1/2}$. (Note that many of the scales associated with δ here are lower limits on the loop size ℓ; e.g., $\delta/(G\mu)$ is the smallest loop size for which the Nambu-Goto approximation is valid.) These two sets of scales are shifted relative to each other depending on ℓ. More complicated combinations of ℓ and δ are of course possible; e.g., the evaporation timescales for cusp-collapse and circular-collapse PBHs are $(G\mu)^8\ell^3/\delta^2$ and $(G\mu)^2\ell^3/\delta^2$ respectively.

threshold (2.131), we can estimate the associated timescale by setting equation (2.134) equal to this threshold, giving

$$\Delta t_{\mathrm{pbh}} \sim G\mu\ell. \tag{2.135}$$

This shows that the PBH is formed on an extremely short timescale: shorter than the loop oscillation period by a factor of $G\mu$, and shorter than the timescale for linear backreaction by a factor of $(G\mu)^2$. (See figure 2.12 for an overview of the different time- and length-scales in the loop-PBH system.) Even if nonlinear backreaction is in principle strong enough to prevent the cusp from forming, it seems unlikely that it can act on a short enough timescale to do so, meaning that backreaction seems unlikely to be able prevent PBH formation.[13]

Equation (2.135) is a very important result for us, as it shows that we should expect the collapse process to occur some short but finite time before the peak of the cusp GW signal at $t = 0$, thereby suppressing the GW emission at frequencies $f \gtrsim 1/\Delta t$. As we showed in section 2.3.9, this loss of power at very high frequencies cures the nonlinear memory divergence that would otherwise occur for cusps.

2.5.6 Radiation from the collapse

Our analysis thus far has neglected the effects of gravitational radiation during the collapse. Radiation is likely to be important, as the collapse is ultrarelativistic and highly nonspherical. In general, one would expect the final mass, linear momentum, and angular momentum of the PBH to be of the form

$$M = M_0(1 - \epsilon_M), \quad P = P_0(1 - \epsilon_P), \quad J = J_0(1 - \epsilon_J), \tag{2.136}$$

where a zero subscript denotes the naïve, zero-radiation quantity calculated above, and the ϵ_i are three numbers between zero and unity, describing the efficiency with which each quantity is radiated away. In the context of circular loop collapse, Hawking [307] calculated a theoretical upper bound on ϵ_M of $1 - \sqrt{1/2} \approx 29\%$ by explicitly constructing a marginally outer trapped surface in the spacetime of the collapsing loop and requiring that this surface be enclosed by the event horizon of the final PBH. This argument depends heavily on the circular symmetry of the loop, and no such construction seems possible in our case.

Despite the lack of symmetries here, we can make some interesting statements by requiring that the final PBH spin $\chi = J/(GM^2)$ be less than or equal to unity; otherwise the PBH would be 'overspun' to reveal a naked singularity, violating cosmic censorship [452, 575]. Since $\chi_0 = 2/3$, we can write

$$\chi = \frac{2}{3}\frac{(1 - \epsilon_J)}{(1 - \epsilon_M)^2} \leq 1. \tag{2.137}$$

We see that, so long as $\epsilon_J \lesssim 2\epsilon_M$, the final spin parameter of the PBH is larger than the naïve value 2/3, which shows that the upper bound (2.137) is likely to be useful. In general, we expect $\epsilon_J \lesssim \epsilon_M$; see e.g. Durrer [246], in which the rate at which loops radiate angular momentum is shown to be typically an order of magnitude smaller than the rate at which they radiate mass. If $\epsilon_J = \epsilon_M$, then equation (2.137)

---

[13]See also Thompson [539], who examines backreaction on cusps and argues geometrically that cusps form *'no matter how strong the gravitational field near a cusp'*.

gives $\epsilon_M \leq 1/3 \approx 33\%$. In the limit where $\epsilon_J \rightarrow 0$, the bound is even stronger, $\epsilon_M \leq 1 - \sqrt{2/3} \approx 18\%$. Since we expect $0 < \epsilon_J < \epsilon_M$, the true upper bound for cusp collapse is likely to lie somewhere between these two extremes.

Interestingly, numerical relativity simulations of circular loop collapse performed by Helfer *et al.* [310] found $\epsilon_M \lesssim 2\%$, well below Hawking's bound (see also Aurrekoetxea *et al.* [88]). The authors suggest that this is due to the symmetry of the circular collapse, which means the horizon is nearly spherical when it first forms, suppressing the total radiation. The initial horizon in our case is likely to be highly distorted, meaning that ϵ_M is likely to be closer to its upper bound. Of course, it is possible that cusp collapse radiates angular momentum much more efficiently than is typical for loops as a whole, in which case ϵ_J could be larger than calculated by Durrer [246]. It would then be possible for ϵ_M to be larger than the rough bounds we found above.

We note in passing that radiation of linear momentum ($\epsilon_P > 0$) would lead to a 'rocket effect' [319, 557], in which the loop's centre of mass is given a kick in the opposite direction to the radiation. However, even if this process is maximally efficient, the radiated momentum is at most $P_0 \approx G\mu M_{\text{loop}}$, so the maximum kick is $v \approx G\mu$. This pales in comparison to the rms velocity of points on the loop, $v_{\text{rms}} = \sqrt{1/2}$ that we found in equation (2.19), so the effect is of negligible interest; radiation from elsewhere on the loop quickly cancels out the kick.

2.5.7 Dynamics of the loop-PBH system

Once formed, the PBH is inextricably linked to the surrounding string; the portion of the string enclosed behind the horizon cannot escape, and the portion outside the horizon is topologically forbidden from detaching itself. Since the mass of our PBHs is smaller than that of their parent loops by a factor of $G\mu$, we expect the loop dynamics to be largely unaffected by the presence of the PBH, at least on timescales $\sim \ell$. In particular, this means that despite its ultrarelativistic velocity, the PBH cannot drag the rest of the loop along with it—instead, we expect the loop's tension to act on the PBH to decelerate it, and for the loop to continue oscillating in essentially the same motion as before. This could mean that cusp-collapse PBHs do not trace the DM distribution, as their parent loops could easily drag them out of DM haloes. (This possibility was also pointed out by Vilenkin *et al.* [566], albeit for a different PBH formation scenario.)

Most cusp-collapse PBHs are very small, and decay rapidly through Hawking radiation [305]. In particular, the evaporation timescale (0.102) for the minimum mass (2.124) is $\approx t_{\text{Pl}}(m_{\text{pbh}}/m_{\text{Pl}})^3 \approx 10^{-27}\,\text{s} \times (G\mu/10^{-11})^{-3/2}$. It is unclear what effect the loop has on the evaporation process, and vice versa. The PBH cannot maintain its mass at the minimum value in equation (2.124) by accreting the loop, since this would correspond to the loop losing mass at a rate

$$\frac{\mathrm{d}m}{\mathrm{d}t} \approx \left(\frac{m_{\text{Pl}}}{m_{\text{pbh}}}\right)^2 \frac{m_{\text{Pl}}}{t_{\text{Pl}}} \approx \mu \approx \frac{M_{\text{loop}}}{\ell}, \tag{2.138}$$

i.e. the loop would have to lose all of its mass within a single oscillation period. This seems very unlikely, given the limited gravitational influence of the PBH—the timescale for an object to free-fall from a distance $\sim \ell$ into the PBH is

$$t_{\text{ff}} \sim \frac{\ell^{3/2}}{\sqrt{Gm_{\text{pbh}}}} \approx \frac{\ell}{(G\mu)^{3/2}}, \tag{2.139}$$

so even neglecting the loop's kinetic energy, it would take many oscillation periods for it to be accreted. We are therefore forced to allow the PBH to decay to sizes smaller than the loop width. It is hard to envisage a way for the topologically-stable field configuration around the string to be disrupted by the PBH evaporation, so the most likely outcome seems to be that the PBH simply vanishes from the loop.[14]

PBHs with rest mass larger than $m_* \approx 5 \times 10^{14} \, \text{g} \approx 3 \times 10^{-19} M_\odot$ evaporate very slowly, and lose a negligible fraction of their mass within a Hubble time [176]. It is therefore interesting to consider how these non-evaporating PBHs interact with their parent loops on cosmological timescales. For the simplest case of an infinitely long straight string, explicit solutions for the metric of a BH threaded by a cosmic string have been constructed for the Nambu-Goto case [84] and for Abelian-Higgs strings [45, 148, 187, 234, 286, 292]. For solutions where the BH is rotating, the string is assumed to be aligned with the spin axis. In each case the solution is static, and the string represents a form of stable long-range hair on the BH. Since the string is static, its only gravitational effect is to induce a conical singularity along its axis, with a deficit angle $\sim G\mu$ [281, 564]. This deficit angle means that the string-BH solution is not asymptotically flat, which explains how it evades the no-hair theorem [108, 490]. The deficit angle can also modify the BH's quasi-normal mode spectrum [189, 192].

The relevance of these results is somewhat limited in our case, as the string emanating from the PBH is not static, but continues to oscillate relativistically. Perhaps even more importantly, the string is not locally aligned with the PBH's spin axis, so does not puncture the PBH at its poles like the cases studied in this literature [45, 84, 148, 187, 234, 286, 292]. Instead, due to the geometry of the cusp, the two points where the string punctures the horizon lie in—or very close to—the equatorial plane, and are separated by a small angle $\sim G\mu$ (as we showed in section 2.5.3). Being in the equatorial plane, one would expect relativistic frame-dragging to pull the string into a spiral configuration around the PBH spin axis on scales $\sim Gm_{\text{pbh}}$. (On larger scales $\sim \ell$, the string tension easily overcomes the frame-dragging forces.) This spiralling of the string around the PBH, combined with the very small separation between the two points at which it punctures the horizon, makes it seem likely that the string intersects itself near to the PBH. The PBH would thus be chopped off from the rest of the loop, leaving it with only a small segment of string still attached, which it would rapidly accrete. For sufficiently small loops, we have shown in equation (2.127) that the PBH is likely to be immediately chopped off at the moment of formation.

It is interesting to ask whether two PBHs connected to the same loop could have a greater chance of merging due to the loop dynamics; a similar effect has been demonstrated for the annihilation of monopole-antimonopole pairs connected by strings (so-called 'cosmic necklaces') [114, 515]. However, the two PBHs would likely be separated by a distance $\sim \ell$ much larger than their size, so based on the

---

[14]Bonjour *et al.* [148] and Gregory *et al.* [292] found that Abelian-Higgs string-BH systems can exhibit interesting 'flux expulsion' effects when the BH is smaller than the string width; however, these results are only valid for extremal Kerr and Reissner-Nordström BHs, and it is not clear whether they have any bearing on our sub-extremal PBHs, or on the evaporation process.

discussion above we would expect the PBHs to be chopped off before the loop has the chance to pull them together.

There are clearly many uncertainties in how cusp-collapse PBHs affect the loop network, but our very rough arguments here suggest that small PBHs rapidly evaporate to leave the loop essentially unchanged (although its dynamics are affected by the radiation), while large PBHs are likely to be cut off from loop by string self-intersections on small lengthscales $\sim Gm_{\mathrm{pbh}} \sim (G\mu)^3 \ell$.

2.6 Observational consequences

Having derived the nonlinear GW memory waveforms emitted by cusps and kinks, and having investigated PBH formation as a potential mechanism for curing the memory divergence we encountered for cusps, we are now ready to ask what implications these results have for observations. We begin by looking at the impact that cusp collapse has on the *primary* GW emission from cosmic strings, and how this changes existing bounds on $G\mu$, before exploring the detection prospects of the memory GW emission. We then investigate the mass spectrum of the PBHs formed through cusp collapse, showing that the evaporation of these PBHs leads to novel bounds on $G\mu$ (albeit weaker than those coming from GW observations), and arguing that they form a unique BH population which could act as an observational 'smoking gun' for cosmic strings.

2.6.1 Consequences of cusp collapse for GW searches

As a first approximation, we can model the effects of cusp collapse by transforming the standard waveform (2.36) to the time domain, and truncating it at some time $t_{\mathrm{pbh}} < t_0$ when the horizon forms. In the limit $t_{\mathrm{pbh}} \to t_0$ where the PBH forms at the peak of the cusp signal, the resulting frequency-domain waveform is exactly half of the standard one, with the other half corresponding to the truncated part of the signal at $t \geq t_0$. Based on the discussion around equation (2.135), we expect $t_0 - t_{\mathrm{pbh}} \sim G\mu\ell \ll \ell/2$, so the waveform is truncated slightly before t_0, as shown in figure 2.13. This leads to a loss of power at frequencies above $f_{\mathrm{pbh}} \sim 1/(G\mu\ell)$, which cures the memory divergence, as we saw in section 2.3.9.

A further contribution to the signal comes from the quasinormal ringing of the PBH. We can describe this very approximately by including only the leading-order $(\ell, m, n) = (2, 2, 0)$ quasinormal mode (QNM), writing

$$h(t) \approx C \frac{Gm_{\mathrm{pbh}}}{r} \exp[\mathrm{i}(\omega t + \phi) - t/\tau], \tag{2.140}$$

where C and ϕ are unknown real constants. Using the fitting formulae in Berti *et al.* [120], we take the real and imaginary parts of the $(2, 2, 0)$ QNM for a PBH with spin $\chi = 2/3$ as

$$\omega \approx 0.5214/(Gm_{\mathrm{pbh}}), \quad 1/\tau \approx 0.1715/(Gm_{\mathrm{pbh}}). \tag{2.141}$$

Since $Gm_{\mathrm{pbh}} \approx (G\mu)^3 \ell$, we see that the ringdown signal is associated with extremely high frequencies,

$$\omega \approx 5 \times 10^{24} \, \mathrm{Hz} \times \left(\frac{\ell}{\mathrm{pc}}\right)^{-1} \left(\frac{G\mu}{10^{-11}}\right)^{-3}. \tag{2.142}$$
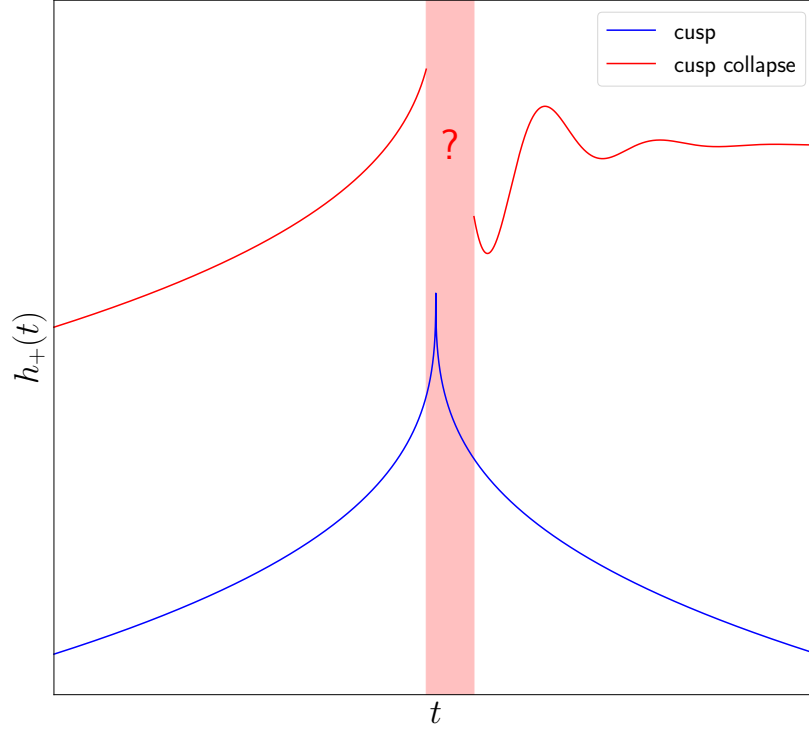
Figure 2.13: A heuristic illustration of the different GW signals from collapsing and non-collapsing cusps. Shown in blue is the standard time-domain cusp waveform (2.36), which is symmetric around the peak. The cusp collapse waveform, in red, is truncated just before the peak, and eventually culminates in the QNM ringing of the final PBH. The uncertain period inbetween is denoted with a question mark. Note that the QNM frequency (2.142) is much higher than depicted here, and that this figure is only for illustrative purposes.

Our ignorance about the exact details of the collapse means that we cannot hope to construct an accurate phase-coherent waveform like those found by Aurrekoetxea *et al.* [88] for circular loop collapse. However, by accounting for the truncation of the cusp signal and the PBH ringdown, we can obtain a reasonable first approximation to the (logarithmic, one-sided) GW energy spectrum,

$$\epsilon_{\text{gw}} \approx \frac{(2/3)^{1/3}}{4} \frac{(\pi A_{\text{c}})^2 G\mu}{(f\ell)^{1/3}} \Theta(1 - G\mu\ell f) + \frac{C^2\ell}{4}(\pi f)^3 (G\mu)^5 \left[\mathcal{L}^2\left(2\pi f; \omega, \tfrac{1}{\tau}\right) + \mathcal{L}^2\left(2\pi f; -\omega, \tfrac{1}{\tau}\right) \right]. \tag{2.143}$$

The first term here is reduced by a factor $1/4$ compared to equation (2.38) (due to a factor $1/2$ in each power of the strain) and is truncated at $f \sim 1/(G\mu\ell)$, while the second part is the ringdown contribution, written in terms of the Lorentzian

$$\mathcal{L}(x; x_0, \gamma) \equiv \frac{\gamma/\pi}{\gamma^2 + (x - x_0)^2}. \tag{2.144}$$

We can fix the constant C by setting the total energy radiated by the ringdown term equal to $G\mu\epsilon_M$, where ϵ_M is the collapse radiation efficiency introduced in equation (2.136), and the factor $G\mu$ translates
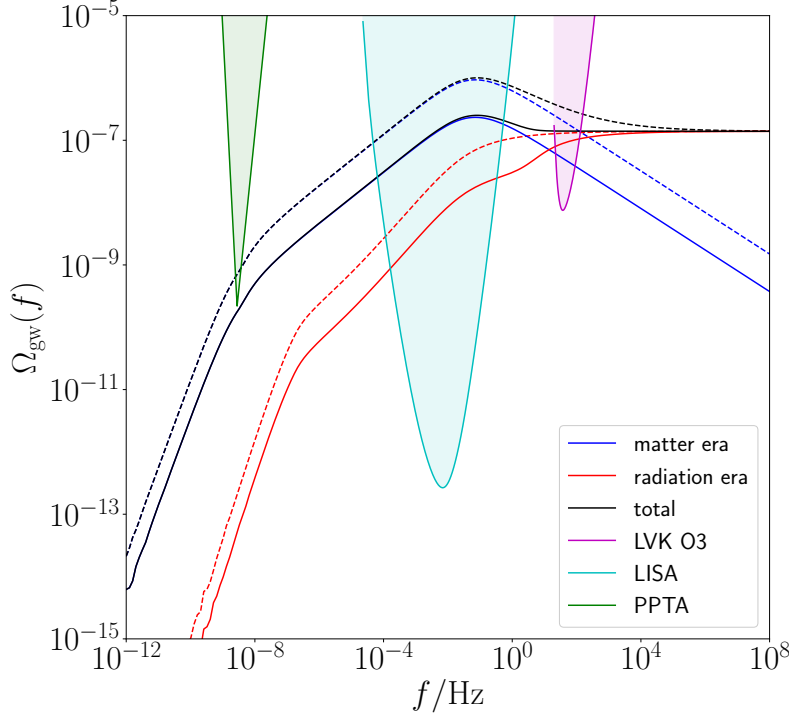
Figure 2.14: The GWB spectrum from cusps on cosmic string loops. Solid lines include the effects of cusp collapse using equation (2.143), while dotted lines correspond to the standard case without collapse (2.38). The magenta, cyan, and green curves show the power-law integrated sensitivities of LIGO/Virgo's third observing run [32, 36, 37], the Parkes PTA [512, 562], and LISA [171, 520], respectively. We use 'model 3' of the loop network [396, 483] with $G\mu = 3 \times 10^{-11}$, illustrating how the PPTA bound is weakened due to cusp collapse. At high frequencies the spectra with and without cusp collapse become identical; the frequency at which this changeover occurs decreases for smaller values of $G\mu$, meaning that the LIGO/Virgo bound on model 3 is the same in both cases.

between the loop's mass and the PBH's relativistic mass (i.e. rest mass plus kinetic energy), with ϵ_M defined as a fraction of the latter. This gives

$$C = \sqrt{\frac{64\pi\epsilon_M}{1+\omega^2\tau^2}\frac{\tau/Gm_{\mathrm{pbh}}}{G\mu}} \approx 10.70 \times (\epsilon_M/G\mu)^{1/2}. \tag{2.145}$$

We assume a value of $\epsilon_M = 10\%$, which is consistent with the upper bounds we found in section 2.5.6, and is comparable to the mass-radiation fraction found in numerical simulations of other ultrarelativistic, strong-gravity phenomena [248, 522]. The total fraction of the loop's mass radiated by the cusp is approximately

$$\int_{2/\ell}^{\infty} \frac{\mathrm{d}f}{f}\epsilon_{\mathrm{gw}} \approx \begin{cases} 14.9\,G\mu & \text{cusp,} \\ (3.71 + \epsilon_M)\,G\mu & \text{cusp collapse,} \end{cases} \tag{2.146}$$

which shows that the radiation from collapsing cusps is comparable to, but strictly less than, that from non-collapsing cusps.

We can account for the effect of cusp collapse on the GWB from cosmic string loops by replacing the GW energy spectrum used in equation (2.49) with that given here in equation (2.143). A representative

example of the resulting GWB spectrum is shown in figure 2.14. At low frequencies Ω_{gw} is reduced by a factor of $1/4$ compared to the standard spectrum, which relaxes the constraints on $G\mu$ coming from LIGO/Virgo [24] and from PTAs [131, 388, 512, 562]. At high frequencies this factor $1/4$ difference vanishes, as the signal is dominated by loops which are too small to undergo cusp collapse. The changeover between these two regimes depends on the value of $G\mu$, as this sets the size of the smallest loops that can undergo cusp collapse through equation (2.123); for smaller values of $G\mu$ the changeover happens at lower frequencies. At very high frequencies the QNM emission from PBHs forming in the matter era gives rise to a strong peak, but this is dwarfed by the radiation-era plateau, making it unobservable (regardless, these frequencies are well beyond the reach of any current or planned GW experiment).

2.6.2 DETECTION PROSPECTS FOR THE NONLINEAR MEMORY

Having calculated the nonlinear GW memory waveforms associated with cusps and kinks, it is natural to ask whether these signals are detectable with current or future GW observatories. Clearly the divergent behaviour diagnosed for cusps in section 2.3.7 could, in principle, have important observational implications, depending on how the divergence is regulated. For the purposes of this section we assume the divergence is resolved along the lines of the cusp-collapse scenario described in section 2.5. We show below that, under this assumption, the cusp and kink memory signals are suppressed so strongly that they are well beyond the reach of GW observatories, even future third-generation interferometers like Einstein Telescope (ET) [464] and Cosmic Explorer (CE) [472].

We start by calculating the expected detection horizons for individual bursts of GW memory from cusps and kinks—i.e. the maximum distance at which a burst can be detected, on average. We assume a matched-filter search, such that the optimal root-mean-square SNR (averaging over sky location and polarisation angle) for a frequency-domain waveform $\tilde{h}(f)$ which is isotropically averaged over the source inclination is given by [406]

$$\rho_{\mathrm{rms}} = \left[\sum_I \frac{4}{5} \sin^2 \alpha_I \int_0^\infty \mathrm{d}f \, \frac{|\tilde{h}(f)|^2}{P_I(f)} \right]^{1/2}.$$ (2.147)

The sum here is over different GW detectors, with $P_I(f)$ representing the noise power spectral density (PSD) of detector I, and α_I the opening angle between the two interferometer arms (this angle enters through the detector's response function; LIGO, Virgo, and CE have an opening angle of $\pi/2$, while each of ET's three interferometers has an opening angle of $\pi/3$). It is convenient to rewrite this in terms of the fractional energy spectrum, using equation (2.37) to give

$$\rho_{\mathrm{rms}} = \left[\sum_I \frac{2G\mu\ell}{5\pi^2 r^2} \sin^2 \alpha_I \int_0^\infty \mathrm{d}f \, \frac{\bar{\epsilon}(f_{\mathrm{s}})}{f_{\mathrm{s}}^3 P_I(f)} \right]^{1/2},$$ (2.148)

where $r(z)$ is now the comoving distance to the source, and $f_{\mathrm{s}} = (1+z)f$ is the source-frame frequency, with z the redshift. We assume that any cosmic string signal with $\rho_{\mathrm{rms}} \geq 12$ can be confidently detected; in reality, this threshold depends on the distribution of non-Gaussian noise transients ('glitches') in the
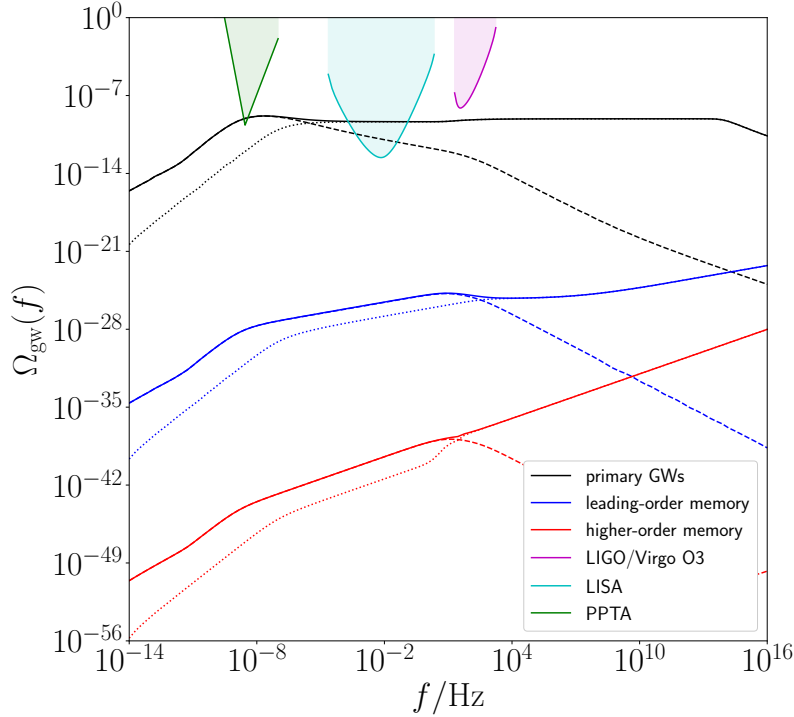
Figure 2.15: Contributions to the GWB energy spectrum $\Omega_{\mathrm{gw}}(f)$ from cosmic strings at different orders in the memory expansion, assuming that the memory divergence is resolved through the cusp-collapse scenario. We see that the memory effect is negligible compared to the primary emission. Here we assume 'model 2' [130, 131] of the string network, and a string tension of $G\mu = 8.1 \times 10^{-11}$ (this is the largest string tension allowed by current constraints in the cusp-collapse scenario [347]; smaller tensions suppress the memory effect even further). The dashed/dotted curves show the contributions from the matter/radiation era, respectively, with the solid curves showing the combined spectra. A distinct change in all three of the radiation-era spectra is visible at around $f \approx 10\,$Hz; at frequencies above this, the spectra are increasingly dominated by small loops which do not undergo cusp-collapse, and this is why the memory effect becomes more prominent (though still undetectable). The sensitivity curves are the same as in figure 2.14.

network, and how closely these are able to mimic the waveforms of interest, but we ignore these details here.

Even assuming an optimistic third-generation GW detector network consisting of Einstein Telescope plus two Cosmic Explorers,[15] we find that given current constraints on the string tension ($G\mu \lesssim 10^{-11}$), the detection horizon for a cusp memory signal is at most $\approx 2\,$pc. This corresponds to a negligibly small detection rate, as very few cosmic strings are expected within a volume of this size. For kink memory the result is even more pessimistic, with a detection horizon of $\approx 0.008\,$AU (a few times larger than the Earth-Moon distance). Larger values of the string tension $G\mu$ would boost the detectability of these memory bursts, but would be in conflict with existing observational results.

In figure 2.15 we show the GWB spectrum from a particular model of the loop network, including the contributions from first-order and higher-order GW memory from cusps and kinks. We see that while the primary GWB reaches a plateau at high frequencies, the memory contributions to the GWB grow with frequency above $\approx 10\,$Hz, which makes sense given the slower fall-off of the cusp memory emission

---

[15] For ET we use the 'ET-D' noise PSD [313], while for CE we use the 'CE-2' noise PSD [295].

at high frequency compared to the primary cusp and kink signals. However, each order in memory is suppressed by an additional factor of $(G\mu)^2$, and this suppression is strong enough to render the memory contribution unobservable at all frequencies.

2.6.3 THE PBH MASS SPECTRUM

We argued in section 2.1.2 that cosmic string loops typically form $N_c \approx 1$ cusp per oscillation period, which means that cusp-collapse PBHs are continuously created by the loop network, from the very early Universe to the present day, resulting in a broad distribution of PBH masses. This contrasts sharply with the standard PBH formation scenario we discussed in section 0.3.4, where the collapse typically occurs at a single early epoch, resulting in a monochromatic PBH mass spectrum.

We can write the comoving number density of cusp-collapse PBHs with rest mass between m and $m + \mathrm{d}m$ as

$$n_{\mathrm{pbh}}(m, t)\, \mathrm{d}m = \int_0^t \mathrm{d}t'\, \frac{2N_c}{\ell_m} n_{\mathrm{loop}}(\ell_m, t')\, \mathrm{d}\ell_m\,, \qquad (2.149)$$

where $n_{\mathrm{loop}}(\ell, t) = (a^3/t^4)\mathcal{F}(\gamma)$ is loop distribution function we encountered in section 2.1.3, and $\ell_m \approx Gm/(G\mu)^3$ is the loop length required to form a PBH of mass m. We assume $N_c = 1$, consistent with much of the literature on cosmic string phenomenology, particularly regarding GW searches [18, 87]. The factor of $2/\ell_m$ here accounts for the oscillation period of the loop which forms the PBH.

Since ℓ_m corresponds to a fixed physical (rather than comoving) scale, PBH production only begins once this scale has entered the horizon. This happens at cosmic time

$$t_i(m) \approx \ell_m \approx 16\,\mathrm{Gyr} \times \frac{m}{10^{-10}M_\odot} \times \left(\frac{G\mu}{10^{-11}}\right)^{-3}, \qquad (2.150)$$

which means that the largest PBHs form at the present day, with mass

$$m_{\mathrm{max}} \approx 0.88 \times 10^{-10}\, M_\odot \times \left(\frac{G\mu}{10^{-11}}\right)^3. \qquad (2.151)$$

Coincidentally, this corresponds to the 'sublunar' mass range—one of the few regimes where there are no constraints preventing PBHs from making up the totality of DM [176, 177] (see figure 0.13). However, the majority of cusp-collapse PBHs have masses much smaller than this.

Inserting equation (2.149) into the expression (0.104) we introduced in section 0.3.4, we find that the mass spectrum of cusp-collapse PBHs is

$$f(m) = \frac{2N_c}{\rho_{\mathrm{cdm}}} \int_0^{t_0} \mathrm{d}t\, \frac{a^3(t)}{t^4} \mathcal{F}\left(\frac{Gm}{(G\mu)^3 t}\right), \qquad (2.152)$$

which is shown in figure 2.16 for a representative value of $G\mu$. In general the integral in equation (2.152) is broken into two parts, corresponding to the different scaling solutions in the matter and radiation eras. Note that equation (2.152) includes only the energy density due to the rest mass of the PBHs; their kinetic energy at formation is larger by a factor of $1/(G\mu)$, though this will eventually be redshifted away. This large kinetic energy will likely lead to interesting phenomenology and constraints which are not
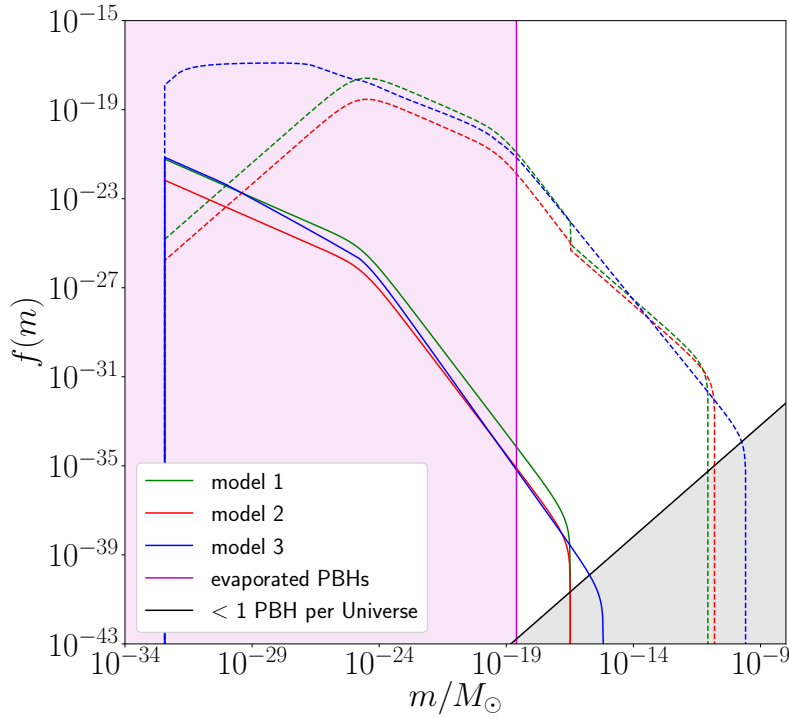
Figure 2.16: The present-day PBH mass spectrum (2.152) for the three standard models of the cosmic string loop network with $G\mu = 10^{-11}$. The solid and dashed lines correspond to PBHs formed in the radiation and matter eras respectively. The cutoffs at small and large masses are given by equations (2.124) and (2.151). The magenta region represents PBHs which have evaporated by the present day. The grey region corresponds to there being less than one PBH of that mass in the observable Universe.

captured by traditional PBH analyses, as these generally assume the PBHs are formed with negligible peculiar velocities. We plan to explore this in future work.

Equation (2.152) does not include evaporation due to Hawking radiation, and is therefore only valid for masses greater than $m_* \approx 5 \times 10^{14}\,\mathrm{g} \approx 3 \times 10^{-19} M_\odot$, with PBHs lighter than this evaporating in less than a Hubble time [176]. Nonetheless, the form of $f(m)$ for masses below m_* can be useful for deriving constraints on the overall mass spectrum due to evaporation effects. For small masses in the radiation era, equation (2.152) approaches a time-independent power law $f(m) = f_*(m/m_*)^{-1/2}$, with f_* a constant depending on $G\mu$ and on the network model. The negative exponent means that the mass spectrum is dominated by very small PBHs, and that the strongest constraints on cusp collapse come from their evaporation. In fact, this is the same power law as the mass spectrum resulting from the collapse of circular loops [337, 400], but with a different pre-factor. In the circular collapse case, the pre-factor depends on the fraction of circular loops, which is unknown; in our case, the pre-factor depends only on

$G\mu$, which we can therefore constrain directly. Using the most up-to-date constraints from James-Turner *et al.* [337], we find[16]

$$G\mu \lesssim \begin{cases} 6.0 \times 10^{-7} & \text{for model 1,} \\ 1.2 \times 10^{-6} & \text{for model 2,} \\ 9.3 \times 10^{-7} & \text{for model 3,} \end{cases} \tag{2.153}$$

which in turn gives a constraint on the total fraction of DM made up by cusp-collapse PBHs,

$$\frac{\rho_{\text{pbh}}}{\rho_{\text{cdm}}} \equiv \int_{m_*}^{\infty} \mathrm{d}m \, \frac{f(m)}{m} \lesssim \begin{cases} 2.0 \times 10^{-10} & \text{for model 1,} \\ 7.4 \times 10^{-11} & \text{for model 2,} \\ 8.0 \times 10^{-10} & \text{for model 3.} \end{cases} \tag{2.154}$$

The constraint (2.153) on $G\mu$ is comparable to those set by CMB analyses [47, 419], and is almost independent of the network model. This constraint is set by the damping of small-scale CMB anisotropies due to PBHs decaying at recombination [182, 597], and is orders of magnitude stronger than the γ-ray constraint [337]. It is likely to become more stringent with future CMB missions, and with similar analyses from upcoming 21 cm intensity-mapping experiments [398, 527]. For now, however, it is unable to compete with the constraints on the order of $G\mu \lesssim 10^{-11}$ coming from GW observations.

2.6.4 A unique BH population

We seen in section 2.5.3 that cusp-collapse PBHs are universally formed with dimensionless spins of $\chi = 2/3$, regardless of the loop size ℓ or the string tension $G\mu$.[17] For sufficiently large rest masses $m \gg m_*$, this initial spin value is not affected by Hawking radiation, and survives to the present day [79]. This is interesting because it means that cusp-collapse PBHs occupy a unique region of the 'Regge plane' [83, 118, 155] (i.e., the BH mass-spin parameter space), as we illustrate in Fig 2.17.

Spins of $\chi \approx 2/3$ are common amongst astrophysical BHs. In particular, this is a very natural value for BHs formed from binary mergers like those observed by LIGO/Virgo [22]; the majority of such binaries are approximately equal-mass with small initial spins [23, 269], which correspond to final spins of $\chi \approx 0.687$ [479]. SMBHs in active galactic nuclei are also observed to have large spins $\chi \gtrsim 0.6$, due to accretion and prior mergers [154]. However, such astrophysical processes are unable to create subsolar-mass BHs, which dominate the cusp-collapse PBH mass spectrum for realistic values of $G\mu$, cf. figure 2.16.

On the other hand, subsolar masses are generally possible in other PBH formation mechanisms, but these mechanisms are unable to generate spins $\chi \approx 2/3$ like those resulting from cusp collapse. 'Conventional' PBHs formed from collapsing overdensities during radiation domination are typically born with small spins of order $\chi \sim 0.01$ [193, 229, 428]. These initially low-spinning PBHs can acquire large spins through accretion, saturating the Thorne bound $\chi \approx 0.998$ [541], but this process only takes place

---

[16] These constraints are phrased in terms of a normalisation constant c_{string}, with the CMB constraint giving $c_{\text{string}} < 2 \times 10^{-12}$. This can be translated to our mass spectrum using $c_{\text{string}} = 2f_* \Omega_{\text{cdm}}$.

[17] Note that this is the 'naïve' zero-radiation value, and that a fully general-relativistic calculation would likely give a different value for the final spin. However, our argument in this section still holds, provided that the true value of χ is significantly larger than zero and less than unity.
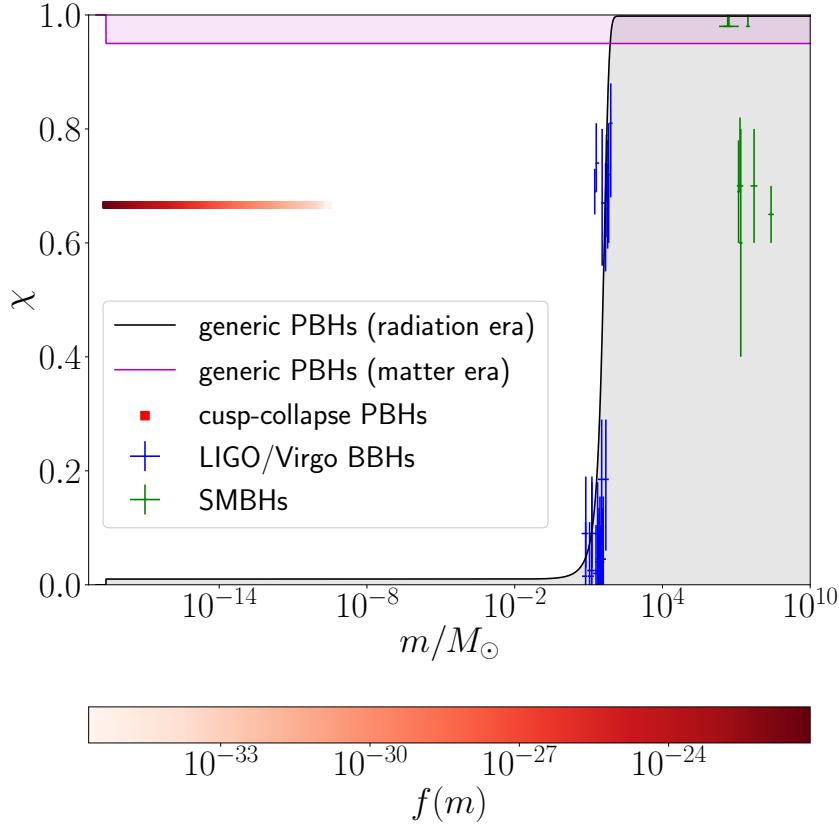
Figure 2.17: The location of various primordial and astrophysical BH populations in the Regge plane (BH mass-spin parameter space). Blue crosses show the initial and final BHs for each of the ten binary BH mergers in LIGO/Virgo's first GW Transient Catalogue (GWTC-1) [22]; the spin distribution is noticeably bimodal, with initial BHs having low spins $\chi \lesssim 0.2$ and final BHs have large spins $\chi \approx 0.7$. (The spins of the initial BHs are not confidently measured in GWTC-1, so the spin values and uncertainties plotted here are merely heuristic and are estimated from the inspiralling binary's effective aligned spin χ_{eff}.) Green crosses show the SMBHs catalogued in Brenneman *et al.* [154]. The grey region shows the expected parameter space for 'conventional' PBHs formed from overdensities collapsing during radiation domination, including the effects of accretion as calculated in De Luca *et al.* [230]. The magenta region shows a possible population of near-extremal PBHs formed during a period of early matter domination. The red region shows cusp-collapse PBHs, shaded according to their mass spectrum for $G\mu = 10^{-11}$, cf. figure 2.16. All of the PBH populations are cut off at $m_* \approx 3 \times 10^{-19} M_\odot$, due to evaporation through Hawking radiation.

within a Hubble time if the PBHs in question are sufficiently massive, $m \gtrsim 50\,M_\odot$ [230]. Subsolar-mass PBHs are extremely inefficient at accreting matter, and remain essentially non-spinning.

Subsolar-mass PBHs can have large spins if they form from collapsing overdensities during a hypothetical period of early matter domination [179, 299, 385, 462], as radiation pressure is then unable to dissipate angular momentum during the collapse. However, these PBHs are expected to have near-extremal spins $\chi \approx 1$, which are easily distinguishable from the $\chi = 2/3$ prediction of cusp collapse.

We therefore see that any observation of a subsolar-mass BH with a large (but non-extremal) spin $\chi \approx 2/3$ would be incompatible with any of the other BH formation mechanisms mentioned here, and would be a 'smoking gun' signature of cusp collapse, and of cosmic strings more generally.

2.7 Summary and outlook

In this chapter, we have thoroughly explored the nonlinear GW memory associated with cusps and kinks on Nambu-Goto cosmic strings, deriving detailed analytical waveform models for the memory GWs, including the 'memory of the memory' and other higher-order memory effects. These are among the first memory observables computed for a cosmological source of GWs, with previous literature having focused almost entirely on astrophysical sources.

The leading-order cusp memory waveform (2.64) that we have found is strikingly similar to the primary GW signal from the cusp (2.36), with the same characteristic $\sim f^{-4/3}$ frequency power-law. However, one very important difference is that this memory signal is emitted in all directions, unlike the primary signal which is confined to a narrow beam of width $\theta_{\rm b} \sim f^{-1/3}$. As a result, we find that the total GW energy radiated by the cusp memory *diverges* for a Nambu-Goto (i.e., zero-width) loop. This divergence can be regularised by introducing a cutoff at the scale of the string width $\delta \sim \ell_{\rm Pl}/\sqrt{G\mu}$, but this then introduces powers of $\ell/\delta \gg 1$ to the higher-order memory terms, causing the sum of all memory contributions to diverge for loops of length $\ell \gtrsim \delta/(G\mu)^3$.

In section 2.3.9 we have argued that the most plausible explanation for this unphysical memory divergence is the assumption that the spacetime containing the cosmic string loop is well-described by a flat background with linear perturbations. We have suggested that some strong-gravity mechanism must kick in on loops of length $\ell \gtrsim \delta/(G\mu)^3$ to suppress the high-frequency GW emission from cusps, thereby curing the divergence. In section 2.5, we have investigated exactly such a mechanism: the collapse of a small cosmic string segment near the cusp to form a PBH. Remarkably, this cusp-collapse process is predicted to occur for all loops of length $\ell \gtrsim \delta/(G\mu)^3$—exactly the same loops for which we have diagnosed the memory divergence. We have shown explicitly that if cusp collapse does indeed occur for these loops, the corresponding GW memory is strongly suppressed, and the divergence is cured.

We have also calculated the memory emission associated with kinks, and have shown that this is suppressed due to interference between GWs emitted by the kink at different points in its history. There is thus no situation in which the kink memory signal diverges; this accords with the cusp-collapse description, as kinks are not predicted to form PBHs in that scenario.

In section 2.6 we have investigated the observational consequences of our results, focusing on the implications for the GW background. We find that by requiring the string tension to agree with existing observational bounds, and by invoking cusp collapse to prevent the memory from diverging, the resulting memory signal is very strongly suppressed, and is not likely to be detected by any current or upcoming GW observatories. Meanwhile, the primary signal from cusps is reduced by a factor of $\approx 1/4$ at low frequencies due to cusp collapse, slightly weakening existing bounds on the string tension. Of course, if cusp collapse does not occur, then it is possible that large loops could source much stronger memory signals; however, one would then need an alternative mechanism for resolving the cusp divergence.

One promising avenue for testing these ideas lies in observing the PBHs themselves. By calculating the angular momentum of the string segment captured behind the horizon, we have shown that cusp-collapse PBHs are highly spinning, with dimensionless spin parameter equal to two-thirds of the extremal Kerr value, $\chi = 2/3$. This spin is a universal property of the formation mechanism, and is independent of

the loop size ℓ and string tension $G\mu$. To the best of our knowledge, cusp collapse is the only known primordial or astrophysical mechanism for generating subsolar-mass BHs with large but sub-extremal spins. The observation of such a BH would therefore be a 'smoking gun' signal of cusp collapse, and of cosmic strings more generally. The fact that cusp-collapse PBHs are born with ultrarelativistic velocities will undoubtedly also give rise to some interesting phenomenology, and may allow us to place new constraints on their abundance, aside from those in the standard PBH literature. It would also be very interesting to calculate the merger rate of cusp-collapse PBH binaries, as well as the corresponding GWB spectrum [104, 207, 411, 467, 468, 579], as consistency with LIGO/Virgo observations (in particular the subsolar-mass search [3, 21, 26, 404]) would provide another independent constraint on the string tension; we leave this for future work.

Our work in this chapter demonstrates the importance of considering the nonlinear memory associated with a broader class of GW sources than just compact binaries; we have shown that the memory effect is interesting not just from an observational point of view, but also as a tool for sharpening our theoretical understanding and modelling of said GW sources. As an unexpected by-product of our analysis, we have shown that the standard Nambu-Goto description of cusps is unphysical, and that strong gravity effects (possibly including PBH formation) could play an important rôle in a more complete understanding of their dynamics. This motivates further work to better understand cusps in full GR, including nonlinear gravitational effects beyond those considered here, and thereby understand whether cusp collapse does indeed take place, with the ultimate goal of computing reliable waveform predictions for GW observatories. It would also be very interesting to study the *linear* memory associated with cusps, and to understand how this contributes to the total GW emission; however, such a study would require us to go beyond the Nambu-Goto approach adopted here. More generally, our results motivate a broader examination of the nonlinear memory effect in GW astronomy and cosmology, to see what other surprises may be in store.

3   BINARY SYSTEMS AS DYNAMICAL GRAVITATIONAL-WAVE DETECTORS

In section 0.4 we discussed several current and near-future experimental efforts to detect gravitational waves. These experiments each hold enormous scientific potential, and together will probe a dizzying range of exotic astrophysical and cosmological phenomena throughout the history of the Universe, including those we covered in section 0.3. However, we saw how practical and technical limitations restrict the sensitivity of each experiment to a narrow frequency band, leaving broad swathes of the GW frequency spectrum essentially unexplored, as shown in figure 0.14. These gaps in the GW spectrum could contain signals that are inaccessible to any current or planned GW observatory. Perhaps the clearest example is the stochastic GW signal expected from a first-order phase transition: as we discussed in section 0.3.2, this signal is sharply peaked and could, for a broad range of the physical parameter space, be missed by all of the GW experiments we have described. FOPTs aside, there is also the distinct possibility of these unexplored frequency bands containing completely unexpected signals, beyond just those models that have been proposed in the literature. Discovering one of these 'unknown unknowns' could revolutionise our understanding of the Universe.

This problem motivates us to explore alternative experimental techniques that can bridge the observational gaps that exist in the GW frequency spectrum. In this chapter we explore one such possibility, which is to search for GWs by studying their influence on binary systems. Rather than searching for oscillations in the proper distance between the test masses in an interferometer (as in LIGO/Virgo/KAGRA, etc.), or between pulsars and the Earth (as in PTAs), we consider here the GW-induced oscillations between two freely-falling astronomical bodies in a gravitationally-bound orbit. While these oscillations are extremely challenging to observe *directly* for any realistic binary, we show below that they can leave lasting imprints on the binary's orbit—particularly if they occur at an integer multiple of the binary's orbital frequency, as this causes the perturbations to be resonantly amplified (as we will see below). For persistent GW signals, these imprints accumulate over time, eventually giving rise to observable deviations which can be used to infer the GW amplitude, turning the binary into a *dynamical GW detector* (see figure 3.1). This idea has a long history [122, 194–197, 277, 328, 329, 391, 416, 417, 429, 434, 489, 553], and has been used to search for GWs with the orbit of the binary pulsar B1913+16 [323]. Similar ideas have also been used to search for orbital changes induced by coherently oscillating ultralight DM fields [80, 81, 139, 140, 236, 395]. Nonetheless, this 'binary resonance' effect has received relatively little attention in the GW community, and has not yet been exploited to its full potential.

In this chapter we develop, from first principles, a new formalism for calculating the GW-induced evolution of a generic binary system. While we could, in principle, use binary resonance to search for any GW signal morphology, there are two practical considerations which lead us to narrow our scope here:
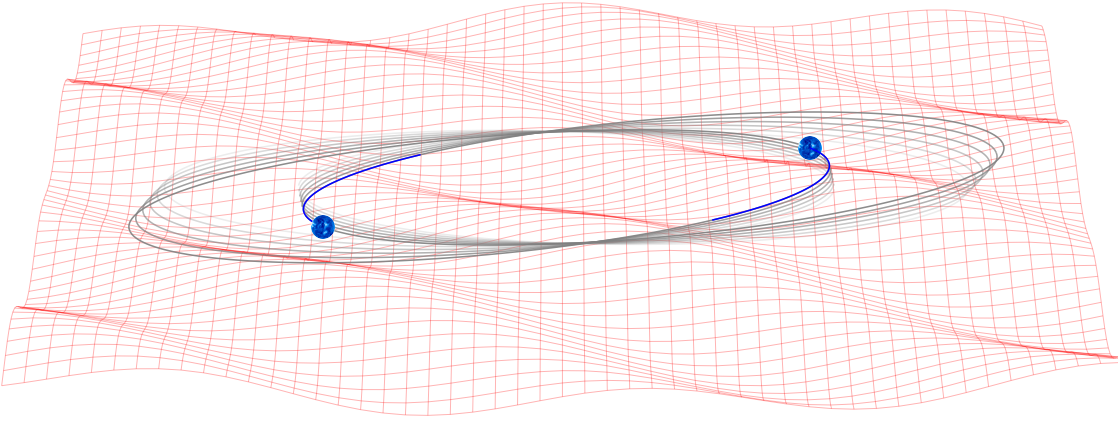
Figure 3.1: Cartoon illustration of the binary resonance effect. Stochastic fluctuations in the background spacetime geometry due to incoming GWs perturb the trajectories of two orbiting masses, causing cumulative changes to their orbital elements.

1. We require the GW signal to be *persistent*, so that the GW-induced perturbations to the binary's orbit can accumulate over time and eventually reach a detectable level.

2. We neglect *narrowband* signals whose power is concentrated in a relatively small range of frequencies, since the probability of the signal's frequency band overlapping with one of the resonant frequencies of the binary is then too small for the searches to have much impact.

As we saw in section 0.2, the Universe provides us with a persistent, broadband GW signal that satisfies these conditions: the stochastic GW background. We therefore focus on the response of a binary system to the GWB.

The stochastic nature of the signal means that we cannot deterministically calculate the evolution of a binary coupled to the GWB; instead we treat each of the binary's orbital elements as a time-dependent random variable, and study how the statistics of the orbital perturbations evolve over time. Our key result is a secularly-averaged *Fokker-Planck equation* (FPE), which improves upon previous results by fully specifying the evolution of the joint probability distribution for all six orbital elements. (Compare with Hui *et al.* [323], who calculated only the variance of the distribution, and focused on just one orbital element: the period.) By comparing solutions of this FPE with high-precision orbital data from various binary systems, one can place stringent new constraints on the GWB at frequencies that are inaccessible to all other current and future GW observatories.

The remainder of this chapter is structured as follows. In section 3.1 we give a brief, self-contained overview of Keplerian orbits, orbital perturbations, and the formalism of osculating orbital elements. This section is a review of standard results, focusing only on the tools we need to tackle the binary resonance problem; for more details on the various points we discuss, we refer the reader to, e.g., Brouwer and Clemence [156], Murray and Dermott [432], or Poisson and Will [459]. In section 3.2 we specialise this formalism to perturbations from the GWB, and develop a FPE for the orbital elements, expressing the coefficients of the equation in terms of GW transfer functions. In section 3.3 we write these coefficients explicitly as functions of the orbital elements, and briefly discuss their properties (the technical details of this calculation are given in appendix D). In section 3.4 we obtain some exact late-time results for the

case where the binary's eccentricity is held fixed at zero, which greatly simplifies the FPE. In section 3.5 we tackle the more complicated general-eccentricity case, consider practical approaches for solving the FPE on observational timescales, and present some example results for the Hulse-Taylor binary pulsar B1913+16. In section 3.6 we develop the necessary statistical formalism to search for GWB-induced orbital perturbations in observational data, discussing how to compute upper limits and sensitivity forecasts, and how to apply these tools to pulsar timing and laser-ranging experiments. In section 3.7 we use this formalism to calculate forecast constraints on the GWB spectrum from our methods. We show that binary resonance searches can fill the 'µHz gap' between LISA and PTAs, probing a unique region of the FOPT parameter space, and potentially helping to shed light on the common process signal detected by NANOGrav.

3.1 Binary dynamics

In this section we introduce the machinery of osculating orbital elements—an extremely useful formalism for describing perturbations to binary orbits. We start by recalling the basic properties of Keplerian orbits (i.e., orbits of two point masses interacting only through Newtonian gravity), before introducing the equations of motion (EoMs) for the osculating elements, and discussing the most important contribution to these EoMs from relativistic effects. We also introduce alternative sets of orbital elements that are useful in cases where the eccentricity or inclination of the orbit are small.

3.1.1 Keplerian orbits

We start with a Keplerian binary, working in cylindrical coordinates $(\hat{\boldsymbol{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\ell}})$ in a frame with the centre of mass fixed at the origin. We also introduce a fixed Cartesian reference frame $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$, such that the line of sight of the observer is in the positive $\hat{\boldsymbol{z}}$ direction. The unperturbed EoM for the separation vector $\boldsymbol{r}$ is

$$\ddot{\boldsymbol{r}} + \frac{GM}{r^2}\hat{\boldsymbol{r}} = \boldsymbol{0}, \tag{3.1}$$

where $r \equiv |\boldsymbol{r}|$ is the radial separation, $\hat{\boldsymbol{r}} \equiv \boldsymbol{r}/r$ is the radial unit vector, and $M \equiv m_1 + m_2$ is the total mass. We define the energy and angular momentum of the binary (in units of the reduced mass $\mu \equiv m_1 m_2/M$) by

$$\mathcal{E} = \frac{1}{2}\dot{\boldsymbol{r}} \cdot \dot{\boldsymbol{r}} - \frac{GM}{r}, \qquad \boldsymbol{\ell} = \boldsymbol{r} \times \dot{\boldsymbol{r}} = r^2 \dot{\theta}\,\hat{\boldsymbol{\ell}}. \tag{3.2}$$

For the total angular momentum we write $\ell \equiv |\boldsymbol{\ell}| = r^2 \dot{\theta}$. These are all conserved under the Newtonian EoM (3.1), since

$$\dot{\mathcal{E}} = \dot{\boldsymbol{r}} \cdot \left(\ddot{\boldsymbol{r}} + \frac{GM}{r^2}\hat{\boldsymbol{r}}\right) = 0, \qquad \dot{\boldsymbol{\ell}} = \boldsymbol{r} \times \ddot{\boldsymbol{r}} = \boldsymbol{r} \times \left(-\frac{GM}{r^2}\hat{\boldsymbol{r}}\right) = \boldsymbol{0}, \tag{3.3}$$

where we used $\dot{r} = \dot{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}}$. The fact that $\dot{\boldsymbol{\ell}} = \boldsymbol{0}$ means that the binary orbit is confined to a fixed 2D plane, which is specified with respect to the $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ reference frame by two angles: the inclination I, which is the angle between the binary's angular momentum vector $\boldsymbol{\ell}$ and the observer's line of sight $\hat{\boldsymbol{z}}$, and the

longitude of ascending node Ω, which is the angle between $\hat{\boldsymbol{x}}$ and the point where the orbit passes through the reference plane with positive velocity in the $\hat{\boldsymbol{z}}$ direction. (These angles are illustrated in figure 3.2.)

We can find the shape of the orbit in this plane by integrating equation (3.1), yielding a family of elliptical solutions

$$r = \frac{\ell^2/(GM)}{1 + e \cos \psi}, \qquad \mathcal{E} = -\frac{GM}{2a}, \qquad \ell = \sqrt{GMa(1 - e^2)}, \tag{3.4}$$

with the shape of the ellipse described by its semi-major axis a and eccentricity e. Here we have introduced the true anomaly ψ as the angular position of the orbit within the orbital plane, defined such that the pericentre (minimum separation) occurs at $\psi = 0$. The point at which this occurs is defined by the argument of pericentre $\omega \equiv \theta - \psi$, which is measured relative to the ascending node. Since ω is constant, the orbit is closed and the motion of the binary is periodic in time, with period P related to the semi-major axis by Kepler's third law,

$$\frac{GM}{a^3} = \left(\frac{2\pi}{P}\right)^2. \tag{3.5}$$

In what follows, we work entirely in terms of the period rather than the semi-major axis, as the former is more closely linked to the resonant frequencies of the orbit.

The five constants $(P, e, I, \Omega, \omega)$ are almost enough information to specify a particular Keplerian orbit; all that remains is to specify the time at which the binary is at pericentre, t_0. In practice, it is more convenient to replace t_0 with the compensated mean anomaly,

$$\varepsilon \equiv \frac{2\pi}{P}(t - t_0) - \int_0^t \mathrm{d}t' \, \frac{2\pi}{P(t')}, \tag{3.6}$$

as this is more well-behaved when the orbit is perturbed [156]. Note that in the absence of perturbations this reduces to $\varepsilon = -2\pi t_0/P$. We call the set $(P, e, I, \Omega, \omega, \varepsilon)$ the *orbital elements* of a binary.

3.1.2 Perturbations and osculating orbits

Now suppose the binary is acted upon by some small perturbing force [140, 161, 323, 432]

$$\ddot{\boldsymbol{r}} + \frac{GM}{r^2}\hat{\boldsymbol{r}} = \delta\ddot{\boldsymbol{r}} = r(\mathcal{F}_r\hat{\boldsymbol{r}} + \mathcal{F}_\theta\hat{\boldsymbol{\theta}} + \mathcal{F}_\ell\hat{\boldsymbol{\ell}}). \tag{3.7}$$

(The factor of r here is included for later convenience, as the GW perturbations we consider are proportional to the orbital separation.) Inserting this into equation (3.3) gives

$$\dot{\mathcal{E}} = \dot{\boldsymbol{r}} \cdot \delta\ddot{\boldsymbol{r}} = r\dot{r}\mathcal{F}_r + r^2\dot{\theta}\mathcal{F}_\theta, \qquad \dot{\boldsymbol{\ell}} = \boldsymbol{r} \times \delta\ddot{\boldsymbol{r}} = r^2\mathcal{F}_\theta\hat{\boldsymbol{\ell}} - r^2\mathcal{F}_\ell\hat{\boldsymbol{\theta}}, \tag{3.8}$$

so the binary's energy and angular momentum are no longer constant. As a result, the binary is no longer described by a fixed set of orbital elements. However, the orbit is still tangent to some Keplerian ellipse at each moment in time. We therefore define $(P, e, I, \Omega, \omega, \varepsilon)$ as functions of time which track the evolution of this tangent ellipse; these are called the *osculating* orbital elements.
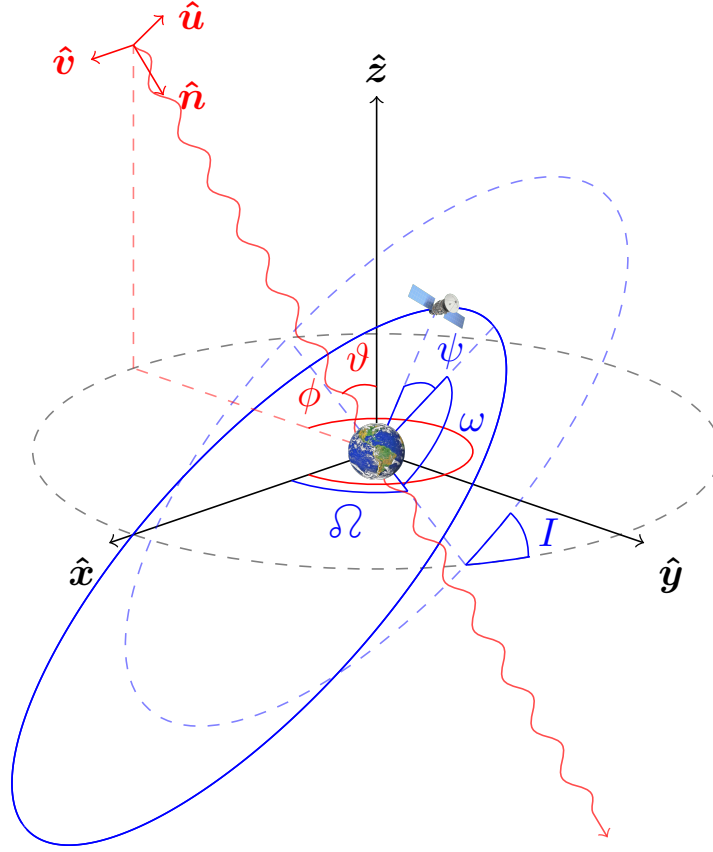
Figure 3.2: Schematic diagram of a Keplerian orbit. The plane of the orbit (shown in blue) is defined relative to the fixed reference frame ($\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}$) (shown in black) by the inclination I and the longitude of ascending node Ω, while the orientation of the orbit within this plane is specified by the argument of pericentre ω. The true anomaly ψ acts as an angular coordinate for the position of the orbit within the orbital plane, measured relative to the pericentre. The polarisation tensors describing the effects of an incoming plane GW are described in terms of the basis ($\hat{\boldsymbol{n}}, \hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}$) (shown in red), which is related to the reference frame by the angles ϑ and ϕ. (Note that the wavelength of the GW is not shown to scale here—our analysis assumes a wavelength much larger than the size of the orbit.)

By writing the osculating elements as functions of the energy $\mathcal{E}$ and angular momentum $\boldsymbol{\ell}$ we can translate equation (3.8) into a set of EoMs for the osculating elements. A full derivation along these lines is given in, e.g., Burns [161] or Murray and Dermott [432]; here we simply quote the resulting set of equations,

$$
\dot{P} = \frac{3P^2\gamma}{2\pi}\left[\frac{e\sin\psi\mathcal{F}_r}{1+e\cos\psi}+\mathcal{F}_\theta\right], \qquad \dot{e} = \frac{\dot{P}\gamma^2}{3Pe}-\frac{P\gamma^5\mathcal{F}_\theta}{2\pi e(1+e\cos\psi)^2},
$$

$$
\dot{I} = \frac{P\gamma^3\cos\theta\mathcal{F}_\ell}{2\pi(1+e\cos\psi)^2}, \qquad \dot{\Omega} = \frac{\tan\theta}{\sin I}\dot{I},
$$

$$
\dot{\omega} = \frac{P\gamma^3}{2\pi e}\left[\frac{(2+e\cos\psi)\sin\psi\mathcal{F}_\theta}{(1+e\cos\psi)^2}-\frac{\cos\psi\mathcal{F}_r}{1+e\cos\psi}\right]-\cos I\dot{\Omega},
$$

$$
\dot{\varepsilon} = -\frac{P\gamma^4\mathcal{F}_r}{\pi(1+e\cos\psi)^2}-\gamma(\cos I\dot{\Omega}+\dot{\omega}),
$$

(3.9)

where we have defined the dimensionless angular momentum,

$$\gamma \equiv \sqrt{1 - e^2} = \frac{\ell}{\sqrt{GMa}}. \tag{3.10}$$

Note that the size and shape of the orbit (determined by P and e) is only affected by forces within the plane (i.e., $\mathcal{F}_r$ and $\mathcal{F}_\theta$), while the plane of the orbit (determined by I and Ω) is only affected by forces normal to the plane (i.e., $\mathcal{F}_\ell$). The radial and angular phases of the orbit (determined by ω and ε) are affected by both.

3.1.3 Secular averaging of the perturbations

The EoMs (3.9) for the osculating elements are somewhat cumbersome to solve, as they each oscillate over the course of an orbital period due to their dependence on the true anomaly $\psi(t)$. In many situations however (including the present application to GW detection), the perturbing forces $\mathcal{F}_\alpha$ are small, and the resulting changes to the orbital elements therefore occur very slowly, on timescales much longer than the orbital period. It is then possible to decouple the fast oscillations of the true anomaly from the much more gradual evolution of the orbital elements by calculating a *secular average* of the EoMs,

$$\langle \dot{X} \rangle_{\text{sec}} \equiv \int_{t_0}^{t_0+P} \frac{\mathrm{d}t}{P} \dot{X}, \tag{3.11}$$

where $X \in (P, e, I, \Omega, \omega, \varepsilon)$. The notation $\langle \cdots \rangle_{\text{sec}}$ here is chosen to avoid confusion with plain angle brackets $\langle \cdots \rangle$ we use below to denote an ensemble average of a random variable. This procedure integrates out the fast oscillations that occur on timescales $\leq P$, leaving just the long-timescale dynamics of the system that we are interested in. Note that we treat the orbital elements themselves as constant inside the integral, which only makes sense in the limit where they are slowly-varying.

It is usually easier to evaluate the secular average if we replace the time integral in equation (3.11) with an integral over the true anomaly. We can do this by noticing that

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \ell/r^2 = \frac{2\pi}{P\gamma^3}(1 + e\cos\psi)^2, \tag{3.12}$$

where we have used equations (3.2), (3.4), and (3.5). The secular average can therefore be written as

$$\langle \dot{X} \rangle_{\text{sec}} = \int_0^{2\pi} \frac{\mathrm{d}\psi}{2\pi} \frac{\gamma^3 \dot{X}}{(1 + e\cos\psi)^2}. \tag{3.13}$$

In order to apply this average to the EoMs (3.9), we need to know the form of the perturbing forces $\mathcal{F}_\alpha$, which in general will vary over the course of the orbit.

3.1.4 Secular evolution due to relativistic effects

In section 3.2, we use equation (3.9) to calculate the perturbations to the osculating orbital elements caused by resonance with the GWB. However, we can use the same set of equations to calculate the perturbations caused by relativistic corrections to the equations of motion, which are particularly important for binaries

with short periods. Following Poisson and Will [459], we write the relativistic force components to leading post-Newtonian (PN) order as

$$\mathcal{F}_r = \left(\frac{2\pi}{P}\right)^2 \frac{v_P^2}{\gamma^8}(1 + e\cos\psi)^3\left[3 - \eta - e^2(1 + 3\eta) + e(2 - 4\eta)\cos\psi + e^2\frac{8 - \eta}{2}\sin^2\psi\right],$$

$$\mathcal{F}_\theta = \left(\frac{2\pi}{P}\right)^2 \frac{2ev_P^2}{\gamma^8}\sin\psi(1 + e\cos\psi)^4(2 - \eta), \qquad \mathcal{F}_\ell = 0,$$

(3.14)

where

$$v_P \equiv \left(\frac{2\pi GM}{P}\right)^{1/3} = \sqrt{\frac{GM}{a}} \tag{3.15}$$

is the binary's rms velocity, and we recall from section 0.3.1 that $\eta \equiv \mu/M$ is the dimensionless mass ratio. Inserting these expressions into equation (3.13), we find that the only non-vanishing secular perturbations to the osculating elements are

$$\langle\dot{\omega}\rangle_{\text{sec}} = \frac{6\pi v_P^2}{P\gamma^2}, \qquad \langle\dot{\varepsilon}\rangle_{\text{sec}} = \frac{2\pi v_P^2}{P}\left(6 - 7\eta - \frac{15 - 9\eta}{\gamma}\right), \tag{3.16}$$

where the first line is the famous *perihelion precession*, which provided some of the earliest observational evidence for GR. These are both 1PN corrections (i.e., order v_P^2).

Note that the above implies that the binary's period and eccentricity are conserved at 1PN order. However, since these are typically the most precisely-measured orbital elements, it is worth including their leading-order relativistic evolution, even if this is at higher PN order and is thus much smaller than the terms in equation (3.16) for most binaries. As we saw in section 0.3.1, this leading-order evolution of the period and eccentricity is due to GW radiation reaction, and can be calculated by applying simple energy-balance arguments to the unperturbed Keplerian orbit [455], giving

$$\langle\dot{P}\rangle_{\text{sec}} = -\frac{192\pi\eta v_P^5}{5\gamma^7}\left(1 + \tfrac{73}{24}e^2 + \tfrac{37}{96}e^4\right), \qquad \langle\dot{e}\rangle_{\text{sec}} = -\frac{608\pi\eta v_P^5}{15P\gamma^5}\left(e + \tfrac{121}{304}e^3\right). \tag{3.17}$$

These are just the period and eccentricity decay rates we found in equations (0.86) and (0.87), albeit with slightly different notation here. We see that both are 2.5PN effects (i.e., order v_P^5). The inclination and longitude of ascending node are both conserved at this PN order.

3.1.5 SMALL-ECCENTRICITY AND SMALL-INCLINATION ORBITS

For binaries where the eccentricity is small (say, $e \lesssim 10^{-3}$) the argument of pericentre ω becomes poorly-defined, as it becomes increasingly difficult to distinguish the point of closest approach as the orbit becomes increasingly circular. This in turn means that the compensated mean anomaly ε becomes poorly-defined, as this is defined relative to the time at which the binary is at pericentre, t_0. These issues can be resolved by defining an alternative set of orbital elements,

$$\zeta \equiv e\sin\omega, \qquad \kappa \equiv e\cos\omega, \qquad \xi \equiv \omega + \varepsilon. \tag{3.18}$$

The first two quantities here are sometimes called the 'Laplace-Lagrange eccentric parameters', while the latter is the compensated mean argument. We can then describe the orbit of a near-circular binary in terms of the alternative set of osculating elements $(P, \zeta, \kappa, I, \Omega, \xi)$ [387, 432]. These evolve according to

$$\dot{\zeta} = \dot{e}\sin\omega + \dot{\omega}e\cos\omega, \qquad \dot{\kappa} = \dot{e}\cos\omega - \dot{\omega}e\sin\omega, \qquad \dot{\xi} = \dot{\omega} + \dot{\varepsilon}, \tag{3.19}$$

with $\dot{e}$, $\dot{\Omega}$, $\dot{\omega}$, and $\dot{\varepsilon}$ given by equation (3.9). For the relativistic perturbations (3.14), we thus have

$$\langle\dot{\zeta}\rangle_{\text{sec}} = \frac{6\pi v_P^2}{P}\left(\kappa - \tfrac{304}{45}\eta v_P^3 \zeta\right), \qquad \langle\dot{\kappa}\rangle_{\text{sec}} = -\frac{6\pi v_P^2}{P}\left(\zeta + \tfrac{304}{45}\eta v_P^3 \kappa\right), \qquad \langle\dot{\xi}\rangle_{\text{sec}} = -\frac{4\pi v_P^2}{P}(3 - \eta),$$
$$\tag{3.20}$$

where we have neglected $O(e^2)$ terms. Similarly, Ω is ill-defined for orbits with very small inclination, so in this case we define

$$p \equiv I\sin\Omega, \qquad q \equiv I\cos\Omega, \qquad \lambda \equiv \Omega + \xi, \tag{3.21}$$

and describe the orbit using $(P, \zeta, \kappa, p, q, \lambda)$.

3.2 Resonant gravitational-wave perturbations

In this section we calculate the evolution of the osculating orbital elements of a binary system due to resonance with the GWB. We start by specifying the perturbing force associated with an incoming plane GW in the limit where the wavelength is much larger than the size of the orbit (i.e., the small-antenna limit). This allows us to write down a *Langevin equation* describing individual random realisations of the stochastic evolution of the osculating elements. Using the statistical properties of the GWB, we then derive a secularly-averaged *Fokker-Planck equation* which describes the evolution of the full statistical distribution of the orbital elements over timescales much longer than the binary period.

3.2.1 Coupling to the gravitational-wave polarisation modes

Using the results of section 0.1.5—in particular, equation (0.30)—we can express the response of a binary to an impinging plane GW in the proper detector frame as [406, 429]

$$\delta\ddot{r}^i = \frac{1}{2}\ddot{h}^{ij}r_j, \tag{3.22}$$

so that the resulting evolution of the binary is described in terms of the perturbing force terms

$$\mathcal{F}_r = \frac{1}{2}\ddot{h}_{ij}\hat{r}^i\hat{r}^j, \qquad \mathcal{F}_\theta = \frac{1}{2}\ddot{h}_{ij}\hat{r}^i\hat{\theta}^j, \qquad \mathcal{F}_\ell = \frac{1}{2}\ddot{h}_{ij}\hat{r}^i\hat{\ell}^j, \tag{3.23}$$

where $h_{ij}(t)$ is the TT part of the metric perturbation at the position of the binary's centre of mass. Carrying out a plane-wave decomposition as discussed in section 0.1.4,

$$h_{ij}(t) = \int_{S^2} \mathrm{d}^2\hat{\boldsymbol{n}}\, e_{ij}^A(\hat{\boldsymbol{n}})h_A(t, \hat{\boldsymbol{n}}), \tag{3.24}$$

these force terms can thus be written as

$$\mathcal{F}_\alpha = \frac{1}{2} \int_{S^2} \mathrm{d}^2\hat{\boldsymbol{n}}\, e_{ij}^A \hat{r}^i \hat{\alpha}^j \ddot{h}_A,$$

(3.25)

where α runs over the cylindrical coordinates (r, θ, ℓ).

As we discussed in section 0.1.5, equation (3.22) is only correct in the limit where the GW wavelength λ is much larger than the distance between the two objects in question, which in this case is set by the semi-major axis of the binary's orbit, $a \ll \lambda$. Using Kepler's third law (3.5), this condition can be rewritten as $fP \ll 1/v_P$, where f is the GW frequency. Since we are interested in GW frequencies which are harmonics of the binary period ($f = n/P$ for some $n \in \mathbb{Z}_+$), this tells us that the analysis below, based on equation (3.22), is only valid for harmonics that satisfy $n \ll 1/v_P$. This is not an impediment, since in the cases we are interested in the binary is sub-relativistic, $v_P \ll 1$, and as we will see in later sections, the strongest GW contribution to the binary's evolution typically comes from the lowest few harmonics anyway.

3.2.2 Langevin formulation

The stochastic evolution of the binary is described by the evolution equations derived in section 3.1.2, with the perturbing force terms given by equation (3.25). All this can be rewritten as a coupled set of nonlinear ordinary differential equations (ODEs),

$$\dot{X}_i(\boldsymbol{X}, t) = V_i(\boldsymbol{X}) + \Gamma_i(\boldsymbol{X}, t),$$

(3.26)

where X_i runs over the set of orbital elements (either $(P, e, I, \Omega, \omega, \varepsilon)$, or the small-eccentricity and/or small-inclination alternatives described in section 3.1.5), V is the deterministic drift term due to other perturbations (e.g., the relativistic effects described in section 3.1.4), and Γ is the stochastic diffusion term due to resonance with the GWB. ODEs of this form, in which one of the source terms is a random variable rather than a deterministic function, are called *Langevin equations* [484]. Rather than possessing a unique, deterministic solution for a given set of initial conditions, Langevin equations are characterised by a large number of distinct solutions, corresponding to different random realisations of the stochastic source term Γ.

In our case, Γ can be written as

$$\Gamma_i(\boldsymbol{X}, t) = \int_{S^2} \mathrm{d}^2\hat{\boldsymbol{n}}\, T_i^A(X, t, \hat{\boldsymbol{n}}) \ddot{h}_A(t, \hat{\boldsymbol{n}}),$$

(3.27)

where the T_i^A are a set of transfer functions describing the coupling between the GWB strain and the orbital elements. For example, using equations (3.9) and (3.25), the transfer functions for the period P are

$$T_P^A = \frac{3P^2\gamma}{4\pi} \left(\frac{e \sin\psi}{1 + e \cos\psi} \hat{r}^i + \hat{\theta}^i \right) \hat{r}^j e_{ij}^A.$$

(3.28)

However, the exact form of the transfer functions is unimportant for this section—all that matters is that we can write the stochastic term in the form (3.27). We write out the full set of transfer functions (in terms of their Fourier components) in appendix D.

The transfer functions are explicitly time-dependent via their dependence on trigonometric functions of the true anomaly $\psi(t)$, due to the variation of the binary's response over the course of each orbit. If we neglect the secular evolution of the orbital elements on long timescales $X/\dot{X} \gg P$, this means that the transfer functions are *periodic* functions of time, with period equal to the binary's orbital period P. This quasi-periodicity allows us to approximate each of the transfer functions as a Fourier series with period P,

$$T_i^A(\boldsymbol{X}, t, \hat{\boldsymbol{n}}) = \sum_{n=-\infty}^{+\infty} e^{-2\pi i n t/P} T_{i,n}^A(\boldsymbol{X}, \hat{\boldsymbol{n}}). \tag{3.29}$$

We therefore see that the binary's response to an incoming GW is characterised by its harmonic frequencies $f = n/P$. Over long timescales, these frequencies will change due to the secular evolution of the binary period P. The Fourier modes $T_{i,n}^A$ will also vary on these timescales, as they are functions of the orbital elements.

3.2.3 From the Langevin equation to the Fokker-Planck equation

Solving the Langevin equation (3.26) gives individual random trajectories of the binary through parameter space, corresponding to different random realisations of the GWB. However, we are more interested in the ensemble of all possible random trajectories. We can describe this ensemble in terms of the time-dependent distribution function (DF) for the orbital elements, $W(\boldsymbol{X}, t)$, which is defined such that the probability of the orbital elements $\boldsymbol{X}$ belonging to any region $\mathcal{X}$ of parameter space at time t is given by the corresponding integral over the DF,

$$\mathrm{Pr}(\boldsymbol{X} \in \mathcal{X}|t) = \int_{\mathcal{X}} \mathrm{d}\boldsymbol{X}\, W(\boldsymbol{X}, t). \tag{3.30}$$

This DF can either be interpreted as the probability density function for the stochastic orbital elements of an individual binary, or as the cumulative distribution for the orbital elements of a population of multiple binaries.

Formally, we can write down the time evolution of the DF in terms of the Kramers-Moyal (KM) forward expansion [280, 484],

$$\frac{\partial W}{\partial t} = \sum_{n=1}^{\infty} (-)^n \frac{\partial^n}{\partial X_{i_1} \partial X_{i_2} \cdots \partial X_{i_n}} \left(D_{i_1 i_2 \cdots i_n}^{(n)} W \right), \tag{3.31}$$

where repeated indices are summed over. This is determined by the KM coefficients,

$$D_{i_1 i_2 \cdots i_n}^{(n)}(\boldsymbol{X}, t) \equiv \lim_{\tau \to 0} \frac{1}{\tau n!} \left\langle \prod_{j=1}^{n} \left[X_{i_j}(t+\tau) - X_{i_j}(t) \right] \right\rangle, \tag{3.32}$$
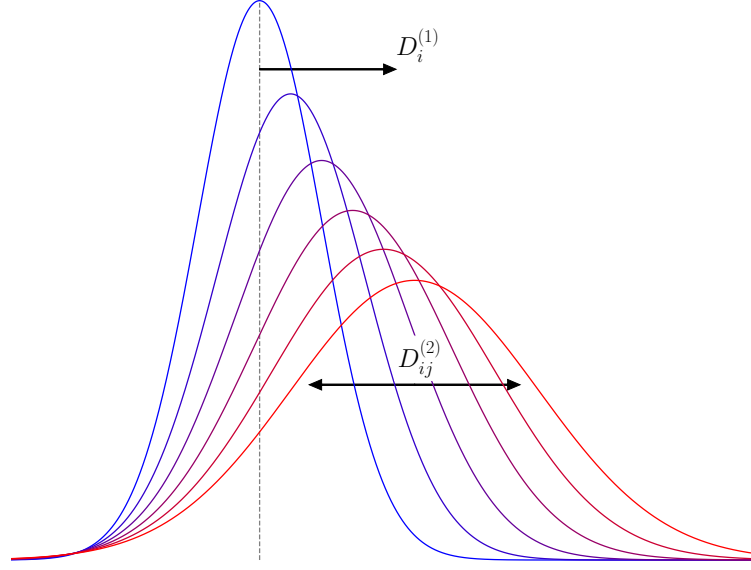
Figure 3.3: Heuristic illustration of a distribution function evolving according to the Fokker-Planck equation. The drift vector $D_i^{(1)}$ drives the bulk movement of the distribution in one direction, while the diffusion matrix $D_{ij}^{(2)}$ causes the distribution to become broader, as the randomness of the stochastic perturbation propagates into an increasing randomness in the distribution.

with angle brackets indicating an ensemble average under the distribution W at time t. Since we take the GWB as Gaussian, the KM coefficients for orders $n \geq 3$ all vanish,[1] leaving just the first two coefficients,

$$
\begin{aligned}
D_i^{(1)} &= \lim_{\tau \to 0} \frac{1}{\tau} \langle X_i(t + \tau) - X_i(t) \rangle, \\
D_{ij}^{(2)} &= \lim_{\tau \to 0} \frac{1}{2\tau} \langle [X_i(t + \tau) - X_i(t)][X_j(t + \tau) - X_j(t)] \rangle,
\end{aligned}
\tag{3.33}
$$

which we call the *drift vector* and the *diffusion matrix*, respectively. The KM forward expansion (3.31) then becomes the *Fokker-Planck equation* (FPE),

$$
\frac{\partial W}{\partial t} = -\partial_i (D_i^{(1)} W) + \partial_i \partial_j (D_{ij}^{(2)} W),
\tag{3.34}
$$

with $\partial_i \equiv \partial/\partial X_i$. The effect of each set of coefficients is illustrated in figure 3.3.

We can calculate the KM coefficients by directly integrating the Langevin equation (3.26), from some initial time t where the orbital elements are 'sharp' (i.e., known exactly rather than randomly distributed), $X_i(t) \equiv x_i$, over some small time interval τ,

$$
X_i(t + \tau) - x_i = \int_t^{t+\tau} \mathrm{d}t' \, [V_i(\boldsymbol{X}(t')) + \Gamma_i(\boldsymbol{X}(t'), t')].
\tag{3.35}
$$

---

[1]One could derive this explicitly by using the approach described later in this section to show that $D_{ijk}^{(3)} = 0$, as the Pawula theorem [450, 484] then implies that all higher-order coefficients $n > 3$ must vanish in order to guarantee that the DF is normalised.

(This derivation closely follows that in section 3.3.2 of Risken [484], but is generalised to allow the noise term Γ to depend on the system's parameters X_i, rather than just being a function of time.) Both terms under the integral on the right-hand side are random, due to the random spread in the orbital elements for all times $t' > t$. However, we can express these in terms of the sharp values x_i by Taylor expanding,

$$
\begin{aligned}
V_i(\boldsymbol{X}(t')) &= V_i(\boldsymbol{x}) + \partial_j V_i(\boldsymbol{x})[X_j(t') - x_j] + \cdots , \\
\Gamma_i(\boldsymbol{X}(t'), t') &= \Gamma_i(\boldsymbol{x}, t') + \partial_j \Gamma_i(\boldsymbol{x}, t')[X_j(t') - x_j] + \cdots .
\end{aligned}
\tag{3.36}
$$

Wherever $X_i(t') - x_i$ appears on the right-hand side of equation (3.36) we can insert equation (3.35) and iterate, such that the only orbital elements that appear are the sharp initial values x_i. Then the only randomness is through the stochastic term Γ_i, whose statistics are completely specified in terms of the GWB moments (3.45).

The first few terms in the iterative expansion are

$$
\begin{aligned}
X_i(t + \tau) - x_i &= \int_t^{t+\tau} \mathrm{d}t' \, V_i(\boldsymbol{x}) + \int_t^{t+\tau} \mathrm{d}t' \, \partial_j V_i(\boldsymbol{x}) \int_t^{t'} \mathrm{d}t'' \, [V_j(\boldsymbol{x}) + \Gamma_j(\boldsymbol{x}, t'')] + \cdots \\
&+ \int_t^{t+\tau} \mathrm{d}t' \, \Gamma_i(\boldsymbol{x}, t') + \int_t^{t+\tau} \mathrm{d}t' \, \partial_j \Gamma_i(\boldsymbol{x}, t') \int_t^{t'} \mathrm{d}t'' \, [V_j(\boldsymbol{x}) + \Gamma_j(\boldsymbol{x}, t'')] + \cdots .
\end{aligned}
\tag{3.37}
$$

By virtue of the time-independence of the sharp drift term $V_i(\boldsymbol{x})$, this immediately simplifies to

$$
\begin{aligned}
X_i(t + \tau) - x_i &= \tau V_i(\boldsymbol{x}) + \frac{1}{2}\tau^2 V_j(\boldsymbol{x})\partial_j V_i(\boldsymbol{x}) + \partial_j V_i(\boldsymbol{x}) \int_t^{t+\tau} \mathrm{d}t' \int_t^{t'} \mathrm{d}t'' \, \Gamma_j(\boldsymbol{x}, t'') + \cdots \\
&+ \int_t^{t+\tau} \mathrm{d}t' \, \Gamma_i(\boldsymbol{x}, t') + V_j(x) \int_t^{t+\tau} \mathrm{d}t' \, (t' - t)\partial_j \Gamma_i(\boldsymbol{x}, t') \\
&+ \int_t^{t+\tau} \mathrm{d}t' \, \partial_j \Gamma_i(\boldsymbol{x}, t') \int_t^{t'} \mathrm{d}t'' \, \Gamma_j(\boldsymbol{x}, t'') + \cdots .
\end{aligned}
\tag{3.38}
$$

Taking the first moment of equation (3.38), all terms linear in Γ_i vanish, and we are left with

$$
\begin{aligned}
\langle X_i(t + \tau) - x_i \rangle &= \tau V_i(\boldsymbol{x}) + \frac{1}{2}\tau^2 V_j(\boldsymbol{x})\partial_j V_i(\boldsymbol{x}) + \cdots \\
&+ \int_t^{t+\tau} \mathrm{d}t' \int_t^{t'} \mathrm{d}t'' \, \langle \Gamma_j(\boldsymbol{x}, t'')\partial_j \Gamma_i(\boldsymbol{x}, t') \rangle + \cdots ,
\end{aligned}
\tag{3.39}
$$

so that the drift vector is given by

$$
D_i^{(1)} = V_i + \lim_{\tau \to 0} \frac{1}{\tau} \int_t^{t+\tau} \mathrm{d}t' \int_t^{t'} \mathrm{d}t'' \, \langle \Gamma_j(\boldsymbol{x}, t'')\partial_j \Gamma_i(\boldsymbol{x}, t') \rangle.
\tag{3.40}
$$

Similarly, when taking the second moment of equation (3.38) and taking the $\tau \to 0$ limit, the only term that survives for the diffusion matrix is

$$
D_{ij}^{(2)} = \lim_{\tau \to 0} \frac{1}{2\tau} \int_t^{t+\tau} \mathrm{d}t' \int_t^{t+\tau} \mathrm{d}t'' \, \langle \Gamma_i(\boldsymbol{x}, t')\Gamma_j(\boldsymbol{x}, t'') \rangle.
\tag{3.41}
$$

The fact that both sets of KM coefficients are quadratic in Γ makes sense, since this is the smallest power of Γ which does not vanish when taking an expectation value. The difference, however, is that the drift vector has a derivative acting on one factor of Γ, while the diffusion matrix has no derivatives. This is because the quadratic-in-Γ term appearing in the drift vector corresponds to the second-order term in the Taylor expansion we wrote down in equation (3.36), while in the diffusion matrix this is simply the product of two first-order Taylor expansion terms.

3.2.4 Secular drift and diffusion

By calculating the KM coefficients from equations (3.40) and (3.41), we can now (in principle) obtain the full DF for the orbital elements by integrating the FPE (3.34). However, in order to calculate the KM coefficients, we must first evaluate the ensemble averages $\langle \Gamma^2 \rangle$ and $\langle \Gamma \partial \Gamma \rangle$ that appear in the integrands of equations (3.40) and (3.41) above. We do this by rewriting these as ensemble averages over the GWB strain. Using equations (3.27) and (3.29), we find

$$
\begin{aligned}
\langle \Gamma_i(\boldsymbol{x}, t')\Gamma_j(\boldsymbol{x}, t'') \rangle &= \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \int_{S^2} \mathrm{d}^2\hat{\boldsymbol{n}} \int_{S^2} \mathrm{d}^2\hat{\boldsymbol{n}}' \, \mathrm{e}^{2\pi\mathrm{i}(nt'-mt'')/P} \\
&\quad \times T_{i,n}^{A*}(\boldsymbol{x}, \hat{\boldsymbol{n}}) T_{j,m}^{A'}(\boldsymbol{x}, \hat{\boldsymbol{n}}') \langle \ddot{\bar{h}}_A(t', \hat{\boldsymbol{n}}) \ddot{\bar{h}}_{A'}(t'', \hat{\boldsymbol{n}}') \rangle, \\
\langle \Gamma_j(\boldsymbol{x}, t'')\partial_j\Gamma_i(\boldsymbol{x}, t') \rangle &= \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \int_{S^2} \mathrm{d}^2\hat{\boldsymbol{n}} \int_{S^2} \mathrm{d}^2\hat{\boldsymbol{n}}' \, \mathrm{e}^{2\pi\mathrm{i}(nt'-mt'')/P} \\
&\quad \times T_{j,m}^{A'}(\boldsymbol{x}, \hat{\boldsymbol{n}}')\partial_j T_{i,n}^{A*}(\boldsymbol{x}, \hat{\boldsymbol{n}}) \langle \ddot{\bar{h}}_A(t', \hat{\boldsymbol{n}}) \ddot{\bar{h}}_{A'}(t'', \hat{\boldsymbol{n}}') \rangle,
\end{aligned}
\tag{3.42}
$$

so that both terms are determined by the 2-point statistics of the strain.

In section 0.2.3, we derived the second moment of the Fourier transform of the strain for a GWB that is Gaussian, stationary, unpolarised, and isotropic, with no nontrivial phase correlations,

$$
\langle \breve{h}_A(f, \hat{\boldsymbol{n}}) \breve{h}_{A'}^*(f', \hat{\boldsymbol{n}}') \rangle = \frac{3}{4} H_0^2 (2\pi|f|)^{-3} \Omega(f) \delta_{AA'} \delta(f-f') \delta^{(2)}(\hat{\boldsymbol{n}}, \hat{\boldsymbol{n}}').
\tag{3.43}
$$

Here we are interested in the statistics of the second time derivative of the strain, which is related to the Fourier components by

$$
\ddot{\bar{h}}_A = \frac{\mathrm{d}^2}{\mathrm{d}t^2} \int_{-\infty}^{+\infty} \mathrm{d}f \, \mathrm{e}^{-2\pi\mathrm{i}ft} \breve{h}_A = -\int_{-\infty}^{+\infty} \mathrm{d}f \, \mathrm{e}^{-2\pi\mathrm{i}ft} (2\pi f)^2 \breve{h}_A,
\tag{3.44}
$$

such that

$$
\langle \ddot{\bar{h}}_A(t, \hat{\boldsymbol{n}}) \ddot{\bar{h}}_{A'}(t', \hat{\boldsymbol{n}}') \rangle = 3\pi H_0^2 \delta_{AA'} \delta(\hat{\boldsymbol{n}}, \hat{\boldsymbol{n}}') \int_0^\infty \mathrm{d}f \, \cos[2\pi f(t-t')] f \Omega(f).
\tag{3.45}
$$

Inserting this into equation (3.42), we obtain

$$\langle \Gamma_i(\boldsymbol{x}, t') \Gamma_j(\boldsymbol{x}, t'') \rangle = 3\pi H_0^2 \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \int_{S^2} \mathrm{d}^2 \hat{\boldsymbol{n}}\, T_{i,n}^{A*} T_{j,m}^{A}$$

$$\times \int_0^\infty \mathrm{d}f\, \mathrm{e}^{2\pi\mathrm{i}(nt'-mt'')/P} \cos[2\pi f(t'-t'')] f\,\Omega(f),$$

$$\langle \Gamma_j(\boldsymbol{x}, t'') \partial_j \Gamma_i(\boldsymbol{x}, t') \rangle = 3\pi H_0^2 \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \int_{S^2} \mathrm{d}^2 \hat{\boldsymbol{n}}\, T_{j,m}^{A} \partial_j T_{i,n}^{A*}$$

$$\times \int_0^\infty \mathrm{d}f\, \mathrm{e}^{2\pi\mathrm{i}(nt'-mt'')/P} \cos[2\pi f(t'-t'')] f\,\Omega(f). \tag{3.46}$$

In order to derive the corresponding KM coefficients using equations (3.40) and (3.41), we see that we must evaluate two oscillatory time integrals,

$$\int_t^{t+\tau} \mathrm{d}t' \int_t^{t+\tau} \mathrm{d}t''\, \mathrm{e}^{-2\pi\mathrm{i}(f-n/P)t'} \mathrm{e}^{2\pi\mathrm{i}(f-m/P)t''},$$

$$\int_t^{t+\tau} \mathrm{d}t' \int_t^{t'} \mathrm{d}t''\, \mathrm{e}^{-2\pi\mathrm{i}(f-n/P)t'} \mathrm{e}^{2\pi\mathrm{i}(f-m/P)t''}, \tag{3.47}$$

where we have converted the cosine in equation (3.46) into an exponential. To do so, we recall that the timescales $\leq P$ associated with the binary's resonant frequencies are much shorter than the secular timescales $X/\dot{X}$ we are interested in.[2] Even though we will later take the limit $\tau \to 0$, all we really require to derive a FPE valid on secular timescales $\sim X/\dot{X}$ is that $\tau \ll X/\dot{X}$, and since we have $P \ll X/\dot{X}$, we can consistently also demand that $\tau \gg P$. In this limit, the first integral in equation (3.47) is approximated by

$$\int_t^{t+\tau} \mathrm{d}t' \int_t^{t+\tau} \mathrm{d}t''\, \mathrm{e}^{-2\pi\mathrm{i}(f-n/P)t'} \mathrm{e}^{2\pi\mathrm{i}(f-m/P)t''} \simeq \tau \delta_{mn} \delta(f-n/P). \tag{3.48}$$

For the second integral, notice that we can write

$$\tau \delta_{mn} \delta(f-n/P) \simeq \int_t^{t+\tau} \mathrm{d}t' \int_t^{t+\tau} \mathrm{d}t''\, \mathrm{e}^{-2\pi\mathrm{i}(f-n/P)t'} \mathrm{e}^{2\pi\mathrm{i}(f-m/P)t''}$$

$$= \int_t^{t+\tau} \mathrm{d}t' \int_t^{t'} \mathrm{d}t''\, \mathrm{e}^{-2\pi\mathrm{i}(f-n/P)t'} \mathrm{e}^{2\pi\mathrm{i}(f-m/P)t''} \tag{3.49}$$

$$+ \int_t^{t+\tau} \mathrm{d}t' \int_{t'}^{t+\tau} \mathrm{d}t''\, \mathrm{e}^{-2\pi\mathrm{i}(f-n/P)t'} \mathrm{e}^{2\pi\mathrm{i}(f-m/P)t''},$$

and that in the limit $\tau \gg P$ the latter two terms are approximately equal to each other, so that

$$\int_t^{t+\tau} \mathrm{d}t' \int_t^{t'} \mathrm{d}t''\, \mathrm{e}^{-2\pi\mathrm{i}(f-n/P)t'} \mathrm{e}^{2\pi\mathrm{i}(f-m/P)t''} \simeq \frac{1}{2}\tau \delta_{mn} \delta(f-n/P). \tag{3.50}$$

---

[2]It would also be interesting to study the dynamics on sub-orbital timescales, as was done by Rozner *et al.* [488] and Desjacques *et al.* [236] for the case of a perturbing ultralight dark matter field. We leave this for future work.

Plugging this back in to equations (3.40), (3.41), and (3.46), we are able to write the KM coefficients directly in terms of the GW transfer functions,

$$
\begin{aligned}
D_i^{(1)} &= V_i + \frac{3\pi}{2} H_0^2 \sum_{n=1}^{\infty} \int_{S^2} \mathrm{d}^2\hat{\boldsymbol{n}} \, \frac{n\Omega_n}{P} \operatorname{Re}\left(T_{j,n}^A \partial_j T_{i,n}^{A*} \right), \\
D_{ij}^{(2)} &= \frac{3\pi}{2} H_0^2 \sum_{n=1}^{\infty} \int_{S^2} \mathrm{d}^2\hat{\boldsymbol{n}} \, \frac{n\Omega_n}{P} \operatorname{Re}\left(T_{i,n}^{A*} T_{j,n}^A \right),
\end{aligned}
\tag{3.51}
$$

where we have used $T_{i,-n} = T_{i,n}^*$, and where $\Omega_n \equiv \Omega(n/P)$ is the GWB energy density at the binary's n^{th} harmonic frequency. (We always take only the real part when defining the KM coefficients, so for brevity we will leave this implicit from now on.) Equation (3.51) describes the secular drift and diffusion of the orbital elements over timescales much longer than the orbital period P. We neglect the $\langle \cdots \rangle_{\text{sec}}$ notation used in section 3.1 for brevity, but the interpretation here is exactly the same.

Note that the drift vector in equation (3.51) includes not just the expected deterministic drift V_i, but also a stochastic term. As we show below, this 'noise-induced drift' leads to a net evolution of the *mean values* of the orbital elements, and not just their variance. This is somewhat counter-intuitive, and justifies the careful derivation presented in this section; otherwise, it would be tempting to assume that the GWB affects only the variance of the orbital elements (as was assumed in Hui *et al.* [323], for example). One can understand this drift as being the result of a 'diffusion gradient' due to the nonlinear coupling between the GWB and the binary: the orbital elements change in response to the GW strain, thereby changing the values of the transfer functions and modifying the response to the strain, with the resulting feedback loop creating a preferred direction in parameter space for the orbital elements to evolve towards. The partial derivative acting on the transfer function in equation (3.51) shows that this term would vanish in the linear case (i.e., where the transfer functions are constant), and therefore that this is a purely nonlinear effect.

We can also understand this stochastic drift as being a consequence of the GWB strain having a *finite correlation time*; i.e., there is some $\tau_{\text{c}} > 0$ such that

$$
\langle \ddot{h}_{ij}(t) \ddot{h}_{ij}(t+\tau) \rangle > 0 \quad \text{for all} \quad |\tau| < \tau_{\text{c}},
\tag{3.52}
$$

(with no implied summation over the spatial indices). Physically speaking, this is because we cannot have arbitrarily high GW frequencies contributing to the GWB spectrum $\Omega(f)$, so equation (3.45) automatically satisfies equation (3.52). However, since this correlation time is much shorter than the secular timescales we are interested in, we can implicitly take $\tau_{\text{c}} \to 0$ at the end of the calculation. This is the Stratonovich prescription [529] for calculating stochastic integrals, which differs from the Itô prescription [331] in which the correlation time is assumed to be zero from the start. Surprisingly, these two prescriptions give different physical predictions when the transfer functions depend on the state of the system, with the stochastic drift term we derived above being present only in the Stratonovich case and not in the Itô case. This can present a problem in systems with a white-noise stochastic term, $\langle \Gamma(t)\Gamma(t') \rangle = \delta(t - t')$, which formally has zero correlation time, as one must determine which of the two prescriptions is more appropriate. However, we stress that for our case here equation (3.45) shows

that there must be a nonzero correlation time, so the Stratonovich prescription is the only sensible choice. (We could get equation (3.45) to be proportional to $\delta(t - t')$ by having $\Omega(f) \propto 1/f$ at arbitrarily high frequencies, but this would cause the total GW energy density to diverge.) The stochastic drift term appearing in equation (3.51) is therefore an unambiguous prediction for a binary coupled to the GWB. For further discussion of the Itô and Stratonovich prescriptions and the origin of noise-induced drift see, e.g., section 3.3.3 of Risken [484].

3.3 Calculating the Kramers-Moyal coefficients

In this section we use equation (3.51) to derive explicit expressions for the KM coefficients as functions of the binary orbital elements and the GWB spectrum. We start with the most general expressions, before specialising to the cases of small eccentricity and small inclination.

3.3.1 General orbits

We present here the secular KM coefficients for general eccentricity $e \in (0, 1)$. The details of this calculation are lengthy and unimportant for the final results, so we quote only the final expressions here, with some further details given in appendix D. In particular, section D.1 derives the necessary projections of the GW polarisation tensors onto the binary's cylindrical coordinate basis; section D.2 uses these projections to write down the transfer functions for all six orbital elements $(P, e, I, \Omega, \omega, \varepsilon)$ in terms of functions of the eccentricity called *Hansen coefficients*, which we introduce below; section D.3 derives the equations describing how each of the KM coefficients transforms under a change of the reference frame, allowing us to select a particular frame which simplifies the calculations; section D.4 presents the values of the KM coefficients in this frame in terms of the Hansen coefficients; and section D.5 writes out all of the necessary Hansen coefficients explicitly as functions of eccentricity.

Putting all of these ingredients together, we write the full set of KM coefficients explicitly in terms of the Hansen coefficients, which are defined as

$$
\begin{aligned}
C_n^{lm}(e) &\equiv \left\langle \exp\left(\frac{2\pi \mathrm{i} n t}{P}\right) \frac{\cos m\psi}{(1 + e\cos\psi)^l} \right\rangle_{\mathrm{sec}}, \\
S_n^{lm}(e) &\equiv \left\langle \exp\left(\frac{2\pi \mathrm{i} n t}{P}\right) \frac{\sin m\psi}{(1 + e\cos\psi)^l} \right\rangle_{\mathrm{sec}}, \\
E_n^{lm}(e) &\equiv C_n^{lm}(e) + S_n^{lm}(e).
\end{aligned}
\tag{3.53}
$$

The secular diffusion matrix $D_{ij}^{(2)}$ is thus given by

$$
D_{PP}^{(2)} = \frac{27 P^3 \gamma^2}{20} \sum_{n=1}^{\infty} n H_0^2 \Omega_n \left[\left| E_n^{02} + \frac{e}{2}(E_n^{11} - E_n^{13}) \right|^2 - \frac{(e S_n^{11})^2}{3} \right],
$$

$$
D_{Pe}^{(2)} = \frac{\gamma^2 D_{PP}^{(2)}}{3Pe} - \frac{9 P^2 \gamma^6}{40} \sum_{n=1}^{\infty} n H_0^2 \Omega_n E_n^{22} \left(\frac{2}{e} E_n^{02} + E_n^{11} - E_n^{13} \right)^*,
$$

$$
D_{ee}^{(2)} = \frac{3 P \gamma^6}{20 e^2} \sum_{n=1}^{\infty} n H_0^2 \Omega_n \left[\left| E_n^{02} + \frac{e}{2}(E_n^{11} - E_n^{13}) - \gamma^2 E_n^{22} \right|^2 - \frac{(e S_n^{11})^2}{3} \right],
$$
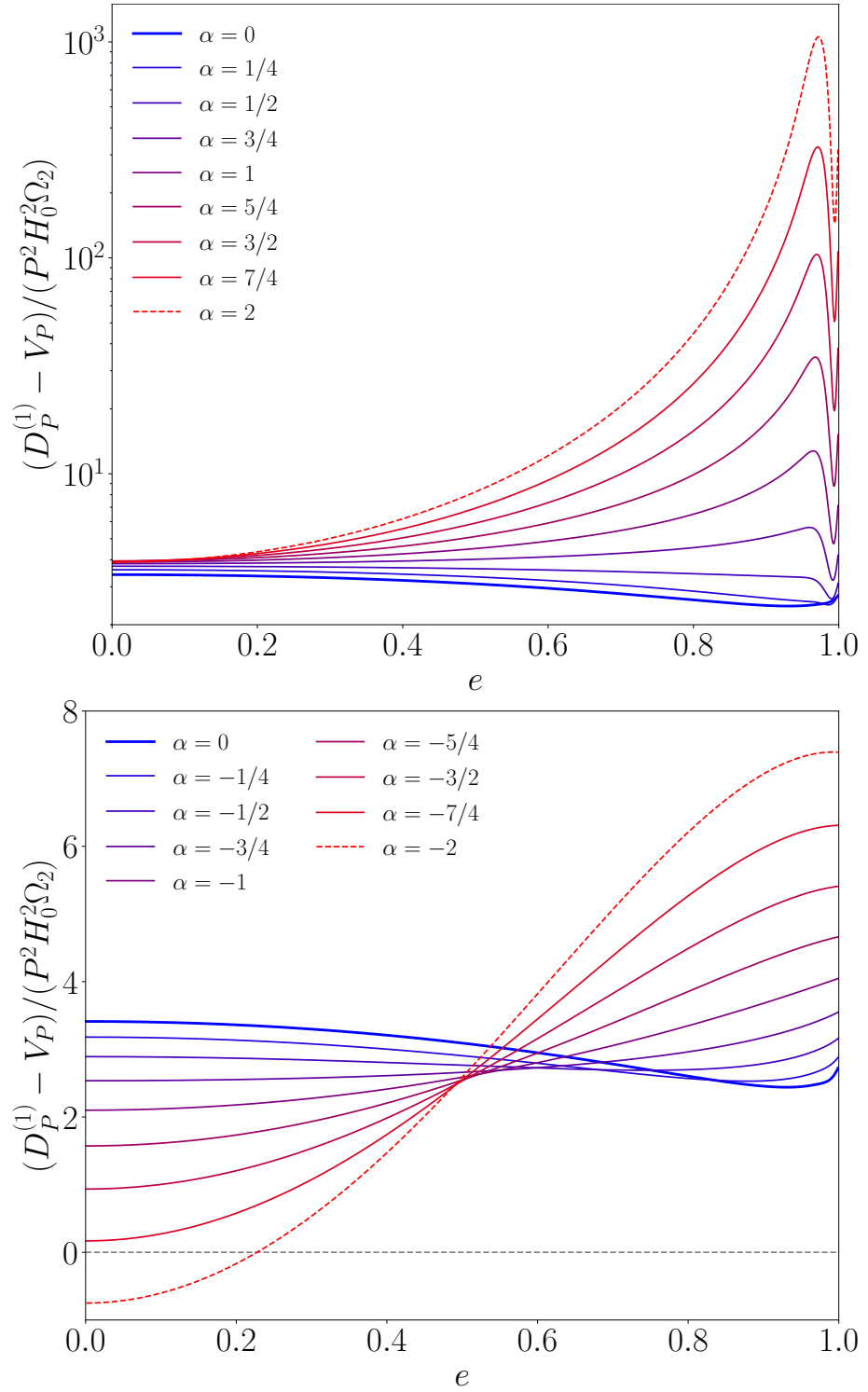
Figure 3.4: Stochastic parts of the secular period drift coefficient $D_P^{(1)}$ as a function of eccentricity for various power-law GWB spectra, $\Omega(f) \sim f^\alpha$. The top panel shows positive power-law indices, $\alpha = 0, \ldots, 2$, while the bottom panel shows negative indices, $\alpha = 0, \ldots, -2$.

$$D_{II}^{(2)} = \frac{3P\gamma^6}{80} \sum_{n=1}^{\infty} nH_0^2 \Omega_n \left[\left|E_n^{20}\right|^2 + \left|E_n^{22}\right|^2 + 2\cos 2\omega E_n^{20}(E_n^{22})^* \right],$$

$$D_{I\Omega}^{(2)} = \frac{3P\gamma^6}{40}\frac{\sin 2\omega}{\sin I} \sum_{n=1}^{\infty} nH_0^2 \Omega_n E_n^{20}(E_n^{22})^*, \qquad D_{I\omega}^{(2)} = -\frac{3P\gamma^6}{40}\frac{\sin 2\omega}{\tan I} \sum_{n=1}^{\infty} nH_0^2 \Omega_n E_n^{20}(E_n^{22})^*,$$

$$D_{\Omega\Omega}^{(2)} = \frac{3P\gamma^6}{80\sin^2 I} \sum_{n=1}^{\infty} nH_0^2 \Omega_n \left[\left|E_n^{20}\right|^2 + \left|E_n^{22}\right|^2 - 2\cos 2\omega E_n^{20}(E_n^{22})^* \right], \qquad D_{\Omega\omega}^{(2)} = -\cos I D_{\Omega\Omega}^{(2)},$$

$$D_{\omega\omega}^{(2)} = \cos^2 I D_{\Omega\Omega}^{(2)}$$
$$+ \frac{3P\gamma^6}{80e^2} \sum_{n=1}^{\infty} nH_0^2 \Omega_n \left[\left| E_n^{11} + E_n^{13} + 2E_n^{21} - 2E_n^{23} + \frac{e}{2}\left(E_n^{20} - E_n^{24}\right) \right|^2 + \frac{4}{3}\left(C_n^{11}\right)^2 \right],$$

$$D_{\omega\varepsilon}^{(2)} = -\frac{3P\gamma^7}{80e^2} \sum_{n=1}^{\infty} nH_0^2 \Omega_n \left[\left| E_n^{11} + E_n^{13} + 2(E_n^{21} - E_n^{23}) + \frac{e}{2}(E_n^{20} - 4E_n^{22} - E_n^{24}) \right|^2 \right.$$
$$\left. + \frac{4}{3}C_n^{11}(C_n^{11} - 2eC_n^{20}) - 4\left|eE_n^{22}\right|^2 \right],$$

$$D_{\varepsilon\varepsilon}^{(2)} = \frac{3P\gamma^8}{80e^2} \sum_{n=1}^{\infty} nH_0^2 \Omega_n \left[\left| E_n^{11} + E_n^{13} + 2(E_n^{21} - E_n^{23}) + \frac{e}{2}(E_n^{20} - 8E_n^{22} - E_n^{24}) \right|^2 \right.$$
$$\left. + \frac{4}{3}(C_n^{11} - 2eC_n^{20})^2 \right],$$

$$D_{PI}^{(2)} = D_{P\omega}^{(2)} = D_{P\Omega}^{(2)} = D_{P\varepsilon}^{(2)} = D_{eI}^{(2)} = D_{e\omega}^{(2)} = D_{e\Omega}^{(2)} = D_{e\varepsilon}^{(2)} = D_{I\varepsilon}^{(2)} = D_{\Omega\varepsilon}^{(2)} = 0,$$

$$(3.54)$$

while the secular drift vector $D_i^{(1)}$ is given by

$$D_P^{(1)} = V_P + \frac{9P^2\gamma^2}{4} \sum_{n=1}^{\infty} nH_0^2 \Omega_n \left[\left| E_n^{02} + \frac{e}{2}(E_n^{11} - E_n^{13}) \right|^2 - \frac{1 + 4e^2}{15}(S_n^{11})^2 - \frac{e\gamma^2}{15} S_n^{11\prime} S_n^{11} \right.$$
$$+ \frac{\gamma^2}{10e}\left(E_n^{02} + \frac{e}{2}(E_n^{11} - E_n^{13}) \right)$$
$$\times \left(3E_n^{11} + E_n^{13} + eE_n^{20} + 4E_n^{21} + 4eE_n^{22} - 4E_n^{23} - eE_n^{24} + 2E_n^{02\prime} + e(E_n^{11\prime} - E_n^{13\prime}) \right)^*$$
$$\left. - \frac{\gamma^4}{10e}E_n^{22}\left(E_n^{11} - E_n^{13} + 2E_n^{02\prime} + e(E_n^{11\prime} - E_n^{13\prime}) \right)^* \right],$$

$$D_e^{(1)} = V_e - \frac{P\gamma^6}{20} \sum_{n=1}^{\infty} nH_0^2 \Omega_n \left[S_n^{11\prime} S_n^{11} + \frac{3}{e^3}\left(E_n^{02} + \frac{e}{2}(E_n^{11} - E_n^{13}) - \gamma^2 E_n^{22} \right) \right.$$
$$\times \left(E_n^{02} - E_n^{22} - e(E_n^{11} + E_n^{13} + 2E_n^{21} - 2E_n^{23} + E_n^{02\prime} - \gamma^2 E_n^{22\prime}) \right.$$
$$\left.\left. - \frac{e^2}{2}(E_n^{20} + 10E_n^{22} - E_n^{24} + E_n^{11\prime} - E_n^{13\prime}) \right)^* \right],$$

$$(3.55)$$

$$D_I^{(1)} = \frac{\sin 2I}{2} D_{\Omega\Omega}^{(2)}, \qquad D_\Omega^{(1)} = -2\cot I D_{I\Omega}^{(2)},$$

$$D_\omega^{(1)} = V_\omega + \frac{3P\gamma^6}{40}\sin 2\omega \frac{2 - \sin^2 I}{\sin^2 I} \sum_{n=1}^{\infty} nH_0^2 \Omega_n E_n^{20}(E_n^{22})^*, \qquad D_\varepsilon^{(1)} = V_\varepsilon.$$

The deterministic drift terms here are those given in section 3.1.4,
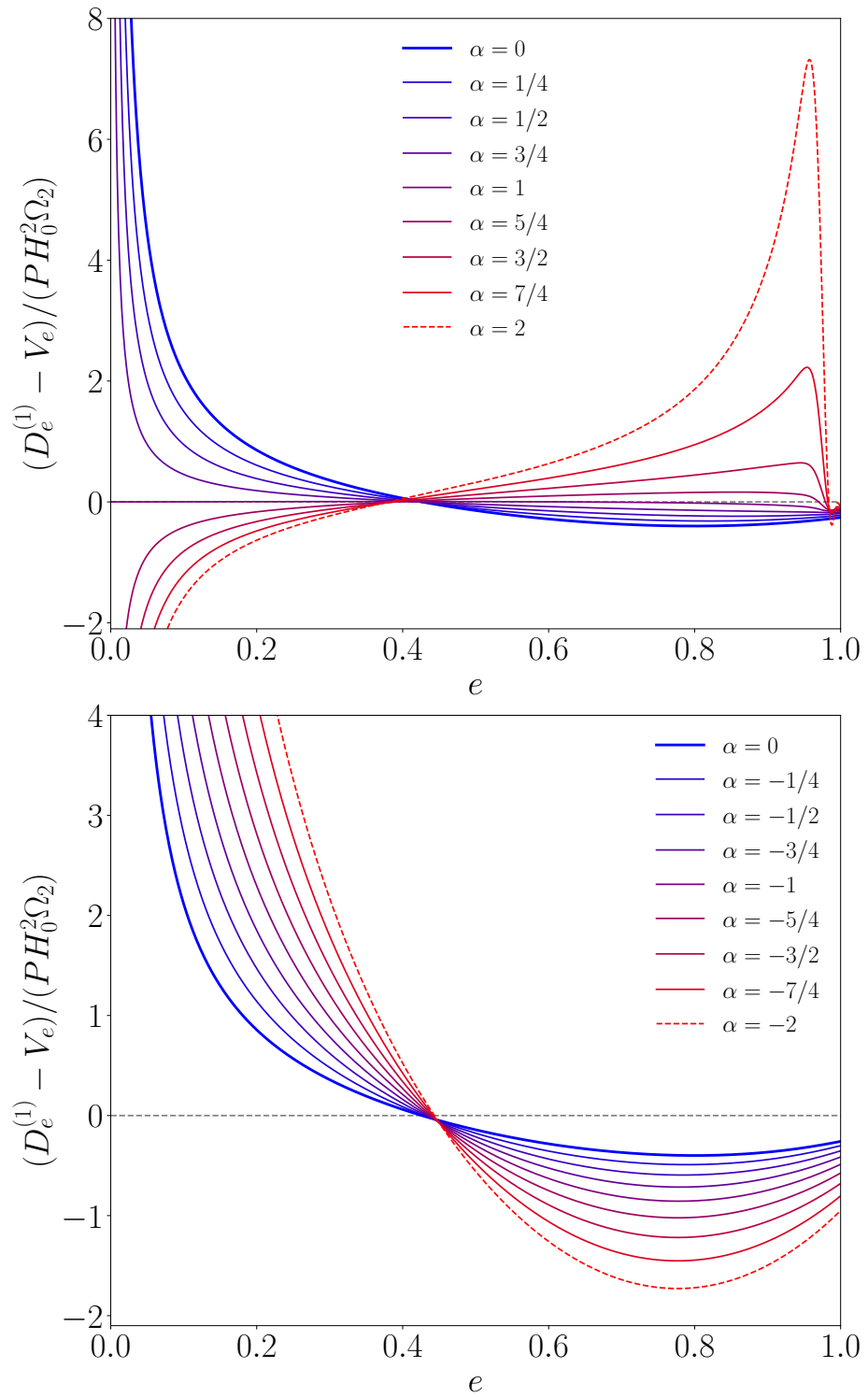
Figure 3.5: Same as figure 3.4, but for the eccentricity drift coefficient $D_e^{(1)}$.

$$V_P = -\frac{192\pi\eta v_P^5}{5\gamma^7}\left(1 + \tfrac{73}{24}e^2 + \tfrac{37}{96}e^4\right), \qquad V_e = -\frac{608\pi\eta v_P^5}{15P\gamma^5}\left(e + \tfrac{121}{304}e^3\right),$$

$$V_\omega = \frac{6\pi v_P^2}{P\gamma^2}, \qquad V_\varepsilon = \frac{2\pi v_P^2}{P}\left(6 - 7\eta - \frac{15 - 9\eta}{\gamma}\right),$$

(3.56)

and primes on the Hansen coefficients denote derivatives with respect to eccentricity.

The values of some of these drift and diffusion coefficients as functions of eccentricity and harmonic frequency are shown in figures 3.4–3.10; we focus on the coefficients pertaining to the period and eccentricity of the binary, as these are the most important for our envisaged observational applications. These coefficients, describing the evolution of the size and shape of the orbit, are independent of the remaining four orbital elements ($I, \Omega, \omega, \varepsilon$), which all describe the orientation of the orbit in space. In figures 3.4–3.8 we plot the coefficients as functions of eccentricity, assuming various power-law GWB spectra $\Omega(f) \sim f^\alpha, \alpha \in [-2, +2]$. There are two particularly striking features that are worth mentioning: the first is that $D_{Pe}^{(2)}$ and $D_{ee}^{(2)}$ both vanish as $e \to 1$, regardless of the GWB spectrum; and the second is that $D_e^{(1)}$ changes sign at an eccentricity $e \approx 0.4$ that is approximately (though not exactly) independent of the GWB spectrum; the significance of this near-universal crossover eccentricity is not immediately clear.

In figures 3.9 and 3.10 we plot the contributions from each of the binary's harmonic frequencies $f = n/P$. We see that the evolution of the period is driven by the $n = 2$ harmonic for near-circular binaries, with all other harmonics having zero contribution to $D_{PP}^{(2)}$ and $D_P^{(1)}$ in the circular limit $e \to 0$. This mirrors the frequency content of GW *emission* from binaries, which is also dominated by the $n = 2$ harmonic for small eccentricity, as we saw in section 0.3.1. As we discussed there, the significance of $n = 2$ can be understood by recognising that advancing a circular orbit in time by $P/2$ (i.e. the inverse of the $n = 2$ harmonic) is equivalent to exchanging the positions of the two bodies, resulting in a setup which has the same GW emission and absorption properties. The eccentricity evolution, on the other hand, is dominated by the $n = 1$ and $n = 3$ harmonics when e is small. In all cases, the contribution from higher harmonics becomes stronger for larger eccentricities, as the Fourier spectrum of the binary's response to GWs becomes richer. (Again, this is the same qualitative pattern that one finds in the case of GW *emission* by the binary.) In the limit $e \to 1$, we see a very simple pattern emerge in which each coefficient approaches a power law in the harmonic number n (except for $D_{ee}^{(2)}$, which vanishes in this limit).

The qualitative link between the GW emission and absorption spectra of binaries noted above is intriguing, and raises the question of whether this relationship can be established more formally. This might lead to deeper insights into binary–GW interactions, for example by allowing us to prove something akin to a fluctuation-dissipation theorem for this system. It would also be interesting to make contact with existing results on the scattering of GWs by binaries [75]. However, on the face of it there are several important differences between the GW absorption and emission processes: for example, the masses of the orbiting bodies are crucial in determining the radiated GW flux, but have no influence at all on GW absorption, since the GW-induced oscillations in the orbital separation are independent of the masses. (This last statement is a manifestation of the equivalence principle.) We leave a more thorough exploration of these questions for future work.
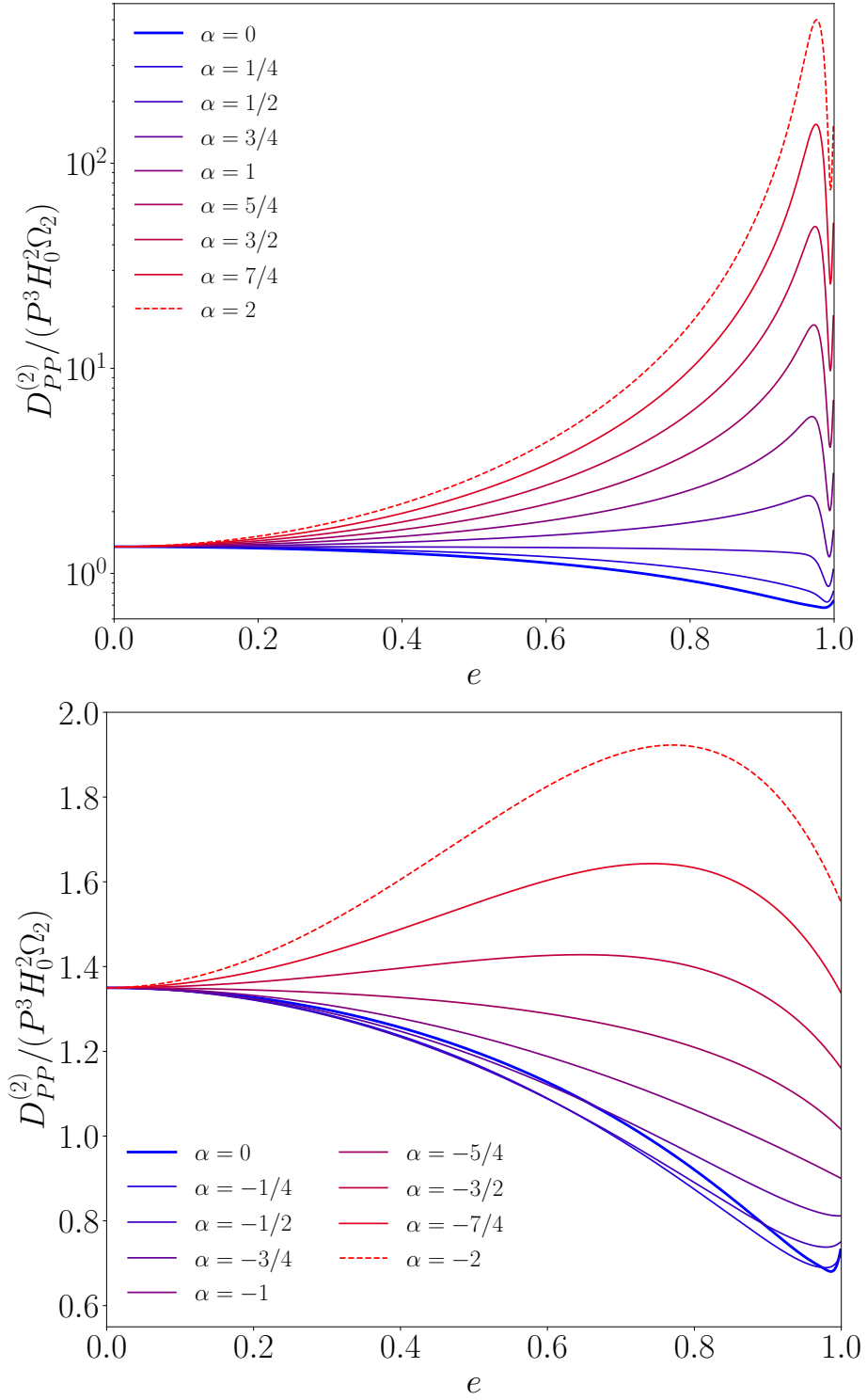
Figure 3.6: The secular period-period diffusion coefficients $D_{PP}^{(2)}$ as a function of eccentricity for various power-law GWB spectra, $\Omega(f) \sim f^\alpha$. The top panel shows positive power-law indices, $\alpha = 0, \ldots, 2$, while the bottom panel shows negative indices, $\alpha = 0, \ldots, -2$.

Inserting the coefficients in equations (3.55) and (3.54) into equation (3.34), we obtain a FPE which completely describes the secular evolution of a general binary system under GWB resonance; this is the main result of our analysis. Sections 3.4, 3.5, and 3.6 explore various strategies for solving this equation, and for using these solutions to place constraints on the GWB spectrum. Before moving on, we derive some simplified expressions for the KM coefficients in the cases where the eccentricity and/or the inclination are small.

3.3.2 SMALL-ECCENTRICITY AND SMALL-INCLINATION ORBITS

For binaries with very small eccentricity, we want to recast the results above in terms of the alternative orbital elements $(P, \zeta, \kappa, I, \Omega, \xi)$, as defined in equation (3.18). We do this using the coordinate transformation laws for the KM coefficients (see, e.g., section 4.9 of Risken [484] for a derivation),

$$D_i^{(1)} = \frac{\partial X_i}{\partial X_{i'}} D_{i'}^{(1)} + \frac{\partial^2 X_i}{\partial X_{i'} \partial X_{j'}} D_{i'j'}^{(2)}, \qquad D_{ij}^{(2)} = \frac{\partial X_i}{\partial X_{i'}} \frac{\partial X_j}{\partial X_{j'}} D_{i'j'}^{(2)}, \tag{3.57}$$

where summation over the primed indices is implied. Neglecting terms of order $e^2 \sim \zeta^2 \sim \zeta\kappa \sim \kappa^2$, we find the drift coefficients

$$D_P^{(1)} = V_P + \frac{3P^2}{160} H_0^2 (-79\Omega_1 + 288\Omega_2 - 27\Omega_3),$$

$$D_\zeta^{(1)} = V_\zeta + \frac{P}{160} \frac{\zeta}{\zeta^2 + \kappa^2} H_0^2 (25\Omega_1 - 27\Omega_3), \qquad D_\kappa^{(1)} = V_\kappa + \frac{P}{160} \frac{\kappa}{\zeta^2 + \kappa^2} H_0^2 (25\Omega_1 - 27\Omega_3),$$

$$D_I^{(1)} = \frac{3P}{80} H_0^2 \Omega_2 \cot^2 I, \qquad D_\Omega^{(1)} = 0, \qquad D_\xi^{(1)} = V_\xi,$$

$$\tag{3.58}$$

where the deterministic drift terms are given by

$$V_\zeta = \frac{6\pi v_P^2}{P} \left(\kappa - \tfrac{304}{45} \eta v_P^3 \zeta \right), \qquad V_\kappa = -\frac{6\pi v_P^2}{P} \left(\zeta + \tfrac{304}{45} \eta v_P^3 \kappa \right), \qquad V_\xi = -\frac{4\pi v_P^2}{P} (3 - \eta), \quad \tag{3.59}$$

and the diffusion coefficients are

$$D_{PP}^{(2)} = \frac{27P^3}{20} H_0^2 \Omega_2, \qquad D_{P\zeta}^{(2)} = -\frac{3P\zeta}{160} H_0^2 (25\Omega_1 + 12\Omega_2 - 27\Omega_3),$$

$$D_{P\kappa}^{(2)} = -\frac{3P\kappa}{160} H_0^2 (25\Omega_1 + 12\Omega_2 - 27\Omega_3), \qquad D_{\zeta\zeta}^{(2)} = D_{\kappa\kappa}^{(2)} = \frac{P}{160} H_0^2 (29\Omega_1 + 9\Omega_3),$$

$$D_{\zeta\Omega}^{(2)} = -\frac{3P\kappa}{80} H_0^2 \Omega_2 \frac{\cos I}{\sin^2 I}, \qquad D_{\zeta\xi}^{(2)} = -\frac{P\kappa}{320} H_0^2 \left[203\Omega_1 - 12\Omega_2 (20 + \cot^2 I) + 63\Omega_3 \right],$$

$$D_{\kappa\Omega}^{(2)} = \frac{3P\zeta}{80} H_0^2 \Omega_2 \frac{\cos I}{\sin^2 I}, \qquad D_{\kappa\xi}^{(2)} = \frac{P\zeta}{320} H_0^2 \left[203\Omega_1 - 12\Omega_2 (20 + \cot^2 I) + 63\Omega_3 \right],$$

$$D_{II}^{(2)} = \frac{3P}{80} H_0^2 \Omega_2, \qquad D_{\Omega\Omega}^{(2)} = \frac{3P}{80} H_0^2 \Omega_2 \csc^2 I, \qquad D_{\Omega\xi}^{(2)} = -\frac{3P}{80} H_0^2 \Omega_2 \frac{\cos I}{\sin^2 I},$$

$$D_{\xi\xi}^{(2)} = \frac{3P}{80} H_0^2 \Omega_2 (16 + \cot^2 I), \qquad D_{PI}^{(2)} = D_{P\Omega}^{(2)} = D_{P\xi}^{(2)} = D_{\zeta\kappa}^{(2)} = D_{\zeta I}^{(2)} = D_{\kappa I}^{(2)} = D_{I\Omega}^{(2)} = D_{I\xi}^{(2)} = 0.$$
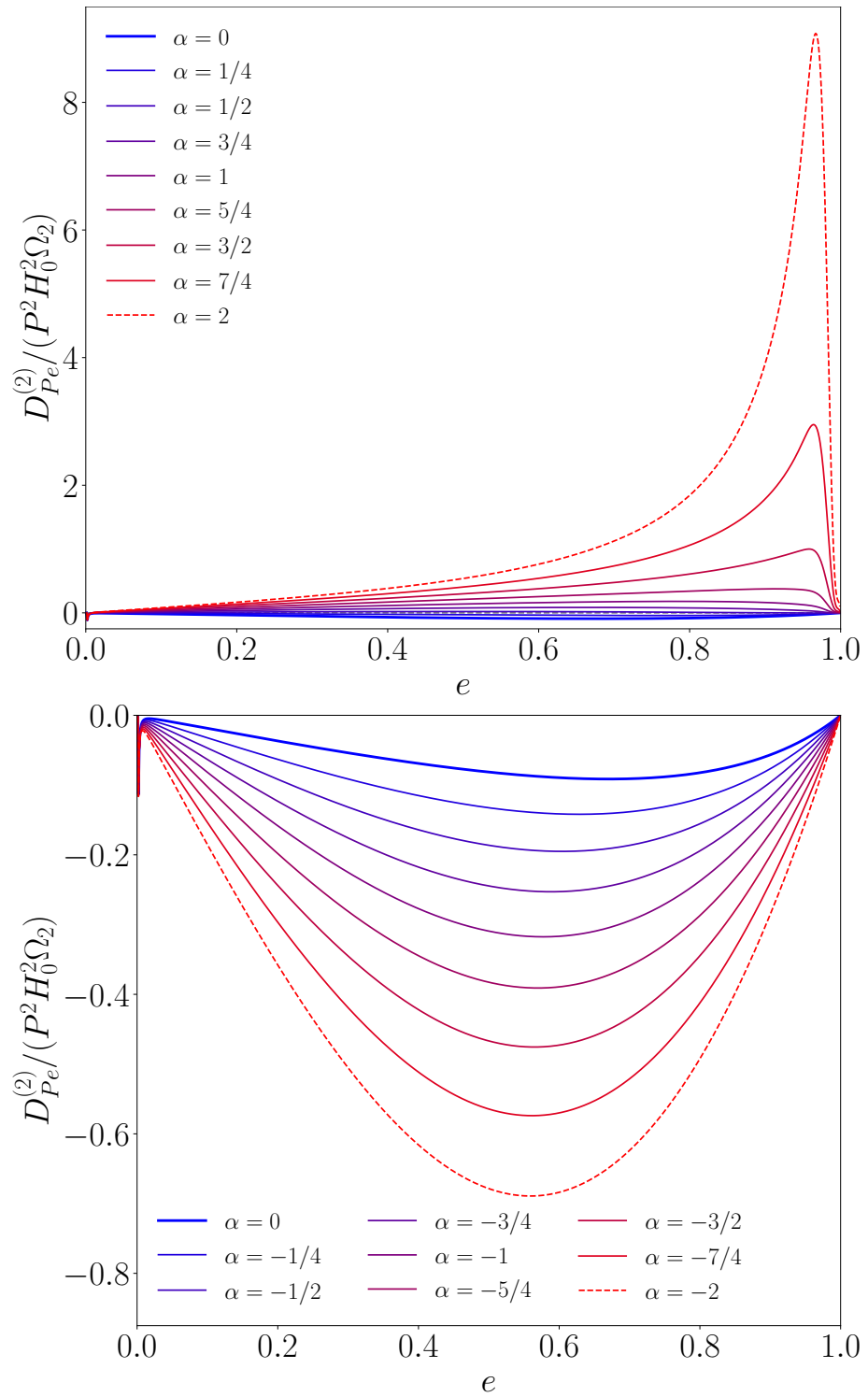
$$\tag{3.60}$$

Figure 3.7: Same as figure 3.6, but for the period-eccentricity diffusion coefficient $D_{Pe}^{(2)}$.

(We have confirmed that the $D_{PP}^{(2)}$ coefficient here matches the corresponding result in Hui *et al.* [323].)

We see from equation (3.58) that the stochastic drift term for the period P is usually positive, and can thus be interpreted physically as describing the softening of the binary due to the absorption of energy from the GWB (the term can become negative if Ω_2 is significantly smaller than Ω_1 and/or Ω_3, but GWB spectra typically vary sufficiently slowly with frequency that this does not occur). Interestingly, this implies that the net secular drift of the binary period (deterministic plus stochastic) generally changes sign at some critical value of P; e.g., for a scale-invariant GWB $\Omega(f) =$ constant, this value is given by

$$P_c = \left(\frac{1024\pi\eta}{91H_0^2\Omega}\right)^{3/11}(2\pi G\mathcal{M})^{5/11} \approx 95 \text{ yr} \times \left(\frac{\Omega}{10^{-6}}\frac{1/4}{\eta}\right)^{-3/11}\left(\frac{M}{m_\odot}\right)^{5/11}, \tag{3.61}$$

which corresponds to a semi-major axis of

$$a_c \approx 21 \text{ au} \times \left(\frac{\Omega}{10^{-6}}\frac{1/4}{\eta}\right)^{-2/11}\left(\frac{M}{m_\odot}\right)^{1/3}. \tag{3.62}$$

Binaries with $P < P_c$ will decay through GW emission, decreasing their period further, whereas binaries with $P > P_c$ undergo a net softening through GWB absorption, leading to a further increase in their period. The point $P = P_c$ is thus an unstable fixed point of the Langevin equation for P.[3] Note however that random diffusion due to $D_{PP}^{(2)}$ acts on a similar timescale (see figure 3.11 and equations (3.72) and (3.73)), and can easily push the system either side of this critical point.

We find a similar phenomenon for the eccentricity. Transforming back from the Laplace-Lagrange variables ζ, κ for now, equations (3.58) and (3.60) give

$$D_e^{(1)} = V_e + \frac{9P}{80e}H_0^2(3\Omega_1 - \Omega_3), \tag{3.63}$$

so that the (usually positive) stochastic drift diverges as $e \to 0$, while the (always negative) deterministic part vanishes as $e \to 0$. The net eccentricity drift thus changes sign at a critical value, just as it does for the period. For example, assuming a scale-invariant GWB spectrum, this critical value is

$$e_c = \sqrt{\frac{27H_0^2\Omega}{4864\pi\eta v_P^5}} \approx 5.9 \times 10^{-5} \times \left(\frac{\Omega}{10^{-6}}\frac{1/4}{\eta}\right)^{1/2}\left(\frac{M}{m_\odot}\right)^{-5/6}\left(\frac{P}{\text{yr}}\right)^{11/6}. \tag{3.64}$$

Since $\partial_e D_e^{(1)} < 0$ at this point, e_c is a stable fixed point of the corresponding Langevin equation: binaries with larger eccentricity will tend to circularise through GW emission until they reach e_c, while binaries with smaller eccentricity will on average have their eccentricity excited through GWB resonance. This is particularly interesting from an observational point of view, as it suggests that eccentricities smaller than

---

[3]We can understand this instability through a thermodynamic analogy. In the absence of the GWB, a binary system radiates GW energy with increasing intensity as it inspirals; its dynamical 'temperature' grows as it loses energy, meaning that it has a *negative heat capacity*—see footnote 34 of chapter 0. Similarly, the GWB acts as a heat reservoir and imparts energy to the binary (on average), therefore slowing the orbital motion and decreasing the system's temperature. A conventional thermodynamic system with positive heat capacity would equilibrate at a point where the heat loss from GW radiation balanced the heat gain from the GWB. Instead, the binary undergoes a runaway increase or decrease in its temperature due to its negative heat capacity.
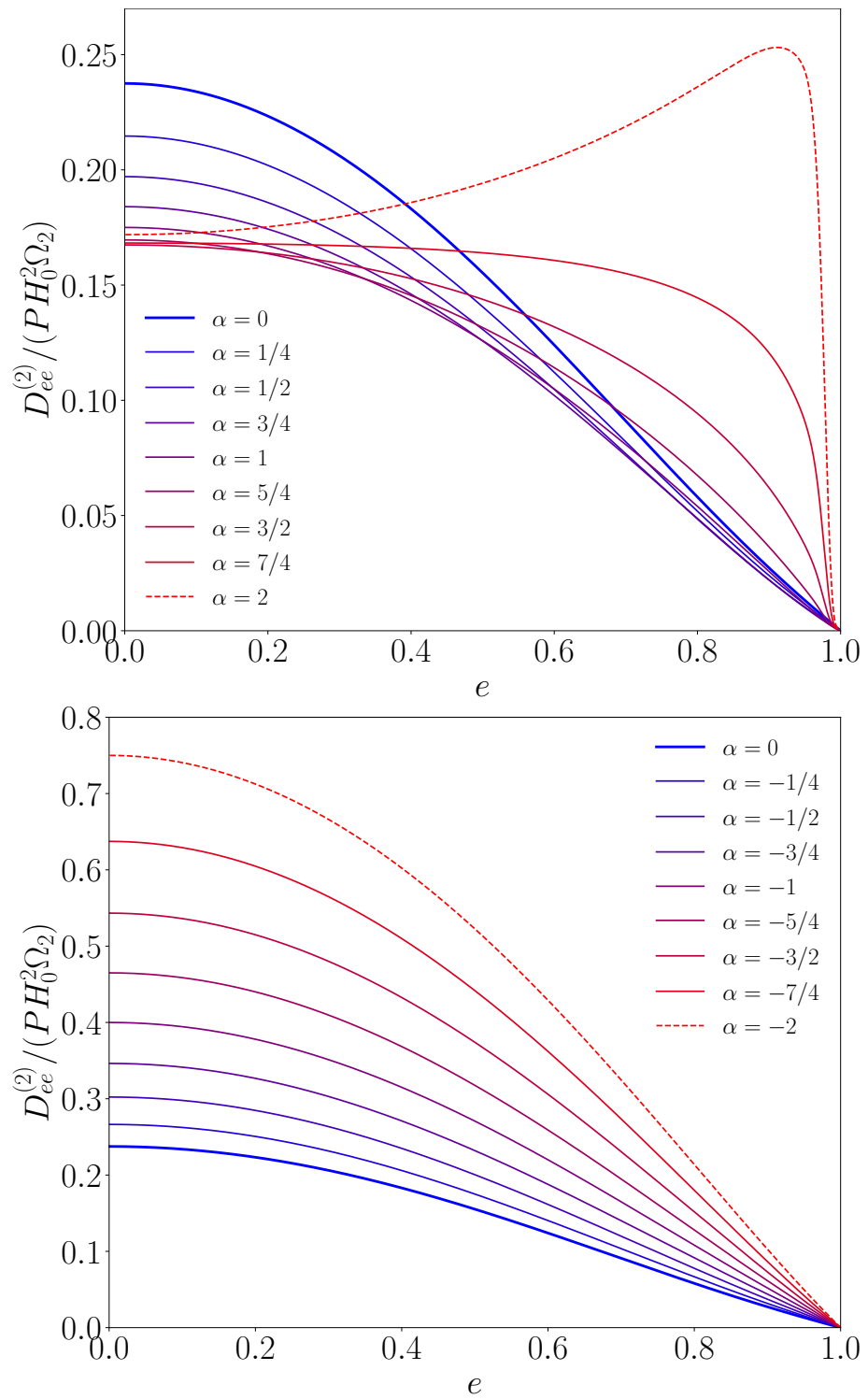
Figure 3.8: Same as figure 3.6, but for the eccentricity-eccentricity diffusion coefficient $D_{ee}^{(2)}$.

e_c might be less frequently observed in sufficiently old systems, though random diffusion due to $D_{ee}^{(2)}$ can still push systems below this point.

We also see from equation (3.58) that the stochastic drift term for the inclination changes sign at $I = \pi/2$ (this is also true in the general-eccentricity case, see equation (3.55)). Since $\partial_I D_I^{(1)} < 0$ at $I = \pi/2$, this is a stable fixed point of the corresponding Langevin equation; stochastic drift will, on average, drive binaries toward $I = \pi/2$. This effect is counteracted, however, by binaries diffusing away from $I = \pi/2$. As we show explicitly in section 3.4, on extremely long timescales these two effects balance each other, leaving an isotropic distribution for the inclination.

In the limit where the binary's inclination is also small, we rewrite the KM coefficients in terms of the orbital elements $(P, \zeta, \kappa, p, q, \lambda)$. This gives the drift coefficients

$$D_p^{(1)} = -\frac{Pp}{40} H_0^2 \Omega_2, \qquad D_q^{(1)} = -\frac{Pq}{40} H_0^2 \Omega_2, \tag{3.65}$$

and the diffusion coefficients,

$$D_{\zeta\lambda}^{(2)} = -\frac{P\kappa}{320} H_0^2 (203\Omega_1 - 240\Omega_2 + 63\Omega_3), \qquad D_{\kappa\lambda}^{(2)} = \frac{P\zeta}{320} H_0^2 (203\Omega_1 - 240\Omega_2 + 63\Omega_3),$$

$$D_{pp}^{(2)} = D_{qq}^{(2)} = \frac{3P}{80} H_0^2 \Omega_2, \qquad D_{p\lambda}^{(2)} = \frac{3Pq}{160} H_0^2 \Omega_2 \qquad D_{q\lambda}^{(2)} = -\frac{3Pp}{160} H_0^2 \Omega_2 \qquad D_{\lambda\lambda}^{(2)} = \frac{3P}{5} H_0^2 \Omega_2,$$

$$D_{Pp}^{(2)} = D_{Pq}^{(2)} = D_{P\lambda}^{(2)} = D_{\zeta\kappa}^{(2)} = D_{\zeta p}^{(2)} = D_{\zeta q}^{(2)} = D_{\kappa p}^{(2)} = D_{\kappa q}^{(2)} = D_{pq}^{(2)} = 0,$$

$$\tag{3.66}$$

where the coefficients not listed are identical to those in equations (3.58) and (3.60). Note that this implies $D_I^{(1)} = 3PH_0^2 \Omega_2/(80I)$, which diverges as $I \to 0$. This means that binaries are quickly excited away from zero inclination, similar to what happens for the eccentricity as $e \to 0$.

3.4 SOME EXACT RESULTS FOR CIRCULAR BINARIES

We have shown that the osculating orbital elements of a binary coupled to the GWB evolve according to a nonlinear six-dimensional FPE, for which no analytical solution is generally available. However, in the small-eccentricity limit $e \to 0$, we found in equations (3.58) and (3.60) that the drift and diffusion of the binary period P are independent of all of the other orbital elements. This allows us to treat P separately by solving the one-dimensional FPE

$$\frac{\partial W}{\partial t} = -\frac{\partial J}{\partial P}, \tag{3.67}$$

where $W(P, t)$ is now the single-variable DF for P, marginalised over the other orbital elements, and where

$$J(P, t) \equiv D^{(1)} W - \partial_P (D^{(2)} W) \tag{3.68}$$

is the *probability current* [280, 484]. (We have suppressed the P subscripts on the drift and diffusion coefficients for this single-variable case.) In this section, we derive some exact results for this simplified equation. These results highlight the power of our Fokker-Planck formalism, which allows us to answer these questions about the full shape of the DF in a way that previous analyses are unable to.
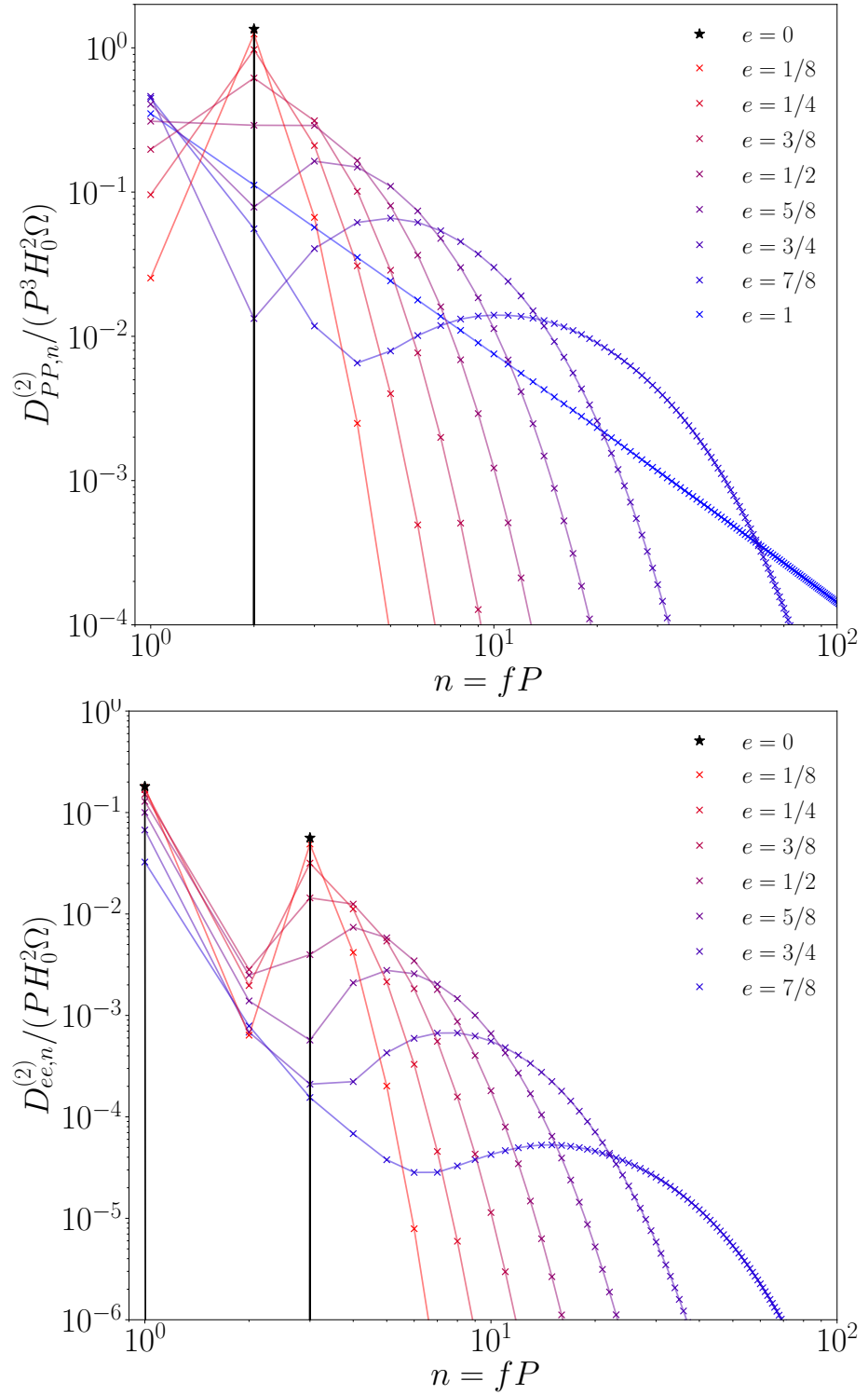
Figure 3.9: Contributions to the secular diffusion coefficients $D_{PP}^{(2)}$ (top panel) and $D_{ee}^{(2)}$ (bottom panel) from different harmonic frequencies, for binaries with various eccentricities $e = 0, \ldots, 1$. (The subscript 'n' here indicates that we have extracted the contribution from the nth harmonic.) We set $\Omega(f) = \text{constant}$ here, so that each harmonic receives equal weighting; alternative GWB spectra will give different weighting to each harmonic. Note that nearly all of the harmonics vanish in the circular case, $e = 0$, so that the few harmonics that do contribute appear as vertical lines. Note also that $D_{ee}^{(2)} = 0$ when $e \to 1$.

3.4.1 QUASI-STATIONARY PERIOD DISTRIBUTION

The simplest kind of solution to look for is a stationary (i.e., time-independent) distribution, corresponding to constant probability current throughout the parameter space. Setting J = constant in equation (3.68), we can use the integrating factor (which is defined up to an arbitrary constant multiplicative factor)

$$I(P) \equiv \exp\left(\int \mathrm{d}P \, \frac{D^{(1)}}{D^{(2)}}\right) \tag{3.69}$$

to obtain

$$W = \frac{I(P)}{D^{(2)}}\left(C - J \int \frac{\mathrm{d}P}{I(P)}\right), \tag{3.70}$$

with C a constant which, for a given value of J, is fixed by the normalisation of the DF. Clearly, the functional form of equation (3.70) depends on the GWB energy spectrum $\Omega(f)$. As a simple example, consider a scale-invariant GWB, $\Omega(f)$ = constant, which has

$$D^{(1)} = V_P + \frac{273}{80}P^2 H_0^2 \Omega, \qquad D^{(2)} = \frac{27}{20}P^3 H_0^2 \Omega, \qquad I(P) \propto P^{91/36} \exp\left[\frac{91}{132}\left(\frac{P_c}{P}\right)^{11/3}\right]. \tag{3.71}$$

A full exploration of spectra beyond this simple scale-invariant case is beyond our scope here, but we expect the qualitative results of this section to be reasonably robust to this choice.

We can fix J and C by imposing the appropriate boundary conditions. For some minimum value of P the binary merges or is tidally disrupted, whereas for some maximum value the binary becomes gravitationally unbound, so at both extremes we require absorbing boundary conditions—i.e., the DF must go to zero at both boundaries. Systems with absorbing boundary conditions do not admit nonzero stationary solutions [280]; formally, the conditional probability of the binary having period P at time t, given that it initially had period P_0 at time zero, obeys $\lim_{t\to\infty} W(P, t|P_0, 0) = 0$ across the entire parameter space. Intuitively, this is because all of the initial probability mass is eventually absorbed by one or other of the boundaries. We can also understand this in terms of the instability discussed in section 3.3.2; binaries either side of the critical period P_c undergo a runaway evolution away from this point, reaching one of the two boundaries within finite time, thus leaving an empty distribution in the limit $t \to \infty$.

In practice, however, the timescale over which the binary evolves is set by its period, and binaries with short periods will approach stationarity much faster than binaries with long periods. More concretely, for a flat GWB spectrum we have a drift timescale

$$\tau_{\mathrm{drift}} \equiv \frac{P}{|D^{(1)}|} \simeq \begin{cases} \dfrac{80}{273 P H_0^2 \Omega}, & P \gg P_c, \\ \dfrac{5 P^{8/3}}{192\pi\eta(2\pi G M)^{5/3}}, & P \ll P_c, \end{cases} \tag{3.72}$$

and a diffusion timescale

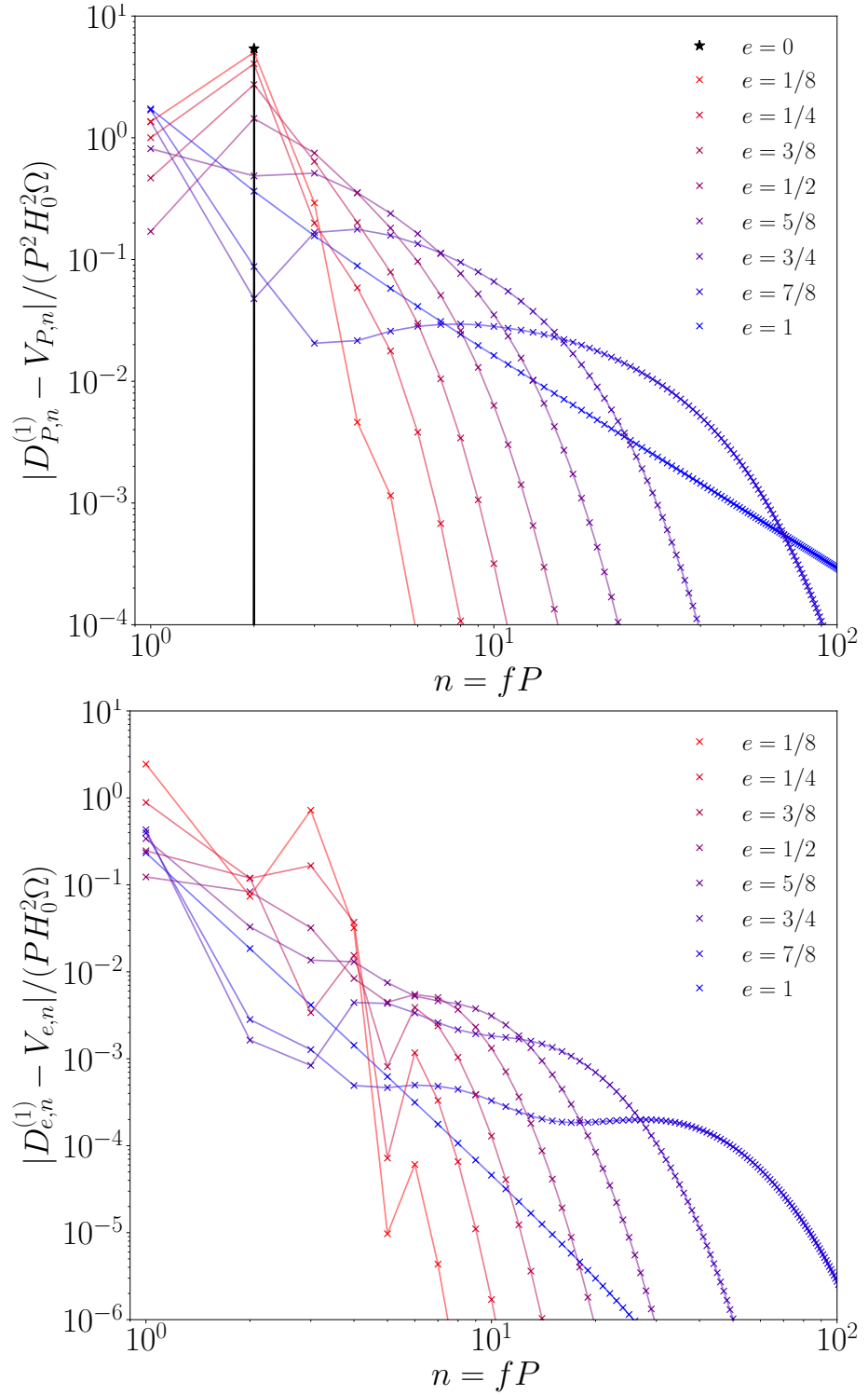$$\tau_{\mathrm{diff}} \equiv \frac{P^2}{|D^{(2)}|} = \frac{20}{27 P H_0^2 \Omega}. \tag{3.73}$$

Figure 3.10: Contributions to the stochastic parts of the secular drift coefficients $D_P^{(1)}$ (top panel) and $D_e^{(1)}$ (bottom panel) from different harmonic frequencies, for binaries with various eccentricities $e = 0, \ldots, 1$. (The subscript 'n' here indicates that we have extracted the contribution from the nth harmonic.) We show the absolute values, as the drift coefficients have both positive and negative contributions. Note that $D_e^{(1)} \to +\infty$ as $e \to 0$.
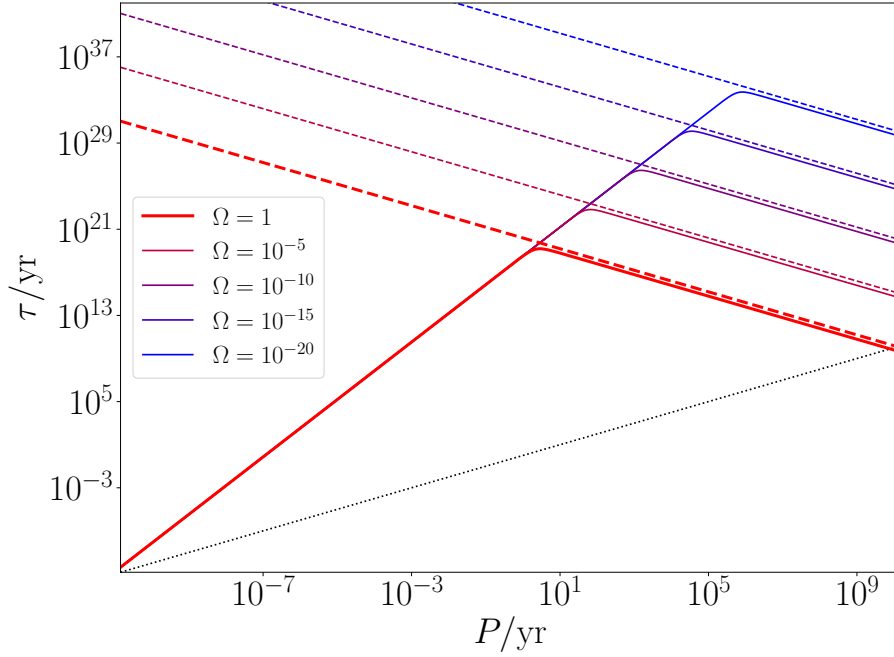
Figure 3.11: Evolution timescales for the DF of a circular binary with total mass $M = m_\odot$ immersed in a scale-invariant GWB $\Omega(f) =$ constant. Solid curves show the drift timescale (3.72), which is dominated by GW emission at 'short' periods $P \lesssim P_c$ and by GW absorption from the GWB at 'long' periods $P \gtrsim P_c$. Dashed curves show the diffusion timescale (3.73). The dotted black curve shows $\tau = P$; our secular-averaging assumption only holds in the region above this curve.

As shown in figure 3.11, the fastest timescale is many orders of magnitude shorter near the lower boundary than it is near the upper boundary. It may be reasonable, therefore, to look for 'quasi-stationary' solutions, where equilibrium is established near the lower boundary, but where the boundary condition for long periods is neglected. We achieve this by choosing J and C such that the DF goes to zero at the lower boundary, and such that the DF is normalised, but without enforcing any condition at the upper boundary.

As a simple example, consider a scale-invariant GWB spectrum $\Omega(f) =$ constant, and fix the DF to be zero at the period corresponding to the binary's innermost stable circular orbit,

$$P_{\text{ISCO}} \equiv 6^{3/2} \times 2\pi G M, \tag{3.74}$$

with the binary assumed to merge at periods shorter than this. The corresponding quasi-stationary distribution is then given in terms of the dimensionless variable $\varrho \equiv P/P_{\text{ISCO}}$ by

$$W_{\text{qs}} \propto \exp\left(\frac{\lambda}{\varrho^{11/3}}\right)\left[\frac{E_{77/132}(\lambda \varrho^{-11/3})}{\varrho^2} - \frac{E_{77/132}(\lambda)}{\varrho^{17/36}}\right], \tag{3.75}$$

where $E_n(z) \equiv \int_1^\infty \mathrm{d}t\, \mathrm{e}^{-zt} t^{-n}$ is the exponential integral function, and

$$\lambda \equiv \frac{91}{132}\left(\frac{P_c}{P_{\text{ISCO}}}\right)^{11/3} = \frac{\sqrt{2}\eta (GMH_0)^{-2}}{8019\sqrt{3}\pi\Omega} \gg 1 \tag{3.76}$$

is a dimensionless constant which quantifies the strength of the deterministic drift V_P relative to the secular diffusion $D^{(2)}$. While the functional form of equation (3.75) is somewhat opaque, one can show that it approximately interpolates between $W_{qs} \sim P^{5/3}$ for $P_{ISCO} \ll P \ll P_c$ and $W_{qs} \sim P^{-17/36}$ for $P \gg P_c$ (this broken-power-law behaviour is clearly seen in figure 3.12). We normalise equation (3.75) by integrating up to some maximum period, which we choose to be the age of the Universe. By substituting equation (3.75) into equation (3.68) and evaluating at $\varrho = 1$ where $W = 0$ and $\partial W/\partial P > 0$, we see that the probability current is strictly negative, $J < 0$. Since we specified that $J = $ constant, this means there is a uniform net flow towards shorter periods throughout the parameter space, regardless of how soft the binary is.

For hard binaries with $P < P_c$ it is obvious that we should have $J < 0$, as the negative deterministic drift is the most important effect and quickly drives the binary toward merger. For soft binaries with $P > P_c$ the negative probability current is less immediately obvious; it shows us that, while there is a net drift $D^{(1)}$ pushing these binaries towards longer periods, on average they are nonetheless expected to flow towards shorter periods. We can understand this somewhat counter-intuitive behaviour by including diffusion as well as drift effects. Indeed, note that by equation (3.68), the condition $J < 0$ is equivalent to

$$\frac{\partial(\ln D^{(2)}W)}{\partial(\ln P)} > \frac{P D^{(1)}}{D^{(2)}} = \frac{91}{36}\left[1 - (P/P_c)^{-11/3}\right], \tag{3.77}$$

where the RHS tends to a positive constant value in the $P \gg P_c$ region we are interested in. What this is telling us is that, despite the positive drift coefficient, we can still have a negative probability current if and only if the diffusion coefficient $D^{(2)}$ grows sufficiently quickly with P, as this makes it sufficiently likely for the binary's random walk to wander below P_c and then rapidly approach short periods through GW emission. Interestingly, we find that for the quasi-stationary distribution we have $\partial(\ln D^{(2)}W_{qs})/\partial(\ln P) \simeq 91/36$, so this diffusive effect is *only just* strong enough to cause a net negative probability current.

We can repeat this process for any given GWB spectrum to write down a corresponding quasi-stationary solution for the period of a circular binary. However, the time taken to relax to this distribution is extremely long for typical GWB spectra (see figure 3.11), so these solutions are physically uninteresting in most cases. Besides, the assumption of a perfectly circular binary is overly simplistic, as we have shown in section 3.3.2 that the eccentricity distribution relaxes away from zero on shorter timescales. Nonetheless, the approach in this section is still useful for building intuitive understanding of the dynamics of the full DF, and may be useful for, e.g., studies of the orbital element distributions of old populations of binaries, where the full shape of the distribution is vitally important. Finding quasi-stationary solutions for the full multivariate FPE is much more challenging, and for eccentric binaries there is no guarantee that such a solution with the appropriate lower boundary condition even exists.

3.4.2 Mean coalescence time

For any one-dimensional FPE, it is possible to write down an explicit formula for the *mean first passage time* at either of its boundaries—i.e., the average time taken for an individual random trajectory to reach that boundary, as a function of the initial position [280]. In our case, this is a useful tool for understanding how
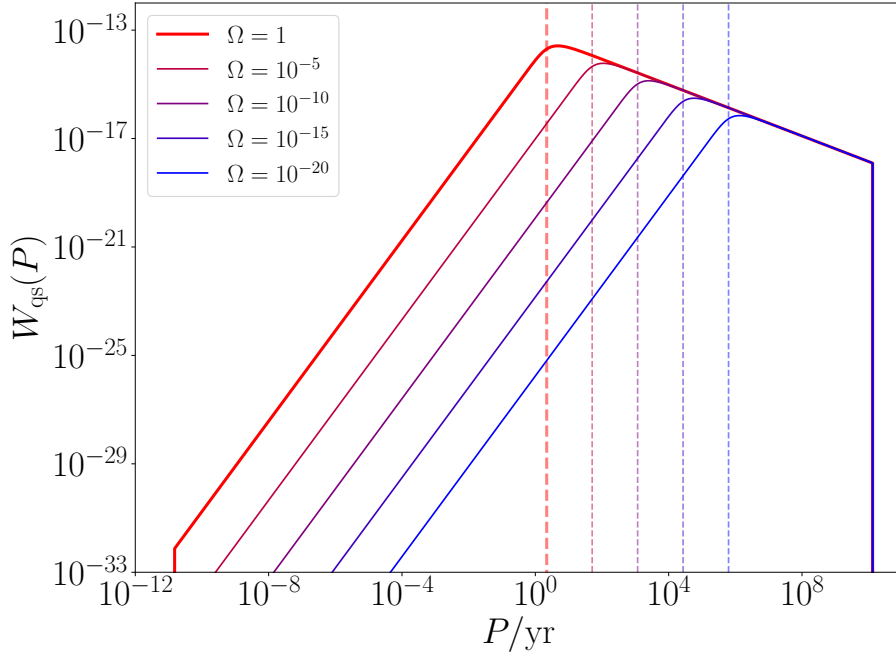
Figure 3.12: The quasi-stationary distribution (3.75) for the period P of a binary with mass $M = m_\odot$ coupled to a scale-invariant GWB, $\Omega(f) = $ constant. The upper and lower cutoffs are due to the age of the Universe and the binary's ISCO period (3.74), respectively. The dashed vertical lines indicate the value of P_c for each Ω, as defined by equation (3.61)—this is roughly the peak of the distribution in each case. Binaries with periods $P < P_c$ decay deterministically through GW emission and are removed from the distribution, whereas binaries with $P > P_c$ are supported against decay by resonant absorption of the GWB. As we lower the GWB intensity Ω, the resonance becomes weaker, and the binaries must have longer periods to avoid decay.

the presence of a particular GWB spectrum impacts upon the eventual fate of a binary system. Applying this to our lower absorbing boundary at the ISCO, we thus have the mean coalescence time of a binary coupled to the GWB, as a function of its initial period P_i,

$$\langle \tau(P_i) \rangle = \int_{P_{\mathrm{ISCO}}}^{P_i} \mathrm{d}P \int_P^{P_{\max}} \mathrm{d}P' \, \frac{I(P')}{I(P)D^{(2)}(P')}, \tag{3.78}$$

where $I(P)$ is the integrating factor defined in equation (3.69).

Returning to the example of a scale-invariant GWB, this becomes

$$\langle \tau(P_i) \rangle = \int_1^{\varrho_i} \frac{\mathrm{d}\varrho}{\varrho^{91/36}} \int_\varrho^{\varrho_{\max}} \frac{\mathrm{d}\varrho'}{\varrho'^{17/36}} \, \frac{20 \, \mathrm{e}^{\lambda(\varrho'^{-11/3} - \varrho^{-11/3})}}{27 P_{\mathrm{ISCO}} H_0^2 \Omega}. \tag{3.79}$$

This double integral is challenging to evaluate in general. However, we can easily verify equation (3.79) by showing that it reproduces the standard expression for the coalescence time due to deterministic GW emission in cases where the GWB resonance is weak. Setting $P_i \ll P_c$, we can safely take the limit $\lambda \to \infty$ and extract the leading-order and next-to-leading-order terms,

$$\langle \tau(P_i) \rangle \simeq \frac{405 GM}{16\eta} \left(\frac{P_i}{P_{\mathrm{ISCO}}} \right)^{8/3} \left[1 - \frac{96}{91} \mathrm{e}^{-\frac{91}{132}(P_c/P_i)^{11/3}} \frac{P_{\max}^{151/36}}{P_i^{19/36} P_c^{11/3}} \right]. \tag{3.80}$$

The leading-order term agrees with the deterministic coalescence time one finds by integrating V_P, as expected. There is a small negative contribution from the next-to-leading-order term, which indicates that GWB resonance slightly speeds up the coalescence in this regime. We can understand this by noticing that the term associated with $D^{(2)}$ in equation (3.68) is always negative for the quasi-stationary distribution, so that diffusion always has a net negative contribution to the probability current and thus, on average, always help drive binaries towards merger.

3.4.3 Including the remaining orbital elements

Having found a quasi-stationary period distribution, it is now relatively easy to obtain stationary distributions for the remaining orbital elements (I, Ω, ξ), so long as we hold ζ and κ fixed at zero. To do so, we write the full FPE as

$$\frac{\partial W}{\partial t} = -\partial_i J_i, \qquad J_i \equiv D_i^{(1)} W - \partial_j (D_{ij}^{(2)} W), \tag{3.81}$$

where W is now interpreted as the multivariate DF over the four orbital elements (P, I, Ω, ξ), and the four corresponding probability currents are

$$
\begin{aligned}
J_P &= D_P^{(1)} W - \partial_P (D_{PP}^{(2)} W), &\qquad J_I &= D_I^{(1)} W - D_{II}^{(2)} \partial_I W, \\
J_\Omega &= -D_{\Omega\Omega}^{(2)} \partial_\Omega W - D_{\Omega\xi}^{(2)} \partial_\xi W, &\qquad J_\xi &= V_\xi W - D_{\Omega\xi}^{(2)} \partial_\Omega W - D_{\xi\xi}^{(2)} \partial_\xi W.
\end{aligned}
\tag{3.82}
$$

We have used equations (3.58) and (3.60) to simplify these expressions, in particular using the independence of the diffusion terms from most of the orbital elements to take them outside the partial derivatives.

Let us first consider the inclination I. By definition, this is constrained to lie in the interval $[0, \pi]$. Unlike for the period P, a binary reaching one of the boundaries of this interval is not removed from the distribution; instead of absorbing boundary conditions, we have *reflecting* boundary conditions, i.e., the probability current J_I vanishes at both boundaries. However, if the distribution is stationary then $J_I = \text{constant}$, so the current must vanish everywhere on $[0, \pi]$. Setting $J_I = 0$ in equation (3.82) gives

$$\partial_I \ln W = D_I^{(1)}/D_{II}^{(2)} = \cot I. \tag{3.83}$$

Integrating this we find $W \propto \sin I$, which corresponds to an isotropic distribution (since $\cos I$ is uniformly distributed). This makes intuitive sense: on long timescales, GWB resonance causes the binary to 'forget' its initial orbital plane, such that the resulting stationary distribution is spherically symmetric. This also agrees with our finding in section 3.3.2 that $D_I^{(1)} = 0$ at $I = \pi/2$, which is the mean inclination of an isotropic distribution.

For Ω and ξ it is natural to impose periodic boundary conditions for the DF and for the probability currents,

$$
\begin{aligned}
W(P, I, \Omega, \xi, t) &= W(P, I, \Omega + 2\pi, \xi, t) = W(P, I, \Omega, \xi + 2\pi, t), \\
J_i(P, I, \Omega, \xi, t) &= J_i(P, I, \Omega + 2\pi, \xi, t) = J_i(P, I, \Omega, \xi + 2\pi, t).
\end{aligned}
\tag{3.84}
$$

Stationarity requires $\partial_\Omega J_\Omega = \partial_\xi J_\xi = 0$. By inspection, we see that this is achieved if $\partial_\Omega W = \partial_\xi W = 0$, so that the DF depends only on P and I. This corresponds to Ω and ξ being uniformly distributed, which

also satisfies the periodic boundary conditions; Ω then has zero probability current, while ξ has a uniform negative current due to the deterministic drift V_ξ, which depends only on the period.

3.5 Solving the full Fokker-Planck equation

We now consider the full FPE for all six orbital elements $(P, e, I, \Omega, \omega, \varepsilon)$. Allowing nonzero eccentricity $e > 0$ leads to much more complicated KM coefficients, and means that the results of section 3.4 are no longer applicable. Nonetheless, we can use the fact that diffusion of the orbital elements due to the GWB takes place on very long timescales $\tau_{\text{diff}} \sim 1/(P H_0^2 \Omega) \gg P$ (see figure 3.11). This allows us to develop some useful approximate solution schemes for much shorter observational timescales.

3.5.1 Perturbative short-time solution

The FPE can be written as an operator equation

$$\frac{\partial W}{\partial t} = L_{\text{FP}} W, \qquad L_{\text{FP}}(\boldsymbol{X}) \equiv -\partial_i D_i^{(1)} + \partial_i \partial_j D_{ij}^{(2)}. \tag{3.85}$$

This has the formal solution[4] $W(\boldsymbol{X}, t) = \exp(t L_{\text{FP}}) W(\boldsymbol{X}, 0)$, which for short times $t \ll \tau_{\text{diff}}$ can be expanded as

$$W(\boldsymbol{X}, t) = \left[1 + t L_{\text{FP}}(\boldsymbol{X}, 0) + O(t/\tau_{\text{diff}})^2\right] W(\boldsymbol{X}, 0), \tag{3.86}$$

where the right-hand side depends only on data at time zero.

Suppose that at time zero the binary's orbital elements take on the 'sharp' values x_i. The initial condition for the DF is then $W(\boldsymbol{X}, 0) = \delta^{(6)}(\boldsymbol{X} - \boldsymbol{x})$. By using a Fourier representation of the delta function, we can evaluate equation (3.86) to find [484]

$$W(\boldsymbol{X}, t) = \frac{1}{\sqrt{\det 4\pi t D^{(2)}}} \exp\left\{-\frac{\left[D^{(2)}\right]_{ij}^{-1}}{4t}\left(X_i - x_i - D_i^{(1)} t\right)\left(X_j - x_j - D_j^{(1)} t\right)\right\} + O\left(\frac{t}{\tau_{\text{diff}}}\right)^2, \tag{3.87}$$

i.e., on short timescales, the DF is a multivariate Gaussian with mean $x_i + D_i^{(1)} t$ and covariance matrix $2t D_{ij}^{(2)}$. Here $\left[D^{(2)}\right]_{ij}^{-1}$ represents the elements of the inverse of the diffusion matrix, and both the drift vector and diffusion matrix are evaluated at $(\boldsymbol{X}, t) = (\boldsymbol{x}, 0)$.

3.5.2 Evolution of moments of the orbital elements

The short-time expansion shows that on observational timescales the DF of the orbital elements is approximately Gaussian, and is therefore completely characterised by its first two moments: the mean and the covariance matrix, which we write as

$$\bar{X}_i \equiv \langle X_i \rangle, \qquad C_{ij} \equiv \text{Cov}[X_i, X_j] = \left\langle \left(X_i - \bar{X}_i\right)\left(X_j - \bar{X}_j\right)\right\rangle. \tag{3.88}$$

---

[4] Here we take advantage of the time-independence of the secular KM coefficients. For time-dependent coefficients, one would need to instead construct a time-ordered Dyson series [484], though this gives the same result for short times.

It is therefore useful to take moments of the FPE to find the time evolution of these quantities, rather than attempting to calculate the full time-dependent DF. In doing so, we can calculate the backreaction of perturbations on the evolution of the binary, obtaining corrections to the linear growth found in equation (3.87).

The first moment of the FPE gives

$$\frac{\mathrm{d}\bar{X}_i}{\mathrm{d}t} = \frac{\partial}{\partial t} \int \mathrm{d}\boldsymbol{X}\, X_i W = \int \mathrm{d}\boldsymbol{X}\, X_i \frac{\partial W}{\partial t} = -\int \mathrm{d}\boldsymbol{X}\, X_i \partial_j \left(D_j^{(1)} W \right) + \int \mathrm{d}\boldsymbol{X}\, X_i \partial_j \partial_k \left(D_{jk}^{(2)} W \right).$$
(3.89)

We integrate by parts, and assume that the DF falls off fast enough that all boundary terms vanish, leaving

$$\frac{\mathrm{d}\bar{X}_i}{\mathrm{d}t} = \int \mathrm{d}\boldsymbol{X}\, D_i^{(1)} W = \left\langle D_i^{(1)} \right\rangle.$$
(3.90)

This fall-off assumption is very reasonable here, as the diffusion rate is extremely small for realistic binaries, so the DF will only have support very near to the mean value. Doing the same for second moment gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\langle X_i X_j \right\rangle = \int \mathrm{d}\boldsymbol{X}\, X_i X_j \frac{\partial W}{\partial t} = -\int \mathrm{d}\boldsymbol{X}\, X_i X_j \partial_k \left(D_k^{(1)} W \right) + \int \mathrm{d}\boldsymbol{X}\, X_i X_j \partial_k \partial_\ell \left(D_{k\ell}^{(2)} W \right)$$
$$= \int \mathrm{d}\boldsymbol{X}\, \left(X_i D_j^{(1)} + X_j D_i^{(1)} \right) W + 2\int \mathrm{d}\boldsymbol{X}\, D_{ij}^{(2)} W = \left\langle X_i D_j^{(1)} \right\rangle + \left\langle X_j D_i^{(1)} \right\rangle + 2\left\langle D_{ij}^{(2)} \right\rangle.$$
(3.91)

We can combine these to give the evolution equation for the covariance matrix,

$$\frac{\mathrm{d}C_{ij}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\left\langle X_i X_j \right\rangle - \bar{X}_i \bar{X}_j \right) = \left\langle X_i D_j^{(1)} \right\rangle + \left\langle X_j D_i^{(1)} \right\rangle + 2\left\langle D_{ij}^{(2)} \right\rangle - \left\langle X_i \right\rangle \left\langle D_j^{(1)} \right\rangle - \left\langle X_j \right\rangle \left\langle D_i^{(1)} \right\rangle$$
$$= \mathrm{Cov}\left[X_i, D_j^{(1)} \right] + \mathrm{Cov}\left[X_j, D_i^{(1)} \right] + 2\left\langle D_{ij}^{(2)} \right\rangle.$$
(3.92)

3.5.3 THE SLOW-DIFFUSION APPROXIMATION

Equations (3.90) and (3.92) fully describe the evolution of the mean and covariance of the orbital elements. However, they are given in terms of ensemble averages over nonlinear functions of the orbital elements, which we cannot perform without knowing the full DF. Even if we were to evaluate them approximately by assuming a Gaussian distribution, the resulting expressions would be very cumbersome.

In the case where the variance is small and any given orbital element X_i is 'close' to its mean value $\langle X_i \rangle$ (in a probabilistic sense), one can instead Taylor expand an arbitrary function of the elements around the mean,

$$f(\boldsymbol{X}) = f(\bar{\boldsymbol{X}}) + \left(X_i - \bar{X}_i \right) \partial_i f(\bar{\boldsymbol{X}}) + \frac{1}{2}\left(X_i - \bar{X}_i \right) \left(X_j - \bar{X}_j \right) \partial_i \partial_j f(\bar{\boldsymbol{X}}) + \cdots,$$
(3.93)

so that the mean of the function is approximated by

$$\langle f(\boldsymbol{X}) \rangle \simeq f(\bar{\boldsymbol{X}}) + \frac{1}{2}C_{ij}\partial_i \partial_j f(\bar{\boldsymbol{X}}).$$
(3.94)

(Note that the first-order term in the expansion vanishes when taking the mean, as the first central moment is identically zero.)

We can justify using equation (3.94) by noting that the diffusion matrix calculated in section 3.3 is very small in most physical situations. To keep track of how this smallness propagates into the evolution equations, we introduce a formal small parameter ϵ (which we will later set to unity), writing $D_{ij}^{(2)} \rightarrow \epsilon D_{ij}^{(2)}$. For sharp initial conditions $\bar{X}_i = x_i$, $C_{ij} = 0$, we see from equation (3.92) that $C_{ij} = O(\epsilon)$, as the $\mathrm{Cov}[\boldsymbol{X}, D^{(1)}]$ terms are initially zero. We therefore also write $C_{ij} \rightarrow \epsilon C_{ij}$. We thus see that equation (3.94) is justified if we neglect terms of order ϵ^2. We call this the *slow-diffusion* approximation, as it relies on the fact that the timescale τ_{diff} over which the covariance grows is long compared to the observation time. Note that the stochastic contribution to the drift vector is generally of the same order as the diffusion matrix, so we also write

$$D_i^{(1)} = V_i + \epsilon \delta D_i^{(1)}, \tag{3.95}$$

where the stochastic term $\delta D_i^{(1)}$ is suppressed by a factor of ϵ.

Applying this approximation to equations (3.90) and (3.92), we obtain the moment evolution equations to first order in ϵ,

$$\begin{aligned} \frac{\mathrm{d}\bar{X}_i}{\mathrm{d}t} &= V_i + \epsilon \delta D_i^{(1)} + \frac{1}{2}\epsilon C_{jk}\partial_j\partial_k V_i + O(\epsilon^2), \\ \epsilon \frac{\mathrm{d}C_{ij}}{\mathrm{d}t} &= 2\epsilon D_{ij}^{(2)} + \epsilon C_{ik}\partial_k V_j + \epsilon C_{jk}\partial_k V_i + O(\epsilon^2), \end{aligned} \tag{3.96}$$

where the drift vector and diffusion matrix are both evaluated at $\bar{X}$.

We see that the evolution equation (3.96) for the mean orbital elements includes both $O(\epsilon^0)$ and $O(\epsilon^1)$ terms. For numerical reasons, it is convenient to separate these. We therefore write the mean orbital elements as

$$\bar{X}(t) = \bar{X}_0(t) + \epsilon \delta \bar{X}(t), \tag{3.97}$$

where $\bar{X}_0$ represents the values the elements would take in the absence of the GWB, which obey the deterministic evolution equation

$$\frac{\mathrm{d}\bar{X}_{0,i}}{\mathrm{d}t} = V_i(\bar{X}_0). \tag{3.98}$$

(Note that this separation of deterministic and stochastic parts of the drift is always possible, regardless of the size of the deterministic term.) Meanwhile, $\delta\bar{X}$ represents the deviation in the mean due to GWB resonance, and evolves according to

$$\epsilon\frac{\mathrm{d}\delta\bar{X}_i}{\mathrm{d}t} = V_i(\bar{X}) - V_i(\bar{X}_0) + \epsilon\delta D_i^{(1)}(\bar{X}) + \frac{1}{2}\epsilon C_{jk}\partial_j\partial_k V_i(\bar{X}) + O(\epsilon^2). \tag{3.99}$$

We can Taylor-expand the terms that are evaluated at $\bar{X} = \bar{X}_0 + \epsilon\delta\bar{X}$ to give

$$\epsilon\frac{\mathrm{d}\delta\bar{X}_i}{\mathrm{d}t} = \epsilon\delta\bar{X}_j\partial_j V_i(\bar{X}_0) + \epsilon\delta D_i^{(1)}(\bar{X}_0) + \frac{1}{2}\epsilon C_{jk}\partial_j\partial_k V_i(\bar{X}_0) + O(\epsilon^2). \tag{3.100}$$
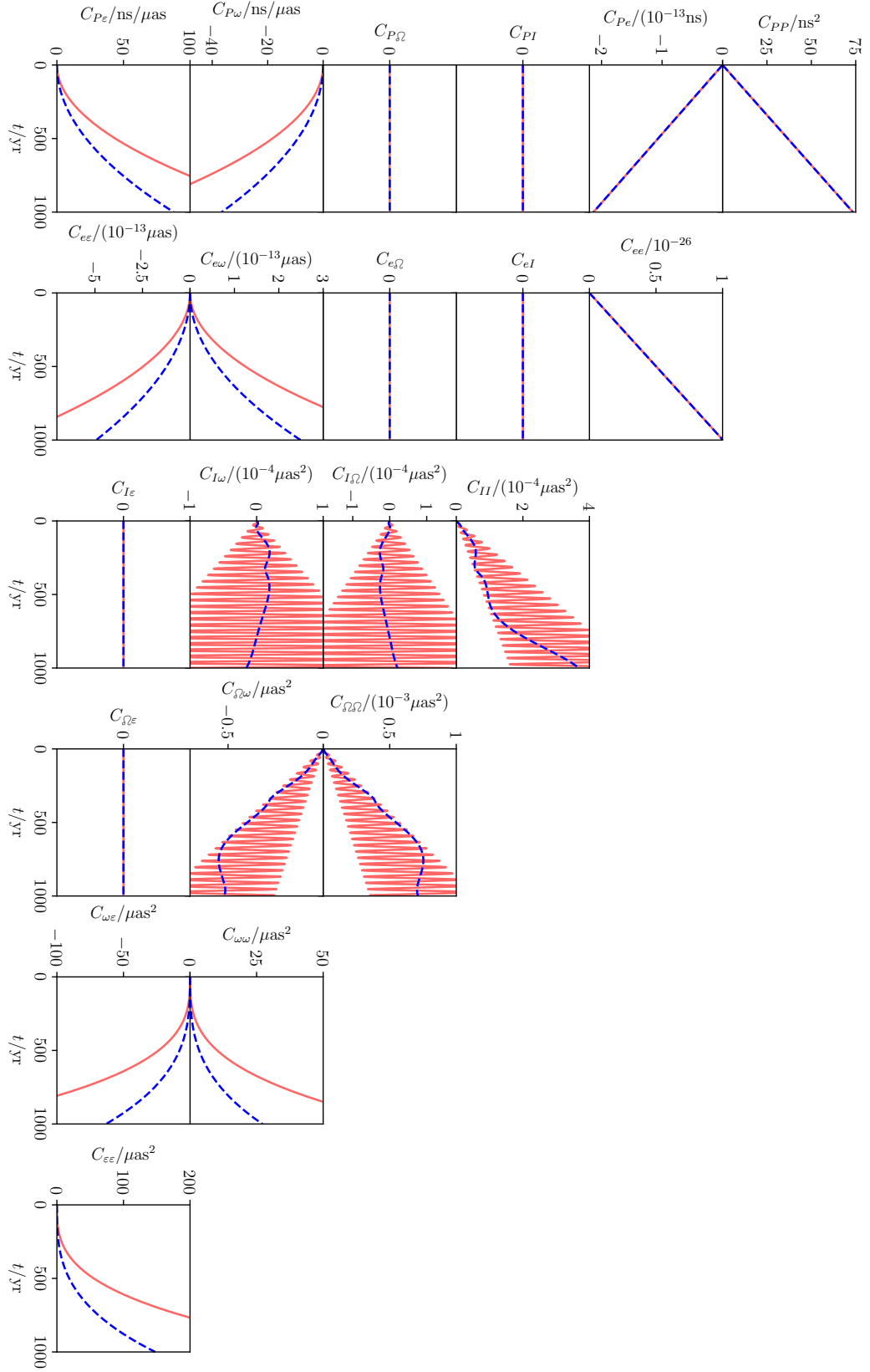
Figure 3.13: The covariance matrix $C_{ij}(t)$ of the orbital elements of the binary pulsar B1913+16 (the Hulse-Taylor system) over a 1 kyr interval, assuming a scale-invariant GWB spectrum $\Omega(f) = 10^{-5}$, and including the first 400 harmonics. The blue dashed curves show numerical solutions of the evolution equations (3.101), while the pale red curves show the naive solution $t \times dC_{ij}/dt$, which is exact only when $dC_{ij}/dt = $ constant.

With the appropriate terms identified, we can send $\epsilon \rightarrow 1$. Our full set of evolution equations, to leading order in the slow-diffusion approximation, reads

$$\frac{\mathrm{d}\bar{X}_{0,i}}{\mathrm{d}t} = V_i,$$

$$\frac{\mathrm{d}\delta\bar{X}_i}{\mathrm{d}t} \simeq \delta D_i^{(1)} + \delta\bar{X}_j \partial_j V_i + \frac{1}{2}C_{jk}\partial_j\partial_k V_i, \qquad (3.101)$$

$$\frac{\mathrm{d}C_{ij}}{\mathrm{d}t} \simeq 2D_{ij}^{(2)} + C_{ik}\partial_k V_j + C_{jk}\partial_k V_i,$$

with all terms evaluated at the deterministic mean elements $\bar{X}_0$. By writing the FPE in this form, we have replaced a six-dimensional, second-order PDE with 33 coupled, one-dimensional, first-order ODEs: six each for the deterministic mean elements $\bar{X}_{0,i}$ and the perturbations $\delta\bar{X}_i$, with the remaining 21 coming from the independent components of the 6×6 symmetric matrix C_{ij}. In figures 3.13 and 3.14 and in section 3.7 below, we explore the behaviour of these equations numerically with a purpose-built Python code `gwresonance`,[5] which uses a fifth-order Runge-Kutta method, as implemented in the `scipy.integrate` library [569].

It is interesting to note that, starting from sharp initial conditions, some elements of the covariance matrix remain fixed at zero under equation (3.101):

$$C_{PI} = C_{P\Omega} = C_{eI} = C_{e\Omega} = C_{I\varepsilon} = C_{\Omega\varepsilon} = 0, \qquad (3.102)$$

i.e., stochastic perturbations to these pairs of orbital elements remain statistically uncorrelated at first order in the slow-diffusion approximation. This is due to the vanishing of the corresponding elements of the diffusion matrix $D_{ij}^{(2)}$ in equation (3.54), as well as the fact that $V_I = V_\Omega = 0$. However, all other elements of the covariance matrix generically grow over time.

In figure 3.13, we show an example of an integration of the evolution equations (3.101) for the Hulse-Taylor binary pulsar B1913+16 [324]. We see that, in this example, the period-eccentricity sector of the covariance matrix grows linearly with time,

$$C_{PP} \sim C_{Pe} \sim C_{ee} \sim t, \qquad (3.103)$$

indicating that the right-hand side of equation (3.101) is dominated by the $D^{(2)}$ term for these components. In contrast, many of the components involving the argument of pericentre and compensated mean anomaly (specifically $C_{P\omega}$, $C_{P\varepsilon}$, $C_{e\omega}$, $C_{e\varepsilon}$, $C_{\omega\omega}$, $C_{\omega\varepsilon}$, and $C_{\varepsilon\varepsilon}$) have relatively smaller values for the diffusion matrix, and are instead driven by the $C\partial V$ terms, leading them to grow like $\sim t^2$ (since the components of the covariance matrix sourcing them grow like $C \sim t$). The components C_{II}, $C_{I\Omega}$, $C_{I\omega}$, $C_{\Omega\Omega}$, and $C_{\Omega\omega}$ evolve more erratically; this is due to the presence of $\sin 2\omega$ and $\cos 2\omega$ terms in the corresponding components of the diffusion matrix (see equation (3.54)), which oscillate on a timescale $\pi/\langle\dot{\omega}\rangle_{\text{sec}} \approx 43$ yr due to the perihelion precession of the Hulse-Taylor system [585, 586]. Finally, the remaining six components of the covariance matrix are zero throughout the integration time, as expected from equation (3.102).
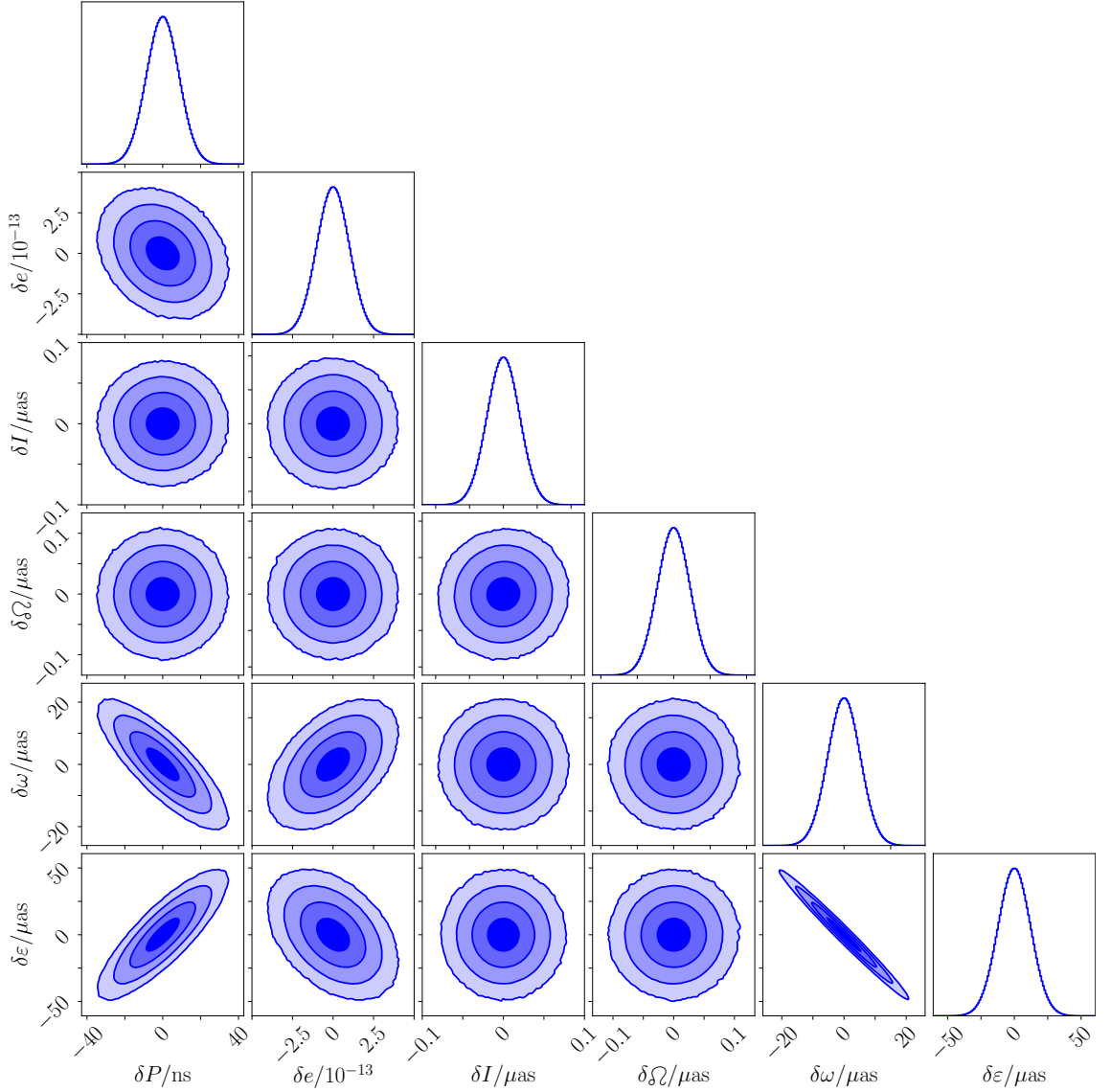
---

[5] https://github.com/alex-c-jenkins/gw-resonance

Figure 3.14: Corner plot showing the distribution of the orbital elements of B1913+16 at the end of the 1 kyr numerical integration shown in figure 3.13.

In figure 3.14 we show the distribution of the orbital elements at the end of the integration in figure 3.13. We see that the orbital elements are, on the whole, weakly correlated with each other, with the notable exceptions of the pairs (P, ω), (P, e), and particularly (ω, ε), which are highly covariant. Note that these distributions include the overall offset due to stochastic drift, but that this is less important than diffusion in this case, and is therefore harder to distinguish by eye.

3.5.4 GROWTH RATE OF NON-GAUSSIANITY

By using equation (3.101) we have neglected all higher-order moments of the distribution, and thus fail to capture any non-Gaussianities in the DF. We can measure the departure from Gaussianity by tracking the evolution of the third central moment,

$$S_{ijk} \equiv \left\langle (X_i - \bar{X}_i)(X_j - \bar{X}_j)(X_k - \bar{X}_k) \right\rangle, \tag{3.104}$$

191

as this vanishes identically for a Gaussian distribution. In particular, for each individual orbital element X_i we have, to leading order in the slow-diffusion approximation,

$$\dot{S}_{iii} \simeq -3\bar{X}_i C_{ij}\partial_j V_i \,, \tag{3.105}$$

(with no summation over i). Unlike the equations for the first two moments, the right-hand side of equation (3.105) is initially zero—non-Gaussianity is only sourced once the covariance has had a chance to grow. We also see that non-Gaussianity can only be sourced at leading order for orbital elements which have a nonzero deterministic drift V_i; the inclination I and longitude of ascending node Ω remain Gaussian at leading order. This justifies our assumption that the distribution is Gaussian on observational timescales. However, studies of, e.g., the properties of old binary populations will require us to drop this assumption and solve for the full DF. The results of section 3.4 are an important first step in studying cases like these.

3.6 Statistical formalism for gravitational-wave searches

In this section we describe how to estimate the sensitivity of a given binary to a generic GWB spectrum using two different high-precision probes of binary dynamics: timing of millisecond pulsars, and laser ranging experiments. This will allow us to compute forecasts for the upper limits that each observational probe will be able to place on various GWB spectra. To do so, we develop a novel likelihood-ratio statistic which accounts for the fact that not only the *measured* orbital elements, but also the *true* orbital elements, are random variables.

3.6.1 Distribution of the observed orbital elements

Consider a set of observations of a binary system, which are broken up into intervals much longer than the orbital period, but much shorter than the orbital diffusion timescale. These observations give us a discrete series of measurements of the orbital elements $X(t)$, one for each interval, with the intervals being labelled by t. (It is necessary to split up the data into intervals, since each high-precision measurement of the orbital elements requires a large number of individual data points.) We assume these measurements are made with Gaussian noise and zero bias. The measured values $\hat{X}(t)$ are thus drawn from a multivariate Gaussian distribution centred on the true values $X(t)$, with log-likelihood $\mathcal{L} = \ln p$ given by

$$-2\mathcal{L}(\hat{X}|X) = \sum_t \left[\ln \det 2\pi\mathsf{M} + (\hat{X} - X)^\mathsf{T} \mathsf{M}^{-1}(\hat{X} - X) \right], \tag{3.106}$$

where the measurement noise is described by the covariance matrix M. The form of the covariance matrix depends on how sensitive the timing residuals are to each orbital element—we discuss this further below.

The true orbital elements $\boldsymbol{X}(t)$ are themselves random due to the uncertainty caused by GWB resonance. As discussed in section 3.5, we can usually approximate the distribution of the elements as Gaussian (at least on observational timescales), so that the log-likelihood reads

$$-2\mathcal{L}(\boldsymbol{X}|\Omega) = \sum_t \left[\ln\det 2\pi\mathsf{C} + (\boldsymbol{X} - \bar{\boldsymbol{X}})^{\mathsf{T}}\mathsf{C}^{-1}(\boldsymbol{X} - \bar{\boldsymbol{X}})\right], \tag{3.107}$$

where the mean values $\bar{\boldsymbol{X}}(t)$ and covariance matrix $\mathsf{C}(t)$ both depend on the GWB spectrum $\Omega(f)$, and are computed as described in section 3.5.

We can marginalise over the unknown 'true' elements $\boldsymbol{X}(t)$ to obtain a likelihood function for the measured elements $\hat{\boldsymbol{X}}(t)$ for a given GWB spectrum,

$$-2\mathcal{L}(\hat{\boldsymbol{X}}|\Omega) \equiv -2\ln\left[\int \mathrm{d}\boldsymbol{X}\, p(\hat{\boldsymbol{X}}|\boldsymbol{X})p(\boldsymbol{X}|\Omega)\right] = \sum_t \left[\ln\det 2\pi\mathsf{N} + (\hat{\boldsymbol{X}} - \bar{\boldsymbol{X}})^{\mathsf{T}}\mathsf{N}^{-1}(\hat{\boldsymbol{X}} - \bar{\boldsymbol{X}})\right],$$
$$\tag{3.108}$$

where $\mathsf{N} \equiv \mathsf{M} + \mathsf{C}$ is the combined covariance matrix, incorporating the measurement uncertainty as well as the stochasticity of the orbital elements.

3.6.2 Likelihood-ratio test

Given a set of measured orbital elements $\hat{\boldsymbol{X}}$, we can phrase the GWB detection problem as a hypothesis test, where

- $\mathcal{H}_0$ is the *null hypothesis*, that there is no GWB present, $\Omega = 0$;

- $\mathcal{H}_\Omega$ is the *alternative hypothesis*, that there is a GWB present, $\Omega \neq 0$.

The simplest version of this problem is when we are searching for a GWB with a fixed spectral shape (e.g., a power law), in which case the only unknown is a single parameter setting the overall amplitude, which we denote Ω. (Concretely, this parameter Ω should then refer to the amplitude of the GWB at some fixed reference frequency, c.f. equation (0.74).)

A very natural way of carrying out such a hypothesis test is by using the log-likelihood-ratio statistic

$$\Lambda(\hat{\boldsymbol{X}}) \equiv 2\max_{\Omega>0}\left[\mathcal{L}(\hat{\boldsymbol{X}}|\Omega) - \mathcal{L}(\hat{\boldsymbol{X}}|0)\right], \tag{3.109}$$

i.e., we compare the likelihood (3.108) in the case of the null hypothesis $\Omega = 0$ with the maximum value of the likelihood as a function of Ω in the case of the alternative hypothesis. Since the maximum value of the likelihood over the entire range of values of Ω is always greater than or equal to its value at $\Omega = 0$, we see that $\Lambda \geq 0$.

Using equation (3.108), along with the fact that $\mathsf{N} = \mathsf{M}$ when $\Omega = 0$, we find

$$\Lambda(\hat{\boldsymbol{X}}) = \max_{\Omega>0}\sum_t \left[(\hat{\boldsymbol{X}} - \bar{\boldsymbol{X}}_0)^{\mathsf{T}}\mathsf{M}^{-1}(\hat{\boldsymbol{X}} - \bar{\boldsymbol{X}}_0) - (\hat{\boldsymbol{X}} - \bar{\boldsymbol{X}})^{\mathsf{T}}\mathsf{N}^{-1}(\hat{\boldsymbol{X}} - \bar{\boldsymbol{X}}) - \ln\det\left(\mathsf{I} + \mathsf{M}^{-1}\mathsf{C}\right)\right],$$
$$\tag{3.110}$$

where $\bar{X}_0$ is the value of $\bar{X}$ when $\Omega = 0$, and where we have used

$$\frac{\det N}{\det M} = \det M^{-1}N = \det\left(I + M^{-1}C\right). \tag{3.111}$$

Given a set of measurements $\hat{X}(t)$, we can thus compute the likelihood ratio statistic using equation (3.110) by solving the FPE to find $\bar{X}(t)$ and $C(t)$ for a large number of possible of values of Ω in order to maximise the likelihood. If the resulting value of Λ is large enough, then we reject the null hypothesis and claim a detection of the GWB.

An obvious question is: What value of Λ is 'large enough'? To define the detection threshold, we need to know the distribution of Λ in the case where $\mathcal{H}_0$ is true. Comparing the observed value of Λ with this null distribution then allows us to directly infer the statistical significance of the results. Since Λ is a complicated function of the data $\hat{X}$, it is difficult to find an exact distribution, even though we have fully specified how the data are distributed. However, in the limit where we have a large number of measurements (i.e., the data cover a large number of time intervals t), Wilks' theorem tells us that Λ follows a chi-squared distribution [184], $\lim_{n_t \to \infty} \Lambda \sim \chi_1^2$, where n_t is the number of time segments. (In this case the chi-squared distribution has one degree of freedom, as $\mathcal{H}_\Omega$ has one more free parameter than $\mathcal{H}_0$; we could imagine using a more complicated model for the GWB with k parameters, in which case the appropriate distribution would be χ_k^2.) In this limit, we can therefore set the threshold for detecting the GWB at a given confidence level according to the corresponding p-value of χ_1^2; e.g., a detection with 95% confidence would require $\Lambda \geq 3.841$.

3.6.3 Sensitivity forecasts

We can also use the likelihood-ratio statistic discussed above to estimate the sensitivity of future observing campaigns to different GWB spectra. To do so, we simply compute the expected value of Λ under the GWB hypothesis $\mathcal{H}_\Omega$, and find the smallest value of Ω for which this expectation surpasses the detection threshold—this tells us the weakest GWB that we can expect to detect with a given set of observations.

Let us use $\langle \cdots \rangle_\Omega$ to denote an expectation value under $\mathcal{H}_\Omega$. By definition, we have

$$\langle \hat{X}_i \rangle_\Omega = \bar{X}_i, \qquad \left\langle \left(\hat{X}_i - \bar{X}_i\right)\left(\hat{X}_j - \bar{X}_j\right) \right\rangle_\Omega = N_{ij}, \tag{3.112}$$

where the mean vector $\bar{X}$ and covariance matrix C here are computed using the true underlying value of Ω. In principle, is different from the value $\hat{\Omega}$ that maximises the likelihood, and it is the latter which determines the values for $\bar{X}$ and C that appear in the statistic Λ. However, in the $n_t \to \infty$ limit discussed above we have $\hat{\Omega} \to \Omega$ (in statistics parlance, the maximum-likelihood estimator is efficient [184]), so the two are interchangeable. It is thus straightforward to show that, in this limit, the expected value of Λ under $\mathcal{H}_\Omega$ is

$$\langle \Lambda \rangle_\Omega = \sum_t \left[\delta\bar{X}^\mathsf{T} M^{-1} \delta\bar{X} + \mathrm{tr}\, M^{-1}C - \ln\det\left(I + M^{-1}C\right) \right]. \tag{3.113}$$

We see that the stochastic drift in the orbital elements appears quadratically here, which means that it typically contributes less to the detectability of the GWB than diffusion, which enters linearly through the covariance matrix.

Once we have specified the covariance matrix M for our observations, we can compute equation (3.113) and find the smallest value of Ω that we can expect to detect.

3.6.4 Application to pulsar timing

While the formalism we have developed above is applicable to a very broad class of astrophysical binary systems, one of the main applications is in the case where one member of the binary is a millisecond pulsar (MSP) [397]. Analysis of the timing data from this MSP then allows us to determine its orbit with incredible precision, with uncertainties as small as a few parts per billion in some cases. These precision measurements allow us to search for the very small stochastic perturbations to the orbit described above. (The same principle has been used to set novel constraints on ultralight dark matter, due to its resonant effects on binary pulsar orbits [80, 81, 139, 140, 236, 395].) Here we follow the approach of Blandford and Teukolsky [136] (see also Epstein [256] and Damour and Deruelle [226]) to estimate the covariance matrix M for the orbital elements of a binary pulsar.

For each of the n_t observation intervals used to construct the likelihood ratio statistic above, one must observe some number $n_{\rm obs}$ of pulse arrival times, compare these arrival times with a timing formula for the binary, and thereby infer the binary's orbital elements at that time from the timing residuals. In the limit $n_{\rm obs} \to \infty$, the resulting covariance matrix M describing the uncertainty in the orbital elements is given by the inverse of the *Fisher matrix*,

$$ \mathsf{M} = \mathsf{F}^{-1}, \qquad \mathsf{F}_{ij} \equiv -\big\langle \partial_i \partial_j \mathcal{L}(\boldsymbol{t}|\boldsymbol{X}) \big\rangle, \tag{3.114} $$

where $\mathcal{L}$ is the log-likelihood describing the distribution of arrival times $\boldsymbol{t} = (t_1, t_2, \ldots, t_{n_{\rm obs}})$ as a function of the orbital elements $\boldsymbol{X}$, and the angle brackets here denote an expectation value under that distribution. We make the standard assumptions that the times-of-arrival (ToAs) of the pulses form a set of uncorrelated Gaussian random variables with constant variance σ^2 (i.e., the timing noise is independent of the binary's orbit), and with mean values given by the timing formula,

$$ \langle t_a \rangle \equiv \mathcal{T}_a(\boldsymbol{X}), \qquad \mathrm{Cov}[t_a, t_b] \equiv \delta_{ab}\sigma^2, \tag{3.115} $$

so that the log-likelihood is given by

$$ -2\mathcal{L}(\boldsymbol{t}|\boldsymbol{X}) = \sum_{a=1}^{n_{\rm obs}} \ln 2\pi\sigma^2 + \frac{1}{\sigma^2}(t_a - \mathcal{T}_a)^2. \tag{3.116} $$

The Fisher matrix for a likelihood of this form (as derived in, e.g., Tegmark *et al.* [538]) is

$$ \mathsf{F}_{ij} \equiv \frac{1}{\sigma^2} \sum_{a=1}^{n_{\rm obs}} \frac{\partial \mathcal{T}_a}{\partial X_i} \frac{\partial \mathcal{T}_a}{\partial X_j}. \tag{3.117} $$

To evaluate the Fisher matrix, we therefore need the derivatives of the timing formula $\mathcal{T}$ with respect to each of the orbital elements. Using the standard Blandford-Teukolsky timing formula [136], we write these as

$$
\begin{aligned}
\frac{\partial \mathcal{T}}{\partial P} &= \frac{v_P \sin I}{1 + m_1/m_2} \frac{E - e \sin E}{2\pi(1 - e \cos E)} (\sin \omega \sin E - \gamma \cos \omega \cos E), \\
\frac{\partial \mathcal{T}}{\partial e} &= -\frac{P}{2\pi} \frac{v_P \sin I}{1 + m_1/m_2} \left[\sin \omega (1 + \sin^2 E) + \frac{\cos \omega \sin E}{\gamma} (e - \gamma^2 \cos E) \right], \\
\frac{\partial \mathcal{T}}{\partial I} &= \frac{P}{2\pi} \frac{v_P \cos I}{1 + m_1/m_2} [\sin \omega (\cos E - e) + \gamma \cos \omega \sin E], \\
\frac{\partial \mathcal{T}}{\partial \Omega} &= 0, \\
\frac{\partial \mathcal{T}}{\partial \omega} &= \frac{P}{2\pi} \frac{v_P \sin I}{1 + m_1/m_2} [\cos \omega (\cos E - e) - \gamma \sin \omega \sin E], \\
\frac{\partial \mathcal{T}}{\partial \varepsilon} &= -\frac{P}{2\pi} \frac{v_P \sin I}{1 + m_1/m_2} \frac{\sin \omega \sin E - \gamma \cos \omega \cos E}{1 - e \cos E},
\end{aligned}
\tag{3.118}
$$

where the *eccentric anomaly* $E(t)$ is defined by

$$
r = a(1 - e \cos E), \qquad \cos \psi = \frac{\cos E - e}{1 - e \cos E}, \tag{3.119}
$$

and acts as an alternative to the true anomaly ψ as a way of parameterising the orbital ellipse. We see that the timing formula does not depend on the longitude of ascending node, $\partial \mathcal{T}/\partial \Omega = 0$, which means that Ω cannot be determined with pulsar timing (physically this is because Ω corresponds to a rotation around the line-of-sight axis, and thus does not affect the observable motion parallel to the line of sight).

In the limit of many ToAs, $n_{\text{obs}} \to \infty$, and assuming that these are distributed uniformly in time, we can replace the sum in equation (3.117) with an integral to obtain

$$
\mathsf{F}_{ij} \simeq \frac{n_{\text{obs}}}{T_{\text{obs}} \sigma^2} \int_0^{T_{\text{obs}}} dt \, \frac{\partial \mathcal{T}}{\partial X_i} \frac{\partial \mathcal{T}}{\partial X_j} \simeq \frac{n_{\text{obs}}}{T_{\text{obs}} \sigma^2} \frac{P}{2\pi} \int_0^{2\pi T_{\text{obs}}/P} dE \, (1 - e \cos E) \frac{\partial \mathcal{T}}{\partial X_i} \frac{\partial \mathcal{T}}{\partial X_j}, \tag{3.120}
$$

where T_{obs} is the time interval over which the n_{obs} pulse measurements are made, and where we have used Kepler's equation,

$$
\frac{dE}{dt} = \frac{2\pi/P}{1 - e \cos E}. \tag{3.121}
$$

For simplicity, we assume that T_{obs} is an integer multiple of the binary period P, and begins when the binary is at pericentre; if this is not the case, then the following formulae contain additional phase factors which do not affect the order of magnitude of the results.

We find that it is convenient to define

$$
\mathsf{F}_{ij} = \frac{n_{\text{obs}} P^2}{16 \pi^2 \sigma^2} \left(\frac{v_P \sin I}{1 + m_1/m_2} \right)^2 \bar{\mathsf{F}}_{ij}, \tag{3.122}
$$

pulling out some factors which appear in all of the Fisher matrix elements. Inserting the derivatives (3.118) into equation (3.120), we find that this factorised form of the Fisher matrix is given by

$$\bar{\mathsf{F}}_{PP} = \frac{8e}{P^2}(1 + \tfrac{1}{4}e^2) + \frac{\cos 2\omega}{P^2}f_1(e) + \frac{2\pi T_{\text{obs}}}{P^3}\sin 2\omega f_2(e) + \frac{8\pi^2 T_{\text{obs}}^2}{3P^4}\left[1 - \tfrac{1}{4}\cos 2\omega f_3(e)\right],$$

$$\bar{\mathsf{F}}_{Pe} = \frac{4}{P}(1 + \tfrac{3}{8}e)\left(1 - \tfrac{1}{6}e^2\right) + \frac{\pi T_{\text{obs}} e}{P^2 \gamma}\sin 2\omega - \frac{8\cos 2\omega}{3P}\left(1 + \tfrac{3}{4}e + \tfrac{1}{6}e^2 + \tfrac{1}{16}e^3\right),$$

$$\bar{\mathsf{F}}_{PI} = \frac{\cot I}{P}\left[2e(1 + \tfrac{1}{4}e) + \cos 2\omega\left(1 - 2e - \tfrac{3}{2}e^2\right)\right],$$

$$\bar{\mathsf{F}}_{P\omega} = -\frac{\sin 2\omega}{P}(1 - 2e - \tfrac{3}{2}e^2) - \frac{2\pi T_{\text{obs}}\gamma}{P^2},$$

$$\bar{\mathsf{F}}_{P\varepsilon} = -\frac{\sin 2\omega}{P}f_2(e) - \frac{2\pi T_{\text{obs}}}{P^2}\left[1 - \tfrac{1}{4}\cos 2\omega f_3(e)\right],$$

$$\bar{\mathsf{F}}_{ee} = \frac{5}{\gamma^2}\left[1 - \tfrac{3}{4}e^2 - \tfrac{1}{20}e^4 - \tfrac{9}{10}\cos 2\omega\left(1 - \tfrac{23}{18}e^2 + \tfrac{1}{18}e^4\right)\right],$$

$$\bar{\mathsf{F}}_{eI} = 3e\cot I\left[1 + \tfrac{1}{12}e^2 - \tfrac{11}{6}\cos 2\omega\left(1 - \tfrac{1}{22}e^2\right)\right],$$

$$\bar{\mathsf{F}}_{e\omega} = \tfrac{11}{2}e\sin 2\omega\left(1 - \tfrac{1}{22}e^2\right), \qquad \bar{\mathsf{F}}_{e\varepsilon} = \frac{e}{\gamma}\sin 2\omega,$$

$$\bar{\mathsf{F}}_{II} = 2\cot^2 I\left(1 + \tfrac{3}{2}e^2 - \tfrac{5}{2}e^2\cos 2\omega\right), \qquad \bar{\mathsf{F}}_{I\omega} = 5e^2\cot I\sin 2\omega, \qquad \bar{\mathsf{F}}_{I\varepsilon} = 0,$$

$$\bar{\mathsf{F}}_{\omega\omega} = 2 + 3e^2 + 5e^2\cos 2\omega, \qquad \bar{\mathsf{F}}_{\omega\varepsilon} = 2\gamma, \qquad \bar{\mathsf{F}}_{\varepsilon\varepsilon} = 4\frac{\sin^2\omega + \gamma\cos^2\omega}{1 + \gamma},$$

(3.123)

where the $f_i(e)$ are functions of eccentricity, which are given to $O(e^{14})$ by

$$f_1(e) \simeq 1 + \tfrac{8}{9}e - \tfrac{3}{16}e^2 - \tfrac{448}{225}e^3 - \tfrac{175}{288}e^4 - \tfrac{11584}{11025}e^5 - \tfrac{2975}{9216}e^6 - \tfrac{67264}{99225}e^7 - \tfrac{96733}{460800}e^8$$
$$- \tfrac{5818432}{12006225}e^9 - \tfrac{278579}{1843200}e^{10} - \tfrac{149726912}{405810405}e^{11} - \tfrac{20910823}{180633600}e^{12},$$

$$f_2(e) \simeq 1 + \tfrac{4}{3}e + \tfrac{5}{8}e^2 + \tfrac{2}{15}e^3 + \tfrac{1}{48}e^4 + \tfrac{1}{210}e^5 - \tfrac{29}{768}e^6 - \tfrac{31}{1260}e^7 - \tfrac{359}{7680}e^8$$
$$- \tfrac{3559}{110880}e^9 - \tfrac{469}{10240}e^{10} - \tfrac{19087}{576576}e^{11} - \tfrac{6099}{143360}e^{12},$$

(3.124)

$$f_3(e) \simeq e^2 + \tfrac{1}{2}e^4 + \tfrac{5}{16}e^6 + \tfrac{7}{32}e^8 + \tfrac{21}{128}e^{10} + \tfrac{33}{256}e^{12}.$$

We have checked numerically that these expressions are accurate to within $\approx 2\%$ even at large eccentricity, $e = 0.95$; for smaller eccentricities, the accuracy is even better.

For binaries with small eccentricity ($e \lesssim 10^{-3}$), we can change to the alternative orbital elements $(P, \zeta, \kappa, I, \Omega, \xi)$ to find

$$\bar{\mathsf{F}}_{PP} = \frac{1}{P^2}\left[-\frac{\zeta^2 - \kappa^2}{\zeta^2 + \kappa^2} + \frac{4\pi T_{\text{obs}}}{P}\frac{\zeta\kappa}{\zeta^2 + \kappa^2} + \frac{8\pi^2 T_{\text{obs}}^2}{3P^2}\right], \qquad \bar{\mathsf{F}}_{P\zeta} = \frac{4\zeta}{3P}\frac{5\zeta^2 + 6\kappa^2}{(\zeta^2 + \kappa^2)^{3/2}},$$

$$\bar{\mathsf{F}}_{P\kappa} = \frac{4\kappa}{3P}\frac{\kappa^2}{(\zeta^2 + \kappa^2)^{3/2}}, \qquad \bar{\mathsf{F}}_{PI} = -\frac{\cot I}{P}\frac{\zeta^2 - \kappa^2}{\zeta^2 + \kappa^2}, \qquad \bar{\mathsf{F}}_{P\xi} = -\frac{2}{P}\left(\frac{\zeta\kappa}{\zeta^2 + \kappa^2} + \frac{\pi T_{\text{obs}}}{P}\right),$$

$$\bar{\mathsf{F}}_{\zeta\zeta} = \frac{19}{2}, \qquad \bar{\mathsf{F}}_{\zeta\kappa} = \frac{3}{4}\zeta\kappa, \qquad \bar{\mathsf{F}}_{\zeta I} = \frac{1}{2}\zeta\cot I\left(15 + \frac{2\zeta^2}{\zeta^2 + \kappa^2}\right), \qquad \bar{\mathsf{F}}_{\zeta\xi} = 2\kappa\frac{\zeta^2 - \kappa^2}{\zeta^2 + \kappa^2},$$

(3.125)

$$\bar{\mathsf{F}}_{\kappa\kappa} = \frac{1}{2}, \qquad \bar{\mathsf{F}}_{\kappa I} = -\frac{1}{2}\kappa\cot I\left(5 - \frac{2\zeta^2}{\zeta^2 + \kappa^2}\right), \qquad \bar{\mathsf{F}}_{\kappa\xi} = \zeta\left(\frac{5}{2} - \frac{\zeta^2}{\zeta^2 + \kappa^2}\right),$$

$$\bar{\mathsf{F}}_{II} = 2\cot^2 I, \qquad \bar{\mathsf{F}}_{I\xi} = 0, \qquad \bar{\mathsf{F}}_{\xi\xi} = 2.$$

Assuming some values for the observation interval T_{obs}, number of ToAs n_{obs}, and rms timing noise σ, we can thus use the Fisher matrix elements given above to calculate the expected likelihood ratio (3.113) for a given GWB spectrum, and therefore compute upper limit forecasts.

3.6.5 APPLICATION TO LASER RANGING

Another extremely precise observational probe of binary dynamics is *laser ranging* (LR), in which laser pulses are fired at retroreflectors on bodies orbiting the Earth; typically the Moon (Lunar Laser Ranging, LLR) [431] or artificial satellites (Satellite Laser Ranging, SLR) [203]. By measuring the round-trip times of these pulses, the size of the orbit can be tracked over time with millimetre precision.

We estimate the sensitivity of a generic LR experiment in a very similar way to our treatment of pulsar timing above. Individual range measurements are assumed to be unbiased, with uncorrelated Gaussian noise of variance σ^2. We write their mean value at time t in terms of the eccentric anomaly (3.119), and define the Fisher matrix in the limit of many uniformly-spaced observations analogously to equation (3.120),

$$\mathsf{F}_{ij} \simeq \frac{n_{\mathrm{obs}}}{T_{\mathrm{obs}}\sigma^2} \frac{P}{2\pi} \int_0^{2\pi T_{\mathrm{obs}}/P} \mathrm{d}E\,(1 - e\cos E)\frac{\partial r}{\partial X_i}\frac{\partial r}{\partial X_j}. \tag{3.126}$$

The necessary partial derivatives are given by

$$\frac{\partial r}{\partial P} = \frac{2a}{3P}(1 - e\cos E) - \frac{ae\sin E}{P}\frac{E - e\sin E}{1 - e\cos E}, \qquad \frac{\partial r}{\partial e} = a\frac{e - \cos E}{1 - e\cos E}, \qquad \frac{\partial r}{\partial \varepsilon} = \frac{ae\sin E}{1 - e\cos E}, \tag{3.127}$$

where we have used the time-integrated form of equation (3.121),

$$E - e\sin E = \frac{2\pi t}{P} + \varepsilon. \tag{3.128}$$

Note that we have neglected the three angles describing the orientation of the orbital plane in space, (I, Ω, ω), as these are not directly measured by the round-trip times of the laser pulses.

Inserting equation (3.127) into equation (3.126), we obtain the Fisher matrix,

$$\begin{aligned}
\mathsf{F}_{PP} &\simeq \frac{4n_{\mathrm{obs}}a^2}{9P^2\sigma^2}\left[\frac{3\pi^2 T_{\mathrm{obs}}^2}{2P^2}(e^2 + \tfrac{1}{4}e^4) + 1 + 3e + \tfrac{27}{16}e^2 + 4e^3 + \tfrac{315}{256}e^4\right], \\
\mathsf{F}_{Pe} &\simeq \frac{3n_{\mathrm{obs}}a^2}{4P\sigma^2}\left(e + \tfrac{8}{9}e^2 + \tfrac{5}{8}e^3 + \tfrac{8}{45}e^4\right), \qquad \mathsf{F}_{P\varepsilon} \simeq -\frac{\pi n_{\mathrm{obs}}T_{\mathrm{obs}}a^2}{2P^2\sigma^2}\left(e^2 + \tfrac{1}{4}e^4\right), \\
\mathsf{F}_{ee} &= \frac{n_{\mathrm{obs}}a^2}{\sigma^2}\left(2 - \gamma - \frac{1-\gamma}{e^2}\right), \qquad \mathsf{F}_{\varepsilon\varepsilon} = \frac{n_{\mathrm{obs}}a^2}{\sigma^2}(1 - \gamma).
\end{aligned} \tag{3.129}$$

We have neglected $O(e^6)$ terms for the first three entries here, since the Moon and the artificial satellites that are typically tracked with SLR have small eccentricities (e.g., the Lunar eccentricity is $e \approx 0.055$).

3.7 BRIDGING THE µHz GAP

As we argued at the beginning of this chapter, the various gaps between the sensitive frequency bands of current and near-future GW experiments present a serious problem for GW astronomy, as failing

to explore these gaps could mean missing out on discovering any number of exciting signals hiding at these frequencies. A particularly large and well-known gap in the GW landscape occurs at roughly 10^{-7}–10^{-4} Hz, between the sensitive bands of pulsar timing arrays and future space-based interferometers such as LISA (see figure 0.14). Accessing these frequencies is challenging, requiring 'detectors' of astronomical scale that are nonetheless sensitive to the subtle effects of GWs. For example, one proposal is to construct an extremely long-baseline space-based interferometer, roughly the size of the Earth's orbit around the Sun (the µAres concept [506]); however, such ideas are currently very futuristic. Instead, we can take advantage of the fact that these µHz frequencies correspond to the orbital frequencies of binaries with periods ranging from days to years, allowing us to fill this gap with binary resonance searches. For shorter periods on the order of hours, we can also begin to explore the LISA band in the decade before LISA flies.

In this section, we use the formalism developed in the preceding sections to explore the GWB constraints that are possible with current and near-future observational data. Our main results are based on three different high-precision probes of binary orbital dynamics:

MSP: Timing of binary millisecond pulsars (MSPs), with periods between $P \approx 1.5$ hr and $P \approx 5.3$ yr [410];

LLR: Laser-ranging measurements of the Moon's orbit around the Earth ($P \approx 27$ days) [431];

SLR: Laser-ranging measurements of the orbits of artificial satellites around the Earth, in particular the LAGEOS-1 satellite ($P \approx 3.8$ hr) [203], as this has been regularly producing laser-ranging data for longer than any other satellite mission.

To get a sense of how strong we can expect our forecast constraints to be, it is instructive to carry out a back-of-the-envelope calculation in which the rms perturbation to the orbital period after time T is $\sigma_P = \sqrt{2T D_{PP}^{(2)}}$. Taking the LLR case as an example, for a SGWB intensity $\Omega_{\mathrm{gw}} = 10^{-5}$ and an observation period of $T = 15$ yr, this gives $\sigma_P \sim 1\,\mu$s. This corresponds to a rms perturbation to the semi-major axis of $\sigma_a = (2a/3P)\sigma_P \sim 0.1$ mm. Given that each LLR 'normal point' measurement determines the Earth-Moon distance to within ~ 3 mm, we see that a campaign of ~ 1000 such measurements should be capable of detecting this signal (assuming that the total measurement uncertainty scales like the inverse square root of the number of independent measurements).

To obtain our forecasts, we numerically evolve the evolution equations (3.101) for the first and second moments of the orbital elements of each of these systems using our Python code `gwresonance`. We start all integrations from sharp initial conditions, and include the first 400 harmonics when dynamically computing the KM coefficients (although the evolution is almost always dominated by the first three harmonics; c.f. figure 3.16). By integrating over the duration of a given observing campaign, and specifying the precision with which a given experiment can measure the orbital elements (via the Fisher matrix; either equation (3.123) for LR or equation (3.129) for MSPs), we can compute the expectation value of the likelihood-ratio statistic (3.113), $\langle \Lambda \rangle_\Omega$. We thus estimate the detection threshold for a given experiment and for a given GWB power-law index α by finding the smallest GWB amplitude such that $\langle \Lambda \rangle_\Omega \geq 3.841$, using a numerical root-finding procedure. We then iterate this procedure over different power-law indices, $\alpha = -10, -9.75, -9.5, \ldots, +10$, to construct the PI curves for these searches [545],

as defined in section 0.2.5. These curves represent the sensitivity of each binary to the GWB, under the assumption that the GWB spectrum is reasonably well-modelled as a power law with $|\alpha| \leq 10$ in that binary's sensitive frequency band.

For each of our binary resonance probes (MSP, LLR, and SLR), we calculate two PI curves: one which reflects the data currently available in 2021, and one which should be achievable by 2038, by which time LISA is expected to have completed its nominal 4-year mission [73]. Before we present the results of these calculations, we first discuss the details of how we model the sensitivity of each set of observations, some of the details of which are given in table 3.1.

3.7.1 Binary pulsars

We extract the orbital elements of 322 binary MSPs from the ATNF pulsar catalogue[6] [410], discarding 106 due to incomplete information, as well as the extremely wide binary J2032+4127, whose 46 yr period [399] means that the system has completed less than one complete orbit since its discovery in 2009 [42]. For the remaining 215 MSPs, we extract the period P, eccentricity e, and argument of pericentre ω; for near-circular systems $e \leq 10^{-3}$ the latter two are replaced by the Laplace-Lagrange parameters ζ, κ, as these are more numerically stable when e is very small. The strongest GW constraints typically come from binaries with longer periods, although the sensitivity also depends on the eccentricity and argument of pericentre in a more complicated way.

The inclinations of binary MSPs are generally poorly-determined due to a degeneracy with the (often unknown) masses of the pulsar and its companion. For most of the 215 systems, we assume a pulsar mass of $m_p = 1.35\, m_\odot$ and an inclination of $I = \pi/3$, as this corresponds to the median value of the companion mass m_c, which we extract from the catalogue. These assumptions are necessary to calculate our predictions for the orbital evolution, since the two masses set the deterministic decay of the period and eccentricity through GW radiation; in practice one would marginalise over the uncertainties in the masses, and propagate this uncertainty in to the model for the orbital evolution. In order to refine our results, we replace these assumed values with more accurate mass and inclination determinations from the literature for the following MSPs, which produce the best GWB bounds from our sample: J0737−3039A [381] (the double pulsar), B1913+16 [585] (the Hulse-Taylor system), B2127+11C [335], B1534+12 [273], J1829+2456 [298], J1439−5501 [397], B2303+46 [369], J0045−7319 [362], J1903+0327 [276], J1740−3052 [403], and B1259−63 [426].

Using these orbital elements and masses, we integrate the moment evolution equations (3.101) from sharp initial conditions, with the initial time set to the year in which each system was discovered. With these details specified, the GWB sensitivity is then set by the number of ToAs per observing interval, n_{obs}, and the rms timing noise σ associated with each ToA. We assume each ToA corresponds to a 10-minute pulse integration time. For our 2021 sensitivity curves, we assume each system is monitored for two weeks every year, with ToAs being gathered for two hours every day within this period; this corresponds to the data cadence for B1913+16 [323], and gives 168 ToAs per year. We further assume $\sigma = 1\,\mu s$. For our 2038 sensitivity curves, we assume an observing campaign of 365 ToAs per year (i.e., 10 minutes of observations

---

[6]https://www.atnf.csiro.au/research/pulsar/psrcat/

	P/day	e	I/rad	ω/rad	$m_1/m_\odot$	$m_2/m_\odot$	T_{obs}/yr
Binary pulsars							
J0737−3039A [381]	0.1023	0.08778	1.548	3.574	1.338	1.249	18
B1913+16 [585]	0.3230	0.6171	0.8223	5.106	1.440	1.389	46
B2127+11C [335]	0.3353	0.6814	1.047	6.027	1.358	1.354	31
B1534+12 [273]	0.4207	0.2737	1.347	4.945	1.333	1.346	30
J1829+2456 [298]	1.176	0.1391	1.32	4.013	1.306	1.569	17
J1439−5501 [397]	2.118	0.00004985	1.047	4.831	1.299	1.376	17
B2303+46 [369]	12.34	0.6584	1.047	0.6122	1.16	1.37	36
J0045−7319 [362]	51.17	0.8079	0.77	2.012	1.4	8.8	30
J1740−3052 [403]	231.0	0.5789	0.93	3.118	1.4	20.0	20
B1259−63 [426]	1237.0	0.8699	2.69	2.420	1.35	4.140	29
J1638−4725 [410]	1941.0	0.955	1.047	1.545	1.35	8.079	15
Laser ranging							
LLR (APOLLO) [431]	27.32	0.0549	0.0899	—	3.003×10^{-6}	3.692×10^{-8}	15
SLR (LAGEOS-1) [203, 327]	0.1563	0.0045	1.917	—	3.003×10^{-6}	2.047×10^{-28}	46
Solar system [353, 432]							
Mercury	87.97	0.2056	0.1108	—	1.0	1.660×10^{-7}	4.5×10^9
Venus	224.7	0.006773	0.03760	—	1.0	2.448×10^{-6}	4.5×10^9
Earth	365.3	0.01671	0.02743	—	1.0	3.003×10^{-6}	4.5×10^9
Mars	687.0	0.09341	0.02847	—	1.0	3.227×10^{-7}	4.5×10^9
Jupiter	4333.0	0.04839	0.005619	—	1.0	9.545×10^{-4}	4.5×10^9
Saturn	10760.0	0.05415	0.01615	—	1.0	2.858×10^{-4}	4.5×10^9
Uranus	30710.0	0.04717	0.01736	—	1.0	4.365×10^{-5}	4.5×10^9
Neptune	60220.0	0.008586	0.01284	—	1.0	5.150×10^{-5}	4.5×10^9
Pluto	90610.0	0.2488	0.2715	—	1.0	6.385×10^{-9}	4.5×10^9
KBO 79360 Sila-Nunam	105200.0	0.007505	0.03939	—	1.0	5.451×10^{-12}	4.5×10^9
KBO 523678	112200.0	0.007206	0.06268	—	1.0	—	4.5×10^9

Table 3.1: Table of masses and orbital elements of the binary systems used to generate our results in figures 3.15–3.19. We show a maximum of four significant digits, though in many cases these quantities have been measured with much greater precision. For the binary pulsars, m_1 and m_2 refer to the masses of the pulsar and its companion respectively, while for the other binaries they are chosen so that $m_1 \geq m_2$.

per pulsar per day) with $\sigma = 80$ ns, which is the forecast 10-minute ToA uncertainty of next-generation radio telescopes like the SKA [393].

It is important to note that our 2038 bounds are based only on known pulsars. However, the SKA and other future radio telescopes are expected to discover large numbers of new pulsars [339], some of which may be in binaries with orbits that are particularly sensitive probes of GWB resonance. We make no assumptions about these as-yet undiscovered pulsars, meaning that our 2038 bounds are conservative in this sense.

3.7.2 Laser ranging experiments

For our LLR results we use the Lunar orbital elements, Lunar mass, and Earth mass tabulated in Murray and Dermott [432]. We base our 2021 sensitivity calculations on the APOLLO experiment, which has been observing since 2006, collecting roughly 260 'normal point' range measurements per year with a rms uncertainty of $\sigma \approx 3$ mm [431]. For our 2038 sensitivity curve, we assume an observation campaign which collects 1040 normal points per year (four times the current level) with an order-of-magnitude improvement in precision, $\sigma = 0.3$ mm (this would likely require the installation of new retroreflectors on the Lunar surface [431], as the degradation of the existing reflectors is currently the main impediment to LLR sensitivity improvements). We emphasise that including only the APOLLO experiment represents a conservative estimate of LLR sensitivity, as this excludes other experiments which have been collecting LLR data since 1969 (albeit with much less precision than the APOLLO data).

For our SLR results we focus on the LAGEOS-I satellite, with a start date of 1976, and using the satellite mass and orbital elements tabulated on the International Laser Ranging Service LAGEOS webpage.[7] We assume that 50,000 normal points are collected per year for our 2021 sensitivity curve [327], rising to 200,000 per year by 2038 (again, a factor of four increase), and assume the same normal point uncertainties as for LLR in both cases.

We note that any futuristic GW mission in the solar system focusing on the frequency band of interest here may face the challenge of modelling the gravity gradient noise from asteroids, as recently pointed out by Fedderke *et al.* [267]. However, this is several orders of magnitude too small to affect the forecasts we present here. Nonetheless, this is indicative of the kinds of systematic uncertainties that are likely to impact our proposed laser-ranging GW searches. In principle, any non-GW-induced evolution in the Moon's orbit can be modelled as part of the deterministic drift V_i, but this is likely to include nuisance terms which will need to be marginalised over, potentially reducing the sensitivity of the search. For now we neglect these potential systematics, and leave a more detailed study of their impact on laser-ranging GW searches for future work.

3.7.3 Results

Our binary resonance PI curves are shown in figure 3.15, alongside the sensitivities of the current and near-future GW experiments discussed in section 0.4. There are various other constraints at lower frequencies not shown here, including those from CMB B-mode searches [48, 433] and CMB spectral

---

[7] https://ilrs.gsfc.nasa.gov/missions/satellite_missions/current_missions/lag1_general.html
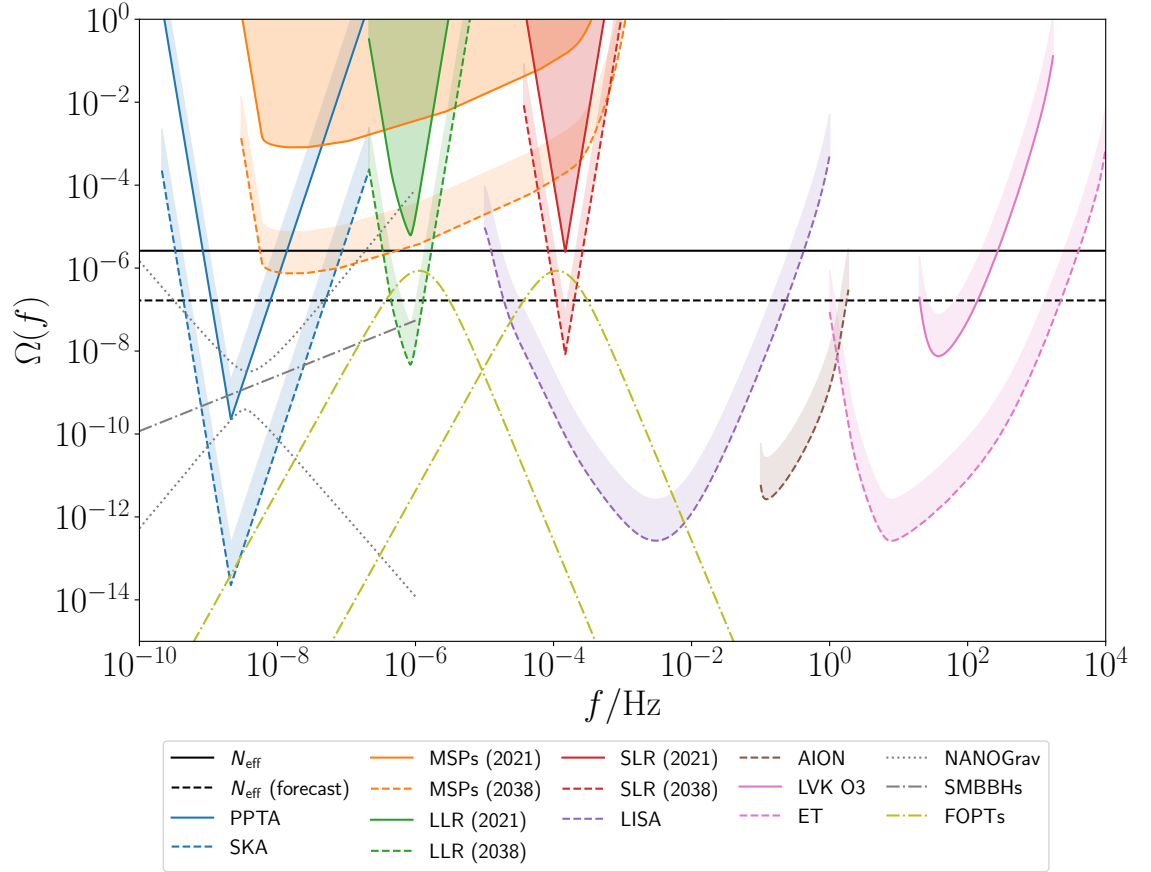
Figure 3.15: Power-law integrated sensitivity curves of current and near-future GW experiments. Solid curves indicate existing results from the LIGO/Virgo/KAGRA Collaboration (LVK) [36], pulsar timing by the Parkes PTA [388], and indirect constraints from N_{eff} [441], as well as expected present-day sensitivities of binary resonance searches with binary millisecond pulsars (MSPs), Lunar laser ranging (LLR), and satellite laser ranging (SLR), which are presented for the first time here. Dashed curves indicate our binary resonance forecast sensitivities for 2038, along with expected bounds from Einstein Telescope (ET) [464], LISA [73], the Square Kilometre Array (SKA) [339], and the proposed km-scale atom interferometer AION [91], as well as improved N_{eff} constraints [441]. The grey dotted curves indicate a range of possible signals associated with the common process (CP) detected by NANOGrav [86], while the overlaid dash-dotted curve shows the median inferred amplitude for the NANOGrav CP when assuming a $\Omega \sim f^{2/3}$ spectrum, as expected for inspiralling supermassive binary black holes (SMBBHs). The yellow dash-dotted curves show two first-order phase transition (FOPT) spectra at temperatures $T_* = 2\,\text{GeV}$ and $200\,\text{GeV}$, peaking at $f \approx 1\,\mu\text{Hz}$ and $\approx 100\,\mu\text{Hz}$ respectively.

distortions [375], as well as potential future constraints in the frequency band we are interested in, e.g. from astrometry [78, 228, 278, 430, 581], helioseismology [394], modulation of 'continuous-wave' GW signals [163], the μAres concept [506], the moon's normal modes [301, 338], and high-cadence PTA observations [454, 580]. However, all these constraints are either very futuristic, not applicable to stochastic GW signals, or not strong enough to be competitive with our forecasts.

Figure 3.16 shows a couple of examples of how the shapes of these PI curves depend on the 'comb' of independent constraints at each of the binary's resonant frequencies. For low-eccentricity cases such as the Earth-Moon system ($e \approx 0.055$), the $n = 2$ harmonic is by far the most sensitive, giving a PI curve which is sharply peaked at this frequency. On the other hand, high-eccentricity systems such as the binary

pulsar J1638-4725 ($e \approx 0.955$) can have sensitivity out to harmonics of order $n \sim 100$ or more, giving much broader PI curves.

We find that laser-ranging experiments are already able to place cosmologically relevant bounds with present data; LLR has an expected sensitivity of $\Omega(f) \geq 6.2 \times 10^{-6}$ at $f = 0.85\,\mu$Hz (95% confidence upper limit), while SLR with the LAGEOS satellite is sensitive to $\Omega(f) \geq 2.4 \times 10^{-6}$ at $f = 0.15$ mHz. These represent by far the most sensitive direct GWB searches to date in the broad frequency band between ground-based interferometers at $f \gtrsim 10$ Hz and PTAs at $f \sim$ nHz, a full three orders of magnitude stronger than existing constraints from the Cassini spacecraft [82] and the Earth's normal modes [217], and are competitive with the indirect, integrated N_{eff} constraint we discussed in section 0.4.3 which currently sets $\int \mathrm{d}(\ln f)\,\Omega(f) \leq 2.6 \times 10^{-6}$ for $f \gtrsim 10^{-15}$ [441]. By the end of the LISA mission in 2038, we expect these bounds to improve to $\Omega(f) \geq 4.8 \times 10^{-9}$ for LLR and $\Omega(f) \geq 8.3 \times 10^{-9}$ for SLR, significantly better than the N_{eff} constraint, which is expected to reach $\int \mathrm{d}(\ln f)\,\Omega(f) \leq 1.7 \times 10^{-7}$ by that time [441].

The frequencies $f = 0.85\,\mu$Hz and $f = 0.15$ mHz mentioned above correspond to the $n = 2$ harmonics of the Earth-Moon and Earth-LAGEOS systems, respectively. The corresponding sensitivity curves are strongly peaked in both cases, since the coupling to the $n = 2$ harmonic is by far the strongest for low-eccentricity orbits like that of the Moon ($e \approx 0.055$) and LAGEOS ($e \approx 0.0045$), as we saw in section 3.3.2. The next most sensitive frequency in both cases is the $n = 1$ harmonic, which is sensitive to $\Omega(f) \geq 3.2 \times 10^{-4}$ for LLR and $\Omega(f) \geq 2.2 \times 10^{-2}$ for SLR at present, improving to $\Omega(f) \geq 2.5 \times 10^{-7}$ and $\Omega(f) \geq 7.5 \times 10^{-5}$ respectively by 2038. (See figure 3.16 for the individual sensitivities of each harmonic of the Earth-Moon system.)

While binary pulsars are not able to compete with the laser-ranging experiments in terms of sheer sensitivity, they do provide useful bounds across a much wider frequency band, spanning nearly five decades in frequency from ≈ 6 nHz up to ≈ 0.2 mHz. This is partly due to the broad range of orbital periods of various systems, and partly to the large eccentricities of many of these binaries, which gives them sensitivity to much higher harmonics. The overall binary pulsar sensitivity curves shown in figure 3.15 are computed by combining the overlapping PI curves of all 215 binaries that we extract from the ATNF pulsar catalogue [410], as illustrated in figure 3.17. The most stringent sensitivity from this combined curve is $\Omega(f) \geq 8.2 \times 10^{-4}$ at frequencies $f = 14$–25 nHz with present data, expected to reach $\Omega(f) \geq 7.5 \times 10^{-7}$ by 2038. These forecasts are much stronger than the upper limit set by Hui *et al.* [323], which relied exclusively on data from the Hulse-Taylor system B1913+16; this makes sense in light of figure 3.17, which shows us that this system does not produce particularly strong constraints compared to some of the other binaries in our sample.

3.7.4 Implications for phase transitions and NANOGrav

Figure 3.15 also shows various potential GWB signals around the μHz band probed by our proposed binary resonance searches. The most important to mention here are the phase transition spectra that we first discussed in section 0.3.2, partly because FOPTs are a robust prediction of many well-motivated extensions to the Standard Model, and partly because the spectral shape of a FOPT signal highlights the constraining power of binary resonance searches. While binary resonance probes are not competitive with
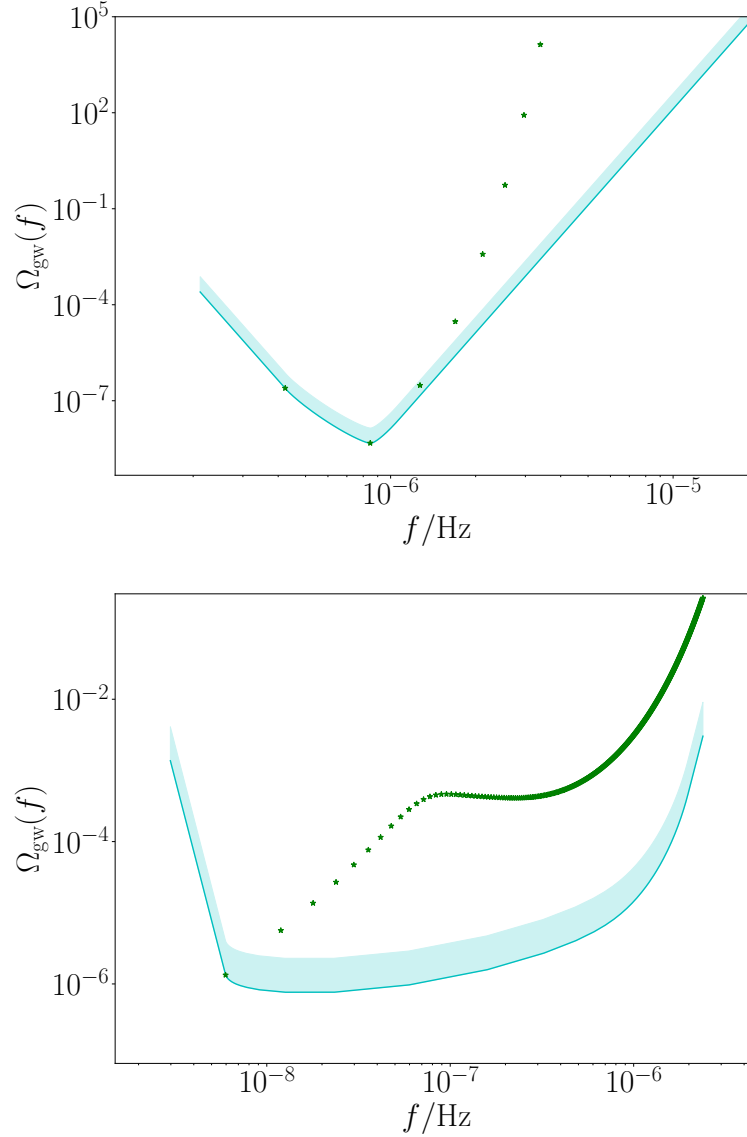
Figure 3.16: Comparison of the continuous PI curves (cyan) and discrete frequency 'comb' constraints (green points) for two binary systems at 2038 sensitivity: the Earth-Moon system in the top panel, and the binary pulsar J1638-4725 in the bottom panel. The mismatch between the discrete harmonics and the PI curves is partly due to the collective constraining power of the large number of harmonics at high frequencies (where the linearly-spaced harmonics become more finely spaced on a logarithmic scale), and partly due to the range of power-law indices considered (e.g., the APOLLO constraints degrade faster than $\Omega \sim f^{10}$ for harmonics $n > 3$).

GW interferometers and PTAs in searching for GWB spectra which are roughly flat over many decades in frequency (e.g., from cosmic strings), they can prove extremely useful for spectra that are confined to a narrow frequency band. Here we focus on the GWB signal due to sound waves in the plasma, as described by equations (0.94)–(0.96), which is thought to be the dominant contribution for most FOPTs [169]. As we mentioned previously, one also generally expects contributions from bubble wall collisions and from turbulent flows in the plasma. These can only enhance the spectra we have investigated, meaning that our results here should be viewed as conservative bounds.
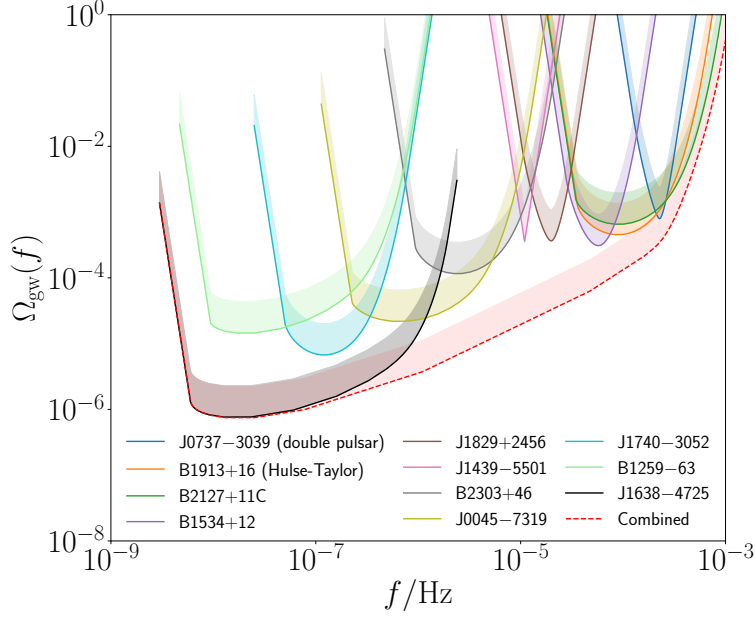
Figure 3.17: GWB PI curves from 11 binary pulsar systems at 2038 sensitivity. The red dashed curve shows the combined bound from these systems plus 204 others in the ATNF catalogue [410], and corresponds to the red dashed curve in figure 3.15. The assorted shapes of the curves shown here depend on the binary orbital parameters (particularly the eccentricity), and illustrate the utility of our formalism in accurately capturing the response of each system to the GWB.

In figure 3.18 we perform a scan over the FOPT parameters $(T_*, \alpha, \beta/H_*, v_w)$ for transitions occurring between $T_* = 10^{-3}$ GeV and 10^7 GeV, identifying regions of parameter space where the corresponding GWB signal is expected to be detected by binary resonance searches and other GW probes at 2038 sensitivity. We use equations (0.94)–(0.96), subject to the requirement that the mean bubble separation,

$$R_* = \frac{(8\pi)^{1/3}}{\beta} \max(v_w, c_s), \tag{3.130}$$

is smaller than the Hubble scale $1/H_*$ (with $c_s = 1/\sqrt{3}$ the speed of sound in the plasma). For the efficiency parameter κ which appears in equation (0.96), we use the fitting functions in the appendix of Espinosa *et al.* [258], while for the sound wave lifetime we take [255]

$$\tau_{sw} = R_* \times \left(\frac{3\kappa}{4} \frac{\alpha}{1 + \alpha} \right)^{-1/2}. \tag{3.131}$$

In order to compute the N_{eff} constraints, we use the integrated form of this spectrum,

$$\int_{-\infty}^{+\infty} \mathrm{d}(\ln f)\, \Omega(f) = \frac{343\sqrt{7/3}}{360} \Omega(f_*) \approx 1.46\Omega(f_*). \tag{3.132}$$

(Strictly speaking this is an overestimate, as it includes frequencies $f \lesssim 10^{-15}$ Hz that do not contribute to N_{eff}; however, this has negligible effect on the results in practice.)
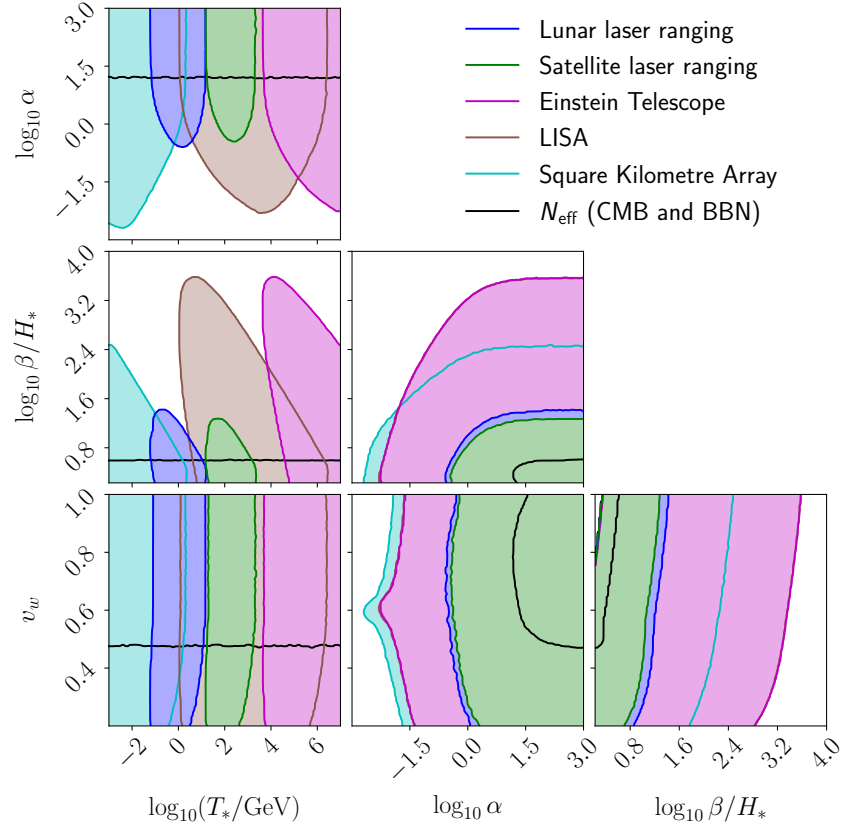
Figure 3.18: Exclusion regions of the FOPT parameter space for various GWB searches at 2038 sensitivity. Here T_* is the temperature at which the FOPT occurs, α is the energy density released by the FOPT in units of the radiation density at the transition epoch, β/H_* is the inverse duration of the transition in units of the Hubble rate at the transition epoch, and v_w is the bubble wall velocity.

We use the Markov chain Monte-Carlo sampler `emcee`[8] [274] to explore the FOPT parameter space, using the following priors (and discarding any samples for which the mean bubble separation (3.130) is larger than the horizon, $R_* H_* > 1$).

- Transition temperature T_*: log-uniform in $[10^{-3}, 10^7]$ GeV.

- Transition strength α: log-uniform in $[10^{-3}, 10^3]$.

- Inverse duration β/H_*: log-uniform in $[10^0, 10^4]$.

- Bubble wall velocity v_w: uniform in $[0.2, 1]$.

The resulting exclusion regions in figure 3.18 show FOPTs which can be detected at $\geq 95\%$ confidence. We find that LLR and SLR are able to probe significant regions of the FOPT parameter space at $T_* \sim$ GeV and ~ 100 GeV respectively. While SLR is less sensitive than LISA and will provide only complementary information, LLR will probe a region of the parameter space that is not accessible by any other planned GW experiment, thus providing a unique and valuable contribution to the search for phase transitions in the

---

[8]https://emcee.readthedocs.io/en/stable/

early Universe. FOPTs are only one example of a strongly-peaked GWB spectrum, but they demonstrate that binary resonance searches (and LLR in particular) have unique GW discovery potential.

Another potential GWB signal shown in figure 3.15 is the stochastic common process detected by NANOGrav in their 12.5-year PTA dataset [86], which we discussed in section 0.4.2. Assuming that the signal seen by NANOGrav is indeed due to GWs, and that its spectrum can be extrapolated into the µHz band, we find that present-day LLR data are able to probe some of the steeper spectra allowed by the NANOGrav data (roughly $\Omega \sim f^{1.8}$), which could correspond to a strongly blue-tilted inflationary tensor spectrum [384, 558].[9] If instead we assume that the NANOGrav signal follows the $\Omega \sim f^{2/3}$ scaling expected from inspiralling SMBBHs, we find that the spectrum should be detectable with 2038 LLR data. This provides further motivation for the binary resonance searches we propose, showing that LLR can probe the nature of GW signals detected in the nHz band by NANOGrav and other PTAs.[10]

3.7.5 SOLAR SYSTEM BOUNDS

All of the binary resonance searches discussed above rely on precision measurements of orbital elements over observational timescales of years to decades. However, our theoretical framework can also be used to study the GWB-induced evolution of binaries on much longer timescales, e.g. the evolution of planetary orbits since the formation of the Solar System ~ 4.5 Gyr ago. This amplifies the size of the effect we are interested in, as the deviations in the orbital elements typically grow like the square root of the elapsed time. However, this also entails a loss of precision, as the initial values of the orbital elements are unknown.

In figure 3.19 we show GWB constraints from the observed orbital elements of the eight Solar System planets, along with the dwarf planet Pluto and 110 classical Kuiper Belt Objects (KBOs). We find that these are all orders of magnitude weaker than the precision binary resonance constraints possible with binary pulsars and laser ranging, with the strongest limit of $\Omega \leq 6.6 \times 10^3$ at $f = 0.13$ nHz coming from 523678 (2013 XB$_{26}$), a classical KBO on a very low-eccentricity orbit [353].

To produce these constraints, we integrate the evolution equations (3.101) over the age of the solar system (~ 4.5 Gyr), and compare the present-day periods, eccentricities, and inclinations of various solar system bodies to the rms changes in each of these predicted due to binary resonance,

$$\sigma_i = \sqrt{\delta \bar{X}_i^2 + C_{ii}}, \tag{3.133}$$

(no summation over the repeated index). Since the GWB tends to drive binaries towards longer periods, higher eccentricities, and larger inclinations, we can infer an upper limit on the GWB intensity by requiring equation (3.133) to be less than the present-day values of each of these quantities. In doing so, we account for the redshifting of GWs over cosmological timescales, setting

$$\Omega = \Omega_0 \times (1+z)^4, \qquad f = f_0 \times (1+z), \tag{3.134}$$

---

[9]Such spectra can avoid the existing LVK and N_{eff} constraints if one allows for a nonstandard thermal history [384].

[10]It is also possible for the spectrum to drop off before reaching µHz frequencies, so that one cannot naively extrapolate the spectrum from NANOGrav frequencies as we have here; however, in this scenario a non-detection of the GWB with LLR would reveal this drop-off, and would thus still provide valuable observational information.
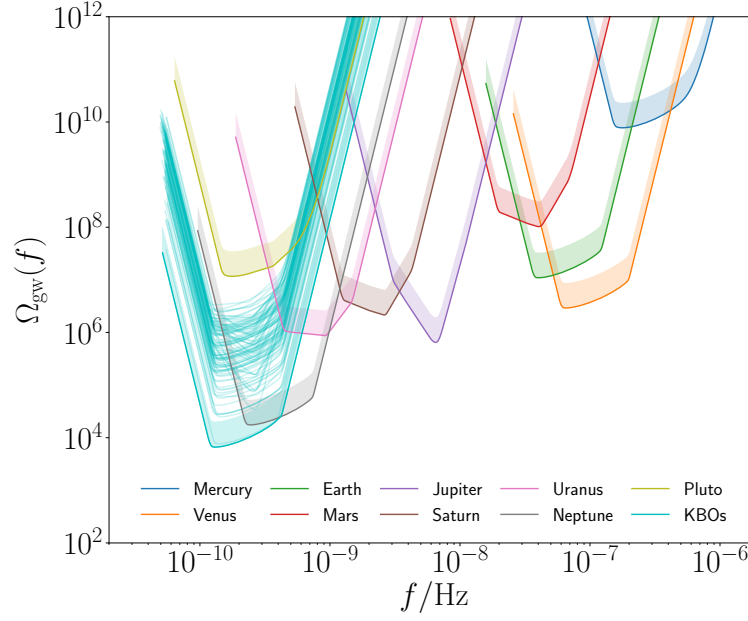
Figure 3.19: GWB PI curves inferred from the present-day orbital elements of various Solar System bodies. The faint cyan curves show constraints from 110 individual KBOs from the NASA/JPL Small-Body Database, while the solid cyan curve shows the combined KBO constraint.

with '0' subscripts denoting the present-day values that we place bounds on. Since the solar system formed at redshift $z \approx 0.41$, this can affect the final bounds by roughly a factor of $(1 + z)^4 \approx 3.9$. The redshifting of the GW frequency also broadens the shape of the resulting PI curve.

We extract the present-day orbital elements and masses of the eight planets and Pluto from Murray and Dermott [432], as well as those of 110 dynamically cold 'classical' KBOs from the NASA/JPL Small-Body Database [353].[11] The individual PI curves of the KBOs are combined to give an overall PI curve for the Kuiper Belt constraint, which is dominated by 523678 (2013 XB26) at low frequencies, and by 79360 Sila-Nunam (1997 CS$_{29}$) at high frequencies, primarily due to their low eccentricities $e \approx 0.007$.

3.8 Summary and outlook

In this chapter we have developed a powerful new formalism for calculating the statistical evolution of binary systems coupled to the GWB, deriving a secular FPE which captures the full probability distribution of all six orbital elements on timescales much longer than the binary period. The KM coefficients describing this FPE, given in equations (3.55) and (3.54), and illustrated in figures 3.4–3.10, encapsulate the rich dynamical structure that arises from the interactions between tensorial GW perturbations and elliptical orbits.

The full FPE is a six-dimensional nonlinear PDE, and it is therefore challenging to find exact solutions. Nonetheless, we have extracted some qualitative features of the late-time behaviour in section 3.4 by fixing the eccentricity to zero. This analysis illustrates one of the key advantages of our formalism over previous approaches: its ability to capture the full shape of the DF. We find that, while the stochastic drift effect

---

[11] https://ssd.jpl.nasa.gov/sbdb.cgi

due to the SGWB tends to increase the energy of the binary and counteract its orbital decay through GW emission, diffusion tends to have the opposite effect, so that the net influence of SGWB resonance is to drive the binary towards merger. At the same time, the GWB perturbations act to erase any memory of the binary's initial configuration in space, driving the orbit towards an isotropic distribution in the inclination I, and uniform distributions in the other angular variables. We also find, however, that there is a strong drive towards larger eccentricities in the $e \to 0$ limit, so these results with $e = 0$ must be taken with a pinch of salt.

We have also developed, in section 3.5, a practical approach for numerically integrating the full FPE with general eccentricity $e \in (0, 1)$ on observational timescales, taking advantage of the fact that these are typically much shorter than the diffusion timescale. The resulting set of equations (3.101) are illustrated in figures 3.13 and 3.14, using the binary pulsar B1913+16 as an example. Combined with the statistical tools developed in section 3.6, this allows us to calculate GWB sensitivity curves for probes of binary dynamics such as pulsar timing and Lunar/satellite laser ranging. We have presented these results in section 3.7, showing that our methods can improve upon existing bounds in the μHz frequency band by several orders of magnitude, thereby constraining unique regions of the FOPT parameter space, and possibly shedding light on the nature of the NANOGrav common process signal.

These results motivate further work to develop binary resonance into a precision tool for GW astronomy. The most important task will be to develop the necessary data analysis pipelines to conduct GW searches with pulsar-timing and laser-ranging data, in a way that fully utilises the theoretical developments in this work. However, there also many possible avenues for developing our theoretical formalism further. One important problem is to develop practical numerical integration schemes that go beyond the approach in section 3.5 and capture non-Gaussian features in the DF, thereby taking full advantage of the Fokker-Planck approach (this will be particularly important for studies of populations of binaries [99, 322]). It would also be interesting to relax some of our assumptions, for example abandoning the secular-averaging approach and attempting to capture the evolution of binaries within a single orbital period, or perhaps relaxing some of the usual assumptions about the GW strain statistics to develop searches for GWBs that are non-Gaussian, anisotropic, or have nonstandard polarisation content. One could also explore the sensitivity of binaries to narrowband sources, using a similar approach to Blas *et al.* [139] to consider GW frequencies between the binary's resonant frequencies. There is also no reason to restrict ourselves to just binaries; in future work, we plan to consider the GWB-driven evolution of other gravitationally-bound systems such as hierarchical triples or many-body systems such as galaxies and globular clusters. Finally, our work could be extended even further by considering other stochastic perturbing fields which may exist in the Universe, such as ultralight scalars [185, 243, 275, 357].

4   CONCLUSION

> 'I may not have gone where I intended to go,
> but I think I have ended up where I needed to be.'
>
> — Douglas Adams

The goal of this thesis has been to investigate some of the many ways in which gravitational-wave observations can give us fascinating new insights into the fundamental laws that govern the Universe; from the statistical distribution of matter on the largest scales, down to the particles and interactions that make up the fabric of the cosmos. In particular, we have focused on three exciting frontiers for GW cosmology: (i) probing late-time cosmology using anisotropies in the astrophysical GW background, (ii) exploring the intersection between nonlinear gravity and particle physics beyond the Standard Model with cosmic strings, and (iii) searching for first-order phase transitions and other cosmological GW signals with binary resonance.

We began in chapter 1 by motivating the importance of treating the AGWB as anisotropic, bringing it into line with other cosmological observables such as the CMB, and thereby giving us access to novel information that cannot be extracted from the isotropic component alone. The underlying hope here is that, since the compact binaries that make up the AGWB signal are formed from stellar evolution, they can act as tracers of the distribution of galaxies and matter on large scales, possibly revealing cosmological information complementary to that provided by existing observational probes. With this aim in mind, we have developed a model for the angular power spectrum of the AGWB, combining well-motivated astrophysical recipes for computing the CBC rate density with a large and detailed mock galaxy catalogue based on the Millennium N-body simulation. We find that the AGWB has strong angular fluctuations, ranging from a few percent on the largest angular scales up to nearly 100% at higher multipoles $\ell \sim 100$. While these anisotropies are well below the sensitivity of present GW searches, we have shown that they are likely to be detected by next-generation experiments such as Einstein Telescope and Cosmic Explorer, providing a direct measurement of the statistical clustering of GW sources. Unfortunately, the information that we will be able to extract from such a detection is likely to be limited by the presence of *shot noise*, which we have shown is a natural consequence of the AGWB consisting of a finite number of transient signals. Though we have developed an optimal data-analysis method for inferring the true angular power spectrum in the presence of shot noise, the uncertainty associated with these analyses is still much larger than the cosmic variance limit, making it difficult to extract precise cosmological information from the signal.

There are, however, several exciting roads ahead for anisotropic AGWB searches. One approach—which we have already highlighted in section 1.4—is to cross-correlate the AGWB with other cosmological

probes, which has the dual advantage of boosting the detectability of the signal and reducing the amplitude of the shot noise. While there have been several preliminary studies into such cross-correlation searches, there is still plenty of work to be done, both in modelling the expected signal and its dependence on various astrophysical and cosmological factors, and implementing efficient statistical methods for inferring the cross-correlation spectrum from GW strain data. This future work may include going beyond two-point statistics such as the angular power spectrum, and trying to capture the full non-Gaussian statistics of the AGWB, perhaps by computing the bispectrum and trispectrum. Another exciting possibility is to search for the *kinematic dipole* in the AGWB (and, indeed, in other stochastic backgrounds). As we showed in section 1.1.5, this dipole encodes information about the frequency spectra of the sources contributing to the background, so that, if measured, it could prove extremely useful in discriminating between different possible sources of a given GWB signal.

In chapter 2, we investigated the GW emission from Nambu-Goto cosmic strings, an extremely promising and well-motivated prediction of many models of particle physics beyond the Standard Model. In particular, we calculated the *nonlinear GW memory* associated with cusps and kinks on Nambu-Goto loops, i.e., the hereditary GW signal sourced by the energy-momentum of the initial cusp and kink GW emission. If observed, this nonlinear memory would provide a fascinating test of the fundamental field-theoretic properties of general relativity, as well as probing the nonlinear gravitational dynamics of the cosmic strings themselves. We have obtained simple analytical waveforms for the nonlinear memory emitted by cusps and kinks, including the 'memory of the memory' and other higher-order effects. Surprisingly, while the memory from kinks is strongly suppressed, the memory from cusps *diverges* for sufficiently large loop sizes. This divergence is unphysical, and, we have argued, indicates a breakdown in the weak-field treatment of Nambu-Goto loop dynamics that is ubiquitous in the literature, requiring some kind of strong-gravity mechanism to reduce the high-frequency GW emission from the cusp and thereby resolve the divergence.

In section 2.5, we have argued that one possible tentative resolution to the cusp divergence would be for the offending cusps to collapse and form primordial black holes. The formation of an event horizon truncates the cusp's GW spectrum before the signal reaches its peak, which, as we have explicitly demonstrated, suppresses the cusp's memory emission and cures the divergence. Using heuristic arguments based on the hoop conjecture, we show that such a situation arises very naturally for the exact same range of loop lengths that give rise to the memory divergence. If these cusp-collapse PBHs do indeed form, we show that they are likely to be born with very small masses, large (but non-extremal) spins, and relativistic velocities, making them very distinct from all other known populations of BHs (both astrophysical and primordial). Thus, while the nonlinear memory is not itself observable in the cusp-collapse scenario, the PBHs formed this way could act as a 'smoking gun' signature of cosmic strings.

The most pressing next steps associated with this work are clearly to ascertain whether or not cusp collapse is the correct picture for resolving the nonlinear memory divergence, and if not, whether the nonlinear memory could in fact leave a substantial observational imprint on GW observations or not. This will likely require numerical relativity simulations, as the relativistic velocities and lack of isometries associated with cusps makes detailed analytical treatments very challenging without resorting to the weak-field approximation. Another interesting avenue to pursue would be to calculate the *linear* GW

memory associated with matter radiation from cusps, though this would require going beyond the simple Nambu-Goto approximation we have adopted here.

In chapter 3, the final of our three forays into GW cosmology, we have developed an exciting new GW search method based on monitoring the orbits of binary systems. The key principle is that impinging GWs with frequency equal to an integer multiple of the binary's orbital frequency interact resonantly with the system, leaving a lasting imprint on the size, shape, and orientation of the orbit. Stochastic GW backgrounds are ideal targets for these searches, as they are *persistent*, thus giving rise to perturbations in the orbital elements that grow over time, and *broadband*, meaning that they are much more likely to overlap with a given binary's resonant frequencies than a narrowband signal. However, due to the phase incoherence of the GWB, we are unable to deterministically model the corresponding evolution of the orbital elements over time. We therefore develop, from first principles, a Fokker-Planck equation that tracks the time-dependent probability distribution of each of the orbital elements.

In section 3.7, we have applied these theoretical tools to forecast the sensitivity of various observational probes of binary dynamics to the GWB, focusing on the timing of binary millisecond pulsars and laser ranging of the Moon and artificial satellites. We have shown that these searches can reach impressive sensitivity levels in the 'µHz gap' between the sensitivities of LISA and pulsar timing arrays, down to the level of $\Omega \sim 10^{-8}$ by the end of LISA's nominal 4-year mission in the late 2030s. This level of sensitivity would allow us to search for a range of exciting GW signals in this thus-far-unexplored frequency range; for example, ruling out a unique region of the parameter space for first-order phase transitions that cannot be reached by other planned GW experiments, as well as shedding some light on the potential GWB signal seen by NANOGrav in their 12.5-year dataset.

There are numerous exciting research directions that these binary resonance searches could take us in. Perhaps the most important is to develop full data analysis pipelines for analysing pulsar-timing and laser-ranging data, allowing us to set upper limits on (or make detections of!) the GWB spectrum. It would also be very interesting to extend our formalism to study how other gravitationally-bound systems, such as galaxies or stellar triples, are perturbed by the GWB.

Clearly, each of the three core topics in this thesis is ripe for further theoretical and observational development as we step further into the era of GW astronomy. As existing detectors are upgraded, and next-generation instruments begin to come online, we can look forward to an ever-growing number of detected GW signals, and perhaps, eventually, the first signs of exotic new physics. While it is not yet clear exactly what this fast-approaching treasure trove of GW data will teach us about cosmology and fundamental physics, we can be confident that there *are* exciting discoveries in store, just as there have been every time humans have used a new tool for looking at the night sky.

A DIRECTIONAL GWB SENSITIVITY FORECAST FOR THE EINSTEIN TELESCOPE

In section 1.2.2 (and, in particular, figure 1.7), we showed that the lowest multipoles of the AGWB angular power spectrum are likely to be detected by the Einstein Telescope, a third-generation ground-based GW interferometer. This statement depends purely on the detector geometry and expected noise spectrum of the ET proposal, and is relatively straightforward to derive. However, since, to my knowledge, this calculation has not been presented explicitly anywhere in the literature, I include it here for completeness. We focus purely on the directional GWB searches that are possible by cross-correlating data between each of ET's three interferometers (ignoring for now the potential issue of correlated noise sources), and use the formalism of Thrane *et al.* [543] to derive the corresponding Fisher matrix for the C_ℓ's. In practice, ET will likely be joined by other third-generation observatories such as Cosmic Explorer [295, 472], and it should therefore be possible to increase the sensitivity further (both in terms of amplitude and in terms of the number of accessible multipoles) by conducting cross-correlation searches between these different detectors. However, the resulting sensitivity then depends on the relative locations and orientations of the detectors, which are not yet known, so we focus for now on searches using only ET data.

The *directional* overlap reduction function between two Michelson interferometers operating in the small-antenna limit $f \ll 1/L$ (where L is the arm length) is given by [487]

$$\gamma_{IJ}(f, \hat{\boldsymbol{r}}) = \frac{1}{2} \mathcal{D}_I^A(\hat{\boldsymbol{r}}) \mathcal{D}_J^A(\hat{\boldsymbol{r}}) \cos[2\pi f \hat{\boldsymbol{r}} \cdot \Delta \boldsymbol{x}], \tag{A.1}$$

which is just the integrand of the isotropic ORF (0.73) we encountered in section 0.2.5, $\Gamma_{IJ}(f) = \int_{S^2} \mathrm{d}^2 \hat{\boldsymbol{r}} / (4\pi) \gamma_{IJ}(f, \hat{\boldsymbol{r}})$. The detector response functions $\mathcal{D}_I^A$ encode the coupling of interferometer I to a plane wave with polarisation A arriving from the direction $\hat{\boldsymbol{r}}$. These are given in terms of the unit vectors $\hat{\boldsymbol{u}}_I$ and $\hat{\boldsymbol{v}}_I$ that point along each of the arms,

$$\mathcal{D}_I^A = \frac{1}{2} e_{ij}^A (\hat{u}_I^i \hat{u}_I^j - \hat{v}_I^i \hat{v}_I^j), \tag{A.2}$$

where e_{ij}^A are the polarisation tensors (0.19), which are given in terms of the standard spherical polar unit vectors $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}$,

$$e_{ij}^+ = \hat{\theta}_i \hat{\theta}_j - \hat{\phi}_i \hat{\phi}_j, \qquad e_{ij}^\times = 2\hat{\theta}_{(i} \hat{\phi}_{j)}. \tag{A.3}$$

(ℓ, m)	$[\gamma_{12}]_{\ell m}$	$[\gamma_{13}]_{\ell m}$	$[\gamma_{23}]_{\ell m}$
$(0, 0)$	$-\dfrac{3\sqrt{\pi}}{20}$	$-\dfrac{3\sqrt{\pi}}{20}$	$-\dfrac{3\sqrt{\pi}}{20}$
$(2, 0)$	$-\dfrac{3}{14}\sqrt{\dfrac{\pi}{5}}$	$-\dfrac{3}{14}\sqrt{\dfrac{\pi}{5}}$	$-\dfrac{3}{14}\sqrt{\dfrac{\pi}{5}}$
$(4, 0)$	$-\dfrac{\sqrt{\pi}}{280}$	$-\dfrac{\sqrt{\pi}}{280}$	$-\dfrac{\sqrt{\pi}}{280}$
$(4, \pm4)$	$-\dfrac{e^{\mp 2\pi i/3}}{4}\sqrt{\dfrac{\pi}{70}}$	$-\dfrac{1}{4}\sqrt{\dfrac{\pi}{70}}$	$-\dfrac{e^{\pm 2\pi i/3}}{4}\sqrt{\dfrac{\pi}{70}}$

Table A.1: Non-zero spherical harmonic components for each of the three ORFs.

For ET, we have three co-located (i.e. $\Delta \boldsymbol{x} = 0$) interferometers, each having two arms which make up two sides of an equilateral triangle [464]. We can therefore write

$$\mathcal{D}_1^A = \frac{1}{2}e_{ij}^A(\hat{u}_1^i\hat{u}_1^j - \hat{u}_2^i\hat{u}_2^j), \qquad \mathcal{D}_2^A = \frac{1}{2}e_{ij}^A(\hat{u}_2^i\hat{u}_2^j - \hat{u}_3^i\hat{u}_3^j), \qquad \mathcal{D}_3^A = \frac{1}{2}e_{ij}^A(\hat{u}_3^i\hat{u}_3^j - \hat{u}_1^i\hat{u}_1^j), \quad \text{(A.4)}$$

and define a Cartesian coordinate frame in which the arm vectors are given by

$$\hat{\boldsymbol{u}}_1 = (1, 0, 0), \qquad \hat{\boldsymbol{u}}_2 = -\frac{1}{2}(1, \sqrt{3}, 0), \qquad \hat{\boldsymbol{u}}_3 = -\frac{1}{2}(1, -\sqrt{3}, 0), \qquad \text{(A.5)}$$

In the same Cartesian basis, we have

$$\hat{\boldsymbol{\theta}} = (\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta), \qquad \hat{\boldsymbol{\phi}} = (-\sin\phi, \cos\phi, 0). \qquad \text{(A.6)}$$

The three overlap reduction functions are therefore

$$\begin{aligned}
\gamma_{12} &= -\frac{3}{128}(1 + \cos^2\theta)^2 \, \text{Re}\left\{1 + 2e^{4i\phi+2\pi i/3}\right\} - \frac{3}{32}\cos^2\theta \, \text{Re}\left\{1 - 2e^{4i\phi+2\pi i/3}\right\}, \\
\gamma_{13} &= -\frac{3}{128}(1 + \cos^2\theta)^2 \, \text{Re}\left\{1 + 2e^{4i\phi}\right\} - \frac{3}{32}\cos^2\theta \, \text{Re}\left\{1 - 2e^{4i\phi}\right\}, \\
\gamma_{23} &= -\frac{3}{128}(1 + \cos^2\theta)^2 \, \text{Re}\left\{1 + 2e^{4i\phi-2\pi i/3}\right\} - \frac{3}{32}\cos^2\theta \, \text{Re}\left\{1 - 2e^{4i\phi-2\pi i/3}\right\}.
\end{aligned} \qquad \text{(A.7)}$$

Note that the fact that the detectors are co-located makes the ORFs frequency-independent (up until the point where the small-antenna approximation breaks down, $f \approx 1/L \approx 1/(10\,\text{km}) \approx 30\,\text{kHz}$). We are interested in ET's response to different spherical harmonics of the GWB, so we decompose equation (A.7) into spherical harmonic components, $[\gamma_{IJ}]_{\ell m} \equiv \int_{S^2} d^2\hat{\boldsymbol{n}} \, Y_{\ell m}(\hat{\boldsymbol{n}}) \gamma_{IJ}(\hat{\boldsymbol{n}})$. The resulting nonzero components are shown in table A.1. We see that each ET pair is only sensitive to four independent spherical harmonic components of the GWB: those corresponding to (ℓ, m) values of $(0, 0)$, $(2, 0)$, $(4, 0)$, and $(4, \pm4)$.

	$\ell = 0$	$\ell = 2$	$\ell = 4$
$\mathrm{UL}_{95\%}$	2.625×10^{-25}	2.143×10^{-26}	2.917×10^{-24}

Table A.2: Resulting 95% confidence upper limits on the C_ℓ's from ET alone.

Given these ORF spherical harmonic components, we can calculate the Fisher matrix [487, 543] for a power-law AGWB $\Omega \sim f^{2/3}$,

$$\Gamma_{\ell m, \ell' m'} = \left(\frac{3H_0^2}{2\pi^2 f_{\mathrm{ref}}^3}\right)^2 \sum_{I > J} \sum_{t,f} [\gamma_{IJ}^*]_{\ell m}(f, t)[\gamma_{IJ}]_{\ell' m'}(f, t) \frac{(f/f_{\mathrm{ref}})^{-14/3}}{P_I(f, t)P_J(f, t)}, \qquad (A.8)$$

where f_{ref} is the reference frequency at which the angular power spectrum is estimated, P_I is the noise PSD in interferometer I, the first sum is over detector pairs, and the second sum is over frequency bins and time intervals. For concreteness, we take the time intervals as having length $\tau = 100$ s, so that we have a frequency resolution of $1/\tau = 0.01$ Hz; this means that the frequency-bin sum is over $f_n = n/\tau$, with $n \in \mathbb{Z}^+$. As we have seen above, the ORFs are time- and frequency-independent. We further assume that the noise PSDs are time-independent (i.e., stationary). The Fisher matrix then simplifies to

$$\Gamma_{\ell m, \ell' m'} = \left(\frac{3H_0^2}{2\pi^2 f_{\mathrm{ref}}^3}\right)^2 \frac{T}{\tau} \sum_{IJ} [\gamma_{IJ}^*]_{\ell m}[\gamma_{IJ}]_{\ell' m'} \sum_{n=1}^{\infty} \frac{(n/\tau f_{\mathrm{ref}})^{-14/3}}{P_I(n/\tau)P_J(n/\tau)}. \qquad (A.9)$$

Using the design ET-D noise PSD curve [313] with a reference frequency $f_{\mathrm{ref}} = 25$ Hz, this becomes

$$\Gamma_{\ell m, \ell' m'} = \left(2.509 \times 10^{25} \times \frac{T}{1\,\mathrm{year}}\right) \sum_{IJ} [\gamma_{IJ}^*]_{\ell m}[\gamma_{IJ}]_{\ell' m'}. \qquad (A.10)$$

The resulting uncertainty on the angular power spectrum is given in terms of the inverse Fisher matrix Γ^{-1} by [543]

$$\mathrm{Var}[C_\ell] = \frac{2}{(2\ell + 1)^2} \sum_{m,m'} \left|(\Gamma^{-1})_{\ell m, \ell m'}\right|^2, \qquad (A.11)$$

with the corresponding 95% confidence upper limits (assuming a half-Gaussian distribution, with no support below zero),

$$\mathrm{UL}_{95\%} = \mathrm{erf}^{-1}(0.95)\sqrt{2\mathrm{Var}[C_\ell]} \approx 1.960\sqrt{\mathrm{Var}[C_\ell]}. \qquad (A.12)$$

In practice, the Fisher matrix is singular due to 'blind-spots' in the detector response, and cannot be inverted. We instead use the Moore-Penrose pseudoinverse, which effectively ignores the spherical harmonics that cannot be measured, and provides the unique minimum-chi-squared solution to the deconvolution problem. The resulting upper limits are shown in table A.2 and in figure 1.7.

B Compound Poisson shot noise model

In section 1.3.2 we introduced the compound Poisson random variable $\Lambda \equiv \sum_{i=1}^{N} \lambda_i$ as a model of the CBC event count in a given spatial volume element δV, where the number of galaxies $N \sim \text{Pois}[\bar{n}\delta V]$ and the number of CBCs in each galaxy $\lambda_i \sim \text{Pois}[R\tau_s]$ are both Poisson random variables.[1] (We recall that $\bar{n}$ is the mean galaxy number density, R is the CBC event rate per galaxy, and $\tau_s = \tau/(1+z)$ is the source-frame time interval corresponding to an observation time τ.) In this appendix we derive a few of the relevant statistical properties of this model. We focus initially on a single point in the galaxy/CBC parameter space, $\boldsymbol{\zeta}$, so that the event rates per galaxy are all i.i.d.

B.1 Cumulants of the distribution

We can characterise the distribution of Λ by its cumulants,

$$\kappa_n \equiv \left. \frac{\mathrm{d}^n K_\Lambda(x)}{\mathrm{d}x^n} \right|_{x=0}, \tag{B.1}$$

where $K_\Lambda(x)$ is the cumulant-generating function (CGF),

$$K_\Lambda(x) \equiv \ln\langle e^{\Lambda x}\rangle. \tag{B.2}$$

The first two cumulants, κ_1, κ_2, are just the mean and the variance, while higher cumulants κ_n represent the 'connected components' of the nth moments of the distribution, and vanish if the distribution is Gaussian. We can always write each of the moments $\langle \Lambda^n \rangle$ as a polynomial in the cumulants; for example, the fourth moment is

$$\langle \Lambda^4 \rangle = \kappa_1^4 + 6\kappa_2\kappa_1^2 + 3\kappa_2^2 + 4\kappa_3\kappa_1 + \kappa_4. \tag{B.3}$$

We can find the full set of cumulants for Λ by evaluating the CGF (B.2) explicitly. As a warm-up exercise before we do this, we start by finding the CGF for each of the per-galaxy CBC event counts, $\lambda \sim \text{Pois}[\bar{\lambda}]$,

$$K_\lambda(x) \equiv \ln\langle e^{\lambda x}\rangle = \ln\left[\sum_{k=0}^{\infty} \Pr(\lambda = k)e^{kx}\right] = \ln\left[\sum_{k=0}^{\infty} \frac{\bar{\lambda}^k e^{-\bar{\lambda}}}{k!}e^{kx}\right] = \ln e^{\bar{\lambda}(e^x-1)} = \bar{\lambda}(e^x - 1). \tag{B.4}$$

---

[1]A more sophisticated approach would account for the statistical properties of haloes [213]. However, this simple assumption is sufficient for the calculation here.

(Similarly, $K_N(x) = \bar{N}(e^x - 1)$, as N is also Poisson-distributed.) Note that this has the property that all of the cumulants are equal, $\kappa_1 = \kappa_2 = \cdots = \bar{\lambda}$ (in a sense, this is the defining feature of the Poisson distribution). The CGF for the full compound distribution is then given by a very similar calculation,

$$
\begin{aligned}
K_\Lambda(x) &= \ln\left[\sum_{k=0}^{\infty} \Pr(N = k)\left\langle\prod_{i=1}^{k} e^{\lambda_i x}\right\rangle\right] = \ln\left[\sum_{k=0}^{\infty} \Pr(N = k)\left\langle e^{\lambda x}\right\rangle^k\right] \\
&= \ln\left[\sum_{k=0}^{\infty} \frac{\bar{N}^k e^{-\bar{N}}}{k!} e^{k K_\lambda(x)}\right] = \ln e^{\bar{N}(e^{K_\lambda(x)} - 1)} = \bar{N}\left(e^{K_\lambda(x)} - 1\right) = K_N(K_\lambda(x)),
\end{aligned}
\tag{B.5}
$$

where in the second equality we have used the fact that the λ's are i.i.d. It is straightforward to take derivatives of equation (B.5) to obtain the cumulants; the first four are

$$
\kappa_1 = \bar{N}\bar{\lambda} \equiv \bar{\Lambda}, \qquad \kappa_2 = \bar{\Lambda}(1+\bar{\lambda}) \equiv \mathrm{Var}[\Lambda], \qquad \kappa_3 = \bar{\Lambda}(1+3\bar{\lambda}+\bar{\lambda}^2), \qquad \kappa_4 = \bar{\Lambda}(1+7\bar{\lambda}+6\bar{\lambda}^2+\bar{\lambda}^3),
\tag{B.6}
$$

where $\bar{N} = \bar{n}\delta V$ and $\bar{\lambda} = R\tau_s$. Note that the expression for the variance here corresponds to equation (1.56), with the two terms representing temporal and spatial shot noise, respectively. We can simplify things significantly by noting that the CBC rate per galaxy is typically on the order of $R \sim \mathrm{Myr}^{-1}$, while the source-frame observing time τ_s is at most a few years, so we have $\bar{\lambda} \sim 10^{-6}$. This means that, to a very good approximation, we can take

$$
\kappa_1 \simeq \kappa_2 \simeq \kappa_3 \simeq \kappa_4 \simeq \cdots \simeq \bar{\Lambda},
\tag{B.7}
$$

i.e., Λ is approximately Poisson-distributed.

B.2 Fourth moment of the noisy spherical harmonic components

In order to calculate the variance of our improved C_ℓ estimator (1.67) in the presence of shot noise, we need to evaluate the fourth moment of the noisy SHCs, $\Omega_{\ell m}^i$. We can do this by generalising our results above to a set of distinct volume elements observed at different times δV_i, whose total CBC counts Λ_i are i.i.d.. To do so, we start with equation (B.3) and insert a Kronecker symbol $\delta_{ij\ldots}$ into each term in the sum, enforcing the statistical independence of each of the different volume elements,

$$
\begin{aligned}
\langle\Lambda_i\Lambda_j\Lambda_k\Lambda_l\rangle_S &= \kappa_1^4 + \kappa_2\kappa_1^2\left(\delta_{ij} + \delta_{ik} + \delta_{il} + \delta_{jk} + \delta_{jl} + \delta_{kl}\right) \\
&\quad + \kappa_2^2\left(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) + \kappa_3\kappa_1\left(\delta_{ijk} + \delta_{ijl} + \delta_{ikl} + \delta_{jkl}\right) + \kappa_4\delta_{ijkl}.
\end{aligned}
\tag{B.8}
$$

Here δ_{ijk} is a generalisation of the Kronecker delta δ_{ij}, which is equal to unity if $i = j = k$ and vanishes otherwise (and similarly for δ_{ijkl}). Note that we have re-introduced the notation for the shot noise average $\langle\ldots\rangle_S$ from section 1.3.5, as it is important to distinguish between this and the cosmological average $\langle\ldots\rangle_\Omega$ when calculating the variance of the C_ℓ estimator under both ensembles.

Rewriting in terms of the comoving CBC rate density, $\mathcal{R}(\boldsymbol{r}, \boldsymbol{\zeta}) \equiv nR = \lim_{\delta V \to 0} \Lambda/(\tau_s \delta V)$, and introducing $\boldsymbol{\zeta}$ to represent the parameters of the CBC (masses, spins, …) and of the galaxy (star formation rate, metallicity, …), we obtain

$$
\begin{aligned}
\langle \mathcal{R}_i \mathcal{R}_j \mathcal{R}_k \mathcal{R}_l \rangle_S &= \bar{\mathcal{R}}_i \bar{\mathcal{R}}_j \bar{\mathcal{R}}_k \bar{\mathcal{R}}_l + \left[\bar{\mathcal{R}}_i \bar{\mathcal{R}}_j \bar{\mathcal{R}}_k (1 + \bar{\lambda}_k) \frac{\delta_{kl}}{\tau_{s,k}} + 5 \text{ perms.} \right] \\
&+ \left[\bar{\mathcal{R}}_i \bar{\mathcal{R}}_k (1 + \bar{\lambda}_i)(1 + \bar{\lambda}_k) \frac{\delta_{ij}}{\tau_{s,i}} \frac{\delta_{kl}}{\tau_{s,k}} + 2 \text{ perms.} \right] \\
&+ \left[\bar{\mathcal{R}}_i \bar{\mathcal{R}}_j (1 + 3\bar{\lambda}_j + \bar{\lambda}_j^2) \frac{\delta_{jkl}}{\tau_{s,j}^2} + 3 \text{ perms.} \right] + \bar{\mathcal{R}}_i (1 + 7\bar{\lambda}_i + 6\bar{\lambda}_i^2 + \bar{\lambda}_i^3) \frac{\delta_{ijkl}}{\tau_{s,i}^3},
\end{aligned}
\tag{B.9}
$$

where here the Kronecker symbols are shorthand for

$$
\delta_{ij} \to \delta_{ij} \delta^3(\boldsymbol{r}_i - \boldsymbol{r}_j) \delta(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j),
\tag{B.10}
$$

i.e., two GW sources must coincide in space, in time, and in parameter space, in order to contribute to the shot noise.

Now, using equation (1.50) to write the SHCs in terms of the CBC rate density, we find

$$
\begin{aligned}
&\left\langle \Omega_{\ell m}^i \Omega_{\ell m}^{j*} \Omega_{\ell m'}^{i'*} \Omega_{\ell m'}^{j'} \right\rangle_S \\
&= \left(\frac{2G}{3} \right)^4 \int \mathrm{d}^3 \boldsymbol{r}_i \, \mathrm{d}^3 \boldsymbol{r}_j \, \mathrm{d}^3 \boldsymbol{r}_{i'} \, \mathrm{d}^3 \boldsymbol{r}_{j'} \, Y_{\ell m}^{i*} Y_{\ell m}^{j} Y_{\ell m'}^{i'} Y_{\ell m'}^{j'*} (1 + z_i)^{-2} (1 + z_j)^{-2} (1 + z_{i'})^{-2} (1 + z_{j'})^{-2} \\
&\times \left(\frac{r_H^4}{r_i r_j r_{i'} r_{j'}} \right)^2 \int \mathrm{d}\boldsymbol{\zeta}_i \, \mathrm{d}\boldsymbol{\zeta}_j \, \mathrm{d}\boldsymbol{\zeta}_{i'} \, \mathrm{d}\boldsymbol{\zeta}_{j'} \, \frac{\mathrm{d}E_i}{\mathrm{d}(\ln f_{s,i})} \frac{\mathrm{d}E_j}{\mathrm{d}(\ln f_{s,j})} \frac{\mathrm{d}E_{i'}}{\mathrm{d}(\ln f_{s,i'})} \frac{\mathrm{d}E_{j'}}{\mathrm{d}(\ln f_{s,j'})} \langle \mathcal{R}_i \mathcal{R}_j \mathcal{R}_{i'} \mathcal{R}_{j'} \rangle_S,
\end{aligned}
\tag{B.11}
$$

where $Y_{\ell m}^i$ is shorthand for $Y_{\ell m}(\hat{\boldsymbol{r}}_i)$, etc. Using (B.9) and the properties of the spherical harmonics, we get

$$
\begin{aligned}
&\left\langle \Omega_{\ell m}^i \Omega_{\ell m}^{j*} \Omega_{\ell m'}^{i'*} \Omega_{\ell m'}^{j'} \right\rangle_S \\
&= |\Omega_{\ell m}|^2 |\Omega_{\ell m'}|^2 + \mathcal{W}_\tau |\Omega_{\ell m}|^2 \left[\delta_{i'j'} + \delta_{m,-m'}(\delta_{i'j} + \delta_{ij'}) \right] + \mathcal{W}_\tau |\Omega_{\ell m'}|^2 \left[\delta_{ij} + \delta_{mm'}(\delta_{ii'} + \delta_{jj'}) \right] \\
&+ \mathcal{W}_\tau^2 (\delta_{ij} \delta_{i'j'} + \delta_{mm'} \delta_{ii'} \delta_{jj'} + \delta_{m,-m'} \delta_{ij'} \delta_{i'j}) + c_{\ell m}^{(3)} \mathcal{X}_\tau \Omega_{\ell 0} \delta_{m'0} (\delta_{ii'j} + \delta_{ijj'}) \\
&+ c_{\ell m'}^{(3)} \mathcal{X}_\tau \Omega_{\ell 0} \delta_{m0} (\delta_{ii'j'} + \delta_{i'jj'}) + c_{\ell m m'}^{(4)} \mathcal{Y}_\tau \delta_{ii'jj'},
\end{aligned}
\tag{B.12}
$$

where the shorthand (B.10) is now no longer used. The coefficients on the 3- and 4-point terms are

$$
\begin{aligned}
c_{\ell m}^{(3)} &\equiv (-)^m \frac{(2\ell + 1)^{3/2}}{\sqrt{4\pi}} \begin{pmatrix} \ell & \ell & \ell \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell & \ell & \ell \\ 0 & m & -m \end{pmatrix}, \\
c_{\ell m m'}^{(4)} &\equiv \sum_{\ell'=0}^{2\ell} (-)^{m+m'} \frac{(2\ell' + 1)(2\ell + 1)^2}{4\pi} \begin{pmatrix} \ell & \ell & \ell' \\ -m & m & 0 \end{pmatrix} \begin{pmatrix} \ell & \ell & \ell' \\ m' & -m' & 0 \end{pmatrix} \begin{pmatrix} \ell & \ell & \ell' \\ 0 & 0 & 0 \end{pmatrix}^2,
\end{aligned}
\tag{B.13}
$$

and are just combinations of Wigner $3j$ symbols. In addition to the 2-point shot noise power $\mathcal{W}_\tau$ we defined before, we now have 3-point and 4-point terms that appear,

$$
\begin{aligned}
\mathcal{W}_\tau &\equiv \left(\frac{2G}{3}\right)^2 \frac{r_H^4}{\tau} \int \frac{\mathrm{d}r}{(1+z)^3 r^2} \int \mathrm{d}\zeta\, \bar{\mathcal{R}}(1+\bar{\lambda}) \left[\frac{\mathrm{d}E}{\mathrm{d}(\ln f_s)}\right]^2, \\
\mathcal{X}_\tau &\equiv \left(\frac{2G}{3}\right)^3 \frac{r_H^6}{\tau^2} \int \frac{\mathrm{d}r}{(1+z)^4 r^4} \int \mathrm{d}\zeta\, \bar{\mathcal{R}}(1+3\bar{\lambda}+\bar{\lambda}^2) \left[\frac{\mathrm{d}E}{\mathrm{d}(\ln f_s)}\right]^3, \\
\mathcal{Y}_\tau &\equiv \left(\frac{2G}{3}\right)^4 \frac{r_H^8}{\tau^3} \int \frac{\mathrm{d}r}{(1+z)^5 r^6} \int \mathrm{d}\zeta\, \bar{\mathcal{R}}(1+7\bar{\lambda}+6\bar{\lambda}^2+\bar{\lambda}^3) \left[\frac{\mathrm{d}E}{\mathrm{d}(\ln f_s)}\right]^4,
\end{aligned}
\tag{B.14}
$$

where $\bar{\lambda}(r, \zeta) = \bar{\mathcal{R}}\tau_s/\bar{n} \ll 1$. (Note that the first line here is just a different way of writing equation (1.59).)

Taking the cosmological average of (B.12) and subtracting the 2nd moments, we therefore find

$$
\begin{aligned}
\mathrm{Cov}\left[\Omega^i_{\ell m}\Omega^{j*}_{\ell m}, \Omega^{i'}_{\ell m'}\Omega^{j'*}_{\ell m'}\right]_{S,\Omega} &= \delta_{mm'}(C_\ell + \delta_{ii'}\mathcal{W}_\tau)(C_\ell + \delta_{jj'}\mathcal{W}_\tau) \\
&\quad + \delta_{m,-m'}(C_\ell + \delta_{ij'}\mathcal{W}_\tau)(C_\ell + \delta_{i'j}\mathcal{W}_\tau) + c^{(4)}_{\ell m m'}\mathcal{Y}_\tau \delta_{ii'jj'}.
\end{aligned}
\tag{B.15}
$$

Note that the term proportional to $\mathcal{X}_\tau$ has vanished, due to the SHCs having zero mean for $\ell > 0$ (recall that we do not attempt to estimate C_0, due to the degeneracy with the unknown mean intensity $\bar{\Omega}$). The term proportional to $\mathcal{Y}_\tau$ is associated with the fourth cumulant κ_4, and is therefore a sign of the non-Gaussian nature of the shot noise fluctuations (as $\kappa_n = 0$ for all $n > 2$ in the Gaussian case). However, this term does not end up contributing to the variance of our improved estimator $\hat{C}_\ell$, as this includes only off-diagonal pairs $i \neq j$ and $i' \neq j'$ by design.

C

HIGHER-ORDER MEMORY FROM COSMIC STRINGS

C.1 UNDERSTANDING THE ORIGIN OF THE HIGHER-ORDER MEMORY DIVERGENCE

Here we derive a condition (C.5) on a generic GW signal $h(t)$ which, we argue heuristically, is necessary for the associated nonlinear memory to diverge. We find this condition by considering a simple toy model in which we can tune the 'sharpness' of the signal to find where the divergence sets in. We then show that this condition predicts the divergence for cusps on 'large' cosmic string loops, while also providing an explanation for why the memory from compact binaries never diverges.

GAUSSIAN PULSE AS A TOY MODEL

Our goal here is to derive a condition on how 'sharp' a putative GW signal must be to give rise to a divergent nonlinear memory expansion. To investigate this, we use a toy model in which the primary GW is a Gaussian pulse which reaches our detector at retarded time $t = 0$,

$$h^{(0)}(t) = \frac{A}{r} \exp\left(-\frac{t^2}{2\sigma^2}\right). \tag{C.1}$$

This has amplitude A and width σ, both of which have dimensions of length. We ignore the polarisation and angular pattern of the pulse, though these can play an important role (e.g., for the case of kinks, as discussed in section 2.4 above), and focus on the pulse's behaviour as a function of time. The Fourier transform of equation (C.1) is

$$\tilde{h}^{(0)}(f) = \frac{A}{r}\sqrt{2\pi}\sigma \exp\left(-\frac{(2\pi f \sigma)^2}{2}\right), \tag{C.2}$$

so we see that the pulse has a characteristic frequency of $\sim 1/\sigma$. The pulse is smooth and infinitely differentiable for all finite σ; however, by making σ small, we can make the pulse arbitrarily 'sharp' (equivalently, we can make the characteristic frequency arbitrarily large). This sharpness can be quantified by the maximum value of the time derivative of the strain,

$$\max_t |r\dot{h}^{(0)}(t)| = e^{-1/2}A/\sigma \sim [\text{amplitude}] \times [\text{characteristic frequency}], \tag{C.3}$$

i.e. if the dimensionless ratio A/σ is much greater than unity the pulse is 'sharp', and if A/σ is much less than unity the pulse is 'soft'.

The pulse's first-order memory can be written as

$$r\dot{h}^{(1)} = C|r\dot{h}^{(0)}|^2 = C\left(\frac{At}{\sigma^2}\right)^2 \exp\left(-\frac{t^2}{\sigma^2}\right), \tag{C.4}$$

where C is some constant arising from the angular integral, which we assume to be $O(1)$ for all orders in the memory expansion. (This assumption can easily be violated: e.g., if the integrand has vanishing TT component then $C = 0$.) We immediately see that the memory signal is larger than the primary GW signal near the arrival time ($|t| \sim \sigma$) if and only if $A/\sigma \gg 1$, or equivalently,

$$\max_t |r\dot{h}(t)| \gg 1. \tag{C.5}$$

Using equation (0.43), we see that this is equivalent to $\max_t |\dot{E}_{\text{gw}}| \gg 1/G = m_{\text{Pl}}/t_{\text{Pl}}$, i.e. the memory is greater than the primary signal if the GW energy flux is trans-Planckian.

Moving on to the second-order memory, there are two contributions: the self-energy of the first-order memory, and its cross-energy with the primary signal,

$$r\dot{h}^{(2)} = Cr^2(|\dot{h}^{(1)} + \dot{h}^{(0)}|^2 - |\dot{h}^{(0)}|^2) = C|r\dot{h}^{(1)}|^2 + 2Cr^2\dot{h}^{(1)}\dot{h}^{(0)}$$
$$= C^3\left(\frac{At}{\sigma^2}\right)^4 \exp\left(-\frac{2t^2}{\sigma^2}\right) - 2C^2\left(\frac{At}{\sigma^2}\right)^3 \exp\left(-\frac{3t^2}{2\sigma^2}\right). \tag{C.6}$$

The general pattern is straightforward from here. Treating the 'sharp' regime $A/\sigma \gg 1$ and the 'soft' regime $A/\sigma \ll 1$ separately (analogous to the separation between the 'large loop' and 'small loop' regimes for cusps), we find for all $n \geq 1$,

$$r\dot{h}^{(n)} \simeq \begin{cases} \frac{1}{C}\left[\frac{CAt}{\sigma^2}\exp\left(-\frac{t^2}{2\sigma^2}\right)\right]^{2^n}, & A/\sigma \gg 1 \\ \frac{1}{4C}\left[-\frac{2CAt}{\sigma^2}\exp\left(-\frac{t^2}{2\sigma^2}\right)\right]^{n+1}, & A/\sigma \ll 1 \end{cases} \tag{C.7}$$

which shows that the memory expansion diverges if equation (C.5) holds.

Of course, there are many ways in which this argument could fail, some of which we have already mentioned (in particular, if $C \ll 1$). However, the general takeaway is that if the time derivative of a GW strain signal is large, then a memory divergence *might* occur. For $\max_t |r\dot{h}| \ll 1$ on the other hand, it seems very likely that the memory expansion must always converge. If this is indeed the case, then we could interpret equation (C.5) as a necessary, but not sufficient, condition for the memory divergence.

Application to cosmic strings

How does this fit into our results for cosmic string cusps? Neglecting numerical constants, the time-domain cusp waveform looks like

$$h_{\rm c}^{(0)}(t) \sim -\frac{G\mu}{r}\ell^{2/3}|t|^{1/3} + \text{constant}, \tag{C.8}$$

so the time derivative *diverges* at $t = 0$ due to the absolute value function. If we introduce a cutoff at the string width scale δ, we find that $\max_t |r\dot{h}_{\rm c}^{(0)}| \sim G\mu(\ell/\delta)^{2/3}$. Naively applying the condition (C.5), we would thus expect the cusp memory signal to diverge for loops of length $\ell \gtrsim \delta/(G\mu)^{3/2}$. However, this does *not* happen, for the simple reason that there is a small factor $C \ll 1$ arising from the angular integral, and as mentioned above this causes the condition (C.5) to fail.

We can circumvent this issue by considering the first-order memory signal from the cusp as the source of the divergence: this should give $C = O(1)$ as it has a broad emission pattern, and is not concentrated into a narrow beam. Indeed, using equation (2.66) we find that the maximum time derivative for the first-order cusp memory signal is $\max_t |r\dot{h}_{\rm c}^{(1)}| \sim (G\mu)^2(\ell/\delta)^{2/3}$. Applying the condition (C.5) again, we find that the cusp memory signal should diverge if $\ell \gtrsim \delta/(G\mu)^3$, which is exactly what we find from the careful analysis in section 2.3.8. This supports the idea that equation (C.5) gives a necessary condition for the memory divergence, which is generally applicable if the factor C arising from the angular integral is not too small.

Application to compact binaries

Here we use a very simple heuristic analysis to show that $|r\dot{h}(t)|$ is at most $O(1)$ for compact binary coalescences, so that the memory divergence condition (C.5) never holds. We neglect numerical factors throughout.

The GW signal from a CBC can be written as $h(t) = \mathcal{A}(t)e^{i\phi(t)}$, where the leading-order Newtonian contributions to the amplitude and phase are [406]

$$\mathcal{A} \sim \frac{(G\mathcal{M})^{5/4}}{r\tau^{1/4}}, \qquad \phi \sim \left(\frac{\tau}{G\mathcal{M}}\right)^{5/8}, \tag{C.9}$$

with τ the time until coalescence and $\mathcal{M} \equiv \eta^{3/5}M$ the chirp mass, where $\eta \leq 1/4$ is the dimensionless mass ratio. The time derivative of the strain is therefore

$$|r\dot{h}(t)| = \sqrt{|r\dot{\mathcal{A}}|^2 + |r\dot{\phi}\mathcal{A}|^2} \sim \sqrt{\left(\frac{G\mathcal{M}}{\tau}\right)^{5/2} + \left(\frac{G\mathcal{M}}{\tau}\right)^{5/4}}. \tag{C.10}$$

Formally, this diverges in the Newtonian analysis as $\tau \to 0$. However, introducing a cutoff at the ISCO radius truncates the signal at $\tau_{\rm min} \sim \eta^{-8/5}G\mathcal{M}$, so that $\max_t |r\dot{h}(t)| \sim \sqrt{\eta^2 + \eta^4}$. Since η is at most $O(1)$, we see that the condition (C.5) is never met. This makes complete sense, given that we know that the memory from CBC signals cannot diverge. As a by-product, this expression leads us to conjecture that equal-mass binaries ($\eta = 1/4$) should give rise to stronger higher-order memory effects than extreme mass-ratio inspirals ($\eta \ll 1$).
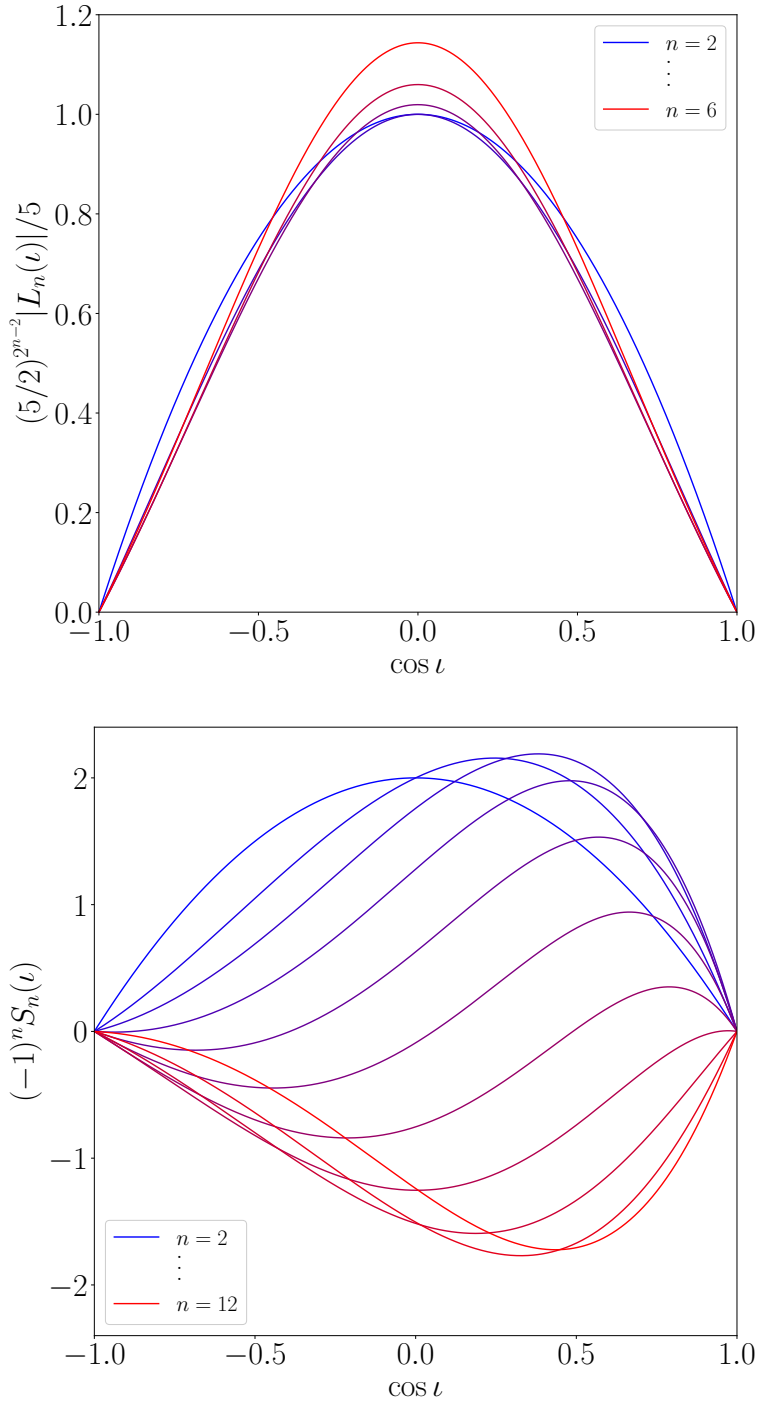
Figure C.1: Functions describing the strength of the nth-order cusp memory signal as a function of inclination, with $L_n(I)$ representing 'large' loops $\ell \gg \ell_*$, and $S_n(I)$ representing 'small' loops $\ell \ll \ell_*$.

C.2 ANGULAR INTEGRALS FOR HIGHER-ORDER CUSP MEMORY

Here we derive the angular patterns of the higher-order cusp memory in the 'large-loop' ($\ell \gg \ell_*$) and 'small-loop' ($\ell \ll \ell_*$) limits, which are described by the functions $L_n(I)$ and $S_n(I)$ respectively, with I

the inclination between the cusp beaming direction and the observer's line of sight. These are defined by inserting the nth-order memory formula (2.82) into the iterative relation (2.79), which gives

$$L_{n+1}(I) = \int_{\hat{r}'} |L_n(I')|^2, \qquad S_{n+1}(I) = \int_{\hat{r}'} 6(1 + \cos I') S_n(I'), \qquad \text{(C.11)}$$

for all $n \geq 2$, with $L_2(I) = S_2(I) = 2 \sin^2 I$. (The integral symbol $\int_{\hat{r}'}$ is defined in equation (2.54).)

We find that L_n and S_n can be written as polynomials in $\cos I$ (of order 2^{n-1} and n, respectively). The iterative process (C.11) can therefore be carried out by evaluating a simpler family of integrals, $C_n(I) \equiv \int_{\hat{r}'} \cos^n I'$. To do so, we choose our $\hat{r}' = (\theta', \phi')$ coordinates such that θ' is zero along the line of sight, so that $\cos \theta' \equiv \hat{r} \cdot \hat{r}'$, with the cusp beaming direction defined relative to the line of sight by $\hat{r}_c \cdot \hat{r} = \cos I$. We then have $\cos I' \equiv \hat{r}_c \cdot \hat{r}' = \cos \theta' \cos I - \cos \phi' \sin \theta' \sin I$, and equation (2.54) becomes

$$\int_{\hat{r}'} = \int_{S^2} \frac{\mathrm{d}^2 \hat{r}'}{4\pi} (1 + \cos \theta') e^{-2\mathrm{i}\phi'}, \qquad \text{(C.12)}$$

so that we can use a binomial expansion of $\cos^n I' = (\cos \theta' \cos I - \cos \phi' \sin \theta' \sin I)^n$ to write

$$C_n(I) = \sum_{k=0}^{n} \binom{n}{k} (-\sin I)^k \cos^{n-k} I \int_{-1}^{+1} \frac{\mathrm{d}x}{2} (1+x) x^{n-k} (1-x^2)^{k/2} \int_0^{2\pi} \frac{\mathrm{d}\phi'}{2\pi} e^{-2\mathrm{i}\phi'} \cos^k \phi', \quad \text{(C.13)}$$

where we have set $x = \cos \theta'$. Using the Beta function identity

$$B(a,b) \equiv \int_0^1 \mathrm{d}x \, x^{a-1}(1-x)^{b-1} = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \qquad \text{(C.14)}$$

we then obtain, for all $n \geq 0$,

$$
\begin{aligned}
C_{2n}(I) &= \sum_{k=0}^{n} \frac{\sqrt{\pi}(n-k)(2n)! \cos^{2k} I \sin^{2(n-k)} I}{2^{2n+1} k! (n-k+1)! \Gamma(n+\frac{3}{2})}, \\
C_{2n+1}(I) &= \sum_{k=0}^{n} \frac{\sqrt{\pi}(n-k)(2n+1)! \cos^{2k+1} I \sin^{2(n-k)} I}{2^{2n+2} k! (n-k+1)! \Gamma(n+\frac{5}{2})}.
\end{aligned}
\qquad \text{(C.15)}
$$

Calculating each iteration of equation (C.11) is then reduced to writing down the appropriate linear combination of $C_n(I)$ for different n.

The resulting expressions for $S_n(I)$ and $L_n(I)$ for the first few $n \geq 2$ are shown in Tables C.1 and C.2 respectively. We find empirically that the large-loop angular functions can be approximated by $|L_n(I)| \approx 5(2/5)^{2^{n-2}} \sin^2 I$ with an accuracy of $\sim 10\%$, as is illustrated in the top panel of figure C.1. The small-loop angular functions $S_n(I)$ do not seem to follow such a simple pattern.

C.3 ANGULAR INTEGRALS FOR KINK MEMORY

Here we compute the integral $K_n(I)$ defined in equation (2.94) for all angular modes $n \in \mathbb{Z}$. Note that while the integrand is complex, the integral itself is always real, as the imaginary part of the integrand is

n	$S_n(I)$
2	$2\sin^2 I$
3	$-\dfrac{2\sin^2 I}{5}(5 + 3\cos I)$
4	$\dfrac{2\sin^2 I}{25}(25 + 24\cos I + 3\cos 2I)$
5	$-\dfrac{2\sin^2 I}{175}(154 + 201\cos I + 42\cos 2I + 3\cos 3I)$
6	$\dfrac{\sin^2 I}{12250}(15575 + 28032\cos I + 7932\cos 2I + 960\cos 3I + 45\cos 4I)$
7	$-\dfrac{\sin^2 I}{245000}\big[466422\cos I + 172824\cos 2I + 5(29702 + 5619\cos 3I + 450\cos 4I + 15\cos 5I)\big]$
8	$\dfrac{\sin^2 I}{245000}\big(-30194 + 299024\cos I + 157771\cos 2I$ $\qquad\qquad +32552\cos 3I + 3490\cos 4I + 200\cos 5I + 5\cos 6I\big)$
9	$-\dfrac{\sin^2 I}{13475000}\big[4934669\cos I + 6432382\cos 2I + 5\big(-2137388 + 347047\cos 3I$ $\qquad\qquad\qquad\qquad +46508\cos 4I + 3565\cos 5I + 154\cos 6I + 3\cos 7I\big)\big]$
10	$\dfrac{\sin^2 I}{10375750000}\big[-5257311104\cos I + 2472686192\cos 2I$ $\qquad +5\big(214599584\cos 3I + 36187180\cos 4I$ $\qquad -7(383512899 - 491360\cos 5I - 28400\cos 6I - 960\cos 7I - 15\cos 8I)\big)\big]$
11	$-\dfrac{\sin^2 I}{1888386500000}\big[-2359095296230\cos I - 55089058544\cos 2I + 116827655852\cos 3I$ $\qquad +27832319320\cos 4I + 3261899900\cos 5I$ $\qquad +49(-59685302066 + 4749680\cos 6I + 214995\cos 7I + 5850\cos 8I + 75\cos 9I)\big]$
12	$\dfrac{\sin^2 I}{122745122500000}\big[-213800187213104\cos I - 34153675430026\cos 2I$ $\qquad +5\big(-37491558286234 + 292405153376\cos 3I$ $\qquad +238091184664\cos 4I + 37145938400\cos 5I + 3229010267\cos 6I$ $\qquad +179962104\cos 7I + 6563970\cos 8I + 147000\cos 9I + 1575\cos 10I\big)\big]$

Table C.1: The first few small-loop angular functions, as defined by equation (C.11) with $S_2(I) = 2\sin^2 I$. Note that $S_n(I)$ is a polynomial in $\cos I$ of order n.

n	$L_n(I)$
2	$2\sin^2 I$
3	$-\dfrac{2\sin^2 I}{15}(5 - \cos 2I)$
4	$\dfrac{\sin^2 I}{141750}\left[3737\cos 2I - 262\cos 4I + 7(-2070 + \cos 6I)\right]$
5	$\dfrac{\sin^2 I}{123092512042500000}\left[91170878299511\cos 2I - 88445052038863\cos 4I + 627703754313\cos 6I - 31676232034\cos 8I + 1106354755\cos 10I + 1001(-326445712830 - 24241\cos 12I + 245\cos 14I)\right]$
6	$\dfrac{\sin^2 I}{437593783647405440659455095457213000000000000000000}\Big[-25238376246201225173073537764237047796445\cos 4I + 22676810284676710047197528844734128748260\cos 6I - 17390645487685587069032755384117598990046\cos 8I + 11470781661141147474713389812292222447740\cos 10I - 6538676998608577029462451915641765109\cos 12I + 3221106161657437764475529017764750040\cos 14I - 1363891120977935353046627931115110580\cos 16I + 4908821919722365197294034447127344\cos 18I + 1463(-10081167364753857595829662265\cos 20I + 2462468614242487202858018 92\cos 22I + 1495(-31450804813245988910714\cos 24I + 44031911264492955804\cos 26I + 6525(-569906751862925826959924173050076 - 61436527080929\cos 28I + 2711997768840\cos 30I)))\Big]$

Table C.2: The first few large-loop angular functions, as defined by equation (C.11) with $L_2(I) = 2\sin^2 I$. Note that $L_n(I)$ is a polynomial in $\cos I$ of order 2^{n-1}, such that the number of terms grows exponentially with n. Despite this apparent complexity, we find that these formulae can be approximated by the simple expression $|L_n(I)| \approx 5(2/5)^{2^{n-2}}\sin^2 I$.
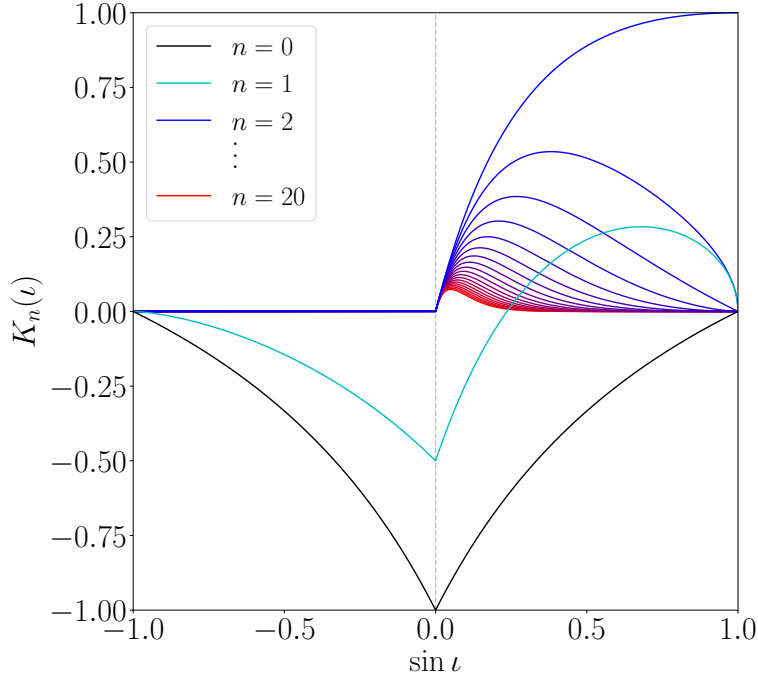
Figure C.2: The angular integral $K_n(I)$ for non-negative n, as given by equation (C.19). The corresponding curves for negative n are obtained by reflecting around $I = 0$.

an odd function of ϕ. Note also that we can focus on non-negative $n \geq 0$ by exploiting the symmetry property $K_{-n}(I) = K_n(-I)$.

For some special values of I we can obtain the full spectrum immediately from equation (2.94),

$$K_n(0) = -\delta_{n,0} - \frac{1}{2}(\delta_{n,1} + \delta_{n,-1}), \qquad K_n(\pm\pi/2) = \delta_{n,\pm 2}. \tag{C.16}$$

For general values $I \in [-\pi/2, +\pi/2]$, we proceed by expanding the denominator of equation (2.94) using the geometric series and the binomial theorem,

$$\frac{1}{1 - \cos I \cos \phi} = \sum_{k=0}^{\infty} \cos^k I \cos^k \phi = \sum_{k=0}^{\infty} \sum_{m=0}^{k} \binom{k}{m} \frac{\cos^k I}{2^k} e^{-i(k-2m)\phi}. \tag{C.17}$$

This converges everywhere on $I \in [-\pi/2, +\pi/2]$ except for $I = 0$, in which case we use equation (C.16) instead. We can then integrate term-by-term to extract the contribution from each order in the series. The result, valid for all $n \geq 0$, is

$$
\begin{aligned}
K_n(I) = \sum_{k=\max(0,2-n)}^{\infty} & \binom{2k+n-2}{k} \left(\frac{\cos I}{2}\right)^{2k+n-2} \left(\frac{1+\sin I}{2}\right)^2 \\
& - 2\sum_{k=0}^{\infty} \binom{2k+n}{k} \left(\frac{\cos I}{2}\right)^{2k+n+2} + \sum_{k=0}^{\infty} \binom{2k+n+2}{k} \left(\frac{\cos I}{2}\right)^{2k+n+2} \left(\frac{1-\sin I}{2}\right)^2.
\end{aligned} \tag{C.18}
$$

The first sum must be carried out separately for the three cases $n = 0$, $n = 1$, and $n > 1$, yielding the following expressions:

$$K_0(I) = \frac{4\sin|I| + \cos 2I - 3}{2\cos^2 I},$$

$$K_1(I) = \frac{\sin|I| - 1}{2\sin|I|}\left[\cos I + (\sin I - 1)\frac{1 - \sin|I| + \sin I(1 + 2\sin I + \sin|I|)}{\cos^3 I}\right],$$

$$K_n(I) = \begin{cases} \frac{4\sin I}{\cos^2 I}\left(\frac{\cos I}{1 + \sin I}\right)^n, & I > 0 \\ 0, & I \leq 0 \end{cases}, \qquad \text{for } n > 1.$$

(C.19)

These clearly agree with equation (C.16) for $I = \pm\pi/2$. Note that for all n, the integral $K_n(I)$ is not differentiable at $I = 0$; this is related to the fact that the series in equation (C.17) diverges at $I = 0$. However, despite this formal divergence, we see that equation (C.19) agrees with equation (C.16) in the limit $I \to 0$, whether this limit is taken from above or from below. We therefore use equation (C.19) over the full domain $I \in [-\pi/2, +\pi/2]$. The resulting curves for the first few non-negative n are illustrated in figure C.2.

For $n \geq 2$, the integral only has support for $I > 0$, and peaks at an inclination I_* given by $I_* = \sin^{-1}[(n/2) - \sqrt{(n/2)^2 - 1}]$. When calculating the kink memory signal we are interested in the high-frequency regime, which corresponds to large n. In the limit $n \to \infty$, we have $I_* \simeq 1/n$ and $K_n(I_*) \simeq 4/(en)$, meaning that the memory signal is strongly suppressed, and is only observable very close to (but strictly outside of) the plane of the kink.

D DERIVATION OF THE SECULAR KRAMERS-MOYAL COEFFICIENTS

D.1 POLARISATION TENSORS IN THE BINARY'S COORDINATE FRAME

In order to describe the GW polarisation tensors e_{ij}^A, we introduce the orthonormal frame $(\hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{n}})$, where $\hat{\boldsymbol{n}}$ is the GW propagation direction (see figure 3.2). We want to find the components of these basis vectors in the frame of the binary, $(\hat{\boldsymbol{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\ell}})$, as this determines the binary's response to the GW. First, we transform from the GW frame to the fixed reference frame $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ by applying the standard rotations with respect to the zenith ϑ and azimuth ϕ,

$$
\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathsf{R}_\phi \, \mathsf{R}_\vartheta \begin{pmatrix} u \\ v \\ n \end{pmatrix}, \qquad \mathsf{R}_\vartheta = \begin{pmatrix} \cos\vartheta & 0 & \sin\vartheta \\ 0 & 1 & 0 \\ -\sin\vartheta & 0 & \cos\vartheta \end{pmatrix}, \quad \mathsf{R}_\phi = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{D.1}
$$

This reference frame is transformed to the binary frame $(\hat{\boldsymbol{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\ell}})$ with three further rotations, which specify the inclination I, the longitude of ascending node Ω, and the argument of the binary in the orbital plane $\theta = \psi + \omega$ [432],

$$
\begin{pmatrix} r \\ \theta \\ \ell \end{pmatrix} = \mathsf{R}_\theta \, \mathsf{R}_I \, \mathsf{R}_\Omega \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \tag{D.2}
$$

where

$$
\mathsf{R}_\Omega = \begin{pmatrix} \cos\Omega & \sin\Omega & 0 \\ -\sin\Omega & \cos\Omega & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathsf{R}_I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos I & \sin I \\ 0 & -\sin I & \cos I \end{pmatrix}, \qquad \mathsf{R}_\theta = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{D.3}
$$

We thus obtain the desired relationship between the binary frame and the GW frame by applying all five rotations,

$$
\begin{pmatrix} r \\ \theta \\ \ell \end{pmatrix} = \mathsf{R}_\theta \, \mathsf{R}_I \, \mathsf{R}_\Omega \, \mathsf{R}_\phi \, \mathsf{R}_\vartheta \begin{pmatrix} u \\ v \\ n \end{pmatrix}. \tag{D.4}
$$

The various contractions with the polarisation tensors are then given by

$$
e^+_{ij}\hat{r}^i\hat{r}^j = [\sin\vartheta\sin I\sin\theta - \cos\vartheta(\cos\varphi\cos\theta + \sin\varphi\cos I\sin\theta)]^2 - [\cos\varphi\cos I\sin\theta - \sin\varphi\cos\theta]^2,
$$

$$
e^\times_{ij}\hat{r}^i\hat{r}^j = 2(\cos\varphi\cos I\sin\theta - \sin\varphi\cos\theta)[\cos\vartheta(\cos\varphi\cos\theta + \sin\varphi\cos I\sin\theta) - \sin\vartheta\sin I\sin\theta],
$$

$$
e^+_{ij}\hat{r}^i\hat{\theta}^j = (\sin\varphi\cos\theta - \cos\varphi\cos I\sin\theta)(\cos\varphi\cos I\cos\theta + \sin\varphi\sin\theta)
$$
$$
\quad - [\cos\vartheta\cos\varphi\sin\theta + (\sin\vartheta\sin I - \cos\vartheta\sin\varphi\cos I)\cos\theta]
$$
$$
\quad\quad \times [\cos\vartheta(\cos\varphi\cos\theta + \sin\varphi\cos I\sin\theta) - \sin\vartheta\sin I\sin\theta],
$$

$$
e^\times_{ij}\hat{r}^i\hat{\theta}^j = \cos\vartheta\cos 2\varphi\cos I\cos 2\theta + \sin\vartheta\sin I(\sin\varphi\cos 2\theta - \cos\varphi\cos I\sin 2\theta)
$$
$$
\quad + \frac{1}{2}\cos\vartheta\sin 2\varphi(1 + \cos^2 I)\sin 2\theta,
$$

$$
e^+_{ij}\hat{r}^i\hat{\ell}^j = \cos\varphi\sin I(\cos\varphi\cos I\sin\theta - \sin\varphi\cos\theta)
$$
$$
\quad - (\sin\vartheta\cos I + \cos\vartheta\sin\varphi\sin I)[\cos\vartheta(\cos\varphi\cos\theta + \sin\varphi\cos I\sin\theta) - \sin\vartheta\sin I\sin\theta],
$$

$$
e^\times_{ij}\hat{r}^i\hat{\ell}^j = \sin\vartheta(\sin\varphi\cos I\cos\theta - \cos\varphi\cos 2I\sin\theta) - \cos\vartheta\sin I(\cos 2\varphi\cos\theta + \sin 2\varphi\cos I\sin\theta),
$$

$$
\text{(D.5)}
$$

where we define $\varphi \equiv \phi - \Omega$. We recall that ϑ, ϕ are the spherical coordinates of the incoming plane GW, Ω is the longitude of ascending node, and $\theta \equiv \psi + \omega$ is the orbital argument with respect to the ascending node, with ψ the true anomaly and ω the argument of pericentre; these are all illustrated in figure 3.2.

D.2 TRANSFER FUNCTIONS

The Fourier components of the transfer functions are defined in terms of the polarisation tensor contractions discussed in section D.1 by

$$
T^A_{P,n} = \frac{3P^2\gamma}{4\pi}\left\langle\frac{e\sin\psi}{1 + e\cos\psi}e^A_{ij}\hat{r}_i\hat{r}_j + e^A_{ij}\hat{r}_i\hat{\theta}_j\right\rangle_n, \qquad T^A_{e,n} = \frac{\gamma^2}{3Pe}T^A_{P,n} - \frac{P\gamma^5}{4\pi e}\left\langle\frac{e^A_{ij}\hat{r}_i\hat{\theta}_j}{(1 + e\cos\psi)^2}\right\rangle_n,
$$

$$
T^A_{I,n} = \frac{P\gamma^3}{4\pi}\left\langle\frac{\cos\theta}{(1 + e\cos\psi)^2}e^A_{ij}\hat{r}_i\hat{\ell}_j\right\rangle_n, \qquad T^A_{\Omega,n} = \frac{P\gamma^3}{4\pi\sin I}\left\langle\frac{\sin\theta}{(1 + e\cos\psi)^2}e^A_{ij}\hat{r}_i\hat{\ell}_j\right\rangle_n,
$$

$$
T^A_{\omega,n} = \frac{P\gamma^3}{4\pi e}\left\langle\frac{\sin\psi(2 + e\cos\psi)}{(1 + e\cos\psi)^2}e^A_{ij}\hat{r}_i\hat{\theta}_j - \frac{\cos\psi\, e^A_{ij}\hat{r}_i\hat{r}_j}{1 + e\cos\psi}\right\rangle_n - \cos I\, T^A_{\Omega,n},
$$

$$
T^A_{\varepsilon,n} = -\frac{P\gamma^4}{2\pi}\left\langle\frac{e^A_{ij}\hat{r}_i\hat{r}_j}{(1 + e\cos\psi)^2}\right\rangle_n - \gamma\cos I\, T^A_{\Omega,n} - \gamma T^A_{\omega,n},
$$

$$
\text{(D.6)}
$$

where we have introduced the secular averaging operation

$$
\langle\cdots\rangle_n \equiv \int_{t_0}^{t_0+P}\frac{\mathrm{d}t}{P}\exp\left(\frac{2\pi int}{P}\right)(\cdots), \tag{D.7}
$$

which extracts the nth-order Fourier coefficient of a given function of the true anomaly $\psi(t)$, holding the orbital elements fixed (as they vary over much longer timescales). The subscript n distinguishes this from the ensemble average $\langle\cdots\rangle$, and from the secular average $\langle\cdots\rangle_{\text{sec}}$ (with the latter being equivalent to $\langle\cdots\rangle_n$ for $n = 0$).

The Fourier components arising in the eccentric transfer functions can be expressed in terms of *Hansen coefficients*, C_n^{lm}, S_n^{lm} [157]. These are functions of eccentricity that have been used for centuries in celestial mechanics to describe Keplerian motion. We define them here by

$$C_n^{lm}(e) = \left\langle \frac{\cos m\psi}{(1 + e\cos\psi)^l} \right\rangle_n, \qquad S_n^{lm}(e) = \left\langle \frac{\sin m\psi}{(1 + e\cos\psi)^l} \right\rangle_n, \tag{D.8}$$

with explicit expressions for particular sets of (l, m) given in section D.5.

Inserting equation (D.5) into equation (D.6) and using equation (D.8), we can thus write the transfer functions as linear combinations of the Hansen coefficients. These are then the input to computing the KM coefficients.

D.3 Transforming the reference frame

The polarisation tensor contractions in equation (D.5) are completely general and apply to any choice of reference frame. However, the corresponding expressions for the transfer functions are very lengthy, making it prohibitively difficult to calculate the KM coefficients. We circumvent this difficulty by choosing a particular reference frame in which the transfer functions are much simpler, before transforming back to a general reference frame.

Given two reference frames, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ and $(\hat{\boldsymbol{x}}', \hat{\boldsymbol{y}}', \hat{\boldsymbol{z}}')$, we have two corresponding sets of orbital elements for the binary, $X = (P, e, I, \Omega, \omega, \varepsilon)$ and $X' = (P, e, I', \Omega', \omega', \varepsilon)$ (the period, eccentricity, and compensated mean anomaly are the same in both frames, as they do not depend on the orientation of the binary in space). The KM coefficients for the unprimed elements are given in terms of those for the primed ones by (see, e.g., section 4.9 of Risken [484] for a derivation)

$$D_i^{(1)} = \frac{\partial X_i}{\partial X_{i'}} D_{i'}^{(1)} + \frac{\partial^2 X_i}{\partial X_{i'}\partial X_{j'}} D_{i'j'}^{(2)}, \qquad D_{ij}^{(2)} = \frac{\partial X_i}{\partial X_{i'}} \frac{\partial X_j}{\partial X_{j'}} D_{i'j'}^{(2)}, \tag{D.9}$$

where primed indices run over the primed orbital elements. We therefore require the first and second partial derivatives of the unprimed elements with respect to the primed ones. These can be deduced from the following relations, which are derived from section 2.8 of Murray and Dermott [432],

$$\cos I = s_{I'}\sin I'\sin\Omega'\hat{\boldsymbol{x}}'\cdot\hat{\boldsymbol{z}} - s_{I'}\sin I'\cos\Omega'\hat{\boldsymbol{y}}'\cdot\hat{\boldsymbol{z}} + \cos I'\hat{\boldsymbol{z}}'\cdot\hat{\boldsymbol{z}},$$

$$\sin I\sin\Omega = s_{I'}\sin I'\sin\Omega'\hat{\boldsymbol{x}}'\cdot\hat{\boldsymbol{x}} - s_{I'}\sin I'\cos\Omega'\hat{\boldsymbol{y}}'\cdot\hat{\boldsymbol{x}} + \cos I'\hat{\boldsymbol{z}}'\cdot\hat{\boldsymbol{x}},$$

$$\sin I\sin\omega = (\cos\Omega'\cos\omega' - \cos I'\sin\Omega'\sin\omega')\hat{\boldsymbol{x}}'\cdot\hat{\boldsymbol{z}} \tag{D.10}$$

$$+ (\sin\Omega'\cos\omega' + \cos I'\cos\Omega'\sin\omega')\hat{\boldsymbol{y}}'\cdot\hat{\boldsymbol{z}} + \sin I'\sin\omega\hat{\boldsymbol{z}}'\cdot\hat{\boldsymbol{z}},$$

where

$$s_{I'} \equiv \begin{cases} +1 & \text{if } \cos I' > 0, \\ -1 & \text{if } \cos I' < 0. \end{cases} \tag{D.11}$$

We are now free to specify the primed frame such that the KM coefficients are easier to compute. One particularly simple choice is to choose $(\hat{x}', \hat{y}', \hat{z}')$ such that $I' = \pi/4$ and $\Omega' = \omega' = 0$. (It may seem at first that $I' = 0$ is a simpler choice, as the reference frame then coincides with the binary's frame. However, there is a coordinate singularity associated with $I' = 0$ which makes some of the associated coefficients poorly-behaved. Taking $I' = \pi/4$ is much easier, particularly since we then have $\sin I' = \cos I'$, which simplifies many of the resulting expressions.)

Having specified the primed frame, we require the first and second derivatives to transform back to the unprimed frame, which is relevant for a general observer. Using equation (D.10), we find that the nonzero derivatives, evaluated at $(I', \Omega', \omega') = (\frac{\pi}{4}, 0, 0)$, are given by

$$\frac{\partial I}{\partial I'} = \cos \omega, \quad \frac{\partial I}{\partial \Omega'} = -\frac{\sin \omega}{\sqrt{2}}, \quad \frac{\partial \Omega}{\partial I'} = \frac{\sin \omega}{\sin I}, \quad \frac{\partial \Omega}{\partial \Omega'} = \frac{\cos \omega}{\sqrt{2} \sin I}, \quad \frac{\partial \omega}{\partial I'} = -\frac{\sin \omega}{\tan I},$$

$$\frac{\partial \omega}{\partial \Omega'} = \frac{1}{\sqrt{2}} \left(1 - \frac{\cos \omega}{\tan I} \right), \quad \frac{\partial \omega}{\partial \omega'} = 1, \quad \frac{\partial^2 I}{\partial I'^2} = \frac{\sin^2 \omega}{\tan I}, \quad \frac{\partial^2 I}{\partial I' \partial \Omega'} = \frac{\sin \omega}{\sqrt{2}} \left(\frac{\cos \omega}{\tan I} - 1 \right),$$

$$\frac{\partial^2 I}{\partial \Omega'^2} = \frac{\cos \omega}{2} \left(\frac{\cos \omega}{\tan I} - 1 \right), \quad \frac{\partial^2 \Omega}{\partial I'^2} = -\frac{\sin 2\omega}{\sin I \tan I}, \quad \frac{\partial^2 \Omega}{\partial I' \partial \Omega'} = \frac{\cos \omega - \cos 2\omega \cot I}{\sqrt{2} \sin I},$$

$$\frac{\partial^2 \Omega}{\partial \Omega'^2} = \frac{\sin \omega}{\sin I} \left(\frac{\cos \omega}{\tan I} - \frac{1}{2} \right), \quad \frac{\partial^2 \omega}{\partial I'^2} = \sin 2\omega \frac{2 - \sin^2 I}{2 \sin^2 I},$$

$$\frac{\partial^2 \omega}{\partial I' \partial \Omega'} = \frac{1}{2\sqrt{2}} \left(\cos 2\omega \frac{2 - \sin^2 I}{\sin^2 I} - \frac{2 \cos \omega}{\tan I} - 1 \right), \quad \frac{\partial^2 \omega}{\partial \Omega'^2} = \frac{\sin \omega}{\tan I} \left(\frac{1}{2} - \cos \omega \frac{2 - \sin^2 I}{\sin 2I} \right). \tag{D.12}$$

It is straightforward to confirm that these are well-behaved and take on the appropriate values when $(I, \Omega, \omega) \to (I', \Omega', \omega')$.

We thus find that the unprimed drift coefficients are given in terms of those in the primed frame by

$$D_I^{(1)} = \cos \omega \left[D_{I'}^{(1)} + \left(\frac{\cos \omega}{\tan I} - 1 \right) \frac{D_{\Omega'\Omega'}^{(2)}}{2} \right] + \frac{\sin^2 \omega}{\tan I} D_{I'I'}^{(2)},$$

$$D_\Omega^{(1)} = \frac{1}{\sin I} \left[\sin \omega D_{I'}^{(1)} - \frac{\sin 2\omega}{\tan I} D_{I'I'}^{(2)} + \sin \omega \left(\frac{\cos \omega}{\tan I} - \frac{1}{2} \right) D_{\Omega'\Omega'}^{(2)} \right], \tag{D.13}$$

$$D_\omega^{(1)} = \frac{\sin \omega}{\sin I} \left[-\cos I D_{I'}^{(1)} + \cos \omega \frac{2 - \sin^2 I}{\sin I} D_{I'I'}^{(2)} + \left(\cos I - \cos \omega \frac{2 - \sin^2 I}{\sin I} \right) \frac{D_{\Omega'\Omega'}^{(2)}}{2} \right],$$

with the diffusion coefficients given by

$$
D_{II}^{(2)} = \cos^2\omega D_{I'I'}^{(2)} + \frac{\sin^2\omega}{2}D_{\Omega'\Omega'}^{(2)}, \qquad D_{I\Omega}^{(2)} = \frac{\sin 2\omega}{2\sin I}\left(D_{I'I'}^{(2)} - \frac{1}{2}D_{\Omega'\Omega'}^{(2)}\right),
$$

$$
D_{I\omega}^{(2)} = \sin\omega\left[-\frac{\cos\omega}{\tan I}D_{I'I'}^{(2)} + \frac{1}{2}\left(\frac{\cos\omega}{\tan I} - 1\right)D_{\Omega'\Omega'}^{(2)} - \frac{1}{\sqrt{2}}D_{\Omega'\omega'}^{(2)}\right],
$$

$$
D_{\Omega\Omega}^{(2)} = \frac{1}{\sin^2 I}\left(\sin^2\omega D_{I'I'}^{(2)} + \frac{\cos^2\omega}{2}D_{\Omega'\Omega'}^{(2)}\right),
$$

$$
D_{\Omega\omega}^{(2)} = \frac{1}{\sin I}\left[-\frac{\sin^2\omega}{\tan I}D_{I'I'}^{(2)} + \frac{\cos\omega}{2}\left(1 - \frac{\cos\omega}{\tan I}\right)D_{\Omega'\Omega'}^{(2)} + \frac{\cos\omega}{\sqrt{2}}D_{\Omega'\omega'}^{(2)}\right],
$$

$$
D_{\omega\omega}^{(2)} = \frac{\sin^2\omega}{\tan^2 I}D_{I'I'}^{(2)} + \frac{1}{2}\left(1 - \frac{\cos\omega}{\tan I}\right)^2 D_{\Omega'\Omega'}^{(2)} + \sqrt{2}\left(1 - \frac{\cos\omega}{\tan I}\right)D_{\Omega'\omega'}^{(2)} + D_{\omega'\omega'}^{(2)},
$$

$$
D_{\omega\varepsilon}^{(2)} = D_{\omega'\varepsilon}^{(2)}, \qquad D_{XI}^{(2)} = D_{X\Omega}^{(2)} = D_{X\omega}^{(2)} = D_{X\varepsilon}^{(2)} = D_{I\varepsilon}^{(2)} = D_{\Omega\varepsilon}^{(2)} = 0,
$$

(D.14)

where X here stands for P or e.

D.4 KRAMERS-MOYAL COEFFICIENTS IN THE PRIMED FRAME

Here we give the KM coefficients evaluated in the primed coordinate frame $(I', \Omega', \omega') = (\frac{\pi}{4}, 0, 0)$. To calculate these, we first evaluate the polarisation tensor contractions from section D.1 in the primed frame, and insert these into equation (D.6) to find the appropriate GW transfer functions, expressing the secular averages in terms of Hansen coefficients that are listed below in section D.5. These secular transfer functions are then inserted into equation (3.51), integrating over the GW propagation direction $\hat{\boldsymbol{n}} = (\vartheta, \phi)$ to obtain the KM coefficients. The resulting expressions for the drift vector are

$$
D_P^{(1)} = V_P + \frac{9P^2\gamma^2}{80}\sum_{n=1}^{\infty} nH_0^2\Omega_n\left[2E_n^{02}\left(10eE_n^{02} + (3+7e^2)E_n^{11} + (1-11e^2)E_n^{13} - 2e\gamma^2 E_n^{22}\right)^*\right.
$$

$$
+ E_n^{11}\left(e(3+2e^2)E_n^{11} - 2e(1+4e^2)E_n^{13} - 2\gamma^2 E_n^{22}\right)^* - e(1-6e^2)\left|E_n^{13}\right|^2
$$

$$
+ 8\gamma^2\left(E_n^{02} + \frac{e}{2}(E_n^{11} - E_n^{13})\right)\left(E_n^{21} - E_n^{23} + \frac{e}{4}(E_n^{20} + 6E_n^{22} - E_n^{24})\right)^*
$$

$$
+ 4\gamma^2\left(E_n^{02} + \frac{e}{2}(E_n^{11} - E_n^{13}) - \gamma^2 E_n^{22}\right)\left(E_n^{02\prime} + \frac{e}{2}(E_n^{11} - E_n^{13})\prime\right)^*
$$

$$
\left. + 2\gamma^2 E_n^{13}\left(E_n^{22}\right)^* - \frac{4e}{3}S_n^{11}\left((1+4e^2)S_n^{11} + e\gamma^2 S_n^{11\prime}\right)\right],
$$

$$
D_e^{(1)} = V_e + \frac{3P\gamma^6}{20e^2}\sum_{n=1}^{\infty} nH_0^2\Omega_n\left[\frac{1}{e}\left|E_n^{02} + \frac{e}{2}(E_n^{11} - E_n^{13})\right|^2 + \frac{\gamma^4}{e}(S_n^{22})^2 - \frac{e^2}{e}S_n^{11\prime}S_n^{11}\right.
$$

$$
+ \left(E_n^{02} + \frac{e}{2}(E_n^{11} - E_n^{13}) - \gamma^2 E_n^{22}\right)\left(-\frac{2}{e}E_n^{02} + \frac{1}{2}(E_n^{11} + 3E_n^{13}) + 2(E_n^{21} - E_n^{23}) + E_n^{02\prime}\right.
$$

$$
\left.\left.- \gamma^2 E_n^{22\prime} + \frac{e}{2}(E_n^{20} + 12E_n^{22} - E_n^{24} + (E_n^{11} - E_n^{13})\prime)\right)^*\right],
$$

$$
D_{I'}^{(1)} = \frac{1}{2}D_{\Omega'\Omega'}^{(2)}, \qquad D_{\Omega'}^{(1)} = 0, \qquad D_{\omega'}^{(1)} = V_{\omega'}, \qquad D_\varepsilon^{(1)} = V_\varepsilon,
$$

(D.15)

while the nonzero components of the diffusion matrix are

$$D_{PP}^{(2)} = \frac{27P^3\gamma^2}{20} \sum_{n=1}^{\infty} nH_0^2\Omega_n \left[\left| E_n^{02} + \frac{e}{2}(E_n^{11} - E_n^{13}) \right|^2 - \frac{(eS_n^{11})^2}{3} \right],$$

$$D_{Pe}^{(2)} = \frac{\gamma^2 D_{PP}^{(2)}}{3Pe} - \frac{9P^2\gamma^6}{40} \sum_{n=1}^{\infty} nH_0^2\Omega_n E_n^{22} \left(\frac{2}{e}E_n^{02} + E_n^{11} - E_n^{13} \right)^*,$$

$$D_{ee}^{(2)} = \frac{3P\gamma^6}{20e^2} \sum_{n=1}^{\infty} nH_0^2\Omega_n \left[\left| E_n^{02} + \frac{e}{2}(E_n^{11} - E_n^{13}) - \gamma^2 E_n^{22} \right|^2 - \frac{(eS_n^{11})^2}{3} \right],$$

$$D_{I'I'}^{(2)} = \frac{3P\gamma^6}{80} \sum_{n=1}^{\infty} nH_0^2\Omega_n \left| E_n^{20} + E_n^{22} \right|^2, \qquad D_{\Omega'\Omega'}^{(2)} = -\sqrt{2}D_{\Omega'\omega'}^{(2)} = \frac{3P\gamma^6}{40} \sum_{n=1}^{\infty} nH_0^2\Omega_n \left| E_n^{20} - E_n^{22} \right|^2,$$

$$D_{\omega'\omega'}^{(2)} = \frac{1}{2}D_{\Omega'\Omega'}^{(2)} + \frac{3P\gamma^6}{80e^2} \sum_{n=1}^{\infty} nH_0^2\Omega_n \left[\left| E_n^{11} + E_n^{13} + 2E_n^{21} - 2E_n^{23} + \frac{e}{2}\left(E_n^{20} - E_n^{24} \right) \right|^2 + \frac{4}{3}\left(C_n^{11} \right)^2 \right],$$

$$D_{\omega'\varepsilon}^{(2)} = -\frac{3P\gamma^7}{80e^2} \sum_{n=1}^{\infty} nH_0^2\Omega_n \left[\left| E_n^{11} + E_n^{13} + 2(E_n^{21} - E_n^{23}) + \frac{e}{2}(E_n^{20} - 4E_n^{22} - E_n^{24}) \right|^2 \right.$$

$$\left. + \frac{4}{3}C_n^{11}(C_n^{11} - 2eC_n^{20}) - 4e^2|E_n^{22}|^2 \right],$$

$$D_{\varepsilon\varepsilon}^{(2)} = \frac{3P\gamma^8}{80e^2} \sum_{n=1}^{\infty} nH_0^2\Omega_n \left[\left| E_n^{11} + E_n^{13} + 2(E_n^{21} - E_n^{23}) + \frac{e}{2}(E_n^{20} - 8E_n^{22} - E_n^{24}) \right|^2 + \frac{4}{3}(C_n^{11} - 2eC_n^{20})^2 \right].$$

$$(\text{D.16})$$

These KM coefficients can be transformed back to a general reference frame using equations (D.13) and (D.14), resulting in the expressions given in section 3.3.

D.5 Hansen coefficients

Using various formulae given in Brumberg [157], we write the Hansen coefficients as

$$C_n^{lm}(e) = \frac{{}_2F_1(-m-l-1, m-l-1; 1; \beta^2)}{2\gamma^{2l}(1+\beta^2)^{l+1}} [J_{n-m}(ne) + J_{n+m}(ne)]$$

$$+ \sum_{k=1}^{\infty} \left\{ \frac{(m-l-1)_k \beta^k}{2k!\gamma^{2l}(1+\beta^2)^{l+1}} {}_2F_1(-m-l-1, k+m-l-1; k+1; \beta^2)[J_{n-m-k}(ne) + J_{n+m+k}(ne)] \right.$$

$$\left. + \frac{(-m-l-1)_k \beta^k}{2k!\gamma^{2l}(1+\beta^2)^{l+1}} {}_2F_1(k-l-m-1, -l+m-1; k+1; \beta^2)[J_{n-m+k}(ne) + J_{n+m-k}(ne)] \right\},$$

$$S_n^{lm}(e) = \frac{\mathrm{i}\,{}_2F_1(-m-l-1, m-l-1; 1; \beta^2)}{2\gamma^{2l}(1+\beta^2)^{l+1}} [J_{n-m}(ne) - J_{n+m}(ne)]$$

$$+ \sum_{k=1}^{\infty} \left\{ \frac{\mathrm{i}(m-l-1)_k \beta^k}{2k!\gamma^{2l}(1+\beta^2)^{l+1}} {}_2F_1(-m-l-1, k+m-l-1; k+1; \beta^2)[J_{n-m-k}(ne) - J_{n+m+k}(ne)] \right.$$

$$\left. + \frac{\mathrm{i}(-m-l-1)_k \beta^k}{2k!\gamma^{2l}(1+\beta^2)^{l+1}} {}_2F_1(k-m-l-1, m-l-1; k+1; \beta^2)[J_{n-m+k}(ne) - J_{n+m-k}(ne)] \right\},$$

$$(\text{D.17})$$

where we define the expansion variable $\beta \equiv e/(1 + \gamma)$, and where $(\cdots)_k$ is a rising Pochhammer symbol, defined by $(n)_k \equiv n(n + 1)(n + 2) \cdots (n + k - 1)$, while $_2F_1(a, b; c; z)$ is a hypergeometric function, and $J_n(z)$ is a Bessel function of the first kind. We see that in the circular case $e = 0$ we have $\beta = 0$, $_2F_1(a, b; c; 0) = 1$, and $J_n(0) = \delta_{n,0}$, and these expressions in equation (D.17) simplify to $C_n^{lm} = \frac{1}{2}(\delta_{n,m} + \delta_{n,-m})$ and $S_n^{lm} = \frac{i}{2}(\delta_{n,m} - \delta_{n,-m})$, which can be confirmed by directly integrating equation (D.8). Note also that for general eccentricity $e \in (0, 1)$, from the definition of the Pochhammer symbol, the sums over k in equation (D.17) terminate if and only if $m \le l+1$, otherwise the corresponding Hansen coefficients will have an infinite number of terms.

Using equation (D.17), we can directly compute all the Hansen coefficients that appear in equations (3.55) and (3.54), obtaining the cosine coefficients

$$C_n^{02} = -\left[\frac{1 + \gamma - \frac{2}{3}e^2(1 + \frac{1}{4}\gamma)}{(1 + \gamma)^4}\right]12\gamma^2 J_n + \sum_{k=0}^{\infty} \frac{\beta^k(1 - \beta^2)^3}{2(1 + \beta^2)}(J_{n-k-2} + J_{n+k+2}),$$

$$C_n^{11} = \frac{1}{2n\gamma^2}(J_{n-1} - J_{n+1}),$$

$$C_n^{13} = -\left[\frac{1 + \gamma - \frac{9}{8}e^2(1 + \frac{5}{9}\gamma) + \frac{1}{5}e^4}{\beta\gamma^2(1 + \gamma)^6}\right]20e^4 J_n + \sum_{k=0}^{\infty} \frac{\beta^k(1 - \beta^2)^5}{2(1 + \beta^2)^2\gamma^2}(J_{n-k-3} + J_{n+k+3})$$

$$- \left[\frac{1 - \gamma - \frac{9}{4}e^2(1 - \frac{7}{9}\gamma) + \frac{3}{4}e^4(1 - \frac{5}{3}\gamma)}{e^3\beta\gamma^2}\right]2(J_{n-1} + J_{n+1})$$

$$- \left[\frac{1 - \gamma - 3e^2(1 - \frac{5}{6}\gamma) + 3e^4(1 - \frac{5}{8}\gamma) - e^6(1 - \frac{15}{32}\gamma)}{e^5\gamma^2}\right]8(J_{n-2} + J_{n+2}),$$

$$C_n^{20} = -\left[\frac{n^2(1 - \frac{1}{2}e^2) - 3n + 2}{n^4 e^2\gamma^4}\right]8J_{n-2} + \left(\frac{n - 1}{n^3 e\gamma^4}\right)4J_{n-3},$$

$$C_n^{21} = \left[\frac{n^3(1 - \frac{7}{4}e^2 + \frac{3}{4}e^4) - n^2(1 - \frac{7}{2}e^2 + e^4) - 4n(1 + \frac{1}{2}e^2) + 4}{n^4 e^3\gamma^4}\right]4J_{n-2}$$

$$- \left[\frac{n^2(1 - \frac{3}{2}e^2 + \frac{1}{2}e^4) + n(1 + e^2) - 2}{n^3 e^2\gamma^4}\right]2J_{n-3},$$

$$C_n^{22} = -\left[\frac{n^3(1 - \frac{7}{4}e^2 + \frac{3}{4}e^4) - 2n^2(1 - \frac{13}{8}e^2 + \frac{7}{16}e^4) - n(1 + \frac{1}{2}e^2) + 2 - e^2}{n^4 e^4\gamma^4}\right]16J_{n-2}$$

$$+ \left[\frac{n^2(1 - \frac{3}{2}e^2 + \frac{1}{2}e^4) + \frac{1}{2}ne^2 - 1 + \frac{1}{2}e^2}{n^3 e^3\gamma^4}\right]8J_{n-3},$$

$$C_n^{23} = \left[\frac{n^2\gamma^6 + 3n(1 - \frac{5}{4}e^2 + \frac{1}{4}e^4) + 2 - \frac{3}{2}e^2}{n^2 e^3\gamma^4}\right]4J_n - \left(\frac{1 - \frac{5}{4}e^2 + \frac{1}{4}e^4}{n^2 e^2\gamma^4}\right)12J_{n-1},$$

$$C_n^{24} = \frac{8\gamma^3}{(1 + \gamma)^4}(J_{n-4} + J_{n+4}) + \sum_{k=0}^{\infty} \frac{\beta^k(1 + \gamma)^3}{16\gamma^4}\left[(1 - \beta^2)^7(J_{n-k-4} + J_{n+k+4})\right.$$

$$\left. + \frac{(-7)_k}{k!}\,_2F_1(1, k - 7; k + 1; \beta^2)(J_{n+k-4} + J_{n-k+4})\right],$$

$$\text{(D.18)}$$

and the sine coefficients,

$$
S_n^{02} = -\left[\frac{1 - \gamma - 2e^2(1 - \frac{3}{4}\gamma) + e^4}{e^3}\right] 2\mathrm{i}(J_{n-1} - J_{n+1}) + \sum_{k=0}^{\infty} \frac{\mathrm{i}\beta^k(1 - \beta^2)^3}{2(1 + \beta^2)}(J_{n-k-2} - J_{n+k+2}),
$$

$$
S_n^{11} = \frac{\mathrm{i}J_n}{ne\gamma},
$$

$$
S_n^{13} = -\left[\frac{1 - 3e^2(1 - \frac{5}{6}\gamma) + 3e^4(1 - \frac{5}{8}\gamma) - e^6}{e^4\gamma^2}\right] 4\mathrm{i}(J_{n-1} - J_{n+1})
$$
$$
- \left[\frac{1 - \gamma - 3e^2(1 - \frac{5}{6}\gamma) + 3e^4(1 - \frac{5}{8}\gamma) - e^6(1 - \frac{15}{32}\gamma)}{e^5\gamma^2}\right] 8\mathrm{i}(J_{n-2} - J_{n+2})
$$
$$
+ \sum_{k=0}^{\infty} \frac{\mathrm{i}\beta^k(1 - \beta^2)^5}{2(1 + \beta^2)^2\gamma^2}(J_{n-k-3} - J_{n+k+3}),
$$

$$
S_n^{21} = \frac{-\mathrm{i}}{n^2 e\gamma^3}\left[(n\gamma^2 + 2)J_n - 2eJ_{n-1}\right],
$$

$$
S_n^{22} = \left[\frac{n^3(1 - \frac{5}{4}e^2 + \frac{1}{4}e^4) - 2n^2(1 - \frac{9}{8}e^2) - n(1 + e^2) + 2}{n^4 e^4\gamma^3}\right] 16\mathrm{i}J_{n-2} - \left[\frac{n^3\gamma^2 + n^2 e^2 - n}{n^4 e^3\gamma^3}\right] 8\mathrm{i}J_{n-3},
$$

$$
S_n^{23} = \frac{15e^2}{8\gamma^3}\mathrm{i}(J_{n-1} - J_{n+1}) - \frac{3e}{2\gamma^3}\mathrm{i}(J_{n-2} - J_{n+2}) + \frac{1 - \frac{1}{4}e^2}{2\gamma^3}\mathrm{i}(J_{n-3} - J_{n+3}),
$$

$$
S_n^{24} = \frac{8\mathrm{i}\gamma^3}{(1 + \gamma)^4}(J_{n-4} - J_{n+4})
$$
$$
+ \sum_{k=0}^{\infty} \frac{\beta^k(1 + \gamma)^3}{16\gamma^4}\left[(1 - \beta^2)^7(J_{n-k-4} - J_{n+k+4}) + \frac{(-7)_k}{k!}{}_2F_1(1, k - 7; k + 1; \beta^2)(J_{n+k-4} - J_{n-k+4})\right].
$$

$$\tag{D.19}$$

Note that all the Bessel functions appearing in equations (D.18) and (D.19) have their argument equal to ne, even if they are not of order n; we suppress this argument for brevity.

Acronyms

AGWB	Astrophysical gravitational-wave background
AION	Atom Interferometer Observatory and Network
BAO	Baryon acoustic oscillations
BBH	Binary black hole
BBN	Big-bang nucleosynthesis
BH	Black hole
BHNS	Black hole-neutron star
BNS	Binary neutron star
CBC	Compact binary coalescence
CDM	Cold dark matter
CE	Cosmic Explorer
CGF	Cumulant-generating function
CP	Common process
DM	Dark matter
EFE	Einstein field equation
EM	Electromagnetic
EMRI	Extreme mass-ratio inspiral
EoM	Equation of motion
EPTA	European Pulsar Timing Array
ET	Einstein Telescope
FLRW	Friedmann-Lemaître-Robertson-Walker
FOPT	First-order phase transition
FPE	Fokker-Planck equation
GR	General relativity
GUT	Grand unified theory
GW	Gravitational wave
GWB	Gravitational-wave background
GWTC	Gravitational-wave transient catalogue
i.i.d.	Independent and identically distributed
IFO	Interferometer
IPTA	International Pulsar Timing Array
ISCO	Innermost stable circular orbit
ISW	Integrated Sachs-Wolfe

KAGRA	Kamioka Gravitational-Wave Detector
KM	Kramers-Moyal
LHC	Large Hadron Collider
LIGO	Laser Interferometer Gravitational-Wave Observatory
LISA	Laser Interferometer Space Antenna
LLR	Lunar laser ranging
LR	Laser ranging
LSS	Large-scale structure
LVK	LIGO/Virgo/KAGRA
MAGIS	Matter-wave Atomic Gradiometer Interferometric Sensor
MHD	Magnetohydrodynamic
MSP	Millisecond pulsar
MVUE	Minimum-variance unbiased estimator
NANOGrav	North American Nanohertz Observatory for Gravitational Waves
NS	Neutron star
ODE	Ordinary differential equation
ORF	Overlap reduction function
PBH	Primordial black hole
PDE	Partial differential equation
PDF	Probability density function
PI	Power-law integrated
PN	Post-Newtonian
PPTA	Parkes Pulsar Timing Array
PSD	Power spectral density
PTA	Pulsar timing array
QNM	Quasinormal mode
rms	Root-mean-square
SFR	Star formation rate
SGWB	Stochastic gravitational-wave background
SHC	Spherical harmonic component
SKA	Square Kilometre Array
SLR	Satellite laser ranging
SNR	Signal-to-noise ratio
SW	Sachs-Wolfe
ToA	Time of arrival
TT	Transverse-traceless
VEV	Vacuum expectation value
WD	White dwarf
ZFL	Zero-frequency limit

Bibliography

'There's more to life than books you know,
but not much more.'

Morrissey

[1] J. Aasi, *et al.* (LIGO, Virgo), *Searching for stochastic gravitational waves using data from the two colocated LIGO Hanford detectors*, Phys. Rev. D **91** (2015), 022003, arXiv:1410.6211 [gr-qc]

[2] J. Aasi, *et al.* (LIGO), *Advanced LIGO*, Class. Quant. Grav. **32** (2015), 074001, arXiv:1411.4547 [gr-qc]

[3] B. Abbott, *et al.* (LIGO), *Search for gravitational waves from primordial black hole binary coalescences in the galactic halo*, Phys. Rev. D **72** (2005), 082002, arXiv:gr-qc/0505042

[4] B. P. Abbott, *et al.* (LIGO), *First LIGO search for gravitational wave bursts from cosmic (super)strings*, Phys. Rev. D **80** (2009), 062002, arXiv:0904.4718 [astro-ph.CO]

[5] B. P. Abbott, *et al.* (LIGO, Virgo), *Observation of Gravitational Waves from a Binary Black Hole Merger*, Phys. Rev. Lett. **116** (2016), 061102, arXiv:1602.03837 [gr-qc]

[6] B. P. Abbott, *et al.* (LIGO, Virgo), *Tests of general relativity with GW150914*, Phys. Rev. Lett. **116** (2016), 221101, [Erratum: Phys. Rev. Lett. 121, 129902 (2018)], arXiv:1602.03841 [gr-qc]

[7] B. P. Abbott, *et al.* (LIGO, Virgo), *GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence*, Phys. Rev. Lett. **116** (2016), 241103, arXiv:1606.04855 [gr-qc]

[8] B. P. Abbott, *et al.* (LIGO, Virgo), *Upper Limits on the Stochastic Gravitational-Wave Background from Advanced LIGO's First Observing Run*, Phys. Rev. Lett. **118** (2017), 121101, [Erratum: Phys.Rev.Lett. 119, 029901 (2017)], arXiv:1612.02029 [gr-qc]

[9] B. P. Abbott, *et al.* (LIGO, Virgo), *Directional Limits on Persistent Gravitational Waves from Advanced LIGO's First Observing Run*, Phys. Rev. Lett. **118** (2017), 121102, arXiv:1612.02030 [gr-qc]

[10] B. P. Abbott, *et al.* (LIGO, Virgo), *GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2*, Phys. Rev. Lett. **118** (2017), 221101, [Erratum: Phys.Rev.Lett. 121, 129901 (2018)], arXiv:1706.01812 [gr-qc]

[11] B. P. Abbott, *et al.* (LIGO, Virgo), *GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence*, Phys. Rev. Lett. **119** (2017), 141101, arXiv:1709.09660 [gr-qc]

[12] B. P. Abbott, *et al.* (LIGO, Virgo), *GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral*, Phys. Rev. Lett. **119** (2017), 161101, arXiv:1710.05832 [gr-qc]

[13] B. P. Abbott, *et al.* (LIGO, Virgo, Fermi GBM, INTEGRAL, IceCube, AstroSat Cadmium Zinc Telluride Imager Team, IPN, Insight-Hxmt, ANTARES, Swift, AGILE Team, 1M2H Team, Dark Energy Camera GW-EM, DES, DLT40, GRAWITA, Fermi-LAT, ATCA, ASKAP, Las Cumbres Observatory Group, OzGrav, DWF (Deeper Wider Faster Program), AST3, CAASTRO, VIN-ROUGE, MASTER, J-GEM, GROWTH, JAGWAR, CaltechNRAO, TTU-NRAO, NuSTAR, Pan-STARRS, MAXI Team, TZAC Consortium, KU, Nordic Optical Telescope, ePESSTO, GROND, Texas Tech University, SALT Group, TOROS, BOOTES, MWA, CALET, IKI-GW Follow-up, H.E.S.S., LOFAR, LWA, HAWC, Pierre Auger, ALMA, Euro VLBI Team, Pi of Sky, Chandra Team at McGill University, DFN, ATLAS Telescopes, High Time Resolution Universe Survey, RIMAS, RATIR, SKA South Africa/MeerKAT), *Multi-messenger Observations of a Binary Neutron Star Merger*, Astrophys. J. Lett. **848** (2017), L12, arXiv:1710.05833 [astro-ph.HE]

[14] B. P. Abbott, *et al.* (LIGO, Virgo, Fermi-GBM, INTEGRAL), *Gravitational Waves and Gamma-rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A*, Astrophys. J. Lett. **848** (2017), L13, arXiv:1710.05834 [astro-ph.HE]

[15] B. P. Abbott, *et al.* (LIGO, Virgo, 1M2H, Dark Energy Camera GW-E, DES, DLT40, Las Cumbres Observatory, VINROUGE, MASTER), *A gravitational-wave standard siren measurement of the Hubble constant*, Nature **551** (2017), 85, arXiv:1710.05835 [astro-ph.CO]

[16] B. P. Abbott, *et al.* (LIGO, Virgo), *GW170608: Observation of a 19-solar-mass Binary Black Hole Coalescence*, Astrophys. J. Lett. **851** (2017), L35, arXiv:1711.05578 [astro-ph.HE]

[17] B. P. Abbott, *et al.* (LIGO, Virgo), *GW170817: Implications for the Stochastic Gravitational-Wave Background from Compact Binary Coalescences*, Phys. Rev. Lett. **120** (2018), 091101, arXiv:1710.05837 [gr-qc]

[18] B. P. Abbott, *et al.* (LIGO, Virgo), *Constraints on cosmic strings using data from the first Advanced LIGO observing run*, Phys. Rev. D **97** (2018), 102002, arXiv:1712.01168 [gr-qc]

[19] B. P. Abbott, *et al.* (LIGO, Virgo), *Search for Tensor, Vector, and Scalar Polarizations in the Stochastic Gravitational-Wave Background*, Phys. Rev. Lett. **120** (2018), 201102, arXiv:1802.10194 [gr-qc]

[20] B. P. Abbott, *et al.* (LIGO, Virgo), *GW170817: Measurements of neutron star radii and equation of state*, Phys. Rev. Lett. **121** (2018), 161101, arXiv:1805.11581 [gr-qc]

[21] B. P. Abbott, *et al.* (LIGO, Virgo), *Search for Subsolar-Mass Ultracompact Binaries in Advanced LIGO's First Observing Run*, Phys. Rev. Lett. **121** (2018), 231103, arXiv:1808.04771 [astro-ph.CO]

[22] B. P. Abbott, *et al.* (LIGO, Virgo), *GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs*, Phys. Rev. X **9** (2019), 031040, arXiv:1811.12907 [astro-ph.HE]

[23] B. P. Abbott, *et al.* (LIGO, Virgo), *Binary Black Hole Population Properties Inferred from the First and Second Observing Runs of Advanced LIGO and Advanced Virgo*, Astrophys. J. Lett. **882** (2019), L24, arXiv:1811.12940 [astro-ph.HE]

[24] B. P. Abbott, *et al.* (LIGO, Virgo), *Search for the isotropic stochastic background using data from Advanced LIGO's second observing run*, Phys. Rev. D **100** (2019), 061101, arXiv:1903.02886 [gr-qc]

[25] B. P. Abbott, *et al.* (LIGO, Virgo), *Directional limits on persistent gravitational waves using data from Advanced LIGO's first two observing runs*, Phys. Rev. D **100** (2019), 062001, arXiv:1903.08844 [gr-qc]

[26] B. P. Abbott, *et al.* (LIGO, Virgo), *Search for Subsolar Mass Ultracompact Binaries in Advanced LIGO's Second Observing Run*, Phys. Rev. Lett. **123** (2019), 161102, arXiv:1904.08976 [astro-ph.CO]

[27] B. P. Abbott, *et al.* (LIGO, Virgo, KAGRA), *Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA*, Living Rev. Rel. **23** (2020), 3, arXiv:1304.0670 [gr-qc]

[28] B. P. Abbott, *et al.* (LIGO, Virgo), *GW190425: Observation of a Compact Binary Coalescence with Total Mass ∼ 3.4$M_\odot$*, Astrophys. J. Lett. **892** (2020), L3, arXiv:2001.01761 [astro-ph.HE]

[29] R. Abbott, *et al.* (LIGO, Virgo), *GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object*, Astrophys. J. Lett. **896** (2020), L44, arXiv:2006.12611 [astro-ph.HE]

[30] R. Abbott, *et al.* (LIGO, Virgo), *GW190521: A Binary Black Hole Merger with a Total Mass of 150$M_\odot$*, Phys. Rev. Lett. **125** (2020), 101102, arXiv:2009.01075 [gr-qc]

[31] R. Abbott, *et al.* (LIGO, Virgo), *Properties and Astrophysical Implications of the 150 $M_\odot$ Binary Black Hole Merger GW190521*, Astrophys. J. Lett. **900** (2020), L13, arXiv:2009.01190 [astro-ph.HE]

[32] R. Abbott, *et al.* (LIGO, Virgo, KAGRA), *Data for "Upper Limits on the Isotropic Gravitational-Wave Background from Advanced LIGO's and Advanced Virgo's Third Observing Run"*, `https://dcc.ligo.org/LIGO-G2001287/public` (2021)

[33] R. Abbott, *et al.* (LIGO, Virgo), *GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run*, Phys. Rev. X **11** (2021), 021053, arXiv:2010.14527 [gr-qc]

[34] R. Abbott, *et al.* (LIGO, Virgo), *Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog*, Phys. Rev. D **103** (2021), 122002, arXiv:2010.14529 [gr-qc]

[35] R. Abbott, *et al.* (LIGO, Virgo), *Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog*, Astrophys. J. Lett. **913** (2021), L7, arXiv:2010.14533 [astro-ph.HE]

[36] R. Abbott, *et al.* (LIGO, Virgo, KAGRA), *Upper limits on the isotropic gravitational-wave background from Advanced LIGO and Advanced Virgo's third observing run*, Phys. Rev. D **104** (2021), 022004, arXiv:2101.12130 [gr-qc]

[37] R. Abbott, *et al.* (LIGO, Virgo, KAGRA), *Constraints on Cosmic Strings Using Data from the Third Advanced LIGO–Virgo Observing Run*, Phys. Rev. Lett. **126** (2021), 241102, arXiv:2101.12248 [gr-qc]

[38] R. Abbott, *et al.* (LIGO, Virgo, KAGRA), *Search for anisotropic gravitational-wave backgrounds using data from Advanced LIGO's and Advanced Virgo's first three observing runs*, Phys. Rev. D **104** (2021), 022005, arXiv:2103.08520 [gr-qc]

[39] R. Abbott, *et al.* (LIGO, Virgo, KAGRA), *Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences*, Astrophys. J. Lett. **915** (2021), L5, arXiv:2106.15163 [astro-ph.HE]

[40] R. Abbott, *et al.* (LIGO, Virgo), *GWTC-2.1: Deep Extended Catalog of Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run* (2021), arXiv:2108.01045 [gr-qc]

[41] R. Abbott, *et al.* (LIGO, VIRGO, KAGRA), *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run* (2021), arXiv:2111.03606 [gr-qc]

[42] A. A. Abdo, *et al.* (Fermi-LAT), *Detection of 16 Gamma-Ray Pulsars Through Blind Frequency Searches Using the Fermi LAT*, Science **325** (2009), 840, arXiv:1009.0748 [astro-ph.GA]

[43] M. Abe, P. Adamson, M. Borcean, D. Bortoletto, K. Bridges, S. P. Carman, S. Chattopadhyay, J. Coleman, N. M. Curfman, K. DeRose, T. Deshpande, S. Dimopoulos, C. J. Foot, J. C. Frisch, B. E. Garber, S. Geer, V. Gibson, J. Glick, P. W. Graham, S. R. Hahn, R. Harnik, L. Hawkins, S. Hindley, J. M. Hogan, Y. Jiang, M. A. Kasevich, R. J. Kellett, M. Kiburg, T. Kovachy, J. D. Lykken, J. March-Russell, J. Mitchell, M. Murphy, M. Nantel, L. E. Nobrega, R. K. Plunkett, S. Rajendran, J. Rudolph, N. Sachdeva, M. Safdari, J. K. Santucci, A. G. Schwartzman, I. Shipsey, H. Swan, L. R. Valerio, A. Vasonis, Y. Wang, and T. Wilkason, *Matter-wave Atomic Gradiometer Interferometric Sensor (MAGIS-100)*, Quantum Sci. Technol. **6** (2021), 044003, arXiv:2104.02835 [physics.atom-ph]

[44] F. Acernese, *et al.* (Virgo), *Advanced Virgo: a second-generation interferometric gravitational wave detector*, Class. Quant. Grav. **32** (2015), 024001, arXiv:1408.3978 [gr-qc]

[45] A. Achucarro, R. Gregory, and K. Kuijken, *Abelian Higgs hair for black holes*, Phys. Rev. D **52** (1995), 5729, arXiv:gr-qc/9505039

[46] M. R. Adams and N. J. Cornish, *Discriminating between a Stochastic Gravitational Wave Background and Instrument Noise*, Phys. Rev. D **82** (2010), 022002, arXiv:1002.1291 [gr-qc]

[47] P. A. R. Ade, *et al.* (Planck), *Planck 2013 results. XXV. Searches for cosmic strings and other topological defects*, Astron. Astrophys. **571** (2014), A25, arXiv:1303.5085 [astro-ph.CO]

[48] P. A. R. Ade, *et al.* (BICEP2, Planck), *Joint Analysis of BICEP2/Keck Array and Planck Data*, Phys. Rev. Lett. **114** (2015), 101301, arXiv:1502.00612 [astro-ph.CO]

[49] S. Adhikari, M. Fishbach, D. E. Holz, R. H. Wechsler, and Z. Fang, *The Binary–Host Connection: Astrophysics of Gravitational-Wave Binaries from Host Galaxy Properties*, Astrophys. J. **905** (2020), 21, arXiv:2001.01025 [astro-ph.GA]

[50] P. Adshead, N. Afshordi, E. Dimastrogiovanni, M. Fasiello, E. A. Lim, and G. Tasinato, *Multimessenger cosmology: Correlating cosmic microwave background and stochastic gravitational wave background measurements*, Phys. Rev. D **103** (2021), 023532, arXiv:2004.06619 [astro-ph.CO]

[51] P. Adshead, E. Martinec, and M. Wyman, *Gauge fields and inflation: Chiral gravitational waves, fluctuations, and the Lyth bound*, Phys. Rev. D **88** (2013), 021302, arXiv:1301.2598 [hep-th]

[52] P. Adshead, E. Martinec, and M. Wyman, *Perturbations in Chromo-Natural Inflation*, JHEP **09** (2013), 087, arXiv:1305.2930 [hep-th]

[53] K. Aggarwal, Z. Arzoumanian, P. T. Baker, A. Brazier, P. R. Brook, S. Burke-Spolaor, S. Chatterjee, J. M. Cordes, N. J. Cornish, F. Crawford, H. T. Cromartie, K. Crowter, M. DeCesar, P. B. Demorest, T. Dolch, J. A. Ellis, R. D. Ferdman, E. C. Ferrara, E. Fonseca, N. Garver-Daniels, P. Gentile, D. Good, J. S. Hazboun, A. M. Holgado, E. A. Huerta, K. Islo, R. Jennings, G. Jones, M. L. Jones, D. L. Kaplan, L. Z. Kelley, J. S. Key, M. T. Lam, T. J. W. Lazio, L. Levin, D. R. Lorimer, J. Luo, R. S. Lynch, D. R. Madison, M. A. McLaughlin, S. T. McWilliams, C. M. F. Mingarelli, C. Ng, D. J. Nice, T. T. Pennucci, N. S. Pol, S. M. Ransom, P. S. Ray, X. Siemens, J. Simon, R. Spiewak, I. H. Stairs, D. R. Stinebring, K. Stovall, J. K. Swiggum, S. R. Taylor, M. Vallisneri, R. van Haasteren, S. J. Vigeland, C. A. Witt, and W. W. Zhu (NANOGrav), *The NANOGrav 11 yr Data Set: Limits on Gravitational Wave Memory*, Astrophys. J. **889** (2020), 38, arXiv:1911.08488 [astro-ph.HE]

[54] N. Aghanim, *et al.* (Planck), *Planck 2018 results. I. Overview and the cosmological legacy of Planck*, Astron. Astrophys. **641** (2020), A1, arXiv:1807.06205 [astro-ph.CO]

[55] N. Aghanim, *et al.* (Planck), *Planck 2018 results. VI. Cosmological parameters*, Astron. Astrophys. **641** (2020), A6, arXiv:1807.06209 [astro-ph.CO]

[56] O. D. Aguiar, *The Past, Present and Future of the Resonant-Mass Gravitational Wave Detectors*, Res. Astron. Astrophys. **11** (2011), 1, arXiv:1009.1138 [astro-ph.IM]

[57] A. Ain, J. Suresh, and S. Mitra, *Very fast stochastic gravitational wave background map making using folded data*, Phys. Rev. D **98** (2018), 024001, arXiv:1803.08285 [gr-qc]

[58] P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan, A. M. Sintes, J. T. Whelan, B. Brügmann, P. Diener, N. Dorband, J. Gonzalez, M. Hannam, S. Husa, D. Pollney, L. Rezzolla, L. Santamaría, U. Sperhake, and J. Thornburg, *A Template bank for gravitational waveforms from coalescing binary black holes. I. Non-spinning binaries*, Phys. Rev. D **77** (2008), 104017, [Erratum: Phys.Rev.D 79, 129901 (2009)], arXiv:0710.2335 [gr-qc]

[59] P. Ajith, M. Hannam, S. Husa, Y. Chen, B. Brügmann, N. Dorband, D. Müller, F. Ohme, D. Pollney, C. Reisswig, L. Santamaría, and J. Seiler, *Inspiral-merger-ringdown waveforms for black-hole binaries with non-precessing spins*, Phys. Rev. Lett. **106** (2011), 241101, arXiv:0909.2867 [gr-qc]

[60] S. Akiba, M. Nakada, C. Yamaguchi, and K. Iwamoto, *Gravitational Wave Memory from the Relativistic Jet of Gamma-Ray Bursts*, Publ. Astron. Soc. Jap. **65** (2013), 59, arXiv:1303.6566 [astro-ph.HE]

[61] T. Akutsu, *et al.* (KAGRA), *KAGRA: 2.5 Generation Interferometric Gravitational Wave Detector*, Nature Astron. **3** (2019), 35, arXiv:1811.08079 [gr-qc]

[62] V. Alba and J. Maldacena, *Primordial gravity wave background anisotropies*, JHEP **03** (2016), 115, arXiv:1512.01531 [hep-th]

[63] S. Alexander and N. Yunes, *Chern-Simons Modified General Relativity*, Phys. Rept. **480** (2009), 1, arXiv:0907.2562 [hep-th]

[64] S. H.-S. Alexander, M. E. Peskin, and M. M. Sheikh-Jabbari, *Leptogenesis from gravity waves in models of inflation*, Phys. Rev. Lett. **96** (2006), 081301, arXiv:hep-th/0403069

[65] Y. Ali-Haïmoud, T. L. Smith, and C. M. F. Mingarelli, *Fisher formalism for anisotropic gravitational-wave background searches with pulsar timing arrays*, Phys. Rev. D **102** (2020), 122005, arXiv:2006.14570 [gr-qc]

[66] Y. Ali-Haïmoud, T. L. Smith, and C. M. F. Mingarelli, *Insights into searches for anisotropies in the nanohertz gravitational-wave background*, Phys. Rev. D **103** (2021), 042009, arXiv:2010.13958 [gr-qc]

[67] B. Allen, *The Stochastic gravity wave background: Sources and detection*, in *Les Houches School of Physics: Astrophysical Sources of Gravitational Radiation* (1996), arXiv:gr-qc/9604033

[68] B. Allen, *Gravitational wave stochastic background from cosmological particle decay*, Phys. Rev. Res. **2** (2020), 012034, arXiv:1910.08213 [gr-qc]

[69] B. Allen and J. D. Romano, *Detecting a stochastic background of gravitational radiation: Signal processing strategies and sensitivities*, Phys. Rev. D **59** (1999), 102001, arXiv:gr-qc/9710117

[70] D. Alonso, C. R. Contaldi, G. Cusin, P. G. Ferreira, and A. I. Renzini, *Noise angular power spectrum of gravitational wave background experiments*, Phys. Rev. D **101** (2020), 124048, arXiv:2005.03001 [astro-ph.CO]

[71] D. Alonso, G. Cusin, P. G. Ferreira, and C. Pitrou, *Detecting the anisotropic astrophysical gravitational wave background in the presence of shot noise through cross-correlations*, Phys. Rev. D **102** (2020), 023002, arXiv:2002.02888 [astro-ph.CO]

[72] P. Amaro-Seoane, J. R. Gair, M. Freitag, M. Coleman Miller, I. Mandel, C. J. Cutler, and S. Babak, *Astrophysics, detection and science applications of intermediate- and extreme mass-ratio inspirals*, Class. Quant. Grav. **24** (2007), R113, arXiv:astro-ph/0703495

[73] P. Amaro-Seoane, *et al.* (LISA Consortium), *Laser Interferometer Space Antenna* (2017), arXiv:1702.00786 [astro-ph.IM]

[74] K. N. Ananda, C. Clarkson, and D. Wands, *The Cosmological gravitational wave background from primordial density perturbations*, Phys. Rev. D **75** (2007), 123518, arXiv:gr-qc/0612013

[75] L. Annulli, L. Bernard, D. Blas, and V. Cardoso, *Scattering of scalar, electromagnetic and gravitational waves from binary systems*, Phys. Rev. D **98** (2018), 084001, arXiv:1809.05108 [gr-qc]

[76] J. Antoniadis, *et al.*, *The International Pulsar Timing Array second data release: Search for an isotropic Gravitational Wave Background*, Mon. Not. Roy. Astron. Soc. (2022), 10.1093/mnras/stab3418, arXiv:2201.03980 [astro-ph.HE]

[77] F. Antonini and F. A. Rasio, *Merging black hole binaries in galactic nuclei: implications for advanced-LIGO detections*, Astrophys. J. **831** (2016), 187, arXiv:1606.04889 [astro-ph.HE]

[78] S. Aoyama, D. Yamauchi, M. Shiraishi, and M. Ouchi, *Gaia 400,894 QSO constraint on the energy density of low-frequency gravitational waves* (2021), arXiv:2105.04039 [gr-qc]

[79] A. Arbey, J. Auffinger, and J. Silk, *Evolution of primordial black hole spin due to Hawking radiation*, Mon. Not. Roy. Astron. Soc. **494** (2020), 1257, arXiv:1906.04196 [astro-ph.CO]

[80] J. M. Armaleo, D. López Nacir, and F. R. Urban, *Binary pulsars as probes for spin-2 ultralight dark matter*, JCAP **01** (2020), 053, arXiv:1909.13814 [astro-ph.HE]

[81] J. M. Armaleo, D. López Nacir, and F. R. Urban, *Pulsar timing array constraints on spin-2 ULDM*, JCAP **09** (2020), 031, arXiv:2005.03731 [astro-ph.CO]

[82] J. W. Armstrong, L. Iess, P. Tortora, and B. Bertotti, *Stochastic gravitational wave background: Upper limits in the* 10^{-6} Hz–10^{-3} Hz *band*, Astrophys. J. **599** (2003), 806

[83] A. Arvanitaki and S. Dubovsky, *Exploring the String Axiverse with Precision Black Hole Physics*, Phys. Rev. D **83** (2011), 044026, arXiv:1004.3558 [hep-th]

[84] M. Aryal, L. H. Ford, and A. Vilenkin, *Cosmic Strings and Black Holes*, Phys. Rev. D **34** (1986), 2263

[85] Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, S. J. Chamberlin, S. Chatterjee, B. Christy, J. M. Cordes, N. J. Cornish, P. B. Demorest, X. Deng, T. Dolch, J. A. Ellis, R. D. Ferdman, E. Fonseca, N. Garver-Daniels, F. Jenet, G. Jones, V. M. Kaspi, M. Koop, M. T. Lam, T. J. W. Lazio, L. Levin, A. N. Lommen, D. R. Lorimer, J. Luo, R. S. Lynch, D. R. Madison, M. A. McLaughlin, S. T. McWilliams, D. J. Nice, N. Palliyaguru, T. T. Pennucci, S. M. Ransom, X. Siemens, I. H. Stairs, D. R. Stinebring, K. Stovall, J. Swiggum, M. Vallisneri, R. van Haasteren, Y. Wang, and W. W. Zhu (NANOGrav), *NANOGrav Constraints on Gravitational Wave Bursts with Memory*, Astrophys. J. **810** (2015), 150, arXiv:1501.05343 [astro-ph.GA]

[86] Z. Arzoumanian, *et al.* (NANOGrav), *The NANOGrav 12.5 yr Data Set: Search for an Isotropic Stochastic Gravitational-wave Background*, Astrophys. J. Lett. **905** (2020), L34, arXiv:2009.04496 [astro-ph.HE]

[87] P. Auclair, J. J. Blanco-Pillado, D. G. Figueroa, A. C. Jenkins, M. Lewicki, M. Sakellariadou, S. Sanidas, L. Sousa, D. A. Steer, J. M. Wachter, and S. Kuroyanagi, *Probing the gravitational wave background from cosmic strings with LISA*, JCAP **04** (2020), 034, arXiv:1909.00819 [astro-ph.CO]

[88] J. C. Aurrekoetxea, T. Helfer, and E. A. Lim, *Coherent Gravitational Waveforms and Memory from Cosmic String Loops*, Class. Quant. Grav. **37** (2020), 204001, arXiv:2002.05177 [gr-qc]

[89] S. Babak, J. Gair, A. Sesana, E. Barausse, C. F. Sopuerta, C. P. L. Berry, E. Berti, P. Amaro-Seoane, A. Petiteau, and A. Klein, *Science with the space-based interferometer LISA. V: Extreme mass-ratio inspirals*, Phys. Rev. D **95** (2017), 103012, arXiv:1703.09722 [gr-qc]

[90] D. C. Backer, S. R. Kulkarni, C. Heiles, M. M. Davis, and W. M. Goss, *A millisecond pulsar*, Nature **300** (1982), 615

[91] L. Badurina, E. Bentine, D. Blas, K. Bongs, D. Bortoletto, T. Bowcock, K. Bridges, W. Bowden, O. Buchmuller, C. Burrage, J. Coleman, G. Elertas, J. Ellis, C. Foot, V. Gibson, M. G. Haehnelt, T. Harte, S. Hedges, R. Hobson, M. Holynski, T. Jones, M. Langlois, S. Lellouch, M. Lewicki, R. Maiolino, P. Majewski, S. Malik, J. March-Russell, C. McCabe, D. Newbold, B. Sauer, U. Schneider, I. Shipsey, Y. Singh, M. A. Uchida, T. Valenzuela, M. van der Grinten, V. Vaskonen, J. Vossebeld, D. Weatherill, and I. Wilmut, *AION: An Atom Interferometer Observatory and Network*, JCAP **05** (2020), 011, arXiv:1911.11755 [astro-ph.CO]

[92] L. Baiotti and L. Rezzolla, *Binary neutron star mergers: a review of Einstein's richest laboratory*, Rept. Prog. Phys. **80** (2017), 096901, arXiv:1607.03540 [gr-qc]

[93] T. Baker, E. Bellini, P. G. Ferreira, M. Lagos, J. Noller, and I. Sawicki, *Strong constraints on cosmological gravity from GW170817 and GRB 170817A*, Phys. Rev. Lett. **119** (2017), 251301, arXiv:1710.06394 [astro-ph.CO]

[94] A. Ball, M. Pate, A.-M. Raclariu, A. Strominger, and R. Venugopalan, *Measuring color memory in a color glass condensate at electron–ion colliders*, Annals Phys. **407** (2019), 15, arXiv:1805.12224 [hep-ph]

[95] S. Banagiri, A. Criswell, T. Kuan, V. Mandic, J. D. Romano, and S. R. Taylor, *Mapping the gravitational-wave sky with LISA: a Bayesian spherical harmonic approach*, Mon. Not. Roy. Astron. Soc. **507** (2021), 5451, arXiv:2103.00826 [astro-ph.IM]

[96] L. Barack and A. Pound, *Self-force and radiation reaction in general relativity*, Rept. Prog. Phys. **82** (2019), 016904, arXiv:1805.10385 [gr-qc]

[97] J. M. Bardeen, *Gauge Invariant Cosmological Perturbations*, Phys. Rev. D **22** (1980), 1882

[98] N. Barnaby, E. Pajer, and M. Peloso, *Gauge Field Production in Axion Inflation: Consequences for Monodromy, non-Gaussianity in the CMB, and Gravitational Waves at Interferometers*, Phys. Rev. D **85** (2012), 023525, arXiv:1110.3327 [astro-ph.CO]

[99] E. D. Barr, P. C. C. Freire, M. Kramer, D. J. Champion, M. Berezina, C. G. Bassa, A. G. Lyne, and B. W. Stappers, *A Massive Millisecond Pulsar in an Eccentric Binary*, Mon. Not. Roy. Astron. Soc. **465** (2017), 1711, arXiv:1611.03658 [astro-ph.HE]

[100] N. Bartolo, D. Bertacca, R. Caldwell, C. R. Contaldi, G. Cusin, V. De Luca, E. Dimastrogiovanni, M. Fasiello, D. G. Figueroa, G. Franciolini, A. C. Jenkins, M. Peloso, M. Pieroni, A. Renzini, A. Ricciardone, A. Riotto, M. Sakellariadou, L. Sorbo, G. Tasinato, J. Torrado, S. Clesse, and S. Kuroyanagi, *Probing Anisotropies of the Stochastic Gravitational Wave Background with LISA* (2022), arXiv:2201.08782 [astro-ph.CO]

[101] N. Bartolo, D. Bertacca, V. De Luca, G. Franciolini, S. Matarrese, M. Peloso, A. Ricciardone, A. Riotto, and G. Tasinato, *Gravitational wave anisotropies from primordial black holes*, JCAP **02** (2020), 028, arXiv:1909.12619 [astro-ph.CO]

[102] N. Bartolo, D. Bertacca, S. Matarrese, M. Peloso, A. Ricciardone, A. Riotto, and G. Tasinato, *Anisotropies and non-Gaussianity of the Cosmological Gravitational Wave Background*, Phys. Rev. D **100** (2019), 121501, arXiv:1908.00527 [astro-ph.CO]

[103] N. Bartolo, D. Bertacca, S. Matarrese, M. Peloso, A. Ricciardone, A. Riotto, and G. Tasinato, *Characterizing the cosmological gravitational wave background: Anisotropies and non-Gaussianity*, Phys. Rev. D **102** (2020), 023527, arXiv:1912.09433 [astro-ph.CO]

[104] N. Bartolo, C. Caprini, V. Domcke, D. G. Figueroa, J. García-Bellido, M. C. Guzzetti, M. Liguori, S. Matarrese, M. Peloso, A. Petiteau, A. Ricciardone, M. Sakellariadou, L. Sorbo, and G. Tasinato, *Science with the space-based interferometer LISA. IV: Probing inflation with gravitational waves*, JCAP **12** (2016), 026, arXiv:1610.06481 [astro-ph.CO]

[105] N. Bartolo and G. Orlando, *Parity breaking signatures from a Chern-Simons coupling during inflation: the case of non-Gaussian gravitational waves*, JCAP **07** (2017), 034, arXiv:1706.04627 [astro-ph.CO]

[106] I. Bartos, B. Kocsis, Z. Haiman, and S. Márka, *Rapid and Bright Stellar-mass Binary Black Hole Mergers in Active Galactic Nuclei*, Astrophys. J. **835** (2017), 165, arXiv:1602.03831 [astro-ph.HE]

[107] R. A. Battye and E. P. S. Shellard, *String radiative back reaction*, Phys. Rev. Lett. **75** (1995), 4354, arXiv:astro-ph/9408078

[108] J. D. Bekenstein, *Black hole hair: 25 - years after* , in *2nd International Sakharov Conference on Physics* (1996), 216–219, arXiv:gr-qc/9605059

[109] K. Belczynski, D. E. Holz, T. Bulik, and R. O'Shaughnessy, *The first gravitational-wave source from the isolated evolution of two 40-100 Msun stars*, Nature **534** (2016), 512, arXiv:1602.04531 [astro-ph.HE]

[110] K. Belczynski, J. Klencki, C. E. Fields, A. Olejak, E. Berti, G. Meynet, C. L. Fryer, D. E. Holz, R. O'Shaughnessy, D. A. Brown, T. Bulik, S. C. Leung, K. Nomoto, P. Madau, R. Hirschi, E. Kaiser, S. Jones, S. Mondal, M. Chruslinska, P. Drozda, D. Gerosa, Z. Doctor, M. Giersz, S. Ekstrom, C. Georgy, A. Askar, V. Baibhav, D. Wysocki, T. Natan, W. M. Farr, G. Wiktorowicz, M. Coleman Miller, B. Farr, and J.-P. Lasota, *Evolutionary roads leading to low effective spins, high black hole masses, and O1/O2 rates for LIGO/Virgo binary black holes*, Astron. Astrophys. **636** (2020), A104, arXiv:1706.07053 [astro-ph.HE]

[111] N. Bellomo, D. Bertacca, A. C. Jenkins, S. Matarrese, A. Raccanelli, T. Regimbau, A. Ricciardone, and M. Sakellariadou, `CLASS_GWB`: *robust modeling of the astrophysical gravitational wave background anisotropies* (2021), arXiv:2110.15059 [gr-qc]

[112] C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, N. Odegard, K. M. Smith, R. S. Hill, B. Gold, M. Halpern, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, J. Dunkley, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright (WMAP), *Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results*, Astrophys. J. Suppl. **208** (2013), 20, arXiv:1212.5225 [astro-ph.CO]

[113] D. P. Bennett and F. R. Bouchet, *Evidence for a Scaling Solution in Cosmic String Evolution*, Phys. Rev. Lett. **60** (1988), 257

[114] V. Berezinsky and A. Vilenkin, *Cosmic necklaces and ultrahigh-energy cosmic rays*, Phys. Rev. Lett. **79** (1997), 5202, arXiv:astro-ph/9704257

[115] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro, *Large scale structure of the universe and cosmological perturbation theory*, Phys. Rept. **367** (2002), 1, arXiv:astro-ph/0112551

[116] P. de Bernardis, P. A. R. Ade, J. J. Bock, J. R. Bond, J. Borrill, A. Boscaleri, K. Coble, B. P. Crill, G. De Gasperis, P. C. Farese, P. G. Ferreira, K. Ganga, M. Giacometti, E. Hivon, V. V. Hristov, A. Iacoangeli, A. H. Jaffe, A. E. Lange, L. Martinis, S. Masi, P. V. Mason, P. D. Mauskopf, A. Melchiorri, L. Miglio, T. Montroy, C. B. Netterfield, E. Pascale, F. Piacentini, D. Pogosyan, S. Prunet, S. Rao, G. Romeo, J. E. Ruhl, F. Scaramuzzi, D. Sforna, and N. Vittorio (Boomerang), *A Flat universe from high resolution maps of the cosmic microwave background radiation*, Nature **404** (2000), 955, arXiv:astro-ph/0004404

[117] D. Bertacca, A. Ricciardone, N. Bellomo, A. C. Jenkins, S. Matarrese, A. Raccanelli, T. Regimbau, and M. Sakellariadou, *Projection effects on the observed angular spectrum of the astrophysical stochastic gravitational wave background*, Phys. Rev. D **101** (2020), 103513, arXiv:1909.11627 [astro-ph.CO]

[118] E. Berti, E. Barausse, V. Cardoso, L. Gualtieri, P. Pani, U. Sperhake, L. C. Stein, N. Wex, K. Yagi, T. Baker, C. P. Burgess, F. S. Coelho, D. Doneva, A. D. Felice, P. G. Ferreira, P. C. C. Freire, J. Healy, C. Herdeiro, M. Horbatsch, B. Kleihaus, A. Klein, K. Kokkotas, J. Kunz, P. Laguna, R. N. Lang, T. G. F. Li, T. Littenberg, A. Matas, S. Mirshekari, H. Okawa, E. Radu, R. O'Shaughnessy, B. S. Sathyaprakash, C. V. D. Broeck, H. A. Winther, H. Witek, M. E. Aghili, J. Alsing, B. Bolen, L. Bombelli, S. Caudill, L. Chen, J. C. Degollado, R. Fujita, C. Gao, D. Gerosa, S. Kamali, H. O. Silva, J. G. Rosa, L. Sadeghian, M. Sampaio, H. Sotani, and M. Zilhao, *Testing General Relativity with Present and Future Astrophysical Observations*, Class. Quant. Grav. **32** (2015), 243001, arXiv:1501.07274 [gr-qc]

[119] E. Berti, V. Cardoso, and A. O. Starinets, *Quasinormal modes of black holes and black branes*, Class. Quant. Grav. **26** (2009), 163001, arXiv:0905.2975 [gr-qc]

[120] E. Berti, V. Cardoso, and C. M. Will, *On gravitational-wave spectroscopy of massive black holes with the space interferometer LISA*, Phys. Rev. D **73** (2006), 064030, arXiv:gr-qc/0512160

[121] G. Bertone, D. Hooper, and J. Silk, *Particle dark matter: Evidence, candidates and constraints*, Phys. Rept. **405** (2005), 279, arXiv:hep-ph/0404175

[122] B. Bertotti, *Is the Solar System Gravitationally Closed?*, Astrophys. Lett. **14** (1973), 51, https://ui.adsabs.harvard.edu/abs/1973ApL....14...51B

[123] L. Bian, R.-G. Cai, J. Liu, X.-Y. Yang, and R. Zhou, *Evidence for different gravitational-wave sources in the NANOGrav dataset*, Phys. Rev. D **103** (2021), L081301, arXiv:2009.13893 [astro-ph.CO]

[124] L. Bieri and D. Garfinkle, *An electromagnetic analogue of gravitational wave memory*, Class. Quant. Grav. **30** (2013), 195009, arXiv:1307.5098 [gr-qc]

[125] J. Binney and S. Tremaine, *Galactic Dynamics*, 2 ed. (Princeton University Press, 2008)

[126] S. Bird, I. Cholis, J. B. Muñoz, Y. Ali-Haïmoud, M. Kamionkowski, E. D. Kovetz, A. Raccanelli, and A. G. Riess, *Did LIGO detect dark matter?*, Phys. Rev. Lett. **116** (2016), 201301, arXiv:1603.00464 [astro-ph.CO]

[127] O. Birnholtz and T. Piran, *Gravitational wave memory from gamma ray bursts' jets*, Phys. Rev. D **87** (2013), 123007, arXiv:1302.5713 [astro-ph.HE]

[128] J. Blaizot, Y. Wadadekar, B. Guiderdoni, S. T. Colombi, E. Bertin, F. R. Bouchet, J. E. G. Devriendt, and S. Hatton, *MoMaf: The Mock Map Facility*, Mon. Not. Roy. Astron. Soc. **360** (2005), 159, arXiv:astro-ph/0309305

[129] L. Blanchet and T. Damour, *Hereditary effects in gravitational radiation*, Phys. Rev. D **46** (1992), 4304

[130] J. J. Blanco-Pillado, K. D. Olum, and B. Shlaer, *The number of cosmic string loops*, Phys. Rev. D **89** (2014), 023512, arXiv:1309.6637 [astro-ph.CO]

[131] J. J. Blanco-Pillado, K. D. Olum, and X. Siemens, *New limits on cosmic strings from gravitational wave observation*, Phys. Lett. B **778** (2018), 392, arXiv:1709.02434 [astro-ph.CO]

[132] J. J. Blanco-Pillado, K. D. Olum, and A. Vilenkin, *No black holes from cosmic string cusps* (2021), arXiv:2101.05040 [astro-ph.CO]

[133] J. J. Blanco-Pillado, K. D. Olum, and J. M. Wachter, *Gravitational backreaction near cosmic string kinks and cusps*, Phys. Rev. D **98** (2018), 123507, arXiv:1808.08254 [gr-qc]

[134] J. J. Blanco-Pillado, K. D. Olum, and J. M. Wachter, *Gravitational backreaction simulations of simple cosmic string loops*, Phys. Rev. D **100** (2019), 023535, arXiv:1903.06079 [gr-qc]

[135] J. J. Blanco-Pillado, K. D. Olum, and J. M. Wachter, *Comparison of cosmic string and superstring models to NANOGrav 12.5-year results*, Phys. Rev. D **103** (2021), 103512, arXiv:2102.08194 [astro-ph.CO]

[136] R. Blandford and S. A. Teukolsky, *Arrival-time analysis for a pulsar in a binary system*, Astrophys. J. **205** (1976), 580

[137] D. Blas and A. C. Jenkins, *Bridging the μHz gap in the gravitational-wave landscape with binary resonance*, Phys. Rev. Lett. **128** (2021), 101103, arXiv:2107.04601 [astro-ph.CO]

[138] D. Blas and A. C. Jenkins, *Detecting stochastic gravitational waves with binary resonance*, Phys. Rev. D **105** (2022), 064021, arXiv:2107.04063 [gr-qc]

[139] D. Blas, D. López Nacir, and S. Sibiryakov, *Secular effects of ultralight dark matter on binary pulsars*, Phys. Rev. D **101** (2020), 063016, arXiv:1910.08544 [gr-qc]

[140] D. Blas, D. L. Nacir, and S. Sibiryakov, *Ultralight Dark Matter Resonates with Binary Pulsars*, Phys. Rev. Lett. **118** (2017), 261102, arXiv:1612.06789 [hep-ph]

[141] S. Blasi, V. Brdar, and K. Schmitz, *Has NANOGrav found first evidence for cosmic strings?*, Phys. Rev. Lett. **126** (2021), 041305, arXiv:2009.06607 [astro-ph.CO]

[142] O. M. Boersma, D. A. Nichols, and P. Schmidt, *Forecasts for detecting the gravitational-wave memory effect with Advanced LIGO and Virgo*, Phys. Rev. D **101** (2020), 083026, arXiv:2002.01821 [astro-ph.HE]

[143] A. Bohé, L. Shao, A. Taracchini, A. Buonanno, S. Babak, I. W. Harry, I. Hinder, S. Ossokine, M. Pürrer, V. Raymond, T. Chu, H. Fong, P. Kumar, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, M. A. Scheel, and B. Szilágyi, *Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors*, Phys. Rev. D **95** (2017), 044028, arXiv:1611.03703 [gr-qc]

[144] G. Boileau, N. Christensen, R. Meyer, and N. J. Cornish, *Spectral separation of the stochastic gravitational-wave background for LISA: Observing both cosmological and astrophysical backgrounds,* Phys. Rev. D **103** (2021), 103529, arXiv:2011.05055 [gr-qc]

[145] G. Boileau, A. C. Jenkins, M. Sakellariadou, R. Meyer, and N. Christensen, *Ability of LISA to detect a gravitational-wave background of cosmological origin: The cosmic string case,* Phys. Rev. D **105** (2022), 023510, arXiv:2109.06552 [gr-qc]

[146] G. Boileau, A. Lamberts, N. Christensen, N. J. Cornish, and R. Meyer, *Spectral separation of the stochastic gravitational-wave background for LISA in the context of a modulated Galactic foreground,* Mon. Not. Roy. Astron. Soc. **508** (2021), 803, arXiv:2105.04283 [gr-qc]

[147] H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner, *Gravitational waves in general relativity. 7. Waves from axisymmetric isolated systems,* Proc. Roy. Soc. Lond. A **269** (1962), 21

[148] F. Bonjour, R. Emparan, and R. Gregory, *Vortices and extreme black holes: The Question of flux expulsion,* Phys. Rev. D **59** (1999), 084022, arXiv:gr-qc/9810061

[149] R. J. Bontz and R. H. Price, *The spectrum of radiation at low frequencies,* Astrophys. J. **228** (1979), 560

[150] M. Boudaud and M. Cirelli, *Voyager 1 $e^\pm$ Further Constrain Primordial Black Holes as Dark Matter,* Phys. Rev. Lett. **122** (2019), 041104, arXiv:1807.03075 [astro-ph.HE]

[151] V. B. Braginsky and L. P. Grishchuk, *Kinematic Resonance and Memory Effect in Free Mass Gravitational Antennas,* Sov. Phys. JETP **62** (1985), 427, http://www.jetp.ac.ru/cgi-bin/e/index/e/62/3/p427?a=list

[152] V. B. Braginsky and K. S. Thorne, *Gravitational-wave bursts with memory and experimental prospects,* Nature **327** (1987), 123

[153] M. Braglia and S. Kuroyanagi, *Probing prerecombination physics by the cross-correlation of stochastic gravitational waves and CMB anisotropies,* Phys. Rev. D **104** (2021), 123547, arXiv:2106.03786 [astro-ph.CO]

[154] L. W. Brenneman, C. S. Reynolds, M. A. Nowak, R. C. Reis, M. Trippe, A. C. Fabian, K. Iwasawa, J. C. Lee, J. M. Miller, R. F. Mushotzky, K. Nandra, and M. Volonteri, *The Spin of the Supermassive Black Hole in NGC 3783,* Astrophys. J. **736** (2011), 103, arXiv:1104.1172 [astro-ph.HE]

[155] R. Brito, S. Ghosh, E. Barausse, E. Berti, V. Cardoso, I. Dvorkin, A. Klein, and P. Pani, *Gravitational wave searches for ultralight bosons with LIGO and LISA,* Phys. Rev. D **96** (2017), 064050, arXiv:1706.06311 [gr-qc]

[156] D. Brouwer and G. M. Clemence, *Methods of celestial mechanics* (Academic Press, New York, 1961)

[157] V. A. Brumberg, *Analytical Techniques of Celestial Mechanics* (Springer-Verlag, Berlin, 1995)

[158] W. Buchmuller, V. Domcke, and K. Schmitz, *From NANOGrav to LIGO with metastable cosmic strings*, Phys. Lett. B **811** (2020), 135914, arXiv:2009.10649 [astro-ph.CO]

[159] A. Buonanno and T. Damour, *Gravitational, dilatonic and axionic radiative damping of cosmic strings*, Phys. Rev. D **60** (1999), 023517, arXiv:gr-qc/9801105

[160] L. M. Burko and G. Khanna, *Climbing up the memory staircase: Equatorial zoom-whirl orbits*, Phys. Rev. D **102** (2020), 084035, arXiv:2007.12545 [gr-qc]

[161] J. A. Burns, *Elementary derivation of the perturbation equations of celestial mechanics*, Am. J. Phys. **44** (1976), 944

[162] A. Burrows and J. Hayes, *Pulsar recoil and gravitational radiation due to asymmetrical stellar collapse and explosion*, Phys. Rev. Lett. **76** (1996), 352, arXiv:astro-ph/9511106

[163] M. J. Bustamante-Rosell, J. Meyers, N. Pearson, C. Trendafilova, and A. Zimmerman, *Gravitational wave timing array*, Phys. Rev. D **105** (2022), 044005, arXiv:2107.02788 [gr-qc]

[164] R. R. Caldwell and P. Casper, *Formation of black holes from collapsed cosmic string loops*, Phys. Rev. D **53** (1996), 3002, arXiv:gr-qc/9509012

[165] R. R. Caldwell and E. Gates, *Constraints on cosmic strings due to black holes formed from collapsed cosmic string loops*, Phys. Rev. D **48** (1993), 2581, arXiv:hep-ph/9306221

[166] T. Callister, A. S. Biscoveanu, N. Christensen, M. Isi, A. Matas, O. Minazzoli, T. Regimbau, M. Sakellariadou, J. Tasson, and E. Thrane, *Polarization-based Tests of Gravity with the Stochastic Gravitational-Wave Background*, Phys. Rev. X **7** (2017), 041058, arXiv:1704.08373 [gr-qc]

[167] M. Campanelli, C. O. Lousto, Y. Zlochower, and D. Merritt, *Maximum gravitational recoil*, Phys. Rev. Lett. **98** (2007), 231102, arXiv:gr-qc/0702133

[168] A. Campoleoni, D. Francia, and C. Heissenberg, *Electromagnetic and color memory in even dimensions*, Phys. Rev. D **100** (2019), 085015, arXiv:1907.05187 [hep-th]

[169] C. Caprini, M. Chala, G. C. Dorsch, M. Hindmarsh, S. J. Huber, T. Konstandin, J. Kozaczuk, G. Nardini, J. M. No, K. Rummukainen, P. Schwaller, G. Servant, A. Tranberg, and D. J. Weir, *Detecting gravitational waves from cosmological phase transitions with LISA: an update*, JCAP **03** (2020), 024, arXiv:1910.13125 [astro-ph.CO]

[170] C. Caprini and D. G. Figueroa, *Cosmological Backgrounds of Gravitational Waves*, Class. Quant. Grav. **35** (2018), 163001, arXiv:1801.04268 [astro-ph.CO]

[171] C. Caprini, D. G. Figueroa, R. Flauger, G. Nardini, M. Peloso, M. Pieroni, A. Ricciardone, and G. Tasinato, *Reconstructing the spectral shape of a stochastic gravitational wave background with LISA*, JCAP **11** (2019), 017, arXiv:1906.09244 [astro-ph.CO]

[172] C. Caprini, M. Hindmarsh, S. Huber, T. Konstandin, J. Kozaczuk, G. Nardini, J. M. No, A. Petiteau, P. Schwaller, G. Servant, and D. J. Weir, *Science with the space-based interferometer eLISA. II: Gravitational waves from cosmological phase transitions*, JCAP **04** (2016), 001, arXiv:1512.06239 [astro-ph.CO]

[173] C. Caprini and L. Sorbo, *Adding helicity to inflationary magnetogenesis*, JCAP **10** (2014), 056, arXiv:1407.2809 [astro-ph.CO]

[174] G. Capurri, A. Lapi, C. Baccigalupi, L. Boco, G. Scelfo, and T. Ronconi, *Intensity and anisotropies of the stochastic gravitational wave background from merging compact binaries in galaxies*, JCAP **11** (2021), 032, arXiv:2103.12037 [gr-qc]

[175] R. Carballo-Rubio, F. Di Filippo, S. Liberati, and M. Visser, *Causal hierarchy in modified gravity*, JHEP **12** (2020), 055, arXiv:2005.08533 [gr-qc]

[176] B. Carr, K. Kohri, Y. Sendouda, and J. Yokoyama, *Constraints on primordial black holes*, Rept. Prog. Phys. **84** (2021), 116902, arXiv:2002.12778 [astro-ph.CO]

[177] B. Carr and F. Kühnel, *Primordial Black Holes as Dark Matter: Recent Developments*, Ann. Rev. Nucl. Part. Sci. **70** (2020), 355, arXiv:2006.02838 [astro-ph.CO]

[178] B. Carr, F. Kühnel, and M. Sandstad, *Primordial Black Holes as Dark Matter*, Phys. Rev. D **94** (2016), 083504, arXiv:1607.06077 [astro-ph.CO]

[179] B. Carr, T. Tenkanen, and V. Vaskonen, *Primordial black holes from inflaton and spectator field perturbations in a matter-dominated era*, Phys. Rev. D **96** (2017), 063507, arXiv:1706.03746 [astro-ph.CO]

[180] B. J. Carr, *The Primordial black hole mass spectrum*, Astrophys. J. **201** (1975), 1

[181] B. J. Carr and S. W. Hawking, *Black holes in the early Universe* , Mon. Not. Roy. Astron. Soc. **168** (1974), 399

[182] B. J. Carr, K. Kohri, Y. Sendouda, and J. Yokoyama, *New cosmological constraints on primordial black holes*, Phys. Rev. D **81** (2010), 104019, arXiv:0912.5297 [astro-ph.CO]

[183] B. Carter and R. A. Battye, *Nondivergence of gravitational selfinteractions for Goto-Nambu strings*, Phys. Lett. B **430** (1998), 49, arXiv:hep-th/9803012

[184] G. Casella and R. L. Berger, *Statistical Inference*, 2 ed. (Duxbury, 2002)

[185] G. P. Centers, J. W. Blanchard, J. Conrad, N. L. Figueroa, A. Garcon, A. V. Gramolin, D. F. Jackson Kimball, M. Lawson, B. Pelssers, J. A. Smiga, A. O. Sushkov, A. Wickenbrock, D. Budker, and A. Derevianko, *Stochastic fluctuations of bosonic dark matter*, Nature Commun. **12** (2021), 7321, arXiv:1905.13650 [astro-ph.CO]

[186] J. L. Cervantes-Cota, S. Galindo-Uribarri, and G.-F. Smoot, *A Brief History of Gravitational Waves*, Universe **2** (2016), 22, arXiv:1609.09400 [physics.hist-ph]

[187] A. Chamblin, J. M. A. Ashbourn-Chamblin, R. Emparan, and A. Sornborger, *Can extreme black holes have (long) Abelian Higgs hair?*, Phys. Rev. D **58** (1998), 124014, arXiv:gr-qc/9706004

[188] G. F. Chapline, *Cosmological effects of primordial black holes*, Nature **253** (1975), 251

[189] S. Chen, B. Wang, and R. Su, *Quasinormal modes and late-time tails in the background of Schwarzschild black hole pierced by a cosmic string: Scalar, electromagnetic and gravitational perturbations*, Int. J. Mod. Phys. A **23** (2008), 2505, arXiv:gr-qc/0701088

[190] S. Chen, *et al.*, *Common-red-signal analysis with 24-yr high-precision timing of the European Pulsar Timing Array: Inferences in the stochastic gravitational-wave background search*, Mon. Not. Roy. Astron. Soc. **508** (2021), 4970, arXiv:2110.13184 [astro-ph.HE]

[191] D. F. Chernoff, É. É. Flanagan, and B. Wardell, *Gravitational backreaction on a cosmic string: Formalism*, Phys. Rev. D **99** (2019), 084036, arXiv:1808.08631 [gr-qc]

[192] M. H.-Y. Cheung, L. W.-H. Poon, A. K.-W. Chung, and T. G. F. Li, *Ringdown spectroscopy of rotating black holes pierced by cosmic strings*, JCAP **02** (2021), 040, arXiv:2002.01695 [gr-qc]

[193] T. Chiba and S. Yokoyama, *Spin Distribution of Primordial Black Holes*, PTEP **2017** (2017), 083E01, arXiv:1704.06573 [gr-qc]

[194] C. Chicone, B. Mashhoon, and D. G. Retzloff, *Gravitational ionization: Periodic orbits of binary systems perturbed by gravitational radiation* , Ann. Inst. H. Poincare Phys. Theor. **64** (1996), 87, arXiv:gr-qc/9508065

[195] C. Chicone, B. Mashhoon, and D. G. Retzloff, *On the ionization of a Keplerian binary system by periodic gravitational radiation*, J. Math. Phys. **37** (1996), 3997, [Erratum: J.Math.Phys. 38, 544 (1997)], arXiv:gr-qc/9605031

[196] C. Chicone, B. Mashhoon, and D. G. Retzloff, *Gravitational ionization: A Chaotic net in the Kepler system*, Class. Quant. Grav. **14** (1997), 699, arXiv:gr-qc/9608054

[197] C. Chicone, B. Mashhoon, and D. G. Retzloff, *Sustained resonance: A Binary system perturbed by gravitational radiation*, J. Phys. A **33** (2000), 513, arXiv:gr-qc/9911123

[198] R. Chornock, E. Berger, D. Kasen, P. S. Cowperthwaite, M. Nicholl, V. A. Villar, K. D. Alexander, P. K. Blanchard, T. Eftekhari, W. Fong, R. Margutti, P. K. G. Williams, J. Annis, D. Brout, D. A. Brown, H.-Y. Chen, M. R. Drout, B. Farr, R. J. Foley, F. A. Frieman, C. L. Fryer, K. Herner, D. E. Holz, R. Kessler, T. Matheson, B. D. Metzger, E. Quataert, A. Rest, M. Sako, D. M. Scolnic, N. Smith, and M. Soares-Santos, *The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. IV. Detection of Near-infrared Signatures of r-process Nucleosynthesis with Gemini-South*, Astrophys. J. Lett. **848** (2017), L19, arXiv:1710.05454 [astro-ph.HE]

[199] N. Christensen, *Measuring the stochastic gravitational radiation background with laser interferometric antennas*, Phys. Rev. D **46** (1992), 5250

[200] N. Christensen, *Stochastic Gravitational Wave Backgrounds*, Rept. Prog. Phys. **82** (2019), 016903, arXiv:1811.08797 [gr-qc]

[201] D. Christodoulou, *Nonlinear nature of gravitation and gravitational wave experiments*, Phys. Rev. Lett. **67** (1991), 1486

[202] D. Christodoulou and S. Klainerman, *The global nonlinear stability of the Minkowski space* (Princeton University Press, 1993)

[203] I. Ciufolini, A. Paolozzi, E. C. Pavlis, R. Koenig, J. Ries, V. Gurzadyan, R. Matzner, R. Penrose, G. Sindoni, C. Paris, H. Khachatryan, and S. Mirzoyan, *A test of general relativity using the LARES and LAGEOS satellites and a GRACE Earth gravity model*, Eur. Phys. J. C **76** (2016), 120, arXiv:1603.09674 [gr-qc]

[204] S. Clark, B. Dutta, Y. Gao, Y.-Z. Ma, and L. E. Strigari, *21 cm limits on decaying dark matter and primordial black holes*, Phys. Rev. D **98** (2018), 043006, arXiv:1803.09390 [astro-ph.HE]

[205] S. Clark, B. Dutta, Y. Gao, L. E. Strigari, and S. Watson, *Planck Constraint on Relic Primordial Black Holes*, Phys. Rev. D **95** (2017), 083006, arXiv:1612.07738 [astro-ph.CO]

[206] S. Clesse and J. García-Bellido, *The clustering of massive Primordial Black Holes as Dark Matter: measuring their mass distribution with Advanced LIGO*, Phys. Dark Univ. **15** (2017), 142, arXiv:1603.05234 [astro-ph.CO]

[207] S. Clesse and J. García-Bellido, *Detecting the gravitational wave background from primordial black hole dark matter*, Phys. Dark Univ. **18** (2017), 105, arXiv:1610.08479 [astro-ph.CO]

[208] H. Collins, *Gravity's shadow: The search for gravitational waves*, 2 ed. (University of Chicago Press, 2004)

[209] C. Conneely, A. H. Jaffe, and C. M. F. Mingarelli, *On the Amplitude and Stokes Parameters of a Stochastic Gravitational-Wave Background*, Mon. Not. Roy. Astron. Soc. **487** (2019), 562, arXiv:1808.05920 [astro-ph.CO]

[210] C. R. Contaldi, *Anisotropies of Gravitational Wave Backgrounds: A Line Of Sight Approach*, Phys. Lett. B **771** (2017), 9, arXiv:1609.08168 [astro-ph.CO]

[211] C. R. Contaldi, J. Magueijo, and L. Smolin, *Anomalous CMB polarization and gravitational chirality*, Phys. Rev. Lett. **101** (2008), 141101, arXiv:0806.3082 [astro-ph]

[212] C. R. Contaldi, M. Pieroni, A. I. Renzini, G. Cusin, N. Karnesis, M. Peloso, A. Ricciardone, and G. Tasinato, *Maximum likelihood map-making with the Laser Interferometer Space Antenna*, Phys. Rev. D **102** (2020), 043502, arXiv:2006.03313 [astro-ph.CO]

[213] A. Cooray and R. K. Sheth, *Halo Models of Large Scale Structure*, Phys. Rept. **372** (2002), 1, arXiv:astro-ph/0206508

[214] E. J. Copeland, D. Haws, and M. Hindmarsh, *Classical theory of radiating strings*, Phys. Rev. D **42** (1990), 726

[215] J. M. Cordes and F. A. Jenet, *Detecting gravitational wave memory with pulsar timing*, Astrophys. J. **752** (2012), 54

[216] N. J. Cornish and J. D. Romano, *When is a gravitational-wave signal stochastic?*, Phys. Rev. D **92** (2015), 042001, arXiv:1505.08084 [gr-qc]

[217] M. Coughlin and J. Harms, *Constraining the gravitational wave energy density of the Universe using Earth's ring*, Phys. Rev. D **90** (2014), 042005, arXiv:1406.1147 [gr-qc]

[218] J. D. E. Creighton and W. G. Anderson, *Gravitational-wave physics and astronomy: An introduction to theory, experiment and data analysis*, Wiley Series in Cosmology (Wiley, 2011)

[219] P. Creminelli and F. Vernizzi, *Dark Energy after GW170817 and GRB170817A*, Phys. Rev. Lett. **119** (2017), 251302, arXiv:1710.05877 [astro-ph.CO]

[220] D. Croon, D. McKeen, N. Raj, and Z. Wang, *Subaru-HSC through a different lens: Microlensing by extended dark matter structures*, Phys. Rev. D **102** (2020), 083021, arXiv:2007.12697 [astro-ph.CO]

[221] G. Cusin, I. Dvorkin, C. Pitrou, and J.-P. Uzan, *First predictions of the angular power spectrum of the astrophysical gravitational wave background*, Phys. Rev. Lett. **120** (2018), 231101, arXiv:1803.03236 [astro-ph.CO]

[222] G. Cusin, I. Dvorkin, C. Pitrou, and J.-P. Uzan, *Properties of the stochastic astrophysical gravitational wave background: astrophysical sources dependencies*, Phys. Rev. D **100** (2019), 063004, arXiv:1904.07797 [astro-ph.CO]

[223] G. Cusin, I. Dvorkin, C. Pitrou, and J.-P. Uzan, *Stochastic gravitational wave background anisotropies in the mHz band: astrophysical dependencies*, Mon. Not. Roy. Astron. Soc. **493** (2020), L1, arXiv:1904.07757 [astro-ph.CO]

[224] G. Cusin, C. Pitrou, and J.-P. Uzan, *Anisotropy of the astrophysical gravitational wave background: Analytic expression of the angular power spectrum and correlation with cosmological observations*, Phys. Rev. D **96** (2017), 103019, arXiv:1704.06184 [astro-ph.CO]

[225] G. Cusin, C. Pitrou, and J.-P. Uzan, *The signal of the gravitational wave background and the angular correlation of its energy density*, Phys. Rev. D **97** (2018), 123527, arXiv:1711.11345 [astro-ph.CO]

[226] T. Damour and N. Deruelle, *General relativistic celestial mechanics of binary systems II. The post-Newtonian timing formula*, Ann. Inst. Henri Poincaré Phys. Théor. **44** (1986), 263, http://adsabs.harvard.edu/abs/1986AIHS...44..263D

[227] T. Damour and A. Vilenkin, *Gravitational wave bursts from cusps and kinks on cosmic strings*, Phys. Rev. D **64** (2001), 064008, arXiv:gr-qc/0104026

[228] J. Darling, A. E. Truebenbach, and J. Paine, *Astrometric Limits on the Stochastic Gravitational Wave Background*, Astrophys. J. **861** (2018), 113, arXiv:1804.06986 [astro-ph.IM]

[229] V. De Luca, V. Desjacques, G. Franciolini, A. Malhotra, and A. Riotto, *The initial spin probability distribution of primordial black holes*, JCAP **05** (2019), 018, arXiv:1903.01179 [astro-ph.CO]

[230] V. De Luca, G. Franciolini, P. Pani, and A. Riotto, *The Evolution of Primordial Black Holes and their Final Observable Spins*, JCAP **04** (2020), 052, arXiv:2003.02778 [astro-ph.CO]

[231] V. De Luca, G. Franciolini, P. Pani, and A. Riotto, *Primordial Black Holes Confront LIGO/Virgo data: Current situation*, JCAP **06** (2020), 044, arXiv:2005.05641 [astro-ph.CO]

[232] V. De Luca, G. Franciolini, and A. Riotto, *NANOGrav Data Hints at Primordial Black Holes as Dark Matter*, Phys. Rev. Lett. **126** (2021), 041303, arXiv:2009.08268 [astro-ph.CO]

[233] G. De Lucia and J. Blaizot, *The hierarchical formation of the brightest cluster galaxies*, Mon. Not. Roy. Astron. Soc. **375** (2007), 2, arXiv:astro-ph/0606519

[234] M. H. Dehghani, A. M. Ghezelbash, and R. B. Mann, *Abelian Higgs hair for AdS - Schwarzschild black hole*, Phys. Rev. D **65** (2002), 044010, arXiv:hep-th/0107224

[235] W. DeRocco and P. W. Graham, *Constraining Primordial Black Hole Abundance with the Galactic 511 keV Line*, Phys. Rev. Lett. **123** (2019), 251102, arXiv:1906.07740 [astro-ph.CO]

[236] V. Desjacques, E. Grishin, and Y. B. Ginat, *Axion oscillations in binary systems: angle-action surgery*, Astrophys. J. **901** (2020), 85, arXiv:2003.10552 [gr-qc]

[237] S. L. Detweiler, *Pulsar timing measurements and the search for gravitational waves*, Astrophys. J. **234** (1979), 1100

[238] R. H. Dicke, P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson, *Cosmic Black-Body Radiation*, Astrophys. J. **142** (1965), 414

[239] E. Dimastrogiovanni, M. Fasiello, and T. Fujita, *Primordial Gravitational Waves from Axion-Gauge Fields Dynamics*, JCAP **01** (2017), 019, arXiv:1608.04216 [astro-ph.CO]

[240] R. d'Inverno, *Introducing Einstein's relativity* (Clarendon Press, Oxford, 1992)

[241] A. K. Divakarla, E. Thrane, P. D. Lasky, and B. F. Whiting, *Memory Effect or Cosmic String? Classifying Gravitational-Wave Bursts with Bayesian Inference*, Phys. Rev. D **102** (2020), 023010, arXiv:1911.07998 [gr-qc]

[242] V. Domcke, J. García-Bellido, M. Peloso, M. Pieroni, A. Ricciardone, L. Sorbo, and G. Tasinato, *Measuring the net circular polarization of the stochastic gravitational wave background with interferometers*, JCAP **05** (2020), 028, arXiv:1910.08052 [astro-ph.CO]

[243] J. A. Dror, H. Murayama, and N. L. Rodd, *Cosmic axion background*, Phys. Rev. D **103** (2021), 115004, arXiv:2101.09287 [hep-ph]

[244] M. R. Drout, A. L. Piro, B. J. Shappee, C. D. Kilpatrick, J. D. Simon, C. Contreras, D. A. Coulter, R. J. Foley, M. R. Siebert, N. Morrell, K. Boutsia, F. Di Mille, T. W.-S. Holoien, D. Kasen, J. A. Kollmeier, B. F. Madore, A. J. Monson, A. Murguia-Berthier, Y.-C. Pan, J. X. Prochaska, E. Ramirez-Ruiz, A. Rest, C. Adams, K. Alatalo, E. Bañados, J. Baughman, T. C. Beers, R. A. Bernstein, T. Bitsakis, A. Campillay, T. T. Hansen, C. R. Higgs, A. P. Ji, G. Maravelias, J. L. Marshall, C. Moni Bidin, J. L. Prieto, K. C. Rasmussen, C. Rojas-Bravo, A. L. Strom, N. Ulloa, J. Vargas-González, Z. Wan, and D. D. Whitten, *Light Curves of the Neutron Star Merger GW170817/SSS17a: Implications for R-Process Nucleosynthesis*, Science **358** (2017), 1570, arXiv:1710.05443 [astro-ph.HE]

[245] G. Duchêne and A. Kraus, *Stellar Multiplicity*, Ann. Rev. Astron. Astrophys. **51** (2013), 269, arXiv:1303.3028 [astro-ph.SR]

[246] R. Durrer, *Gravitational Angular Momentum Radiation of Cosmic Strings*, Nucl. Phys. B **328** (1989), 238

[247] R. Durrer, *The Cosmic Microwave Background* (Cambridge University Press, Cambridge, 2008)

[248] W. E. East and F. Pretorius, *Ultrarelativistic black hole formation*, Phys. Rev. Lett. **110** (2013), 101101, arXiv:1210.0443 [gr-qc]

[249] M. Ebersold and S. Tiwari, *Search for nonlinear memory from subsolar mass compact binary mergers*, Phys. Rev. D **101** (2020), 104041, arXiv:2005.03306 [gr-qc]

[250] A. Einstein, *The Foundation of the General Theory of Relativity*, Annalen Phys. **49** (1916), 769

[251] A. Einstein, *Über Gravitationswellen (on gravitational waves)* , Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) **1918** (1918), 154

[252] J. Ellis, M. Fairbairn, M. Lewicki, V. Vaskonen, and A. Wickens, *Intergalactic Magnetic Fields from First-Order Phase Transitions*, JCAP **09** (2019), 019, arXiv:1907.04315 [astro-ph.CO]

[253] J. Ellis, M. Fairbairn, M. Lewicki, V. Vaskonen, and A. Wickens, *Detecting circular polarisation in the stochastic gravitational-wave background from a first-order cosmological phase transition*, JCAP **10** (2020), 032, arXiv:2005.05278 [astro-ph.CO]

[254] J. Ellis and M. Lewicki, *Cosmic String Interpretation of NANOGrav Pulsar Timing Data*, Phys. Rev. Lett. **126** (2021), 041304, arXiv:2009.06555 [astro-ph.CO]

[255] J. Ellis, M. Lewicki, and J. M. No, *Gravitational waves from first-order cosmological phase transitions: lifetime of the sound wave source*, JCAP **07** (2020), 050, arXiv:2003.07360 [hep-ph]

[256] R. Epstein, *The binary pulsar: Post-Newtonian timing effects*, in *General Relativity and Gravitation 1977* (1977), 139, https://ui.adsabs.harvard.edu/abs/1977grg..conf..139E

[257] R. Epstein, *The Generation of Gravitational Radiation by Escaping Supernova Neutrinos*, Astrophys. J. **223** (1978), 1037

[258] J. R. Espinosa, T. Konstandin, J. M. No, and G. Servant, *Energy Budget of Cosmological First-order Phase Transitions*, JCAP **06** (2010), 028, arXiv:1004.4187 [hep-ph]

[259] F. B. Estabrook and H. D. Wahlquist, *Response of Doppler spacecraft tracking to gravitational radiation*, Gen. Rel. Grav. **6** (1975), 439, `https://ui.adsabs.harvard.edu/abs/1975GReGr.
..6..439E`

[260] L. Evans and P. Bryant (editors), *LHC Machine*, JINST **3** (2008), S08001

[261] J. M. Ezquiaga and M. Zumalacárregui, *Dark Energy After GW170817: Dead Ends and the Road Ahead*, Phys. Rev. Lett. **119** (2017), 251304, arXiv:1710.05901 [astro-ph.CO]

[262] M. Favata, *Gravitational-wave memory revisited: memory from the merger and recoil of binary black holes*, J. Phys. Conf. Ser. **154** (2009), 012043, arXiv:0811.3451 [astro-ph]

[263] M. Favata, *Post-Newtonian corrections to the gravitational-wave memory for quasi-circular, inspiralling compact binaries*, Phys. Rev. D **80** (2009), 024002, arXiv:0812.0069 [gr-qc]

[264] M. Favata, *Nonlinear gravitational-wave memory from binary black hole mergers*, Astrophys. J. Lett. **696** (2009), L159, arXiv:0902.3660 [astro-ph.SR]

[265] M. Favata, *The gravitational-wave memory effect*, Class. Quant. Grav. **27** (2010), 084036, arXiv:1003.3486 [gr-qc]

[266] M. Favata, *The Gravitational-wave memory from eccentric binaries*, Phys. Rev. D **84** (2011), 124013, arXiv:1108.3121 [gr-qc]

[267] M. A. Fedderke, P. W. Graham, and S. Rajendran, *Gravity Gradient Noise from Asteroids*, Phys. Rev. D **103** (2021), 103017, arXiv:2011.13833 [gr-qc]

[268] H. A. Feldman, N. Kaiser, and J. A. Peacock, *Power spectrum analysis of three-dimensional redshift surveys*, Astrophys. J. **426** (1994), 23, arXiv:astro-ph/9304022

[269] M. Fishbach and D. E. Holz, *Picky Partners: The Pairing of Component Masses in Binary Black Hole Mergers*, Astrophys. J. Lett. **891** (2020), L27, arXiv:1905.12669 [astro-ph.HE]

[270] É. É. Flanagan, *The Sensitivity of the laser interferometer gravitational wave observatory (LIGO) to a stochastic background, and its dependence on the detector orientations*, Phys. Rev. D **48** (1993), 2389, arXiv:astro-ph/9305029

[271] É. É. Flanagan and D. A. Nichols, *Conserved charges of the extended Bondi-Metzner-Sachs algebra*, Phys. Rev. D **95** (2017), 044002, arXiv:1510.03386 [hep-th]

[272] R. Flauger, N. Karnesis, G. Nardini, M. Pieroni, A. Ricciardone, and J. Torrado, *Improved reconstruction of a stochastic gravitational wave background with LISA*, JCAP **01** (2021), 059, arXiv:2009.11845 [astro-ph.CO]

[273] E. Fonseca, I. H. Stairs, and S. E. Thorsett, *A Comprehensive Study of Relativistic Gravity using PSR B1534+12*, Astrophys. J. **787** (2014), 82, arXiv:1402.4836 [astro-ph.HE]

[274] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, *emcee: The MCMC Hammer*, Publ. Astron. Soc. Pac. **125** (2013), 306, arXiv:1202.3665 [astro-ph.IM]

[275] J. W. Foster, N. L. Rodd, and B. R. Safdi, *Revealing the Dark Matter Halo with Axion Direct Detection*, Phys. Rev. D **97** (2018), 123006, arXiv:1711.10489 [astro-ph.CO]

[276] P. C. C. Freire, C. G. Bassa, N. Wex, I. H. Stairs, D. J. Chamion, S. M. Ransom, P. Lazarus, V. M. Kaspi, J. W. T. Hessels, M. Kramer, J. M. Cordes, J. P. W. Verbiest, P. Podsiadlowski, D. J. Nice, J. S. Deneva, D. R. Lorimer, B. W. Stappers, M. A. McLaughlin, and F. Camilo, *On the nature and evolution of the unique binary pulsar J1903+0327*, Mon. Not. Roy. Astron. Soc. **412** (2011), 2763, arXiv:1011.5809 [astro-ph.GA]

[277] T. Futamase and T. Matsuda, *Resonance between primordial gravitational waves and gravitationally bound systems*, Prog. Theor. Phys. **61** (1979), 86

[278] J. Garcia-Bellido, H. Murayama, and G. White, *Exploring the early Universe with Gaia and Theia*, JCAP **12** (2021), 023, arXiv:2104.04778 [hep-ph]

[279] J. García-Bellido, A. D. Linde, and D. Wands, *Density perturbations and black hole formation in hybrid inflation*, Phys. Rev. D **54** (1996), 6040, arXiv:astro-ph/9605094

[280] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, Springer Series in Synergetics, 3 ed. (Springer-Verlag, Berlin, 2004)

[281] D. Garfinkle, *General Relativistic Strings*, Phys. Rev. D **32** (1985), 1323

[282] J. Garriga and M. Sakellariadou, *Effects of friction on cosmic strings*, Phys. Rev. D **48** (1993), 2502, arXiv:hep-th/9303024

[283] J. Garriga and A. Vilenkin, *Black holes from nucleating strings*, Phys. Rev. D **47** (1993), 3265, arXiv:hep-ph/9208212

[284] V. Gayathri, J. Healy, J. Lange, B. O'Brien, M. Szczepanczyk, I. Bartos, M. Campanelli, S. Klimenko, C. Lousto, and R. O'Shaughnessy, *Eccentricity Estimate for Black Hole Mergers with Numerical Relativity Simulations*, Nat. Astron. (2020), 10.1038/s41550-021-01568-w, arXiv:2009.05461 [astro-ph.HE]

[285] M. Geller, A. Hook, R. Sundrum, and Y. Tsai, *Primordial Anisotropies in the Gravitational Wave Background from Cosmological Phase Transitions*, Phys. Rev. Lett. **121** (2018), 201303, arXiv:1803.10780 [hep-ph]

[286] A. M. Ghezelbash and R. B. Mann, *Abelian Higgs hair for rotating and charged black holes*, Phys. Rev. D **65** (2002), 124022, arXiv:hep-th/0110001

[287] N. Giacobbo and M. Mapelli, *The progenitors of compact-object binaries: impact of metallicity, common envelope and natal kicks*, Mon. Not. Roy. Astron. Soc. **480** (2018), 2011, arXiv:1806.00001 [astro-ph.HE]

[288] B. Goncharov, R. M. Shannon, D. J. Reardon, G. Hobbs, A. Zic, M. Bailes, M. Curyło, S. Dai, M. Kerr, M. E. Lower, R. N. Manchester, R. Mandow, H. Middleton, M. T. Miles, A. Parthasarathy, E. Thrane, N. Thyagarajan, X. Xue, X.-J. Zhu, A. D. Cameron, Y. Feng, R. Luo, C. J. Russell, J. Sarkissian, R. Spiewak, S. Wang, J. B. Wang, L. Zhang, and S. Zhang, *On the Evidence for a Common-spectrum Process in the Search for the Nanohertz Gravitational-wave Background with the Parkes Pulsar Timing Array*, Astrophys. J. Lett. **917** (2021), L19, arXiv:2107.12112 [astro-ph.HE]

[289] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelman, *HEALPix - A Framework for high resolution discretization, and fast analysis of data distributed on the sphere*, Astrophys. J. **622** (2005), 759, arXiv:astro-ph/0409513

[290] P. W. Graham, J. M. Hogan, M. A. Kasevich, S. Rajendran, and R. W. Romani (MAGIS), *Mid-band gravitational wave detection with precision atomic sensors* (2017), arXiv:1711.02225 [astro-ph.IM]

[291] A. M. Green and B. J. Kavanagh, *Primordial Black Holes as a dark matter candidate*, J. Phys. G **48** (2021), 4, arXiv:2007.10722 [astro-ph.CO]

[292] R. Gregory, D. Kubiznak, and D. Wills, *Rotating black hole hair*, JHEP **06** (2013), 023, arXiv:1303.0519 [gr-qc]

[293] L. P. Grishchuk, *Amplification of gravitational waves in an istropic universe* , Zh. Eksp. Teor. Fiz. **67** (1974), 825

[294] R. van Haasteren and Y. Levin, *Gravitational-wave memory and pulsar timing arrays*, Mon. Not. Roy. Astron. Soc. **401** (2010), 2372, arXiv:0909.0954 [astro-ph.IM]

[295] E. D. Hall, K. Kuns, J. R. Smith, Y. Bai, C. Wipf, S. Biscans, R. X. Adhikari, K. Arai, S. Ballmer, L. Barsotti, Y. Chen, M. Evans, P. Fritschel, J. Harms, B. Kamai, J. Graef Rollins, D. Shoemaker, B. Slagmolen, R. Weiss, and H. Yamamoto, *Gravitational-wave physics with Cosmic Explorer: Limits to low-frequency sensitivity*, Phys. Rev. D **103** (2021), 122004, arXiv:2012.03608 [gr-qc]

[296] A. J. S. Hamilton, *Power spectrum estimation. 1. Basics* , Lect. Notes Phys. **665** (2008), 415, arXiv:astro-ph/0503603

[297] S. Hanany, P. Ade, A. Balbi, J. Bock, J. Borrill, A. Boscaleri, P. de Bernardis, P. G. Ferreira, V. V. Hristov, A. H. Jaffe, A. E. Lange, A. T. Lee, P. D. Mauskopf, C. B. Netterfield, S. Oh, E. Pascale, B. Rabii, P. L. Richards, G. F. Smoot, R. Stompor, C. D. Winant, and J. H. P. Wu, *MAXIMA-1:*

A Measurement of the cosmic microwave background anisotropy on angular scales of 10 arcminutes to 5 degrees, Astrophys. J. Lett. **545** (2000), L5, arXiv:astro-ph/0005123

[298] H. T. Haniewicz, R. D. Ferdman, P. C. C. Freire, D. J. Champion, K. A. Bunting, D. R. Lorimer, and M. A. McLaughlin, *Precise mass measurements for the double neutron star system J1829+2456*, Mon. Not. Roy. Astron. Soc. **500** (2020), 4620, arXiv:2007.07565 [astro-ph.SR]

[299] T. Harada, C.-M. Yoo, K. Kohri, and K.-I. Nakao, *Spins of primordial black holes formed in the matter-dominated phase of the Universe*, Phys. Rev. D **96** (2017), 083517, [Erratum: Phys.Rev.D 99, 069904 (2019)], arXiv:1707.03595 [gr-qc]

[300] J. Harms, C. Mahrdt, M. Otto, and M. Priess, *Subtraction-noise projection in gravitational-wave detector networks*, Phys. Rev. D **77** (2008), 123010, arXiv:0803.0226 [gr-qc]

[301] J. Harms, *et al.* (LGWA), *Lunar Gravitational-wave Antenna*, Astrophys. J. **910** (2021), 1, arXiv:2010.13726 [gr-qc]

[302] G. M. Harry (LIGO), *Advanced LIGO: The next generation of gravitational wave detectors*, Class. Quant. Grav. **27** (2010), 084006

[303] S. Hawking, *Gravitationally collapsed objects of very low mass* , Mon. Not. Roy. Astron. Soc. **152** (1971), 75

[304] S. W. Hawking, *Black hole explosions*, Nature **248** (1974), 30

[305] S. W. Hawking, *Particle Creation by Black Holes*, Commun. Math. Phys. **43** (1975), 199, [Erratum: Commun.Math.Phys. 46, 206 (1976)]

[306] S. W. Hawking, *Black Holes From Cosmic Strings*, Phys. Lett. B **231** (1989), 237

[307] S. W. Hawking, *Gravitational radiation from collapsing cosmic string loops*, Phys. Lett. B **246** (1990), 36

[308] T. He, V. Lysov, P. Mitra, and A. Strominger, *BMS supertranslations and Weinberg's soft graviton theorem*, JHEP **05** (2015), 151, arXiv:1401.7026 [hep-th]

[309] O. Heaviside, *A gravitational and electromagnetic analogy*, The Electrician **31** (1893), 281, http://sergf.ru/Heavisid.htm

[310] T. Helfer, J. C. Aurrekoetxea, and E. A. Lim, *Cosmic String Loop Collapse in Full General Relativity*, Phys. Rev. D **99** (2019), 104028, arXiv:1808.06678 [gr-qc]

[311] R. W. Hellings and G. S. Downs, *Upper limits on the isotropic gravitational radiation background from pulsar timing analysis*, Astrophys. J. Lett. **265** (1983), L39

[312] G. Cañas Herrera, O. Contigiani, and V. Vardanyan, *Cross-correlation of the astrophysical gravitational-wave background with galaxy clustering*, Phys. Rev. D **102** (2020), 043513, arXiv:1910.08353 [astro-ph.CO]

[313] S. Hild, *et al.*, *Sensitivity Studies for Third-Generation Gravitational Wave Observatories*, Class. Quant. Grav. **28** (2011), 094013, arXiv:1012.0908 [gr-qc]

[314] M. Hindmarsh, J. Lizarraga, J. Urrestilla, D. Daverio, and M. Kunz, *Scaling from gauge and scalar radiation in Abelian Higgs string networks*, Phys. Rev. D **96** (2017), 023525, arXiv:1703.06696 [astro-ph.CO]

[315] M. Hindmarsh, S. Stuckey, and N. Bevis, *Abelian Higgs Cosmic Strings: Small Scale Structure and Loops*, Phys. Rev. D **79** (2009), 123504, arXiv:0812.1929 [hep-th]

[316] M. B. Hindmarsh and T. W. B. Kibble, *Cosmic strings*, Rept. Prog. Phys. **58** (1995), 477, arXiv:hep-ph/9411342

[317] M. B. Hindmarsh, M. Lüben, J. Lumma, and M. Pauly, *Phase transitions in the early universe*, SciPost Phys. Lect. Notes **24** (2021), 1, arXiv:2008.09136 [astro-ph.CO]

[318] G. Hobbs, *The Parkes Pulsar Timing Array*, Class. Quant. Grav. **30** (2013), 224007, arXiv:1307.2629 [astro-ph.IM]

[319] C. J. Hogan and M. J. Rees, *Gravitational interactions of cosmic strings*, Nature **311** (1984), 109

[320] S. C. Hotinli, M. Kamionkowski, and A. H. Jaffe, *The search for anisotropy in the gravitational-wave background with pulsar-timing arrays*, Open J. Astrophys. **2** (2019), 8, arXiv:1904.05348 [astro-ph.CO]

[321] M. Hübner, C. Talbot, P. D. Lasky, and E. Thrane, *Measuring gravitational-wave memory in the first LIGO/Virgo gravitational-wave transient catalog*, Phys. Rev. D **101** (2020), 023011, arXiv:1911.12496 [astro-ph.HE]

[322] C. Y. Hui, K. Wu, Q. Han, A. K. H. Kong, and P. H. T. Tam, *On the orbital properties of millisecond pulsar binaries*, Astrophys. J. **864** (2018), 30, arXiv:1807.09001 [astro-ph.HE]

[323] L. Hui, S. T. McWilliams, and I.-S. Yang, *Binary systems as resonance detectors for gravitational waves*, Phys. Rev. D **87** (2013), 084009, arXiv:1212.2623 [gr-qc]

[324] R. A. Hulse and J. H. Taylor, *Discovery of a pulsar in a binary system*, Astrophys. J. Lett. **195** (1975), L51

[325] L. Iacconi, M. Fasiello, H. Assadullahi, E. Dimastrogiovanni, and D. Wands, *Interferometer Constraints on the Inflationary Field Content*, JCAP **03** (2020), 031, arXiv:1910.12921 [astro-ph.CO]

[326] L. Iacconi, M. Fasiello, H. Assadullahi, and D. Wands, *Small-scale Tests of Inflation*, JCAP **12** (2020), 005, arXiv:2008.00452 [astro-ph.CO]

[327] ILRS, *International laser ranging service: Lageos*, https://ilrs.gsfc.nasa.gov/missions/satellite_missions/current_missions/lag1_general.html (accessed May 2021)

[328] L. Iorio, *Orbital effects of a monochromatic plane gravitational wave with ultra-low frequency incident on a gravitationally bound two-body system* (2011), 10.14293/A2199-1006.01.SOR-ASTRO.WXWIL.v1, arXiv:1104.4853 [gr-qc]

[329] L. Iorio and M. L. Ruggiero, *Perturbations of the orbital elements due to the magnetic-like part of the field of a plane gravitational wave*, Int. J. Mod. Phys. D **30** (2021), 2150088, arXiv:2106.14462 [gr-qc]

[330] K. Islo, J. Simon, S. Burke-Spolaor, and X. Siemens, *Prospects for Memory Detection with Low-Frequency Gravitational Wave Detectors* (2019), arXiv:1906.11936 [astro-ph.HE]

[331] K. Itô, *Stochastic integral*, Proc. Imp. Acad. **20** (1944), 519

[332] P. Ivanov, P. Naselsky, and I. Novikov, *Inflation and primordial black holes as dark matter*, Phys. Rev. D **50** (1994), 7173

[333] J. D. Jackson, *Classical Electrodynamics* (Wiley, 1998)

[334] J. D. Jackson and L. B. Okun, *Historical roots of gauge invariance*, Rev. Mod. Phys. **73** (2001), 663, arXiv:hep-ph/0012061

[335] B. A. Jacoby, P. B. Cameron, F. A. Jenet, S. B. Anderson, R. N. Murty, and S. R. Kulkarni, *Measurement of Orbital Decay in the Double Neutron Star Binary PSR B2127+11C*, Astrophys. J. Lett. **644** (2006), L113, arXiv:astro-ph/0605375

[336] A. H. Jaffe and D. C. Backer, *Gravitational waves probe the coalescence rate of massive black hole binaries*, Astrophys. J. **583** (2003), 616, arXiv:astro-ph/0210148

[337] C. James-Turner, D. P. B. Weil, A. M. Green, and E. J. Copeland, *Constraints on the cosmic string loop collapse fraction from primordial black holes*, Phys. Rev. D **101** (2020), 123526, arXiv:1911.12658 [astro-ph.CO]

[338] K. Jani and A. Loeb, *Gravitational-Wave Lunar Observatory for Cosmology* (2020), arXiv:2007.08550 [gr-qc]

[339] G. Janssen, G. Hobbs, M. McLaughlin, C. Bassa, A. T. Deller, M. Kramer, K. Lee, C. Mingarelli, P. Rosado, S. Sanidas, A. Sesana, L. Shao, I. Stairs, B. W. Stapper, and J. Verbiest, *Gravitational wave astronomy with the SKA*, PoS **AASKA14** (2015), 037, arXiv:1501.00127 [astro-ph.IM]

[340] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003)

[341] R. Jeannerot, J. Rocher, and M. Sakellariadou, *How generic is cosmic string formation in SUSY GUTs*, Phys. Rev. D **68** (2003), 103514, arXiv:hep-ph/0308134

[342] A. C. Jenkins, R. O'Shaughnessy, M. Sakellariadou, and D. Wysocki, *Anisotropies in the astrophysical gravitational-wave background: The impact of black hole distributions*, Phys. Rev. Lett. **122** (2019), 111101, arXiv:1810.13435 [astro-ph.CO]

[343] A. C. Jenkins, A. G. A. Pithis, and M. Sakellariadou, *Can we detect quantum gravity with compact binary inspirals?*, Phys. Rev. D **98** (2018), 104032, arXiv:1809.06275 [gr-qc]

[344] A. C. Jenkins, J. D. Romano, and M. Sakellariadou, *Estimating the angular power spectrum of the gravitational-wave background in the presence of shot noise*, Phys. Rev. D **100** (2019), 083501, arXiv:1907.06642 [astro-ph.CO]

[345] A. C. Jenkins and M. Sakellariadou, *Anisotropies in the stochastic gravitational-wave background: Formalism and the cosmic string case*, Phys. Rev. D **98** (2018), 063509, arXiv:1802.06046 [astro-ph.CO]

[346] A. C. Jenkins and M. Sakellariadou, *Shot noise in the astrophysical gravitational-wave background*, Phys. Rev. D **100** (2019), 063508, arXiv:1902.07719 [astro-ph.CO]

[347] A. C. Jenkins and M. Sakellariadou, *Primordial black holes from cusp collapse on cosmic strings* (2020), arXiv:2006.16249 [astro-ph.CO]

[348] A. C. Jenkins and M. Sakellariadou, *Nonlinear gravitational-wave memory from cusps and kinks on cosmic strings*, Class. Quant. Grav. **38** (2021), 165004, arXiv:2102.12487 [gr-qc]

[349] A. C. Jenkins, M. Sakellariadou, T. Regimbau, and E. Slezak, *Anisotropies in the astrophysical gravitational-wave background: Predictions for the detection of compact binaries by LIGO and Virgo*, Phys. Rev. D **98** (2018), 063501, arXiv:1806.01718 [astro-ph.CO]

[350] A. D. Johnson, S. J. Kapadia, A. Osborne, A. Hixon, and D. Kennefick, *Prospects of detecting the nonlinear gravitational wave memory*, Phys. Rev. D **99** (2019), 044045, arXiv:1810.09563 [gr-qc]

[351] N. Jokela, K. Kajantie, and M. Sarkkinen, *Memory effect in Yang-Mills theory*, Phys. Rev. D **99** (2019), 116003, arXiv:1903.10231 [hep-th]

[352] N. Jokela, K. Kajantie, and M. Sarkkinen, *Memory effect in Yang-Mills theory with an axion*, Phys. Rev. D **104** (2021), 074029, arXiv:2012.14160 [hep-ph]

[353] JPL, *Small-body database browser*, `https://ssd.jpl.nasa.gov/sbdb.cgi` (accessed May 2021)

[354] T. Kahniashvili, G. Gogoberidze, and B. Ratra, *Polarized cosmological gravitational waves from primordial helical turbulence*, Phys. Rev. Lett. **95** (2005), 151301, arXiv:astro-ph/0505628

[355] N. Kaiser and A. Stebbins, *Microwave Anisotropy Due to Cosmic Strings*, Nature **310** (1984), 391

[356] K. Kajantie, M. Laine, K. Rummukainen, and M. E. Shaposhnikov, *Is there a hot electroweak phase transition at m(H) larger or equal to m(W)?*, Phys. Rev. Lett. **77** (1996), 2887, arXiv:hep-ph/9605288

[357] T. Kalaydzhyan and N. Yu, *Searching for stochastic background of ultra-light fields with atomic sensors*, Universe **4** (2018), 99, arXiv:1801.07577 [astro-ph.CO]

Bibliography

[358] M. Kamionkowski, A. Kosowsky, and A. Stebbins, *A Probe of primordial gravity waves and vorticity*, Phys. Rev. Lett. **78** (1997), 2058, arXiv:astro-ph/9609132

[359] M. Kamionkowski, A. Kosowsky, and M. S. Turner, *Gravitational radiation from first order phase transitions*, Phys. Rev. D **49** (1994), 2837, arXiv:astro-ph/9310044

[360] D. Kapec, V. Lysov, S. Pasterski, and A. Strominger, *Higher-dimensional supertranslations and Weinberg's soft graviton theorem*, Ann. Math. Sci. Appl. **02** (2017), 69, arXiv:1502.07644 [gr-qc]

[361] A. Kashlinsky, R. G. Arendt, F. Atrio-Barandela, N. Cappelluti, A. Ferrara, and G. Hasinger, *Looking at cosmic near-infrared background radiation anisotropies*, Rev. Mod. Phys. **90** (2018), 025006, arXiv:1802.07774 [astro-ph.CO]

[362] V. M. Kaspi, M. Bailes, R. N. Manchester, B. W. Stappers, and J. F. Bell, *Evidence from a precessing pulsar orbit for a neutron-star birth kick*, Nature **381** (1996), 584, https://ui.adsabs.harvard.edu/abs/1996Natur.381..584K

[363] S. Katsanevas, *et al.*, *Lunar Seismic and Gravitational-Wave Antenna*, Ideas for exploring the Moon with a large European lander (ESA) (2020), https://ideas.esa.int/servlet/hype/IMT?documentTableId=4508760703174401O&userAction=Browse&templateName=&documentId=a315450fae481074411ef65e4c5b7746

[364] B. J. Kavanagh, *bradkav/pbhbounds: Release version* (2019), doi:10.5281/zenodo.3538999, https://doi.org/10.5281/zenodo.3538999

[365] B. J. Kavanagh, D. Gaggero, and G. Bertone, *Merger rate of a subdominant population of primordial black holes*, Phys. Rev. D **98** (2018), 023536, arXiv:1805.09034 [astro-ph.CO]

[366] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, vol. 1 (Prentice Hall, 1993)

[367] D. Kennefick, *Prospects for detecting the Christodoulou memory of gravitational waves from a coalescing compact binary and using it to measure neutron star radii*, Phys. Rev. D **50** (1994), 3587

[368] D. Kennefick, *Traveling at the speed of thought: Einstein and the quest for gravitational waves* (Princeton University Press, 2007)

[369] M. H. van Kerkwijk and S. R. Kulkarni, *A massive white dwarf companion to the eccentric binary pulsar system PSR B2303+46*, Astrophys. J. Lett. **516** (1999), L25, arXiv:astro-ph/9901149

[370] N. Khera, B. Krishnan, A. Ashtekar, and T. De Lorenzo, *Inferring the gravitational wave memory for binary coalescence events*, Phys. Rev. D **103** (2021), 044012, arXiv:2009.06351 [gr-qc]

[371] T. W. B. Kibble, *Topology of Cosmic Domains and Strings*, J. Phys. A **9** (1976), 1387

[372] T. W. B. Kibble and N. Turok, *Selfintersection of Cosmic Strings*, Phys. Lett. B **116** (1982), 141

[373] H. Kim, *A constraint on light primordial black holes from the interstellar medium temperature,* Mon. Not. Roy. Astron. Soc. **504** (2020), 5475, arXiv:2007.07739 [hep-ph]

[374] L. Kisslinger and T. Kahniashvili, *Polarized Gravitational Waves from Cosmological Phase Transitions,* Phys. Rev. D **92** (2015), 043006, arXiv:1505.03680 [astro-ph.CO]

[375] T. Kite, A. Ravenni, S. P. Patil, and J. Chluba, *Bridging the gap: spectral distortions meet gravitational waves,* Mon. Not. Roy. Astron. Soc. **505** (2021), 4396, arXiv:2010.00040 [astro-ph.CO]

[376] A. Klein, E. Barausse, A. Sesana, A. Petiteau, E. Berti, S. Babak, J. Gair, S. Aoudia, I. Hinder, F. Ohme, and B. Wardell, *Science with the space-based interferometer eLISA: Supermassive black hole binaries,* Phys. Rev. D **93** (2016), 024003, arXiv:1511.05581 [gr-qc]

[377] K. D. Kokkotas and B. G. Schmidt, *Quasinormal modes of stars and black holes,* Living Rev. Rel. **2** (1999), 2, arXiv:gr-qc/9909058

[378] K. Kotake, K. Sato, and K. Takahashi, *Explosion mechanism, neutrino burst, and gravitational wave in core-collapse supernovae,* Rept. Prog. Phys. **69** (2006), 971, arXiv:astro-ph/0509456

[379] S. J. Kovács and K. S. Thorne, *The Generation of Gravitational Waves. 4. Bremsstrahlung,* Astrophys. J. **224** (1978), 62

[380] M. Kramer and D. J. Champion, *The European Pulsar Timing Array and the Large European Array for Pulsars,* Class. Quant. Grav. **30** (2013), 224009

[381] M. Kramer, I. H. Stairs, R. N. Manchester, M. A. McLaughlin, A. G. Lyne, R. D. Ferdman, M. Burgay, D. R. Loreimer, A. Possenti, N. D'Amico, J. M. Sarkissian, G. B. Hobbs, J. E. Reynolds, P. C. C. Freire, and F. Camilo, *Tests of general relativity from timing the double pulsar,* Science **314** (2006), 97, arXiv:astro-ph/0609417

[382] S. Kumar, R. Sundrum, and Y. Tsai, *Non-Gaussian stochastic gravitational waves from phase transitions,* JHEP **11** (2021), 107, arXiv:2102.05665 [astro-ph.CO]

[383] S. Kuroyanagi, K. Takahashi, N. Yonemaru, and H. Kumamoto, *Anisotropies in the gravitational wave background as a probe of the cosmic string network,* Phys. Rev. D **95** (2017), 043531, arXiv:1604.00332 [astro-ph.CO]

[384] S. Kuroyanagi, T. Takahashi, and S. Yokoyama, *Blue-tilted inflationary tensor spectrum and reheating in the light of NANOGrav results,* JCAP **01** (2021), 071, arXiv:2011.03323 [astro-ph.CO]

[385] F. Kühnel, *Enhanced Detectability of Spinning Primordial Black Holes,* Eur. Phys. J. C **80** (2020), 243, arXiv:1909.04742 [astro-ph.CO]

[386] R. Laha, *Primordial Black Holes as a Dark Matter Candidate Are Severely Constrained by the Galactic Center 511 keV γ -Ray Line,* Phys. Rev. Lett. **123** (2019), 251101, arXiv:1906.09994 [astro-ph.HE]

[387] C. Lange, F. Camilo, N. Wex, M. Kramer, D. C. Backer, A. G. Lyne, and O. Doroshenko, *Precision timing measurements of PSR J1012+5307*, Mon. Not. Roy. Astron. Soc. **326** (2001), 274, arXiv:astro-ph/0102309

[388] P. D. Lasky, C. M. F. Mingarelli, T. L. Smith, J. J. T. Giblin, E. Thrane, D. J. Reardon, R. Caldwell, M. Bailes, N. D. Ramesh Bhat, S. Burke-Spolaor, S. Dai, J. Dempsey, G. Hobbs, M. Kerr, Y. Levin, R. N. Manchester, S. Osłowski, V. Ravi, P. A. Rosado, R. M. Shannon, R. Spiewak, W. van Straten, L. Toomey, J. Wang, L. Wen, X. You, and X. Zhu, *Gravitational-wave cosmology across 29 decades in frequency*, Phys. Rev. X **6** (2016), 011035, arXiv:1511.05994 [astro-ph.CO]

[389] P. D. Lasky, E. Thrane, Y. Levin, J. Blackman, and Y. Chen, *Detecting gravitational-wave memory with LIGO: implications of GW150914*, Phys. Rev. Lett. **117** (2016), 061102, arXiv:1605.01415 [astro-ph.HE]

[390] G. Lemson (VIRGO), *Halo and Galaxy Formation Histories from the Millennium Simulation: Public release of a VO-oriented and SQL-queryable database for studying the evolution of galaxies in the Lambda-CDM cosmogony* (2006), arXiv:astro-ph/0608019

[391] B. Linet, *Absorption of gravitational waves by nearly Newtonian systems*, Gen. Rel. Grav. **14** (1982), 479

[392] J. Liu, R.-G. Cai, and Z.-K. Guo, *Large Anisotropies of the Stochastic Gravitational Wave Background from Cosmic Domain Walls*, Phys. Rev. Lett. **126** (2021), 141303, arXiv:2010.03225 [astro-ph.CO]

[393] K. Liu, J. P. W. Verbiest, M. Kramer, B. W. Stappers, W. van Straten, and J. M. Cordes, *Prospects for High-Precision Pulsar Timing*, Mon. Not. Roy. Astron. Soc. **417** (2011), 2916, arXiv:1107.3086 [astro-ph.HE]

[394] I. Lopes and J. Silk, *Nearby Stars as Gravitational Wave Detectors*, Astrophys. J. **807** (2015), 135, arXiv:1507.03212 [astro-ph.SR]

[395] D. López Nacir and F. R. Urban, *Vector Fuzzy Dark Matter, Fifth Forces, and Binary Pulsars*, JCAP **10** (2018), 044, arXiv:1807.10491 [astro-ph.CO]

[396] L. Lorenz, C. Ringeval, and M. Sakellariadou, *Cosmic string loop distribution on all length scales and at any redshift*, JCAP **10** (2010), 003, arXiv:1006.0931 [astro-ph.CO]

[397] D. R. Lorimer, *Binary and Millisecond Pulsars*, Living Rev. Rel. **11** (2008), 8, arXiv:0811.0762 [astro-ph]

[398] M. Lucca, N. Schöneberg, D. C. Hooper, J. Lesgourgues, and J. Chluba, *The synergy between CMB spectral distortions and anisotropies*, JCAP **02** (2020), 026, arXiv:1910.04619 [astro-ph.CO]

[399] A. Lyne, B. Stappers, M. Keith, P. Ray, M. Kerr, F. Camilo, and T. Johnson, *The Binary Nature of PSR J2032+4127*, Mon. Not. Roy. Astron. Soc. **451** (2015), 581, arXiv:1502.01465 [astro-ph.HE]

[400] J. H. MacGibbon, R. H. Brandenberger, and U. F. Wichoski, *Limits on black hole formation from cosmic string loops*, Phys. Rev. D **57** (1998), 2158, arXiv:astro-ph/9707146

[401] C. S. Machado, W. Ratzinger, P. Schwaller, and B. A. Stefanek, *Audible Axions*, JHEP **01** (2019), 053, arXiv:1811.01950 [hep-ph]

[402] D. R. Madison, J. M. Cordes, and S. Chatterjee, *Assessing Pulsar Timing Array Sensitivity to Gravitational Wave Bursts with Memory*, Astrophys. J. **788** (2014), 141, arXiv:1404.5682 [astro-ph.HE]

[403] E. C. Madsen, I. H. Stairs, M. Kramer, F. Camilo, G. B. Hobbs, G. H. Janssen, A. G. Lyne, R. N. Manchester, A. Possenti, and B. W. Stappers, *Timing the main-sequence-star binary pulsar J1740-3052*, Mon. Not. Roy. Astron. Soc. **425** (2012), 2378, arXiv:1207.2202 [astro-ph.HE]

[404] R. Magee, A.-S. Deutsch, P. McClincy, C. Hanna, C. Horst, D. Meacher, C. Messick, S. Shandera, and M. Wade, *Methods for the detection of gravitational waves from subsolar mass ultracompact binaries*, Phys. Rev. D **98** (2018), 103024, arXiv:1808.04772 [astro-ph.IM]

[405] M. Maggiore, *Gravitational wave experiments and early universe cosmology*, Phys. Rept. **331** (2000), 283, arXiv:gr-qc/9909001

[406] M. Maggiore, *Gravitational Waves. Vol. 1: Theory and Experiments*, Oxford Master Series in Physics (Oxford University Press, 2007)

[407] M. Maggiore, C. Van Den Broeck, N. Bartolo, E. Belgacem, D. Bertacca, M.-A. Bizouard, M. Branchesi, S. Clesse, S. Foffa, J. García-Bellido, S. Grimm, J. Harms, T. Hinderer, S. Matarrese, C. Palomba, M. Peloso, A. Ricciardone, and M. Sakellariadou, *Science Case for the Einstein Telescope*, JCAP **03** (2020), 050, arXiv:1912.02622 [astro-ph.CO]

[408] A. Malhotra, E. Dimastrogiovanni, M. Fasiello, and M. Shiraishi, *Cross-correlations as a Diagnostic Tool for Primordial Gravitational Waves*, JCAP **03** (2021), 088, arXiv:2012.03498 [astro-ph.CO]

[409] R. N. Manchester, *The International Pulsar Timing Array*, Class. Quant. Grav. **30** (2013), 224010, arXiv:1309.7392 [astro-ph.IM]

[410] R. N. Manchester, G. B. Hobbs, A. Teoh, and M. Hobbs, *The Australia Telescope National Facility pulsar catalogue*, Astron. J. **129** (2005), 1993, arXiv:astro-ph/0412641

[411] V. Mandic, S. Bird, and I. Cholis, *Stochastic Gravitational-Wave Background due to Primordial Binary Black Hole Mergers*, Phys. Rev. Lett. **117** (2016), 201102, arXiv:1608.06699 [astro-ph.CO]

[412] A. Margalit, C. R. Contaldi, and M. Pieroni, *Phase decoherence of gravitational wave backgrounds*, Phys. Rev. D **102** (2020), 083506, arXiv:2004.01727 [astro-ph.CO]

[413] B. Margalit and B. D. Metzger, *Constraining the Maximum Mass of Neutron Stars From Multi-Messenger Observations of GW170817*, Astrophys. J. Lett. **850** (2017), L19, arXiv:1710.05938 [astro-ph.HE]

[414] K. Martinovic, C. Badger, M. Sakellariadou, and V. Mandic, *Searching for parity violation with the LIGO-Virgo-KAGRA network*, Phys. Rev. D **104** (2021), L081101, arXiv:2103.06718 [gr-qc]

[415] K. Martinovic, P. M. Meyers, M. Sakellariadou, and N. Christensen, *Simultaneous estimation of astrophysical and cosmological stochastic gravitational-wave backgrounds with terrestrial detectors*, Phys. Rev. D **103** (2021), 043023, arXiv:2011.05697 [gr-qc]

[416] B. Mashhoon, *On tidal resonance*, Astrophys. J. **223** (1978), 285, `https://ui.adsabs.harvard.edu/abs/1978ApJ...223..285M`

[417] B. Mashhoon, B. J. Carr, and B. L. Hu, *The influence of cosmological gravitational waves on a Newtonian binary system*, Astrophys. J. **246** (1981), 569, `https://ui.adsabs.harvard.edu/abs/1981ApJ...246..569M`

[418] D. Matsunami, L. Pogosian, A. Saurabh, and T. Vachaspati, *Decay of Cosmic String Loops Due to Particle Radiation*, Phys. Rev. Lett. **122** (2019), 201301, arXiv:1903.05102 [hep-ph]

[419] J. D. McEwen, S. M. Feeney, H. V. Peiris, Y. Wiaux, C. Ringeval, and F. R. Bouchet, *Wavelet-Bayesian inference of cosmic strings embedded in the cosmic microwave background*, Mon. Not. Roy. Astron. Soc. **472** (2017), 4081, arXiv:1611.10347 [astro-ph.IM]

[420] M. A. McLaughlin, *The North American Nanohertz Observatory for Gravitational Waves*, Class. Quant. Grav. **30** (2013), 224008, arXiv:1310.0758 [astro-ph.IM]

[421] L. O. McNeill, E. Thrane, and P. D. Lasky, *Detecting Gravitational Wave Memory without Parent Signals*, Phys. Rev. Lett. **118** (2017), 181103, arXiv:1702.01759 [astro-ph.IM]

[422] D. Meacher, E. Thrane, and T. Regimbau, *Statistical properties of astrophysical gravitational-wave backgrounds*, Phys. Rev. D **89** (2014), 084063, arXiv:1402.6231 [astro-ph.CO]

[423] G. Mentasti and M. Peloso, *ET sensitivity to the anisotropic Stochastic Gravitational Wave Background*, JCAP **03** (2021), 080, arXiv:2010.00486 [astro-ph.CO]

[424] P. M. Meyers, K. Martinovic, N. Christensen, and M. Sakellariadou, *Detecting a stochastic gravitational-wave background in the presence of correlated magnetic noise*, Phys. Rev. D **102** (2020), 102005, arXiv:2008.00789 [gr-qc]

[425] H. Middleton, A. Sesana, S. Chen, A. Vecchio, W. Del Pozzo, and P. A. Rosado, *Massive black hole binary systems and the NANOGrav 12.5 yr results*, Mon. Not. Roy. Astron. Soc. **502** (2021), L99, arXiv:2011.01246 [astro-ph.HE]

[426] J. C. A. Miller-Jones, A. T. Deller, R. M. Shannon, R. Dodson, J. Moldón, M. Ribó, G. Dubus, S. Johnston, J. M. Paredes, S. M. Ransom, and J. A. Tomsick, *The geometric distance and binary orbit of PSR B1259–63*, Mon. Not. Roy. Astron. Soc. **479** (2018), 4849, arXiv:1804.08402 [astro-ph.HE]

[427] S. E. de Mink and I. Mandel, *The chemically homogeneous evolutionary channel for binary black hole mergers: rates and properties of gravitational-wave events detectable by advanced LIGO*, Mon. Not. Roy. Astron. Soc. **460** (2016), 3545, arXiv:1603.02291 [astro-ph.HE]

[428] M. Mirbabayi, A. Gruzinov, and J. Noreña, *Spin of Primordial Black Holes*, JCAP **03** (2020), 017, arXiv:1901.05963 [astro-ph.CO]

[429] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (W. H. Freeman, San Francisco, 1973)

[430] C. J. Moore, D. P. Mihaylov, A. Lasenby, and G. Gilmore, *Astrometric Search Method for Individually Resolvable Gravitational Wave Sources with Gaia*, Phys. Rev. Lett. **119** (2017), 261102, arXiv:1707.06239 [astro-ph.IM]

[431] T. W. Murphy, *Lunar laser ranging: the millimeter challenge*, Rept. Prog. Phys. **76** (2013), 076901, arXiv:1309.6294 [gr-qc]

[432] C. D. Murray and S. F. Dermott, *Solar System Dynamics* (Cambridge University Press, 2000)

[433] T. Namikawa, S. Saga, D. Yamauchi, and A. Taruya, *CMB Constraints on the Stochastic Gravitational-Wave Background at Mpc scales*, Phys. Rev. D **100** (2019), 021303, arXiv:1904.02115 [astro-ph.CO]

[434] L. A. Nelson and W. Y. Chau, *Orbital perturbations of a gravitationally bound two-body system with the passage of gravitational waves*, Astrophys. J. **254** (1982), 735

[435] A. Neronov, A. Roper Pol, C. Caprini, and D. Semikoz, *NANOGrav signal from magnetohydrodynamic turbulence at the QCD phase transition in the early Universe*, Phys. Rev. D **103** (2021), 041302, arXiv:2009.14174 [astro-ph.CO]

[436] D. A. Nichols, *Spin memory effect for compact binaries in the post-Newtonian approximation*, Phys. Rev. D **95** (2017), 084048, arXiv:1702.03300 [gr-qc]

[437] H. Niikura, M. Takada, S. Yokoyama, T. Sumi, and S. Masaki, *Constraints on Earth-mass primordial black holes from OGLE 5-year microlensing events*, Phys. Rev. D **99** (2019), 083503, arXiv:1901.07120 [astro-ph.CO]

[438] A. Nitz, I. Harry, D. Brown, C. M. Biwer, J. Willis, T. D. Canton, C. Capano, T. Dent, L. Pekowsky, A. R. Williamson, G. S. C. Davies, S. De, M. Cabero, B. Machenschalk, P. Kumar, D. Macleod, S. Reyes, D. Finstad, F. Pannarale, T. Massinger, S. Kumar, M. Tápai, L. Singer, S. Khan, S. Fairhurst, A. Nielsen, S. Singh, S. J. Kapadia, B. U. V. Gadre, and I. Dorrington, *gwastro/pycbc: Pycbc release 1.18.1* (2021), doi:10.5281/zenodo.4849433, https://doi.org/10.5281/zenodo.4849433

[439] S. Ölmez, V. Mandic, and X. Siemens, *Anisotropies in the Gravitational-Wave Stochastic Background*, JCAP **07** (2012), 009, arXiv:1106.5555 [astro-ph.CO]

[440] C. D. Ott, *The Gravitational Wave Signature of Core-Collapse Supernovae*, Class. Quant. Grav. **26** (2009), 063001, arXiv:0809.0695 [astro-ph]

[441] L. Pagano, L. Salvati, and A. Melchiorri, *New constraints on primordial gravitational waves from Planck 2015*, Phys. Lett. B **760** (2016), 823, arXiv:1508.02393 [astro-ph.CO]

[442] D. N. Page, *Particle Emission Rates from a Black Hole. 2. Massless Particles from a Rotating Hole*, Phys. Rev. D **14** (1976), 3260

[443] D. N. Page, *Particle Emission Rates from a Black Hole: Massless Particles from an Uncharged, Nonrotating Hole*, Phys. Rev. D **13** (1976), 198

[444] S. Panda, S. Bhagwat, J. Suresh, and S. Mitra, *Stochastic gravitational wave background mapmaking using regularized deconvolution*, Phys. Rev. D **100** (2019), 043541, arXiv:1905.08276 [gr-qc]

[445] A. Parida, J. Suresh, S. Mitra, and S. Jhingan, *Component separation map-making for stochastic gravitational wave background* (2019), arXiv:1904.05056 [gr-qc]

[446] S. Pasterski, *Asymptotic Symmetries and Electromagnetic Memory*, JHEP **09** (2017), 154, arXiv:1505.00716 [hep-th]

[447] S. Pasterski, A. Strominger, and A. Zhiboedov, *New Gravitational Memories*, JHEP **12** (2016), 053, arXiv:1502.06120 [hep-th]

[448] M. Pate, A.-M. Raclariu, and A. Strominger, *Color Memory: A Yang-Mills Analog of Gravitational Wave Memory*, Phys. Rev. Lett. **119** (2017), 261602, arXiv:1707.08016 [hep-th]

[449] M. E. Pati and C. M. Will, *Post-Newtonian gravitational radiation and equations of motion via direct integration of the relaxed Einstein equations. 1. Foundations*, Phys. Rev. D **62** (2000), 124015, arXiv:gr-qc/0007087

[450] R. F. Pawula, *Approximation of the Linear Boltzmann Equation by the Fokker-Planck Equation*, Phys. Rev. **162** (1967), 186

[451] P. J. E. Peebles, *The large-scale structure of the universe* (Princeton University Press, 1980)

[452] R. Penrose, *Naked singularities*, Annals N. Y. Acad. Sci. **224** (1973), 125

[453] A. A. Penzias and R. W. Wilson, *A Measurement of excess antenna temperature at 4080-Mc/s*, Astrophys. J. **142** (1965), 419

[454] B. B. P. Perera, B. W. Stappers, S. Babak, M. J. Keith, J. Antoniadis, C. G. Bassa, R. N. Caballero, D. J. Champion, I. Cognard, G. Desvignes, E. Graikou, L. Guillemot, G. H. Janssen, R. Karuppusamy, M. Kramer, P. Lazarus, L. Lentati, K. Liu, A. G. Lyne, J. W. McKee, S. Osłowski, D. Perrodin, S. A. Sanidas, A. Sesana, G. Shaifullah, G. Theureau, J. P. W. Verbiest, and S. R. Taylor, *Improving timing sensitivity in the microhertz frequency regime: limits from PSR J1713+0747 on gravitational waves produced by super-massive black-hole binaries*, Mon. Not. Roy. Astron. Soc. **478** (2018), 218, arXiv:1804.10571 [astro-ph.HE]

[455] P. C. Peters and J. Mathews, *Gravitational radiation from point masses in a Keplerian orbit*, Phys. Rev. **131** (1963), 435

[456] E. S. Phinney, *A Practical theorem on gravitational wave backgrounds* (2001), arXiv:astro-ph/0108028

[457] C. Pitrou, G. Cusin, and J.-P. Uzan, *Unified view of anisotropies in the astrophysical gravitational-wave background*, Phys. Rev. D **101** (2020), 081301, arXiv:1910.04645 [astro-ph.CO]

[458] H. Poincaré, *Sur la dynamique de l'électron*, Rend. Circ. Matem. Palermo **21** (1906), 129

[459] E. Poisson and C. M. Will, *Gravity: Newtonian, Post-Newtonian, Relativistic* (Cambridge University Press, 2014)

[460] D. Pollney and C. Reisswig, *Gravitational memory in binary black hole mergers*, Astrophys. J. Lett. **732** (2011), L13, arXiv:1004.4209 [gr-qc]

[461] A. Polnarev and R. Zembowicz, *Formation of Primordial Black Holes by Cosmic Strings*, Phys. Rev. D **43** (1991), 1106

[462] A. G. Polnarev and M. Yu. Khlopov, *Cosmology, primordial black holes, and supermassive particles*, Sov. Phys. Usp. **28** (1985), 213

[463] V. Poulin, J. Lesgourgues, and P. D. Serpico, *Cosmological constraints on exotic injection of electromagnetic energy*, JCAP **03** (2017), 043, arXiv:1610.10051 [astro-ph.CO]

[464] M. Punturo, *et al.*, *The Einstein Telescope: A third-generation gravitational wave observatory*, Class. Quant. Grav. **27** (2010), 194002

[465] J. M. Quashnock and D. N. Spergel, *Gravitational Selfinteractions of Cosmic Strings*, Phys. Rev. D **42** (1990), 2505

[466] D. Radice, A. Perego, F. Zappa, and S. Bernuzzi, *GW170817: Joint Constraint on the Neutron Star Equation of State from Multimessenger Observations*, Astrophys. J. Lett. **852** (2018), L29, arXiv:1711.03647 [astro-ph.HE]

[467] M. Raidal, C. Spethmann, V. Vaskonen, and H. Veermäe, *Formation and Evolution of Primordial Black Hole Binaries in the Early Universe*, JCAP **02** (2019), 018, arXiv:1812.01930 [astro-ph.CO]

[468] M. Raidal, V. Vaskonen, and H. Veermäe, *Gravitational Waves from Primordial Black Hole Mergers*, JCAP **09** (2017), 037, arXiv:1707.01480 [astro-ph.CO]

[469] W. Ratzinger and P. Schwaller, *Whispers from the dark side: Confronting light new physics with NANOGrav data*, SciPost Phys. **10** (2021), 047, arXiv:2009.11875 [astro-ph.CO]

[470] T. Regimbau, *The astrophysical gravitational wave stochastic background*, Res. Astron. Astrophys. **11** (2011), 369, arXiv:1101.2762 [astro-ph.CO]

[471] T. Regimbau, M. Evans, N. Christensen, E. Katsavounidis, B. Sathyaprakash, and S. Vitale, *Digging deeper: Observing primordial gravitational waves below the binary black hole produced stochastic background*, Phys. Rev. Lett. **118** (2017), 151105, arXiv:1611.08943 [astro-ph.CO]

[472] D. Reitze, R. X. Adhikari, S. Ballmer, B. Barish, L. Barsotti, G. Billingsley, D. A. Brown, Y. Chen, D. Coyne, R. Eisenstein, M. Evans, P. Fritschel, E. D. Hall, A. Lazzarini, G. Lovelace, J. Read, B. S. Sathyaprakash, D. Shoemaker, J. Smith, C. Torrie, S. Vitale, R. Weiss, C. Wipf, and M. Zucker, *Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO* , Bull. Am. Astron. Soc. **51** (2019), 035, arXiv:1907.04833 [astro-ph.IM]

[473] R. A. Remillard and J. E. McClintock, *X-ray Properties of Black-Hole Binaries*, Ann. Rev. Astron. Astrophys. **44** (2006), 49, arXiv:astro-ph/0606352

[474] A. Renzini and C. Contaldi, *Improved limits on a stochastic gravitational-wave background and its anisotropies from Advanced LIGO O1 and O2 runs*, Phys. Rev. D **100** (2019), 063527, arXiv:1907.10329 [gr-qc]

[475] A. I. Renzini and C. R. Contaldi, *Mapping Incoherent Gravitational Wave Backgrounds*, Mon. Not. Roy. Astron. Soc. **481** (2018), 4650, arXiv:1806.11360 [astro-ph.IM]

[476] A. I. Renzini and C. R. Contaldi, *Gravitational Wave Background Sky Maps from Advanced LIGO O1 Data*, Phys. Rev. Lett. **122** (2019), 081102, arXiv:1811.12922 [astro-ph.CO]

[477] A. I. Renzini, B. Goncharov, A. C. Jenkins, and P. M. Meyers, *Stochastic Gravitational-Wave Backgrounds: Current Detection Efforts and Future Prospects*, Galaxies **10** (2022), 34, arXiv:2202.00178 [gr-qc]

[478] A. I. Renzini, J. D. Romano, C. R. Contaldi, and N. J. Cornish, *Comparison of maximum-likelihood mapping methods for gravitational-wave backgrounds*, Phys. Rev. D **105** (2022), 023519, arXiv:2107.02292 [gr-qc]

[479] L. Rezzolla, E. Barausse, E. N. Dorband, D. Pollney, C. Reisswig, J. Seiler, and S. Husa, *On the final spin from the coalescence of two black holes*, Phys. Rev. D **78** (2008), 044002, arXiv:0712.3541 [gr-qc]

[480] C. de Rham and A. J. Tolley, *Causality in curved spacetimes: The speed of light and gravity*, Phys. Rev. D **102** (2020), 084048, arXiv:2007.01847 [hep-th]

[481] A. Ricciardone, L. V. Dall'Armi, N. Bartolo, D. Bertacca, M. Liguori, and S. Matarrese, *Cross-Correlating Astrophysical and Cosmological Gravitational Wave Backgrounds with the Cosmic Microwave Background*, Phys. Rev. Lett. **127** (2021), 271301, arXiv:2106.02591 [astro-ph.CO]

[482] A. G. Riess, S. Casertano, W. Yuan, J. B. Bowers, L. Macri, J. C. Zinn, and D. Scolnic, *Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with ΛCDM*, Astrophys. J. Lett. **908** (2021), L6, arXiv:2012.08534 [astro-ph.CO]

[483] C. Ringeval, M. Sakellariadou, and F. Bouchet, *Cosmological evolution of cosmic string loops*, JCAP **02** (2007), 023, arXiv:astro-ph/0511646

[484] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, 2 ed. (Springer-Verlag, Berlin, 1989)

[485] C. L. Rodriguez, M. Morscher, B. Pattabiraman, S. Chatterjee, C.-J. Haster, and F. A. Rasio, *Binary Black Hole Mergers from Globular Clusters: Implications for Advanced LIGO*, Phys. Rev. Lett. **115** (2015), 051101, [Erratum: Phys.Rev.Lett. 116, 029901 (2016)], arXiv:1505.00792 [astro-ph.HE]

[486] J. D. Romano, *Searches for stochastic gravitational-wave backgrounds* , in *Lecture Notes of the Les Houches Summer School: Gravitational Waves* (2019), arXiv:1909.00269 [gr-qc]

[487] J. D. Romano and N. J. Cornish, *Detection methods for stochastic gravitational-wave backgrounds: a unified treatment*, Living Rev. Rel. **20** (2017), 2, arXiv:1608.06889 [gr-qc]

[488] M. Rozner, E. Grishin, Y. B. Ginat, A. P. Igoshev, and V. Desjacques, *Axion resonances in binary pulsar systems*, JCAP **03** (2020), 061, arXiv:1904.01958 [astro-ph.CO]

[489] V. N. Rudenko, *Test bodies under the effect of gravitational radiation*, Sov. Astron. **19** (1975), 270, https://ui.adsabs.harvard.edu/abs/1975SvA....19..270R

[490] R. Ruffini and J. A. Wheeler, *Introducing the black hole*, Phys. Today **24** (1971), 30

[491] S. Sachdev, T. Regimbau, and B. S. Sathyaprakash, *Subtracting compact binary foreground sources to reveal primordial gravitational-wave backgrounds*, Phys. Rev. D **102** (2020), 024051, arXiv:2002.05365 [gr-qc]

[492] R. K. Sachs, *Gravitational waves in general relativity. 8. Waves in asymptotically flat space-times*, Proc. Roy. Soc. Lond. A **270** (1962), 103

[493] R. K. Sachs and A. M. Wolfe, *Perturbations of a cosmological model and angular variations of the microwave background*, Astrophys. J. **147** (1967), 73

[494] N. Sago, K. Ioka, T. Nakamura, and R. Yamazaki, *Gravitational wave memory of gamma-ray burst jets*, Phys. Rev. D **70** (2004), 104012, arXiv:gr-qc/0405067

[495] J. Sakstein and B. Jain, *Implications of the Neutron Star Merger GW170817 for Cosmological Scalar-Tensor Theories*, Phys. Rev. Lett. **119** (2017), 251303, arXiv:1710.05893 [astro-ph.CO]

[496] P. F. de Salas and S. Pastor, *Relic neutrino decoupling with flavour oscillations revisited*, JCAP **07** (2016), 051, arXiv:1606.06986 [hep-ph]

[497] M. Sasaki, T. Suyama, T. Tanaka, and S. Yokoyama, *Primordial Black Hole Scenario for the Gravitational-Wave Event GW150914*, Phys. Rev. Lett. **117** (2016), 061101, [Erratum: Phys.Rev.Lett. 121, 059901 (2018)], arXiv:1603.08338 [astro-ph.CO]

[498] M. Sasaki, T. Suyama, T. Tanaka, and S. Yokoyama, *Primordial black holes—perspectives in gravitational wave astronomy*, Class. Quant. Grav. **35** (2018), 063001, arXiv:1801.05235 [astro-ph.CO]

[499] M. Satoh, S. Kanno, and J. Soda, *Circular Polarization of Primordial Gravitational Waves in String-inspired Inflationary Cosmology*, Phys. Rev. D **77** (2008), 023526, arXiv:0706.3585 [astro-ph]

[500] P. R. Saulson, *Fundamentals of interferometric gravitational wave detectors* (World Scientific, 1995)

[501] K. Schmitz, *New Sensitivity Curves for Gravitational-Wave Signals from Cosmological Phase Transitions*, JHEP **01** (2021), 097, arXiv:2002.04615 [hep-ph]

[502] B. F. Schutz, *Determining the Hubble Constant from Gravitational Wave Observations*, Nature **323** (1986), 310

[503] E. B. Segalis and A. Ori, *Emission of gravitational radiation from ultrarelativistic sources*, Phys. Rev. D **64** (2001), 064018, arXiv:gr-qc/0101117

[504] U. Seljak and M. Zaldarriaga, *Signature of gravity waves in polarization of the microwave background*, Phys. Rev. Lett. **78** (1997), 2054, arXiv:astro-ph/9609169

[505] P. D. Serpico, V. Poulin, D. Inman, and K. Kohri, *Cosmic microwave background bounds on primordial black holes including dark matter halo accretion*, Phys. Rev. Res. **2** (2020), 023204, arXiv:2002.10771 [astro-ph.CO]

[506] A. Sesana, N. Korsakova, M. Arca Sedda, V. Baibhav, E. Barausse, S. Barke, E. Berti, M. Bonetti, P. R. Capelo, C. Caprini, J. García-Bellido, Z. Haiman, K. Jani, O. Jennrich, P. Johansson, F. M. Khan, V. Korol, A. Lamberts, A. Lupi, A. Mangiagli, L. Mayer, G. Nardini, F. Pacucci, A. Petiteau, A. Raccanelli, S. Rajendran, J. Regan, L. Shao, A. Spallicci, N. Tamanini, M. Volonteri, N. Warburton, K. Wong, and M. Zumalacarregui, *Unveiling the Gravitational Universe at μHz Frequencies*, Exper. Astron. **51** (2021), 1333, arXiv:1908.11391 [astro-ph.IM]

[507] N. Seto, *Prospects for direct detection of circular polarization of gravitational-wave background*, Phys. Rev. Lett. **97** (2006), 151101, arXiv:astro-ph/0609504

[508] N. Seto, *Quest for circular polarization of gravitational wave background and orbits of laser interferometers in space*, Phys. Rev. D **75** (2007), 061302, arXiv:astro-ph/0609633

[509] N. Seto, *Search for Memory and Inspiral Gravitational Waves from Super-Massive Binary Black Holes with Pulsar Timing Arrays*, Mon. Not. Roy. Astron. Soc. **400** (2009), L38, arXiv:0909.1379 [astro-ph.CO]

[510] N. Seto and A. Taruya, *Measuring a Parity Violation Signature in the Early Universe via Ground-based Laser Interferometers*, Phys. Rev. Lett. **99** (2007), 121101, arXiv:0707.0535 [astro-ph]

[511] N. Seto and A. Taruya, *Polarization analysis of gravitational-wave backgrounds from the correlation signals of ground-based interferometers: Measuring a circular-polarization mode*, Phys. Rev. D **77** (2008), 103001, arXiv:0801.4185 [astro-ph]

[512] R. M. Shannon, V. Ravi, L. T. Lentati, P. D. Lasky, G. Hobbs, M. Kerr, R. N. Manchester, W. A. Coles, Y. Levin, M. Bailes, N. D. R. Bhat, S. Burke-Spolaor, S. Dai, M. J. Keith, S. Osłowski, D. J. Reardon, W. van Straten, L. Toomey, J.-B. Wang, L. Wen, J. S. B. Wyithe, and X.-J. Zhu, *Gravitational waves from binary supermassive black holes missing in pulsar observations*, Science **349** (2015), 1522, arXiv:1509.07320 [astro-ph.CO]

[513] A. Sharma and J. Harms, *Searching for cosmological gravitational-wave backgrounds with third-generation detectors in the presence of an astrophysical foreground*, Phys. Rev. D **102** (2020), 063009, arXiv:2006.16116 [gr-qc]

[514] X. Siemens, J. Creighton, I. Maor, S. Ray Majumder, K. Cannon, and J. Read, *Gravitational wave bursts from cosmic (super)strings: Quantitative analysis and constraints*, Phys. Rev. D **73** (2006), 105001, arXiv:gr-qc/0603115

[515] X. Siemens, X. Martin, and K. D. Olum, *Dynamics of cosmic necklaces*, Nucl. Phys. B **595** (2001), 402, arXiv:astro-ph/0005411

[516] J. Silk and M. J. Rees, *Quasars and galaxy formation* , Astron. Astrophys. **331** (1998), L1, arXiv:astro-ph/9801013

[517] L. Smarr, *Gravitational Radiation from Distant Encounters and from Headon Collisions of Black Holes: The Zero Frequency Limit*, Phys. Rev. D **15** (1977), 2069

[518] R. Smith and E. Thrane, *Optimal Search for an Astrophysical Gravitational-Wave Background*, Phys. Rev. X **8** (2018), 021019, arXiv:1712.00688 [gr-qc]

[519] T. L. Smith and R. Caldwell, *Sensitivity to a Frequency-Dependent Circular Polarization in an Isotropic Stochastic Gravitational Wave Background*, Phys. Rev. D **95** (2017), 044036, arXiv:1609.05901 [gr-qc]

[520] T. L. Smith and R. Caldwell, *LISA for Cosmologists: Calculating the Signal-to-Noise Ratio for Stochastic and Deterministic Sources*, Phys. Rev. D **100** (2019), 104055, arXiv:1908.00546 [astro-ph.CO]

[521] G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess, E. S. Cheng, G. De Amici, S. Gulkis, M. G. Hauser, G. Hinshaw, P. D. Jackson, M. Janssen, E. Kaita, T. Kelsall, P. Keegstra, C. Lineweaver, K. Loewenstein, P. Lubin, J. Mather, S. S. Meyer, S. H. Moseley, T. Murdock, L. Rokker, R. F. Silverberg, L. Tenorio, R. Weiss, and D. T. Wilkinson (COBE), *Structure in the COBE differential microwave radiometer first year maps*, Astrophys. J. Lett. **396** (1992), L1

[522] U. Sperhake, V. Cardoso, F. Pretorius, E. Berti, and J. A. Gonzalez, *The High-energy collision of two black holes*, Phys. Rev. Lett. **101** (2008), 161101, arXiv:0806.1738 [gr-qc]

[523] V. Springel, *et al.*, *Simulating the joint evolution of quasars, galaxies and their large-scale distribution*, Nature **435** (2005), 629, arXiv:astro-ph/0504097

[524] M. Srednicki and S. Theisen, *Nongravitational Decay of Cosmic Strings*, Phys. Lett. B **189** (1987), 397

[525] A. A. Starobinsky, *Spectrum of relict gravitational radiation and the early state of the universe*, JETP Lett. **30** (1979), 682

[526] H. Stephani, D. Kramer, M. A. H. MacCallum, C. Hoenselaers, and E. Herlt, *Exact solutions of Einstein's field equations*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, 2003)

[527] P. Stöcker, M. Krämer, J. Lesgourgues, and V. Poulin, *Exotic energy injection with ExoCLASS: Application to the Higgs portal model and evaporating black holes*, JCAP **03** (2018), 018, arXiv:1801.01871 [astro-ph.CO]

[528] M. J. Stott, T. Elghozi, and M. Sakellariadou, *Gravitational Wave Bursts from Cosmic String Cusps and Pseudocusps*, Phys. Rev. D **96** (2017), 023533, arXiv:1612.07599 [hep-th]

[529] R. L. Stratonovich, *Conditional Markov Processes and Their Application to the Theory of Optimal Control*, Modern analytic and computational methods in science and mathematics (Elsevier, 1968)

[530] A. Strominger, *Lectures on the Infrared Structure of Gravity and Gauge Theory* (Princeton University Press, 2018), arXiv:1703.05448 [hep-th]

[531] A. Strominger and A. Zhiboedov, *Gravitational Memory, BMS Supertranslations and Soft Theorems*, JHEP **01** (2016), 086, arXiv:1411.5745 [hep-th]

[532] J. Suresh, D. Agarwal, and S. Mitra, *Jointly setting upper limits on multiple components of an anisotropic stochastic gravitational-wave background*, Phys. Rev. D **104** (2021), 102003, arXiv:2106.09593 [gr-qc]

[533] J. Suresh, A. Ain, and S. Mitra, *Unified mapmaking for an anisotropic stochastic gravitational wave background*, Phys. Rev. D **103** (2021), 083024, arXiv:2011.05969 [gr-qc]

[534] L. Susskind, *Electromagnetic Memory* (2015), arXiv:1507.02584 [hep-th]

[535] L. B. Szabados, *Quasi-Local Energy-Momentum and Angular Momentum in General Relativity*, Living Rev. Rel. **12** (2009), 4

[536] T. Takahashi and J. Soda, *Chiral Primordial Gravitational Waves from a Lifshitz Point*, Phys. Rev. Lett. **102** (2009), 231301, arXiv:0904.0554 [hep-th]

[537] C. Talbot, E. Thrane, P. D. Lasky, and F. Lin, *Gravitational-wave memory: waveforms and phenomenology*, Phys. Rev. D **98** (2018), 064031, arXiv:1807.00990 [astro-ph.HE]

[538] M. Tegmark, A. Taylor, and A. Heavens, *Karhunen-Loeve eigenvalue problems in cosmology: How should we tackle large data sets?*, Astrophys. J. **480** (1997), 22, arXiv:astro-ph/9603021

[539] A. C. Thompson, *Dynamics of Cosmic String*, Phys. Rev. D **37** (1988), 283

[540] K. S. Thorne, *Nonspherical Gravitational Collapse: A Short Review* (W. H. Freeman, San Francisco, 1972), 231–258

[541] K. S. Thorne, *Disk accretion onto a black hole. 2. Evolution of the hole.*, Astrophys. J. **191** (1974), 507

[542] K. S. Thorne, *Gravitational-wave bursts with memory: The Christodoulou effect*, Phys. Rev. D **45** (1992), 520

[543] E. Thrane, S. Ballmer, J. D. Romano, S. Mitra, D. Talukder, S. Bose, and V. Mandic, *Probing the anisotropies of a stochastic gravitational-wave background using a network of ground-based laser interferometers*, Phys. Rev. D **80** (2009), 122002, arXiv:0910.0858 [astro-ph.IM]

[544] E. Thrane, N. Christensen, and R. Schofield, *Correlated magnetic noise in global networks of gravitational-wave interferometers: observations and implications*, Phys. Rev. D **87** (2013), 123009, arXiv:1303.2613 [astro-ph.IM]

[545] E. Thrane and J. D. Romano, *Sensitivity curves for searches for gravitational-wave backgrounds*, Phys. Rev. D **88** (2013), 124032, arXiv:1310.5300 [astro-ph.IM]

[546] P. Tisserand, L. Le Guillou, C. Afonso, J. N. Albert, J. Andersen, R. Ansari, É. Aubourg, P. Bareyre, J. P. Beaulieu, X. Charlot, C. Coutures, R. Ferlet, P. Fouqué, J. F. Glicenstein, B. Goldman, A. Gould, D. Graff, M. Gros, J. Haissinski, C. Hamadache, J. de Kat, T. Lasserre, É. Lesquoy, C. Loup, C. Magneville, J. B. Marquette, É. Maurice, A. Maury, A. Milsztajn, M. Moniez, N. Palanque-Delabrouille, O. Perdereau, Y. R. Rahal, J. Rich, M. Spiro, A. Vidal-Madjar, L. Vigroux, and S. Zylberajch (EROS-2), *Limits on the Macho Content of the Galactic Halo from the EROS-2 Survey of the Magellanic Clouds*, Astron. Astrophys. **469** (2007), 387, arXiv:astro-ph/0607207

[547] V. Tiwari and S. Fairhurst, *The Emergence of Structure in the Binary Black Hole Mass Distribution*, Astrophys. J. Lett. **913** (2021), L19, arXiv:2011.04502 [astro-ph.HE]

[548] A. Tolish, L. Bieri, D. Garfinkle, and R. M. Wald, *Examination of a simple example of gravitational wave memory*, Phys. Rev. D **90** (2014), 044060, arXiv:1405.6396 [gr-qc]

[549] A. Tolish and R. M. Wald, *Retarded Fields of Null Particles and the Memory Effect*, Phys. Rev. D **89** (2014), 064008, arXiv:1401.5831 [gr-qc]

[550] M. Turner, *Gravitational radiation from point-masses in unbound orbits: Newtonian results*, Astrophys. J. **216** (1977), 610

[551] M. Turner and C. M. Will, *Post-Newtonian gravitational bremsstrahlung*, Astrophys. J. **220** (1978), 1107

[552] M. S. Turner, *Gravitational Radiation from Supernova Neutrino Bursts*, Nature **274** (1978), 565

[553] M. S. Turner, *Influence of a weak gravitational wave on a bound system of two point-masses*, Astrophys. J. **233** (1979), 685, `https://ui.adsabs.harvard.edu/abs/1979ApJ...233..685T`

[554] N. Turok, *Grand Unified Strings and Galaxy Formation*, Nucl. Phys. B **242** (1984), 520

[555] C. S. Unnikrishnan, *IndIGO and LIGO-India: Scope and plans for gravitational wave research and precision metrology in India*, Int. J. Mod. Phys. D **22** (2013), 1341010, arXiv:1510.06059 [physics.ins-det]

[556] S. A. Usman, A. H. Nitz, I. W. Harry, C. M. Biwer, D. A. Brown, M. Cabero, C. D. Capano, T. Dal Canton, T. Dent, S. Fairhurst, M. S. Kehl, D. Keppel, B. Krishnan, A. Lenon, A. Lundgren, A. B. Nielsen, L. P. Pekowsky, H. P. Pfeiffer, P. R. Saulson, M. West, and J. L. Willis, *The PyCBC search for gravitational waves from compact binary coalescence*, Class. Quant. Grav. **33** (2016), 215004, arXiv:1508.02357 [gr-qc]

[557] T. Vachaspati and A. Vilenkin, *Gravitational Radiation from Cosmic Strings*, Phys. Rev. D **31** (1985), 3052

[558] S. Vagnozzi, *Implications of the NANOGrav results for inflation*, Mon. Not. Roy. Astron. Soc. **502** (2021), L11, arXiv:2009.13432 [astro-ph.CO]

[559] L. Valbusa Dall'Armi, A. Ricciardone, N. Bartolo, D. Bertacca, and S. Matarrese, *Imprint of relativistic particles on the anisotropies of the stochastic gravitational-wave background*, Phys. Rev. D **103** (2021), 023522, arXiv:2007.01215 [astro-ph.CO]

[560] E. Vangioni, K. A. Olive, T. Prestegard, J. Silk, P. Petitjean, and V. Mandic, *The Impact of Star Formation and Gamma-Ray Burst Rates at High Redshift on Cosmic Chemical Evolution and Reionization*, Mon. Not. Roy. Astron. Soc. **447** (2015), 2575, arXiv:1409.2462 [astro-ph.GA]

[561] V. Vaskonen and H. Veermäe, *Did NANOGrav see a signal from primordial black hole formation?*, Phys. Rev. Lett. **126** (2021), 051303, arXiv:2009.07832 [astro-ph.CO]

[562] J. P. W. Verbiest, *et al.*, *The International Pulsar Timing Array: First Data Release*, Mon. Not. Roy. Astron. Soc. **458** (2016), 1267, arXiv:1602.03640 [astro-ph.IM]

[563] L. Verde, T. Treu, and A. G. Riess, *Tensions between the Early and the Late Universe*, Nature Astron. **3** (2019), 891, arXiv:1907.10625 [astro-ph.CO]

[564] A. Vilenkin, *Gravitational Field of Vacuum Domain Walls and Strings*, Phys. Rev. D **23** (1981), 852

[565] A. Vilenkin, *Cosmic Strings and Domain Walls*, Phys. Rept. **121** (1985), 263

[566] A. Vilenkin, Y. Levin, and A. Gruzinov, *Cosmic strings and primordial black holes*, JCAP **11** (2018), 008, arXiv:1808.00670 [astro-ph.CO]

[567] A. Vilenkin and E. P. S. Shellard, *Cosmic Strings and Other Topological Defects*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, 2000)

[568] G. Vincent, N. D. Antunes, and M. Hindmarsh, *Numerical simulations of string networks in the Abelian Higgs model*, Phys. Rev. Lett. **80** (1998), 2277, arXiv:hep-ph/9708427

[569] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, Nature Methods **17** (2020), 261, https://rdcu.be/b08Wh

[570] J. M. Wachter and K. D. Olum, *Gravitational smoothing of kinks on cosmic string loops*, Phys. Rev. Lett. **118** (2017), 051301, [Erratum: Phys.Rev.Lett. 121, 149901 (2018)], arXiv:1609.01153 [gr-qc]

[571] J. M. Wachter and K. D. Olum, *Gravitational backreaction on piecewise linear cosmic string loops*, Phys. Rev. D **95** (2017), 023519, arXiv:1609.01685 [gr-qc]

[572] R. V. Wagoner, *Low Frequency Gravitational Radtiation From Collapsing Systems*, Phys. Rev. D **19** (1979), 2897

[573] L. A. Wainstein and V. D. Zubakov, *Extraction of Signals from Noise*, Dover books on physics and mathematical physics (Prentice-Hall, Englewood Cliffs, NJ, 1962)

[574] R. M. Wald, *General Relativity* (University of Chicago Press, Chicago, USA, 1984)

[575] R. M. Wald, *Gravitational collapse and cosmic censorship*, in *Black Holes, Gravitational Radiation and the Universe: Essays in Honor of C. V. Vishveshwara* (Springer Netherlands, Dordrecht, 1999), 69–86, arXiv:gr-qc/9710068, https://doi.org/10.1007/978-94-017-0934-7_5

[576] A. Wang, Q. Wu, W. Zhao, and T. Zhu, *Polarizing primordial gravitational waves by parity violation*, Phys. Rev. D **87** (2013), 103512, arXiv:1208.5490 [astro-ph.CO]

[577] J. B. Wang, G. Hobbs, W. Coles, R. M. Shannon, X. J. Zhu, D. R. Madison, M. Kerr, V. Ravi, M. J. Keith, R. N. Manchester, Y. Levin, M. Bailes, N. D. R. Bhat, S. Burke-Spolaor, S. Dai, S. Osłowski, W. van Straten, L. Toomey, N. Wang, and L. Wen, *Searching for gravitational wave memory bursts with the Parkes Pulsar Timing Array*, Mon. Not. Roy. Astron. Soc. **446** (2015), 1657, arXiv:1410.3323 [astro-ph.GA]

[578] S. Wang and K. Kohri, *Probing Primordial Black Holes with Angular Power Spectrum for Anisotropies in Stochastic Gravitational-Wave Background* (2021), arXiv:2107.01935 [gr-qc]

[579] S. Wang, Y.-F. Wang, Q.-G. Huang, and T. G. F. Li, *Constraints on the Primordial Black Hole Abundance from the First Advanced LIGO Observation Run Using the Stochastic Gravitational-Wave Background*, Phys. Rev. Lett. **120** (2018), 191102, arXiv:1610.08725 [astro-ph.CO]

285

Bibliography

[580] Y. Wang, S. D. Mohanty, and Z. Cao, *Extending the frequency reach of pulsar timing array based gravitational wave search without high cadence observations*, Astrophys. J. Lett. **907** (2021), L43, arXiv:2012.04261 [astro-ph.HE]

[581] Y. Wang, K. Pardo, T.-C. Chang, and O. Doré, *Gravitational Wave Detection with Photometric Surveys*, Phys. Rev. D **103** (2021), 084007, arXiv:2010.02218 [gr-qc]

[582] J. Weber, *Detection and Generation of Gravitational Waves*, Phys. Rev. **117** (1960), 306

[583] J. Weber, *Evidence for discovery of gravitational radiation*, Phys. Rev. Lett. **22** (1969), 1320

[584] S. Weinberg, *Cosmology* (Oxford University Press, 2008)

[585] J. M. Weisberg and Y. Huang, *Relativistic Measurements from Timing the Binary Pulsar PSR B1913+16*, Astrophys. J. **829** (2016), 55, arXiv:1606.02744 [astro-ph.HE]

[586] J. M. Weisberg, D. J. Nice, and J. H. Taylor, *Timing Measurements of the Relativistic Binary Pulsar PSR B1913+16*, Astrophys. J. **722** (2010), 1030, arXiv:1011.0718 [astro-ph.GA]

[587] R. Weiss, *Electromagnetically coupled broadband gravitational antenna* (1972), unpublished technical report, https://dcc.ligo.org/LIGO-P720002/public

[588] J. A. Wheeler and K. Ford, *Geons, black holes, and quantum foam: A life in physics* (W. W. Norton, 1998)

[589] A. G. Wiseman and C. M. Will, *Christodoulou's nonlinear gravitational wave memory: Evaluation in the quadrupole approximation*, Phys. Rev. D **44** (1991), 2945

[590] D. Wysocki, J. Lange, and R. O'Shaughnessy, *Reconstructing phenomenological distributions of compact binaries via gravitational wave observations*, Phys. Rev. D **100** (2019), 043012, arXiv:1805.06442 [gr-qc]

[591] H. Yang and D. Martynov, *Testing Gravitational Memory Generation with Compact Binary Mergers*, Phys. Rev. Lett. **121** (2018), 071102, arXiv:1803.02429 [gr-qc]

[592] K. Z. Yang, V. Mandic, C. Scarlata, and S. Banagiri, *Searching for Cross-Correlation Between Stochastic Gravitational Wave Background and Galaxy Number Counts*, Mon. Not. Roy. Astron. Soc. **500** (2020), 1666, arXiv:2007.10456 [astro-ph.CO]

[593] N. Yonemaru, S. Kuroyanagi, G. Hobbs, K. Takahashi, X.-J. Zhu, W. A. Coles, S. Dai, E. Howard, R. Manchester, D. Reardon, C. Russell, R. M. Shannon, N. Thyagarajan, R. Spiewak, and J.-B. Wang, *Searching for gravitational wave bursts from cosmic string cusps with the Parkes Pulsar Timing Array*, Mon. Not. Roy. Astron. Soc. **501** (2021), 701, arXiv:2011.13490 [gr-qc]

[594] N. Yunes, K. Yagi, and F. Pretorius, *Theoretical Physics Implications of the Binary Black-Hole Mergers GW150914 and GW151226*, Phys. Rev. D **94** (2016), 084002, arXiv:1603.08955 [gr-qc]

[595] Y. B. Zel'dovich and A. G. Polnarev, *Radiation of gravitational waves by a cluster of superdense stars*, Sov. Astron. **18** (1974), 17, `https://ui.adsabs.harvard.edu/abs/1974SvA....18...17Z/abstract`

[596] Ya. B. Zel'dovich and I. D. Novikov, *The Hypothesis of Cores Retarded during Expansion and the Hot Cosmological Model*, Sov. Astron. **10** (1967), 602, `https://ui.adsabs.harvard.edu/abs/1967SvA....10..602Z`

[597] L. Zhang, X. Chen, M. Kamionkowski, Z.-G. Si, and Z. Zheng, *Constraints on radiative dark-matter decay from the cosmic microwave background*, Phys. Rev. D **76** (2007), 061301, arXiv:0704.2444 [astro-ph]