MONASH University

DOCTORAL THESIS

Astrophysics of Binary Black Holes at the Dawn of Gravitational-Wave Astronomy

Author:
Colm TALBOT

Supervisor:
Ass. Prof. Eric THRANE

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

School of Physics and Astronomy

February 25, 2020

Copyright Notice

Publications

1. **C. Talbot**, E. Thrane, *Determining the population properties of spinning black holes*, Phys. Rev. D, 96, 023012, arXiv:1704.08370
2. **C. Talbot**, E. Thrane, *Measuring the binary black hole mass spectrum with an astrophysically motivated parameterization*, ApJ., 858, 2, arXiv:1801.02699
3. **C. Talbot**, E. Thrane, P. Lasky, F. Lin, *Gravitational-wave memory: waveforms and phenomenology*, Phys. Rev. D, 98, 0.64031, arXiv:1807.00990
4. E. Thrane, **C. Talbot**, *An introduction to Bayesian inference in gravitational-wave astronomy: parameter estimation, model selection, and hierarchical models*, PASA, 36, E010, arXiv:1809.02293
5. G. Ashton, M. Hübner, P. Lasky, **C. Talbot**, *et al.*, *Bilby: A user-friendly Bayesian inference library for gravitational-wave astronomy*, ApJS., 241, 27, arXiv:1811.02042
6. LIGO/Virgo Collaboration inc. **C. Talbot**, *Binary Black Hole Population Properties Inferred from the First and Second Observing Runs of Advanced LIGO and Advanced Virgo*, ApJL, 882, 2, arXiv:1811.12940
7. LIGO/Virgo Collaboration inc. **C. Talbot**, *Directional limits on persistent gravitational waves using data from Advanced LIGO's first two observing runs*, Phys. Rev. D, 100, 062001, arXiv:1903.08844
8. **C. Talbot**, R. Smith, E. Thrane,, G. Poole, *Parallelized Inference for Gravitational-Wave Astronomy*, Phys. Rev. D, 100, 043030, arXiv:1904.02863
9. E. Payne, **C. Talbot**, E. Thrane, *Higher order gravitational-wave modes with likelihood reweighting*, Phys. Rev. D, 100, 123017 arXiv:1905.05477
10. S. Banagiri, M.Coughlin, J. Clark, P. Lasky, M. A. Bizouard, **C. Talbot**, Eric Thrane, V. Mandic, *Constraining the Gravitational-Wave Afterglow From a Binary Neutron Star Coalescence*, MNRAS, 492, 4, 4945-4951, arxiv:1909.01934
11. F. Hernandez Vivanco, R. Smith, E. Thrane, P. Lasky, **C. Talbot**, V. Raymond, *Measuring the neutron star equation of state with gravitational waves: the first forty binary neutron star mergers*, Phys. Rev. D, 100, 103009, arxiv:1909.02698
12. M. Hübner, **C. Talbot**, P. Lasky, E. Thrane, *Measuring gravitational-wave memory in the first LIGO/Virgo gravitational-wave transient catalog* Phys. Rev. D, 101, 023011 arxiv:1911.12496
13. S. Galaudage, **C. Talbot**, E. Thrane, *Gravitational-wave inference in the catalog era: evolving priors and marginal events* arxiv:1912.09708

I am also an author on many other collaboration papers not listed here for which my contributions vary.

Declaration of Authorship

I, Colm Talbot, hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

The core theme of the thesis is astrophysical inference with gravitational-wave transients. This thesis includes six papers which have been previously published in peer-reviewed journals in Chapters 2, 3, 4, 5, 6, and 7. The ideas, development and writing up of Chapters 3, 4, 5, and 7 in the thesis were the principal responsibility of myself, the student, working within the School of Physics and Astronomy under the supervision of Ass. Prof. Eric Thrane. Chapter 2 was jointly prepared for an invited lecture given by Eric Thrane and subsequently published as a review article. Chapter 6 is the result of substantial code development by myself, Gregory Ashton, and Moritz Hübner, initiated by Paul Lasky, with many other contributors. Due to the collaborative nature of this project it is not examinable under the university rules, however, it is included to preserve the narrative of the thesis.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research. My contribution to each chapter published separately is summarised in the following table.

Signed:

Date: February 25, 2020

I, Ass. Prof. Eric Thrane, hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Signed:

Date: February 25, 2020

Thesis chapter	Publication title	Status	Student contribution	Co-author contribution	Monash student co-authors
2	An introduction to Bayesian inference in gravitational-wave astronomy: parameter estimation, model selection, and hierarchical models	Published (PASA, 2019)	50% - writing	E. Thrane - 50% writing	N
3	Determining the population properties of spinning black holes	Published (PRD, 2017)	80% - method, code development, analysis, writing	E. Thrane - 20% - concept, writing	N
4	Measuring the binary black hole mass spectrum with an astrophysically motivated parameterization	Published (ApJ, 2018)	90% - concept, method, code development, analysis, writing	E. Thrane - 10% - writing	N
5	Gravitational-wave memory: waveforms and phenomenology	Published (PRD, 2018)	75% - concept, method, code development, analysis, writing	E. Thrane/P. Lasky - 20% - concept, writing; F. Lin - 5% - code development	N
7	Parallelized Inference for Gravitational-Wave Astronomy	Published (PRD, 2019)	80% - concept, code development, analysis, writing	R. Smith - 5% - concept, writing; E. Thrane - 5% - writing; G. Poole - 10% - code development	N

This was a triumph.
I’m making a note here; “huge success.”
It’s hard to overstate my satisfaction.

Jonathan Coulton, *Still Alive*

MONASH UNIVERSITY

Abstract

Faculty of Science
School of Physics and Astronomy

Doctor of Philosophy

Astrophysics of Binary Black Holes at the Dawn of Gravitational-Wave Astronomy

by Colm TALBOT

On the 14th of September 2015 gravitational waves from the merger of two black holes (Abbott, 2016h) observed by advanced LIGO (Aasi, 2015) marked the beginning of the era of gravitational-wave astronomy. This first detection, along with further observations by advanced LIGO and Virgo (Acernese, 2015), have revealed a previously unobserved population of black holes tens of times more massive than the sun. In this thesis, I describe how we can build models for the distribution of the masses and spins of binary black holes to gain insights into how massive stars end their lives. In addition to the astrophysical information these events provide, observations of merging black hole binaries provide the best possible tests of strong field effects of general relativity. I introduce a method to compute the gravitational-wave memory effect. By constructing more accurate models for the coalescence of compact binaries, we can enhance our understanding of extreme dynamics in general relativity. I also present new, fast, and flexible software for astrophysical inference, which, combined with the theoretical models developed in this thesis, will allow us to make new scientific discoveries in the burgeoning field of gravitational-wave astronomy.

Copyright Notice

My first thanks go to Eric for giving me this opportunity and enabling my research at every step. I didn't know what to expect when beginning my PhD, but I couldn't have hoped for a more encouraging and fulfilling environment to pursue my research.

I should, however, not neglect all of the other scientists who have shaped me as a researcher. I have benefitted hugely from working as part of the Monash gravitational-wave group, and the wider Monash astrophysics community. Additionally, I should mention all the members of OzGrav and the LIGO/Virgo scientific collaborations for providing such a wide-ranging and open collaborative environment. Having access to such a wealth of information has been invaluable to my education.

Beyond all of the scientific meetings and discussions are the social interactions that came with them and the friends I've made along the way. Especially everyone I have shared an office with for providing just the right combination of quiet working, lively discussion, and walks for coffee.

I am immensely grateful to everyone I've sung with during my stay in Australia, thank you for providing such enjoyable relief from the academic world. Mark, Michael, George, and the rest of the Melbournaires, it has really been a pleasure to sing with you all.

Of course, I must acknowledge the role my family played in getting me where I am today. My parents, who always put my education and well-being first, and Domhnall who is always there when needed. Without your support I wouldn't even have been able to begin this research.

Finally, to Sylvia, for making the difficult times easier, the easy times more enjoyable, and always inspiring me to be better.

This research was supported by a Monash Graduate Scholarship (MGS) and a Monash International Postgraduate Research Scholarship (MIPRS). I also received additional funding from OzGrav, the ARC centre of excellence for gravitational-wave discovery.

Contents

For all the stars that died to make this research possible.

Chapter 1

Introduction

The direct detection of gravitational waves from a binary black hole merger by Advanced LIGO (Aasi, 2015) on the 14th of September 2015 marked the beginning of the era of gravitational-wave astronomy (Abbott, 2016h). Following this, nine additional binary black hole mergers and one binary neutron star merger were seen during the first two observing runs of Advanced LIGO/Advanced Virgo (Acernese, 2015; Abbott, 2019b). Searches for compact binary coalescences by groups outside of the LIGO/Virgo collaborations have independently identified these events and identified a new set of binaries (e.g., Nitz et al., 2018a; Zackay et al., 2019; Venumadhav et al., 2019a; Nitz et al., 2019). These observations have enabled the first measurement of the population of binary black holes in the local Universe (Abbott, 2019a), a new, independent, measurement of the Hubble constant (Abbott, 2017a), and new tests of general relativity (e.g., Abbott, 2016k; Abbott, 2019d; Abbott, 2019e; Isi et al., 2019; Hübner et al., 2020).

1.1 Bayesian Inference

Bayesian inference is how we use observations and experiments to understand the Universe. It relies on Bayes' theorem, which was first implied by Thomas Bayes in a private letter first published posthumously (Bayes, 1763) and expanded upon in Laplace, 1774. However the familiar form

$$p(a|b) = \frac{p(b|a)p(a)}{p(b)} \tag{1.1}$$

was first written a century ago by Dorothy Wrinch and Harold Jeffreys (Wrinch & Jeffreys, 1919; Wrinch & Jeffreys, 1921; Wrinch & Jeffreys, 1923). Bayes' theorem tells us that the probability of one event a given another event b has happened is entirely determined by the probabilities of event b given a, event b, and event a.

In the context of scientific inference we often write this equation in the form

$$p(\theta|d, \mathcal{M}) = \frac{\mathcal{L}(d|\theta, \mathcal{M})\pi(\theta|\mathcal{M})}{\mathcal{Z}(d|\mathcal{M})}. \tag{1.2}$$

I have introduced a new quantity $\mathcal{M}$ to the right hand side of each term. This is our model for the data[1][2][3]. Here we have identified the quantity we want to learn about as θ, the parameters describing our model, e.g., black hole masses, or the maximum mass of black holes, and the quantity that informs our inference as our data, d, e.g., measurements of strain in interferometers. We have also relabeled a number of the terms:

- $p(\theta|d, \mathcal{M})$ is the *posterior* probability of the parameters given the data and our model.
- $\mathcal{L}(d|\theta, \mathcal{M})$ is the *likelihood* of obtaining the data given our model and parameters.
- $\pi(\theta|\mathcal{M})$ is the *prior* probability of the parameters given our model.
- $\mathcal{Z}(d|\mathcal{M})$ is the *evidence* for the data given the model.

For a detailed introduction to Bayesian inference for gravitational-wave astronomy, including definitions of these terms, see Chapter 2. For a general statistical introduction I suggest, e.g., Gelman et al., 2013, and for a history of Bayes' theorem, e.g., Dale, 1982; McGrayne, 2011.

1.2 Gravitational Waveforms

Aside from data, the other essential ingredient for performing inference is the model. When analysing gravitational-wave transients this is the expected effect of gravitational radiation on our interferometers. This consists of two effects:

- the displacement of test masses due to the passage of the gravitational radiation. This is often expressed in terms of the polarisation states $h_{+,\times}$[4].
- the response of the detector to the gravitational wave.

Compact binary coalescences are typically described in terms of three phases: "inspiral" when the two bodies orbit each other, "merger" when the two bodies collide, "ringdown" (sometimes referred to as "post-merger" for binary neutron star inspirals) where the remnant returns to a stationary state, i.e., a rotating, Kerr, black hole. Computing the expected waveform (often referred to as a "template") requires solving the general relativistic

[1] When performing model selection we can invoke a prior on a set of different models to perform model comparison, see, Chapter 2.

[2] Some authors include an additional term $\mathcal{I}$ to represent all of the prior information we implicitly include in our inference.

[3] $\mathcal{H}$ is sometimes used to represent a hypothesis rather than $\mathcal{M}$ for model.

[4] In general relativity only these two "tensor" polarisations are allowed, rather than the full six allowed polarisations allowed in general metric theories of gravity (Nishizawa et al., 2009).

wave equation for the source. However, this waveform cannot be analytically solved through inspiral, merger, and ringdown. Therefore, a range of approximations are used to compute the gravitational radiation. For compact binary coalescences, these range from perturbative expansions about the Newtonian limit in the parameterised post-Newtonian formalism, see, e.g., Blanchet, 2014, or about analytic solutions of general relativity, e.g., quasi-normal ringdown (e.g., Chandrasekhar & Detweiler, 1975; Berti, Cardoso & Starinets, 2009), to the effective one-body approximation for binary systems (e.g., Damour, 2014), and numerical relativity (e.g., Lehner & Pretorius, 2014), computational integration of the relativistic equations of motion.

Each of these approaches has advantages and disadvantages. The expansion based methods provide analytic approximations to the waveform which can, in turn, be Fourier transformed to give closed-form frequency-domain expressions for the waveform allowing for efficient parallelised evaluation of the waveform[5], see, Chapter 7. However, neither of these approximations can model the full waveform.

The inspiral phase is well described by the post-Newtonian expansion. This consists of an expansion in the orbital speed, (v/c)[6], of the components of the binary. However, this is not a convergent series and therefore inevitably breaks down as the binary becomes increasingly relativistic, i,e., during the merger. For a detailed review of the post-Newtonian formalism in the context of binary inspirals see, e.g., Blanchet, 2014. Since the coefficient of each of the post-Newtonian parameters is determined by general relativity, measuring these coefficients is a powerful test of general relativity that has been performed on solar system effects (e.g., Ni, 2017), radio pulsars (e.g., Perrodin & Sesana, 2018), and gravitational-wave sources (e.g., Abbott, 2019e).

After the merger of two black holes, the remnant sheds its "hair" to yield a Kerr black hole during the ringdown phase. In general relativity, the exact frequencies of the ringdown are entirely determined by the mass and spin of the final black hole by the "no-hair" theorem (Israel, 1968; Carter, 1971), and the amplitude and phase by the binary components. Gravitational-wave observations have already allowed new tests of this theorem (e.g., Isi et al., 2019). While this approximation predicts the post-merger emission, it cannot describe the pre-merger phase.

Numerical relativity provides the most accurate approximation we have of the true waveform by numerically solving the equations of motion. However, due to the computational difficulty in performing numerical relativity simulations, generating each waveform takes currently months of computing time.

One limiting factor in the accuracy of gravitational waveforms computed using numerical relativity is determined by the waveform extraction method.

[5]As discussed in Chapter 2 the likelihood typically used to analyse gravitational-wave transients are expressed in the frequency domain due to the convenient description of the noise in terms of a power spectral density.

[6]Because the first correction to the Newtonian (denoted 1PN) equations of motion is proportional to $(v/c)^2$, the expansion is in practice in $(v/c)^2$

Extracting the strain observed by a distant observer from the numerical relativity simulation involves evolving the wave from the wave zone, just outside the merger, to the asymptotic limit. A range of methods are used for this, however, the most commonly used methods are not able to extract the gravitational-wave memory signal, see, e.g. Pollney & Reisswig, 2011; Bishop & Rezzolla, 2016. Gravitational-wave memory is a permanent displacement of freely-falling test masses at infinity (Christodoulou, 1991) sourced by the emitted gravitational radiation. This means that any waveforms based on numerical relativity simulations are missing this term. In Chapter 5, I present a method for calculating the memory signal; a similar approach has been used for calculating binary black hole "merger kicks" (e.g., Gerosa, Hébert & Stein, 2018).

One popular approach to combine these different methods is to "stitch" together a post-Newtonian inspiral, a parameterised fit to the merger signal, and a quasi-normal ringdown with the IMRPhenom family of waveform models (e.g., Hannam et al., 2014; Schmidt, Ohme & Hannam, 2015; Khan et al., 2016; Khan et al., 2019; London et al., 2018). These allow rapid evaluation of a full inspiral-merger-ringdown waveform directly in the frequency domain and have been critical in facilitating past astrophysical inference.

A range of interpolation methods has been developed to rapidly generate high-fidelity gravitational-wave templates at accuracies approaching that of numerical relativity simulations. Surrogate models have found widespread use in gravitational-wave data analysis by interpolating numerical relativity (e.g., Blackman et al., 2017; Blackman et al., 2017; Varma et al., 2019), effective one-body (e.g., Field et al., 2014; Pürrer, 2016), and analytic waveforms (e.g., Pürrer, 2014; Canizares et al., 2015; Smith et al., 2016; Zackay, Dai & Venumadhav, 2018) at known times or frequencies. Gaussian process regression has been employed on a range of waveform models (e.g., Doctor et al., 2017; Huerta et al., 2018). This approach has the advantage that it returns a measure of the uncertainty in the waveform estimate, which is typically neglected. Easter et al., 2019 interpolated the post-merger spectrum of gravitational-waves from a binary neutron star inspiral. Recently Tiglio & Villanueva, 2019 employed symbolic regression on numerical relativity simulations to obtain closed-form expressions for the full inspiral-merger-ringdown of non-spinning binary black holes.

The response of detectors to gravitational waves is, by comparison, much simpler, depending only on geometric terms (e.g., Anderson et al., 2001), although, for long-duration signals the motion of the Earth can complicate the analysis (e.g., Giampieri, 1997; Essick, Vitale & Evans, 2017).

1.3 Astrophysical Population Modelling

While the data for astrophysical population inference is still gravitational-wave strain, we require an extra layer of modeling. Unlike in the case of the gravitational waveform simulation, where we know exactly the 16 parameters which describe a binary black hole waveform[7], we have less prior knowledge of what the model should look like for the distribution of masses and spins[8]. Building models for these distributions is the main focus of Chapters 3 and 4.

The remainder of the thesis is structured as follows. Chapter 2 gives a detailed introduction to Bayesian inference for gravitational-wave astronomy. This covers many of the methods used in the subsequent chapters. Chapters 3 and 4 introduce phenomenological models for the distribution of black hole masses and spins, how we can use these models to extract specific physics, and how many observations we need before we can measure these features. In Chapter 5, I develop a method to compute the expected signal due to gravitational-wave memory from waveforms that omit this term. Chapters 6 and 7 discuss computational frameworks for performing astrophysical inference. I present some closing thoughts on the current state of astrophysical inference with gravitational waves including a review of current results, and future possible directions for astrophysical inference in Chapter 8. In Appendix A, I discuss the effect of applying a time-domain window to gravitational-wave data and how to avoid potential biases in parameter estimation using windowed data.

As Chapters 2-7 are reproductions of previously published papers a number of the literature references and discussion of the observed catalogue of compact binaries are somewhat out of date. Rather than modifying these chapters, I direct readers to this chapter or Chapter 8 for discussions which reflect the state of the field at the time of writing. Additionally, the subject of Bayesian inference for analysing single gravitational-wave transients and astrophysical populations is introduced multiple times throughout the thesis. In order to preserve the flow of those chapters, and enable them to be taken out of context, I have not removed those sections. The introduction in the following chapter should be taken as the authoritative version and those sections in subsequent can safely be skipped by readers reading all chapters in order.

[7]There are nine "intrinsic" parameters which describe the binary components and orbit (two masses, two three-dimensional spin vectors, and orbital eccentricity), and seven "extrinsic" parameters which describe the relative position and orientation of the binary and the detector (distance, two-dimensional sky position, orbital inclination, signal polarisation, and a reference time and orbital phase).

[8]There are more exotic scenarios which require a larger number of parameters to describe a merging binary, e.g, parameterised deviations from general relativity (e.g., Abbott, 2019e)

Chapter 2

An introduction to Bayesian inference in gravitational-wave astronomy: parameter estimation, model selection, and hierarchical models

Published as:
Thrane, E. & Talbot, C., Publ. Astron. Soc. Aust. 36, e010 (2019).

Abstract

This is an introduction to Bayesian inference with a focus on hierarchical models and hyper-parameters. We write primarily for an audience of Bayesian novices, but we hope to provide useful insights for seasoned veterans as well. Examples are drawn from gravitational-wave astronomy, though we endeavor for the presentation to be understandable to a broader audience. We begin with a review of the fundamentals: likelihoods, priors, and posteriors. Next, we discuss Bayesian evidence, Bayes factors, odds ratios, and model selection. From there, we describe how posteriors are estimated using samplers such as Markov Chain Monte Carlo algorithms and nested sampling. Finally, we generalize the formalism to discuss hyper-parameters and hierarchical models. We include extensive appendices discussing the creation of credible intervals, Gaussian noise, explicit marginalization, posterior predictive distributions, and selection effects.

2.1 Preface: why study Bayesian inference?

Bayesian inference is an essential part of modern astronomy. It finds particularly elegant application in the field of gravitational-wave astronomy thanks to the clear predictions of general relativity and the extraordinary simplicity with which compact binary systems are described. An astrophysical black hole is completely characterized by just its mass and its dimensionless spin vector. The gravitational waveform from a black hole binary is typically characterized by just fifteen parameters. Since sources of gravitational waves are so simple, and since we have a complete theory describing how they emit gravitational waves, there is a direct link between data and model. The significant interest in Bayesian inference within the gravitational-wave community reflects the great possibilities of this area of research.

Bayesian inference and parameter estimation are the tools that allow us to make statements about the Universe based on data. In gravitational-wave astronomy, Bayesian inference is the tool that allows us to reconstruct sky maps of where a binary neutron star merged (Abbott, 2017i), to determine that GW170104 merged 880^{+450}_{-390} Mpc away from Earth (Abbott, 2017f), and that the black holes in GW150914 had masses of $35^{+5}_{-3}\,\mathrm{M}_{\odot}$ and $33^{+3}_{-4}\,\mathrm{M}_{\odot}$ (Abbott, 2016i). We use it to determine the Hubble constant (Abbott, 2017a), to study the formation mechanism of black hole binaries (Vitale & Evans, 2017; Stevenson, Berry & Mandel, 2017; Talbot & Thrane, 2017; Gerosa & Berti, 2017; Farr et al., 2017; Wysocki et al., 2018; Lower et al., 2018), and to probe how stars die (Fishbach, Holz & Farr, 2018; Talbot et al., 2018; Abbott, 2019a). Increasingly, Bayesian inference and parameter estimation are the language of gravitational-wave astronomy. In this note, we endeavor to provide a primer on Bayesian inference with examples from gravitational-wave astronomy[1].

Before beginning, we highlight additional resources, useful for researchers interested in Bayesian inference in gravitational-wave astronomy. Sivia & Skilling, 2006 and Gregory, 2005 are useful references that are accessible to physicists and astronomers; see also the Springer Series in Astrostatistics (Manuel et al., 2012; Hilbe, 2013; Chattopadhyay & Chattopadhyay, 2014; Andreon & Weaver, 2015). The chapter in Hilbe, 2013 by Loredo discusses hierarchical models, but refers to them as "multilevel" models (Loredo, 2013). Seasoned veterans may find Gelman et al., 2013 to be a thorough reference.

[1] This review focuses on Bayesian inference applied to audio-band gravitational waves from compact binary coalescence, the only source of gravitational waves yet detected. We note in passing that Bayesian inference has been applied to study gravitational waves from rotating neutron stars (Umstätter et al., 2004; Dupuis & Woan, 2005; Abbott, 2017d), bursting sources (Cornish & Littenberg, 2015; Logue et al., 2012; Powell et al., 2016), and stochastic backgrounds (Mandic et al., 2012; Callister et al., 2017; Abbott, 2018a). Bayesian inference methods have also been developed for space-based observatories observing at millihertz frequencies (Babak et al., 2008; Babak et al., 2010) and for pulsar timing arrays operating at nanohertz frequencies (Lentati et al., 2014; Vigeland & Vallisneri, 2014).

2.2 Fundamentals: likelihoods, priors, and posteriors

A primary aim of modern Bayesian inference is to construct a posterior distribution

$$p(\theta|d). \tag{2.1}$$

Here, θ is the set of model parameters and d is the data associated with a measurement[2]. For illustrative purposes, let us say that θ are the 15 parameters describing a binary black hole coalescence and d is the strain data from a network of gravitational-wave detectors. The posterior distribution $p(\theta|d)$ is the probability density function for the continuous variable θ given the data d. The probability that the true value of θ is between $(\theta', \theta' + d\theta)$ is given by $p(\theta'|d)d\theta'$. It is normalized so that

$$\int d\theta \, p(\theta|d) = 1 \tag{2.2}$$

The posterior distribution is what we use to construct credible intervals that tell us, for example, the component masses of a binary black hole event like GW150914. For details about the construction of credible intervals, see Section 2.6.

According to Bayes theorem, the posterior distribution is given by

$$p(\theta|d) = \frac{\mathcal{L}(d|\theta)\, \pi(\theta)}{\mathcal{Z}}. \tag{2.3}$$

Here, $\mathcal{L}(d|\theta)$ is the likelihood function of the data given the parameters θ, $\pi(\theta)$ is the prior distribution for θ, and $\mathcal{Z}$ is a normalization factor[3,4] called the "evidence"

$$\mathcal{Z} \equiv \int d\theta \mathcal{L}(d|\theta)\, \pi(\theta). \tag{2.4}$$

The likelihood function is something that we choose. It is a description of the measurement. By writing down a likelihood, we implicitly introduce a noise model. For gravitational-wave astronomy, we typically assume a

[2] By referring to "model parameters," we are implicitly acknowledging that we begin with some model. Some authors make this explicit by writing the posterior as $p(\theta|d, M)$ where M is the model. (Other authors sometimes use I to denote the model.) We find this notation clunky and unnecessary since it goes without saying that one must always assume *some* model. If/when we consider two *distinct* models, we add an additional variable to denote the model.

[3] In this document we use different symbols for different distributions: p for posteriors, $\mathcal{L}$ for likelihoods, and π for priors. We advocate this notation since it highlights what is what and makes formulas easy to read. However, it is by no means standard, and some authors will use p for any and all probability distributions.

[4] For now, we treat the evidence as "just" a normalization factor, though, below we see that it plays an important role in model selection, and that it can be understood as a marginalized likelihood.

Gaussian-noise likelihood function that looks something like this

$$\mathcal{L}(d|\theta) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2}\frac{|d-\mu(\theta)|^2}{\sigma^2}\right). \tag{2.5}$$

Here, $\mu(\theta)$ is a template for the gravitational strain waveform given θ and σ is the detector noise. Note that π with no parentheses and no subscript is the mathematical constant, not a prior distribution. There is no square root in the normalisation factor because d is (typically) complex, which means that we are working with a two-dimensional Gaussian—the Whittle likelihood (Whittle, 1951); see also Cornish & Romano, 2013. This likelihood function reflects our assumption that the noise in gravitational-wave detectors is Gaussian[5]. Note that the likelihood function is not normalized with respect to θ and so[6]

$$\int d\theta\, \mathcal{L}(d|\theta) \neq 1. \tag{2.6}$$

For a more detailed discussion of the Gaussian noise likelihood in the context of gravitational-wave astronomy, see Section 2.7.

Like the likelihood function, the prior is something we get to choose. The prior incorporates our belief about θ before we carry out a measurement. In some cases, there is an obvious choice of prior. For example, if we are considering the sky location of a binary black hole merger, it is reasonable to choose an isotropic prior that weights each patch of sky as equally probable. In other situations, the choice of prior is not obvious. For example, before the first detection of gravitational waves, what would have been a suitable choice for the prior on the primary[7] black hole mass $\pi(m_1)$? When we are ignorant about θ, we often express our ignorance by choosing a distribution that is either uniform or log-uniform[8].

While θ may consist of a large number of parameters, we usually want to look at just one or two at a time. For example, the posterior distribution for a

[5] The Gaussian noise assumption is a good starting point for describing the strain noise in gravitational-wave detectors. The combined effect of many random noise processes tends to produce nearly Gaussian strain noise. Of course, the noise description can be generalized to include non-Gaussian glitches, drift over time, and instrumental lines all of which can be described by noise parameters; see, e.g., Littenberg & Cornish, 2015; Röver, Meyer & Christensen, 2011.

[6] Given that the likelihood is not normalized with respect to θ, one might ask in what way it *is* normalized. The answer is that the likelihood is normalized with respect to the *data* d. Before we collect any data, the likelihood describes the chance of getting data d. It is a probability density function with units of inverse data. The integral over all possible d is unity. Once we obtain actual data, d is, of course, fixed.

[7] The "primary" black hole is the heavier of two black holes in a binary, which is contrasted with the lighter "secondary" black hole.

[8] A log uniform distribution is used when we do not know the order of magnitude of some quantity, for example, the energy density of primordial gravitational waves.

binary black hole merger is a fifteen-dimensional[9] function that includes information about black hole masses, sky location, spins, etc. What if we want to look at the posterior distribution for just the primary mass? To answer this question we *marginalize* (integrate) over the parameters that we are not interested in (called "nuisance parameters") so as to obtain a marginalized posterior

$$p(\theta_i|d) = \int \left(\prod_{k\neq i} d\theta_k\right) p(\theta|d) \tag{2.7}$$
$$= \frac{\mathcal{L}(d|\theta_i)\,\pi(\theta_i)}{\mathcal{Z}} \tag{2.8}$$

The quantity $\mathcal{L}(d|\theta_i)$ is called the "marginalized likelihood." It can be expressed like so:

$$\mathcal{L}(d|\theta_i) = \int \left(\prod_{k\neq i} d\theta_k\right) \pi(\theta_k)\,\mathcal{L}(d|\theta) \tag{2.9}$$

When we marginalize over one variable θ_a in order to obtain a posterior on θ_b, we are calculating our best guess for θ_b given uncertainty in θ_a. Speaking somewhat colloquially, if θ_a and θ_b are covariant, then marginalizing over θ_a "injects" uncertainty into the posterior for θ_b. When this happens, the marginalized posterior $p(\theta_b|d)$ is broader than the *conditional posterior* $p(\theta_b|d,\theta_a)$. The conditional posterior $p(\theta_b|d,\theta_a)$ represents a slice through the $p(\theta_b|d)$ posterior at a fixed value of θ_a.

This is nicely illustrated with an example. There is a well-known covariance between the luminosity distance of a merging compact binary from Earth D_L and the inclination angle θ_{JN}. For the binary neutron star coalescence GW170817, we are able to constrain the inclination angle much better when we use the known distance and sky location of the host galaxy compared to the constraint obtained using the gravitational-wave measurement alone[10]. Results from Abbott, 2019c are shown in Fig. 2.1.

2.3 Models, evidence and odds

In Eq. (2.4), reproduced here, we defined the Bayesian evidence:

$$\mathcal{Z} \equiv \int d\theta \mathcal{L}(d|\theta)\,\pi(\theta).$$

[9]There are eight "intrinsic" parameters, which are fundamental properties of the binary: primary mass m_1, secondary mass m_2, primary dimensionless spin vector $\vec{s}_1$, and secondary dimensionless spin vector $\vec{s}_2$. The other seven parameters are "extrinsic," relating to how we view the binary. The extrinsic parameters are: inclination angle ι, polarization angle ψ, phase at coalescence ϕ_c, right ascension RA, declination DEC, luminosity distance D_L, and time of coalescence t.

[10]The viewing angle = $\Theta = \min(\theta_{JN}, 180^\circ - \theta_{JN}$ is constrained to be $< 28^\circ$ with the electromagnetic counterpart, and $< 55^\circ$ without it (Abbott, 2017i)

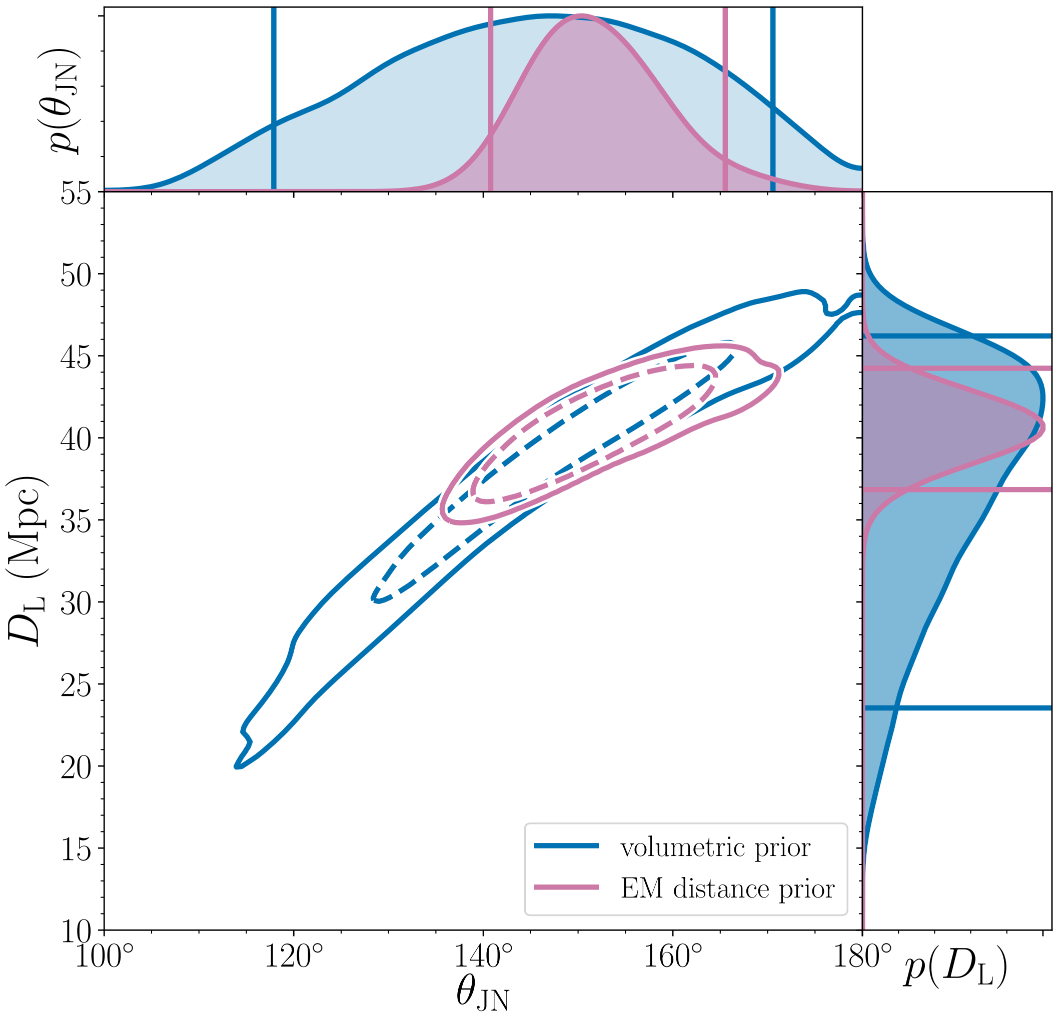

FIGURE 2.1: The joint posterior for luminosity distance and inclination angle for GW170817 from Abbott, 2019c. The blue contours show the credible region obtained using gravitational-wave data alone. The purple contours show the smaller credible region obtained by employing a relatively narrow prior on distance obtained with electromagnetic measurements. Publicly available posterior samples for this plot are available here: Abbott, 2018e.

In practical terms, the evidence is a single number. It usually does not mean anything by itself, but becomes useful when we compare one evidence with another evidence. Formally, the evidence is a likelihood function. Specifically, it is the completely marginalized likelihood function. It is therefore sometimes denoted $\mathcal{L}(d)$ with no θ dependence. However, we prefer to use $\mathcal{Z}$ to denote the fully marginalized likelihood function.

Above, we described how the evidence serves as a normalization constant for the posterior $p(\theta|d)$. However, the evidence is also used to do model selection. Model selection answers the question: which model is statistically preferred by the data and by how much? There are different ways to think about models. Let us return to the case of binary black holes. We may compare a "signal model" in which we suppose that there is a binary black hole signal present in the data with a prior $\pi(\theta)$ to the "noise model," in which we suppose that there is no binary black hole signal present. While the signal model is described by the fifteen binary parameters θ, the noise model is described by no parameters. Thus, we can define a signal evidence $\mathcal{Z}_S$ and a noise evidence $\mathcal{Z}_N$

$$\mathcal{Z}_S \equiv \int d\theta \mathcal{L}(d|\theta)\, \pi(\theta) \tag{2.10}$$

$$\mathcal{Z}_N \equiv \mathcal{L}(d|0), \tag{2.11}$$

where

$$\mathcal{L}(d|0) \equiv \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2}\frac{|d|^2}{\sigma^2}\right). \tag{2.12}$$

The noise evidence $\mathcal{Z}_N$ is sometimes referred to as the "null likelihood."

The ratio of the evidence for two different models is called the Bayes factor. In this example, the signal/noise Bayes factor is

$$\mathrm{BF}_N^S \equiv \frac{\mathcal{Z}_S}{\mathcal{Z}_N}. \tag{2.13}$$

It is often convenient to work with the log of the Bayes factor[11]

$$\log \mathrm{BF}_N^S \equiv \log(\mathcal{Z}_S) - \log(\mathcal{Z}_N). \tag{2.14}$$

When the absolute value of $\log \mathrm{BF}$ is large, we say that one model is preferred over the other. The sign of $\log \mathrm{BF}$ tells us which model is preferred. A threshold of $|\log \mathrm{BF}| = 8$ is often used as the level of "strong evidence" in favor of one hypothesis over another (Jeffreys, 1961).

The signal/noise Bayes factor is just one example of a Bayes factor comparing two models. We can calculate a Bayes factor comparing identical models but with different priors. For example, we can calculate the evidence for a binary black hole with a uniform prior on dimensionless spin and compare that to the evidence obtained using a zero-spin prior. The Bayes factor

[11] A typical log evidence might be -5000, which evaluates to zero when exponentiated on a computer . Functions such as `logsumexp` can be useful for combining evidences.

comparing these models would tell us if the data prefer spin.

$$\mathcal{Z}_{\text{spin}} = \int d\theta \mathcal{L}(d|\theta)\, \pi(\theta) \tag{2.15}$$

$$\mathcal{Z}_{\text{no spin}} = \int d\theta \mathcal{L}(d|\theta)\, \pi_{\text{no spin}}(\theta). \tag{2.16}$$

Where $\pi_{\text{no spin}}(\theta)$ is a prior with zero spins. The spin/no spin Bayes factor is

$$\text{BF}^{\text{spin}}_{\text{no spin}} = \frac{\mathcal{Z}_{\text{spin}}}{\mathcal{Z}_{\text{no spin}}}. \tag{2.17}$$

We may also compare two disparate signal models. For example, we can compare the evidence for a binary black hole waveform predicted by general relativity (model M_A with parameters θ) with a binary black hole waveform predicted by some other theory (model M_B with parameters ν):

$$\mathcal{Z}_A = \int d\theta \mathcal{L}(d|\theta, M_A)\, \pi(\theta) \tag{2.18}$$

$$\mathcal{Z}_B = \int d\nu \mathcal{L}(d|\nu, M_B)\, \pi(\nu). \tag{2.19}$$

The A/B Bayes factor is

$$\text{BF}^A_B = \frac{\mathcal{Z}_A}{\mathcal{Z}_B}. \tag{2.20}$$

Note that the number of parameters in ν can be different from the number of parameters in θ.

Our presentation of model selection so far has been a bit fast and loose. Formally, the correct metric to compare two models is not the Bayes factor, but rather the odds ratio

$$\mathcal{O}^A_B \equiv \frac{\mathcal{Z}_A}{\mathcal{Z}_B} \frac{\pi_A}{\pi_B}. \tag{2.21}$$

The odds ratio is the product of the Bayes factor with the prior odds π_A/π_B, which describes our prior belief about the relative likelihood of hypotheses A and B. In many practical applications, we set the prior odds ratio to unity, and so the odds ratio *is* the Bayes factor. This practice is sensible in many applications where our intuition tells us: until we do this measurement both hypotheses are equally likely[12].

[12]There are some (fairly uncommon) examples where we might choose a different prior odds ratio. For example, we may construct a model in which general relativity (GR) is wrong. We may further suppose that there are multiple different ways in which it could be wrong, each corresponding to a different GR-is-wrong sub-hypothesis. If we calculated the odds ratio comparing one of these GR-is-wrong sub-hypotheses to the GR-is-right hypothesis, we would not assign equal prior odds to both hypotheses. Rather, we would assign at most 50% probability to the entire GR-is-wrong hypothesis, which would then have to be split among the various sub-hypotheses (Callister et al., 2017).

Bayesian evidence encodes two pieces of information. First, the likelihood tells us how well our model fits the data. Second, the act of marginalization tell us about the size of the volume of parameter space we used to carry out a fit. This creates a sort of tension. We want to get the best fit possible (high likelihood) but with a minimum prior volume. A model with a decent fit and a small prior volume often yields a greater evidence than a model with an excellent fit and a huge prior volume. In these cases, the Bayes factor penalizes the more complicated model for being too complicated.

This penalty is called an Occam factor. It is a mathematical formulation of the statement that all else equal, a simple explanation is more likely than a complicated one. If we compare two models where one model is a superset of the other—for example, we might compare general relativity and general relativity with non-tensor modes—and if the data are better explained by the simpler model, the log Bayes factor is typically modest, $\log \mathrm{BF} \approx (-2, -1)$. Thus, it is difficult to completely rule out extensions to existing theories. We just obtain ever tighter constraints on the extended parameter space.

2.4 Samplers

Thanks to the creation of phenomenological gravitational waveforms (called "approximants"), it is now computationally straightforward to make a prediction about what the data d should look like given some parameters θ. That is a forward problem. Calculating the posterior, the probability of parameters θ given the data as in Eq. 2.3, reproduced here, is a classic inverse problem[13]

$$p(\theta|d) = \frac{\mathcal{L}(d|\theta)\, \pi(\theta)}{\mathcal{Z}}.$$

In general, inverse problems are computationally challenging compared to forward problems. To illustrate why let us imagine that we wish to calculate the posterior probability for the fifteen parameters describing a binary black hole merger. If we do this naively, we might create a grid with ten bins in every dimension and evaluate the likelihood at each grid point. Even with this coarse resolution, our calculation suffers from "the curse of dimensionality." It is computationally prohibitive to carry out 10^{15} likelihood evaluations. The problem becomes worse as we add dimensions. As a rule of thumb, brute-force bin approaches become painful once one exceeds three dimensions.

The solution is to use a stochastic sampler, (although recent work has shown progress carrying out these calculations using the alternative technique of iterative fitting, Pankow et al., 2015; Lange, O'Shaughnessy & Rizzo, 2018). Commonly used sampling algorithms can be split into two broad categories of method: Markov-chain Monte Carlo (MCMC) (Metropolis et al.,

[13] We note here a few early papers important in the development of Bayesian inference tools for gravitational-wave astronomy. Initial implementation of MCMC methods for spinning binaries was carried out in van der Sluys et al., 2008a. The first demonstration of Bayesian parameter estimation for spinning binaries was performed in van der Sluys et al., 2008b. Veitch & Vecchio, 2008, demonstrated Bayesian model selection for compact binaries.

1953; Hastings, 1970) and nested sampling (Skilling, 2004). These algorithms generate a list of posterior samples $\{\theta\}$ drawn from the posterior distribution such that the number of samples on the interval $(\theta, \theta + \Delta\theta) \propto p(\theta)$ (Veitch et al., 2015). Some samplers also produce an estimate of the evidence. We can visualize the posterior samples as a spreadsheet. Each column is a different parameter, for example, primary black hole mass, secondary black hole mass, etc. For binary black hole mergers, there are typically fifteen columns. Each row represents a different posterior sample.

Posterior samples have two useful properties. First, they can be used to compute expectation values of quantities of interest since (Hogg & Foreman-Mackey, 2018)

$$\langle f(x) \rangle_{p(x)} = \int dx\, p(x)\, f(x) \approx \frac{1}{n_s} \sum_{k}^{n_s} f(x_k). \tag{2.22}$$

Here $p(x)$ is the posterior distribution that we are sampling, $f(x)$ is some function we want to find the expectation value of, and the sum over k runs over n_s posterior samples. Below, Eq. 2.22 will prove useful simplifying our calculation of the likelihood of data given hyper-parameters.

The second useful property of posterior samples is that, once we have samples from an N-dimensional space, we can generate the marginalized probability for any subset of the parameters by simply selecting the corresponding columns in our spreadsheet. This property is used to help visualize the output of these samplers by constructing "corner plots," which show the marginalized one- and two-dimensional posterior probability distributions for each of the parameters. For an example of a corner plot, see Fig. 2.1. A handy python package exists for making corner plots (Foreman-Mackey, 2016).

2.4.1 MCMC

Markov chain Monte Carlo sampling was first introduced in Metropolis et al., 1953 and extended in Hastings, 1970. For a recent overview of MCMC methods in astronomy, see Sharma, 2017. In MCMC methods, particles undergo a random walk through the posterior distribution where the probability of moving to any given point is determined by the transition probability of the Markov chain. By noting the position of the particles—or "walkers" as they are sometimes called—at each iteration, we generate draws from the posterior probability distribution.

There are some subtleties that must be considered when using MCMC samplers. First, the early-time behavior of MCMC walkers is strongly dependent on the initial conditions. It is therefore necessary to include a "burn-in" phase to ensure that the walker has settled into a steady state before beginning to accumulate samples from the posterior distribution. Once the walker has reached a steady state, the algorithm can continue indefinitely and so it is necessary for the user to define a termination condition. This is typically chosen to be when enough samples have been acquired for the user to believe an accurate representation of the posterior has been obtained. Thus, MCMC requires a degree of artistry, developed from experience.

Additionally, the positions of a walker in a chain are often autocorrelated. Because of this correlation, the positions of the walkers do not represent a faithful sampling from the posterior distribution. If no remedy is applied, the width of the posterior distribution is underestimated. It is thus necessary to "thin" the chain by selecting samples separated by the autocorrelation length of the chain.

Markov chain Monte Carlo walkers can also fail to find multiple modes of a posterior distribution if there are regions of low posterior probability between the modes. However, this can be mitigated by running many walkers which begin exploring the space at different points. This also demonstrates a simple way to parallelize MCMC computations to quickly generate many samples. Many variants of MCMC sampling have been proposed in order to improve the performance of MCMC algorithms with respect to these and other issues. For a more in-depth discussion of MCMC methods see, e.g., chapter 11 of Gelman et al., 2013, or Hogg & Foreman-Mackey, 2018. The most widely used MCMC code in astronomy is EMCEE (Foreman-Mackey et al., 2013)[14].

2.4.2 Nested sampling

The first widely used alternative to MCMC, was introduced by Skilling in 2004. While MCMC methods are designed to draw samples from the posterior distribution, nested sampling is designed to calculate the evidence. Generating samples from the posterior distribution is a by-product of the nested sampling evidence calculation algorithm. By weighting each of the samples used to calculate the evidence by the posterior probability of the sample, nested samples are converted into posterior samples.

Nested sampling works by populating the parameter space with a set of "live points" drawn from the prior distribution. At each iteration, the lowest likelihood point is removed from the set of live points and new samples are drawn from the prior distribution until a point with higher likelihood than the removed point is found. The evidence is evaluated by assigning each removed point a prior volume and then computing the sum of the likelihood multiplied by the prior volume for each sample.

Since the nested sampling algorithm continually moves to higher likelihood regions, it is possible to estimate an upper limit on the evidence at each iteration. This is done by imagining that the entire remaining prior volume has a likelihood equal to that of the highest likelihood live point. This is used to inform the termination condition for the nested sampling algorithm. The algorithm stops when the current estimate of the evidence is above a certain fraction of the estimated upper limit[15]. Unlike MCMC algorithms nested

[14] http://dfm.io/emcee/

[15] In practice this is expressed as the difference between the calculated log evidence and the upper limit of the log evidence.

sampling is not straightforwardly parallelizable, and posterior samples do not accumulate linearly with run time.

2.5 Hyper-parameters and hierarchical models

As more and more gravitational-wave events are detected, it is increasingly interesting to study the *population properties* of binary black holes and binary neutron stars. These are the properties common to all of the events in some set. Examples include the neutron star equation of state and the distribution of black hole masses. Hierarchical Bayesian inference is a formalism, which allows us to go beyond individual events in order to study population properties[16].

The population properties of some set of events is described by the shape of the prior. For example, two population synthesis models might yield two different predictions for the prior distribution of the primary black hole mass $\pi(m_1)$. In order to probe the population properties of an ensemble of events, we make the prior for θ conditional on a set of "hyper-parameters" Λ

$$\pi(\theta|\Lambda). \tag{2.23}$$

The hyper-parameters parameterize the shape of the prior distribution for the parameters θ. An example of a (parameter, hyper-parameter) relationship is (θ = primary black hole mass m_1, Λ = the spectral index of the primary mass spectrum α). In this example

$$\pi(m_1|\alpha) \propto m_1^\alpha. \tag{2.24}$$

A key goal of population inference is to estimate the posterior distribution for the hyper-parameters Λ. In order to do this, we marginalize over the entire parameter space θ in order to obtain a marginalized likelihood.

$$\mathcal{L}(d|\Lambda) = \int d\theta\, \mathcal{L}(d|\theta)\, \pi(\theta|\Lambda). \tag{2.25}$$

Normally, we would call this completely marginalized likelihood an evidence, but because it still depends on Λ, we call it the likelihood for the data d given the hyper-parameters Λ. The hyper-posterior is given simply by

$$p(\Lambda|d) = \frac{\mathcal{L}(d|\Lambda)\, \pi(\Lambda)}{\int d\Lambda\, \mathcal{L}(d|\Lambda)\, \pi(\Lambda)}. \tag{2.26}$$

[16] Possibly the earliest papers proposing to measure *distributions* of gravitational-wave parameters are Mandel & O'Shaughnessy, 2010; Mandel, 2010, while hierarchical Bayesian inference was introduced to study the population properties of sources of gravitational waves in Adams, Cornish & Littenberg, 2012.

Note that we have introduced a hyper-prior $\pi(\Lambda)$, which describes our prior belief about the hyper-parameters Λ. The term in the denominator

$$\mathcal{Z}_\Lambda \equiv \int d\Lambda \, \mathcal{L}(d|\Lambda) \, \pi(\Lambda) \tag{2.27}$$

is the "hyper-evidence," which we denote $\mathcal{Z}_\Lambda$ in order to distinguish it from the regular evidence $\mathcal{Z}_\theta$. In Section 2.9 we discuss posterior predictive distributions (PPD), which represent the updated prior on θ in light of the data d and given some hyper-parameterization.

We now generalize the discussion of hyper-parameters in order to handle the case of N independent events. In this case, the total likelihood for all N events $\mathcal{L}_{\text{tot}}$ is simply the product of each individual likelihood

$$\mathcal{L}_{\text{tot}}(\vec{d}|\vec{\theta}) = \prod_i^N \mathcal{L}(d_i|\theta_i). \tag{2.28}$$

Here, we use vector notation so that $\vec{d}$ is the set of measurements of N events, each of which has its own parameters, which make up the vector $\vec{\theta}$. Since we suppose that every event is drawn from the same population prior distribution—hyper-parameterized by Λ—the total marginalized likelihood is

$$\mathcal{L}_{\text{tot}}(\vec{d}|\Lambda) = \prod_i^N \int d\theta_i \, \mathcal{L}(d_i|\theta_i) \, \pi(\theta_i|\Lambda). \tag{2.29}$$

The associated (hyper-) posterior is

$$p_{\text{tot}}(\Lambda|\vec{d}) = \frac{\mathcal{L}_{\text{tot}}(\vec{d}|\Lambda) \, \pi(\Lambda)}{\int d\Lambda \, \mathcal{L}_{\text{tot}}(\vec{d}|\Lambda) \, \pi(\Lambda)}. \tag{2.30}$$

The denominator, of course, is the total hyper-evidence.

$$\mathcal{Z}_\Lambda^{\text{tot}} = \int d\Lambda \, \mathcal{L}_{\text{tot}}(\vec{d}|\Lambda) \, \pi(\Lambda) \tag{2.31}$$

We may calculate the Bayes factor comparing different hyper-models in the same way that we calculate the Bayes factor for different models.

Examining Eq. 2.31, we see that the total hyper-evidence involves a large number of integrals. For the case of binary black hole mergers, every event has 15 parameters, and so the dimension of the integral is $15N + M$ taking where M is the number of hyper-parameters in Λ. As N gets large, it becomes difficult to sample such a large prior volume all at once. Fortunately, it is possible to break the integral into individual integrals for each event, which are then combined through a process sometimes referred to as "recycling."

It turns out that the total marginalized likelihood in Eq. 2.29 can be written like so

$$\boxed{\mathcal{L}_{\text{tot}}(\vec{d}|\Lambda) = \prod_i^N \frac{\mathcal{Z}_\varnothing(d_i)}{n_i} \sum_k^{n_i} \frac{\pi(\theta_i^k|\Lambda)}{\pi(\theta_i^k|\varnothing)}}. \tag{2.32}$$

Here, the sum over k is a sum over the n_i posterior samples associated with event i. The posterior samples for each event are generated with some default prior $\pi(\theta_k|\varnothing)$. The default prior is ultimately canceled from the final answer, so it not so important what we choose for the default prior so long as it is sufficiently uninformative. Using the $\varnothing$ prior, we obtain an evidence $\mathcal{Z}_\varnothing$. In this way, we are able to analyze each event individually before recycling the posterior samples to obtain a likelihood of the data given Λ.

To see where this formula comes from, we note that

$$p(\theta_i|d_i,\varnothing) = \frac{\mathcal{L}(d_i|\theta_i)\,\pi(\theta_i|\varnothing)}{\mathcal{Z}_\varnothing(d_i)} \tag{2.33}$$

Rearranging terms,

$$\mathcal{L}(d_i|\theta_i) = \mathcal{Z}_\varnothing(d_i)\,\frac{p(\theta_i|d_i,\varnothing)}{\pi(\theta_i|\varnothing)}. \tag{2.34}$$

Plugging this into Eq. 2.29, we obtain[17]

$$\mathcal{L}_{\text{tot}}(\vec{d}|\Lambda) = \prod_i^N \int d\theta_i\, p(\theta_i|d_i,\varnothing)\, \mathcal{Z}_\varnothing(d_i)\, \frac{\pi(\theta_i|\Lambda)}{\pi(\theta_i|\varnothing)}. \tag{2.35}$$

Finally, we use Eq. 2.22 to convert the integral over θ_i to a sum over posterior samples, thereby arriving at Eq. 2.32.

All of the results derived up until this point ignore selection effects where an event with parameters θ_1 is easier to detect than an event with parameters θ_2. There are cases where selection effects are important. For example, the visible volume for binary black hole mergers scales as approximately $V \propto M^{2.1}$, which means that higher mass mergers are relatively easier to detect than lower mass mergers (Fishbach & Holz, 2017). In Section 2.10, we show how this method is extended to accommodate selection effects.

2.6 Credible intervals

It is often convenient to use the posterior to construct "credible intervals," regions of parameter space containing some fraction of posterior probability. (Note that Bayesian inference yields credible intervals while frequentist inference yields *confidence intervals*.) For example, one can plot one-, two-, and three-sigma contours. By definition, a two-sigma credible region

[17] One "recycles" the posterior samples generated using the the $\pi(\theta_i|\varnothing)$ prior in order to do something new with the hyper-parameterized prior $\pi(\theta_i|\Lambda)$.

includes 95% of the posterior probability, but this requirement does not uniquely determine a single credible region. One well-motivated method for constructing confidence intervals is the highest posterior density interval (HPDI) method.

We can visualize the HPDI method as follows. We draw a horizontal line through a posterior distribution and calculate the area of above the line. If we move the line down, the area goes up. If we place the line such that the area is 95%, then the posterior above the line is the HPDI two-sigma credible interval. In general, the HPDI is neither symmetric nor unimodal. The advantage of HPDI over other methods is that it yields the minimum width credible interval. This method is sometimes referred to as "draining the bathtub."

Another commonly used method for calculating credible intervals is to construct symmetric intervals. Symmetric credible intervals are constructed using the cumulative distribution function,

$$P(x) = \int_{-\infty}^{x} dx' \, p(x'). \tag{2.36}$$

The X% credible region is the region

$$\frac{1}{2}\left(1 - \frac{X}{100}\right) < P(x) < \frac{1}{2}\left(1 + \frac{X}{100}\right). \tag{2.37}$$

While symmetric credible intervals are simpler to construct than HPDI, particularly from samples drawn from a distribution, they can be misleading for multi-modal distributions and for distributions which peak near prior boundaries.

Credible intervals are useful for testing and debugging inference projects. Before applying an inference calculation to real data, it is useful to test it on simulated data. The standard test, see, e.g., Sidery et al., 2014, is to simulate data d according to parameters θ_{true} drawn at random from the prior distribution $\pi(\theta)$. Then, we analyze this data in order to obtain a posterior $p(\theta|d)$. The true value should fall inside the 90% credible interval 90% of the time. Testing that this is true provides a powerful validation of the inference algorithm. Note that we do not expect the posterior to peak precisely at θ_{true}, just within the one-sigma region.

2.7 Gaussian noise likelihood

In this section, we introduce additional notation that is helpful for talking about the Gaussian noise likelihood frequently used in gravitational-wave astronomy. In the main body of the manuscript, d has been taken to represent data. Now, we take d to represent the Fourier transform of the strain time series $d(t)$ measured by a gravitational-wave detector. In the language of computer programming,

$$d = \texttt{fft}\left(d(t)\right) / f_s, \tag{2.38}$$

where f_s is the sampling frequency and `fft` is a Fast Fourier transform. The noise in each frequency bin is characterized by the single-sided noise power spectral density $P(f)$, which is proportional to strain squared and which has units of Hz^{-1}.

The likelihood for the data in a single frequency bin j given θ is

$$\mathcal{L}(d_j|\theta) = \frac{1}{2\pi P_j} \exp\left(-2\Delta f \frac{|d_j - \mu_j(\theta)|^2}{P_j}\right). \tag{2.39}$$

Here Δf is the frequency resolution. The factor of $2\Delta f$ comes about from a factor of $1/2$ in the normal distribution and a factor of $4\Delta f$ needed to convert the square of the Fourier transforms into units of one-sided power spectral density. Note that the normalisation factor does not contain a square root because the data are complex, and so the Gaussian is a two-dimensional Whittle likelihood (Whittle, 1951); see also Cornish & Romano, 2013. The template $\mu(\theta)$ is related to the metric perturbation $h_{+,\times}(\theta)$ via antenna response factors $F_{+,\times}$ (Anderson et al., 2001)

$$\mu(\theta) = F_{+}(\mathrm{RA}, \mathrm{DEC}, \psi) h_{+}(\theta) + F_{\times}(\mathrm{RA}, \mathrm{DEC}, \psi) h_{\times}(\theta) \tag{2.40}$$

Gravitational-wave signals are typically spread over many (M) frequency bins. Assuming the noise in each bin is independent, the combined likelihood is a product of the likelihoods for each bin

$$\mathcal{L}(\mathbf{d}|\theta) = \prod_j^M \mathcal{L}(d_j|\theta) \tag{2.41}$$

Here $\mathbf{d}$ is the set of data including all frequency bins and d_j represents the data associated with frequency bin j. If we consider a measurement with multiple detectors, the product over j frequency bins gains an additional index l for each detector. Combining data from different detectors is like combining data from different frequency bins.

It is frequently useful to work with the log likelihood, which allows us to replace products with sums of logs. The log also helps dealing with small numbers. The log likelihood is

$$\begin{aligned}
\log \mathcal{L}(\mathbf{d}|\theta) &= \sum_j^M \log \mathcal{L}(d_j|\theta) \\
&= -\frac{1}{2}\sum_j \log\left(2\pi P_j\right) - 2\Delta f \sum_j \frac{|d - \mu(\theta)|^2}{P_j} \\
&= \Psi - \frac{1}{2}\langle d - \mu(\theta), d - \mu(\theta)\rangle.
\end{aligned}$$

In the last line, we define the noise-weighted inner product[18] (Cutler & Flanagan, 1994)

$$\langle a,b\rangle \equiv 4\Delta f \sum_j \Re\left(\frac{a_j^* b_j}{P_j}\right), \tag{2.42}$$

and the constant

$$\Psi \equiv -\sum_j \log\left(2\pi P_j\right). \tag{2.43}$$

Since constants do not change the shape of the log likelihood we often "leave off" this normalizing term and work with log likelihood minus Ψ. This is permissible as long as we do so consistently because when we take the ratio of two evidences—or equivalently, the difference of two log evidences—the Ψ factor cancels anyway. For the remainder of this section, we set $\Psi = 0$. Now that we have introduced the inner product notation, we are going to stop bold-facing the data d as it is implied that we are dealing with many frequency bins.

Using the inner product notation, we may expand out the log likelihood

$$\begin{aligned}
\log \mathcal{L}(d|\theta) &= -\frac{1}{2}\left[\langle d,d\rangle - 2\langle d,\mu(\theta)\rangle + \langle \mu(\theta),\mu(\theta)\rangle\right] \\
&= -\frac{1}{2}\left[-2\log \mathcal{Z}_N - 2\kappa^2(\theta) + \rho^2_{\text{opt}}(\theta)\right] \\
&= \log \mathcal{Z}_N + \kappa^2(\theta) - \frac{1}{2}\rho^2_{\text{opt}}(\theta).
\end{aligned} \tag{2.44}$$

We see that the log likelihood can be expressed with three terms. The first is proportional to the log noise evidence

$$-2\log \mathcal{Z}_N \equiv \langle d,d\rangle. \tag{2.45}$$

For debugging purposes, it is useful to keep in mind that if we calculate $-\log \mathcal{Z}_N$ on actual Gaussian noise (with $\Psi = 0$), we expect a typical value nearly equal to the number of frequency bins M (multiplied by the number of detectors) since each term in the inner product contributes a value close to unity[19]. We skip over the second term κ^2 for a moment. The third term is the optimal matched filter signal-to-noise ratio squared

$$\rho^2_{\text{opt}} \equiv \langle \mu,\mu\rangle. \tag{2.46}$$

[18] Following the convention of gravitational-wave astronomy, our inner product is real by construction. However, below it will be useful to define a complex-valued inner product; see Eq. 2.58.

[19] Specifically, the distribution of an ensemble of independent $-\ln \mathcal{Z}_N$ is a normal distribution with mean M and width $M^{1/2}$ where M is the number of frequency bins (multiplied by the number of detectors). This follows from the central limit theorem.

Returning now to the second term, we express κ^2 as the product of the matched filter signal-to-noise ratio and the optimal signal-to-noise ratio

$$\begin{aligned}\kappa^2 \equiv& \langle d, \mu \rangle \\ =& \rho_{\rm mf}\, \rho_{\rm opt},\end{aligned} \tag{2.47}$$

where

$$\rho_{\rm mf} \equiv \frac{\langle d, \mu \rangle}{\langle \mu, \mu \rangle^{1/2}}. \tag{2.48}$$

Readers familiar with gravitational-wave astronomy are likely acquainted with the concept of matched filtering, which is the maximum likelihood technique for gravitational-wave detection. By writing the likelihood in this way, we highlight how parameter estimation is related to matched filtering. Rapid evaluation of the likelihood function in Eq. 2.44 has been made possible through reduced order methods (Smith et al., 2016; Pürrer, 2014; Canizares et al., 2013).

2.8 Explicitly Marginalized Likelihoods

The most computationally expensive step in computing the likelihood for compact binary coalescences is creating the waveform template (μ in Eq. 2.5). This is done in two steps. The first step is to use the *intrinsic parameters* to calculate the metric perturbation. The second (much faster) step is to use the *extrinsic parameters* to project the metric perturbation onto the detector response tensor. In some cases, it is possible to reduce the dimensionality of the inverse problem—thereby speeding up calculations and improving convergence—by using a likelihood, which explicitly marginalizes over extrinsic parameters. The improvement is especially marked for comparatively weak signals, which can be important for population studies; see, e.g., Smith & Thrane, 2018. In this section, we show how to calculate $\mathcal{L}_{\rm marge}$—a likelihood, which explicitly marginalize over coalescence time, phase at coalescence, and/or luminosity distance. We continue with notation introduced in Section 2.7.

2.8.1 Time marginalization

In this subsection, we follow Farr, 2014 to derive a likelihood, which explicitly marginalizes over time of coalescence t. Given a waveform with a reference coalescence time of t_0, we can calculate the waveform at some new coalescence time t by multiplying by the appropriate phasor:

$$\mu_j(t) = \mu_j(t_0) \exp\left(-2\pi i j \frac{(t-t_0)}{T}\right). \tag{2.49}$$

Here $T = 1/\Delta f$ is the duration of data segment and j is the index of the frequency bin as in Section 2.7. It is understood that μ is a function of whatever parameters we are not explicitly marginalizing over. We can therefore write κ^2 (see Eq: 2.47) as

$$\begin{aligned}\kappa^2(t) \equiv& \langle d, \mu(t)\rangle \\ =& 4\Delta f \Re \sum_j^M \frac{d_j^* \mu_j(t_0)}{P_j} \exp\left(-2\pi i j \frac{(t-t_0)}{T}\right).\end{aligned} \tag{2.50}$$

However this sum is the discrete Fourier transform. By recasting this equation in terms of the fast Fourier transform `fft`, it is possible to take advantage of a highly optimized tool.

We discretize $t - t_0 = k\Delta t$ where k takes on integer values between 0 and $M = T/\Delta t$. Having made this definition, marginalizing over coalescence time becomes summing over k. The variable κ^2 is a function of (discretized) coalescence time k. We can write in terms of a fast Fourier transform.

$$\begin{aligned}\kappa^2(k) &= 4\Delta f \Re \sum_j^M \frac{d_j^* \mu_j(t_0)}{P_j} \exp\left(-2\pi i j \frac{k}{M}\right) \\ &= 4\Delta f \Re\, \mathtt{fft}_k \left(\frac{d_j^* \mu_j(t_0)}{P_j}\right).\end{aligned} \tag{2.51}$$

Here $\mathtt{fft}_k$ refers to the k bin of a fast Fourier transform.

The other terms in 2.44 are independent of the time at coalescence of the template. The marginalized likelihood is therefore

$$\begin{aligned}\log \mathcal{L}^t_{\mathrm{marg}} &= \log \int_{t_0}^{t_0+T} dt\, \mathcal{L}(\theta, t)\pi(t) \\ &= \log \mathcal{Z}_N - \frac{1}{2}\rho^2_{\mathrm{opt}}(\theta) + \log \int_{t_0}^{t_0+T} dt\, e^{\kappa^2(\theta,t)}\pi(t) \\ &= \log \mathcal{Z}_N - \frac{1}{2}\rho^2_{\mathrm{opt}}(\theta) + \log \sum_k^M e^{\kappa^2(\theta,k)}\pi_k,\end{aligned} \tag{2.52}$$

where $\pi_k = \pi(t)\Delta t$ is the prior on the discretized coalescence time.

Caution should be taken to avoid edge effects. If we employ a naive prior, the waveform will exhibit unphysical wrap-around. Similarly, care must be taken to ensure that the time-shifted waveform is consistent with time-domain data conditioning, e.g., windowing. (This is usually not a problem for confident detections because the coalescence time is well-known and so the segment edges can be avoided.) A good solution is to choose a suitable prior, which is uniform over some values of k, but with some values set to zero in order to prevent the signal from wrapping around the edge of the data segment. Note that Eq. 2.49 breaks down for when the detector changes significantly over T due to the rotation of the Earth. It can also fail in the high signal-to-noise ratio limit when the t array becomes insufficiently fine-grained.

2.8.2 Phase marginalization

In this subsection, we follow Veitch & Del Pozzo, 2013 (see also Veitch et al., 2015) to derive a likelihood, which explicitly marginalizes over phase of coalescence ϕ_c. To begin, we assume a gravitational waveform approximant consisting entirely of the dominant $\ell = 2, |m| = 2$ modes so that[20]

$$\mu = \mu_{22} + \mu_{2-2}, \tag{2.54}$$

This is a valid assumption, e.g., for the widely used waveform approximants—e.g., TAYLORF2 (Damour, Iyer & Sathyaprakash, 2005), IMRPHENOMD (Khan et al., 2016), IMRPHENOMP (Hannam et al., 2014)—but not for waveforms that employ higher order modes, e.g., Blackman et al., 2017. Given this approximation[21],

$$\mu(\phi_c) = e^{2i\phi_c}\mu(\phi_c = 0). \tag{2.56}$$

The optimal signal-to-noise ratio ρ_{opt} is invariant under rotations in ϕ_c. However the matched filter signal-to-noise ratio is not. Thus, the phase-marginalized likelihood is

$$\begin{aligned}\mathcal{L}^{\phi_c}_{\text{marg}} = \mathcal{Z}_N - \exp\left(\frac{1}{2}\rho^2_{\text{opt}}\right) \\ + \int_0^{2\pi} d\phi_c \exp\left(\frac{1}{2}\langle d, \mu(\phi_c)\rangle + \frac{1}{2}\langle \mu(\phi_c), d\rangle\right)\pi(\phi_c).\end{aligned} \tag{2.57}$$

[20] The variables μ_{22} and μ_{2-2} are defined like so

$$\begin{aligned}\mu_{\ell m} \equiv & F_+ \Re\Big(h_{\ell m}(\theta)\ {}_{-2}Y_{\ell m}(\iota, \phi)\Big) \\ & + F_\times \Im\Big(h_{\ell m}(\theta)\ {}_{-2}Y_{\ell m}(\iota, \phi)\Big).\end{aligned} \tag{2.53}$$

They depend on the metric perturbation $h_{\ell m}$ and the antenna response functions $F_{+,\times}$. The variable ${}_{-2}Y_{\ell m}(\iota, \phi)$ is a spin-weighted spherical harmonic function, evaluated the inclination angle ι and aziumuthal angle ϕ of the observer. Without loss of generality, we can set $\phi = 0$, which establishes a coordinate frame. Having defined this frame, we may rotate the binary by the phase of coalescence ϕ_c in order to change the phase of the signal observed at Earth.

[21] We emphasize that the phase at coalescence is distinct from ϕ, the azimuthal angle to the observer in the source frame, which transforms differently

$$\mu(\phi) = e^{2i\phi}\mu_{22}(\phi = 0) + e^{-2i\phi}\mu_{2-2}(\phi = 0). \tag{2.55}$$

The variable ϕ_c calibrates the time evolution of the gravitational waveform observed at Earth, while ϕ describes how the the waveform varies at a fixed time for observers at different spatial locations (corresponding to different azimuthal angles).

Using Eq. 2.56, we can rewrite the phase-marginalized likelihood

$$\mathcal{L}^{\phi_c}_{\text{marg}} = \int_0^{2\pi} d\phi_c \, \exp\Big(\frac{1}{2}\big\langle d, \mu(\phi_c = 0)\big\rangle_{\mathbb{C}} \exp(2i\phi_c) + \\ \frac{1}{2}\big\langle \mu(\phi_c = 0), d\big\rangle_{\mathbb{C}} \exp(-2i\phi_c)\Big)\pi(\phi_c) \\ + \ldots$$

The parts that do not depend on ϕ_c are implied by the ellipsis. Here we introduce the "complex inner product" denoted with a subscript $\mathbb{C}$.

$$\langle a, b\rangle_{\mathbb{C}} \equiv 4\Delta f \sum_j \left(\frac{a_j^* b_j}{P_j}\right), \tag{2.58}$$

which is identical to the regular inner product defined in Eq. 2.42 except we do not take the real part in order to preserve phase information that will be useful later on. Employing a uniform prior on ϕ_c and grouping terms, the integral can be rewritten yet again

$$\mathcal{L}^{\phi_c}_{\text{marg}} = \int_0^{2\pi} \frac{d\phi_c}{2\pi} \exp\Big(A\cos(2\phi_c) + B\sin(2\phi_c)\Big) + \ldots \tag{2.59}$$

where

$$A \equiv \Re\big\langle d, \mu(\phi_c = 0)\big\rangle_{\mathbb{C}} \tag{2.60}$$
$$B \equiv \Im\big\langle d, \mu(\phi_c = 0)\big\rangle_{\mathbb{C}}. \tag{2.61}$$

The integral yields modified Bessel function of the first kind

$$I_0\left(\sqrt{A^2 + B^2}\right) = \frac{1}{2\pi}\int_0^{2\pi} d\phi \, e^{Ac_\phi + Bs_\phi}. \tag{2.62}$$

Thus

$$\begin{aligned}\sqrt{A^2 + B^2} &= \sqrt{\Re\big\langle d, \mu(0)\big\rangle^2_{\mathbb{C}} + \Im\big\langle d, \mu(\phi_c = 0)\big\rangle^2_{\mathbb{C}}} \\ &= \Big|\big\langle d, \mu(\phi_c = 0)\big\rangle_{\mathbb{C}}\Big| \\ &= \Big|\kappa^2_{\mathbb{C}}\Big|,\end{aligned} \tag{2.63}$$

where $\kappa^2_{\mathbb{C}}$ is calculated the same way as κ (Eq. 2.47), except we use a complex inner product. The ϕ_c marginalized likelihood becomes

$$\log \mathcal{L}^{\phi}_{\text{marg}} = \log \mathcal{Z}_N - \frac{1}{2}\rho^2_{\text{opt}} + \log I_0(|\kappa^2_{\mathbb{C}}|). \tag{2.64}$$

We reiterate that this marginalized likelihood is valid only insofar as we trust our initial assumption, that the signal is dominated by $l = 2, |m| = 2$ modes.

2.8.3 Distance marginalization

In this subsection, we follow Singer & Price, 2016 (see also Singer, 2016) to derive a likelihood, which explicitly marginalizes over luminosity distance D_L. Given a waveform at some reference distance $\mu(D_0)$, the waveform at an arbitrary distance is obtained by multiplication of a scale factor

$$\mu_j(D_L) = \mu_j(D_0)\left(\frac{D_0}{D_L}\right). \tag{2.65}$$

As before, it is understood that μ is a function of whatever parameters are not explicitly marginalizing over. Unlike time and phase, distance affects ρ_{opt} in addition to κ^2 (Eq. 2.47),

$$\begin{aligned}\kappa^2(D_L) &= \kappa^2(D_0)\left(\frac{D_0}{D_L}\right),\\ \rho^2_{\text{opt}}(D_L) &= \rho^2_{\text{opt}}(D_0)\left(\frac{D_0}{D_L}\right)^2.\end{aligned} \tag{2.66}$$

Note that κ^2 and ρ_{opt} are implicit functions of whatever parameters we are not explicitly marginalizing over.

At a fixed distance, the likelihood is

$$\log\mathcal{L}(D_L) = \log\mathcal{Z}_N + \kappa^2(D_L) - \frac{1}{2}\rho^2_{\text{opt}}(D_L), \tag{2.67}$$

and the likelihood marginalized over luminosity distance is

$$\log\mathcal{L}^D_{\text{marg}} = \log\mathcal{Z}_N + \log\mathcal{L}_D, \tag{2.68}$$

where

$$\mathcal{L}_D(\kappa^2, \rho_{\text{opt}}) \equiv \int dD_L\, e^{\kappa^2(D_L) - \frac{1}{2}\rho^2_{\text{opt}}(D_L)}\pi(D_L). \tag{2.69}$$

This integral to calculate $\log\mathcal{L}_D$ can be evaluated numerically. This explicitly marginalized form is generally true for all gravitational-waves sources. Its validity is only limited by the resolution of the numerical integral, though, cosmological redshifts adds additional complications, which we discuss in the next subsection. One can construct a pre-computed lookup table $\log\mathcal{L}_D(\rho_{\text{mf}}, \rho_{\text{opt}})$ to facilitate fast and precise evaluation.

2.8.4 Distance marginalization with cosmological effects

There is a caveat for our discussion of distance marginalization in the previous subsection: when considering events at cosmological distances, the prior distributions for lab-frame masses become covariant with luminosity distance D_L due to cosmological redshift. A signal emitted with source-frame mass m_s is observed with lab-frame mass given by

$$m_l = (1+z)m_s. \tag{2.70}$$

In this subsection, "mass" m is shorthand for an array of both primary and secondary mass.

Now we derive an expression for $\mathcal{L}^D_{\text{marg}}$, which can be applied to cosmological distances. We start by specifying the prior on redshift and source-frame mass[22]:

$$\pi(z, m_s) = \pi(z)\pi(m_s). \tag{2.71}$$

Both $\pi(z)$ and $\pi(m_s)$ can be chosen using astrophysically motivated priors; see e.g., Talbot & Thrane, 2018; Fishbach & Holz, 2017; Fishbach, Holz & Farr, 2018. Whatever priors we choose for $\pi(z)$ and $\pi(m_s)$, they imply some prior for the lab-frame mass:

$$\begin{aligned}\pi(z, m_l) =& \pi\big(z, m_l/(1+z)\big)\left|\frac{dm_s}{dm_l}\right| \\ =& (1+z)^{-1}\pi\big(z, m_l/(1+z)\big).\end{aligned} \tag{2.72}$$

Now that we have converted the source-frame prior into a lab-frame prior, we can write down the distance-marginalized (redshift-marginalized) likelihood in terms of lab-frame quantities:

$$\mathcal{L}^z_{\text{marge}}(\kappa^2, \rho_{\text{opt}}) = \int dz\, \mathcal{L}(\kappa^2, \rho_{\text{opt}}, z)\pi(z|m_l), \tag{2.73}$$

where

$$\mathcal{L}(\kappa^2, \rho_{\text{opt}}, z) = \mathcal{Z}_N\, e^{\kappa^2(D_L(z)) - \frac{1}{2}\rho^2_{\text{opt}}(D_L(z))}. \tag{2.74}$$

Note that κ^2 and ρ_{opt} are implicit functions of whatever parameters we are not explicitly marginalizing over.

By creating a grid of z, we can create a look-up table for $\mathcal{L}(\kappa^2, \rho_{\text{opt}}, z)$, which allows for rapid evaluation of Eq. 2.73. However, this means we will also need to create a look-up table for $\pi(z|m_l)$. In order to derive this look-up table, we rewrite the joint prior on redshift and lab-frame mass can be rewritten like so

$$\pi(z, m_l) = \pi(z|m_l)\pi(m_l). \tag{2.75}$$

The marginalized lab-mass prior is

$$\pi(m_l) \equiv \int dz\, \pi(z, m_l), \tag{2.76}$$

[22]Many previous analyses have assumed that this distribution is separable, however this marginalization technique does not require this.

which can be calculated numerically. (We also need this distribution to provide to the sampler.) Thus, the conditional prior we need for our look-up table is:

$$\pi(z|m_l) = \pi(z, m_l)/\pi(m_l). \tag{2.77}$$

With look-up tables for $\mathcal{L}(\kappa^2, \rho_{\rm opt}, z)$ and $\pi(z|m_l)$, the sampler can quickly evaluate $\mathcal{L}^z_{\rm marge}$ by summing over the grid of z:

$$\mathcal{L}^z_{\rm marg}(\kappa^2, \rho_{\rm opt}) = \Delta z \sum_k \mathcal{L}(\kappa^2, \rho_{\rm opt}, z_k)\pi(z_k|m_l), \tag{2.78}$$

where Δz is the spacing of the redshift grid. This allows us to carry out explicit distance marginalization while taking into account cosmological redshift.

2.8.5 Marginalization with multiple parameters

One must take care with the order of operations when implementing these marginalization schemes simultaneously. We describe how to combine the three marginalization techniques described above. The correct procedure is to start with Eq. 2.64 and then marginalize over distance.

$$\begin{aligned} \log \mathcal{L}^{\phi,D}_{\rm marg} =& \log \mathcal{Z}_N \\ &+ \log \int dD_L e^{I_0(|\kappa^2_{\mathbb{C}}(D_L)|) - \frac{1}{2}\rho^2_{\rm opt}(D_L)} \pi(dD_L). \end{aligned} \tag{2.79}$$

Carrying out this integral numerically, one obtains a look-up table $\log \mathcal{L}^{\phi,D}_{\rm marge}(\kappa^2_{\mathbb{C}}, \rho_{\rm opt})$, which marginalizes over ϕ and D_L. Finally, we add in t marginalization by combining the look-up table with a fast Fourier transform

$$\mathcal{L}^{\phi,D,t}_{\rm marg}(\kappa^2_{\mathbb{C}}, \rho_{\rm opt}) = \sum_k \pi_k \, \mathcal{L}^{\phi,D}_{\rm marg}\big(\kappa^2_{\mathbb{C}}(k), \rho_{\rm opt}(k)\big). \tag{2.80}$$

2.8.6 Reconstructing the unmarginalized posterior

While explicitly marginalizing over parameters improves convergence and reduces runtime, the sampler will generate no posterior samples for the marginalized parameters. Sometimes, we want posterior samples for these parameters. In this subsection we explain how it is possible to generate them with an additional post-processing step.

The parameter we are most likely to be interested in reconstructing is the luminosity distance D_L. Let us assume for the moment that this is the only parameter over which we have explicitly marginalized. The first step to calculate the matched filter signal-to-noise ratio $\rho_{\rm mf}$ and optimal signal-to-noise ratio $\rho_{\rm opt}$ for each sample. For one posterior sample k, the likelihood for distance is

$$\mathcal{L}_k(d|D_L) = \mathcal{Z}_N \, e^{\kappa^2(\theta_k, D_L) - \frac{1}{2}\rho^2_{\rm opt}(\theta_k, D_L)}, \tag{2.81}$$

where $\kappa^2(D_L)$ and $\rho_{\rm opt}(D_L)$ are defined in Eq. 2.66. (When comparing with Eq. 2.66, note that we have again made explicit the dependence on θ_k = whatever parameters we are not explicitly marginalizing over.) Since this likelihood is one-dimensional, it is easy to calculate the posterior for sample k using Bayes' theorem:

$$p_k(D_L|d) = \frac{\mathcal{L}(d|D_L)\pi(D_L)}{\int dL\, \mathcal{L}(d|D_L)\pi(D_L)}. \tag{2.82}$$

Using the posterior, one can construct a cumulative posterior distribution for sample k:

$$P_k(D_L|d) = \int dD_L\, p_k(D_L|d). \tag{2.83}$$

The integral can be carried out numerically. The cumulative posterior distribution can be used to generate random values of D_L for each posterior sample.

$$D_L = P_k^{-1}(\texttt{rand}) \tag{2.84}$$

Reconstructing the likelihood or posterior when multiple parameters have been explicitly marginalized over is more complicated. However, one may use the following iterative algorithm.

1. For each sample θ_k marginalize over all originally marginalized parameters except one (λ).
2. Draw a single λ sample from the marginalized likelihood times prior.
3. Add this λ sample to the θ_k and return to step 1, this time not marginalizing over λ.

Alternatively, one can skip the step of generating new samples in distance and calculate the likelihood of the data given D_L marginalized over all other parameters,

$$\begin{aligned}\mathcal{L}(d|D_L) =& \frac{1}{n}\sum_k^n \mathcal{L}_k(d|D_L) \\ =& \frac{\mathcal{Z}_N}{n}\sum_k^n e^{\kappa^2(\theta_k,D_L)-\frac{1}{2}\rho^2_{\rm opt}(\theta_k,D_L)}.\end{aligned} \tag{2.85}$$

This likelihood can be used in Eq. 2.29 to perform population inference on the distribution of source distances and/or redshifts.

2.9 Posterior predictive distributions

The posterior predictive distribution (PPD) represents the updated prior on the parameters θ given the data d. Recall that the hyper-posterior $p(\Lambda|d)$ describes our post-measurement knowledge of the hyper-parameters that describe the shape of the prior distribution $\pi(\theta)$. The PPD answers the question: given this hyper-posterior, what does the distribution of $\pi(\theta)$ look like? More precisely, it is the probability that the next event will have true parameter values θ given what we have learned about the population hyper-parameters Λ

$$p_\Lambda(\theta|d) = \int d\Lambda\, p(\Lambda|d)\, \pi(\theta|\Lambda). \tag{2.86}$$

The Λ subscript helps us distinguish the PPD from the posterior $p(\theta|d)$. The hyper-posterior sample version is

$$p_\Lambda(\theta|d) = \frac{1}{n_s}\sum_k^{n_s} \pi(\theta|\Lambda_k), \tag{2.87}$$

where k runs over n_s hyper-posterior samples. While the PPD is the best guess for what the distribution $\pi(\theta)$ looks like, it does not communicate information about the variability possible in $\pi(\theta)$ given uncertainty in Λ. In order to convey this information, it can be useful to overplot many realizations of $\pi(\theta|\Lambda_k)$ where Λ_k is a randomly selected hyper-posterior sample. An example of a PPD is included in Fig. 2.3.

2.10 Selection Effects

In this section, we discuss how to carry out inference while taking into account selection effects, which arise from the fact that some events are easier to detect than others. We loosely follow the arguments from Abbott, 2016b; however, see also Mandel, Farr & Gair, 2019; Fishbach, Holz & Farr, 2018. In Subsection 2.10.1, we discuss selection effects in the context of an individual detection. In Subsection 2.10.2, we generalize these results to populations of events.

2.10.1 Selection effects with a single event

Some gravitational-wave events are easier to detect than others. All else equal, it is easier to detect binaries if they are closer, higher mass (at least, up until the point that they start to go out of the observing band), and with face-on/off inclination angles. More subtle selection effects arise due to black hole spin (e.g., Ng et al., 2018). Typically, a gravitational-wave event is said to have been detected if it is observed with a matched-filter signal-to-noise ratio—maximized over extrinsic parameters $\theta_{\text{extrinsic}}$—above some threshold

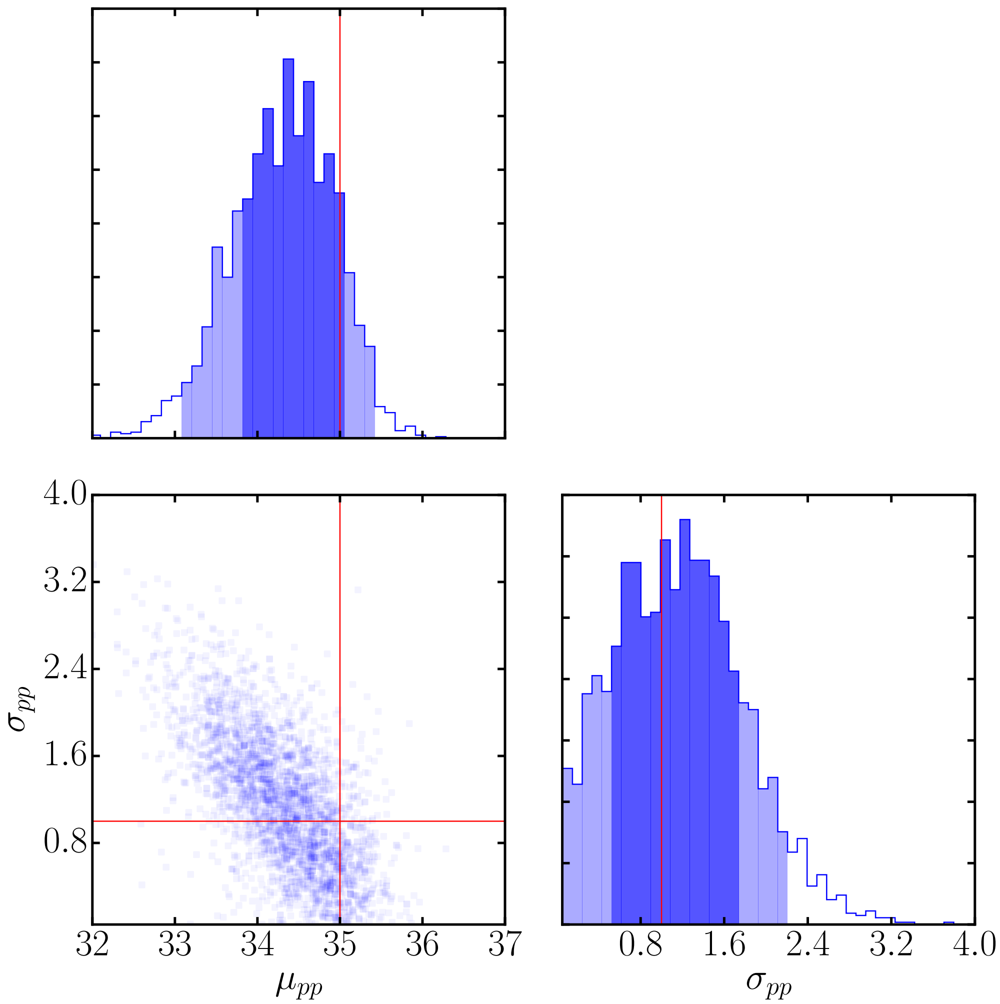

FIGURE 2.2: An example corner plot from Talbot & Thrane, 2018 showing posteriors for hyper-parameters $\mu_{\rm pp}$ and $\sigma_{\rm pp}$. Respectively, these two hyper-parameters describe the mean and width of a peak in the primary mass spectrum due to the presence of pulsational pair instability supernovae.

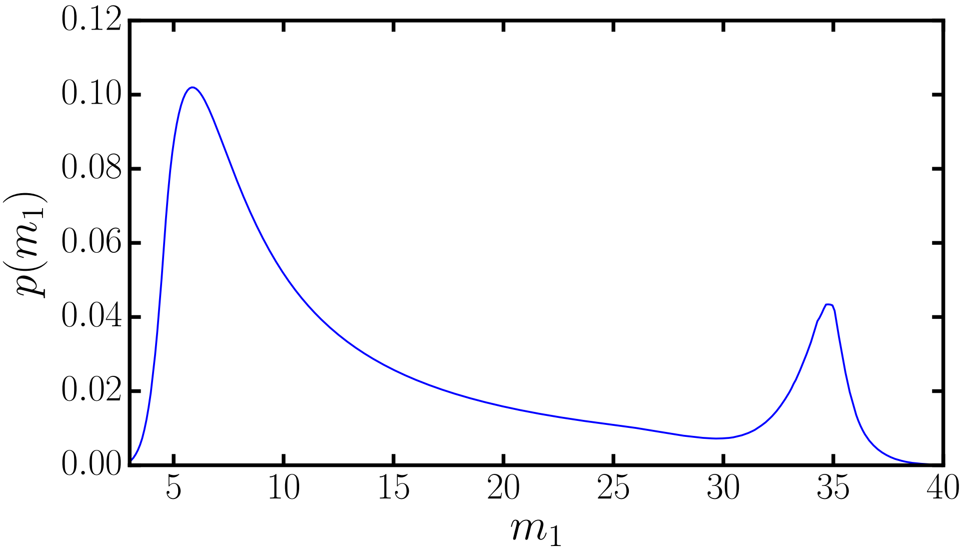

FIGURE 2.3: An example of a posterior predictive distribution (PPD) for primary black hole mass, calculated using the hyper-posterior distributions in the top panel (adapted from Talbot & Thrane, 2018). The PPD has a peak near $m_1 = 35$ because the hyper-posterior for μ_{pp} is maximal near this value. The width of the PPD peak is consistent with the hyper-posterior for σ_{pp}.

ρ_{th}

$$\rho'_{\text{mf}} \equiv \max_{\theta_{\text{extrinsic}}} (\rho_{\text{mf}}) > \rho_{\text{th}}. \tag{2.88}$$

Usually, $\rho_{\text{th}} = 8$ for a single detector or $\rho_{\text{th}} = 12$ for a ≥ 2 detector network.

Focusing on events with a $\rho_{\text{mf}} > \rho_{\text{th}}$ detection forces us to modify the likelihood function

$$\mathcal{L}(\mathbf{d}|\theta, \text{det}) = \begin{cases} \frac{1}{p_{\text{det}}(\theta)} \mathcal{L}(\mathbf{d}|\theta) & \rho'_{\text{mf}}(\theta) \geq \rho_{\text{th}} \\ 0 & \rho'_{\text{mf}}(\theta) < \rho_{\text{th}} \end{cases}, \tag{2.89}$$

where

$$p_{\text{det}}(\theta) \equiv \int_{\rho'_{\text{mf}}(\theta) > \rho_{\text{th}}} d\mathbf{d}\, \mathcal{L}(\mathbf{d}|\theta). \tag{2.90}$$

(Here, we temporarily switch to data=$\mathbf{d}$ to avoid confusing data with the differential d; we switch back to data=d in a moment once we are finished with this normalization constant.) This modification enforces the fact that we are not looking at data with $\rho'_{\text{mf}} < \rho_{\text{th}}$. The p_{det} factor ensures that the likelihood is properly normalized.

There are different ways to calculate p_{det} in practice. The probability density function for ρ_{mf} given θ—the distribution of ρ_{mf} arising from random

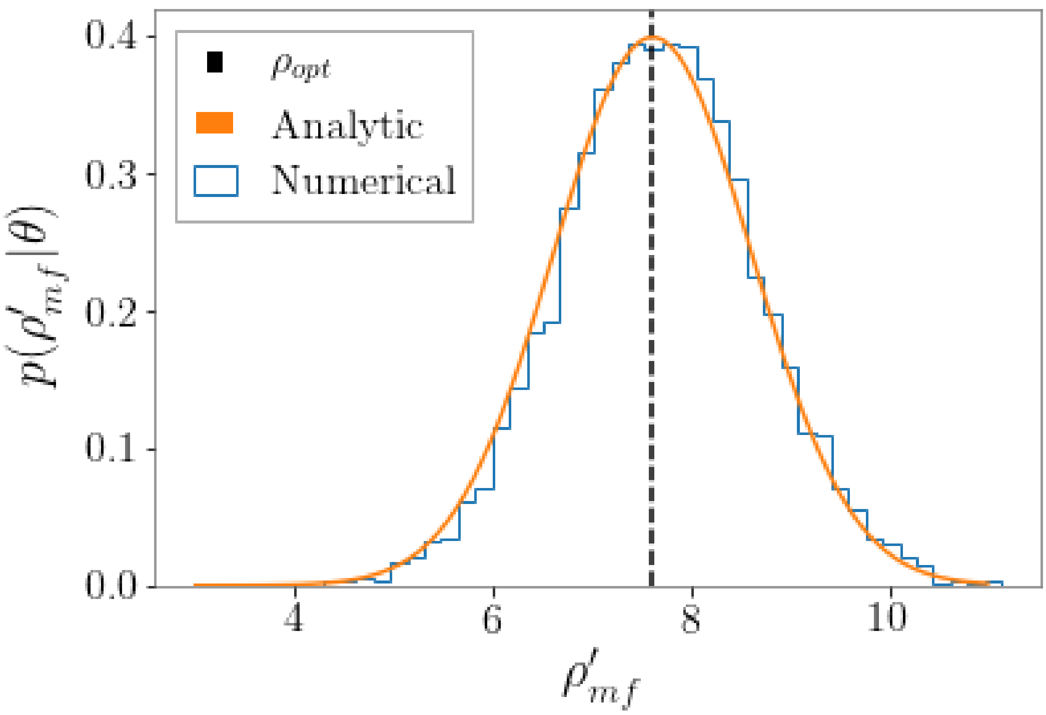

FIGURE 2.4: The distribution of matched filter signal-to-noise ratio maximized over phase for the same template in many noise realisations (blue). The distribution peaks at $\rho_{\mathrm{opt}} = 7.6$ (dashed black). The theoretical distribution (Eq. 2.91) is shown in orange.

noise fluctuations—is a normal distribution with mean ρ_{opt} and unit variance

$$p(\rho'_{\mathrm{mf}}|\theta) = \frac{1}{2\pi}\exp\left(-\frac{1}{2}\left(\rho'_{\mathrm{mf}} - \rho_{\mathrm{opt}}(\theta)\right)^2\right), \tag{2.91}$$

see Fig. 2.4. Thus,

$$p_{\mathrm{det}}(\theta) = \int_{\rho_{\mathrm{th}}}^{\infty} dx \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(x - \rho_{\mathrm{opt}}(\theta)\right)^2\right) \tag{2.92}$$

$$= \frac{1}{2}\operatorname{erfc}\left(\frac{\rho_{\mathrm{th}} - \rho_{\mathrm{opt}}(\theta)}{\sqrt{2}}\right). \tag{2.93}$$

Alternatively, if we are interested in the selection effects of intrinsic parameters, one may express p_{det} as the ratio of the “visible volume” $\mathcal{V}(\theta)$ to the total spacetime volume $\mathcal{V}_{\mathrm{tot}}$

$$p_{\mathrm{det}}(\theta) = \frac{\mathcal{V}(\theta)}{\mathcal{V}_{\mathrm{tot}}}. \tag{2.94}$$

The visible volume is typically calculated numerically with injected signals.

2.10.2 Selection effects with a population of events

When considering a population of events, Eq. 2.89 generalizes to

$$\mathcal{L}(d, N|\Lambda, \text{det}) = \begin{cases} \frac{1}{p_{\text{det}}(\Lambda|N)} \mathcal{L}(d, N|\Lambda, R). & \rho_{\text{mf}} \geq \rho_{\text{th}} \\ 0 & \rho_{\text{mf}} < \rho_{\text{th}} \end{cases}. \tag{2.95}$$

In analogy to Eq. 2.94, the p_{det} normalization factor can be calculated using the visible volume as a function of the hyper-parameters Λ

$$\mathcal{V}(\Lambda) \equiv \int d\theta \mathcal{V}(\Lambda) \pi(\theta|\Lambda). \tag{2.96}$$

Naively, one might expect that

$$p_{\text{det}}(\Lambda|N) = \left(\frac{\mathcal{V}(\Lambda)}{\mathcal{V}_{\text{tot}}}\right)^N, \tag{2.97}$$

but this expression is incorrect because it does not marginalize over the Poisson-distributed rate, which ends up changing the answer. Marginalizing over the rate, we obtain

$$\begin{aligned} p_{\text{det}}(\Lambda|N) &= \int dR \left(\frac{\mathcal{V}(\Lambda)}{\mathcal{V}_{\text{tot}}}\right)^N \pi(N|R)\pi(R) \\ &= \int dR \left(\frac{\mathcal{V}(\Lambda)}{\mathcal{V}_{\text{tot}}}\right)^N \left[e^{-R\mathcal{V}(\Lambda)} \frac{\mathcal{V}(\Lambda)^N R^N}{N!}\right] \pi(R) \\ &= \left(\frac{\mathcal{V}(\Lambda)}{\mathcal{V}_{\text{tot}}}\right)^N \left[\int dR\, e^{-R\mathcal{V}(\Lambda)} \frac{\mathcal{V}(\Lambda)^N R^N}{N!}\right] \pi(R). \end{aligned} \tag{2.98}$$

Note that p_{det} depends on our prior for the rate R. If we choose a uniform-in-log prior $\pi(R) \propto 1/R$, we obtain

$$p_{\text{det}}(\Lambda|N) \propto \left(\frac{\mathcal{V}(\Lambda)}{\mathcal{V}_{\text{tot}}}\right)^N, \tag{2.99}$$

which reproduces the results from Abbott, 2019a. Note that

$$\mathcal{L}(d|\Lambda, \text{det}) \neq \int d\theta \mathcal{L}(d|\theta, \text{det})\, \pi(\theta|\Lambda). \tag{2.100}$$

Addendum

Since the publication of this paper additional compact binary coalescences have been observed and significant progress has been made in theoretical and observational analysis. The reader is directed to Chapters 1 and 8 for an overview of the field at the time of writing.

Chapter 3

Determining the population properties of spinning black holes

Published as:
Talbot, C. & Thrane, E., Phys. Rev. D 96, 023012 (2017).

Abstract

There are at least two formation scenarios consistent with the first gravitational-wave observations of binary black hole mergers. In field models, black hole binaries are formed from stellar binaries that may undergo common envelope evolution. In dynamic models, black hole binaries are formed through capture events in globular clusters. Both classes of models are subject to significant theoretical uncertainties. Nonetheless, the conventional wisdom holds that the distribution of spin orientations of dynamically merging black holes is nearly isotropic while field-model black holes prefer to spin in alignment with the orbital angular momentum. We present a framework in which observations of black hole mergers can be used to measure ensemble properties of black hole spin such as the typical black hole spin misalignment. We show how to obtain constraints on population hyperparameters using minimal assumptions so that the results are not strongly dependent on the uncertain physics of formation models. These data-driven constraints will facilitate tests of theoretical models and help determine the formation history of binary black holes using information encoded in their observed spins. We demonstrate that the ensemble properties of binary detections can be used to search for and characterize the properties of two distinct populations of black hole mergers.

3.1 Introduction

At present, merging black holes are the only directly detected source of gravitational waves (Abbott, 2016h; Abbott, 2016i; Abbott, 2016f; Abbott, 2016c; Abbott, 2016g; Abbott, 2016b; Abbott, 2017b; Abbott, 2016a; Abbott, 2016e; Abbott, 2017f). A variety of mechanisms by which black hole binaries can form have been proposed. These mechanisms might yield significantly different distributions of the intrinsic parameters of binaries (Rodriguez et al., 2016b). In this work we focus on the distribution of spin orientations to probe black hole binary formation mechanisms. We consider two mechanisms which are expected to dominate, the field and dynamical models (see, e.g., Mandel & O'Shaughnessy, 2010 for a detailed review).

In dynamical models, the binary forms when two black holes become gravitationally bound in dense stellar environments such as globular clusters (Heggie, 1975). Due to mass segregation such clusters arrange themselves with more massive objects being found in the center and less massive objects on the outside. This means that binaries are expected to have mass ratios close to unity (Sigurdsson & Hernquist, 1993). It is expected that the spins of the two companions will be isotropically oriented (Rodriguez et al., 2016b).

The distribution of spin orientations in field models is subject to more theoretical uncertainty (e.g., Postnov & Yungelson, 2014). In field models, a stellar binary forms and the components of the binary then coevolve. Although such stars are expected to form with their angular momenta aligned with the total angular momentum of the binary, there are exceptions (e.g., Albrecht et al., 2011; Albrecht et al., 2014). If binaries are formed with misaligned spins, tidal interactions and mass transfer processes between the stars can align the angular momenta of the stars with the total angular momentum of the binary (e.g., Bardeen & Petterson, 1975; Hut, 1981). When the first star explodes in a supernova and collapses to form a black hole, a natal kick may be imparted on the two companions due to asymmetry of the explosion (e.g., Janka, 2012), increasing misalignment between spin and angular momentum vectors. The subsequent evolution of the secondary, possibly involving a common envelope phase, can reverse this misalignment (Ivanova et al., 2013). This is followed by the supernova of the secondary, which may give each black hole another kick and some additional degree of misalignment. The net effect is to leave the population of black hole spin orientations distributed about the angular momentum vector of the binary with some unknown typical misalignment angle (Kalogera, 2000; Repetto, Davies & Sigurdsson, 2012; Rodriguez et al., 2016b; O'Shaughnessy, Gerosa & Wysocki, 2017).

Following the formation of the black hole binary (either through dynamical capture or common evolution) the spin orientation of nonaligned spinning black holes changes due to precession. Isotropic spin orientation distributions are expected to remain isotropic throughout such evolution (Bogdanović, Reynolds & Miller, 2007). However, anisotropic distributions, such as those predicted by field models, may change significantly (Schnittman, 2004; Kesden et al., 2015; Gerosa et al., 2015). Here, we are interested in the distribution of spin orientations at the moment the binary enters LIGO's observing band. We therefore measure our spin orientations at $f_{\mathrm{ref}} = 20\,\mathrm{Hz}$. Advanced LIGO's observing band will eventually extend down to 10 Hz, but we use 20 Hz here for the sake of convenience. One may use the spin orientation at f_{ref} to reverse engineer the spin alignment distribution at the moment of formation, but this is not our present goal.

In this paper, we use Bayesian hierarchical modeling (e.g., Gelman et al., 2013) and model selection to infer the parameters describing the distribution of spins of black hole binaries. We construct a mixture model, which treats the fraction of dynamical mergers, the fraction of isolated binary mergers, and the typical spin misalignment of the primary and secondary black holes as free parameters. We apply the model to simulated data (including noise) to show that we can both detect the presence of distinct populations, and also measure hyperparameters describing typical spin misalignment.

Our method builds on a body of research using gravitational waves to study the ensemble properties of compact binaries. In Stevenson, Ohme & Fairhurst, 2015, it was shown that Bayesian model selection can be used to distinguish between formation channels using nonparametrized mass distributions. Clustering was used in Mandel et al., 2017 to show that model-independent statements about the existence of distinct mass subpopulations can be made with an ensemble of detections. In Fishbach, Holz & Farr, 2017; Gerosa & Berti, 2017, it was shown that the spin magnitude distribution can be used to determine whether observed merging black holes formed through hierarchical mergers of smaller black holes. Hierarchical merger models predict an isotropic distribution of black hole spin orientations since all binaries form through dynamical capture.

Vitale et al., 2017b showed that model selection can be used to distinguish between models which predict mutually exclusive spin orientations of merging compact binaries, both binary black holes and neutron star black hole binaries. In order to generate two distinct populations with different spin distributions, binaries were generated with random spin angles. Those with tilt angles (between the black hole spin and the Newtonian orbital angular momentum) $< 10^\circ$ were considered to be a fieldlike binary while those with tilt angles $> 10^\circ$ were considered to be dynamiclike. The authors showed that, after ~ 100 detections, one can recover the proportion of binaries in each population to within $\sim 10\%$ at 1σ.

Stevenson, Berry & Mandel, 2017 used Bayesian hierarchical modeling to recover the proportion of binaries taken from a set of four populations distributed according to astrophysically motivated, spin orientation distributions with fixed spin magnitudes ($a_i = 0.7$). Unlike Vitale et al., 2017b, the populations overlap so that even precise knowledge of a binary's spin parameters does not provide certain knowledge about its parent population. Of the four populations, three are different distributions predicted by population synthesis models of isolated binary evolution and the fourth is the isotropic distribution predicted for dynamic formation. They achieve a similar result to Vitale et al., 2017b, measuring the relative proportion of different populations at the $\sim$ 10% level after 100 events. They also demonstrate that their two “extreme hypotheses" (perfect alignment and isotropy) can be ruled out at $> 5\sigma$ after as few as five events if they are not good descriptions of nature.

We build on these studies by employing a (hyper)parametrized model of the spin orientation distribution for the field model in order to measure not just the fraction of binaries from different populations, but also properties of the field model. In particular, we aim to *measure* the typical black hole misalignment for black hole binaries formed in the field. The advantage of this approach is that our modeling employs a broadly accepted idea from theoretical modeling (black holes in field binaries should be somewhat aligned) without assuming less certain details about the size of the misalignment. Since our model is agnostic with respect to the detailed physics of binary formation and subsequent evolution, the resulting methodology is robust against theoretical bias and provides a *measurement* of black hole spin misalignment for binaries formed in the field.

The remainder of the paper is organized as follows. In the next section we review how the properties of merging binary black holes are recovered from observed data and briefly discuss the current observational results. We then introduce a useful parametrization to describe an admixture of field and dynamical black hole mergers. We follow this with a description hierarchical inference. We then present the results of a proof-of-principle study using simulated data. We introduce a new tool for visualizing spin orientations, spin maps. Finally, closing thoughts are provided.

3.2 Gravitational-wave parameter estimation

In order to determine the parameters describing the sources of gravitational waves Θ from gravitational-wave strain data h, we employ Bayesian inference. Merging binary black hole waveforms are described by 15 parameters: two masses $\{m_1, m_2\}$, two three-dimensional spin vectors $\{\mathbf{S}_1, \mathbf{S}_2\}$, and seven additional parameters to specify the position and orientation of the source relative to Earth. It is possible that in both the field and dynamical formation models the presence of a third companion will induce eccentricity when the binary enters LIGO’s observing band through Lidov-Kozai cycles (Antonini, Toonen & Hamers, 2017; Toonen, Hamers & Zwart, 2016; Lidov, 1962; Kozai, 1962; Wen, 2003). However, we consider only circular binaries. Most gravitational-wave parameter estimation results obtained to date have been obtained using the Bayesian parameter estimation code LALINFERENCE (Veitch et al., 2015). For our study we use the LALINFERENCE implementation of nested sampling (Skilling, 2004). We employ reduced order modeling and reduced order quadrature (Smith et al., 2016) to limit the computational time of the analysis.

Performing parameter estimation over this 15-dimensional space is computationally intensive. In order to maximize the efficiency sampling this high-dimensional space, the effect of the two spin vectors on the waveform is approximately represented using two spin parameters (Schmidt, Ohme & Hannam, 2015),

$$\begin{aligned}\chi_{\mathrm{eff}} =& \frac{a_1\cos(\theta_1) + q a_2\cos(\theta_2)}{1+q} \\ \chi_p =& \max\left(a_1\sin(\theta_1), \left(\frac{4q+3}{4+3q}\right) q a_2 \sin(\theta_2)\right).\end{aligned} \tag{3.1}$$

Here (a_1, a_2) are the dimensionless spin magnitudes, $q = m_2/m_1 < 1$ is the mass ratio and (θ_1, θ_2) are the angles between the spin angular momenta and the Newtonian orbital angular momentum of the binary. The variable χ_{eff} is "the effective spin parameter." When $\chi_{\mathrm{eff}} > 0$, the binary merges at a higher frequency than for $\chi_{\mathrm{eff}} = 0$ and hence spends more time in the observing band (Campanelli, Lousto & Zlochower, 2006). Similarly, binaries with $\chi_{\mathrm{eff}} < 0$ spend less time in the observing band. The variable χ_p describes the precession of the binary, which is manifest as a long-period modulation of the signal (Apostolatos et al., 1994).

Using numerical relativity to compute all of the waveforms necessary for parameter estimation is computationally prohibitive. Parameter estimation therefore relies on "approximants," which can be used for rapid waveform estimation. We use the IMRPHENOMP approximant (Hannam et al., 2014), which has been used in many recent parameter estimation studies, including parameter estimation for recently observed binaries (e.g., Abbott, 2016i; Abbott, 2017b; Abbott, 2017f). IMRPHENOMP approximates a generically precessing binary waveform using χ_{eff} and χ_p. Parameter estimation of the confirmed binary black hole detections, GW150914 (Abbott, 2016g; Abbott, 2016i), GW151226 (Abbott, 2016f) and GW170104 (Abbott, 2017f), yield (slightly) informative posterior distributions for χ_{eff}. However, the posterior distributions for χ_p show no significant deviation from the prior.

The observed distribution of these two effective spin parameters will depend on the mass and spin magnitude distributions of black holes. The distributions are expected to differ for binaries formed through different mechanisms (Kalogera, 2000). We do not consider these effects. Instead we work directly with the spin orientations of each black hole. For our purposes, it will be useful to define two additional variables:

$$\begin{aligned} z_1 =& \cos(\theta_1) \\ z_2 =& \cos(\theta_2). \end{aligned} \tag{3.2}$$

Instead of working with χ_{eff} and χ_p, we work with distributions of z_1, z_2. We note that $z_i \approx 1$ corresponds to aligned spin while $z_i \approx -1$ corresponds to antialigned spin and $z_i = 0$ corresponds to black holes spinning in the orbital plane.

3.3 Models

For the purpose of this work we ignore the detailed formation history used in population synthesis studies. Instead, we introduce a simple parametrization designed to capture the salient features of the field and dynamic models. More sophisticated parametrizations are possible and will (eventually) be necessary to accurately describe realistic populations. However, we believe this is a suitable starting point given current theoretical uncertainty.

We hypothesize that the distribution of $\{z_1, z_2\}$ can be approximated as an admixture of two populations. The first population is described by a truncated Gaussian peaked at $(z_1, z_2) = (1, 1)$ with width (σ_1, σ_2). This is our proxy for the population formed in the field. The Gaussian shape mimics the form of distributions predicted by population synthesis models, which are clustered about $z = 1$ with some unknown spread. The second population is uniform in (z_1, z_2), this represents the dynamically formed population. The relative abundances of each population are given by ξ (field) and $1 - \xi$ (dynamic). Thus, according to our parametrization, the true distribution of black hole mergers can be approximately described as follows:

$$p_0(z_1, z_2) = \frac{1}{4}$$

$$p_1(z_1, z_2) = \frac{2}{\pi} \frac{1}{\sigma_1} \frac{e^{-(z_1-1)^2/2\sigma_1^2}}{\operatorname{erf}\left(\sqrt{2}/\sigma_1\right)} \frac{1}{\sigma_2} \frac{e^{-(z_2-1)^2/2\sigma_2^2}}{\operatorname{erf}\left(\sqrt{2}/\sigma_2\right)} \tag{3.3}$$

$$p(z_1, z_2) = (1 - \xi) p_0 + \xi p_1 \tag{3.4}$$

Here, $p_0(z_1, z_2)$ is the true dynamic-only distribution, $p_1(z_1, z_2)$ is the true field-only distribution, and $p(z_1, z_2)$ is the true distribution for all black hole binaries. These distributions depend on three hyperparameters: two widths (σ_1, σ_2) and one fraction ξ.

For each of our population hyperparameters $\{\sigma_1, \sigma_2, \xi\}$, we choose uniform prior distributions between 0 and 1. For ξ this covers the full allowed range of values. For σ, this prior is chosen to be consistent with the most conservative estimates on spin misalignments predicted by field models (isotropically distributed kicks with the same velocity distribution as neutron stars, isotropic full kicks in Rodriguez et al., 2016b). In Fig. 3.1, we plot p_1 for various values of σ.

There are two interesting limiting cases. We note that $p_1(z|\sigma) \to \delta(z - 1)$ as $\sigma \to 0$. This corresponds to perfect alignment of black hole spins. We also note that $p_1(z|\sigma) \to p_0$ as $\sigma \to \infty$. Thus, depending on the choice of prior, the dynamical model is degenerate with the field model evaluated at one point in hyperparameter space. A consequence of this limiting behavior is that it is far more difficult to distinguish samples drawn from a broad aligned distribution ($\sigma = 1$), than an *almost* perfectly aligned distribution ($\sigma = 0.01$). It is simple to extend this model to include more terms describing additional subpopulations or alter the form of the existing terms to better fit physically motivated distributions.

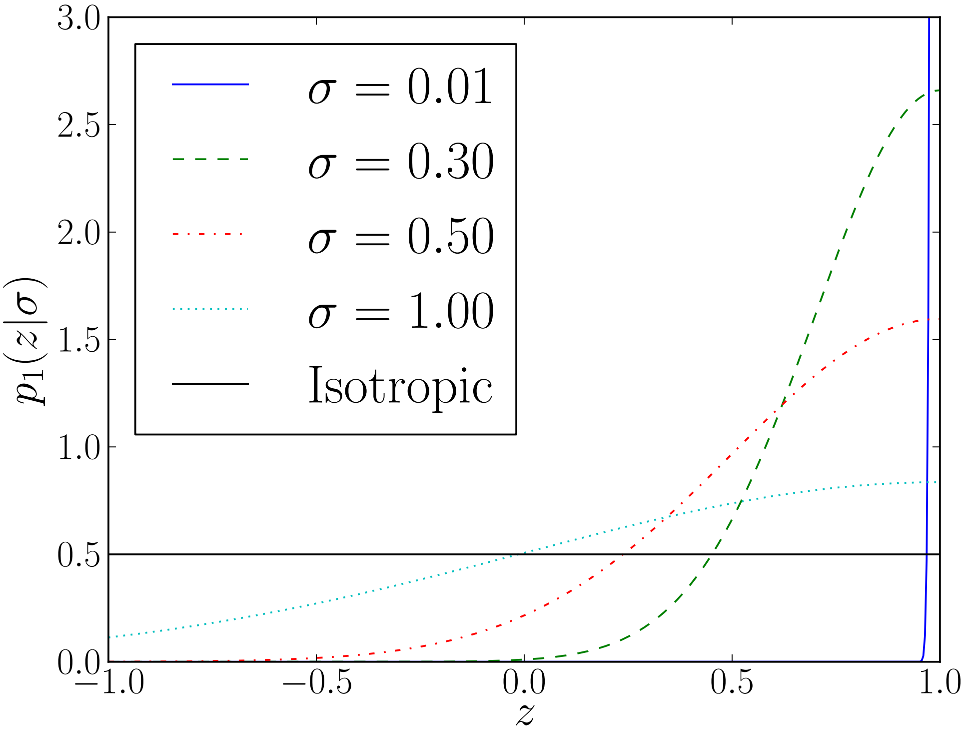

FIGURE 3.1: The distribution of z for our field model proxy with varying σ; see Eq. 3.3. By sending $\sigma \to 0$, we obtain perfect alignment and by sending $\sigma \to \infty$, we obtain an isotropic distribution.

3.4 Bayesian hierarchical modeling

Bayesian hierarchical modeling involves splitting a Bayesian inference problem into multiple stages. In the case of merging compact binaries these steps are as follows:

i Perform gravitational-wave parameter estimation as described above. We adopt priors that are uniform in spin magnitude and isotropic in spin orientations.

ii Assume the population from which events are drawn is described by hyperparameters Λ. Calculate a likelihood function for the data given Λ by marginalizing over the parameters for individual events Θ.

iii Combine multiple events to derive a joint likelihood for Λ.

iv Use the joint likelihood to derive posterior distributions for Λ, which, in turn, may be used to construct Bayes factors or odds ratios comparing different population models and confidence intervals on hyperparameters.

Step (i) produces a set of n_k posterior samples $\{\Theta_i\}$, sampled according to the likelihood of the binary having each set of parameters, $p(\Theta|h)$. This step is computationally expensive and requires the application of a specialized tool such as LALINFERENCE. In Step (ii), we estimate Λ using the posterior samples $\{z_i\}$. Our likelihood requires marginalization over z, for each event. Since LALINFERENCE approximates the posterior for Θ with a list of posterior sample points, the marginalization integral over (z_1, z_2) can be approximated by summing the probability of each sample in the LALINFERENCE posterior chain for our population model[1] (see, e.g., MacKay, 2002, Chapter 29 for details).

Step (iii): To combine data from N events, we multiply the likelihoods:

$$\begin{aligned}\mathcal{L}_k(h_k|\Lambda) &\propto \int dz_1 dz_2\, p\left(z_1, z_2|h_k\right) p\left(z_1, z_2|\Lambda\right) \\ &\propto \frac{1}{n_k}\sum_{\alpha=1}^{n_k} p\left(z_{\alpha 1}, z_{\alpha 2}|\Lambda\right) \end{aligned} \tag{3.5}$$

$$\mathcal{L}(\{h_k\}|\Lambda) = \prod_{k=1}^{N} \mathcal{L}_k(h_k|\Lambda). \tag{3.6}$$

Here, $\mathcal{L}_k(h_k|\Lambda)$ is the likelihood function for the kth event with strain data h_k. The joint likelihood function $\mathcal{L}(\{h_k\}|\Lambda)$ combines data from all N measurements to arrive at the best possible constraints on Λ.

Step (iv): At last, we arrive at the posterior distribution for Λ, $p(\Lambda|\{h_k\})$. Combining the joint likelihood $\mathcal{L}(\{h_k\}|\Lambda)$ with a prior distribution for the

[1] Stictly speaking, it is necessary to divide by the prior used in the single event parameter estimation in the following equations, i.e., $p\left(z_1, z_2|\Lambda\right)$ becomes $p\left(z_1, z_2|\Lambda\right) / p\left(z_1, z_2|LAL\right)$. Since the original prior is uniform in our case, we neglect this term.

hyperparameters Λ, $\pi(\Lambda|H)$, for a particular population model, H, we obtain

$$\begin{aligned} p(\Lambda|\{h_k\}) &= \frac{\mathcal{L}(\{h_k\}|\Lambda)\pi(\Lambda|H)}{Z(\{h_k\}|H)} \\ &\propto \frac{\pi(\Lambda|H)}{Z(\{h_k\}|H)} \prod_{k=1}^{N} \frac{1}{n_k} \sum_{\alpha=1}^{n_k} p\left(z_{\alpha 1}, z_{\alpha 2}|\Lambda\right) \\ &\propto \prod_{k=1}^{N} \sum_{\alpha=1}^{n_k} p(z_{\alpha 1}, z_{\alpha 2}|\Lambda). \end{aligned} \tag{3.7}$$

Here, $Z(\{h_k\}|H)$ is the Bayesian evidence for the data from N observations $\{h_k\}$, for a model H, which is given by marginalizing over the hyperprior space

$$Z(\{h_k\}|H) = \int d\Lambda\, \mathcal{L}(\{h_k\}|\Lambda, H)\, \pi(\Lambda|H). \tag{3.8}$$

From our (hyper)posterior distribution $p(\Lambda|\{h_k\})$, we construct confidence intervals for our hyperparameters.

The odds ratio of two models is:

$$\mathcal{O}^i_j = \frac{Z(\{h_k\}|H_i)p(H_i)}{Z(\{h_k\}|H_j)p(H_j)}. \tag{3.9}$$

We use the odds ratio to select between different models. Here, the $p(H_i)$ are the prior probabilities assigned to each model. In our study, we assign equal probabilities to each model. Thus, the odds ratio is equivalent to the Bayes factor:

$$B^i_j = \frac{Z(\{h_k\}|H_i)}{Z(\{h_k\}|H_j)}. \tag{3.10}$$

We impose a somewhat arbitrary, but commonly used threshold of $|\ln(B)| > 8$ ($\sim 3.6\sigma$) to define the point at which one model is significantly preferred over another.

Now that we have derived a number of statistical tools, it is worthwhile to pause and consider what astrophysical questions we can answer with them.

i. If $p(\sigma_1, \sigma_2|\{h_k\})$ excludes $\sigma_1 = \sigma_2 = \infty$, then it necessarily follows that $p(\xi|\{h_k\})$ excludes $\xi = 0$, and we may infer that at least *some* binaries merge through fieldlike models.

ii. If $p(\xi|\{h_k\})$ excludes $\xi = 1$, we may infer that not all binaries can be formed via fieldlike models.

iii. If both $\xi = 0$ and $\xi = 1$ are excluded, then we may infer the existence of at least two distinct populations.

iv. If the (σ_1, σ_2) posterior distribution $p(\sigma_1, \sigma_2|\{h_k\})$ excludes $\sigma_1 = \sigma_2 = 0$, we may infer that not all binaries are perfectly aligned.

In this way we can distinguish between different formation channels or specific models, i.e., perfect alignment in case (iv).

We employ Bayes factors to compare our population models. We calculate evidences for three hypotheses:

i. Z_{dyn} – Dynamic formation only, $\xi = 0$.

ii. Z_{field} – Field formation only, $\xi = 1$.

iii. Z_{mix} – Mixture of field and dynamic, $\xi \in [0, 1]$.

We then define three Bayes' factors to compare these three hypotheses:

i. $B^{\text{mix}}_{\text{field}} = Z_{\text{mix}} / Z_{\text{field}}$.

ii. $B^{\text{mix}}_{\text{dyn}} = Z_{\text{mix}} / Z_{\text{dyn}}$.

iii. $B^{\text{field}}_{\text{dyn}} = Z_{\text{field}} / Z_{\text{dyn}}$.

In the next section, we apply these tools to a variety of simulated data sets in order to show under what circumstances we can measure various hyperparameters and carry out model selection.

3.5 Simulated population study

We use a simulated population to test our models. For the sake of simplicity, we construct a somewhat contrived population in which every binary shares some parameters corresponding to the best-fit parameters of GW150914:

- $(m_1, m_2) = (35M_\odot, 30M_\odot)$.
- $d_L = 410\,\text{Mpc}$.
- $(a_1, a_2) = (0.6, 0.6)$.

Here, d_L is luminosity distance and (a_1, a_2) are the black hole spin magnitudes. The remaining extrinsic parameters (sky position and source orientation) are sampled from isotropic distributions. We emphasize that the distance and mass and spin magnitude distributions are not representative of the full population of black hole binaries, which is poorly constrained. These distributions represent a subset of GW150914-like events, chosen for illustrative purposes. In reality, for every GW150914-like event, there are likely to be a large number of more distant (and possibly lower mass) events, which contribute relatively less information about spin.

We inject 160 binary merger signals into simulated Gaussian noise corresponding to Advanced LIGO at design sensitivity (Aasi, 2015; Abbott, 2016j). Of these, we generate 80 distributed according to p_0 and 80 distributed according to p_1; see Eq. (3.3). The injected values of (z_1, z_2) are shown in Fig. 3.2. The red diamonds correspond to the p_0 dynamical model and the blue circles to the p_1 fieldlike model. From these we construct "universes" summarized in Table 3.1. Each universe contains a different mixture of field and dynamical binaries. In every universe, $(\sigma_1, \sigma_2) = (0.3, 0.5)$.

For each universe, we present the results of the methods described above. In Fig. 3.3, we plot the 1σ (dark), 2σ (lighter), and 3σ (lightest) confidence

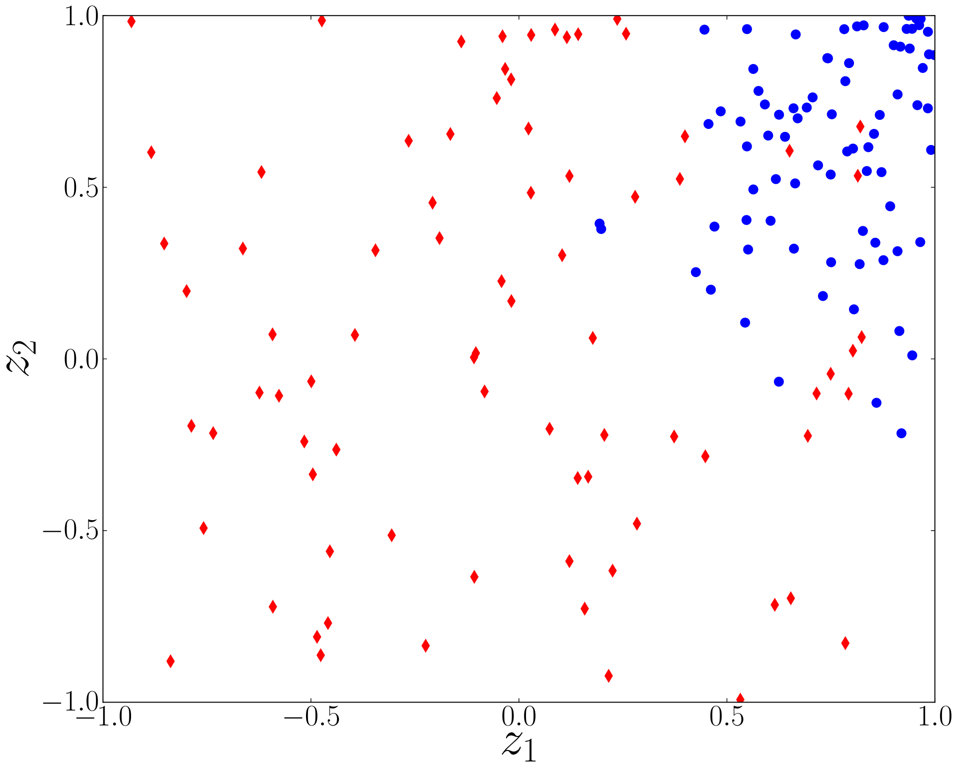

FIGURE 3.2: Simulated spin misalignment parameters (z_1, z_2) for the different populations of binary black holes used in our study. Red diamonds are drawn from the isotropic distribution p_0 while the blue circles are drawn from the aligned distribution, $p_1(z_1, z_2|\sigma_1, \sigma_2 = 0.3, 0.5)$; see Eq. (3.3).

Universe	ξ	σ_1	σ_2
A	0	N/A	N/A
B	0.1	0.3	0.5
C	0.5	0.3	0.5
D	0.9	0.3	0.5
E	1	0.3	0.5

TABLE 3.1: Hyperparameters describing different simulated universes. Here, ξ is the proportion drawn from our aligned model and (σ_1, σ_2) describe the typical misalignment angle; see Eq. (3.3).

regions as a function of the number of GW150914-like events. In Fig. 3.4, we plot the three Bayes factors defined in Eq. (3.10) as a function of the number of GW150914-like events. Each row in Fig. 3.3 and panel in Fig. 3.4 represents a different universe.

First we consider universe A, consisting of only dynamically formed binaries, $\xi = 0$; see the top row of Fig. 3.3. Since all binaries form dynamically in this universe, σ is undefined. We see that after $O(1)$ event we rule out $\xi = 1$ at 3σ (the hypothesis that all binaries form in the field).

Next we consider universe E in which all events are drawn from the aligned model, $\xi = 1$; see the bottom row of Fig. 3.3 and the bottom panel of Fig. 3.4. For this universe, $\sigma_1 = 0.3$, $\sigma_2 = 0.5$. We rule out $\xi = 0$ (dynamical only) at 3σ after $O(1)$ event. The Bayes factors also rule out all binaries forming dynamically after $\lesssim 10$ events. The threshold $|\ln(B)| = 8$ is shown by the dashed line. After 80 events, the 1σ confidence intervals for σ_1 and σ_2 have shrunk to $\sim 30\%$ and the 1σ confidence interval for ξ has shrunk to 3%. The Bayes factor comparing the two-population hypothesis to the purely field hypothesis $B^{\text{mix}}_{\text{field}}$ (the blue line in the bottom panel of Fig. 3.4) does not strongly favor field-only formation.

Universes B, C and D are mixtures of the field and dynamical populations. Of these, B and D have only 10% drawn from the subdominant population. We recover marginally weaker constraints than the corresponding single population universes. The hypothesis that all binaries form through the dominant mechanism is disfavored at 1σ after a few tens of events for universes B and D, establishing a weak preference for the presence of two distinct populations. For some realizations we can rule out both one component models after 80 events, however generally we see a subthreshold preference for the mixture model. This is unsurprising since each one-population model is a subset of our two-population model. For universe C, an equal mixture of events drawn from the field and dynamical populations. Both $\xi = 0$ and $\xi = 1$ are excluded at 3σ after tens of events establishing the presence of two distinct subpopulations.

For all five universes, the presence of a perfectly aligned component ($\sigma = 0$) is excluded after fewer than 20 events. For many realizations this number is < 5. For universes B, C and D (consisting of a mixture of field and dynamical mergers), we can rule out the entire population forming from one of the two channels after 10–40 GW150914-like events. When there is a large contribution from the aligned model, we observe that the allowed region for σ_1 becomes small faster than the allowed region for σ_2. There are two effects, which explain this. First, the secondary black hole's spin has a less significant effect on the waveform (Vitale et al., 2014; Vitale & Evans, 2017; Vitale et al., 2017a). The spin orientation of the secondary is therefore less well constrained for each event. This translates to a larger uncertainty for σ_2 compared to σ_1. Second, the width of the distribution of spin tilts is broader for the secondary black holes. This broader distribution is intrinsically more difficult to resolve.

3.6 Spin maps

In addition to our hierarchical analysis, we present a visualization tool for the distribution of spin orientations. We introduce "spin maps": histograms of posterior spin orientation probability density, averaged over many events, and plotted using a Mollweide projection of the sphere defining the spin orientation, see Fig. 3.5. The maps use HEALPix (Górski et al., 2005). For each posterior sample the latitude is the spin tilt of the primary black hole, θ_1, and the longitude the difference in azimuthal angles of the two black holes, $\Delta\Phi$. The difference in azimuthal angles may give information about the history of the binary, specifically by identifying spin-orbit resonances at $\Delta\Phi = 0, \pi$ (Schnittman, 2004; Gerosa et al., 2013; Gerosa et al., 2014; Kesden et al., 2015; Gerosa et al., 2015; Trifirò et al., 2016). These resonances, if detected, would appear as bands of constant longitude. We do not utilize azimuthal angle in this work and our injected distributions are isotropic in $\Delta\Phi$. In the future, it would also be interesting to produce ensemble spin disk plots (e.g., Fig. 5 of Abbott, 2016i), showing the spin magnitude and orientation for a population of binaries.

The spin maps in Fig. 3.5 include contributions from 80 events for universes A and C (see Table 3.1). This simple representation is useful because it provides qualitative insight into the distribution of spins and helps us to see trends and patterns that might not be obvious from our likelihood formalism. The north pole on these maps corresponds to spin aligned with the total angular momentum of the binary. We see the preference for the spin to be aligned with the angular momentum vector of the binary by the clustering in the northern hemisphere.

3.7 Discussion

The physics underlying the formation of black hole binaries is poorly constrained both theoretically and observationally. We do not know which of the proposed mechanisms is the main source of binary mergers: preferentially aligned mergers formed in the field versus randomly aligned mergers formed dynamically. We are also not confident in the predicted characteristics of binaries formed through either channel. We therefore create a simple (hyper)parametrization, describing the ensemble properties of black

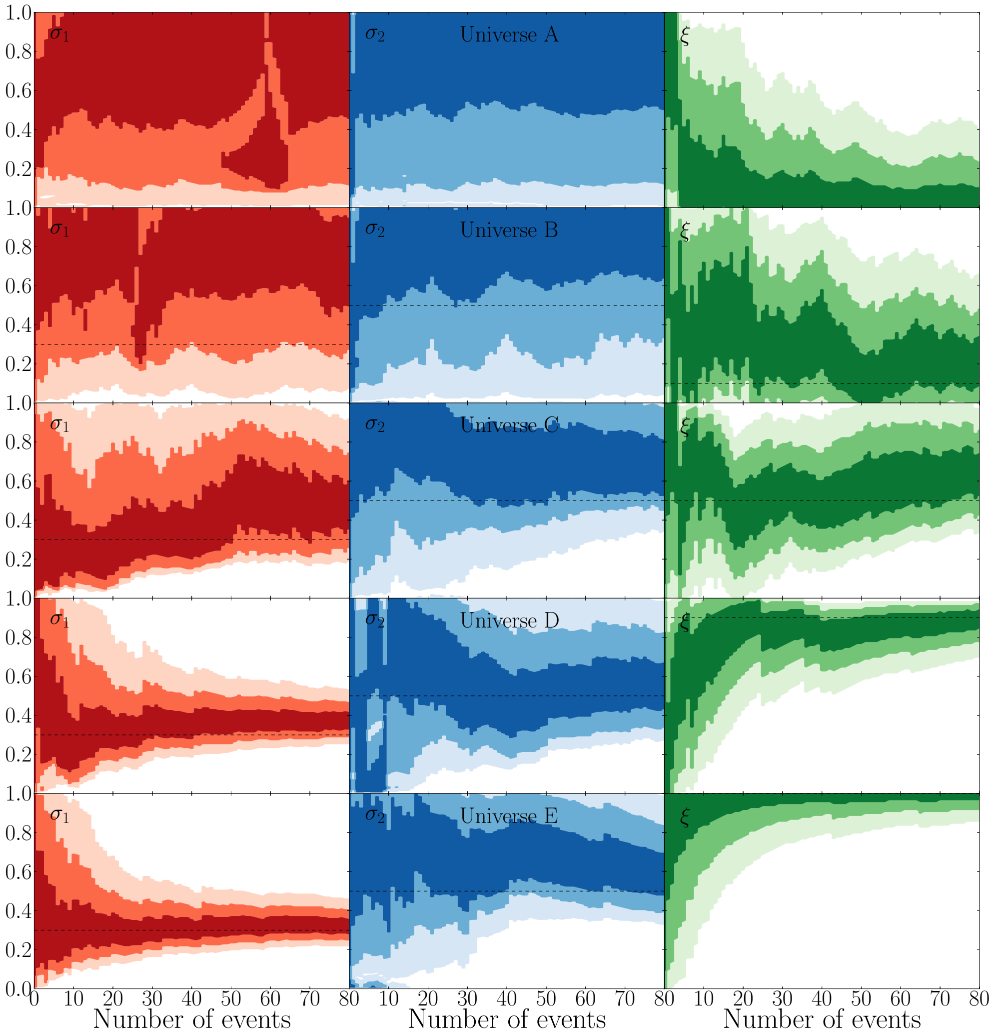

FIGURE 3.3: In each panel we plot 1σ (dark shading), 2σ (medium shading), and 3σ (light shading) confidence for different hyperparameters as a function of the number of events N. Each column represents a different hyperparameter: σ_1 (left), σ_2 (middle), and ξ (right). Each row represents a different universe; see Table 3.1. From top to bottom, the universes are A, B, C, D, and E. The dashed line indicates the true hyperparameter values. The highest likelihood values of the three parameters after 80 events are shown on each panel along with the width of the 1σ confidence interval.

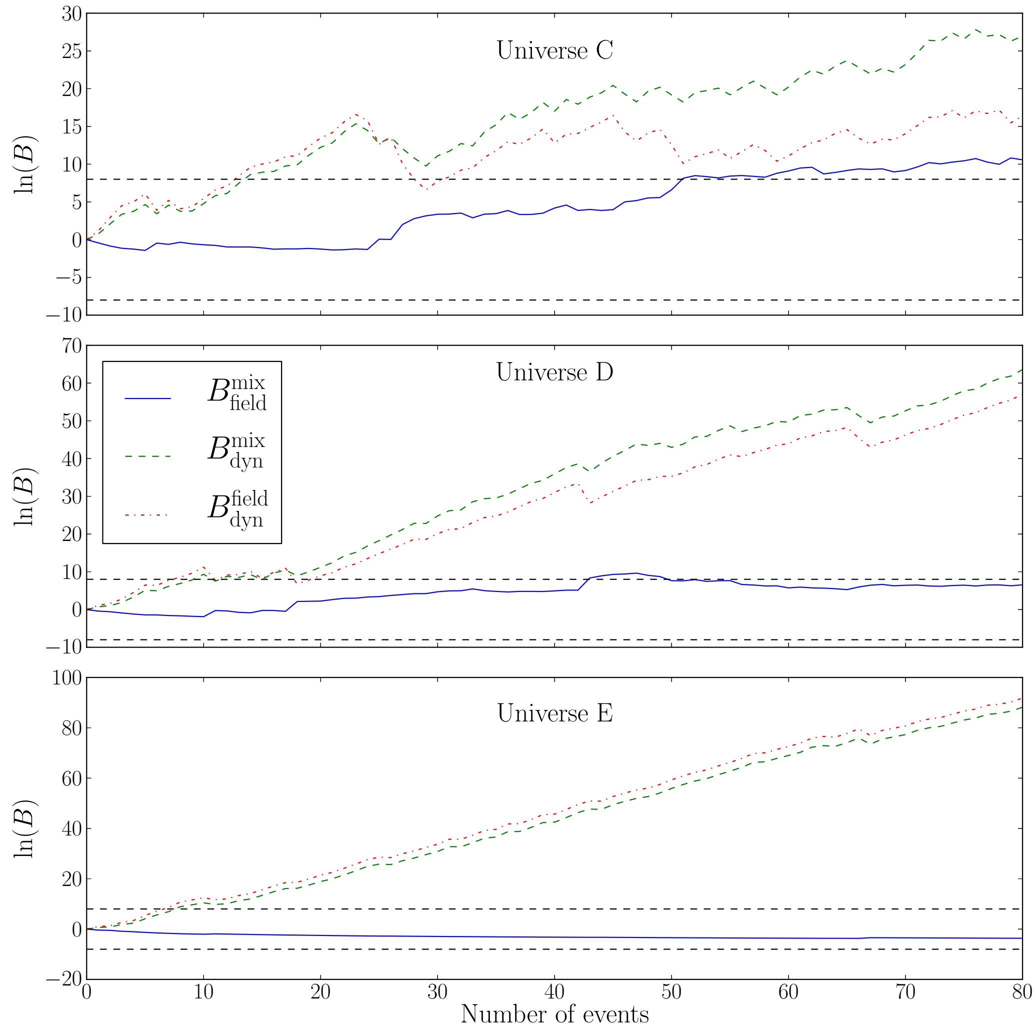

FIGURE 3.4: Log Bayes factors as a function of the number of GW150914-like events. The dot-dashed red line shows $B^{\rm field}_{\rm dyn}$ comparing the pure-field hypothesis to the pure-dynamical hypothesis. The dashed green line shows $B^{\rm mix}_{\rm dyn}$ comparing the two-population hypothesis to the dynamical hypothesis. The solid blue line shows $B^{\rm mix}_{\rm field}$ comparing the two-population hypothesis to the pure-field hypothesis. The dashed lines denotes $|\ln(B)| = 8$, our threshold for distinguishing between models. Each panel is a different universe. The top panel is universe C (equal mixture of field and dynamic). With $\lesssim 40$ events, there is a strong preference for the two-component hypothesis over the pure-dynamic hypothesis. After ~ 50 events there is a preference for the two-component hypothesis over the pure-field hypothesis. The center panel is universe D (majority field with some dynamic). With $\lesssim 10$ events, there is a strong preference for the two-component and pure-field hypotheses over the pure-dynamic hypothesis. There is a preference for the correct two-population hypothesis over the pure-field hypothesis. The bottom panel is universe E (pure field). With $\lesssim 10$ events, there is a strong preference for the two-component and field hypotheses over the dynamic hypothesis. There is a marginal preference for the correct field hypothesis over the two-population hypothesis.

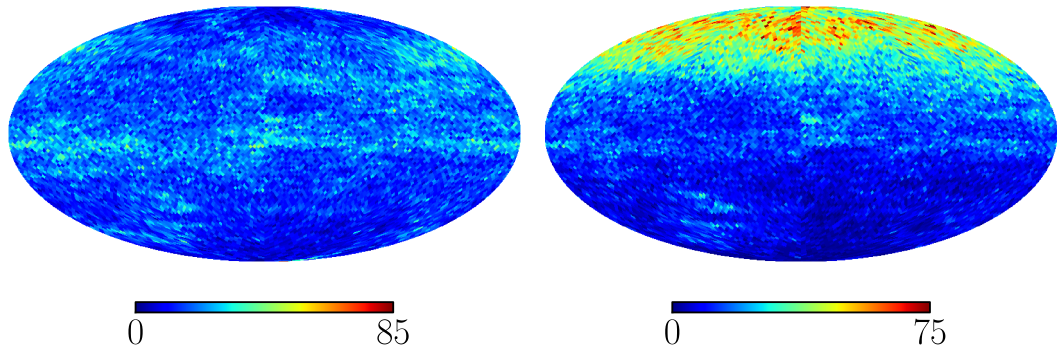

FIGURE 3.5: Spin maps: maps of posterior spin orientation probability density averaged over many realizations. The latitude is the spin tilt of the primary (more massive) black hole. The longitude is the angle between the projection of the black hole spins onto the orbital plane. The color bar is the number of posterior samples per 5 deg^2 HEALPix bin. The left panel shows a spin map for 80 events drawn from universe A (see, Table 3.1) in which every binary merges dynamically. The right panel shows 80 events drawn from universe C drawn in which, on average, 50% of the events are drawn from the dynamical population while 50% are drawn from the field population with $(\sigma_1, \sigma_2) = (0.3, 0.5)$; see, Eq. 3.3. The presence of a preferentially oriented population is seen as clustering around the north pole.

hole binaries. We demonstrate that we can measure hyperparameters describing the spin properties of an ensemble of black hole mergers with multiple populations. Previous work by Vitale et al., 2017b and Stevenson, Berry & Mandel, 2017 demonstrated that the fraction of binaries drawn from different populations can be inferred after O(10) events. We show that after a similar number of events, the shape of the spin-orientation distribution can be inferred using a simple hyperparametrization. We reproduce the finding from Stevenson, Berry & Mandel, 2017, that $O(1)$ event is required to distinguish an isotropically oriented distribution, $\xi = 0$, from a perfectly aligned distribution, $\xi = 1$, $\sigma_1 = \sigma_2 = 0$. After fewer than 40 GW150914-like events we can determine the properties of the dominant formation mechanism for all of our considered scenarios. We also introduce the concept of spin maps, which provide a tool for visualizing the distribution of spin orientations from an ensemble of detections.

One limitation of our study is that, for the sake of simplicity, we employ a population of binaries with masses, distance, and spins fixed to values consistent with GW150914. The advantage of this simple model is that we are able to isolate the effect of spin orientation by holding other parameters fixed. The disadvantage is that the GW150914-like population is not a realistic description of nature. By changing from a population of binaries at a fixed distance to a population distributed uniformly in comoving volume, more events will be required for measurement of population hyperparameters. This is because most events, coming from the edge of the visible volume, will contribute only marginally to our knowledge of these hyperparameters. We assume fixed spin magnitudes of $a_1 = a_2 = 0.6$. For a binary with aligned spins, this would imply $\chi_{\rm eff} = 0.6$. Based on recent LIGO detections, this might be optimistic. For GW151226, $\chi_{\text{eff}} = 0.21^{+0.20}_{-0.10}$. For all other observed events, χ_{eff} is consistent with 0. This implies either that the observed black holes are not spinning rapidly or that the merging black holes observed so far possess significantly misaligned spins (Abbott, 2016b; Abbott, 2017f; Farr et al., 2017). If we have overestimated the typical black hole spin magnitude a, the number of events required to determine the distribution of spin orientation will increase. Implementing a theoretically motivated distribution of these parameters is left to future studies. Another area of future work is extending the method to other physically motivated spin orientation distributions.

Addendum

Since the publication of this paper additional compact binary coalescences have been observed and significant progress has been made in theoretical and observational analysis. The reader is directed to Chapters 1 and 8 for an overview of the field at the time of writing.

Chapter 4

Measuring the binary black hole mass spectrum with an astrophysically motivated parameterization

Published as:
Talbot, C. & Thrane, E., Astrophys. J. 856, 173 (2018).

Abstract

Gravitational-wave detections have revealed a previously unknown population of stellar mass black holes with masses above $20M_{\odot}$. These observations provide a new way to test models of stellar evolution for massive stars. By considering the astrophysical processes likely to determine the shape of the binary black hole mass spectrum, we construct a parameterized model to capture key spectral features that relate gravitational-wave data to theoretical stellar astrophysics. In particular, we model the signature of pulsational pair-instability supernovae, which are expected to cause all stars with initial mass $100M_{\odot} \lesssim M \lesssim 150M_{\odot}$ to form $\sim 40M_{\odot}$ black holes. This would cause a cut-off in the black hole mass spectrum along with an excess of black holes near $40M_{\odot}$. We carry out a simulated data study to illustrate some of the stellar physics that can be inferred using gravitational-wave measurements of binary black holes and demonstrate several such inferences that might be made in the near future. First, we measure the minimum and maximum stellar black hole mass. Second, we infer the presence of a peak due to pair-instability supernovae. Third, we measure the black hole mass ratio distribution. Finally, we show how inadequate models of the black hole mass spectrum lead to biased estimates of the merger rate and the amplitude of the stochastic gravitational-wave background.

4.1 Introduction

The black holes observed by advanced gravitational-wave detectors such as the Laser Interferometer Gravitational-wave Observatory (LIGO) (Aasi, 2015) and Virgo (Acernese, 2015) are widely believed to be formed from massive stars with initial mass, $M \gtrsim 20M_\odot$ (Heger et al., 2003). Gravitational-wave measurements constrain the mass and spin of merging binaries. These measurements, in turn, can be used to better understand the evolution of massive stars. In this paper, we study how gravitational-wave measurements of the black hole mass spectrum can be used to inform our understanding of stellar evolution.

Simulating the final stages of stellar binary evolution is computationally expensive. Additionally, there are significant theoretical uncertainties in key aspects of binary evolution, especially the common envelope phase (Ivanova et al., 2013) and supernova mechanism. For these reasons, populations of compact objects are simulated using population synthesis models (e.g., Dominik et al., 2015; Belczynski et al., 2017; Stevenson et al., 2017). These are phenomenological models calibrated against a small number of more detailed stellar simulations.

At the time of writing, there have been five confirmed detections of binary black hole mergers and one unconfirmed candidate event (e.g., Abbott, 2016h; Abbott, 2016f; Abbott, 2016b; Abbott, 2017f; Abbott, 2017h; Abbott, 2017g). The 90% credible regions for the source masses of the black holes range from $\sim 5M_\odot$ to $\sim 40M_\odot$. Of the observed events, only one (GW151226) provides unambiguous evidence of black hole spin (Abbott, 2016f), although two events (GW150914 and GW170104) show a weak preference for spins anti-aligned with respect to the orbital angular momentum vector (Abbott, 2016c; Abbott, 2017f). The implications of these measurements are currently unclear. Current theories include: most binary black holes are formed dynamically (Rodriguez et al., 2016a; Rodriguez et al., 2018), black holes are subject to large black hole natal kicks (Rodriguez et al., 2016b; O'Shaughnessy, Gerosa & Wysocki, 2017), and/or that large black holes do not form with significant dimensionless spins (Belczynski et al., 2017; Wysocki et al., 2018).

There has been significant work using gravitational-wave data to infer the properties of black hole formation with ensembles of detections. These works range from comparing gravitational-wave data to specific, non-parameterized models (Mandel & O'Shaughnessy, 2010; Stevenson, Ohme & Fairhurst, 2015; Dominik et al., 2015; Belczynski et al., 2016; Stevenson, Berry & Mandel, 2017; Zevin et al., 2017; Belczynski et al., 2017; Miyamoto et al., 2017; Farr et al., 2017; Wysocki et al., 2018; Barrett et al., 2018), to attempts to group the data by binning, clustering or Gaussian mixture modeling (Mandel et al., 2017; Farr, Holz & Farr, 2018; Wysocki, 2017), to fitting physically motivated phenomenological population (hyper)parameters (Kovetz et al., 2017; Talbot & Thrane, 2017; Fishbach & Holz, 2017). In this work, we take the last approach and demonstrate that it is possible to identify physical features in the black hole mass spectrum with an ensemble of detections using phenomenological models, building on work in Kovetz et al., 2017 and Fishbach & Holz, 2017.

Previous attempts to determine the binary black hole mass spectrum have employed one or more of these three approaches. Clustering is applied to a binned mass distribution in Mandel et al., 2017 to demonstrate that a mass gap between neutron star and black hole masses can be identified after $O(100)$ observations. In Zevin et al., 2017; Stevenson, Ohme & Fairhurst, 2015; Barrett et al., 2018, the authors compare population synthesis models with different physical assumptions and show that predicted mass distribution can be distinguished using $O(10)$ of observations.

Previous analyses by the LIGO and Virgo scientific collaborations fit a power-law model with variable spectral index, α. Fishbach & Holz, 2017 point out that, given LIGO/Virgo's additional sensitivity to heavier binary systems, there is a possible dearth of black holes larger than $\sim 40 M_{\odot}$. They suggest that this is due to the occurrence of pulsational pair-instability supernovae (Heger & Woosley, 2002) and propose an extension of the current LIGO analysis where the maximum mass is a free parameter. Pulsational pair-instability supernovae occur in stars with initial masses $100 M_{\odot} \lesssim M \lesssim 150 M_{\odot}$, causing all stars in that mass range to form black holes with mass $\sim 40 M_{\odot}$. In addition to a cut-off in the black hole mass spectrum, we expect that there will be an excess of black holes around the cut-off mass.

Kovetz et al., 2017 model the lower mass limit of black holes and use a Fisher analysis to demonstrate that it should be possible to identify the presence of the neutron-star black hole mass gap. They also model the distribution of mass ratios and propose a test for detecting primordial black holes. A method to simultaneously estimate the binary black hole mass spectrum and the merger rate is presented in Wysocki, 2017. Wysocki, 2017 also considers a Gaussian mixture model for fitting the distribution of compact binary parameters.

The rest of the paper is structured as follows. In section 4.2, we introduce the statistical tools necessary to make statements about the black hole population. We then develop our model in section 4.3 in terms of population (hyper)parameters by considering current observational constraints and predictions from theoretical astrophysics and population synthesis. In section 4.4, we perform a Monte Carlo injection study. We consider how many detections will be necessary to identify different features using Bayesian parameter estimation and model selection. We show how the predicted mass distributions differ when using different (hyper)parameterizations. We also explore some of the consequences of using inadequate (hyper)parameterizations. In particular, we show that inadequate (hyper)parameterization can lead to significant bias in the estimate of the merger rate and the predicted amplitude of the stochastic gravitational-wave background (SGWB). Some closing thoughts are provided in section 4.5.

4.2 Bayesian Inference

4.2.1 Gravitational-wave Detection

A binary black hole system is completely described by 15 parameters, Θ. Recovering these parameters from the observed strain data requires the use of specialized Bayesian parameter inference software, e.g., LALINFERENCE (Veitch et al., 2015). The likelihood of a given set of binary parameters is computed by comparing the strain data to the signal predicted by general relativity. For the analysis presented here, the expected signal is calculated using phenomenological approximations to numerical relativity waveforms (e.g., Hannam et al., 2014; Schmidt, Ohme & Hannam, 2015; Smith et al., 2016). For a given set of strain data, h_i, LALINFERENCE returns a set of n_i samples, $\{\Theta\}$, which are sampled from the posterior distribution, $p\left(\Theta|h_i, H\right)$, of the binary parameters, along with the Bayesian evidence, $\mathcal{Z}(h_i|H)$, where H is the model being tested.

The distribution of binary black hole systems observed by current detectors is not representative of the astrophysical distribution of binary black holes. The observing volume of current gravitational-wave detectors is limited by the instruments' sensitivity. The sensitive volume for a detector to a given binary is primarily determined by the masses of the black holes with spin entering as a higher order effect. More massive systems produce gravitational waves of greater amplitude. However, these more massive systems merge at a lower frequency and, hence, spend less time in the observing band of the detector. Additionally, distant sources undergo cosmological redshift and appear more massive than they actually are. Here, we will deal only with the un-redshifted "source-frame" masses, not the "lab-frame" masses observed by gravitational-wave detectors We note that the source/lab-frame distinction is about cosmological redshift and is not a statement about detectability and/or selection effects.

Accounting for these factors, we calculate $V_{obs}(\Theta)$, the sensitive volume for a binary with parameters Θ, following Abbott, 2016b, using semi-analytic noise models corresponding to different sensitivities (Abbott, 2016j). The noise, and hence sensitivity, in real detectors is time-dependent and so calculating this volume requires averaging over the observing time to obtain a mean sensitive volume $\langle V_{obs}(\Theta)\rangle$ (Abbott, 2016b).

4.2.2 Population Inference

We are interested in inferring population (hyper)parameters describing the distribution of source-frame black hole masses. The formalism to do this is briefly described below (see e.g., Gelman et al., 2013, chapter 29 for a more detailed discussion of hierarchical Bayesian modeling and Mandel, Farr & Gair, 2014 for a discussion of selection biases).

Hierarchical inference of this kind can be cast as a post hoc method of changing from the prior distribution used in the single event parameter estimation to a new prior, which depends on population (hyper)parameters, Λ.

We marginalise over all of the binary parameters while reweighting the posterior samples by the ratio between our (hyper)parameterized model and the prior used to generate the posterior distribution. This marginalisation integral is approximated by summing over the posterior samples for each event. The N events are then combined by multiplying the new marginalised likelihood for the individual events,

$$\mathcal{L}\left(\{h\}_{i=1}^{N}\middle|\Lambda, H\right) \propto \prod_{i}^{N}\sum_{j}^{n_i}\frac{\pi(\Theta_j^i|\Lambda, H)}{\pi(\Theta_j^i|\mathrm{LAL})}. \tag{4.1}$$

Here $\pi(\Theta|\mathrm{LAL})$ is the prior probability distribution used for single event parameter estimation. The distribution $\pi(\Theta_j^i|\Lambda, H)$ is the probability of a binary having parameters Θ in our model; see Sec. 4.3. We do not model any black hole parameters other than source-frame mass and so Θ can be replaced by $m = (m_1, m_2)$, in Eq. 4.1 and all following equations. The prior distribution of source-frame masses in LALINFERENCE is a convolution of a uniform in component mass prior between limits determined by the reduced order model (Smith et al., 2016) and the redshift distribution corresponding to uniform in luminosity distance extending out to $4Gpc$. This distance distribution is converted to redshift assuming ΛCDM cosmology using the results from the *Planck* 2015 data release (Planck Collaboration & Ade, 2016).

We combine this likelihood with $\pi(\Lambda|H)$, the prior for the (hyper)parameters assuming a model H, and the Bayesian evidence for the data given H to obtain the posterior distribution for our (hyper)parameters,

$$p\left(\Lambda\middle|\{h\}_{i=1}^{N}, H\right) = \frac{\mathcal{L}\left(\{h\}_{i=1}^{N}|\Lambda, H\right)\pi(\Lambda|H)}{\mathcal{Z}\left(\{h\}_{i=1}^{N}|H\right)}, \tag{4.2}$$

$$\mathcal{Z}\left(\{h\}_{i=1}^{N}\middle|H\right) = \int d\Lambda\,\mathcal{L}\left(\{h\}_{i=1}^{N}\middle|\Lambda, H\right)\pi(\Lambda|H). \tag{4.3}$$

To perform the (hyper)parameter estimation we use the python implementation of `MultiNest` (Feroz, Hobson & Bridges, 2009; Buchner et al., 2014). Additionally, we calculate the posterior predictive distribution (PPD) of the binary parameters,

$$\begin{aligned} p(m|\{h\}_{i=1}^{N}, H) &= \int d\Lambda \pi(m|\Lambda, H)p(\Lambda|\{h\}_{i=1}^{N}, H) \\ &\approx \frac{1}{n_k}\sum_{k}^{n_k}\pi(m|\Lambda_k, H), \end{aligned} \tag{4.4}$$

where Λ_k are the n_k (hyper)posterior samples. The PPD shows the probability that a subsequent detection will have parameters Θ given the previous data, $\{h\}$.

4.2.3 Model Selection

Model selection is performed in our Bayesian framework by considering Bayes factors,

$$BF^{\alpha}_{\beta} = \frac{\mathcal{Z}\left(\{h\}^{N}_{i=1}|H_{\alpha}\right)}{\mathcal{Z}\left(\{h\}^{N}_{i=1}|H_{\beta}\right)}. \tag{4.5}$$

A large Bayes factor, $BF^{\alpha}_{\beta} \gg 1$, indicates that H_{α} is strongly favored over H_{β}. We adopt a conventional threshold of $\ln BF = 8$ to distinguish between two models.

4.3 Phenomenology

In this section, we develop a parameterization of the black hole mass spectrum using predictions from astrophysics theory, population synthesis models, and electromagnetic observations. In this way, we can relate gravitational-wave measurements to stellar astrophysics. For low-mass systems, the parameter that most strongly affects the observable gravitational waveform is a combination of the component masses known as the chirp mass, $\mathcal{M} = (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$. For high-mass systems, the waveform is primarily determined by the total mass of the system. The mass ratio is more difficult to determine due to covariances between the mass ratio and the spin of the black holes.

The canonical assumed distribution of black hole masses is a power law distribution in the primary mass between some maximum and minimum masses. This power law distribution has three typical parameters: the spectral index α, the minimum mass $m_{\min}$, and the maximum mass $m_{\max}$. The distribution of secondary mass is typically taken to be flat between $m_{\min}$ and m_1. We take this as the starting point for our parameterization.

4.3.1 High-Mass Binaries

The observation of binary black hole mergers through the detection of gravitational waves revealed the presence of a previously unobserved population of black holes with mass $\sim 30 M_{\odot}$ (Abbott, 2016h; Abbott, 2017f; Abbott, 2017h). Since gravitational-wave detectors can observe more massive binaries at greater distances, binaries containing larger black holes are preferentially detected over less massive systems. Fishbach & Holz, 2017 note that, given the observation rate of binaries with mass $\sim 30 M_{\odot}$, it is somewhat surprising that we have not seen more massive black holes. They propose that this is due to a cut-off in the black hole mass spectrum around this mass. By comparing the Bayesian evidence for $m_{\max} = 41 M_{\odot}$ and $m_{\max} = 100 M_{\odot}$ using the first four events, they find tentative support for a cut-off. They further show that it will be possible to identify the presence of an "upper mass gap" with a Bayes Factor of $\gtrsim 150$ ($\ln BF \approx 5$) using 10 detections and the cut-off mass can be measured with ~ 40 detections.

The theoretical motivation for such a cut-off is pulsational pair-instability supernovae (PPSN) (Heger & Woosley, 2002; Woosley & Heger, 2015), whereby large amounts of matter are ejected prior to collapse to form a black hole. The expected result of this process is that all stars with initial mass $100M_\odot \lesssim M \lesssim 150M_\odot$ form black holes with masses $\sim 40M_\odot$. Stars with $150M_\odot \lesssim M \lesssim 250M_\odot$ are expected to undergo pair-instability supernovae (PISN) and leave no remnant. Hence, we expect a gap in the black hole mass spectrum between $\sim 40M_\odot$ and $\sim 250M_\odot$ along with an excess of black holes at some mass $m_{\rm pp} \sim 40M_\odot$. The excess is due to the $100M_\odot \lesssim M \lesssim 150M_\odot$ stars which undergo PPSN. The size, position and shape are determined by the unknown details of PPSN. While the observation of cut-off near $40M_\odot$ could be interpreted as evidence for PPSN, the additional observation of a peak would provide a smoking-gun signature that the highest mass stellar binaries are reduced in mass via PPSN.

Thus, we extend our description of the upper end of the black hole mass spectrum to allow for the possibility of an excess due to PPSN. We model this as a normal distribution with unknown mean $m_{\rm pp} \in [25, 100]M_\odot$ and variance $\sigma_{\rm pp} < 5M_\odot$. It is also necessary to introduce a mixing fraction parameter, λ, which describes the proportion of binaries which are drawn from the normal distribution.

We expect that any PPSN mass peak should be near the high-mass cut-off, this corresponds to $m_{\rm max} \approx m_{\rm pp}$. Also, assuming that the power law is otherwise a good description of the black hole mass spectrum, we expect that the number of black holes in the PPSN peak should be no more than the number of black holes that would have formed had the power law distribution continued to the upper limit of the upper mass gap, determined by the onset of pair-instability supernovae, $m_{\rm PI} \sim 150M_\odot$. We impose this condition by requiring that the extrapolated area which would be under the power-law curve is less than the area contained within the Gaussian. This amounts to a restriction on the allowed values of λ,

$$\lambda \leq \frac{\int_{m_{\rm max}}^{m_{\rm PI}} m^{-\alpha}}{\int_{m_{\rm min}}^{m_{\rm PI}} m^{-\alpha}} = \frac{m_{\rm PI}^{1-\alpha} - m_{\rm max}^{1-\alpha}}{m_{\rm PI}^{1-\alpha} - m_{\rm min}^{1-\alpha}} \approx \left(\frac{m_{\rm min}}{m_{\rm max}}\right)^{\alpha-1}. \tag{4.6}$$

Here, $m_{\rm min}$ is the upper limit of the NS-BH mass gap and $m_{\rm max}$ is the lower limit of the upper mass gap. The variable $m_{\rm PI}$ is the mass above which stars undergo PISN leaving no remnant. Here, we assume $\alpha > 1$ and $m_{\rm PI} \gg m_{\rm max}$.

To hone our intuition, we can plug in plausible values of α, $m_{\rm min}$, $m_{\rm max}$ and $m_{\rm PI}$ to determine a typical value of λ. For example, if we set $m_{\rm min} = 5M_\odot, m_{\rm max} = 50M_\odot$ and $\alpha = 2$, we find $\lambda \sim 0.1$. If we measure a peak consistent with these values, we anticipate that the position and width of the peak can inform our physical understanding of this mechanism. If λ is measured to be inconsistent with this constraint it could indicate that either the extrapolation of the power law is not a valid assumption, or that the peak is not entirely due to PPSN. We note that the fraction of *observed* black holes which formed through PPSN will be larger than λ since the more massive black holes are observable out to a greater distance.

4.3.2 Low-Mass Binaries

The smaller sensitive volume for lower-mass binaries means that it is more difficult to probe the low-mass end of the black hole mass spectrum with gravitational-wave detections. Previous analyses of the black hole mass spectrum from gravitational-wave detections have assumed that the black hole mass spectrum has a sharp cut-off at some minimum mass $m_{\min}$. However, this overestimates the number of low-mass black holes if the distribution of low-mass black holes in merging binaries is the same as that in low-mass X-ray binaries (Özel et al., 2010). Population synthesis models also generically predict that the primary mass distribution peaks above the minimum mass.

We replace the step function at the low-mass end of the black hole mass spectrum with a smoothing function, $S(m, m_{\min}, \delta m)$, which rises from zero at $m_{\min}$ to one at $m_{\min} + \delta m$,

$$S(m, m_{\min}, \delta m) = \left(\exp f(m - m_{\min}, \delta m) + 1\right)^{-1} \tag{4.7}$$

$$f(m, \delta m) = \frac{\delta m}{m} - \frac{\delta m}{m - \delta m}. \tag{4.8}$$

We note that $\delta m = 0$ recovers the step function used in previous analyses. Since the mass distribution is expected to be an increasing function, the peak of $p(m_1)$ occurs below $m_{\min} + \delta m$. We expect $m_{\min} \sim 5M_{\odot}$ and $\delta m \gtrsim 3$ given that the black hole mass spectrum inferred from electromagnetic observations peaks at $8M_{\odot}$, with no black holes less massive than $5M_{\odot}$.

Our model of the distribution of the primary mass can be summarized as

$$p(m_1|\Lambda) = (1 - \lambda) p_{pow}(m_1|\Lambda) + \lambda \, p_{pp}(m_1|\Lambda) \tag{4.9}$$

where

$$p_{pow}(m_1|\Lambda) \propto m_1^{-\alpha} S(m_1, m_{\min}, \delta m) \mathcal{H}(m_{\max} - m_1)$$

encodes the power-law distribution with a smooth turn on at low mass and

$$p_{pp}(m_1|\Lambda) \propto \exp\left(-\frac{(m_1 - m_{\text{pp}})^2}{2\sigma_{\text{pp}}^2}\right) S(m_1, m_{\min}, \delta m).$$

encodes the peak from PPSN.

4.3.3 Mass Ratio

Previous analyses by the LIGO/Virgo scientific collaborations have assumed that the secondary mass is distributed uniformly between a lower limit set by $m_{\min}$ and an upper limit of m_1. This is motivated by observations of the stellar initial binary population, (e.g., Kroupa et al., 2013 and Belloni et al., 2017). In contrast to this, population synthesis models typically predict that the distribution of mass ratios should be biased towards equal mass binaries, (e.g., Belczynski et al., 2017). We model the distribution of the mass ratio as a power-law with spectral index β as in Kovetz et al., 2017; Fishbach & Holz, 2017. For a mass ratio distribution peaked at equal masses, $\beta > 0$. We also

α	Spectral index of m_1 for the power-law distributed component as the mass spectrum.
$m_{\max}$	Maximum mass of the power-law distributed component as the mass spectrum.
λ	Proportion of primary black holes formed via PPSN.
m_{pp}	Mean mass of black holes formed via PPSN.
σ_{pp}	Standard deviation of masses of black holes formed via PPSN.
$m_{\min}$	Minimum black hole mass.
δm	Mass range over which black hole mass spectrum turns on.
β	Spectral index of m_2.

TABLE 4.1: (Hyper)parameters describing the black hole mass spectrum.

impose the same smoothing at the lower limit as we apply to the primary mass. This allows us to write down the conditional probability distribution for secondary masses given a primary mass,

$$p(m_2|m_1,\Lambda) = \left(\frac{m_2}{m_1}\right)^{\beta} S(m_2, m_{\min}, \delta m)\mathcal{H}(m_1 - m_2). \tag{4.10}$$

4.3.4 Summary

A table listing the (hyper)parameters and their physical meaning is provided in Tab. 4.1. Including a factor of V_{obs} to account for selection biases, the probability of detecting a mass pair given our (hyper)parameters, $\Lambda = \{\alpha, m_{\min}, m_{\max}, \delta m, \lambda, m_{\text{pp}}, \sigma_{\text{pp}}, \beta\}$ and under model H, is

$$\pi(m|\Lambda, H) \propto p(m_1|\Lambda, H)p(m_2|m_1, \Lambda, H)V_{obs}(m). \tag{4.11}$$

We consider six different models for our (hyper)prior distribution, corresponding to decreasingly stringent physical assumptions as independent hypotheses, these different prior assumptions can be tested with our Bayesian framework using Bayes factors. In our first model, H_0, we take the power law distribution with maximum mass, $m_{\max} = 100M_\odot$, since the injected data set considered in Sec. 4.4 has a minimum mass of $3M_\odot$, we allow $m_{\min}$ to vary, rather than fixing it at $m_{\min} = 5M_\odot$ as in previous analyses. In H_1 we introduce a uniform prior on $m_{\max}$. In order to determine the relative importance of the different features, we switch to models which include all the effects described above except one. In H_2, H_3 and H_4 we do not include the Gaussian component, mass ratio and low-mass smoothing respectively. Finally, in H_5 we include all of these effects. These prior choices are summarized in Tab. 4.2.

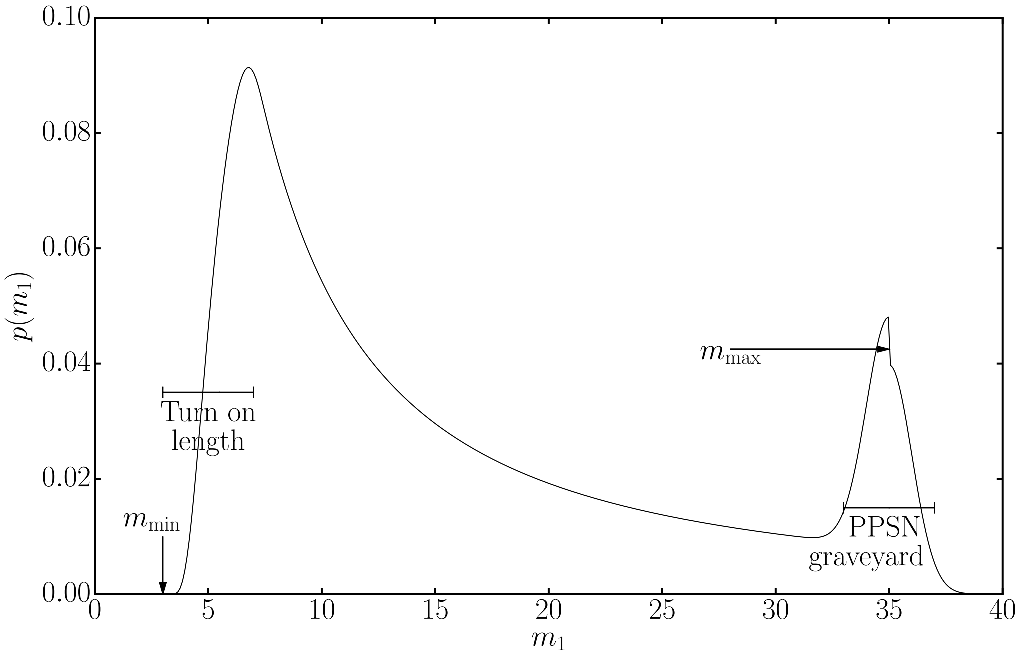

FIGURE 4.1: The astrophysical distribution of source-frame masses assuming our modeled distribution of m_1 with (hyper)parameters as specified in Tab. 4.2 We identify the excess of black holes at $35M_\odot$ due to PPSN. We also see the smooth turn-on at low masses.

4.3.5 Other Effects

As with any phenomenological model, our model has limitations. If the proposed mass gaps exist in the population of black holes formed as the endpoint of stellar evolution there may still be black holes found in these gaps. The remnant of the binary neutron star merger GW170817 has mass $M_{\text{rem}} \lesssim 2.8M_\odot$ (Abbott, 2017i). It is not clear whether this object is a neutron star or a black hole. Similarly, the remnant from binary black hole mergers such as GW150914 is more massive than the suggested upper mass limit due to PPSN, $M_{\text{rem}} = 62^{+4}_{-4}M_\odot$ (Abbott, 2016i). Both of these objects lie within the proposed mas gaps. If either of these mergers happened in a dense environment such as a globular cluster, it is possible that such objects could merge with a new companion (Heggie, 1975; Rodriguez et al., 2018). Similarly, primordial black holes (Hawking, 1971) are not bound by the limitations of stellar evolution.

We do not expect either of these mechanisms to significantly affect the position and shape of an excess due to PPSN, although they complicate the interpretation of the maximum black hole mass. It is possible that black holes formed through repeated mergers could be identified on a case by case basis. For example, black holes formed by a binary black hole merger event are expected to have large dimensionless spins, $a \sim 0.7$ for equal mass non-spinning pre-merger black holes (Scheel et al., 2009).

	α	$m_{\max}$	λ	m_{pp}	σ_{pp}	β	$m_{\min}$	δm
H_0	[-3, 7]	**100**	**0**	N/A	N/A	**0**	[2,10]	**0**
H_1	[-3, 7]	[10,100]	**0**	N/A	N/A	**0**	[2,10]	**0**
H_2	[-3, 7]	[10,100]	**0**	N/A	N/A	[-5,5]	[2,10]	[0,10]
H_3	[-3, 7]	[10,100]	[0,1]	[25,100]	(0,5]	**0**	[2,10]	[0,10]
H_4	[-3, 7]	[10,100]	[0,1]	[25,100]	(0,5]	[-5,5]	[2,10]	**0**
H_5	[-3, 7]	[10,100]	[0,1]	[25,100]	(0,5]	[-5,5]	[2,10]	[0,10]
MC	1.5	35	0.1	35	1	2	3	5

TABLE 4.2: Summary of example models. The prior ranges for our (hyper)parameters in each model are indicated. Each of these distributions is uniform over the stated range. The fixed parameters are in bold. "MC" refers to the values chosen for the simulated universe in Sec. 4.4. These values are chosen to be consistent with current observational data and theoretical predictions of pulsational pair-instability supernovae and population synthesis modeling.

4.4 Monte Carlo Study

We verify that we are able to recover a distribution described by a particular set of (hyper)parameters using a Monte Carlo injection study. We create a simulated universe in which the black hole mass distribution follows our model with (hyper)parameters given in Tab. 4.2. For simplicity, we draw all of the extrinsic parameters from the geometrically determined prior distribution used by LALINFERENCE (Veitch et al., 2015) with luminosity distance extending to $4Gpc$. We draw the masses according to Eq. 4.11 and draw black hole spins uniformly in spin magnitude and isotropically in orientation.

Motivated by the prediction of pulsational pair-instability supernovae, our simulated universe includes a Gaussian component centred at the upper limit of the power law component, $m_{\max} = 35M_{\odot}$. The Gaussian component has a width $\sigma_{\mathrm{pp}} = 1M_{\odot}$ and the mixing fraction $\lambda = 0.1$. The inferred value of α is covariant with other parameters of the model, e.g., for the first four detections, decreasing the maximum black hole mass, decreases the inferred value of α (Fishbach & Holz, 2017). We set $\alpha = 1.5$ for our injection study, which is consistent with that analysis. We choose the spectral index of the secondary mass distribution to be $\beta = 2$ to reflect the preference of population synthesis models to produce near equal mass binaries. We impose a lower mass cut-off of $3M_{\odot}$ with a turn-on of $\delta m = 5M_{\odot}$. These values are chosen to be consistent with current observational data. The distribution of primary masses with this choice of (hyper)parameters is shown in Fig. 4.1. We can see that our model gives us a bimodal distribution with peaks at $\approx 7M_{\odot}$, due to the smooth turn-on, and $35M_{\odot}$, due to PPSN.

To enforce selection effects, we keep only binaries with optimal matched filter signal to noise ratio, $\rho > 8$, in a single Advanced LIGO detector operating at design sensitivity (Abbott, 2016j). We generate a set of 200 events for our simulated universe. Each signal is then injected into a three detector LIGO-Virgo network with all detectors operating at their design sensitivities.

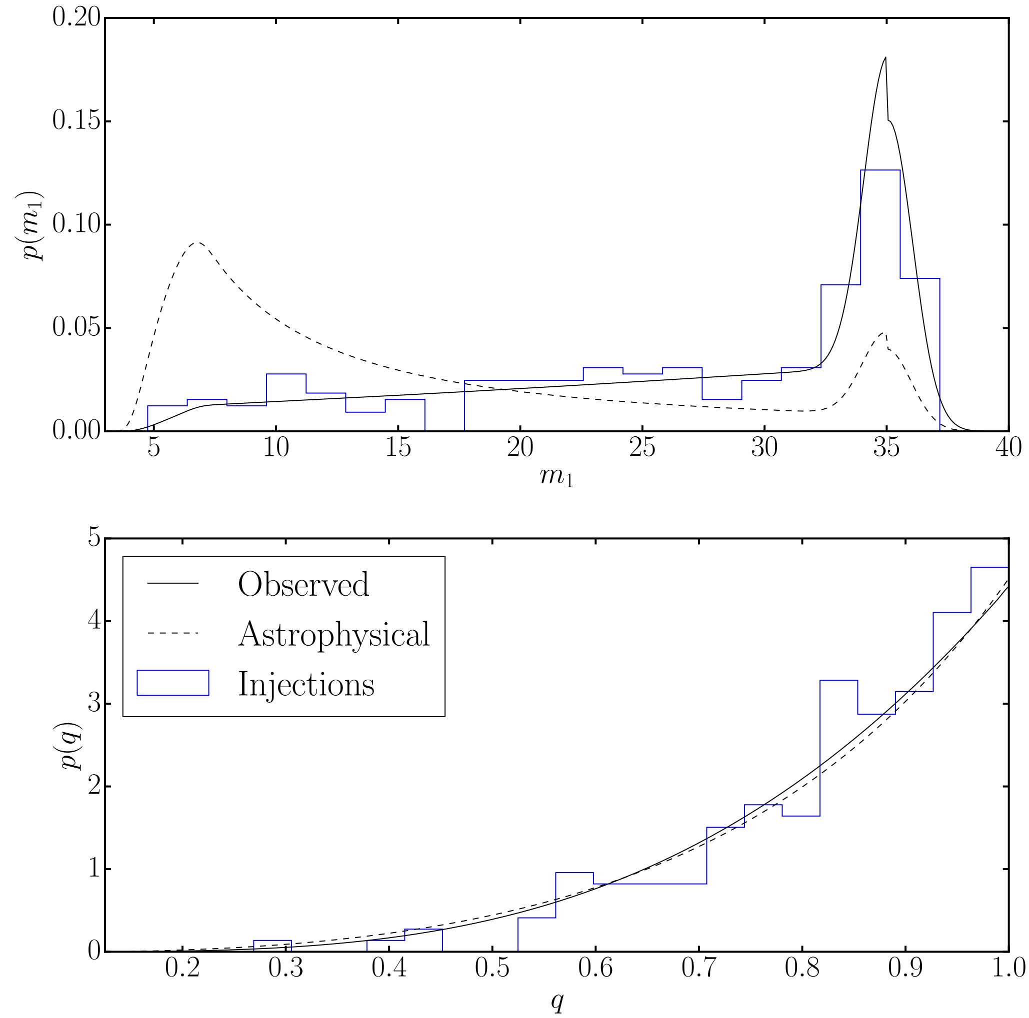

FIGURE 4.2: The distribution of source-frame primary mass and mass ratio ($q \equiv m_2/m_1$) for our simulated universe, see Tab. 4.2. The dashed and solid lines show the distribution before and after accounting for selection biases respectively. The blue histogram indicates the injected values.

Fig. 4.2 shows the distribution of primary masses and mass ratio in our simulated universe before (dashed) and after (solid) accounting for observation bias. The blue histogram indicates the injected values.

Using the recovered posterior distributions for the injected events, we employ the statistical methods described above for each of our models. The Bayes factors comparing H_5 to the others are enumerated in Table 4.3. In Table 4.3 we also give an approximate number of events needed to reach our threshold $\ln BF = 8$, assuming linear growth of $\ln BF$ with number of detections. We consider two cases. "Cosmic" assumes zero measurement error. All uncertainty comes from cosmic variance. "Design" uses posterior samples obtained through running LALINFERENCE for a three detector network operating at design sensitivity. Including measurement errors reduces our resolving power between any pair of models by a significant factor for all

	H_0	H_1	H_2	H_3	H_4
Cosmic $\ln BF_i^5$	253.0	55.0	18.0	31.0	1.0
Design $\ln BF_i^5$	161.0	14.0	5.0	7.0	-1.0
N_{expected}	10	100	300	250	$\gg 200$

TABLE 4.3: The log Bayes factor comparing each of the hypotheses summarized in Tab. 4.2 to the correct model, H_5, given the 200 samples shown in Fig. 4.2 with population (hyper)parameters as specified in Tab. 4.2. "Cosmic" indicates that the masses are used with no measurement error, this represents an upper limit on how well we can differentiate the two distributions. "Design" uses the output of LALInference for a three detector Advanced LIGO/Virgo network operating at design sensitivity. The bottom row gives an approximate number of events to reach our threshold of Design $\ln BF = 8$. Measuring the shape of the low-mass cut-off will require many detections, due to the lower sensitivity at low masses.

the models. Unless otherwise specified we will refer to the Design Bayes factors. Below, we consider the effect of each of the modifications on the mass distribution model.

4.4.1 Upper-Mass Cut-Off

After 200 events, our model without the variable upper-mass cut-off, H_0, is disfavored with a log Bayes factor of ~ 160. We determine how many events are necessary to surpass the threshold of $\ln BF_0^5 = 8$ by considering subsets of our injection set. After 20 detections $\ln BF_0^5 \sim N(\mu = 13.3, \sigma = 2.4)$. Here, $N(\mu, \sigma)$ denotes a normal distribution with mean μ and variance σ^2.

Given this, we expect to be able to identify an upper-mass cut-off in the mass distribution after $\lesssim 20$ events. We note that μ grows linearly with number of events, this scaling is used in Table 4.3 to approximate the number of events to reach $\ln BF = 8$. This is consistent with a similar study by Fishbach & Holz, 2017.

4.4.2 PPSN Peak

After 200 events, the posterior distribution on λ, the fraction of black holes formed through PPSN, is shown in Figure 4.4. We measure the maximum posterior probability point and 95% highest density confidence interval (HDI) to be $\lambda \sim 0.11^{+0.07}_{-0.04}$ (all future confidence regions will be 95% HDI unless specified) and disfavor $\lambda = 0$ at $\gtrsim 3\sigma$. Correspondingly, $\ln BF_2^5 = 4.9$ which is moderate evidence for the existence of the PPSN peak, but below our threshold for a confident detection.

Figure 4.3 shows the one and two-dimensional posterior distribution for the position and width of the PPSN peak. We meausre $m_{\text{pp}} = 34.4^{+1.0}_{-1.2} M_\odot$, $\sigma_{\text{pp}} = 1.2^{+0.9}_{-1.2} M_\odot$. We note that the peak and width of the distribution are covariant, with smaller values of m_{pp} requiring a larger σ_{pp}. This is unsurprising as the highest mass black holes, $\sim 40 M_\odot$, must be accounted for. The posterior distribution on λ shows no significant correlation with m_{pp} or σ_{pp}.

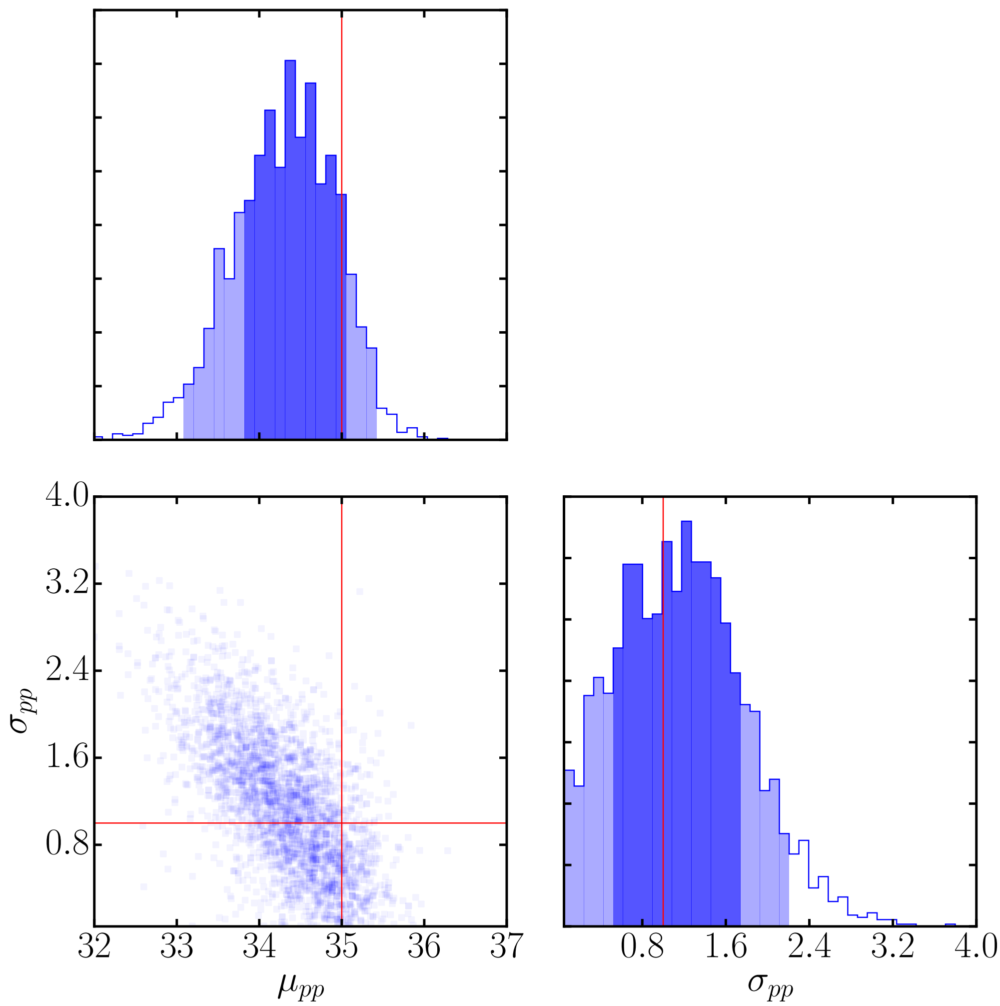

FIGURE 4.3: The posterior on the mean and width of the PPSN peak using our 200 events injected into Gaussian noise. After 200 detections we can measure the position and width of the PPSN peak to within $\sim 1 M_{\odot}$ at 95% confidence. The dark and light shaded regions indicate the one-dimensional 68% and 95% confidence intervals.

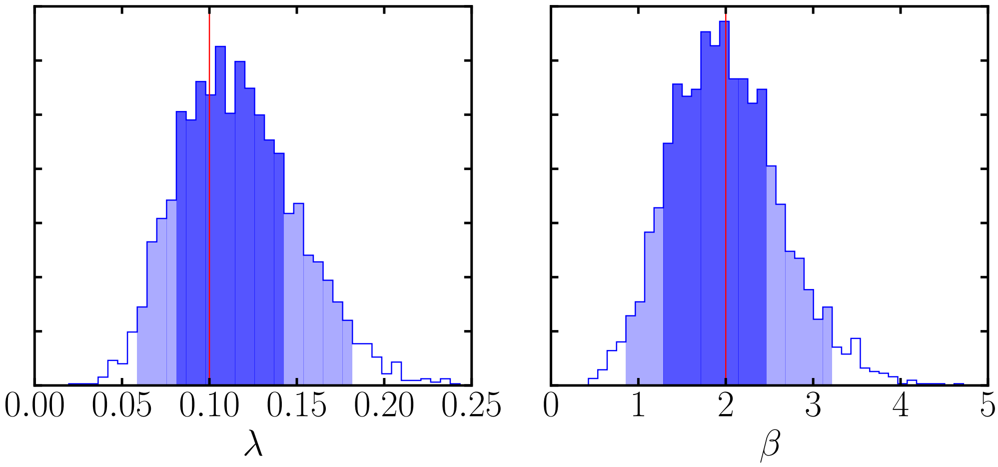

FIGURE 4.4: The posterior on the fraction of black holes formed through PPSN and the power-law index on the mass ratio using our 200 events injected into Gaussian noise. We can measure the fraction of black holes formed through PPSN to be $\lambda \sim 0.11^{+0.07}_{-0.04}$ at 95% confidence. We can determine the spectral index of the mass ratio distribution to within ± 1 at 95% confidence with 200 detections.

4.4.3 Mass Ratio

The posterior distribution on β is shown in Figure 4.4. After 200 events, the 1σ and 2σ confidence intervals on β span 1.1 and 2.2 respectively. We disfavor H_3 with a log Bayes factor of $\ln BF^5_3 = 7.1$ after our 200 injections, just below our threshold of 8.

4.4.4 Low Mass

The (hyper)parameters describing the low-mass end of the distribution are more difficult to measure than the high-mass (hyper)parameters due to the observation bias favoring high-mass systems. After 200 events, there is no evidence for or against the low-mass smoothing described by δm. This is unsurprising since only 9 injected binaries have $m_1 \lesssim 8M_\odot$, above which H_4 and H_5 are identical. Measuring the same events with improved strain sensitivity would not improve our sensitivity to δm. We are limited by cosmic variance: Cosmic $\ln BF^5_4 = 1$. The two-dimensional posterior distribution on $m_{\min}$ and δm, Figure 4.5, shows the correlation between low minimum masses and long turn-on lengths.

4.4.5 Mass Distribution Recovery

As a qualitative measure of the difference between the inferred mass distributions, we plot the posterior predictive distribution for the primary mass given the binaries in our injection study for our models in figure 4.6. The dashed black line indicates the injected distribution. We can see the effect of the different (hyper)parameterizations.

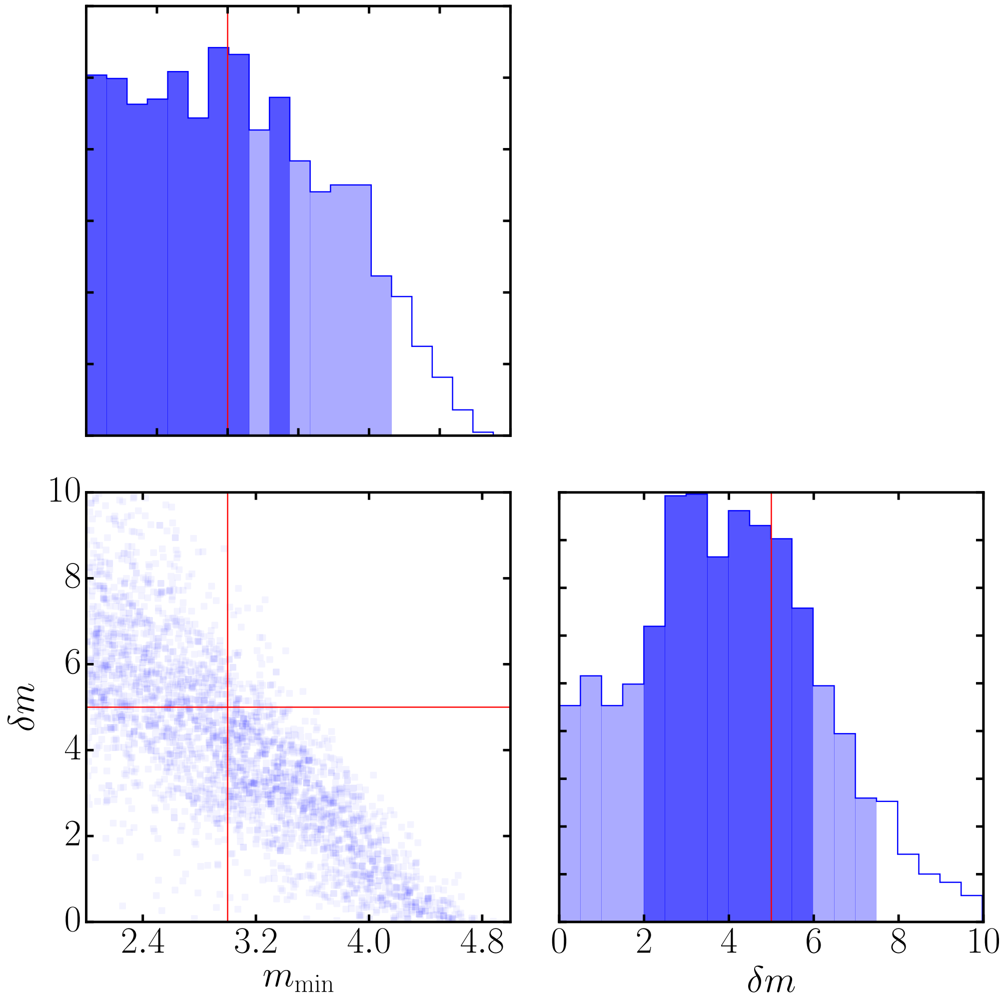

FIGURE 4.5: The posterior on the (hyper)parameters describing the low-mass end of the black holes mass spectrum using our 200 events injected into Gaussian noise. These parameters are difficult to measure as only about 5% of events have $m_1 < 8M_{\odot}$. We can see the clear covariance between a low minimum mass and a long turn-on length.

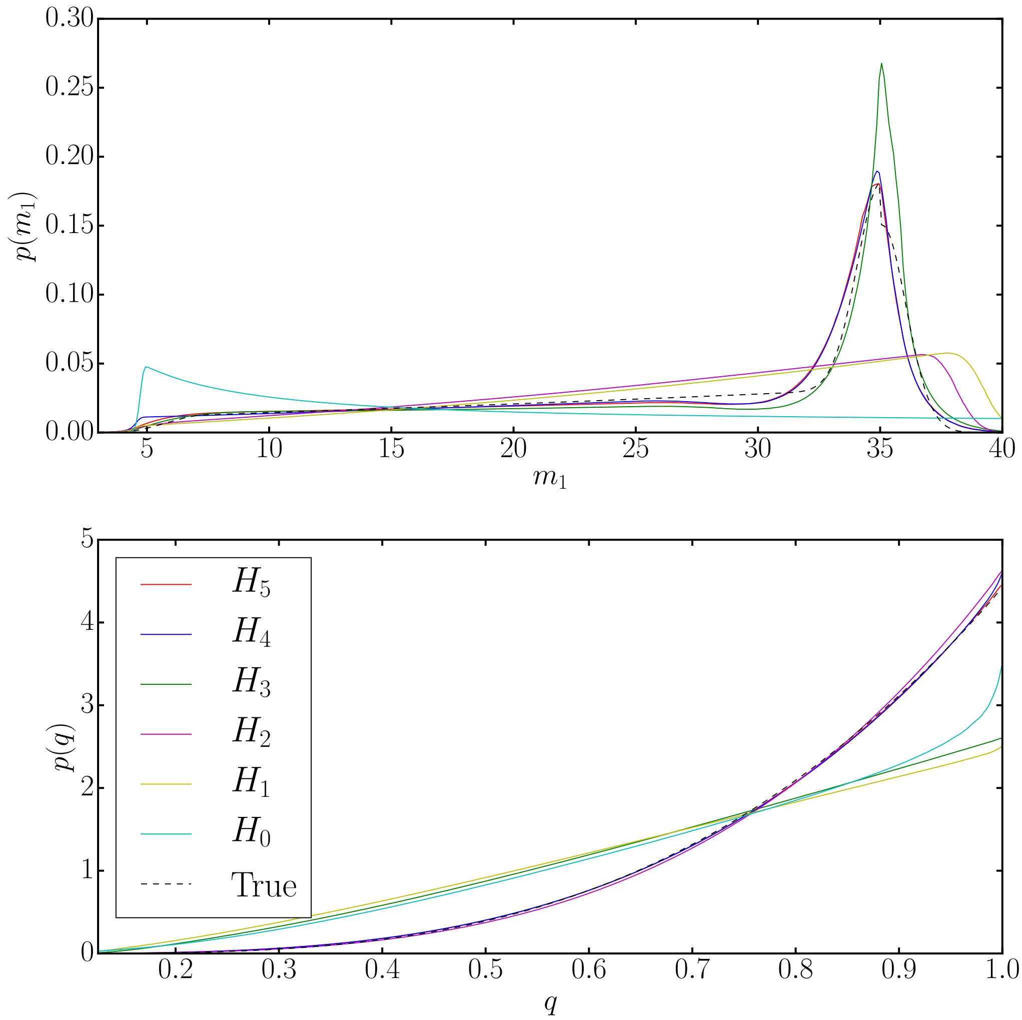

FIGURE 4.6: Posterior predicitive distribution (see Eq. 4.4) for m_1 and q for our models from Tab. 4.2 after 200 injected events. The dashed black line indicates the true distribution and the notches indicate the injected values. These distributions represent the observed distribution of source-frame masses.

In order to accommodate the lack of black holes with $m \gtrsim 40M_{\odot}$, α is overestimated in model H_0. This leads to an overly steep inferred distribution and an overestimate of the total merger rate (see Sec. 4.4.6). Models H_1 and H_2, which do not include the Gaussian component, favor a mass spectrum which is less steep than the injected distribution. This manifests as a positive gradient after accounting for observation bias. If we do not fit the mass ratio power-law index, we overestimate the number of heavy black holes. This bias is seen as an overestimate of λ in H_3 and maximum mass in H_1.

4.4.6 Impact on the Merger Rate

The majority of binary black hole mergers are not individually resolvable by Advanced LIGO/Virgo. Using a (hyper)parameterization which does not accurately describe the true distribution leads to a biased estimate of the fraction of mergers which are individually resolvable and hence the merger rate (Abadie, 2010; Abbott, 2017f). Compact binary coalescences are a Poisson process which can be described by a merger rate $R(\Lambda)$. For a detector with time-independent sensitivity and a model of the distribution of binary black hole systems, the merger rate can be inferred from: the number of observed events N, the sensitive volume of our detectors $V(\Lambda)$, and the observation time T,

$$R(\Lambda) = \frac{N}{V(\Lambda)T}, \tag{4.12}$$

where

$$V(\Lambda) = \int d\Theta \, \pi \left(\Theta|\Lambda\right) V_{obs}(\Theta), \tag{4.13}$$

and $V_{obs}(\Theta)$ is the sensitive volume to a given binary introduced in section Sec. 4.2.

To illustrate the dependence of R on the mass distribution model, we calculate the posterior distribution for the inferred merger rate estimate for the models described in Tab. 4.2. For each model, we compute the posterior distribution for R. During the first observing run of Advanced LIGO, ~ 3 binary black hole mergers were identified in ~ 48 days joint observing time (Abbott, 2016b). We use these values, $N = 3$, $T = (48/365)$ yr to normalize our rate estimates. We neglect the (currently large) Poisson uncertainty in the arrival rate since this will be small once 200 detections have been made. Fig. 4.7 shows the posterior distribution for the merger rate for each of our models. We note that if we assume the power law mass distribution extends out to $100M_{\odot}$, the dash-dotted line, we overestimate the merger rate by a factor of 2-3. This is due to the much larger α required to be consistent with the lack of detections at high masses. A similar result is obtained in Wysocki, 2017 by simultaneously fitting the merger rate and power-law spectral index.

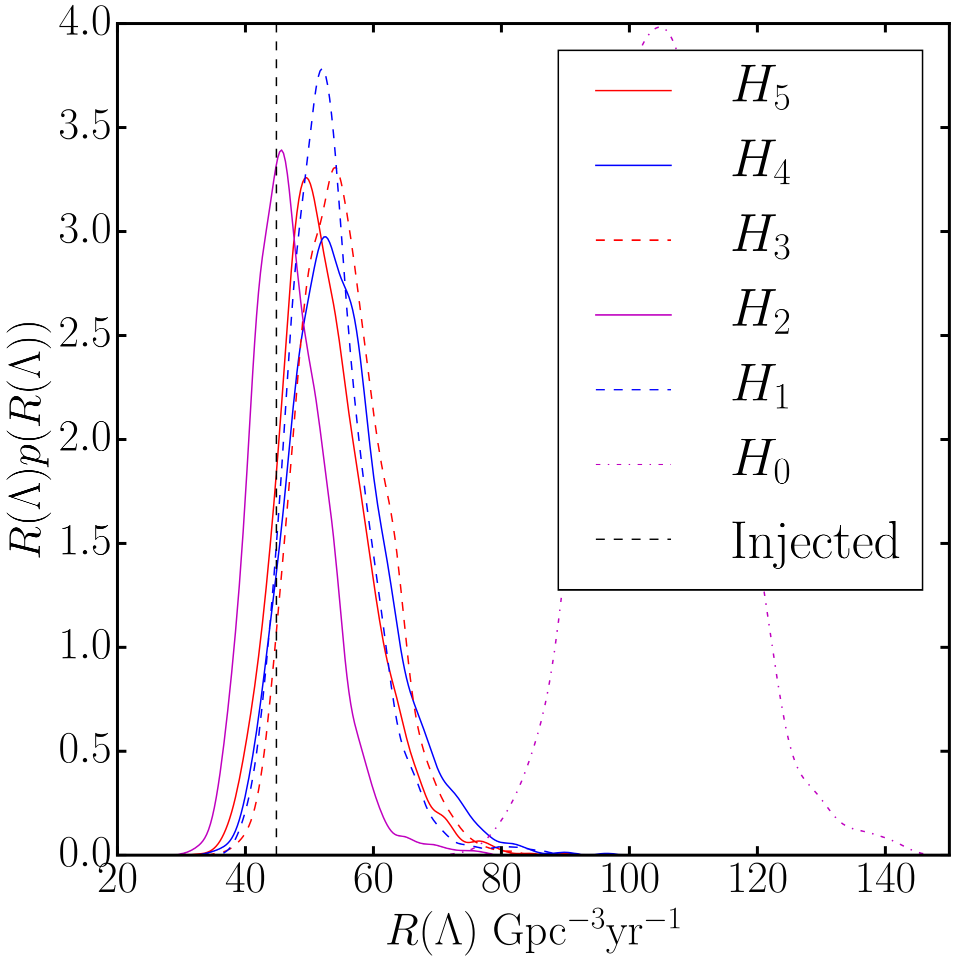

FIGURE 4.7: Posterior distribution for the binary black hole merger rate after 200 simulated events. We assume a detection rate consistent with Advanced LIGO's first observing run, $N/T \approx 23\mathrm{yr}^{-1}$, and ignore Poisson uncertainties. The models are described in Tab. 4.2. The dashed black line indicates the rate for the injected distribution. If we do not fit the maximum mass (the dash-dotted line) the rate is overestimated by a factor of 2-3. The inferred merger rate is not strongly sensitive to any of the other modifications to the mass function.

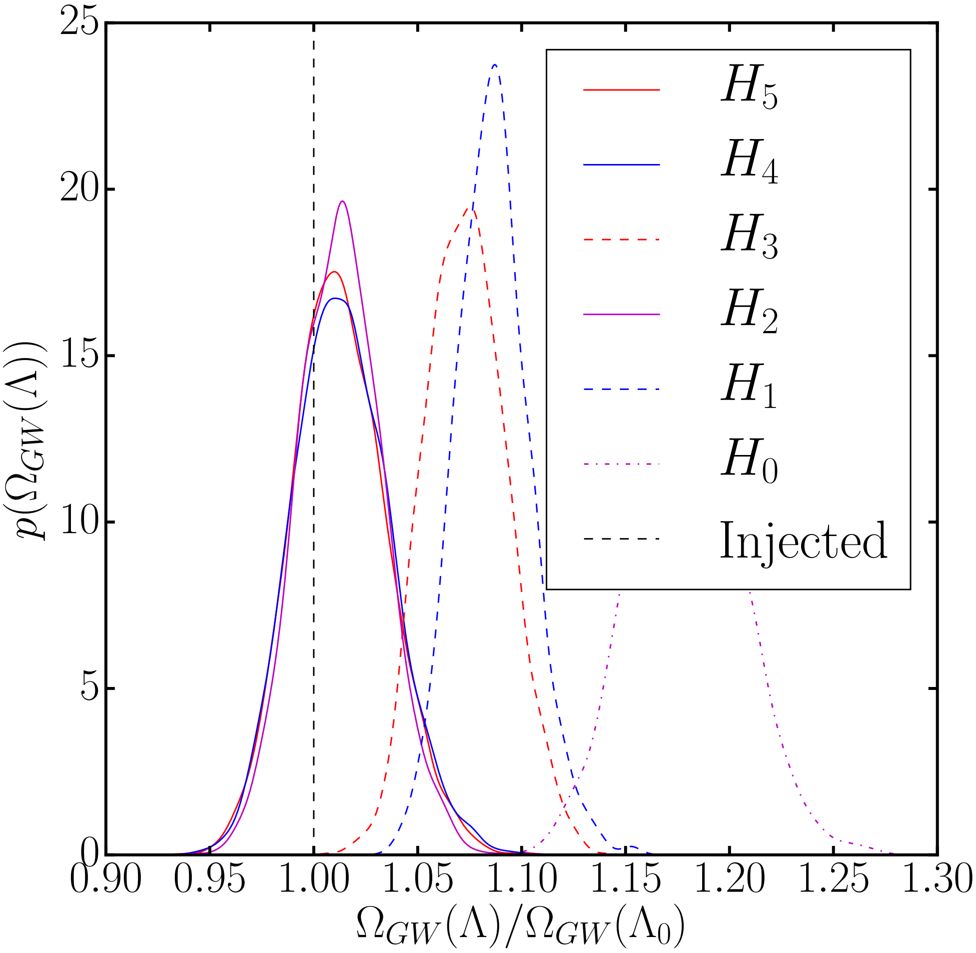

FIGURE 4.8: Posterior distribution for the ratio of the amplitude of the expected stochastic gravitational-wave background to the value using the injected distribution. The models are described in Tab. 4.2. The dashed lines indicate models in which the mass ratio distribution is assumed to be uniform. The dash-dotted line indicates the model in which the maximum black hole mass is fixed to be $100M_{\odot}$. Allowing the maximum mass of the power law component to vary decreases the predicted amplitude of the stochastic background by $\sim 10\%$. Relaxing the assumption that the distribution of secondary masses in uniform between the minimum mass and the primary mass decreases the predicted amplitude by $\sim 10\%$.

4.4.7 Impact on the Stochastic Background

Unresolvable mergers are widely believed to make the dominant contribution to the SGWB (Abbott, 2018c; Abbott, 2016d; Abbott, 2017l). The SGWB is typically characterized by the ratio of the energy density of the universe in gravitational-waves to the energy required to close the universe, Ω_{GW}. The most sensitive frequency of current detectors to the SGWB is $\sim$ 30Hz, this frequency corresponds to the inspiral phase of all binaries relevant to this work. The energy density due to binary black hole mergers depends on the distribution of chirp masses and the merger rate (Zhu et al., 2011),

$$\Omega_{GW} \sim \langle \mathcal{M}^{5/3} \rangle R. \tag{4.14}$$

As seen above, cutting off the mass distribution around $40M_{\odot}$ leads to a reduction in the merger rate, however, this is accompanied by an increase in $\mathcal{M}^{5/3}$. Overall, this leads to an $\sim$ 10% reduction in the expected SGWB as seen in Figure 4.8. We also observe that relaxing the assumption that the secondary mass is uniformly distributed leads to a further $\sim$ 10% reduction in $\Omega_{\rm GW}$. This is because the chirp mass is maximized for equal mass binaries for a given primary mass. These reductions are smaller than the current uncertainty on the amplitude of the background due to Poisson uncertainty in the observed merger rate. The current method of searching for this background is by cross-correlating the strain data from the two LIGO detectors, this method will take more time to resolve a weaker background.

The cross-correlation method is expected to require years of observation before the background can be resolved. Recently, a method involving searching directly for the stochastic background due to binary black hole mergers has been introduced in Smith & Thrane, 2018. This method is expected to be able to detect this component of the background using days of data. Since this method relies on the rate of binary black hole mergers rather than $\Omega_{\rm GW}$ it will be more sensitive to the black hole mass function than cross-correlation searches.

4.5 Discussion

The first gravitational-wave detections are revealing a previously unexplored population of black holes. While we are still in the regime of small-number statistics, the systems observed to date may be suggestive of a cut-off in the black hole mass spectrum at $\sim 40M_{\odot}$. This is consistent with the predicted black hole mass distribution if stars with initial masses $M \gtrsim 100M_{\odot}$ undergo pulsational pair-instability supernovae. We hypothesize that, if this is the cause of the cut-off, then there should be a corresponding excess of black holes at around the same mass. We construct a phenomenological model, which captures this behavior. In agreement with Fishbach & Holz, 2017, we find that the presence of an upper mass cut-off can be identified at high significance with $O(10)$ events.

We highlight several other interesting results that can be obtained using 200 detections at design sensitivity:

1. We will be able to identify the presence of an excess due to PPSN at $\sim 3\sigma$ and constrain the fraction of black holes forming through PPSN to within ~ 0.05 at 95% confidence.

2. We can measure the position and width of the PPSN graveyard to within $\sim 1 M_{\odot}$.

3. We will be able to measure the power-law index on the mass ratio to within $\sim \pm 1$.

Detailed measurement of the low-mass end of the mass distribution will most likely require 1000s of detections and may have to wait for future detectors, e.g., the proposed Einstein Telescope (Punturo et al., 2010) or Cosmic Explorer (Abbott, 2017c).

We demonstrate that neglecting the presence of either a cut-off or a mass peak can lead to a mis-recovery of the astrophysical distribution of black holes in merging binaries. For example, the higher sensitivity of current detectors to high-mass binaries means that in order to fit the upper mass range well, the low-mass distribution is biased. This leads to incorrect estimates of the total binary black hole merger rate and the predicted amplitude of the SGWB. The amplitude of the SGWB is also sensitive to the distribution of mass ratios.

Our analysis assumes that a clear distinction can be made between binary black hole systems and other compact binaries. In reality, if there is not a well-defined mass gap between neutron stars and black holes, it will be non-trivial to distinguish between binary black hole, neutron star-black hole, and binary neutron star systems (Yang, East & Lehner, 2018). Although, differences in, e.g., the spins of the component objects may enable this distinction (Littenberg et al., 2015). Our framework can be naturally expanded to include these other classes of compact binaries.

Addendum

Since the publication of this paper additional compact binary coalescences have been observed and significant progress has been made in theoretical and observational analysis. The reader is directed to Chapters 1 and 8 for an overview of the field at the time of writing.

Chapter 5

Gravitational-wave memory: waveforms and phenomenology

Published as:
Talbot C., Thrane, E., Lasky, P. D. & Lin, F., Phys. Rev. D 98, 064031 (2018).

Abstract

The non-linear gravitational-wave memory effect is a prediction of general relativity in which test masses are permanently displaced by gravitational radiation. We implement a method for calculating the expected memory waveform from an oscillatory gravitational-wave time series. We use this method to explore the phenomenology of gravitational-wave memory using a numerical relativity surrogate model. Previous methods of calculating the memory have considered only the dominant oscillatory ($\ell = 2$, $m = |2|$) mode in the spherical harmonic decomposition or the post-Newtonian expansion. We explore the contribution of higher-order modes and reveal a richer phenomenology than is apparent with $\ell = |m| = 2$ modes alone. We also consider the "memory of the memory" in which the memory is, itself, a source of memory, which leads to a small, $O\left(10^{-4}\right)$, correction to the memory waveform. The method is implemented in the python package GWMEMORY, which is made publicly available.

5.1 Introduction

The non-linear (Christodoulou) gravitational-wave memory is a permanent displacement of freely-falling test masses due to the passage of gravitational waves (Zel'dovich & Polnarev, 1974; Braginsky & Thorne, 1987; Christodoulou, 1991). This memory effect can be understood as the travelling gravitational waves themselves sourcing gravitational radiation. Gravitational-wave memory may be detectable by advanced LIGO (Aasi, 2015) and Virgo (Acernese, 2015) by considering an ensemble of detections (Lasky et al., 2016; McNeill, Thrane & Lasky, 2017), especially if the low-frequency sensitivity can be increased (Yu et al., 2018). Future interferometers such as LISA, Cosmic Explorer (Abbott, 2017c) and Einstein Telescope (Punturo et al., 2010) may be able to resolve the memory effect for individual binaries (Favata, 2009a; Yang & Martynov, 2018). Gravitational-wave memory is also a target for pulsar timing arrays (van Haasteren & Levin, 2010; Pshirkov, Baskaran & Postnov, 2010; Seto, 2009; Cordes & Jenet, 2012; Wang et al., 2015; Arzoumanian et al., 2015).

While extracting memory from numerical relativity simulations is possible (Pollney & Reisswig, 2011), it is time-consuming and dependent on the waveform extraction method (Taylor et al., 2013; Bishop & Rezzolla, 2016; Blackman et al., 2017). The minimal waveform model (MWM) (Favata, 2009a; Favata, 2009b; Favata, 2010) uses analytic expressions for the memory from the inspiral, using the post-Newtonian expansion, and quasi-normal mode ringdown and incorporates uncertainty in the memory sourced during merger with a "fudge factor". The MWM assumes the oscillatory emission is well-described by the $\ell = |m| = 2$ spin-weighted spherical harmonic modes. For binaries with unequal masses and/or large spins this assumption is known to break down (Calderón Bustillo et al., 2016; Calderón Bustillo, Laguna & Shoemaker, 2017). In this work we implement a previously suggested method of calculating the memory which avoids these issues (Wiseman & Will, 1991; Thorne, 1992; Favata, 2010) and explore the phenomenology of the gravitational-wave memory from binary black holes.

For the memory sourced by the $\ell = |m| = 2$ modes it is possible to choose the gauge such that the memory is entirely "+" polarized and the inclination dependence is $\delta h_+ \propto \sin^2 \iota \left(17 + \cos^2 \iota\right)$. The binary inclination, ι, is the angle between the angular momentum vector of the binary and the line-of-sight between the binary and the observer. We make this choice of gauge throughout to emphasize the deviation from the behaviour when including additional oscillatory modes. In this paper, we demonstrate that including additional, "higher-order", modes in the calculation of the gravitational-wave memory leads to $O(10\%)$ corrections to the predicted strain and a richer phenomenology of gravitational-wave memory than previously believed[1].

Since the memory effect is sourced by gravitational radiation, the memory itself contributes to a higher-order memory effect that we call "memory of the memory". We iteratively include higher-order memory terms and

[1]The impact of higher-order oscillatory modes on the memory for non-spinning binaries is also considered in a Masters thesis by Goran Dojcinoski (Favata, 2018)

demonstrate that each memory order is suppressed by a factor of ~ 100 with respect to the previous order.

Measuring gravitational-wave memory will allow new tests of general relativity and alternative theories of gravity. For example, massive graviton theories predict a memory amplitude which is dependent on the mass of the graviton and discretely different from general relativity (Kilicarslan & Tekin, 2019). Additionally, the memory effect is significantly reduced in spacetimes with more than four non-compactified dimensions (Hollands, Ishibashi & Wald, 2017; Satishchandran & Wald, 2018; Garfinkle et al., 2017). Recently, it has been suggested that the inclination dependence of the memory could be used as a test of general relativity (Yang & Martynov, 2018). Given that in this work we demonstrate that including higher-order oscillatory modes changes the inclination dependence, care should be taken to avoid false detection of deviations from general relativity. Indeed, failing to consider higher-order oscillatory modes has been shown to lead to similar false detections of deviation (Pang et al., 2018).

The remainder of the paper is structured as follows. In the following section, we describe a method by which the gravitational-wave memory can be computed from an arbitrary spherical harmonic decomposed time-domain gravitational waveform. We then explore the phenomenology of the gravitational-wave memory describing how the (ℓ, m) content of the oscillatory waveform affects the (ℓ, m) content of the memory. After this, we consider the memory of the memory and demonstrate that the higher-order memory terms are strongly suppressed. Finally, we present some closing thoughts.

5.2 Calculating Gravitational-Wave Memory

The non-linear memory sourced by gravitational waves can be expressed as an integral of the quadrupole moment of the gravitational-wave flux (Wiseman & Will, 1991; Thorne, 1992; Favata, 2010)

$$\delta h_{jk}^{TT}(T_R, \Omega) = \frac{4G}{Rc^4} \int_{-\infty}^{T_R} dt \int_{S^2} d\Omega' \frac{dE}{dtd\Omega'} \left[\frac{n_j n_k}{1 - n^l N_l} \right]^{TT}. \tag{5.1}$$

Here, $n(\Omega')$ is a unit vector, $N(\Omega)$ is the unit line-of-sight vector drawn from the observer at Earth to the source and the energy flux is

$$\frac{dE}{dtd\Omega} = \frac{R^2 c^3}{16\pi G} \left| \dot{h}\left(t, \Omega\right) \right|^2, \tag{5.2}$$

where $\dot{h} \equiv dh/dt$ and h is the gravitational-wave strain. We use Einstein summation convention throughout. The angles $\Omega = (\iota, \phi)$ are the inclination and a reference phase the source (typically the phase at coalescence for compact binaries), T_R is the retarded time, Ω' describes a sphere centered on the source with a radius R, the distance between the source and the observer, and TT denotes the transverse-traceless gauge.

We project onto the polarization basis by contracting with the polarization tensors, e_{+}^{ij}, $e_{\times}^{ij}$ (Anderson et al., 2001)

$$\delta h = \delta h_{+} - i\delta h_{\times} = \frac{1}{2}\delta h_{jk}^{TT}(e_{+}^{jk} - ie_{\times}^{jk}). \tag{5.3}$$

It is convenient to project the gravitational-wave strain $h(t, \Omega)$ onto a basis of spin-weighted spherical harmonics,

$$h(t, \Omega) = h_{+}(t, \Omega) - ih_{\times}(t, \Omega) = h_{\ell m}(t)_{-2}Y_{\ell m}(\Omega). \tag{5.4}$$

This allows us to separate the time-dependence from the angular dependence using the same basis that is regularly used for numerical relativity waveform extraction.

Substituting Equations 5.2 and 5.4 into Equation 5.1, we separate the time and angular integrals

$$\delta h(T_R, \Omega) = \frac{R}{4\pi c} H_{\ell_1 \ell_2 m_1 m_2}(T_R) \Lambda_{\ell_1 \ell_2 m_1 m_2}(\Omega), \tag{5.5}$$

where we have defined

$$H_{\ell_1 \ell_2 m_1 m_2}(-\infty, T_R) \equiv \int_{-\infty}^{T_R} dt \dot{h}_{\ell_1 m_1}(t) \dot{\bar{h}}_{\ell_2 m_2}(t), \tag{5.6}$$

$$\begin{aligned} \Lambda_{\ell_1 \ell_2 m_1 m_2}(\Omega) \equiv \frac{1}{2}(e_{+}^{jk} - ie_{\times}^{jk}) \times \\ \int_{S^2} d\Omega'_{-2}Y_{\ell_1 m_1}(\Omega')_{-2}\bar{Y}_{\ell_2 m_2}(\Omega') \left[\frac{n_j n_k}{1 - n_l N_l}\right]^{TT}. \end{aligned} \tag{5.7}$$

Overbars denote the complex conjugate. We note that $\delta h \propto 1/R$ as $H_{\ell_1 \ell_2 m_1 m_2} \propto 1/R^2$.

We perform one more projection of $\Lambda_{\ell_1 \ell_2 m_1 m_2}$ onto the basis of spin-weighted spherical harmonics to facilitate combination of the oscillatory and memory waveforms,

$$\begin{aligned} \Gamma_{\ell m}^{\ell_1 \ell_2 m_1 m_2} &\equiv \int_{S^2} d\Omega \Lambda_{\ell_1 \ell_2 m_1 m_2 -2}\bar{Y}_{\ell m} \\ &= 2\pi \int_{-1}^{1} d\cos\iota \Lambda_{\ell_1 \ell_2 m_1 m_2}(\iota, 0)_{-2}\bar{Y}_{\ell m_1 - m_2}(\iota, 0), \end{aligned} \tag{5.8}$$

where we have used the fact that $\Lambda_{\ell_1 \ell_2 m_1 m_2} \propto e^{i(m_1 - m_2)\phi}$ to perform the integral over ϕ and evaluate the ι integral at $\phi = 0$. The variable Γ is a purely geometric factor, which we can think of as the coupling constant linking oscillatory "input" modes (l_1, m_1, l_2, m_2) to memory "output" mode (ℓ, m). The coefficients $\Gamma_{\ell m}^{\ell_1 \ell_2 m_1 m_2}$ are independent of the oscillatory waveform and so can be computed in advance to speed up evaluation at runtime. It is then necessary only to compute $H_{\ell_1 \ell_2 m_1 m_2}$ and look up the relevant $\Gamma_{\ell m}^{\ell_1 \ell_2 m_1 m_2}$.

The memory accumulates over the entire lifetime of the binary, however, we are only interested here in the memory sourced from the final moments

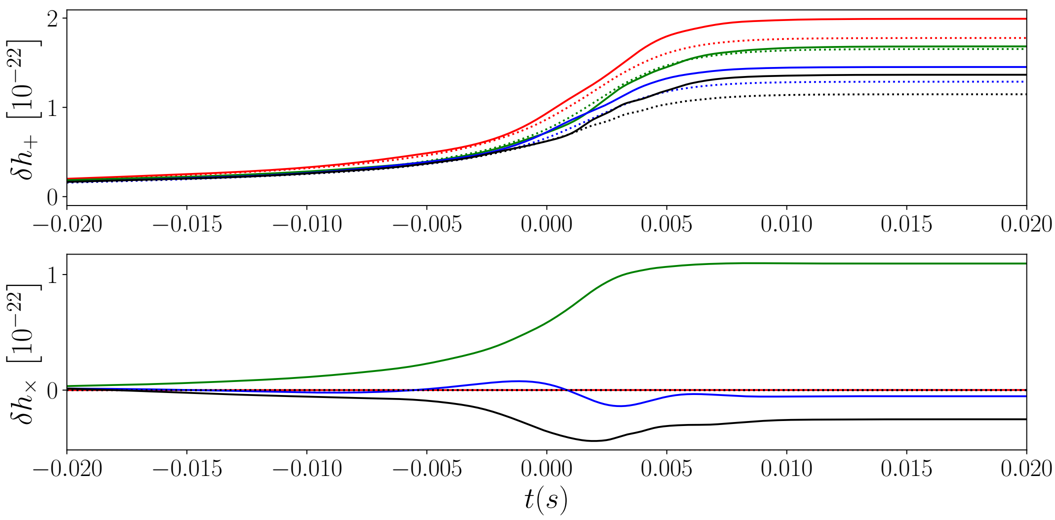

FIGURE 5.1: Including higher-order oscillatory modes significantly affects the predicted memory. Comparison of the $+$ (top panel) and $\times$ (bottom panel) polarizations of the memory time series when using only the $\ell = |m| = 2$ oscillatory modes (dotted) and when using all modes with $\ell \leq 4$ (solid). The colors are for binaries as follows: red is equal-mass ($q = 1$) and non-spinning ($S_1 = S_2 = \vec{0}$), green is equal-mass with precessing spins ($S_{||} = 0$, $S_\perp = 0.8$), blue is unequal-mass and non-spinning, black is unequal-mass ($q \equiv m_1/m_2 = 2$) with precessing spins. In all cases, the late-time memory is different by $O(10\%)$ compared with the $\ell = |m| = 2$ only case and is larger for large mass ratios and large, precessing, spins. For non-spinning binaries, this is due to the excitation of higher-order modes during merger and ringdown. Ignoring the higher-order modes completely removes the predicted $\times$ polarized memory. The systems shown are edge-on ($\iota = \pi/2$, $\phi = 0$) with total mass, $M = 60M_\odot$, at a luminosity distance, $D_L = 400\text{Mpc}$.

of the inspiral, merger and ringdown. Thus, we define the lower-limit of the time integral T_0 to be the time at which the binary enters the sensitive band of our detector, usually taken to be 20Hz for current detectors. Finally, we obtain

$$\delta h_{\ell m} = \frac{R}{4\pi c}\Gamma^{\ell_1 \ell_2 m_1 m_2}_{\ell m}(\Omega) H_{\ell_1 \ell_2 m_1 m_2}(T_0, T_R). \tag{5.9}$$

5.3 Memory Phenomenology

5.3.1 Importance of Higher-Order Modes

Previous studies of the gravitational-wave memory effect from compact binary coalescences have considered only memory sourced by the dominant, $\ell = |m| = 2$ mode of the oscillatory waveform. As mentioned above, in this case the angular dependence is given by (Favata, 2009a)

$$\delta h_+ \propto \sin^2\iota \left(17 + \cos^2\iota\right), \, \delta h_\times = 0. \tag{5.10}$$

This relation breaks down when additional modes are included and the angular dependence of the memory will depend on the relative size of the oscillatory spherical harmonic modes.

For our study, we use a numerical-relativity surrogate model, `NRSur7dq2` (Blackman et al., 2017). This model approximates the strain for all spin-weighted spherical harmonic modes with $2 \leq \ell \leq 4$ and is valid for mass ratios $1 \leq q \equiv m_1/m_2 \leq 2$ and dimensionless spin magnitudes up to 0.8. For all figures we choose a binary with a total mass of $60M_\odot$ at a luminosity distance of 400Mpc with binary inclination and polarization $\iota = \pi/2$, $\phi = 0$, unless otherwise stated. We begin the integration 0.08s before the merger.

The importance of including the higher-order modes in the calculation of memory is demonstrated in Fig. 5.1. We show the expected memory signal when considering only the $\ell = |m| = 2$ oscillatory modes and when using all modes with $\ell \leq 4$. We consider both non-spinning binaries and binaries with significant in-plane spins. The in-plane spins lead to precession of the orbital plane of the binary and have a larger contribution from higher-order oscillatory modes.

We can see that even in the case of an equal-mass non-spinning binary, including the higher-order modes leads to an $O(10\%)$ change in the predicted memory signal. This is due to the excitation of higher-order modes during the merger and ringdown portions of the coalescence. This effect is even more pronounced for precessing, unequal-mass, binaries. We observe that all of the considered systems other than the equal mass, non-spinning binary have a non-zero $\times$ component of the memory when the higher-order modes are included whereas the $\ell = |m| = 2$ memory is entirely plus polarized.

5.3.2 Mode Decomposition of the Oscillatory Waveform

We now explore the effect including additional modes in the oscillatory waveform has on the final amplitude of the memory signal for the binaries in Fig. 5.1. We consider limits on the sum in Equation 5.5 by progressively adding more pairs of spherical harmonic modes. Figure 5.2 shows how the late-time non-linear memory depends on the spherical harmonic modes considered.

We see that for non-spinning binaries (red and blue curves) the most important oscillatory modes are the $\ell = 2, 3$, $|m| = 2$. For unequal mass binaries (blue), there is a contribution from the $\ell = |m| = 3$ modes during merger, this leads to a $\times$-polarized memory component, even in the non-spinning case. Binaries with spins in the orbital plane (green and black curves) precess, this leads to excitation of $|m| \neq 2$ modes due to mode mixing (Hannam et al., 2014). Since there are now terms in our sum where $|m_1| \neq |m_2|$ we see a significant $\times$-polarized component in the memory.

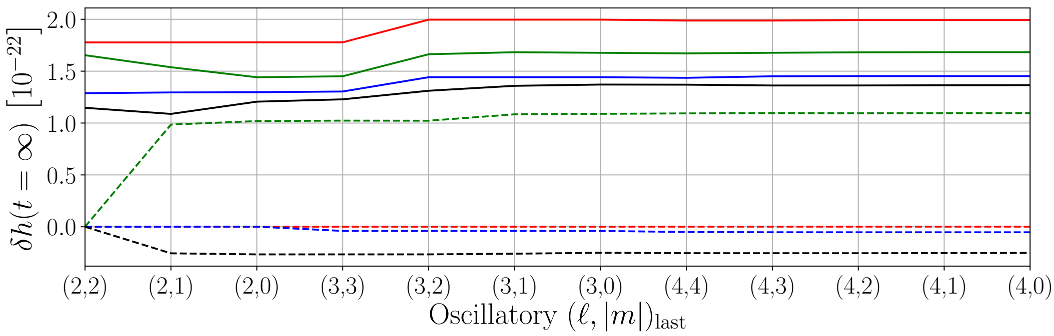

FIGURE 5.2: The late time $+$ (solid) and $\times$ polarizatons (dashed) of the memory amplitudes when including increasing numbers of modes in the oscillatory waveform (left to right). The horizontal axis $(\ell, |m|)_{\text{last}}$ indicates the last two oscillatory modes included in the calculation. The $(\ell = 2, |m| = 2)_{\text{last}}$ modes make the dominant contribution to the $+$ polarization and have no $\times$ component for all spins and mass ratios. When the $(\ell = 2, |m| = 1)_{\text{last}}$ modes are added we see that there is a non-zero $\times$ polarization for spinning systems. Including the $(\ell = 3, |m| = 2)_{\text{last}}$ modes has the largest effect of all the higher-order modes on the $+$ contribution to the late-time memory. The colors are for binaries as follows: red is equal-mass ($q = 1$) and non-spinning ($S_1 = S_2 = \vec{0}$), green is equal-mass with precessing spins ($S_{||} = 0$, $S_{\perp} = 0.8$), blue is unequal-mass and non-spinning, black is unequal-mass ($q \equiv m_1/m_2 = 2$) with precessing spins. The systems shown are edge-on ($\iota = \pi/2$, $\phi = 0$) with total mass, $M = 60M_{\odot}$, at a luminosity distance, $D_L = 400\text{Mpc}$. We note that the memory is not necessarily maximized for edge-on systems when higher-order modes are included.

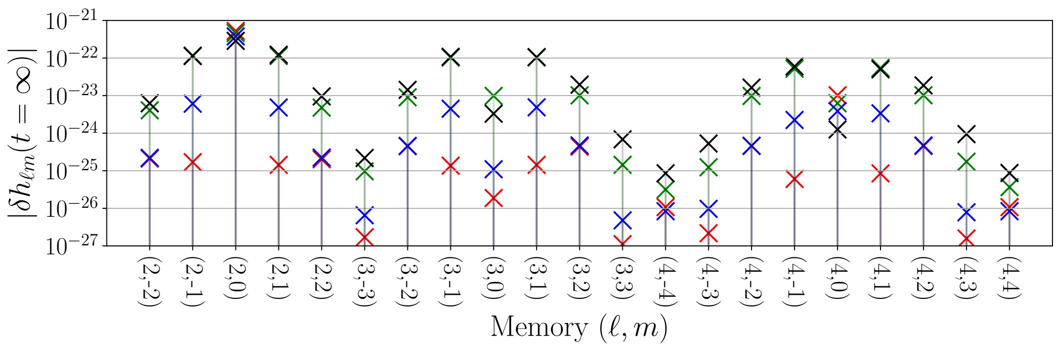

FIGURE 5.3: The spherical harmonic decomposition of the memory waveform for a range of mass ratios and spins. The absolute value of the late-time memory is shown as a function of the (ℓ, m) spherical harmonic decomposition of the memory. The dominant term is the $\ell = 2$, $m = 0$ mode for equal mass non-precessing binaries. For precessing binaries, the $\times$ terms in the memory integral lead to significant azimuthal dependence of the memory. This is seen in the $|m| = 1$ modes. The colors are for binaries as follows: red is equal-mass ($q = 1$) and non-spinning ($S_1 = S_2 = \vec{0}$), green is equal-mass with precessing spins ($S_{||} = 0$, $S_{\perp} = 0.8$), blue is unequal-mass and non-spinning, black is unequal-mass ($q \equiv m_1/m_2 = 2$) with precessing spins. The systems shown are edge-on ($\iota = \pi/2$, $\phi = 0$) with total mass, $M = 60M_{\odot}$, at a luminosity distance, $D_L = 400\text{Mpc}$.

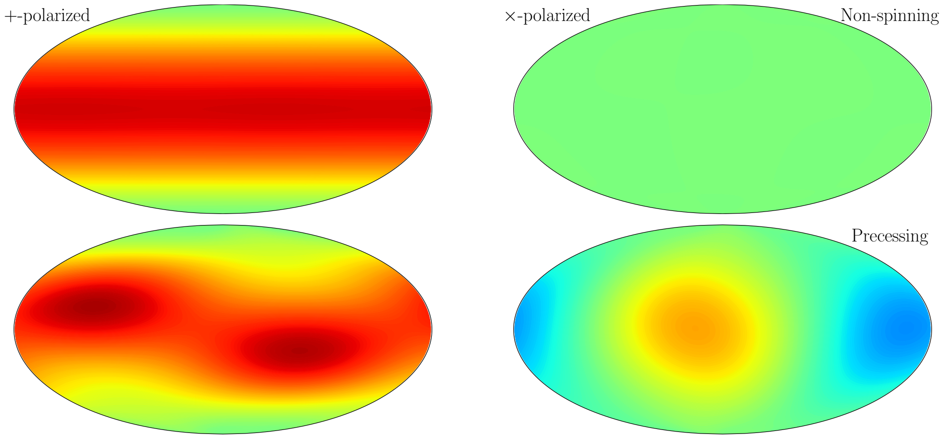

FIGURE 5.4: Angular dependence of the late-time $+$-(left) and $\times$-(right) polarized memory strain as a function of orientation angles ι (polar) and ϕ (azimuth). The top panels show the late-time memory for a non-spinning equal-mass binary and the bottom panels the late-time memory for a precessing equal-mass binary. The top panel follows the analytic expression given an oscillatory waveform containing only the $\ell = |m| = 2$ mode, $\delta h_+ \propto \sin^2\iota(17 + \cos^2\iota)$, $\delta h_\times = 0$. The bottom panel demonstrates how precessing systems give rise to a more complex memory structure.

5.3.3 Mode Decomposition of the Memory Waveform

For convenience, we decompose the memory onto the basis of spin-weighted spherical harmonics. This decomposition is given explicitly in Equation 5.9 where the $\Gamma^{\ell_1\ell_2 m_1 m_2}_{\ell m}$ map the "input" oscillatory modes to the "output" memory modes. Using the coefficients for $\ell = |m| = 2$ we recover the familiar $\sin^2\iota\left(17 + \cos^2\iota\right)$ dependence.

We use the Γ coefficients to decompose the memory onto this basis for the mass ratios and spins considered in Fig. 5.1. The angular spectral content of these memory waveforms is shown in Fig. 5.3. We see that the dominant term is the $\ell = 2$, $m = 0$ mode in all cases. Other modes are more important for higher mass ratios and binaries with large misaligned spins. For precessing sources (green/black) the $|m| = 1$ memory modes are nearly as large as the $m = 0$ modes. While the $m = 0$ contributions to the memory decay rapidly with increasing ℓ, the $|m| > 0$ modes converge more slowly. Therefore, it may be necessary to go consider $\ell > 4$ modes to ensure waveform fidelity at the sub-percent level.

Figure 5.4 shows the angular dependence of the late-time memory as a function of binary inclination (polar) and polarization (azimuth) for an equal mass binary. We consider two cases: non-spinning (top panels) and precessing (bottom). The $|m| = 1$ of the memory can be seen in the precessing case. We also draw the reader's attention to the non-vanishing $\times$ polarized memory for the precessing binary, in contrast to the non-spinning case. We note that the orientation dependence is a function of time as different memory modes grow at different rates, which is the cause of the structure in the memory time-series in Fig. 5.1 for precessing systems.

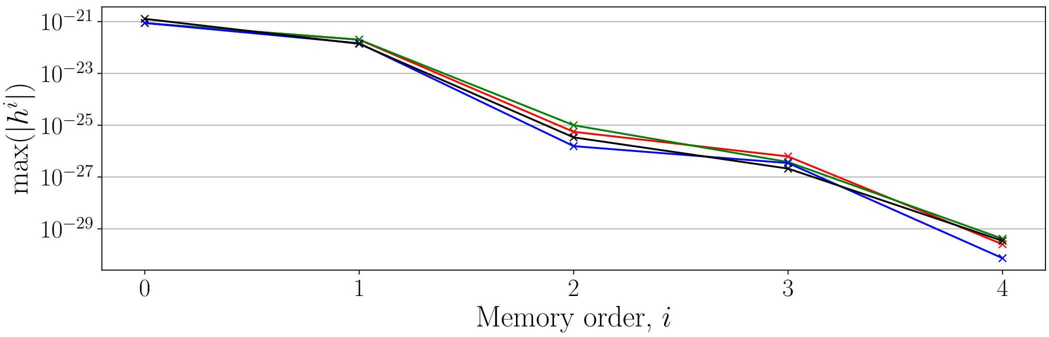

FIGURE 5.5: The maximum of the absolute value of the contribution to the memory entering at the ith iterative order for a range of mass ratios and spins. The peak of the oscillatory waveform corresponds to $i = 0$. The first-order memory is $i = 1$, we note that this is of the same order as the peak oscillatory strain. Each successive order of the memory is then on average two orders of magnitude smaller than the previous. The colors are for binaries as follows: red is equal-mass ($q = 1$) and non-spinning ($S_1 = S_2 = \vec{0}$), green is equal-mass with precessing spins ($S_{||} = 0$, $S_{\perp} = 0.8$), blue is unequal-mass and non-spinning, black is unequal-mass ($q \equiv m_1/m_2 = 2$) with precessing spins. The systems shown are edge-on ($\iota = \pi/2$, $\phi = 0$) with total mass, $M = 60 M_{\odot}$, at a luminosity distance, $D_L = 400$Mpc.

5.4 Memory of the Memory

Since the memory is sourced by gravitational radiation, the memory itself imparts a second-order "memory of the memory". To calculate this we replace $h_{\ell m}$ with $h^{\rm osc}_{\ell m} + \delta h^1_{\ell m}$ in Equation 5.4, where $\delta h^1_{\ell m}$ is the first-order memory. We apply this procedure iteratively to calculate the total strain

$$h_{\ell m} = h^{\rm osc}_{\ell m} + \sum_{i=1}^{\infty} \delta h^i_{\ell m}, \tag{5.11}$$

where δh^i is the contribution to the memory entering at the ith order.

Figure 5.5 shows the relative contribution of the different order memories for the systems considered previously. Each successive order is suppressed by $\sim$ two orders of magnitude with respect to the previous order. We do not expect these contributions to be significant for current detectors. However, the sensitivity of future detectors may be sufficient to measure the memory of the memory.

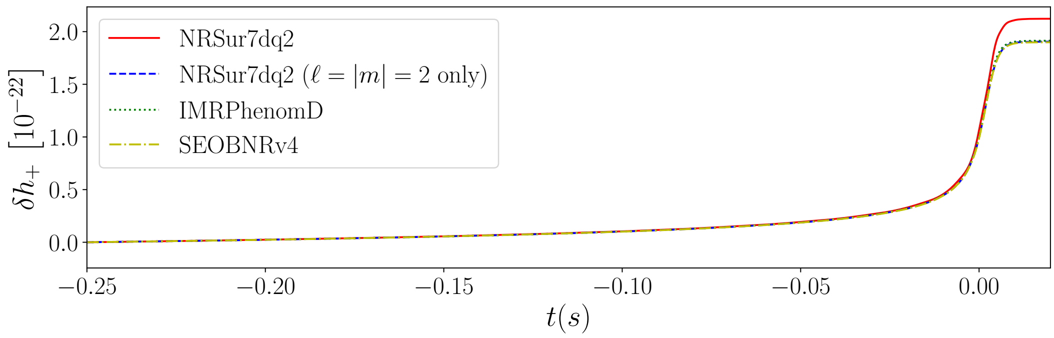

FIGURE 5.6: The plus component of the predicted memory using waveforms generated using different models. We compare the numerical relativity surrogate (NRSur7dq2) used in the rest of the paper, an effective one-body model (SEOBNRv4) and a phenomenological model (IMRPhenomD). The predicted memory agrees for all model when only considering the $\ell = |m| = 2$ oscillatory modes. As demonstrated above, including the higher-order oscillatory modes in the surrogate changes the predicted memory. The system shown is edge-on ($\iota = \pi/2$, $\phi = 0$), non-spinning, with total mass, $M = 60M_\odot$, equal mass, and at a luminosity distance, $D_L = 400$Mpc.

5.5 Memory Calculation Code

We release the Python package GWMEMORY[2] used in this work. The code enables calculation of the memory from arbitrary spherical harmonic decomposed gravitational waveforms along with functionality for creating waveforms using a range of commonly-used waveform families including numerical relativity surrogates, e.g., `NRSur7dq2` (Blackman et al., 2017), waveforms implemented in `LALSuite`[3], and numerical relativity waveforms. Additionally, we include an implementation of the MWM[4].

We have tested our waveform calculator using an aligned-spin effective one-body waveform approximant, `SEOBNRv4` (Bohé et al., 2017), a phenomenological waveform approximant, `IMRPhenomD` (Khan et al., 2016), and a numerical relativity surrogate, `NRSur7dq2` (Blackman et al., 2017). We find that the predicted memory does not strongly depend on the chosen oscillatory waveform family within each waveform's domain of validity, see Figure 5.6. The surrogate model is currently limited to mass ratios $q \leq 2$. The memory for aligned-spin binaries with mass ratio $q > 2$ can be calculated using the aligned-spin waveforms available in `LALSimulation`[5].

[2] https://github.com/ColmTalbot/gwmemory

[3] https://git.ligo.org/lscsoft/lalsuite

[4] We note that the minimal waveform model predicts a memory $\sim 20\%$ larger than our full calculation. We attribute this difference to the continuing development of the effective-one-body waveforms used to calibrate the MWM.

[5] While precessing waveforms are available in `LALSuite` the necessary decomposition into spherical harmonic modes is non-trivial and is not yet supported.

5.6 Discussion

Detection of gravitational waves from binary black hole mergers allows new tests of general relativity. In particular, we may be able to detect the gravitational-wave memory effect with current detectors (Lasky et al., 2016). In order to detect gravitational-wave memory using observations of merging binary black hole systems, it will be necessary to rapidly create high-fidelity frequency-domain memory waveforms for use in Bayesian parameter estimation.

The gravitational-wave memory is generally not extracted from numerical relativity simulations and is thus not modelled by the waveform approximants tuned to these simulations. For this reason, it is necessary to calculate the expected memory waveform from the oscillatory waveform as a post-processing step. We create a python package GWMEMORY to generate the memory waveform directly from arbitrary time-domain oscillatory waveforms.

Using this code, we provide a detailed analysis of the dependence of the observed memory waveform on the spectral content of the oscillatory signal and the binary orientation[6]. We find that the phenomenology of the gravitational-wave memory is richer than previously believed when subdominant oscillatory modes are included in the calculation of the memory. We additionally consider the contribution of the memory waveform to a "memory of the memory". While this effect is interesting from a pedagogical perspective, we find that this effect is small in all considered cases, and can be neglected with the current generation of gravitational-wave detectors.

Addendum

Since the publication of this paper additional compact binary coalescences have been observed and significant progress has been made in theoretical and observational analysis. The reader is directed to Chapters 1 and 8 for an overview of the field at the time of writing.

[6]The code used to generate the plots in this paper, along with demonstration of additional functionality, can be found at `https://github.com/ColmTalbot/gwmemory/examples/GWMemory.ipynb`.

Chapter 6

Bilby: A user-friendly Bayesian inference library for gravitational-wave astronomy

Published as:
Ashton, G., Hübner, M., Lasky, P. D., Talbot, C. *et al.*, Astrophys. J. Suppl. Ser. 241, 27 (2019).

Abstract

Bayesian parameter estimation is fast becoming the language of gravitational-wave astronomy. It is the method by which gravitational-wave data is used to infer the sources' astrophysical properties. We introduce a user-friendly Bayesian inference library for gravitational-wave astronomy, BILBY. This python code provides expert-level parameter estimation infrastructure with straightforward syntax and tools that facilitate use by beginners. It allows users to perform accurate and reliable gravitational-wave parameter estimation on both real, freely-available data from LIGO/Virgo, and simulated data. We provide a suite of examples for the analysis of compact binary mergers and other types of signal model including supernovae and the remnants of binary neutron star mergers. These examples illustrate how to change the signal model, how to implement new likelihood functions, and how to add new detectors. BILBY has additional functionality to do population studies using hierarchical Bayesian modelling. We provide an example in which we infer the shape of the black hole mass distribution from an ensemble of observations of binary black hole mergers.

6.1 Introduction

Bayesian inference underpins gravitational-wave science. Following a detection, Bayesian parameter estimation allows one to estimate the properties of a gravitational-wave source, for example, the masses and spins of the components in a binary merger (e.g., Abbott, 2016i; Abbott, 2016g; Abbott, 2016b; Abbott, 2018d; Abbott, 2019c). If the detection involves neutron stars, Bayesian parameter estimation is used to study the properties of matter at nuclear densities via the signature of tidal physics imprinted on the gravitational waveform (Abbott, 2017i; Abbott, 2019c; Abbott, 2018d). The posterior probability distributions of source parameters such as inclination angle can be used, in turn, to make inferences about electromagnetic phenomena such as gamma-ray bursts (e.g., Abbott, 2017e). Such parameter estimation is also used to measure cosmological parameters such as the Hubble constant (Abbott, 2017a). By combining data from multiple detections, Bayesian inference is used to understand the population properties of gravitational-wave sources (e.g., Abbott, 2016b; Talbot & Thrane, 2018; Wysocki et al., 2018; Smith & Thrane, 2018; Farr et al., 2017; Taylor & Gerosa, 2018; Roulet & Zaldarriaga, 2019, and references therein), which is providing insights into stellar astrophysics. By extending the gravitational-wave signal model, Bayesian inference is used to test general relativity and look for evidence of new physics (Abbott, 2016k; Abbott, 2017h; Abbott, 2017e; Abbott, 2018b; Abbott, 2018a)

The field of gravitational-wave astronomy is growing rapidly. We have entered the "open data era," in which gravitational-wave data has become publicly available (Vallisneri et al., 2015). Since Bayesian parameter estimation is central to gravitational-wave science, there is a need for a robust, user-friendly code that can be used by both gravitational-wave novices and experts alike.

The primary tool currently used by the LIGO and Virgo collaborations for parameter estimation of gravitational-wave signals is LALINFERENCE (Veitch et al., 2015). This pioneering code enabled the major gravitational-wave discoveries achieved during the first two LIGO observing runs (e.g., Abbott, 2016i; Abbott, 2016g; Abbott, 2016b; Abbott, 2018d; Abbott, 2019c). The code itself is now almost a decade old, and years of development have made it hard for beginners to learn, and difficult for experts to modify and adapt to new challenges. More recently, PYCBC INFERENCE (Biwer et al., 2019) was released; a modern, PYTHON-based toolkit designed for compact binary coalescence parameter estimation. This package provides access to several different samplers and builds on the PYCBC package (Nitz et al., 2018a) – an open-source toolkit for gravitational-wave astronomy.

We introduce BILBY, a user-friendly parameter-estimation code for gravitational-wave astronomy. BILBY provides expert-level parameter estimation infrastructure with straightforward syntax and tools that facilitate use by beginners. For example, with minimal user effort, users can download and analyze publicly-available LIGO and Virgo data to obtain posterior distributions for the astrophysical parameters associated with recent detections of binary black holes (Abbott, 2016h; Abbott, 2016f; Abbott, 2016b; Abbott, 2017f; Abbott, 2017g; Abbott, 2017h) and the binary neutron star merger (Abbott, 2017i).

One key functional difference between BILBY and LALINFERENCE/PYCBC INFERENCE is its modularity and adaptability. The core library is not specific to gravitational-wave science and has uses outside of the gravitational-wave community. Ongoing projects include astrophysical inference in multimessenger astronomy, pulsar timing, and x-ray observations of accreting neutron stars. The gravitational-wave specific library is also built in a modular way, enabling users to easily define their own waveform models, likelihood functions, etc. This implies BILBY can be used for more than studying compact binary coalescences—see Sec. 6.5. The modularity further ensures the code will be sufficiently extensible to suit the future needs of the gravitational-wave community. Moreover, we believe the wider astrophysics inference community will find the code useful by virtue of having a common interface and ideas that can be easily adapted to a range of inference problems.

The remainder of this paper is structured to highlight the versatile, yet user-friendly nature of the code. To that end, the paper is example driven. We assume familiarity with the mathematical formalism of Bayesian inference and parameter estimation (priors, likelihoods, evidence, etc.) as well as familiarity with gravitational-wave data analysis (antenna-response functions, power spectral densities, etc.). Readers looking for an introduction to Bayesian inference in general are referred to Skilling, 2004, while gravitational-wave specific introductions to inference can be found in Refs. (Veitch et al., 2015; Thrane & Talbot, 2019). Section 6.2 describes the BILBY design philosophy, and Sec. 6.3 provides an overview of the code including installation instructions in Sec. 6.3.1. Subsequent sections show worked examples. The initial examples are the sort of simple calculations that we expect will be of interest to most casual readers. Subsequent sections deal with increasingly complex applications that are more likely of interest to specialists.

The worked examples are as follows. Section 6.4 is devoted to compact binary coalescences. In 6.4.1, we carry out parameter estimation with publicly-available data to analyze GW150914, the first ever gravitational-wave event. In 6.4.2, we study a simulated binary black hole signal added to Monte Carlo noise. In 6.4.3, we study the matter effects encoded in the gravitational waveforms of a binary neutron star inspiral. In 6.4.4, we show how it is possible to add more sophisticated gravitational waveform phenomenology, for example, by including memory, eccentricity, and higher order modes. In 6.4.5, we study an extended gravitational-wave network with a hypothetical new detector.

Section 6.5 is devoted to signal models for sources that are not compact binary coalescences. In 6.5.1, we perform model selection for gravitational waves from a core collapse supernova. In 6.5.2, we study the case of a post-merger remnant. Section 6.6 is devoted to *hyper*-parameterization, a technique used to study the population properties of an ensemble of events. Closing remarks are provided in Section 6.8.

6.2 BILBY Design Philosophy

Three goals guide the design choices of BILBY. First, we seek to provide a parameter-estimation code that is sufficiently powerful to serve as a workhorse for expert users. Second, we aim to make the code accessible for novices, lowering the bar to work on gravitational-wave inference. Third, we desire to produce a code that will age gracefully; advances in gravitational-wave astronomy and Bayesian inference can be incorporated straightforwardly without resort to inelegant workarounds or massive rewrites. To this end, we adhere to a design philosophy, which we articulate with four principles.

- **Modularity.** Wherever possible, we seek to modularize the code and follow the abstraction principle (Pierce, 2002), reducing the amount of repeated code and easing development. For example, the sampler is a modularized object, so if a problem is initially analyzed using the PYMULTINEST (Buchner, 2014) sampler for example, one can easily switch to the EMCEE (Foreman-Mackey et al., 2013) sampler or even a custom-built gravitational-wave sampler. For example, BILBY accesses samplers through a common interface; as a result it is trivial to easily switch between samplers to compare performance or check convergence issues.

- **Consistency.** We enforce strict style guidelines, including adherence to the PEP8 style guide for PYTHON[1]. As a result, the code is relatively easy-to-follow and intuitive. In order to maintain integrity of the code while responding to the needs of a large and active user base, we employ GITLAB's merge request feature. Updates require approval by two experts. The PEP8 protocol is enforced using continuous integration.

- **Generality.** Wherever possible, we keep the code as general as possible. For example, the gravitational-wave package is separate from the package that passes the likelihood and prior to the sampler. This generality provides flexibility. For example, in Section 6.6, we show how BILBY can be used to carry out population inference, even though the likelihood function is completely different to the one used for gravitational-wave parameter estimation. Moreover, a general design facilitates the transfer of ideas into and out of gravitational-wave astronomy from the greater astro-statistics community.

- **Usability.** We observe that historically, people find it difficult to get started with gravitational-wave inference. In order to lower the bar, we

[1] https://www.python.org/dev/peps/pep-0008/

endeavour to make basic things doable with very few lines of code. We provide a large number of tutorials that can serve as a blueprint for a large variety of real-world problems. Finally, we endeavour to follow the advice of the PEP20 style guide for PYTHON[2]: "There should be one—and preferably only one—obvious way to do it." In other words, once users are familiar with the basic layout of BILBY, they can intuit where to look if they want to, for example, add a new detector (see Section 6.4.5) or include non-standard polarization modes.

6.3 Code Overview

6.3.1 Installation

BILBY is open-source, MIT licensed, and written in python. The simplest installation method is through PyPI[3]. The following command installs from the command line:

```
$ pip install bilby
```

This command downloads and installs the package and dependencies. The source-code itself can be obtained from the `git` repository (Ashton et al., 2018b), which also houses an issue tracker and merge-request tool for those wishing to contribute to code development. Documentation about code installation, functionality, and user syntax is also provided (Ashton et al., 2018a). Scripts to run all examples presented in this work are provided in the `git` repository.

6.3.2 Packages

BILBY has been designed such that logical blocks of code are separated and, wherever possible, code is abstracted away to allow future re-use by other models. At the top level, BILBY has three packages: `core`, `gw`, and `hyper`. The `core` package contains the key functionalities. It passes the user-defined priors and likelihood function to a sampler, harvests the posterior samples and evidence calculated by the sampler, and returns a `result` object providing a common interface to the output of any sampler along with information about the inputs. The `gw` package contains gravitational-wave specific functionality, including waveform models, gravitational-wave specific priors and likelihoods. The `hyper` package contains functionality for the hierarchical Bayesian inference (see Sec. 6.6). A flowchart showing the dependency of different packages and modules is available on the `git` repository (Ashton et al., 2018b).

[2] https://www.python.org/dev/peps/pep-0020/
[3] https://pypi.org/project/BILBY/

The `core` package

The `core` package provides all of the code required for general problems of inference. It provides a unified interface to several different samplers listed below, standard sets of priors including arbitrary user-defined options, and a universal result object that stores all important information from a given simulation.

Prior and likelihood functions are implemented as classes, with a number of standard types implemented in the `core` package: e.g., the `Normal`, `Uniform`, and `LogUniform` priors, and `GaussianLikelihood`, `PoissonLikelihood`, and `ExponentialLikelihood` likelihoods. One can write their own custom prior and likelihood functions by writing a new class that inherits from the parent `Prior` or `Likelihood`, respectively. The user only needs to define how the new prior or likelihood is instantiated and calculated, with all other house-keeping logic being abstracted away from the user.

The prior and likelihood are passed to the function `run_sampler`, which allows the user to quickly change the sampler method between any of the pre-wrapped samplers, and to define specific run-time requirements such as the number of live points, number of walkers, etc. Pre-packaged samplers include Markov Chain Monte Carlo Ensemble samplers `emcee` (Foreman-Mackey et al., 2013), `ptemcee` (Vousden, Farr & Mandel, 2016), `PyMC3` (Salvatier, Wiecki & Fonnesbeck, 2016), and Nested samplers (Skilling, 2004; Skilling, 2006) `MultiNest` (Feroz & Hobson, 2008; Feroz, Hobson & Bridges, 2009; Feroz et al., 2013) (through the PYTHON implementation `pyMultiNest` (Buchner, 2014)), `Nestle` (Barbary, 2015), `Dynesty` (Speagle, 2019), and `CPNest` (Veitch, 2017). The `Sampler` class again allows users to specify their own sampler by following the other examples.

Despite the choice of sampler, the output from BILBY is universal: an `hdf5` file (The HDF Group, 1997) that contains all output including posterior samples, likelihood calculations, injected parameters, evidence calculations, etc. The `Result` object can be used to load in these output files, and also perform common operations such as generating corner plots, and creating plots of the data and maximum posterior fit.

The `gw` package

The `gw` package provides the core functionality for parameter estimation specific to transient gravitational waves. Building on the `core` package, this provides prior specifications unique to such problems, e.g., a prior that is uniform in co-moving volume distance, as well as the standard likelihood used when studying gravitational-wave transients (Veitch et al., 2015 and Eq. 6.1), defined as the `GravitationalWaveTransient` class. The `gw` package also provides an implementation of current gravitational-wave detectors in the detector module, including their location and orientation, as well as different noise power spectral densities for both current and future instruments. Standard waveform approximants are also included in the `source` module, which are handled through the LALSIMULATION package (*Lalsuite*).

The gw package also contains a set of tools to load, clean and analyse gravitational-wave data. Many of these functions are built on the GWpy (Macleod et al., 2018) code base, which are contained within bilby.gw.detector and primarily accessed by instantiating a list of Interferometer objects. This functionality also allows one to implement their own gravitational-wave detector by instantiating a new Interferometer object—we show an explicit example of this in Sec. 6.4.5.

The hyper package

The hyper package contains all required functionality to perform hierarchical Bayesian inference of populations. This includes both a Model module and a HyperparameterLikelihood class. This entire package is discussed in more detail in Sec. 6.6.

6.4 Compact Binary Coalescence

In this section, we show a suite of BILBY examples analyzing binary black hole and binary neutron star signals.

We employ a standard Gaussian noise likelihood $\mathcal{L}$ for strain data d given source parameters θ (van der Sluys et al., 2008a; van der Sluys et al., 2008b; Veitch & Vecchio, 2008):

$$\ln \mathcal{L}(d|\theta) = -\frac{1}{2}\sum_k \left\{ \frac{[d_k - \mu_k(\theta)]^2}{\sigma_k^2} + \ln\left(2\pi\sigma_k^2\right) \right\}, \tag{6.1}$$

where k is the frequency bin index, σ is the noise amplitude spectral density, and $\mu(\theta)$ is the waveform. The waveform is a function of the source parameters θ, which consist of (at least) eight intrinsic parameters (primary mass m_1, secondary mass m_2, primary spin vector $\vec{S}_1$, secondary spin vector $\vec{S}_2$) and seven extrinsic parameters (luminosity distance d_L, inclination angle ι, polarization angle ψ, time of coalescence t_c, phase of coalescence ϕ_c, right ascension and declination ra and dec, respectively. Table 6.1 shows the default priors implemented for binary black hole systems. We show how these priors can be called in Secs. 6.4.1 and 6.4.2. Unless otherwise specified, $\mu(\theta)$ is given using the IMRPhenomP approximant (Hannam et al., 2014). However, the approximant can be easily changed; see Secs. 6.4.2 and 6.4.3. Moreover, it is relatively simple to sample in different parameters than those listed above (e.g., chirp mass and mass ratio instead of m_1 and m_2); examples for doing this are provided in the git repository (Ashton et al., 2018b).

6.4.1 GW150914: the onset of gravitational wave astronomy

The first direct detection of gravitational waves occurred on the 14$^{\text{th}}$ of September, 2015, when the two LIGO detectors (Aasi, 2015) in Hanford, Washington and Livingston, Louisiana detected the coalescence of a binary

variable	unit	prior	minimum	maximum
$m_{1,2}$	$M_\odot$	uniform	5	100
$a_{1,2}$	-	uniform	0	0.8
$\theta_{1,2}$	rad.	sin	0	π
$\delta\phi$, $\phi_{\rm JL}$	rad.	uniform	0	2π
d_L	Mpc	comoving	10^2	5×10^3
ra	rad.	uniform	0	2π
dec	rad.	cos	$-\pi/2$	$\pi/2$
ι	rad.	sin	0	π
ψ	rad.	uniform	0	π
ϕ_c	rad.	uniform	0	2π

TABLE 6.1: Default binary black hole priors. The intrinsic variables are the two black hole masses $m_{1,2}$, their dimensionless spin magnitudes $a_{1,2}$, the tilt angle between their spins and the orbital angular momentum $\theta_{1,2}$, and the two spin vectors describing the azimuthal angle separating the spin vectors $\delta\phi$ and the cone of precession about the system's angular momentum $\phi_{\rm JL}$. The extrinsic parameters are the luminosity distance d_L, the right ascension ra and declination dec, the inclination angle between the observers line of sight and the orbital angular momentum ι, the polarisation angle ψ, and the phase at coalescence ϕ_c. The phase, spins, and inclination angles are all defined at some reference frequency. We do not set a default prior for the coalescence time t_c. 'sin' and 'cos' priors are uniform in cosine and sine, respectively, and 'comoving' implies uniform in comoving volume.

black hole system (Abbott, 2016h). The gravitational waves swept through the two detectors with a $6.9^{+0.5}_{-0.4}$ ms time difference which, when combined with polarization information, allowed for a sky-location reconstruction covering an annulus of 590 deg^2 (Abbott, 2016h). The initially-published masses of the colliding black holes were given as $36^{+5}_{-4}\,M_\odot$ and $29^{+4}_{-4}\,M_\odot$ (Abbott, 2016i). Subsequent analyses with more accurate precessing waveforms constrained the masses to be $35^{+5}_{-3}\,M_\odot$ and $30^{+3}_{-4}\,M_\odot$ at 90% confidence (Abbott, 2016g). The distance to the source is determined to be 440^{+160}_{-180} Mpc (Abbott, 2016g).

In this example, we use BILBY to reproduce the parameter estimation results for GW150914. The data for published LIGO/Virgo events is made available through the Gravitational Wave Open Science Center (Vallisneri et al., 2015). Built-in BILBY functionality downloads and parses this data. We begin with the following two lines.

```
>>> import bilby
>>> interferometers = bilby.gw.detector.get_event_data("GW150914")
```

The first line of code imports the BILBY code-base into the PYTHON environment. The second line returns a set of objects that contain the relevant data segments and associated data products relevant for the analysis for both the LIGO Hanford and Livingston detectors. By default, BILBY downloads and windows the data. A local copy of the data is saved along with diagnostic plots of the gravitational-wave strain amplitude spectral density.

In addition to the data, the two key ingredients for any Bayesian inference calculation are the likelihood and the prior. Default sets of priors can be called from the `gw.prior` module, and we also employ the default Gaussian noise likelihood (Eq. 6.1).

```
>>> prior = bilby.gw.prior.BBHPriorDict(filename="GW150914.prior")
>>> likelihood = bilby.gw.likelihood.get_binary_black_hole_likelihood(
>>>     interferometers)
```

The above code calls the GW150914 prior, which differs from the priors described in Tab. 6.1 in two main ways. Firstly, to speed up the running of the code it restricts the mass priors to between 30 and 50 $M_\odot$ for the primary mass, and 20 and 40 $M_\odot$ for the secondary mass. Moreover, this prior call restricts the time of coalescence to 0.1 seconds before and after the known coalescence time. One can revert to the priors in Tab. 6.1 by replacing the above file call with `filename="binary_black_holes.prior"`, but this would require separately setting a prior for the coalescence time. We show how this can be done in Sec. 6.4.2.

The next step is to call the sampler:

```
>>> result = bilby.core.sampler.run_sampler(likelihood, prior)
```

This line performs parameter inference using the sampler default DYNESTY (Speagle, 2019), with a default 500 live points. This number can be increased by passing the `nlive=` keyword argument to `run_sampler()`. The sampler returns a list of posterior samples, the Bayesian evidence, and metadata, which is stored in an `hdf5` file. One may plot a corner plot showing the posterior distribution for all parameters in the model using the command

```
>>> result.plot_corner()
```

The above example code produces posterior distributions that by eye, agree reasonably well with the parameter uncertainty associated with the published distributions for GW150914. The shape of the likelihood for the extrinsic parameters presents significant challenges for samplers, due to strong degeneracy's between different sky locations, distances, inclination angles, and polarization angles (see e.g., Raymond & Farr, 2014; Farr et al., 2014). For more accurate results, we use the nested sampling package CPNEST (Veitch, 2017), which is invoked by changing the `run_sampler` function above to include the additional argument `sampler='cpnest'`. We also change the number of live points by adding `nlive=5000` to the same function, and specify a keyword argument `maxmcmc=5000`, which is the maximum number of steps the sampler takes before accepting a new sample. To resolve the issue with the phase at coalescence, we analytically marginalize over this parameter (Farr, 2014) by adding the optional `phase_marginalization=True` argument to the instantiation of the likelihood. BILBY has built in analytic marginalization procedures for the time of coalescence (Farr, 2014) and distance (Singer & Price, 2016; Singer, 2016), which can both be invoked using

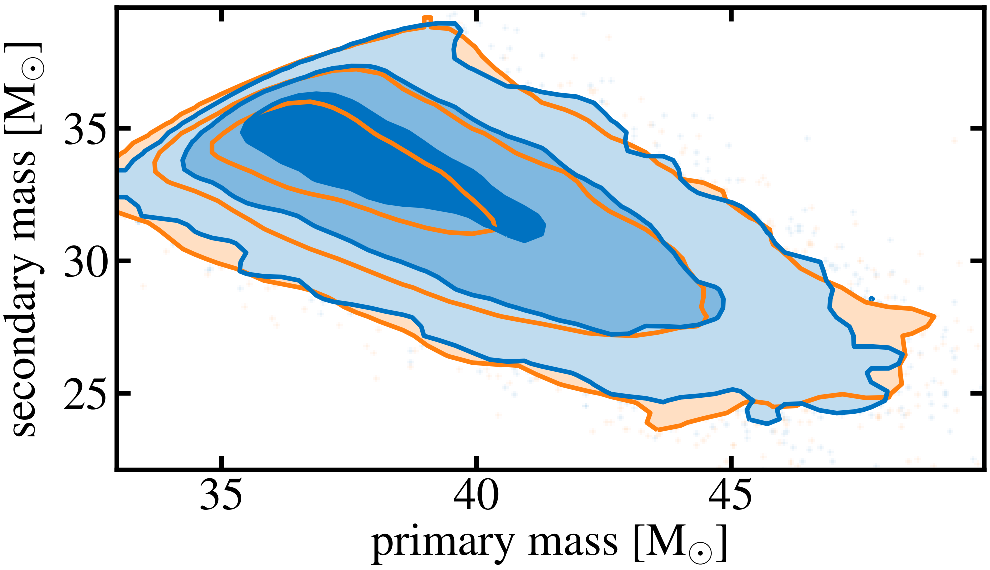

FIGURE 6.1: Marginalised posterior source-mass distributions for the first binary black hole merger detected by LIGO, GW150914. We show the posterior distributions recovered using BILBY (blue), and those using LALINFERENCE (orange), using open data from the Gravitational Wave Open Science Centre (Vallisneri et al., 2015). The five lines of BILBY code required for reproducing the posteriors are shown in Section 6.4.1.

`time_marginalization=True` and `distance_marginalization=True`, respectively. These decrease the run time of the code by minimizing the dimensionality of the parameter space. Posterior distributions can still be determined for these parameters by reconstructing them analytically from the full set of posterior samples e.g., see Thrane & Talbot, 2019.

Using BILBY we can plot marginalized distributions by simply passing the `plot_corner` function the optional `parameters=...` argument. In Fig. 6.1 we show the marginalized, two-dimensional posterior distribution for the masses of the two black holes as calculated using the above BILBY code (shown in blue). In orange we show the LIGO posterior distributions from (Abbott, 2016b), calculated using the LALINFERENCE software (Veitch et al., 2015), and hosted on the Gravitational Wave Open Science Center (Vallisneri et al., 2015).

In Fig. 6.2 we show the marginalized posterior distribution of the luminosity distance and inclination angle, where the BILBY posteriors are again shown in blue, and the LALINFERENCE posteriors in orange. Figure 6.3 shows the sky localisation uncertainty for both BILBY and LALINFERENCE.

The above example does not make use of detector calibration uncertainty, which is an important feature in LIGO data analysis. Such calibration uncertainty is built in to BILBY using the cubic spline parameterization (Farr, Farr & Littenberg, 2014), with example usage in the BILBY repository.

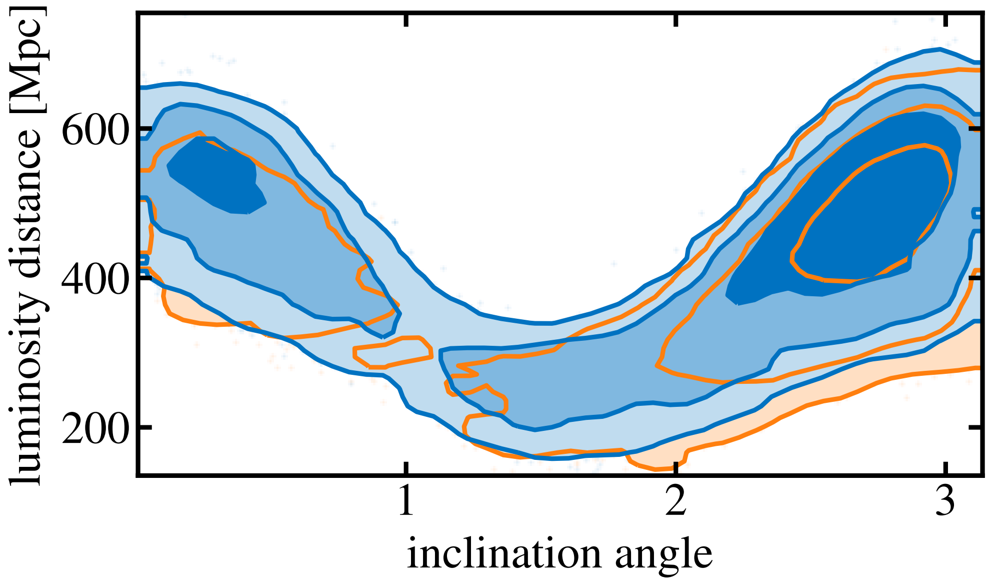

FIGURE 6.2: Marginalized posterior distributions on the binary inclination angle and luminosity distance for the first binary black hole merger detected by LIGO, GW150914. We show the posterior distributions recovered using BILBY (blue), and those using LALINFERENCE (orange), using open data from the Gravitational Wave Open Science Centre (Vallisneri et al., 2015).

6.4.2 Binary black hole merger injection

BILBY supports both the analysis of real data as in the previous section, as well as the ability to inject simulated signals into Monte Carlo data. In the following two sections we inject a binary black hole signal and a binary neutron star signal, respectively, showing how one can easily inject and recover signals and their astrophysical properties.

In this first example[4], we create a binary black hole signal with parameters similar to GW150914 (Abbott, 2016i), albeit at a luminosity distance of $d_L = 2$ Gpc (cf. $d_L \approx 400$ Mpc for GW150914). We inject the signal into a network of LIGO-Livingston, LIGO-Hanford (Aasi, 2015) and Virgo interferometers (Acernese, 2015), each operating at design sensitivity. When doing examples of this nature, it is time intensive to sample over all fifteen parameters in the waveform model. Therefore, to get quick results that can be run on a laptop, we only sample over four parameters in the waveform model: the two black-hole masses $m_{1,2}$, the luminosity distance d_L, and the inclination angle ι. BILBY supports simple functionality to limit or extend the number of parameters included in the likelihood calculation, as shown below.

We begin by setting up a `WaveformGenerator` object using a frequency domain strain model that takes the signal injection parameters and specific waveform arguments such as the waveform approximant as

[4]This example is found in the BILBY `git` repository at `https://git.ligo.org/Monash/bilby/blob/master/examples/injection_examples/basic_tutorial.py`.

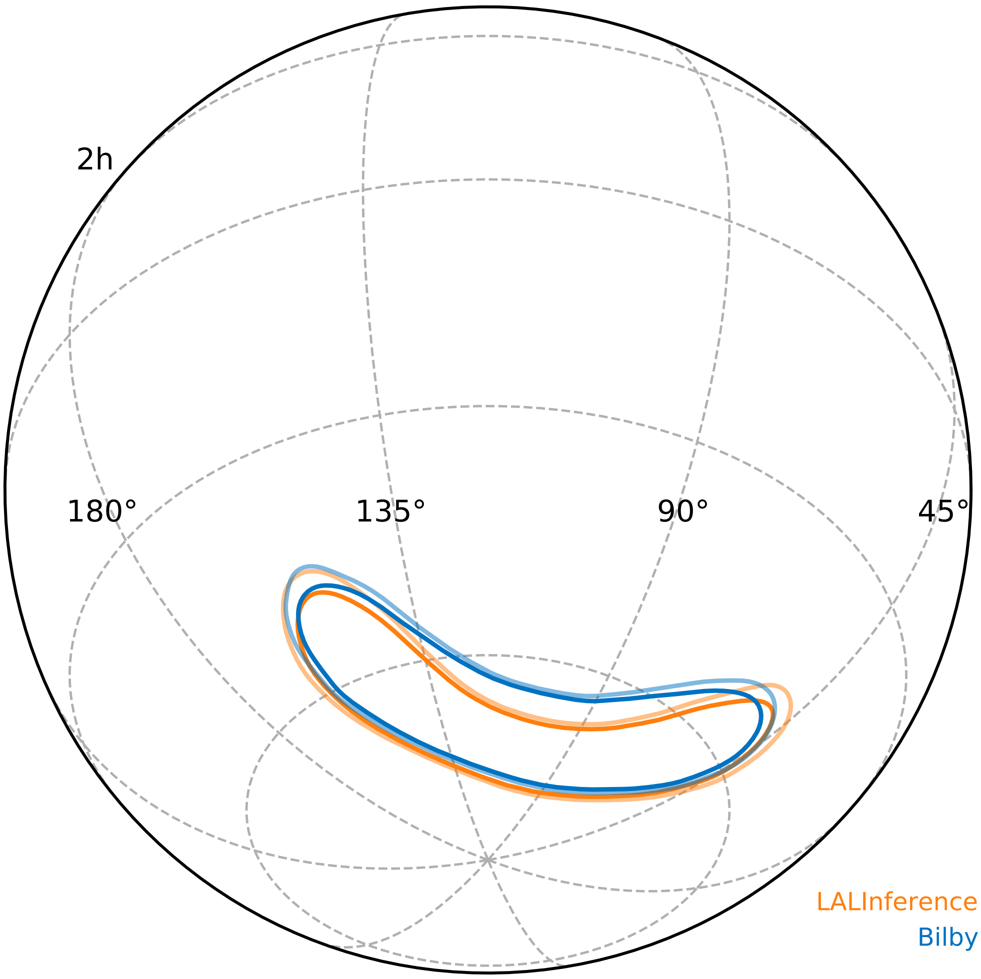

FIGURE 6.3: Sky localisation uncertainty for GW150914. The blue marginalized posterior distributions are those recovered using BILBY, and the orange are those recovered using LALINFERENCE, using open data from the Gravitational Wave Open Science Center (Vallisneri et al., 2015).

arguments. The `WaveformGenerator` also takes data duration and sampling frequency as input parameters. With the source model defined, we now instantiate an `interferometer` object that takes the strain signal from the `WaveformGenerator` and injects it into a noise realisation of the three interferometers. One could choose to do a zero-noise simulation by simply including the flag `zero_noise=True`.

Priors are set up as in the previous open data example, except we call the `binary_black_holes.prior` file instead of the specific prior file for GW150914. Moreover, to hold all but four of the parameters fixed, we set the value of the prior for those other parameters to the injection value. For example, setting

```
>>> prior["a_1"]=0
```

sets the prior on the dimensionless spin magnitude of the primary black hole to a delta-function at zero.

In general, we can change the prior for any parameter with one line of code. For example, to change the prior on the primary mass to be uniform between $m_1 = 25\ \mathrm{M}_\odot$ and $35\ \mathrm{M}_\odot$, say, one includes

```
>>> prior["mass_1"]=bilby.core.prior.Uniform(
>>>     minimum=25, maximum=35, unit=r"$M_\odot$")
```

BILBY knows about many different types of priors that can all be called in this way. For this example we are also required to define priors on the coalescence time, which we define to be a uniform prior with minimum and maximum one second either side of the injection time.

The likelihood is again set up similarly to the open-data example of Sec. 6.4.1, although this time we must pass the `interferometer`, `waveform_generator`, and `prior`. Finally, the sampler can be called in the same way as Sec. 6.4.1; for this example we use the `pyMultiNest` nested sampler (Buchner, 2014).

Figure 6.4 shows the recovered posterior distributions (blue) and the injected parameter values (orange). For this example, using the PYMULTINEST (Buchner, 2014) nested sampling package with 6000 live points took approximately 30 minutes on a laptop to sample fully the four-dimensional parameter space. The parameters in Fig. 6.4 are recovered well with the usual degeneracy present between the luminosity distance and inclination angle of the source, d_L and ι, respectively.

6.4.3 Measuring tidal effects in binary neutron star coalescences

The first detection of binary neutron star coalescence GW170817 was a landmark event signalling the beginning of multimessenger gravitational-wave astronomy (Abbott, 2017i; Abbott, 2017j). Gravitational-wave parameter estimation of the inspiral is what ultimately determined that both objects were likely neutron stars, and provides the best-yet constraints on the nuclear equation of state of matter at supranuclear densities (Abbott, 2017i; Abbott, 2017j; Abbott, 2017e).

One of the key measurements in determining the equation of state from binary neutron star coalescences is that of the tidal parameters. The dimensionless tidal deformability

$$\Lambda = \frac{2k_2}{3}\left(\frac{c^2 R}{Gm}\right)^5, \tag{6.2}$$

is a fixed parameter for a given equation of state and neutron star mass. Here, k_2 is the second Love number, R and m are the neutron star radius and mass, respectively. The binary neutron star merger GW170817 provided constraints of $\Lambda_{1.4} = 190^{+390}_{-120}$ (Abbott, 2018d; De et al., 2018), where the subscript denotes

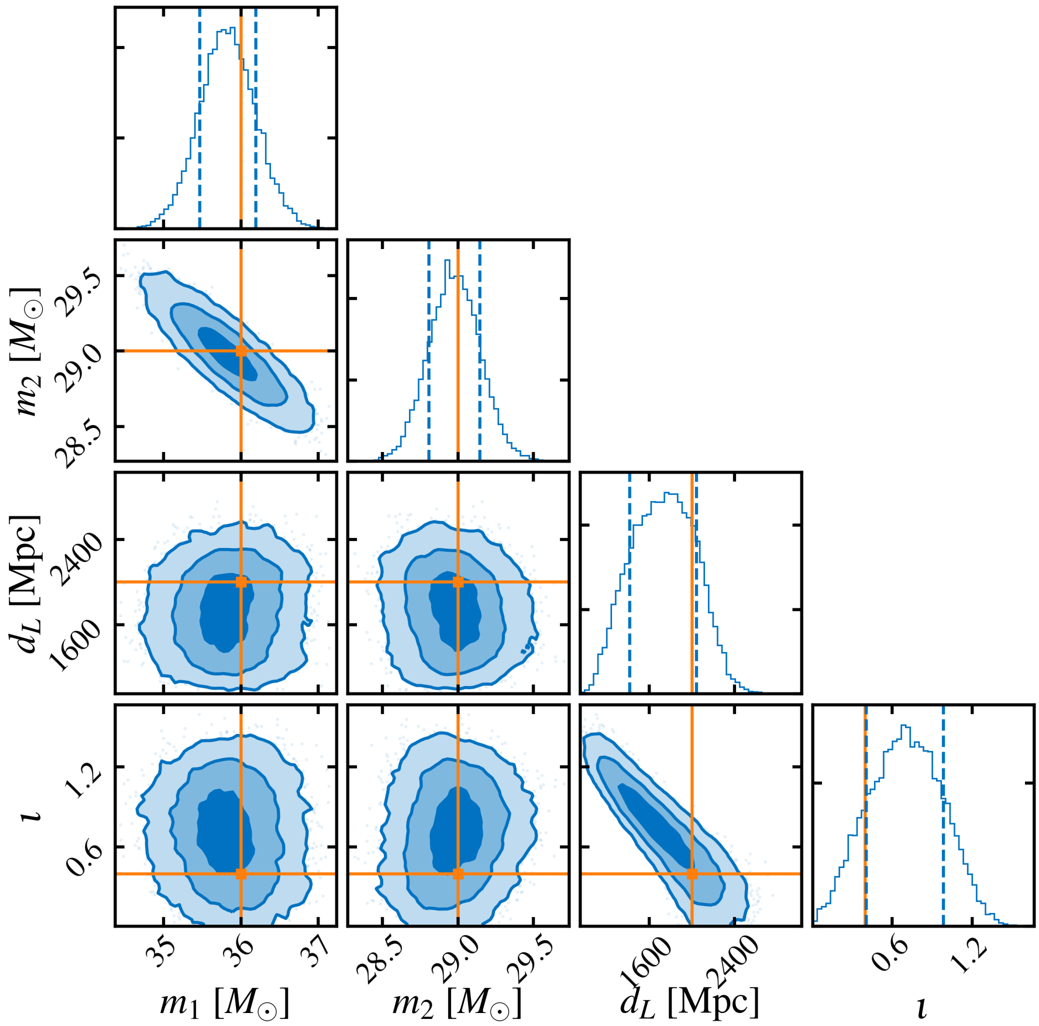

FIGURE 6.4: Injecting and recovering a binary black hole gravitational-wave signal with BILBY. We inject a signal into a three-detector network of LIGO-Livingston, LIGO-Hanford, and Virgo and perform parameter estimation. The posterior distributions are shown in blue and the injected values in orange. To speed up the simulation we only search over the two black hole masses m_1 and m_2, the luminosity distance d_L, and the inclination angle ι.

this is the estimate on Λ assuming a $1.4\,\mathrm{M}_\odot$ neutron star, and the uncertainty is the 90% credible interval.

BILBY can be used to study neutron star coalescences in both real and simulated data. We inject a binary neutron star signal using the `TaylorF2` waveform approximant into a three-detector network of the two LIGO detectors and Virgo, all operating at design sensitivity[5]. Our injected signal is an $m_1 = 1.3\,\mathrm{M}_\odot$, $m_2 = 1.5\,\mathrm{M}_\odot$ binary at $d_L = 50$ Mpc with dimensionless spin parameters $a_{1,2} = 0.02$, and tidal deformabilities $\Lambda_{1,2} = 400$. Setting up such a system in BILBY is equivalent to doing the binary black hole injection

[5] This example is found in the BILBY git repository at https://git.ligo.org/Monash/bilby/blob/master/examples/injection_examples/binary_neutron_star_example.py.

variable	unit	prior	minimum	maximum
$m_{1,2}$	$M_\odot$	uniform	1	2
$a_{1,2}$	-	uniform	-0.05	0.05
$\Lambda_{1,2}$	-	uniform	0	3000
d_L	Mpc	comoving	10	500
ra	rad.	uniform	0	2π
dec	rad.	cos	$-\pi/2$	$\pi/2$
ι	rad.	sin	0	π
ψ	rad.	uniform	0	π
ϕ_c	rad.	uniform	0	2π

TABLE 6.2: Default binary neutron star priors. $\Lambda_{1,2}$ are the tidal deformability parameters of the primary and secondary neutron star defined in Eq. 6.2. For other variable definitions, see Tab. 6.1. Note our commonly-used waveform approximant does not allow misaligned neutron star spins, implying we do not require priors on those spin parameters.

study of Sec. 6.4.2, except we call the `lal_binary_neutron_star` source function, which requires the additional $\Lambda_{1,2}$ arguments. We also have specific binary neutron star priors; the default set can be called using

```
>>> priors = bilby.gw.prior.BNSPriorDict()
```

The standard set of binary neutron star priors are shown in Tab. 6.2. In this example we use the `Dynesty` sampler (Speagle, 2019).

The tidal deformability parameters Λ_1 and Λ_2 are known to be highly correlated. The terms that appear explicitly due to the tidal corrections in the phase evolution are instead $\tilde{\Lambda}$ and $\delta\tilde{\Lambda}$ (Flanagan & Hinderer, 2008) (for definitions of these parameters, see Eqs. (14) and (15) of Lackey & Wade, 2015). We therefore sample in $\tilde{\Lambda}$ and $\delta\tilde{\Lambda}$, instead of Λ_1 and Λ_2. Although we sample in all binary neutron star parameters, we show only the two-dimensional marginalized posterior distribution for $\tilde{\Lambda}$ and $\delta\tilde{\Lambda}$ in Fig. 6.5. The corresponding injected values of $\tilde{\Lambda}$ and $\delta\tilde{\Lambda}$ are shown as the orange vertical and horizontal lines, respectively.

6.4.4 Implementing New Waveforms

The preceding subsections have only given a flavour of what can be achieved with BILBY for compact binary coalescences. It is trivial to implement more complex signal models that include, for example, higher order modes, eccentricity, gravitational-wave memory, non-standard polarizations. Examples showing different signal models are included in the `git` repository (Ashton et al., 2018b). BILBY has already been used in one such application: testing how well the orbital eccentricity of binary black hole systems can be measured with Advanced LIGO and Advanced Virgo (Lower et al., 2018). An example script reproducing those results can be found in the `git` repository (Ashton et al., 2018b).

If a signal model exists in the LAL software (*Lalsuite*), then calling that signal model and defining which parameters to include in the sampler is

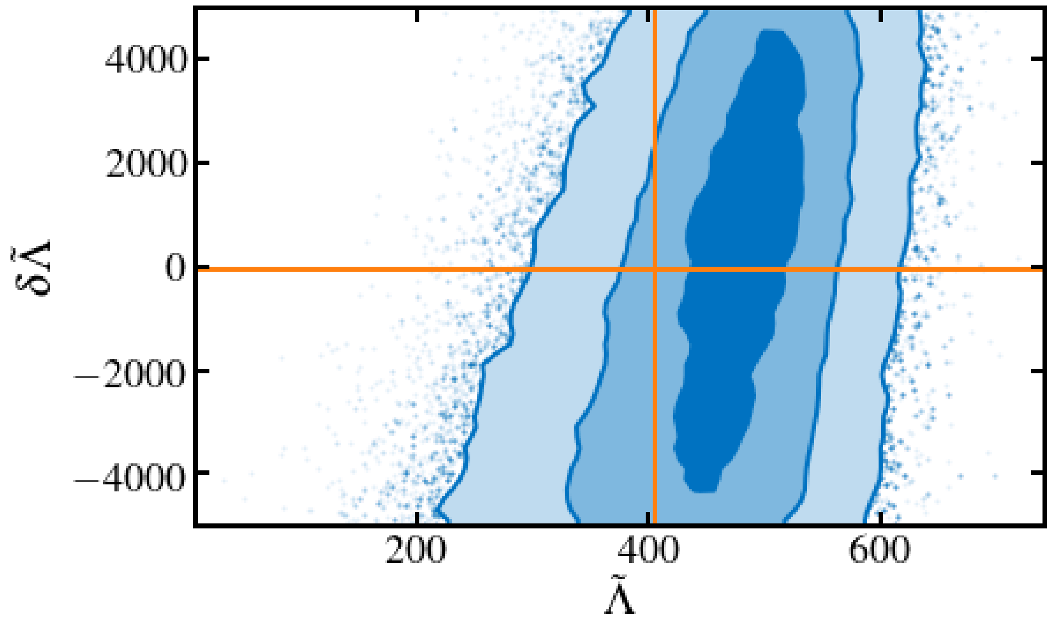

FIGURE 6.5: Injecting and recovering a binary neutron star gravitational-wave signal with BILBY. We inject a signal into the three-detector network, and show here only the marginalized two-dimensional posterior on the two tidal deformability parameters (blue) with the injected values shown in orange.

as simple as the above examples. In Sec. 6.5 we also show how to include a user-defined source model. Moreover, one is free to define and sample models in either the time or frequency domain. We include examples for both cases in the `git` repository. The latter case of using a time-domain source model requires doing little more than selecting the argument `time_domain_source_model` in the `WaveformGenerator`, rather than selecting `frequency_domain_source_model`.

Of course, one may also want to set up the injection and the sampler using two different waveform models, for example to inject a numerical relativity signal into Monte Carlo data and recover it with a waveform approximant (see also Sec. 6.5.1). This is possible by simply instantiating two `WaveformGenerators`, injecting with one and passing the other to the `likelihood`.

6.4.5 Adding detectors to the network

The full network of ground-based gravitational-wave interferometers will soon consist of the two LIGO detectors in the US, Virgo, LIGO-India (Iyer et al., 2011) and the KAGRA detector in Japan (Aso, 2013), all of which are implemented in BILBY. A gravitational-wave interferometer is specified by its geographic coordinates, orientation, and noise power spectral density. By default, BILBY includes descriptions of current detectors including LIGO, Virgo, and KAGRA, as well as proposed future detectors, A+ (Miller et al., 2015), Cosmic Explorer (Abbott, 2017c), and the Einstein Telescope (Punturo et al., 2010). It is also possible to define new detectors, which is useful for developing the science case for proposals and to optimize the design and placement of new detectors. Among other things, this can be used in developing the science-case for interferometer design and placement.

BILBY provides a common interface to define detectors by their geometry, location, and frequency response. By way of example, we place a new four-kilometer-arm interferometer in the Shire of Gingin, located outside of Perth, Australia; the current location of the Australian International Gravitational Observatory (AIGO). We assume a futuristic network configuration of the Australian Observatory together with the two LIGO detectors in Hanford and Livingston, all operating at A+ sensitivity (Miller et al., 2015). We generate A+ power spectral densities in the same script used to run BILBY by using the PYGWINC software (LIGO Laboratory, 2018), which creates an array containing the frequency and noise power spectral density[6] (one could equally use more sophisticated software such as FINESSE (Brown & Freise, 2014) to create more detailed interferometer sensitivity curves). We then create a new `Interferometer` object using `bilby.gw.detector.Interferometer()`, which takes numerous arguments including the position and orientation of the detector, minimum and maximum frequencies, and the power or amplitude noise spectral density. The noise spectral density can be passed as an ascii file containing the frequency and spectral noise density. With the new detector defined, one can again calculate a noise realisation and signal injection in a manner similar to what is done in Sec. 6.4.

In this example we inject a GW150914-like binary black hole inspiral signal at a luminosity distance of d_L=4 Gpc, and recover the masses, sky location, luminosity distance and inclination angle of the system. In this example we use the `Nestle` sampler (Barbary, 2015). Figure 6.6 shows the two-dimensional marginalized posterior for the sky-location uncertainty when including (blue) and not including (orange) the Australian detector in Gingin. In this instance, the sky localisation uncertainty decreases by approximately a factor four when including the third detector.

While this example includes three detectors, it is straightforward to extend this analysis to an arbitrary detector network. The likelihood evaluation simply loops over the number of detectors passed to it and multiplies the likelihood for each detector to get a combined likelihood for each point in the parameter space.

6.5 Alternative signal models

Section 6.4 focuses on compact binary coalescences. However, the BILBY `gw` package enables parameter estimation for any type of signal for which a signal model can be defined. In this section, we show two illustrative examples: the injection and recovery of a core-collapse supernovae signal, and a much-simplified model of a hypermassive neutron star following a binary neutron

[6] This example is found in the BILBY `git` repository at https://git.ligo.org/Monash/bilby/blob/master/examples/injection_examples/Australian_detector.py.

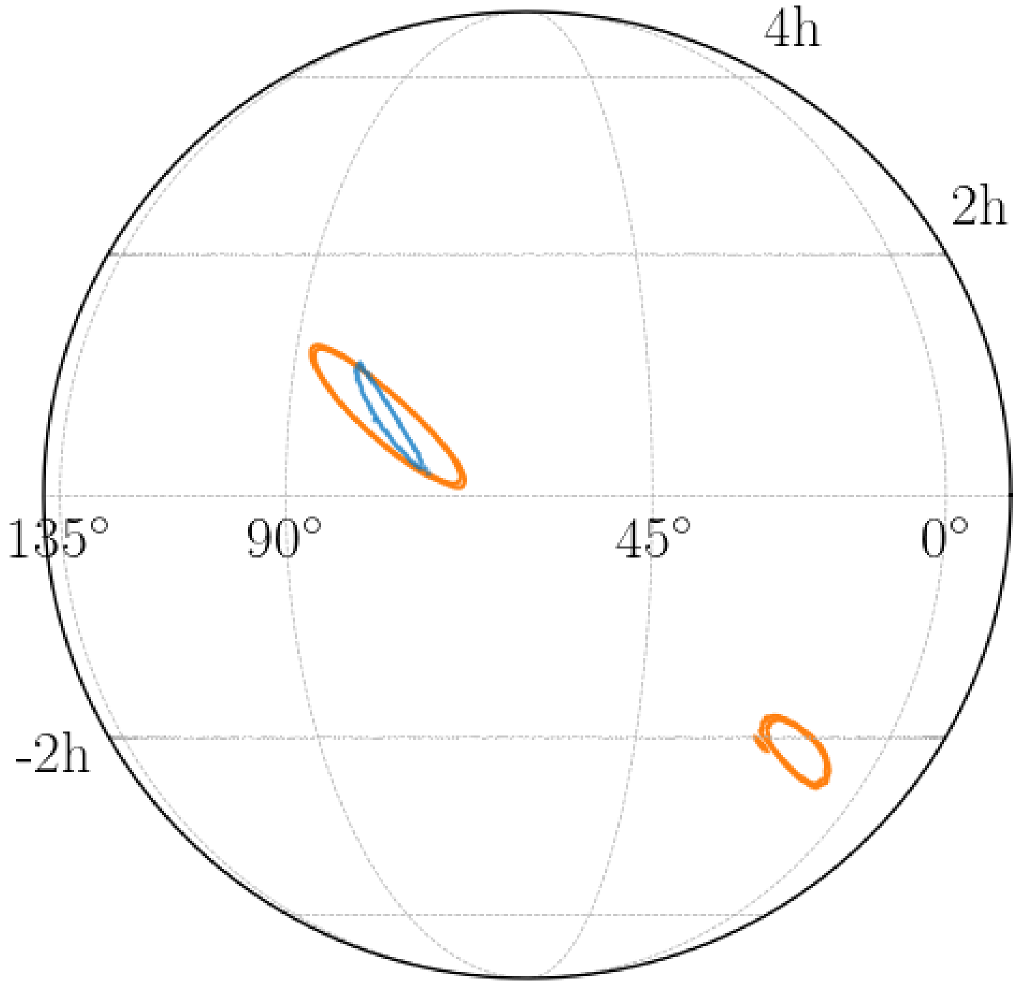

FIGURE 6.6: Sky location uncertainty when including a gravitational-wave detector in Gingin, Australia. Shown are the sky localisations (marginalized two-dimensional posterior distributions) for an injected binary black hole signal using a two-detector network of gravitational-wave interferometers Hanford and Livingston (orange) and a three-detector network that also includes the Australian detector (blue).

star merger. The former example highlights two key pieces of infrastructure; the ability to inject numerical relativity signals, and to develop ones own source model that is not built into BILBY. The latter example highlights the use of a different likelihood function that only uses the amplitude of the signal, and throws away the phase information.

6.5.1 Supernovae

Gravitational-wave signals from core-collapse supernovae are complicated and not well understood in terms of their specific phase evolution. Numerous techniques have been developed to deal with both detection and parameter estimation. One such method for the latter problem involves principal component analysis (Logue et al., 2012; Powell et al., 2016; Powell, Szczepanczyk & Heng, 2017), where the signal is reconstructed using a weighted sum of orthonormal basis vectors. In this example, we inject a gravitational-wave signal from a numerical relativity simulation (Müller, Janka & Wongwathanarat, 2012) and recover the principal components using BILBY[7].

The injection is performed by defining a new signal class that, in this case, simply reads in an ascii text file containing the gravitational-wave strain time series. The injection is then performed in a way akin to the binary black hole and binary neutron star examples in Sec. 6.4. We inject signal L15 from Müller, Janka & Wongwathanarat, 2012, which comes from a three-dimensional simulation of a non-rotating core-collapse supernova with a 15 $M_\odot$ progenitor star. The signal is injected at a distance of 5 kpc in the direction of the galactic center. The amplitude spectral density of the injected signal is shown in Fig. 6.7 as the orange trace.

The signal is reconstructed using principal component analysis, such that the strain is expressed as

$$\tilde{h}(f) = A \sum_{j=1}^{k} \beta_j U_j(f), \tag{6.3}$$

where A is an amplitude factor, β_j and U_j are the complex principal component amplitudes and vectors, respectively. Equation (6.3) is implemented into BILBY as another new signal model that takes the β_j coefficients, luminosity distance (which is a proxy for A), and sky location as inputs. Priors for each of the new parameters are established in the same way as the example with the mass in Sec. 6.4.2. In this case, we set $k = 5$ and use uniform priors between -1 and 1 for each of the β_j's.

Figure 6.7 shows the injected (orange) and recovered (blue) gravitational-wave signal in the frequency domain. The dark blue curve shows the maximum likelihood curve, and the shaded blue region is a superposition of many reconstructed waveforms from the posterior samples.

6.5.2 Neutron star post-merger remnant

There are a number of physical scenarios that can occur following the merger of two neutron stars, including the existence of short- or long-lived neutron star remnants. In the early phases post-merger ($\lesssim$1 s), these neutron stars are highly dynamic, and can emit significant gravitational radiation potentially observable by Advanced LIGO and Virgo at design sensitivity out to $\sim$ 50 Mpc e.g., Clark et al., 2014, and references therein. While the ultimate fate of binary merger GW170817 is unknown, no gravitational waves from a post-merger remnant were found (Abbott, 2017k; Abbott, 2019c), which is not surprising given the interferometers were not operating at design sensitivity and the distances involved.

[7]This example is found in the BILBY git repository at `https://git.ligo.org/Monash/bilby/blob/master/examples/supernova_example/supernova_example.py`.

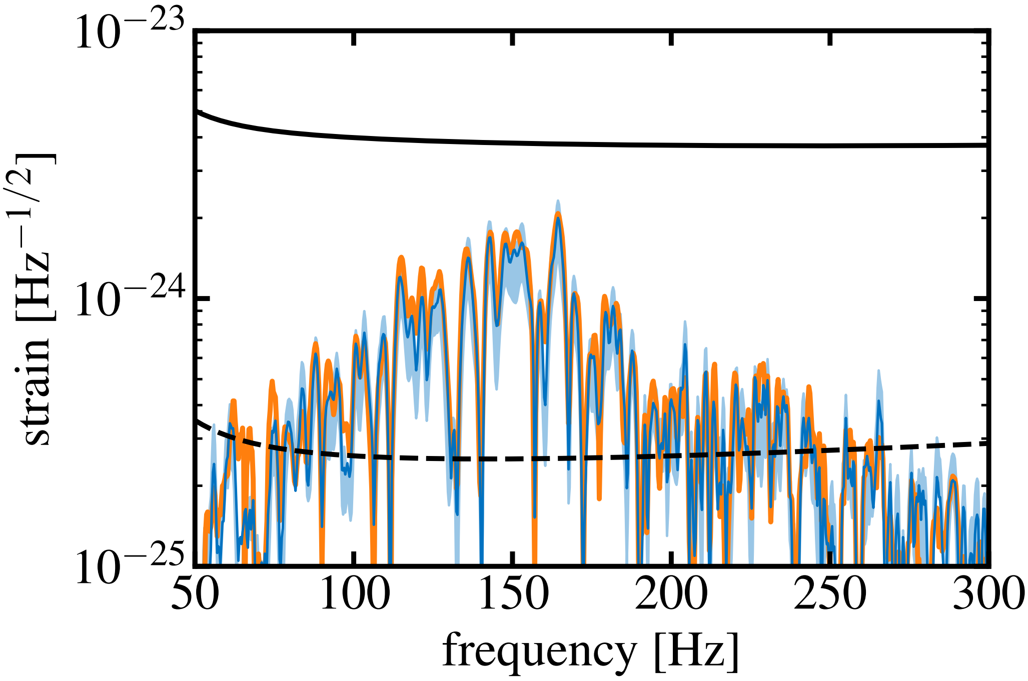

FIGURE 6.7: Parameter estimation reconstruction of a numerical relativity supernova signal. A numerical relativity supernovae signal (orange) is injected into a three-detector network of the two Advanced LIGO detectors and Advanced Virgo, all operating at design sensitivity. The maximum likelihood reconstruction of the signal is shown in dark blue, and the blue light band shows the superposition of many reconstructed waveforms from the posterior samples.

Providing the sensitivity of gravitational-wave interferometers continues to increase, it is possible a gravitational-wave signal from a post-merger remnant could be detected in the relative near future. Such a detection would provide an excellent opportunity to understand the nuclear equation of state of matter at extreme densities, as well as the rich physics of these exotic objects (e.g., Shibata & Taniguchi, 2006; Baiotti, Giacomazzo & Rezzolla, 2008; Read, 2013). Parameter inference of such short-lived signals is in its infancy see, e.g., Chatziioannou et al., 2017, largely due to the paucity of reliable waveforms (Clark et al., 2016; Easter et al., 2019). This is an ongoing challenge due to the expensive nature of numerical relativity simulations and the complex physics that must be included in such simulations.

Simple models that provide approximate gravitational-wave signals fit to a handful of numerical relativity waveforms exist (Messenger et al., 2014; Bose et al., 2018; Easter et al., 2019), which may eventually be used for full parameter inference. The phase evolution of such numerical relativity simulations is rapid, and very difficult to model (Messenger et al., 2014; Easter et al., 2019). However, it is the frequency content of the signal that carries information about the equation of state and the physics of the remnant e.g., Takami, Rezzolla & Baiotti, 2015, and references therein. It is therefore possible that parameter-estimation algorithms may require one to throw away information about the phase, and only keep amplitude spectral content. Such a process requires a different likelihood function than the one that has been used to this point. This therefore provides good motivation for showing how to include a different likelihood function in BILBY code.

We implement a power-spectral density ("burst") likelihood

$$\ln \mathcal{L}(|d| \mid \theta) = \sum_{i=1}^{N} \left[\ln I_0 \left(\frac{|\tilde{h}_i(\theta)||\tilde{d}_i|}{S_n(f_i)} \right) - \frac{|\tilde{h}_i|^2 + |\tilde{d}_i|^2}{2S_n(f_i)} + \ln |\tilde{h}_i(\theta)| - \ln S_n(f_i) \right], \tag{6.4}$$

where I_0 is the zeroth-order modified Bessel function of the first kind. This requires setting up a new Likelihood class, that contains a `log_likelihood` function that reads in the frequency array, noise spectral density and waveform model, and outputs a single likelihood evaluation. Having defined a new likelihood function, one calls the remaining functions in the usual way; the likelihood function is instantiated and passed to the `run_sampler()` command.

We inject a double-peaked Gaussian, shown in Fig. 6.8 as the solid orange curve. We recover this signal using the same model (with a constant noise spectral density), where we use uniform priors for the amplitudes, widths and frequencies of each of the peaks. Figure 6.8 shows the waveform reconstruction for each of the posterior samples, which can be seen to cover the injected signal.

6.6 Population Inference: hyperparameterizations

Individual detections of binary coalescences can provide stunning insights into various physical and astrophysical questions. Increased detector sensitivities imply significantly more events will be detected, enabling statements to also be made about ensemble properties of populations e.g., Abbott, 2016b; Talbot & Thrane, 2018; Wysocki et al., 2018; Smith & Thrane, 2018; Farr, Holz & Farr, 2018; Taylor & Gerosa, 2018; Roulet & Zaldarriaga, 2019, and references therein. Extracting information from a population of events is performed using hierarchical Bayesian inference where the population is described by a set of hyper-parameters, Λ. BILBY has built-in support for calculating Λ from multiple sets of posterior samples from individual events.

BILBY implements the conventional method whereby the posterior samples θ_i^j for each event j are re-weighted according to the ratio of the population model prior $\pi(\theta|\Lambda)$ and the sampling prior $\pi(\theta)$ to obtain the hyper-parameter likelihood

$$\mathcal{L}(h|\Lambda) = \prod_{j}^{N} \frac{\mathcal{Z}_j}{n_j} \sum_{i}^{n_j} \frac{\pi(\theta_i^j|\Lambda)}{\pi(\theta_i^j)}. \tag{6.5}$$

Here, $\mathcal{Z}_j$ is the Bayesian evidence for the data given the original model and n_j is the number of posterior samples in the j^{th} event.

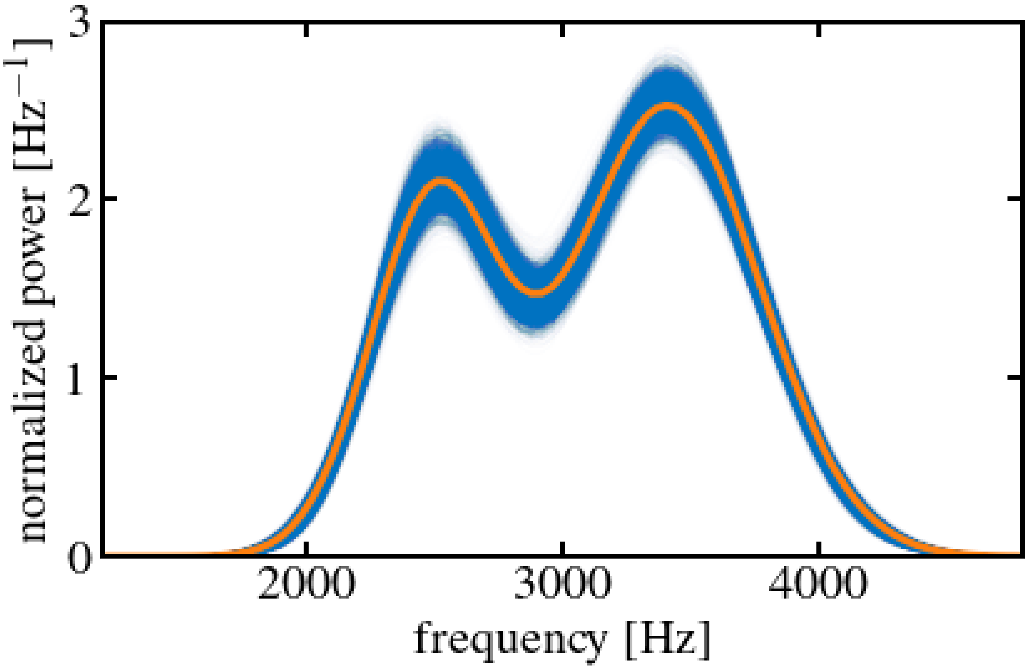

FIGURE 6.8: A proxy post-merger gravitational-wave signal from a short-lived neutron star showing the implementation of a different likelihood function in BILBY. The orange curve is an injected, double-peaked Gaussian signal injected into a constant noise realisation. The blue band shows the waveform reconstructions from the posterior samples using a power-spectrum likelihood function; i.e., one that only uses the amplitude of the signal and ignores the phase.

The BILBY implementation requires the user to define $\pi(\theta|\Lambda)$ and $\pi(\theta)$ which, along with the set of posterior samples θ_i^j, are passed to the `HyperparameterLikelihood` in BILBY's `hyper` package. The hyperparameter priors are then set up in the usual way, and passed to the standard `run_sampler` function.

As a demonstration[8] of this method we reproduce results Talbot & Thrane, 2018 recovering parameters describing a postulated excess of black holes due to pulsational pair-instability supernovae (PPSN) (Heger et al., 2003; Woosley & Heger, 2015). The posterior distribution for the hyperparameters determining the abundance and characteristic mass of black holes formed through this mechanism are shown in Fig. 6.9. The hyperparameter λ is the fraction of binaries where the more massive black hole formed through PPSN, $\mu_{\rm pp}$ is the typical mass of these black holes and $\sigma_{\rm pp}$ determines the width of the "PPSN graveyard".

This model contains seven additional hyperparameters describing the remainder of the distribution of black hole masses that we hold fixed for the purposes of this example. Additional hyperparameters may be added straightforwardly.

[8]This example is found in the BILBY git repository at `https://git.ligo.org/Monash/bilby/blob/master/examples/other_examples/hyper_parameter_example.py`.

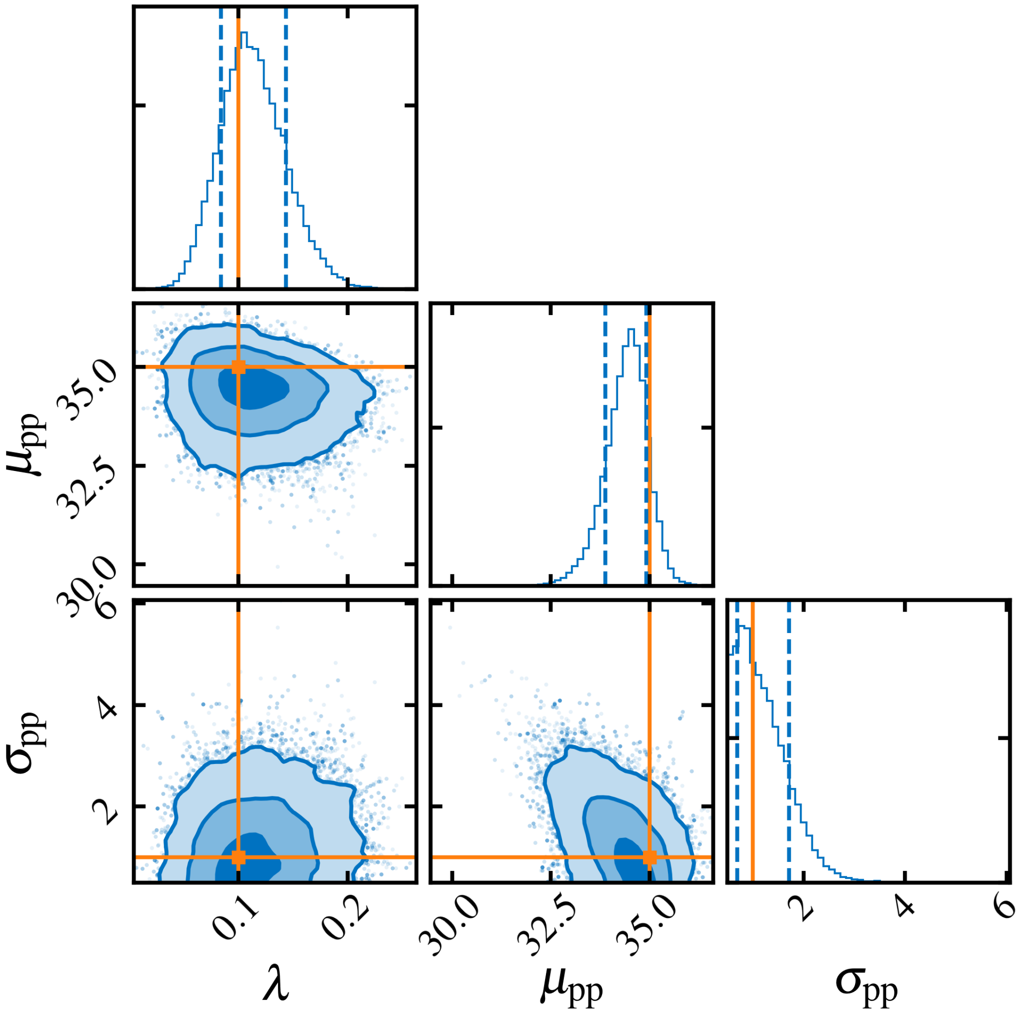

FIGURE 6.9: Population modelling with BILBY hierarchical Bayesian inference module. We show the recovery of parameters describing part of the mass distribution of binary black holes using the model described in Talbot & Thrane, 2018. The population parameters are drawn from values shown in orange, and the posterior distributions for the hyperparameters shown in blue. Here, λ is the fraction of binaries where the more massive black hole formed through pulsational pair-instability supernovae, μ_{pp} and σ_{pp} are the typical mass of these black holes and the width of the "PPSN graveyard", respectively.

6.7 Analysis of arbitrary data: an example

BILBY is more than a tool for gravitational-wave astronomy; it can also be used as a generic and versatile inference package. In the documentation examples, we demonstrate how BILBY can be applied to generic time-domain data from radioactive decay processes. Furthermore, BILBY is currently being used to analyse radio and x-ray data from neutron stars, and to study multi-messenger signals associated with binary neutron star mergers. Here we show an example that calculates posterior distributions for one of the letters in the BILBY logo.

We import an image file containing the letter, map this to an x-y coordinate system and sample in both dimensions with likelihood

$$\ln \mathcal{L} \propto \frac{-1}{xy}, \tag{6.6}$$

assuming uniform priors on both variables. Figure 6.10 shows the posterior distribution for the "B" in the BILBY logo. All letters are shown in Fig 6.11, where the axis labels have been removed. The code for making this plot, and all other posterior distributions in the logo, are available with the `git` repository (Ashton et al., 2018b) in `sample_logo.py`.

6.8 Conclusion

Gravitational-wave astronomy is fast becoming a data-rich field. With the significantly increased activity in the field, there is a developing need for robust, easy-to-use inference software that is also modular and adaptable. We present BILBY: the Bayesian inference library for gravitational-wave astronomy. BILBY is open-source software that can be used to perform Bayesian inference. It is easily applied to data from LIGO/Virgo, including open data available from the Gravitational Wave Open Science Center. We access and manipulate LIGO data using `GWPy` (Macleod et al., 2018). Alternatively, BILBY may be used to study simulated data. BILBY can also be used to perform hierarchical Bayesian inference for population studies.

We present examples highlighting BILBY's functionality and usability, including examples using open data from the first gravitational-wave detection GW150914. Only five lines of code are required to reconstruct the astrophysical parameters of GW150914. One can redo the analysis using different priors, alternative waveform models, and/or a different sampling method with only modest changes. We show how to inject binary black hole and binary neutron star signals into Monte Carlo noise. We show how to define new gravitational-wave detectors.

We emphasise that BILBY is a front-end system that provides a unified interface to a variety of samplers, which are a primary workhorse of Bayesian inference. While numerous off-the-shelf samplers are implemented (see Sec. 6.3.2), to the best of our knowledge there is no universal sampling solution to gravitational-wave parameter estimation problems. BILBY is

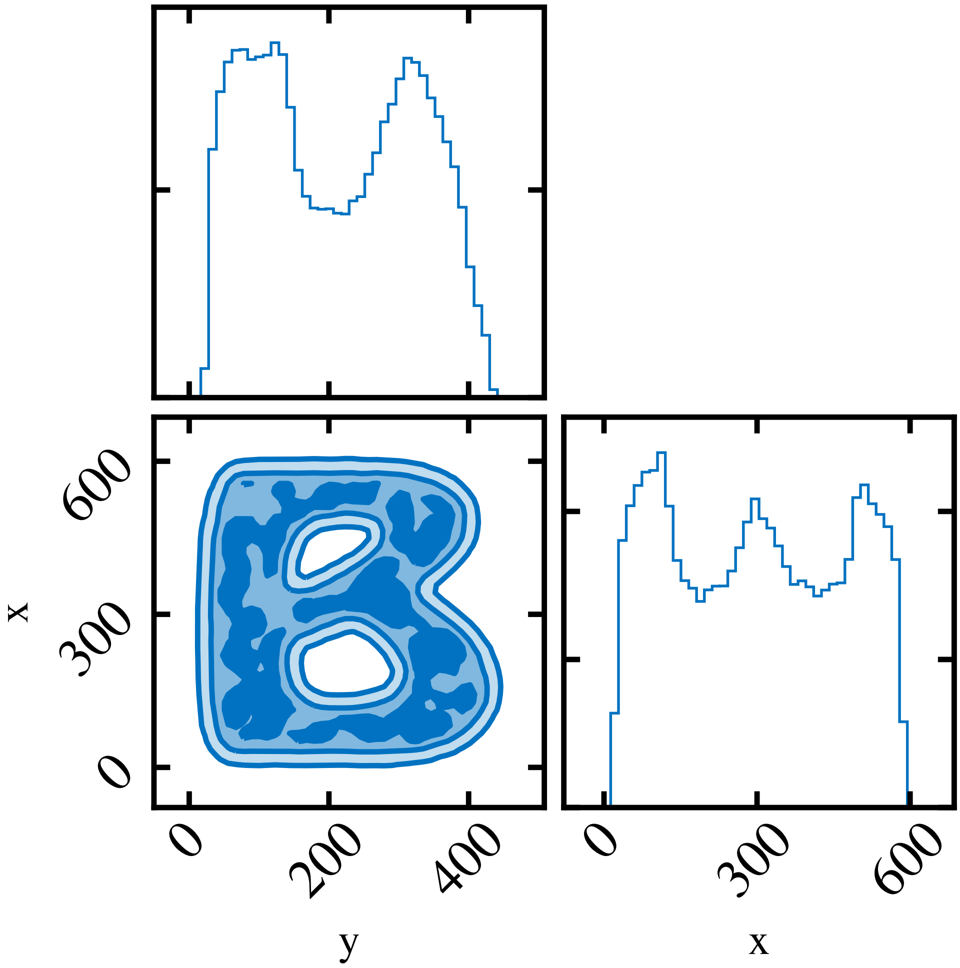

FIGURE 6.10: The 'B' from the BILBY logo, generated using the BILBY package; see Sec. 6.7

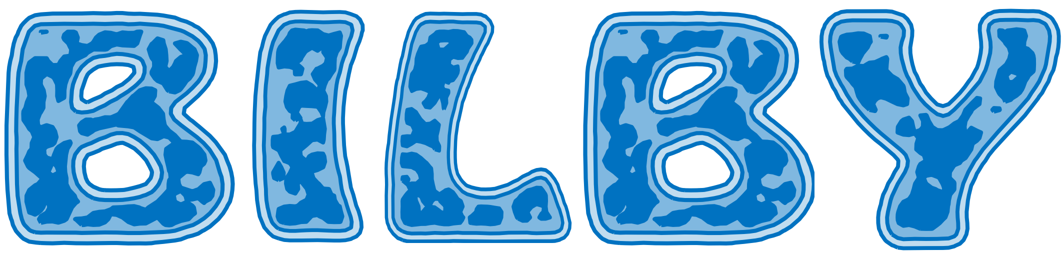

FIGURE 6.11: All letters from the BILBY logo, generated using the BILBY package; see Sec. 6.7

therefore only as good as the implemented samplers; initial studies show that `CPNest` (Veitch, 2017), `Dynesty` (Speagle, 2019), and `emcee` (Foreman-Mackey et al., 2013; Vousden, Farr & Mandel, 2016) sample the extrinsic parameters of binary coalescences more accurately than `Nestle` (Barbary, 2015) and `pyMultiNest` (Buchner, 2014). A systematic comparison of all off-the-shelf and boutique samplers is currently underway using BILBY.

BILBY is designed so as to be applicable to arbitrary signal models, not just compact binary coalescences. To this end we show two examples: one of an injected numerical relativity supernova waveform that we reconstruct using principal component analysis, and another using a proxy for a neutron star post-merger waveform. The former example highlights how one can include their own signal models to perform both injections and signal recoveries, while the latter example demonstrates the ability to add a likelihood function that is different from the standard gravitational-wave transient likelihood.

Addendum

Since the publication of this paper additional compact binary coalescences have been observed and significant progress has been made in theoretical and observational analysis. The reader is directed to Chapters 1 and 8 for an overview of the field at the time of writing. I note that due to a KDE artefact in the plotting routines Figure 6.1 appears to show support for primary mass > secondary mass, which is forbidden by the prior.

Chapter 7

Parallelized Inference for Gravitational-Wave Astronomy

Published as:
Talbot, C., Smith, R., Thrane, E. & Poole, G. B., Phys. Rev. D, 98, 064031 (2019).

Abstract

Bayesian inference is the workhorse of gravitational-wave astronomy, for example, determining the mass and spins of merging black holes, revealing the neutron star equation of state, and unveiling the population properties of compact binaries. The science enabled by these inferences comes with a computational cost that can limit the questions we are able to answer. This cost is expected to grow. As detectors improve, the detection rate will go up, allowing less time to analyze each event. Improvement in low-frequency sensitivity will yield longer signals, increasing the number of computations per event. The growing number of entries in the transient catalog will drive up the cost of population studies. While Bayesian inference calculations are not entirely parallelizable, key components are embarrassingly parallel: calculating the gravitational waveform and evaluating the likelihood function. Graphical processor units (GPUs) are adept at such parallel calculations. We report on progress porting gravitational-wave inference calculations to GPUs. Using a single code—which takes advantage of GPU architecture if it is available—we compare computation times using modern GPUs (NVIDIA P100) and CPUs (Intel Gold 6140). We demonstrate speed-ups of $\sim 50\times$ for compact binary coalescence gravitational waveform generation and likelihood evaluation, and more than $100\times$ for population inference within the lifetime of current detectors. Further improvement is likely with continued development. Our python-based code is publicly available and can be used without familiarity with the parallel computing platform, CUDA.

7.1 Introduction

In the first two observing runs of Advanced LIGO/Virgo, ten binary black hole mergers were detected along with one binary neutron star inspiral (Abbott, 2019b). These observations allowed us to measure the Hubble parameter (Abbott, 2017a), to study matter at extreme densities (Abbott, 2018d), and to probe the underlying distribution of black holes in merging binaries (Abbott, 2019a). Within the lifetime of advanced detectors, we conservatively estimate that hundreds of such observations will be made given inferred merger rates and projected sensitivity (Abbott, 2016j).

Compact binary coalescences are analyzed with Bayesian inference (see, e.g. Gelman et al., 2013 for a general introduction or Thrane & Talbot, 2019 for applications to gravitational waves.). We distinguish between two kinds. We refer to inferring the properties (e.g., the masses, spins, and location) of individual binaries as *single-event inference*. Hierarchical Bayesian inference is then used to infer the ensemble properties (e.g. the shape of the binary black hole mass distribution) of the observed binaries in *population inference*. These are typically performed with stochastic sampling algorithms such as Markov Chain Monte Carlo (MCMC) (Metropolis et al., 1953; Hastings, 1970) or Nested Sampling (Skilling, 2004). These algorithms generate samples from the posterior distribution and possibly a Bayesian evidence which can be used for model selection.

Both single-event inference and population inference require the computation of likelihood functions consisting of many independent operations. For single-event inference, the number of operations per likelihood evaluation is determined by the length of the signals being analyzed, see, Eq. 7.1. As the low-frequency sensitivity of detectors increases, binaries will spend longer in our sensitive frequency range, leading to fast-increasing computational demands. For population inference, the number of operations required per likelihood calculation is proportional to the number of binaries in the population, see, Eq. 7.3. The growing number of observations and the growing duration of the longest signals in the catalog require improved speed for inference algorithms.

Most previous methods for accelerating inference for compact binary coalescences have sped up calculations by reducing the amount of data required to represent the gravitational-wave signal, thereby reducing the number of operations required to evaluate the likelihood, e.g., reduced order methods (Pürrer, 2014; Canizares et al., 2015; Smith et al., 2016), multibanding (Vinciguerra, Veitch & Mandel, 2017), and relative binning (Zackay, Dai & Venumadhav, 2018). Another approach, which we investigate here, is to parallelize the most time-consuming calculations in the likelihood evaluation. While it is difficult to parallelize the actual sampling algorithm, there are embarrassingly parallel calculations within the likelihoods. In this paper, we explore how astrophysical inference can be accelerated by executing parallelizable calculations on graphical processor units.

While we focus here on inference using stochastic samplers, e.g., (Veitch et al., 2015; Biwer et al., 2019; Ashton et al., 2019), it bears mentioning that there are alternative inference schemes, which also face computational challenges, e.g., iterative fitting (Pankow et al., 2015; Lange, O'Shaughnessy & Rizzo, 2018) is one such alternative. This method evaluates far fewer gravitational waveforms, which can significantly reduce computation times for inference when the waveform evaluation is very time-consuming. Recently, it has been shown that this algorithm can also be significantly accelerated by parallelization (Wysocki et al., 2019).

Modern computation is mostly performed on either a central processing unit (CPU) or a graphics processing unit (GPU). CPUs consist of a relatively small number of cores optimized to perform all the tasks necessary for a computer in serial. GPUs, on the other hand, consist of hundreds or thousands of cores, enabling evaluation of many numerical operations simultaneously. This makes them ideal for embarrassingly parallel operations such as manipulating large arrays of numbers. Low-level GPU programming is carried out using the parallel computing platform, CUDA. In this work, we take advantage of the library, CUPY (Okuta et al., 2017).

We present python packages with parallelized versions of the likelihoods necessary for performing both single-event and population inference. The code is designed to run on either a CPU or a GPU, depending on the available hardware. Our GPU-compatible code for single-event inference is available at `github.com/colmtalbot/gpucbc`, our CUDA compatible version of IMRPHENOMPV2 at `https://github.com/ADACS-Australia/ADACS-SS18A-RSmith`. Our GPU population inference code GWPOPULATION is available from the python package manager PYPI and from git at `github.com/colmtalbot/gwpopulation`.

Using our GPU-accelerated code, we investigate the speed-up achieved carrying out gravitational-wave inference calculations using GPUs versus traditional CPUs. We carry out a benchmarking study in which we compare inference code using GPUs to identical code running on CPUs. We compare the runtimes for various tasks including: (i) evaluating a gravitational waveform, (ii) evaluating the single-event likelihood, and (iii) evaluating population likelihood. To carry out this comparison, we use NVIDIA P100 GPUs and Intel Gold 6140 CPUs available on the OzStar supercomputing cluster.

In Section 7.2 we describe methods for parallelizing the evaluation of gravitational waveforms. In section 7.3 we use these parallelized waveforms to consider the possible acceleration for single-event inference. In section 7.4, we investigate the possible acceleration from parallelizing the population inference likelihood. Finally, we provide a summary of our findings and future work in section 7.5.

7.2 Waveform Acceleration

A key ingredient for single-event inference is a model for the waveform, $\tilde{h}(f,\theta)$. Here, $\tilde{h}$ is the discrete Fourier transform of the gravitational-wave strain time series, f is frequency, and θ is the set of parameters (typically 15-17), which determine the waveform, e.g., the masses, spins, and orientation of the binary. This theoretical waveform is compared with the data every time the likelihood function is evaluated. Many commonly used waveform models can be directly evaluated in the frequency domain and the value at each frequency can be evaluated independently (e.g., Ajith et al., 2011; Hannam et al., 2014; Khan et al., 2016; Khan et al., 2019). This makes the waveform model evaluation embarrassingly parallel.

We design two different codes for performing waveform generation on a GPU. First, we implement a GPU version of a simple, post-Newtonian inspiral-only waveform TaylorF2 (Ajith et al., 2011) directly in python using cupy. Secondly, we convert the C code for a phenomenological waveform IMRPhenomPv2 (Hannam et al., 2014) into CUDA[1]. Both of these methods are then compared to the time required to evaluate the same waveform implemented in C within LALSuite (*Lalsuite*).

Technically, the CUDA version of IMRPhenomPv2 does not adhere to our philosophy of keeping the GPU programming entirely "under the hood,", however, writing custom CUDA kernels allows increased optimization of the GPU code. We consider the effect of pre-allocating a reusable memory "buffer" on the GPU. During parameter estimation waveforms of the same length will be generated many times as a waveform evaluation is required for every likelihood evaluation. Since the waveform always has the same size a predefined spot in memory can be reused for every waveform.

In Fig. 7.1, we plot the speed-up (defined as the CPU computation time divided by the GPU computation time) for both of our waveforms. The dashed line indicates no speedup. In blue we show the comparison of our TaylorF2 waveform using cupy with the C version available in lalsuite (*Lalsuite*). We find that, for signals longer than $\sim 10\,\mathrm{s}$, we achieve faster waveform evaluation. The speed-up scales roughly linearly with signal duration so that the waveforms of duration $100\,\mathrm{s}$ are sped up by a factor of ≈ 10. For signals longer than $\sim 1000\,\mathrm{s}$ the GPU queue saturates and the rate of increase of the speedup decreases.

In orange, we plot the speedup for our CUDA implementation of IMRPhenomPv2 compared to the C implementation of the same waveform model in lalsuite. The solid curve includes the use of a pre-allocated memory buffer, while the dashed curve does not. We find that pre-allocating memory buffer increases the performance when the number of frequencies is $\lesssim 10^5$ and leads to accelerations for all waveforms when the number of frequency bins is larger than 100. This suggests that the performance of our cupy waveform could be similarly improved for short signal durations using more sophisticated waveform allocation.

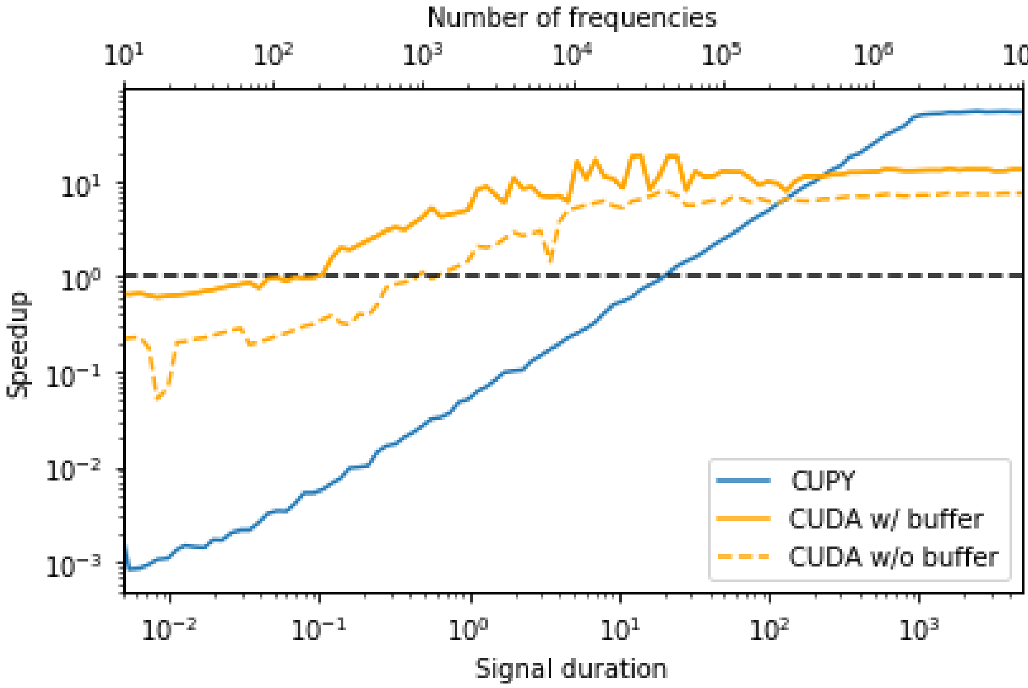

FIGURE 7.1: Speedup comparing our GPU accelerated waveforms with the equivalent versions in LALSuite as a function of the number of frequencies. The corresponding signal duration is shown for comparison assuming a maximum frequency of 2048 Hz. The blue curve shows the comparison of our CUPY implementation of TAYLORF2 with the version available in lalsuite. The GPU version is slower for shorter waveform durations, this is likely because of overheads in memory allocation in CUPY. For binary neutron star inspirals the signal duration from 20 Hz is ~ 100 s, at which signal durations, we see a speedup of $\sim 10\times$. For even longer signals we find an acceleration of up to $80\times$. The orange curves show a comparison of our CUDA implementation of IMRPhenomPv2, the solid curve reuses a pre-allocated memory buffer while the dashed line does not. We see that the memory buffer makes the GPU waveform faster for shorter signals.

7.3 Single-event Likelihood acceleration

The standard likelihood used in single-event inference is

$$\mathcal{L}(d|\theta) = \prod_{i}^{N_{\text{det}}} \prod_{j}^{M} \frac{1}{2\pi S_{ij}} \exp\left(-\frac{2}{T} \frac{|\tilde{d}_{ij} - \tilde{h}_{ij}(\theta)|^2}{S_{ij}} \right), \tag{7.1}$$

where d is the detector strain data, θ are the binary parameters, $\tilde{h}$ is the template for the response of the detector to the gravitational-wave signal as described in Section 7.2, S is the strain power spectral density, the products over i and j are over the M frequency bins and N_{det} detectors in the network respectively.

[1]Specifically, we rewrite in CUDA the function lalsimulation.SimIMRPhenomPFrequencySequence branched from LALSuite at SHA:8cbd1b7187

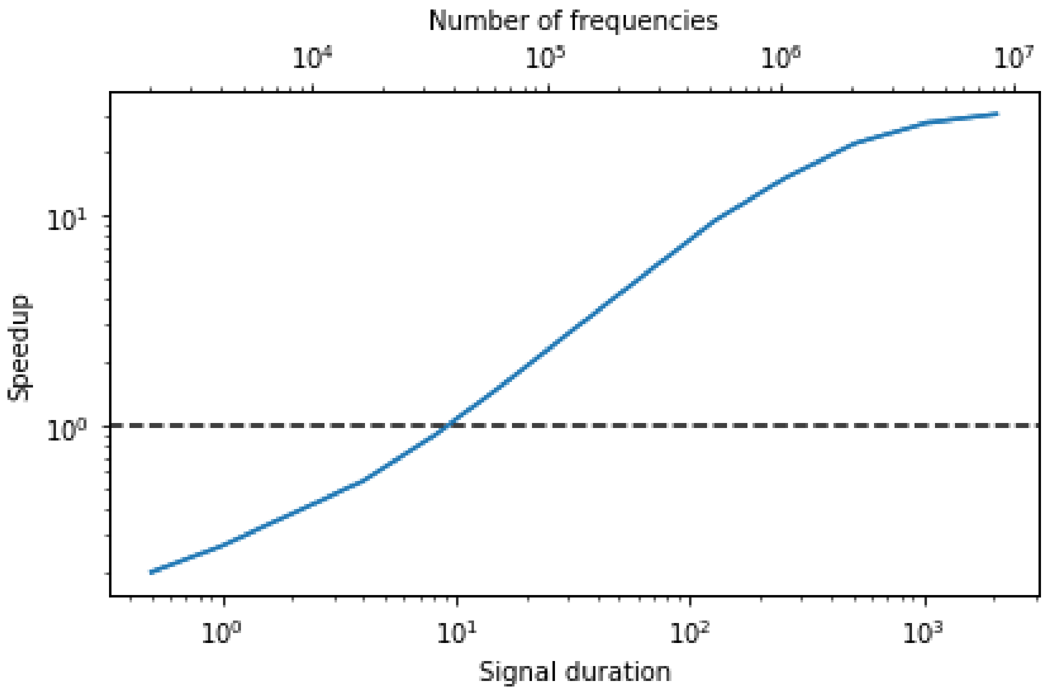

FIGURE 7.2: Speedup of the compact binary coalescence likelihood as a function of the number of frequencies with and without a GPU using our accelerated TaylorF2. The corresponding signal duration is shown for comparison assuming a maximum frequency of 2048 Hz. We see similar performance as for the waveform generation. We find a speedup of an order of magnitude for 128s signals.

Different signals have different durations, and thus, different values for the number of frequency bins M. The more frequency bins, the more embarrassingly parallel calculations to perform, and the more we expect to gain from the use of GPUs. For example, systems with lower masses produce longer duration signals than systems with higher masses, all else equal. In previously published observational papers, the minimum frequency is typically set to 20 Hz for binary black holes, although a larger minimum frequency was used in the analysis of GW170817 (Abbott, 2017i). For high mass binary black holes, this corresponds to $M = 4{,}000 - 32{,}000$ bins. Binary neutron star signals are much longer in duration, requiring $M \approx 10^6$ bins. When current detectors reach design sensitivity, frequencies down to 10 Hz will be used, leading to substantially longer signals and therefore many times more bins. Future detectors are expected to be sensitive to frequencies as low as 5 Hz (Yu et al., 2018).

During each likelihood evaluation, the most expensive step is usually calculating the template. The next most expensive operation is applying a time translation to the frequency domain template in order to align the template with the merger signal. To do this, the frequency-domain waveform is multiplied by a factor of $\exp\left(-2i\pi f \delta t_{\text{det}}\right)$ where t_{det} is different for each detector in the network. This becomes increasingly expensive (with linear scaling) as more detectors are added to the network.

In Fig. 7.2 we show the speedup achieved evaluating the single-event likelihood function using our TAYLORF2 implementation compared to the LALSIMULATION implementation. We find a speedup of an order of magnitude for a 128 s-long signal, the typical duration for binary neutron star analysis with a minimum frequency of 20 Hz. This reduces the total calculation time from a few weeks to a few days.

When the number of frequency bins is relatively small $M < 10^5$, we see a slowdown rather than a speedup. This is due to the same overheads as seen in Section 7.2. Using the CUDA implementation of IMRPHENOMPV2, we would expect to see an acceleration for much shorter signal durations.

In order to demonstrate the improvement achieved analyzing longer-duration signals, we analyze a synthetic binary neutron star inspiral two different ways. First we analyze the final 128 s of inspiral from a frequency of 30 Hz as was done for the LIGO/Virgo analysis of GW170817 (Abbott, 2017i). Then we repeat the same analysis including 512 s worth of inspiral with a minimum frequency of 15 Hz.

The one- and two-dimensional posterior distributions for chirp mass and effective aligned spin $\chi_{\rm eff}$ are shown in Fig. 7.3. Observing more of the early inspiral improves our measurement of both of these parameters. A factor of ~ 2 improvement of the effective spin will facilitate future comparisons with the galactic pulsar population (Zhu et al., 2018).

7.4 Population acceleration

In population inference, we are interested in measuring hyper-parameters, Λ, describing a *population* of binaries (e.g., minimum/maximum black hole mass) rather than the parameters, θ, of each of the individual binaries. The population properties are often described by either phenomenological models (e.g., Vitale et al., 2017b; Talbot & Thrane, 2017; Fishbach & Holz, 2017; Kovetz et al., 2017; Wysocki, 2017; Talbot & Thrane, 2018; Fishbach, Holz & Farr, 2018; Wysocki, Lange & O'Shaughnessy, 2019; Roulet & Zaldarriaga, 2019; Gaebel et al., 2019) or by the results of detailed physical simulations, e.g., population synthesis or N-body dynamical simulations (e.g., Mandel & O'Shaughnessy, 2010; Stevenson, Ohme & Fairhurst, 2015; Belczynski et al., 2016; Gerosa & Berti, 2017; Fishbach, Holz & Farr, 2017; Stevenson, Berry & Mandel, 2017; Zaldarriaga, Kushnir & Kollmeier, 2018; Zevin et al., 2017; Wysocki et al., 2018; Barrett et al., 2018; Qin et al., 2018). In this work, we use the former for examples. However, our methods apply equally to both. The formalism for hierarchical inference including a discussion of selection effects is briefly described below, see, e.g., Farr et al., 2015; Thrane & Talbot, 2019; Mandel, Farr & Gair, 2019 for detailed derivations.

In order to analyze a population of binary black holes, we typically use the following likelihood (Farr et al., 2015),

$$\mathcal{L}_{\rm tot}(\{d_i\}|\Lambda, R) = R^N e^{-RVT(\Lambda)} \prod_i^N \int d\theta_i \mathcal{L}(d_i|\theta_i) p(\theta_i|\Lambda). \tag{7.2}$$

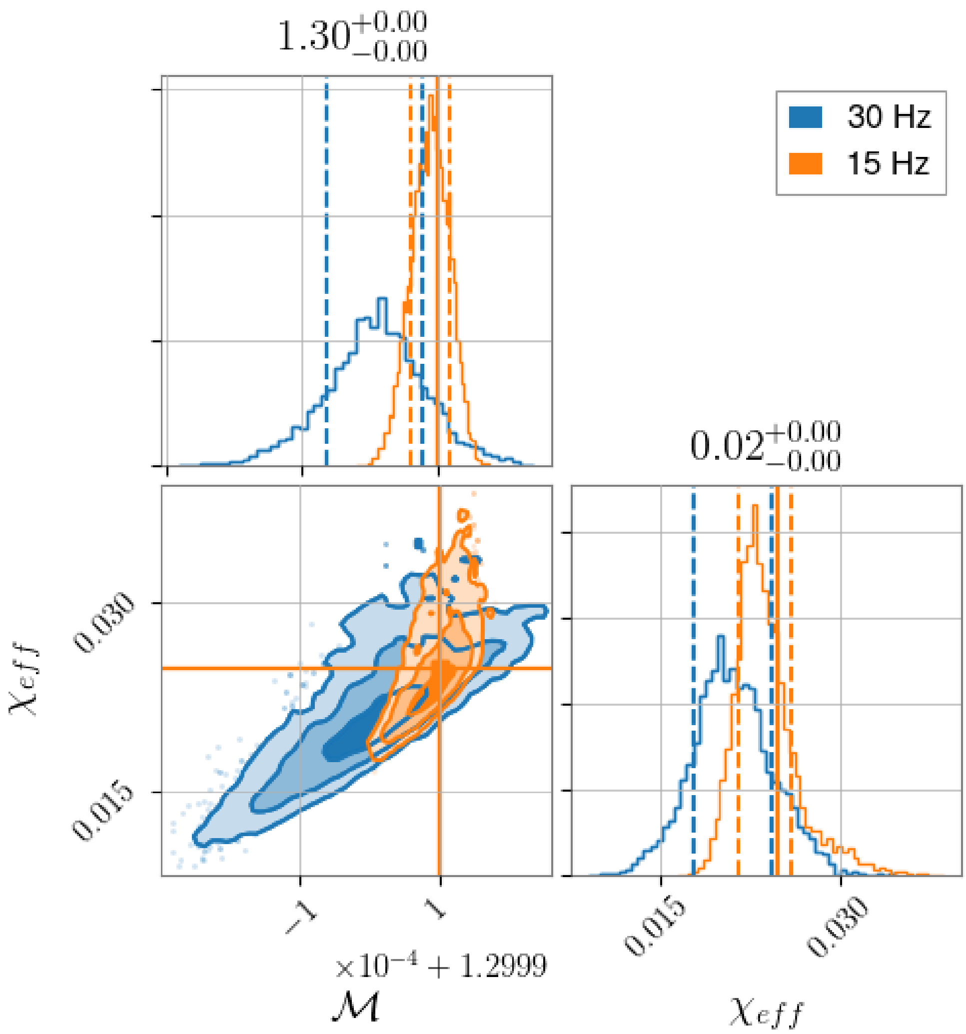

FIGURE 7.3: Posterior distributions for the chirp mass and effective spin of a binary neutron star inspiral similar to GW170817 when beginning the analysis at 30 Hz (blue) and 15 Hz (orange). Analyzing more of the early inspiral enables better measurement of the chirp mass, which leads to an improved measurement of the neutron star spins.

Here, $\mathcal{L}(d_i|\theta)$ is the likelihood of obtaining strain data d_i given binary parameters θ_i as in Eq. 7.1, $p(\theta|\Lambda)$ is our population model, and $VT(\Lambda)$ is the total observed spacetime volume if the population is described by Λ. See, e.g., (Finn & Chernoff, 1993; Dominik et al., 2015; Wycoski & O'Shaughnessy, 2018; Tiwari, 2018) for discussions of methods to calculate $VT(\Lambda)$.

Within GWPOPULATION we currently support the calculation of VT on a regular grid with GPU acceleration. The calculation of VT does not depend on the number of events. However, this grid-based integration is limited to small dimensional spaces, and so the subdominant effects of spin on detectability cannot be included in this method. This is mitigated by performing monte carlo integration. However, the cost of this compation scales as $O(N)$ (Farr, 2019). We are currently developing a method which will enable spin-effects to be included in the calculation of VT wih no increase in cost at runtime (Talbot, in prep.). For the benchmarking performed in this work, we ignore this quantity, but it may be added back in without significant impact on performance.

Since $\mathcal{L}(d_i|\theta)$ is independent of the population model, it can be evaluated independently for each observed binary, as described in Section 7.3. Since we are interested in evaluating an integral of the likelihood over the binary parameters, it is convenient to perform a Monte Carlo integral using samples from the likelihood. This process is known as "recycling".

The Bayesian inference algorithms used for single-event inference typically generate samples from the posterior distribution. Therefore, it is necessary to weight each of the samples by the prior probability distribution $p(\theta_{ij}|\varnothing)$ used during the initial inference step. This yields the following likelihood

$$\mathcal{L}_{\text{tot}}(\{d_i\}|\Lambda, R) \propto R^N e^{-RVT(\Lambda)} \prod_i^N \sum_j^{n_i} \frac{p(\theta_{ij}|\Lambda)}{p(\theta_{ij}|\varnothing)}. \tag{7.3}$$

Where $\{\theta_j\}_i$ is a set of n_i samples drawn from the posterior distribution $p(\theta_i|d_i)$. The evaluation of the population model for each of the posterior samples is embarrassingly parallel.

The first step is to draw an equal number of samples from each posterior so that for each binary parameter we have a single $N \times n_i$ array. These arrays are then transferred to the GPU to evaluate the probabilities, sums, and logarithms. We then need to only transfer a single number, the (log-)likelihood, back to the CPU when the likelihood is evaluated.

We calculate the likelihood evaluation time as a function of the number of posterior samples being recycled. The tests performed in this work use the mass distribution proposed in Talbot & Thrane, 2018, the spin magnitude distribution from Wysocki, Lange & O'Shaughnessy, 2019, and the spin orientation model from Talbot & Thrane, 2017. Fig. 7.4 shows the expected linear scaling in the speedup obtained by using the GPU. At around 3×10^6 samples, the GPU likelihood evaluation time begins increasing and the growth of the relative speedup slows. This is due to GPU queue saturation.

The data released after the second observing run of advanced LIGO/Virgo includes ten binary black hole systems and the shortest posterior contains $\sim 2 \times 10^4$ posterior samples. Using 2×10^5 samples the GPU code is $\approx 10\times$ faster, reducing runtimes from a week to less than a day.

During Advanced LIGO/Virgo's third observing run, beginning April 2019, we can expect to detect tens more binary black hole mergers (Abbott, 2016j). Given the current performance, we would expect the relative speed of the GPU code and the CPU code to continue to scale linearly with the size of the observed population. Within the lifetime of current detectors, we can conservatively assume that we will detect hundreds of events. At this stage using a GPU will accelerate population inference by more than two orders of magnitude.

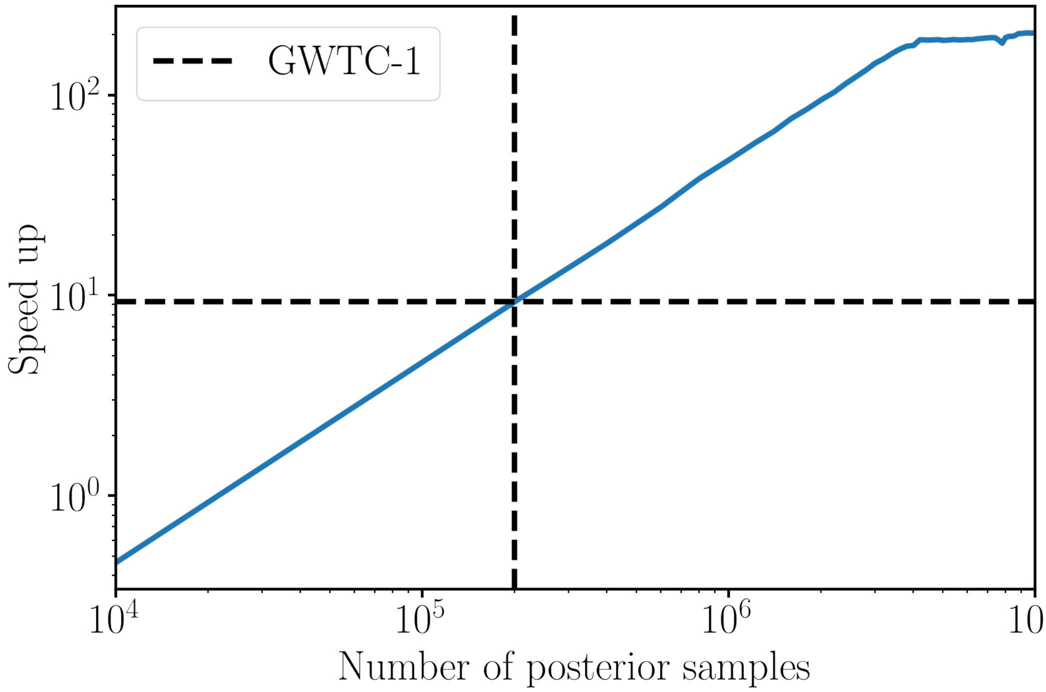

FIGURE 7.4: Ratio of the likelihood evaluation time (top) and likelihood evaluation time (bottom) on CPU and GPU as a function of number of samples. The vertical lines indicate (left to right): the number of samples released as part of the GWTC-1 data release, the anticipated number of samples at design sensitivity, the number of samples for a week of data. With the number of samples available as part of GWTC-1, the speedup is $\sim 10\times$. When the number of samples exceeds 4 million we reach the limits of the available GPUs and the likelihood evaluation time begins to increase. As GPU technology improves we expect that the maximum speedup over single-threaded code will continue to increase.

7.5 Discussion

As the field of gravitational-wave astronomy grows, the quantity of data to be analyzed is rapidly increasing. Thus, it is necessary to constantly improve and accelerate inference algorithms. In this paper, we demonstrate multiple ways in which GPUs can aid in this endeavor. We show that multiple orders of magnitude speedup can be achieved within the lifetime of current detectors in three areas:

- waveform evaluation.
- CBC likelihood evaluation.
- population inference.

Most of these improvements use CUPY, a python interface to CUDA, which acts as a GPU wrapper for existing C code. CUPY has also recently been used for other parameter estimation methods in gravitational-wave astronomy (Wysocki et al., 2019).

We provide two complementary GPU versions of commonly used waveforms, a CUDA implementation of IMRPHENOMPV2 and a python implementation of TAYLORF2. We find that the performance of the CUDA waveform exceeds that of the pure-python waveform for short waveforms when efficient memory allocation is vital. For longer waveforms, memory allocation is less important and the python waveforms give similar or greater speedups than the CUDA implementation. The CUDA implementation of IMRPHENOMPV2 is available at[2]. Future development of parallelized waveforms may enable rapid evaluation of waveforms encoding more of the phenomenology of general relativity, e.g., higher-order modes (Blackman et al., 2017), gravitational-wave kicks (Gerosa, Hébert & Stein, 2018), or gravitational-wave memory (Talbot et al., 2018).

Other than waveform evaluation, the dominant cost for the likelihood used in inference for compact binary coalescences are exponentials to perform frequency domain time-shifts. This is another operation which drastically benefits from parallelization. Using these two methods, we reduce the likelihood evaluation time for binary neutron star mergers by an order of magnitude at current sensitivity and more when current detectors reach their design sensitivity. The code for performing GPU-accelerated single-event parameter estimation can be found at[3].

Other methods for speeding up likelihood evaluation for long signals include reduced order quadrature methods (Smith et al., 2016) and relative binning (Cornish, 2010; Zackay, Dai & Venumadhav, 2018). These methods rely on the waveform being sufficiently well described by a small set of unevenly sampled frequencies. For binary neutron star mergers like GW170817, the signal from ~ 30 Hz to the merger can be described with only $\sim 10^3$ frequencies. Additionally, these methods do not require computing any exponentials at run time. GPU waveforms will have less of a benefit for these cases. However, we may be able to accelerate parameter estimation by an additional factor of a few. This will facilitate more rapid production of sky maps for electromagnetic observers following up on gravitational-wave events.

The computational cost of performing population inference increases linearly with the size of the observed population. Using a GPU to perform the embarrassingly parallel likelihood evaluation we find an acceleration of $\sim 10\times$ using the data in GWTC-1 (Abbott, 2019b) compared to the CPU code. We additionally find that the GPU implementation will outperform the CPU code by more than two orders of magnitude during the lifetime of current detectors. We therefore present GWPOPULATION[4]: a CPU/GPU agnostic framework for performing gravitational-wave population inference. Both of these packages use the framework available within BILBY (Ashton et al., 2019).

Within GWPOPULATION, we include CPU/GPU tools for performing population inference. We provide:

- implementations of many previously proposed binary black hole population models.

[2] adacs-ss18a-rsmith-python.readthedocs.io/en/latest/

[3] github.com/ColmTalbot/gpucbc

[4] github.com/ColmTalbot/gwpopulation, GWPOPULATION is also available through PYPI.

- the likelihoods commonly used in gravitational-wave population inference.
- methods for computing selection biases.

Using our GPU-enabled implementation of a binary neutron inspiral waveform we demonstrate that beginning the analysis at lower frequencies will improve our measurements of the intrinsic parameters of the system. While these results present a tantalizing glimpse of the physics that will be enabled through GPU acceleration, more work is required to realize these gains.

1. When analyzing signals for many minutes it will be necessary to include the effect of the Earth's rotation in our analysis. Including the effect of the Earth's rotation will improve sky localization since the movement of the detector allows triangulation from data taken at different times. While progress on this front has been made in recent years, e.g., (Marsat & Baker, 2018; Liang et al., 2019), a working implementation is not available at this time.
2. Central to the likelihood we use (Eq. 7.1) is an assumption of gaussianity and stationarity of the noise. These assumptions are not generally valid over the lengths of time considered in this paper.

The improvements between current sensitivity and the projected design sensitivities of advanced LIGO/Virgo are largest at low frequencies, so additional upgrades may be required in order to achieve the improvements shown here with real data.

As we enter "the data era" of gravitational-wave astronomy, optimizing Bayesian inference codes become ever more important. When the likelihood evaluation requires a large number of independent operations, GPUs can yield significant benefits.

Addendum

Since the publication of this paper additional compact binary coalescences have been observed and significant progress has been made in theoretical and observational analysis. The reader is directed to Chapters 1 and 8 for an overview of the field at the time of writing.

Chapter 8

Conclusion

In this thesis, I have developed theoretical models for astrophysical populations of compact binaries and created frameworks for enabling astrophysical inference. In this final chapter, I will describe recent results using these models and methods and provide an outlook on future measurements that can be made and development that will be needed to realise them.

8.1 Astrophysical population inference

During the first two observing runs of Advanced LIGO/Virgo, we observed ten binary black hole mergers and a single binary neutron star coalescence (Abbott, 2019b). Subsequent searches for gravitational-wave transients by groups outside of the LIGO/Virgo collaborations have independently identified these events along with a set of new significant triggers (e.g., Nitz et al., 2018a; Zackay et al., 2019; Venumadhav et al., 2019a; Nitz et al., 2019). During the first half of Advanced LIGO/Advanced Virgo's third observing run (O3a) 40 public alerts were released for triggers surpassing a once per month false alarm rate[1]. Seven of these triggers were subsequently retracted due to data quality issues, however, many of the remaining are likely to be confirmed as gravitational-wave transients in the coming months. Given this rate of triggers, we can expect to have many tens, possibly a hundred binary black hole mergers observed by the end of the third observing run.

8.1.1 Binary black holes

In Abbott, 2019a I applied the methods described in Chapters 3 and 4 to the binary black holes in Abbott, 2019b. The main result of this paper is the inference that 99% in binary black hole systems the more massive companion is less massive than $45M_{\odot}$. This is consistent with predictions about the location of a mass cutoff due to (pulsational) pair-instability supernovae. We also identified a low significance, $\ln BF \approx 2$, preference for the presence of a Gaussian deviation from a power-law distribution of black hole masses around

[1] At the time of writing 32 alerts have been issued since the beginning of O3b with 14 retractions

$\sim 35M_\odot$. This represents the first evidence that the black hole mass distribution does not follow a power-law mass distribution. A power-law distribution is often assumed as the base model because the stellar initial mass function follows a power-law over the relevant mass range (e.g., Kroupa et al., 2013). The location of this excess is less massive than most predictions of the location of black holes forming through pulsational pair-instability supernovae which is expcted to be $\gtrsim 40M_\odot$ (e.g., Farmer et al., 2019). This suggests that either our models of these supernovae are inadequate, or that another effect is at play. This is a tantalising glimpse of the astrophysical measurements we will be able to make in the coming years.

Additionally, we can rule out the hypothesis that all black holes have large spins or spins exactly aligned with the orbital angular momentum, but cannot make strong statements about the fraction of events with misaligned spins. We found that the distribution of spin magnitudes is skewed towards small spin magnitudes using a parameterisation from Wysocki, Lange & O'Shaughnessy, 2019. Using a model for the evolution of the binary black hole merger rate with redshift (Fishbach, Holz & Farr, 2018), we can say that the merger rate increases from $z = 0$ to $z = 0.9$ with 93% probability.

8.1.2 Binary neutron stars

The binary neutron star coalescence GW170817 rapidly became one of the most widely observed astronomical transients in recent years, leading to rich new information about the neutron star equation of state (Abbott, 2018d), confirmation that binary neutron stars are progenitors of short gamma-ray bursts (Abbott, 2017e), and a new independent measurement of the rate of expansion of the Universe (Abbott, 2017a). However, this is just the beginning of the physics which will be possible in the coming years as we begin to routinely observe such inspirals. Hernandez Vivanco et al., 2019 used random forest regression to combine multiple binary neutron star observations to infer the neutron star equation of state. Assuming current rate estimates and projected sensitivities, we estimate that we can constrain the radius of a $1.4M_\odot$ neutron star to within 10% by then of the fourth observing run.

8.1.3 Future work

To avoid biasing future population studies, it will be necessary to allow for the possibility that our transient catalogues contain terrestrial, as well as astrophysical, transients. Progress has recently been made towards this through two methods. Gaebel et al., 2019 demonstrate a method to include an expected fraction of terrestrial transients due to lowering the threshold for population inference. Smith & Thrane, 2018 suggest performing parameter estimation on all gravitational-wave data to make a statistical measurement of the binary black hole merger rate. Preliminary studies suggest that this method can be updated to allow the distribution of binary black holes to be measured simultaneously with this statistical measurement.

It is common practice to assign each trigger a probability that it is of astrophysical origin, $p_{\rm astro}$ (e.g., Kapadia et al., 2019; Zackay et al., 2019). Galaudage, Talbot & Thrane, 2019 developed a method to use $p_{\rm astro}$ as computed by the compact binary search pipelines to model the impurity of our catalogues and thus allow us to include more low significance events in our inference.

A current limiting factor in population analyses such as those performed in Abbott, 2019a is evaluating the observational selection function. The quantity typically used

$$\mathcal{V}(\Lambda) = \int dt d\theta p_{\rm det}(\theta) \pi(\theta|\Lambda), \tag{8.1}$$

often referred to as "VT" is the total observed spacetime volume given the population model, $\pi(\theta|\Lambda)$ requires integrating the probaility of detecting any source over the entire observing time. The quantity $p_{\rm det}$ is the probability of detecting a binary with parameters θ. At first order, $\mathcal{V}(\Lambda)$ depends on the assumed model for the mass and redshift distribution of binary black hole mergers. The next most important parameters for computing $\mathcal{V}$ are the aligned components of the component spins (e.g., Campanelli, Lousto & Zlochower, 2006; Scheel et al., 2015; Ng et al., 2018). The effect of spin-orbit precession, tidal deformability, and eccentricity on the observed volume has not been well studied aa they are not included in searches for compact binaries (e.g., Allen et al., 2012; Hooper et al., 2012; Adams et al., 2016; Messick et al., 2017; Nitz et al., 2018b; Venumadhav et al., 2019b; Sachdev et al., 2019), although unmodeled searches can perform better in searches for these binaries (e.g., Klimenko et al., 2016). The other "extrinsic" parameters are assumed to follow known, geometric, distributions.

Current methods for evaluating $\mathcal{V}(\Lambda)$ rely on performing a numerical integral for each likelihood evaluation, using either Gaussian quadrature (e.g., Wycoski & O'Shaughnessy, 2018) or Monte Carlo methods based on injections into real data (e.g., Tiwari, 2018). The number numerical operations to maintain a constant resolution with Gaussian quadrature scales exponentially with the dimensionality and so including subdominant effects, e.g., spin, in the integral is not possible. Monte Carlo integration scales better with increasing dimensionality, however, the number of elements in the sum scales linearly with the number of observed events (Farr, 2019), and by extension the number of numerical operations. This method can be parallelised as described in Chapter 7.

Additionally, I am developing a method to use machine learning to evaluate $\mathcal{V}(\Lambda)$ at runtime in O(ms). This is done by performing an offline training step where $\mathcal{V}(\Lambda)$ is computed in advance through either Gaussian quadrature or Monte Carlo integration for a large number of samples to generate a training set. A neural network is then trained to estimate $\mathcal{V}(\Lambda)$ from this training set. Generating the training data can be arbitrarily parallelised over many computational nodes.

8.2 Accelerating inference

The work in Chapter 7 represents just one attempt at accelerating astrophysical inference to keep pace with our increasing computational demands. Payne, Talbot & Thrane, 2019 applied importance sampling to the likelihood used in compact binary parameter estimation to perform the analysis with a waveform model which is prohibitively expensive for traditional sampling methods. We used this method to perform the first analysis of all ten binary black hole mergers in GWTC-1 using a waveform which contained "higher-order" waveform modes. This method has subsequently been adapted to search for eccentricity in the same catalogue (Romero-Shaw, Lasky & Thrane, 2019).

8.3 Gravitational-wave memory

Using the waveform model developed in Chapter 5 and the method from Payne, Talbot & Thrane, 2019, we performed the first search for gravitational-wave memory using gravitational-wave transients Hübner et al., 2020. While this initial sample of events was insufficient to identify the presence of gravitational-wave memory in the signal, we expect that this prediction of general relativity can be confirmed or falsified within in the next decade. Assuming a detection of gravitational-wave memory is made in the future, we can use the nature of the memory signal analogously to the "inspiral-merger-ringdown consistency tests" currently performed on binary black hole mergers (e.g., Abbott, 2019e).

Appendix A

Windowing effects in parameter estimation

Abstract

The noise power spectral densities of gravitational-wave detectors contain sharp features known as "lines". Due to the presence of these lines, it is necessary to apply a "window" to time-domain data before performing a fast Fourier transform to avoid spectral leakage. This windowing leads to a reduction in power of Gaussian noise by a known, window-dependent factor. To account for this loss of power, it is common to multiply the analysed data by a correction factor. However, this must be done carefully to avoid inadvertant amplification of astrophysical signals. In this appendix, I describe how we compute the correct factor and how it can be safely applied when performing parameter estimation on gravitational-wave transients. Additionally, I show how previous misapplication of these factors has led to a biased recovery of the luminosity distance of the first gravitational-wave transients analysed with LALINFERENCE.

A.1 Background

We measure strain data $d(t)$ in an interferometer. If the data contain an astrophysical signal, we can write this as

$$d = n + h_{\o}, \tag{A.1}$$

where $h_{\o}$ is the gravitational wave strain, and n is the noise in the detector. In both LALINFERENCE (Veitch et al., 2015) and BILBY (Ashton et al., 2019) we analyse this data in the frequency domain. We, therefore, have to take a fast Fourier transform (FFT) of the data

$$\tilde{d}(f) = \texttt{FFT}(d(t)). \tag{A.2}$$

However, to minimise spectral leakage and edge effects, a window function $w(t)$ is applied to the data before performing the FFT. This means that we have the convolution of the frequency domain strain and the frequency domain window

$$\tilde{d}'(f) = \texttt{FFT}(d(t)w(t)) = \tilde{d}(f) * \tilde{w}(f) \tag{A.3}$$

rather than $\tilde{d}$. We assume that the signal is not windowed in the frequency band, i.e.,

$$\tilde{h}'_{\o} = \tilde{h}_{\o}. \tag{A.4}$$

This can be achieved by choosing the segment duration and prior for coalescence time such that the signal is not affected by the window.

The standard likelihood we wish to evaluate is

$$\ln \mathcal{L} = -\frac{|\tilde{d} - \tilde{h}|^2}{2P} - \ln(2\pi P), \tag{A.5}$$

where P is the noise power spectral density (PSD) of n, and $\tilde{h}$ is our template which approximates the frequency-domain gravitational-wave strain. We note that generally $\tilde{h} \neq \tilde{h}'_{\o}$. However, if our template perfectly matches the true signal, $\tilde{h} = \tilde{h}'_{\o}$, this reduces to

$$\ln \mathcal{L}_{\o} = -\frac{|\tilde{n}|^2}{2P} - \ln(2\pi P). \tag{A.6}$$

In addition to this expected behaviour of the likelihood, it is generally accepted that the real and imaginary components of the "whitened residual", $\tilde{n}_{\mathbb{R},\mathbb{I}}/P^{1/2}$, should independently follow a unit normal distribution, i.e.,

$$\frac{\tilde{n}_{\mathbb{R},\mathbb{I}}}{P^{1/2}} \sim \mathcal{N}(\mu = 0, \sigma^2 = 1). \tag{A.7}$$

However, applying the window reduces the power in the signal due to the part of the noise which is reduced at the edges from windowing. Using

Parseval's theorem, this power loss factor can be shown to be

$$\overline{w^2} = \left\langle w(t)^2 \right\rangle \tag{A.8}$$

Equation A.6 can therefore be written

$$\ln \mathcal{L}_\varnothing = -\frac{1}{2}\frac{|\tilde{n}'|^2}{\overline{w^2}P} - \ln(2\pi P), \tag{A.9}$$

$$\ln \mathcal{L} = -\frac{1}{2}\frac{|\tilde{d}' - \tilde{h}|^2}{\overline{w^2}P} = -\frac{1}{2}\frac{|\tilde{n}' + \tilde{h}_\varnothing - \tilde{h}|^2}{\overline{w^2}P} - \ln(2\pi P). \tag{A.10}$$

A.2 Window choice

The window most commonly used for gravitational-wave data analysis is the Tukey window (Harris, 1978),

$$w(t - t_0, \delta t) = \begin{cases} \frac{1}{2}\left[1 + \cos\left(\pi\left(\frac{t}{\delta t} - 1\right)\right)\right] & 0 \leq t \leq \delta t, \\ 1 & \delta t \leq t \leq T - \delta t, \\ \frac{1}{2}\left[1 + \cos\left(\pi\left(\frac{t-T}{\delta t} + 1\right)\right)\right] & T - \delta t \leq t \leq T. \end{cases} \tag{A.11}$$

The "roll-off" parameter, δt, is the duration over which the window rises from the beginning of the segment and rolls-off at the end of the segment. The parameter T is the duration of the analysis segment. The Tukey window is often instead parameterised by the fraction of the window in the rise *and* roll-off, $\alpha \equiv 2\delta t / T$.

For this window, we can analytically integrate and calculate the exact correction factor that we should apply

$$\left\langle \bar{\omega^2} \right\rangle = \frac{1}{T}\int_0^T dt w^2(t, \delta t) = 1 - \frac{5\delta t}{4T} = 1 - \frac{5\alpha}{8} \tag{A.12}$$

In Table A.1 I show $\left\langle \bar{\omega^2} \right\rangle$ for $\delta t = 0.2$ and a range of commonly used durations.

A.3 Implementation

There are two common ways that factors of $\overline{w^2}$ can be included when performing our analysis.

A.3.1 PSD correction

Depending on how P is calculated, the window factor may or may not be included in the definition of P. This is the case for many off-the-shelf estimation routines, including those implemented in SCIPY and MATPLOTLIB, which

T [s]	α	$\langle \bar{\omega^2} \rangle$	$1 - \langle \bar{\omega^2} \rangle^{0.5}$
4	0.1	0.9375	3.175×10^{-2}
8	0.05	0.96875	1.575×10^{-2}
16	0.025	0.984375	7.843×10^{-3}
32	0.0125	0.9921875	3.914×10^{-3}
64	0.00625	0.99609375	1.955×10^{-3}
128	0.003125	0.998046875	9.770×10^{-4}

TABLE A.1: Window normalisation factors for various segment durations assuming a roll-off of $\delta t = 0.2s$. The columns show: segment duration, the Tukey parameter (α), the power-reduction factor ($\langle \bar{\omega^2} \rangle$), and the fractional change in amplitude.

are used in BILBY (Chapter 6)[1]. In the previous section, we assumed that this factor is not included. However, if it is included, then the factor of $\overline{w^2}^{-1}$ also has to be added to the likelihood manually, leading to cancellation with the factor included in Equation A.10[2]. Since we do not apply the correction to the data segment, it should not be applied to the template.

A.3.2 Template correction

It is tempting to apply the correction directly to the data after taking the FFT:

$$\tilde{d} = \tilde{d}' / \sqrt{\overline{w^2}}. \tag{A.13}$$

However, this leads to an undesirable amplification of the signal since there is no power loss in the signal due to windowing; (recall, we were careful to make sure the signal did not overlap with the window edge). It is the *noise* that we must correct, not the signal. We must, therefore, apply the same factor to our template. In this case, we see that we recover the previous likelihood

$$\ln \mathcal{L} = -\frac{1}{2} \frac{|\tilde{d} - \tilde{h} / \sqrt{\overline{w^2}}|^2}{P} \tag{A.14}$$

$$= -\frac{1}{2} \frac{|\tilde{d}' / \sqrt{\overline{w^2}} - \tilde{h} / \sqrt{\overline{w^2}}|^2}{P} \tag{A.15}$$

$$= -\frac{1}{2} \frac{|\tilde{d}' - \tilde{h}|^2}{\overline{w^2} P}. \tag{A.16}$$

This is the method used in LALINFERENCE. During the development of BILBY, I noticed that the luminosity distances returned by the two codes differed by the analysis segment duration-dependent factor described above,

[1] This factor is applied automatically when the user calls `bilby.Interferometer.power_spectral_density_array`.

[2] Another point of caution here is that the window applied when estimating the PSD must be identical to the window applied to the analysis segment to avoid other sources of bias.

see, e.g., Fig. 6.2. The reason for this difference was because LALINFERENCE uses method 2 to perform this correction, correcting the frequency-domain data and template independently. However, the correction was only being applied to the template if it was calculated in the time-domain.

The bug was resolved by merge request[3]. This means that the distance estimated in any analysis of time-domain data (real or simulated) performed before this commit with a frequency-domain waveform model was biased by a known factor, given by the final column of Table A.1. The O1/O2 catalogue paper, GWTC-1, was published after this bug was fixed, therefore the distances in the *final* official data release are unaffected, however, initial data releases were, along with source estimates for all papers published before GWTC-1.

A.4 Tests

We can test the correctness of windowing procedures in two different ways, in order to identify errors associated with the two implementation methods described.

A.4.1 Data whitening

As mentioned in Section A.1, our aim is for the power spectral density to correctly whiten the data, i.e.,

$$\frac{\tilde{n}_{\mathbb{R},\mathbb{I}}}{\overline{w^2}^{1/2} P^{1/2}} \sim \mathcal{N}(\mu = 0, \sigma = 1). \tag{A.17}$$

Here $\tilde{n}_{\mathbb{R},\mathbb{I}}$ are the real and imaginary components of the frequency-domain noise. If the term $\overline{w^2}$ is omitted entirely the noise will be "over-whitened," which will change the standard deviation.

Several methods are used to assess how well the data is whitened by the PSD. The first is a visual comparison of a histogram of the whitened strain, an example of this is shown in Figure A.1. In this figure I show the distribution of the whitened strain when: applying a rectangular window, and applying a Tukey window with $\alpha = 0.5$ with and without including the $\overline{w^2}$ factor.

A more quantitative approach is to compute the Anderson-Darling statistic for the whitened strain,

$$A^2 = N \int_{-\infty}^{\infty} dF \, \frac{(F_s - F)^2}{F(1 - F)}. \tag{A.18}$$

Here F is the expected cumulative distribution, in our case, the cumulative distribution of a unit normal distribution, and F_s is the cumulative distribution of the samples, and N is the number of samples. If the two distributions

[3] https://git.ligo.org/lscsoft/lalsuite/merge_requests/780

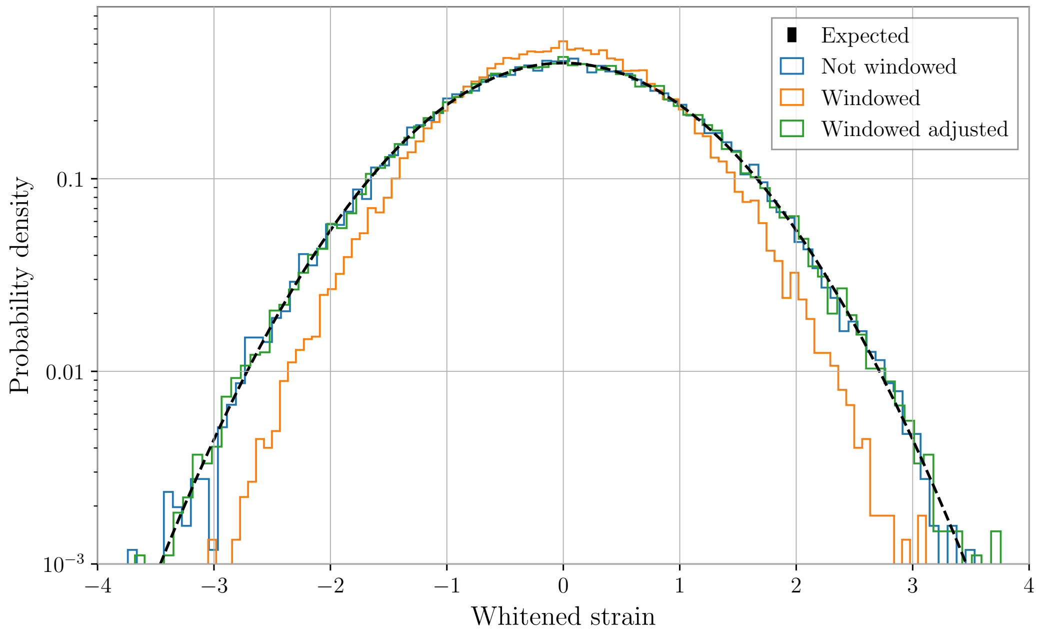

FIGURE A.1: Whitened strain in three cases: no window applied (orange), window applied without accounting for power loss factor (green), and after applying the factor (purple).

being compared are the same then the distribution of the Anderson-Darling statistic should follow a predictable distribution.

When applying a rectangular window, the distribution of the Anderson-Darling statistic is as expected. However, when we apply a Tukey window with $\alpha \gtrsim 0.0125$ to the data, the distribution of the Anderson-Darling statistic is twice as narrow as expected, telling us that the data appears to match the unit normal distribution *too well*, see, Figure A.2. This effect was noticed by Chatziioannou et al., 2019 where the authors attributed this to not resolving spectral lines when using short segment durations. However, the effect can be explained by noting that windowing the data introduces an offset in the amplitude and phase of the frequency domain strain, see Figure A.3. This, in turn, effectively doubles the number of degrees of freedom as both the original real and imaginary information is in each component. Hence we should use $2N$ in Equation A.18. We note that when using sharply rising windows, long segment durations in Fig. A.3, there is significant correlation between the offsets at neighbouring frequencies, however, this does not appear to affect the Anderson-Darling statistic at the level tested here.

A.4.2 Signal recovery

If the correction factor is applied to the data without also applying that factor to the template, we see the ill effects by recovering an injected signal with known parameters in a zero noise realization. Specifically, if the template has

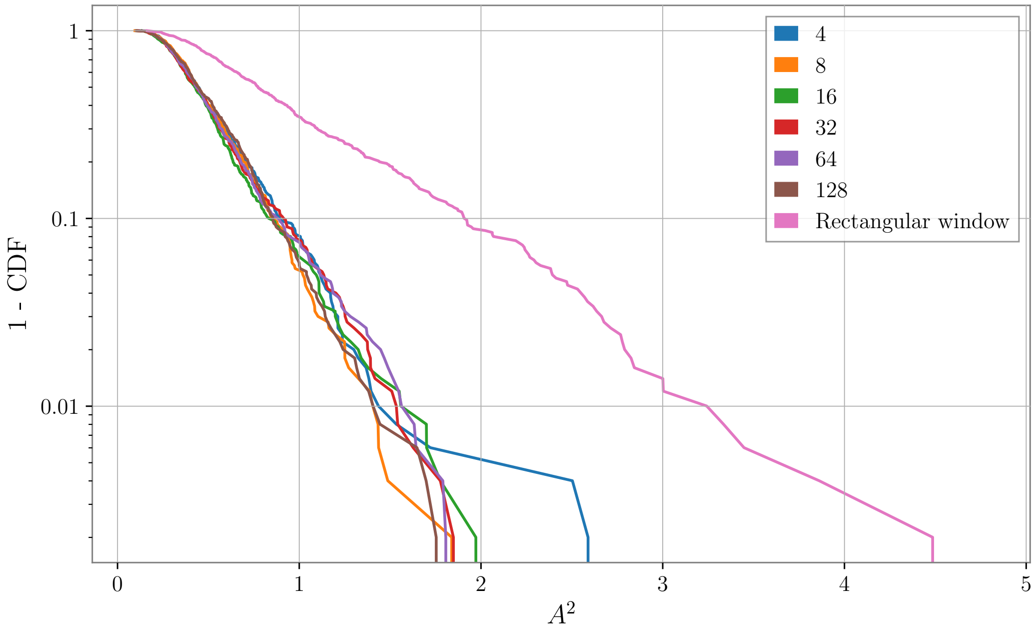

FIGURE A.2: The distribution of the Anderson-Darling statistic for various segment durations.

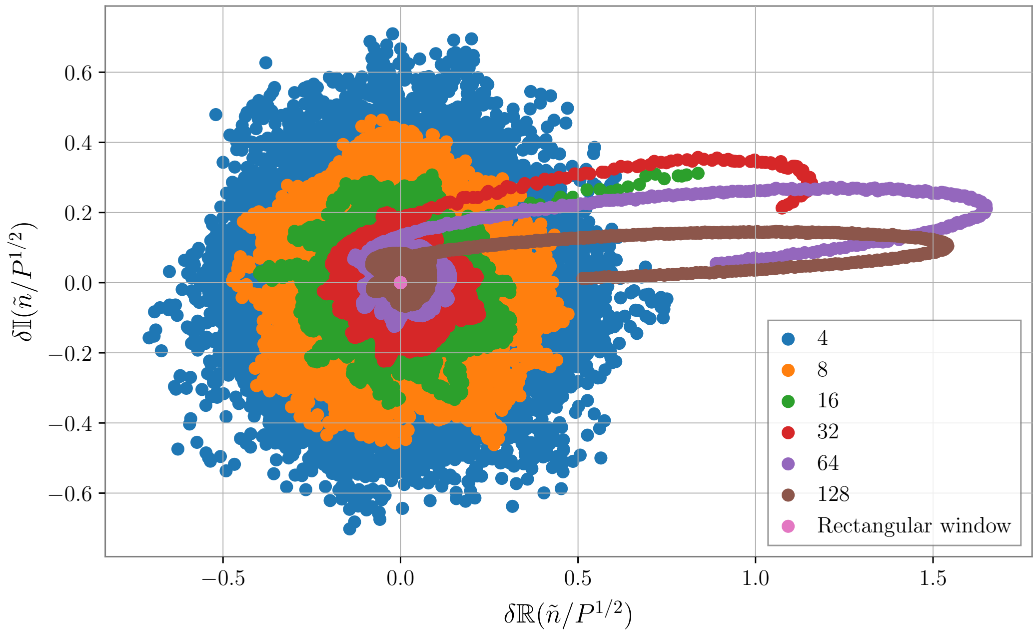

FIGURE A.3: The change in the real and imaginary components of the whitened strain after applying a range of windows. When applying a rectangular window there is no deviation in the strain. For very sharply turning on windows, i.e., small α/long durations the neighbouring frequency bins are highly correlated, leading to the features in the plot. For windows with a larger α, shorter duration, the offsets appear random and uncorrelated.

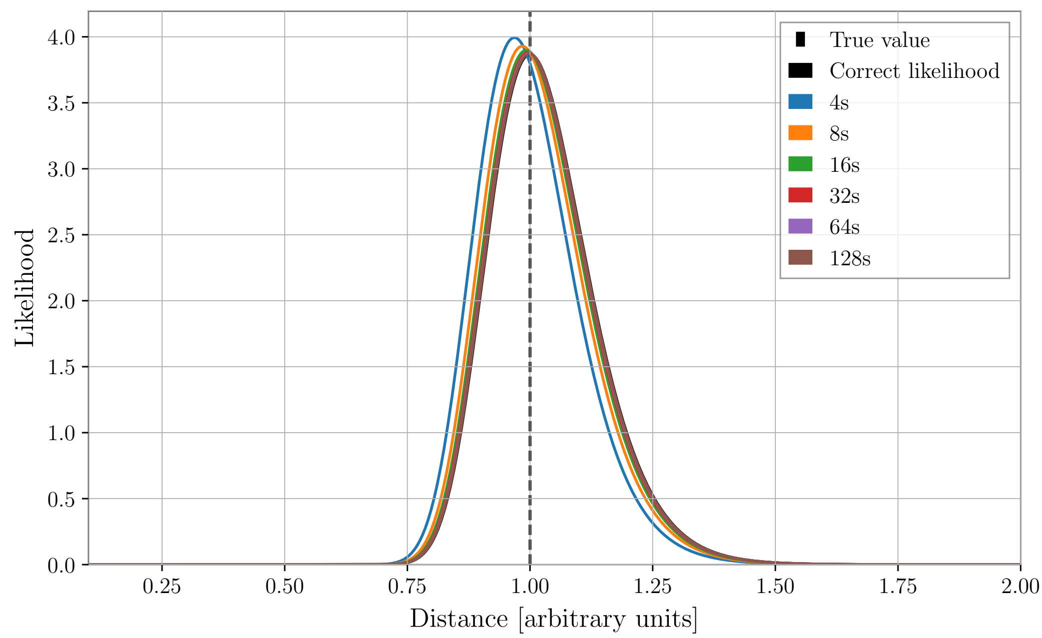

FIGURE A.4: Likelihood of data given distance for a zero-noise realisation when not accounting for the amplification of the signal after correcting the power in the frequency domain strain.

not been corrected to account for the amplification of the signal, the distance will be underestimated by a factor of $\sqrt{\overline{w^2}}$[4].

An example posterior distribution for distance calculated using a range of windows without applying the appropriate correction are shown in Figure A.4. This fractional bias in the maximum likelihood distance is shown in the final column of Table A.1 for a range of windows and the corresponding segment duration for a window roll-off of 0.2s. A real-world example of this bias can be seen in Figure 6.2. The LALINFERENCE posterior for distance is clearly shifted to smaller values compared to the posterior recovered by BILBY.

[4]This is due to the signal amplitude scaling like $1/d_L$.

Bibliography

Aasi, J et al. (2015), (2015), Class. Quantum Grav., 32.7, 74001, DOI: 10.1088/0264-9381/32/7/074001

Abadie, J et al. (Sept. 2010), (Sept. 2010), Class. Quantum Grav., 27.17, 173001, DOI: 10.1088/0264-9381/27/17/173001

Abbott, B. P. et al. (Oct. 2017a), (Oct. 2017a), Nature, 551.7678, 85, DOI: 10.1038/nature24471

—(2018a), (2018a), Phys. Rev. Lett., 120, 201102, DOI: 10.1103/PhysRevLett.120.201102

—(2016a), (2016a), Astrophys. J. Lett., 818.2, L22, DOI: 10.3847/2041-8205/818/2/L22

—(Oct. 2016b), (Oct. 2016b), Phys. Rev. X, 6, 041015, DOI: 10.1103/PhysRevX.6.041015

—(Sept. 2019a), (Sept. 2019a), Astrophys. J., 882.2, L24, DOI: 10.3847/2041-8213/ab3800

—(2017b), (2017b), Class. Quantum Grav., 34.10, 104002, URL: http://stacks.iop.org/0264-9381/34/i=10/a=104002

—(2017c), (2017c), Class. Quantum Grav., 34.4, 44001, DOI: 10.1088/1361-6382/aa51f4

—(2017d), (2017d), Astrophys. J., 839, 12, DOI: 10.3847/1538-4357/aa677f

—(Jan. 2018b), (Jan. 2018b), Phys. Rev. Lett., 120.3, 031104, DOI: 10.1103/PhysRevLett.120.031104

—(Oct. 2017e), (Oct. 2017e), Astrophys. J., 848, L13, DOI: 10.3847/2041-8213/aa920c

—(June 2016c), (June 2016c), Phys. Rev. D, 93.12, 122003, DOI: 10.1103/PhysRevD.93.122003

—(Mar. 2016d), (Mar. 2016d), Phys. Rev. Lett., 116.13, 131102, DOI: 10.1103/PhysRevLett.116.131102

—(Mar. 2016e), (Mar. 2016e), Phys. Rev. Lett., 116.13, 131103, DOI: 10.1103/PhysRevLett.116.131103

—(June 2016f), (June 2016f), Phys. Rev. Lett., 116.24, 241103, DOI: 10.1103/PhysRevLett.116.241103

—(June 2017f), (June 2017f), Phys. Rev. Lett., 118.22, 221101, DOI: 10.1103/PhysRevLett.118.221101

—(Dec. 2017g), (Dec. 2017g), Astrophys. J., 851.2, L35, DOI: 10.3847/2041-8213/aa9f0c

—(Oct. 2017h), (Oct. 2017h), Phys. Rev. Lett., 119.14, 141101, DOI: 10.1103/PhysRevLett.119.141101

—(Feb. 2018c), (Feb. 2018c), Phys. Rev. Lett., 120 (9), 091101, DOI: 10.1103/PhysRevLett.120.091101

Abbott, B. P. et al. (2018d), (2018d), Phys. Rev. Lett., 121.16, 161101, DOI: 10.1103/PhysRevLett.121.161101
—(Oct. 2017i), (Oct. 2017i), Phys. Rev. Lett., 119.16, 161101, DOI: 10.1103/PhysRevLett.119.161101
—(Sept. 2019b), (Sept. 2019b), Phys. Rev. X, 9 (3), 031040, DOI: 10.1103/PhysRevX.9.031040
—(Oct. 2016g), (Oct. 2016g), Phys. Rev. X, 6.4, 41014, DOI: 10.1103/PhysRevX.6.041014
—(Oct. 2017j), (Oct. 2017j), Astrophys. J., 848, L12, DOI: 10.3847/2041-8213/aa91c9
—(Feb. 2016h), (Feb. 2016h), Phys. Rev. Lett., 116.6, 061102, DOI: 10.1103/PhysRevLett.116.061102
—(June 2016i), (June 2016i), Phys. Rev. Lett., 116.24, 241102, DOI: 10.1103/PhysRevLett.116.241102
—(2018e). (2018e), "Properties of the binary neutron star merger GW170817", URL: https://dcc.ligo.org/LIGO-P1800061/public
—(2019c), (2019c), Phys. Rev. X, 9, 011001, DOI: 10.1103/PhysRevX.9.011001
—(2016j), (2016j), Living Reviews in Relativity, 19.1, 1, DOI: 10.1007/lrr-2016-1
—(2017k), (2017k), Astrophys. J., 851, L16, DOI: 10.3847/2041-8213/aa9a35
—(2016k), (2016k), Phys. Rev. Lett., 116.22, 221101, DOI: 10.1103/PhysRevLett.116.221101
—(July 2019d), (July 2019d), Phys. Rev. Lett., 123.1, DOI: 10.1103/PhysRevLett.123.011102
—(Nov. 2019e), (Nov. 2019e), Phys. Rev. D, 100 (10), 104036, DOI: 10.1103/PhysRevD.100.104036
—(Mar. 2017l), (Mar. 2017l), Phys. Rev. Lett., 118.12, 121101, DOI: 10.1103/PhysRevLett.118.121101
Acernese, F et al. (Jan. 2015), (Jan. 2015), Class. Quantum Grav., 32.2, 024001, DOI: 10.1088/0264-9381/32/2/024001
Adams, M. R., Cornish, N. J., & Littenberg, T. B. (Dec. 2012), (Dec. 2012), Phys. Rev. D, 86.12, 124032, DOI: 10.1103/PhysRevD.86.124032
Adams, T., Buskulic, D., Germain, V., et al. (Aug. 2016), (Aug. 2016), Class. Quantum Gravity, 33.17, DOI: 10.1088/0264-9381/33/17/175012
Ajith, P., Hannam, M., Husa, S., et al. (June 2011), (June 2011), Phys. Rev. Lett., 106.24, 241101, DOI: 10.1103/PhysRevLett.106.241101
Albrecht, S., Winn, J. N., Carter, J. A., Snellen, I. A. G., & de Mooij, E. J. W. (2011), (2011), Astrophys. J., 726.2, 68, DOI: 10.1088/0004-637X/726/2/68
Albrecht, S., Winn, J. N., Torres, G., et al. (2014), (2014), Astrophys. J., 785.2, 83, DOI: 10.1088/0004-637X/785/2/83
Allen, B., Anderson, W. G., Brady, P. R., Brown, D. A., & Creighton, J. D. E. (June 2012), (June 2012), Phys. Rev. D, 85.12, 122006, DOI: 10.1103/PhysRevD.85.122006
Anderson, W. G., Brady, P. R., Creighton, J. D. E., & Flanagan, É. É. (Jan. 2001), (Jan. 2001), Phys. Rev. D, 63.4, 042003, DOI: 10.1103/PhysRevD.63.042003
Andreon, S., & Weaver, B. (2015). (2015), Bayesian Methods for the Physical Sciences (First; Switzerland: Springer), DOI: 10.1007/978-3-319-15287-5

Antonini, F., Toonen, S., & Hamers, A. S. (May 2017), (May 2017), Astrophys. J., 841.2, 77, DOI: 10.3847/1538-4357/aa6f5e

Apostolatos, T. A., Cutler, C., Sussman, G. J., & Thorne, K. S. (1994), (1994), Phys. Rev. D, 49.12, 6274, DOI: 10.1103/PhysRevD.49.6274

Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. (Sept. 2015), (Sept. 2015), Astrophys. J., 810.2, 150, DOI: 10.1088/0004-637X/810/2/150

Ashton, G., Hübner, M, Lasky, P., Talbot, C., et al. (2018a). (2018a), Bilby documentation, URL: https://lscsoft.docs.ligo.org/bilby/

—(2018b). (2018b), Bilby repository, URL: https://git.ligo.org/lscsoft/bilby/

Ashton, G., Hübner, M., Lasky, P. D., et al. (Apr. 2019), (Apr. 2019), Astrophys. J.s, 241.2, 27, DOI: 10.3847/1538-4365/ab06fc

Aso, Y. et al. (2013), (2013), Phys. Rev. D, 88.4, 043007, DOI: 10.1103/PhysRevD.88.043007

Babak, S., Baker, J. G., Benacquista, M. J., et al. (2008), (2008), Class. Quantum Grav., 25, 184026, DOI: 10.1088/0264-9381/25/18/184026

Babak, S., Baker, J. G., Benacquista, M. J., et al. (2010), (2010), Class. Quantum Grav., 27, 084009, DOI: 10.1088/0264-9381/27/8/084009

Baiotti, L., Giacomazzo, B., & Rezzolla, L. (Oct. 2008), (Oct. 2008), Phys. Rev. D, 78.8, 084033, DOI: 10.1103/PhysRevD.78.084033

Barbary, K. (2015). (2015), nestle, URL: http://kylebarbary.com/nestle/

Bardeen, J. M., & Petterson, J. A. (Jan. 1975), (Jan. 1975), Astrophys. J., 195, L65, DOI: 10.1086/181711

Barrett, J. W., Gaebel, S. M., Neijssel, C. J., et al. (Nov. 2018), (Nov. 2018), Mon. Not. R. Astron. Soc., 477.4, 4685, DOI: 10.1093/mnras/sty908

Bayes, T. (Jan. 1763), (Jan. 1763), Philos. Trans. R. Soc. London, 53, 370, DOI: 10.1098/rstl.1763.0053

Belczynski, K., Klencki, J., Meynet, G., et al. (June 2017), (June 2017), arXiv: 1706.07053

Belczynski, K., Heger, A., Gladysz, W., et al. (2016), (2016), A&A, 594, A97, DOI: 10.1051/0004-6361/201628980

Belloni, D., Askar, A., Giersz, M., Kroupa, P., & Rocha-Pinto, H. J. (Nov. 2017), (Nov. 2017), Monthly Notices of the Royal Astronomical Society, 471.3, 2812, DOI: 10.1093/mnras/stx1763

Berti, E., Cardoso, V., & Starinets, A. O. (2009), (2009), Class. Quantum Gravity, 26.16, DOI: 10.1088/0264-9381/26/16/163001

Bishop, N. T., & Rezzolla, L. (Dec. 2016), (Dec. 2016), Living Reviews in Relativity, 19.1, 2, DOI: 10.1007/s41114-016-0001-9

Biwer, C. M., Capano, C. D., De, S., et al. (Feb. 2019), (Feb. 2019), Publ. Astron. Soc. Pacific, 131.996, 024503, DOI: 10.1088/1538-3873/aaef0b

Blackman, J., Field, S. E., Scheel, M. A., et al. (2017), (2017), Phys. Rev. D, 95, DOI: 10.1103/PhysRevD.95.104023

Blackman, J., Field, S. E., Scheel, M. A., et al. (2017), (2017), Phys. Rev. D, 96.2, 024058, DOI: 10.1103/PhysRevD.96.024058

Blanchet, L. (Feb. 2014), (Feb. 2014), Living Rev. Relativ., 17, DOI: 10.12942/lrr-2014-2

Bogdanović, T., Reynolds, C. S., & Miller, M. C. (2007), (2007), Astrophys. J., 661.2, L147, DOI: 10.1086/518769

Bohé, A., Shao, L., Taracchini, A., et al. (Feb. 2017), (Feb. 2017), Phys. Rev. D, 95.4, 044028, DOI: 10.1103/PhysRevD.95.044028

Bose, S., Chakravarti, K., Rezzolla, L., Sathyaprakash, B. S., & Takami, K. (2018), (2018), Phys. Rev. Lett., 120.3, 031102, DOI: 10.1103/PhysRevLett.120.031102

Braginsky, V. B., & Thorne, K. S. (1987), (1987), Nature, 327, 123, DOI: 10.1038/327123a0

Brown, D. D., & Freise, A. (2014). (2014), Finesse, DOI: 10.5281/zenodo.821363

Buchner, J., Georgakakis, A., Nandra, K., et al. (Apr. 2014), (Apr. 2014), Astronomy & Astrophysics, 564, A125, DOI: 10.1051/0004-6361/201322971

Buchner, J. et al. (2014), (2014), Astron. Astrophys., 564, A125, DOI: 10.1051/0004-6361/201322971

Calderón Bustillo, J., Laguna, P., & Shoemaker, D. (May 2017), (May 2017), Phys. Rev. D, 95.10, 104038, DOI: 10.1103/PhysRevD.95.104038

Calderón Bustillo, J., Husa, S., Sintes, A. M., & Pürrer, M. (Apr. 2016), (Apr. 2016), Phys. Rev. D, 93.8, 084019, DOI: 10.1103/PhysRevD.93.084019

Callister, T., Biscoveanu, A. S., Christensen, N., et al. (2017), (2017), Phys. Rev. X, 7, 041058, DOI: 10.1103/PhysRevX.7.041058

Campanelli, M, Lousto, C. O., & Zlochower, Y (2006), (2006), Phys. Rev. D - Particles, Fields, Gravitation and Cosmology, 74.4, DOI: 10.1103/PhysRevD.74.041501

Canizares, P., Field, S. E., Gair, J., et al. (2015), (2015), Phys. Rev. Lett., 114.7, 071104, DOI: 10.1103/PhysRevLett.114.071104

Canizares, P., Field, S. E., Gair, J. R., & Tiglio, M. (2013), (2013), Phys. Rev. D, 87, 124005, DOI: 10.1103/PhysRevD.87.124005

Carter, B. (1971), (1971), Phys. Rev. Lett., 26.6, 331, DOI: 10.1103/PhysRevLett.26.331

Chandrasekhar, S., & Detweiler, S. (Aug. 1975), (Aug. 1975), Proc. R. Soc. A Math. Phys. Eng. Sci., 344.1639, 441, DOI: 10.1098/rspa.1975.0112

Chattopadhyay, A. K., & Chattopadhyay, T. (2014). (2014) (First; New York: Springer), DOI: 10.1007/978-1-4939-1507-1

Chatziioannou, K., Clark, J. A., Bauswein, A., et al. (Dec. 2017), (Dec. 2017), Phys. Rev. D, 96.12, 124035, DOI: 10.1103/PhysRevD.96.124035

Chatziioannou, K., Haster, C.-J., Littenberg, T. B., et al. (Nov. 2019), (Nov. 2019), Phys. Rev. D, 100 (10), 104004, DOI: 10.1103/PhysRevD.100.104004

Christodoulou, D. (1991), (1991), Phys. Rev. Lett., 67.12, 1486, DOI: 10.1103/PhysRevLett.67.1486

Clark, J., Bauswein, A., Cadonati, L., et al. (Sept. 2014), (Sept. 2014), Phys. Rev. D, 90.6, 062004, DOI: 10.1103/PhysRevD.90.062004

Clark, J. A., Bauswein, A., Stergioulas, N., & Shoemaker, D. (2016), (2016), Class. Quantum Grav., 33.8, 085003, DOI: 10.1088/0264-9381/33/8/085003

Cordes, J. M., & Jenet, F. A. (June 2012), (June 2012), Astrophys. J., 752.1, 54, DOI: 10.1088/0004-637X/752/1/54

Cornish, N. J., & Romano, J. D. (2013), (2013), Phys. Rev. D, 87, 122003, DOI: 10.1103/PhysRevD.87.122003

Cornish, N. J. (July 2010), (July 2010), arXiv: 1007.4820

Cornish, N. J., & Littenberg, T. B. (2015), (2015), Class. Quantum Grav., 32, 135012, DOI: 10.1088/0264-9381/32/13/135012

Cutler, C., & Flanagan, É. E. (1994), (1994), Phys. Rev. D, 49, 2658, DOI: 10.1103/PhysRevD.49.2658

Dale, A. I. (1982), (1982), Arch. Hist. Exact Sci., 27.1, 23, DOI: 10.1007/BF00348352

Damour, T. (Jan. 2014), (Jan. 2014), in Gen. Relativ. Cosmol. Astrophys. Vol. 177 (Springer Science+Business Media), 111, DOI: 10.1007/978-3-319-06349-2_5, arXiv: 1212.3169

Damour, T., Iyer, B. R., & Sathyaprakash, B. S. (2005), (2005), Phys. Rev. D, 63, 044023, DOI: 10.1103/PhysRevD.63.044023

De, S., Finstad, D., Lattimer, J. M., et al. (Aug. 2018), (Aug. 2018), Phys. Rev. Lett., 121 (9), 091102, DOI: 10.1103/PhysRevLett.121.091102

Doctor, Z., Farr, B., Holz, D. E., & Pürrer, M. (Dec. 2017), (Dec. 2017), Phys. Rev. D, 96.12, DOI: 10.1103/PhysRevD.96.123011

Dominik, M., Berti, E., O 'shaughnessy, R., et al. (2015), (2015), Astrophys. J., 806, 263, DOI: 10.1088/0004-637X/806/2/263

Dupuis, R. J., & Woan, G. (2005), (2005), Phys. Rev. D, 72, 102002, DOI: 10.1103/PhysRevD.72.102002

Easter, P. J., Lasky, P. D., Casey, A. R., Rezzolla, L., & Takami, K. (Aug. 2019), (Aug. 2019), Phys. Rev. D, 100.4, DOI: 10.1103/physrevd.100.043005

Essick, R., Vitale, S., & Evans, M. (Oct. 2017), (Oct. 2017), Phys. Rev. D, 96.8, DOI: 10.1103/PhysRevD.96.084004

Farmer, R., Renzo, M., de Mink, S. E., Marchant, P., & Justham, S. (Dec. 2019), (Dec. 2019), The Astrophysical Journal, 887.1, 53, DOI: 10.3847/1538-4357/ab518b

Farr, B., Holz, D. E., & Farr, W. M. (Feb. 2018), (Feb. 2018), Astrophys. J., 854, L9, DOI: 10.3847/2041-8213/aaaa64

Farr, B., Ochsner, E., Farr, W. M., & O'Shaughnessy, R. (July 2014), (July 2014), Phys. Rev. D, 90.2, 024018, DOI: 10.1103/PhysRevD.90.024018

Farr, W. M. (2014). (2014), "Marginalisation of the Time Parameter in Gravitational Wave Parameter Estimation", URL: https://dcc.ligo.org/T1400460-v2/public

Farr, W. M, Farr, B., & Littenberg, T. (2014). (2014), "Modelling calibration errors in CBC waveforms", URL: https://dcc.ligo.org/LIGO-T1400682/public

Farr, W. M. (May 2019), (May 2019), Res. Notes AAS, 3.5, 66, DOI: 10.3847/2515-5172/ab1d5f

Farr, W. M., Gair, J. R., Mandel, I., & Cutler, C. (Jan. 2015), (Jan. 2015), Phys. Rev. D, 91.2, 023005, DOI: 10.1103/PhysRevD.91.023005

Farr, W. M., Stevenson, S., Miller, M. C., et al. (Aug. 2017), (Aug. 2017), Nature, 548.7667, 426, DOI: 10.1038/nature23453

Favata, M. (2018). (2018), personal communication

Favata, M. (May 2009a), (May 2009a), Astrophysical Journal, 696.2 PART 2, L159, DOI: 10.1088/0004-637X/696/2/L159

—(July 2009b), (July 2009b), Phys. Rev. D, 80.2, 024002, DOI: 10.1103/PhysRevD.80.024002

—(2010), (2010), Class. Quantum Grav., 27.8, 084036, DOI: 10.1088/0264-9381/27/8/084036

Feroz, F., & Hobson, M. P. (Feb. 2008), (Feb. 2008), Mon. Not. R. Astron. Soc., 384, 449, DOI: 10.1111/j.1365-2966.2007.12353.x

Feroz, F., Hobson, M. P., & Bridges, M. (Oct. 2009), (Oct. 2009), Mon. Not. R. Astron. Soc., 398, 1601, DOI: 10.1111/j.1365-2966.2009.14548.x

Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. (2013), (2013), eprint: arXiv:1306.2144

Field, S. E., Galley, C. R., Hesthaven, J. S., Kaye, J., & Tiglio, M. (2014), (2014), Phys. Rev. X, 4.3, DOI: 10.1103/PhysRevX.4.031006

Finn, L. S., & Chernoff, D. F. (Mar. 1993), (Mar. 1993), Phys. Rev. D, 47.6, 2198, DOI: 10.1103/PhysRevD.47.2198

Fishbach, M., Holz, D., & Farr, B. (2017), (2017), Astrophys. J. Lett., 840.2, L24, DOI: 10.3847/2041-8213/aa7045

Fishbach, M., & Holz, D. E. (Sept. 2017), (Sept. 2017), Astrophys. J., 851.2, L25, DOI: 10.3847/2041-8213/aa9bf6

Fishbach, M., Holz, D. E., & Farr, W. M. (Aug. 2018), (Aug. 2018), Astrophys. J., 863.2, L41, DOI: 10.3847/2041-8213/aad800

Flanagan, É. É., & Hinderer, T. (Jan. 2008), (Jan. 2008), Phys. Rev. D, 77.2, 021502, DOI: 10.1103/PhysRevD.77.021502

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2013), (2013), Publications of the Astronomical Society of the Pacific, 125.925, 306, DOI: 10.1086/670067

Foreman-Mackey, D. (2016), (2016), The Journal of Open Source Software, 24, DOI: 10.21105/joss.00024

Gaebel, S. M., Veitch, J., Dent, T., & Farr, W. M. (2019), (2019), Mon. Not. R. Astron. Soc., 484.3, 4008, DOI: 10.1093/mnras/stz225

Galaudage, S., Talbot, C., & Thrane, E. (Dec. 2019), (Dec. 2019), arXiv: 1912.09708 [astro-ph.HE], URL: https://arxiv.org/abs/1912.09708

Garfinkle, D., Hollands, S., Ishibashi, A., Tolish, A., & Wald, R. M. (July 2017), (July 2017), Class. Quantum Grav., 34.14, 145015, DOI: 10.1088/1361-6382/aa777b

Gelman, A, Carlin, J. B., Stern, H. S., et al. (2013). (2013), Bayesian Data Analysis, Third Edition, Chapman & Hall/CRC Texts in Statistical Science (Taylor & Francis), URL: https://books.google.com.au/books?id=ZXL6AQAAQBAJ

Gerosa, D., & Berti, E. (June 2017), (June 2017), Phys. Rev. D, 95 (12), 124046, DOI: 10.1103/PhysRevD.95.124046

Gerosa, D., Hébert, F., & Stein, L. C. (May 2018), (May 2018), Phys. Rev. D, 97.10, 104049, DOI: 10.1103/PhysRevD.97.104049

Gerosa, D., O'Shaughnessy, R., Kesden, M., Berti, E., & Sperhake, U. (2014), (2014), Phys. Rev. D - Particles, Fields, Gravitation and Cosmology, 89.12, DOI: 10.1103/PhysRevD.89.124025

Gerosa, D., Kesden, M., Sperhake, U., Berti, E., & O'Shaughnessy, R. (Sept. 2015), (Sept. 2015), Phys. Rev. D, 92, DOI: 10.1103/PhysRevD.92.064016

Gerosa, D., Kesden, M., Berti, E., O'Shaughnessy, R., & Sperhake, U. (2013), (2013), Phys. Rev. D, 87, DOI: 10.1103/PhysRevD.87.104028

Giampieri, G. (1997), (1997), Mon. Not. R. Astron. Soc., 289.1, 185, DOI: 10.1093/mnras/289.1.185

Górski, K. M., Hivon, E., Banday, A. J., et al. (2005), (2005), Astrophys. J., 622, 759, DOI: 10.1086/427976

Gregory, P (2005). (2005) (First; Cambridge, England: Cambridge University Press)

Hannam, M., Schmidt, P., Bohé, A., et al. (Oct. 2014), (Oct. 2014), Phys. Rev. Lett., 113.15, 151101, DOI: 10.1103/PhysRevLett.113.151101

Harris, F. J. (1978), (1978), Proc. IEEE, 66.1, 51, DOI: 10.1109/PROC.1978.10837

Hastings, W. K. (1970), (1970), Biometrika, 57.1, 97, DOI: 10.2307/2334940

Hawking, S. (Apr. 1971), (Apr. 1971), Monthly Notices of the Royal Astronomical Society, 152.1, 75, DOI: 10.1093/mnras/152.1.75

Heger, A, & Woosley, S. E. (2002), (2002), Astrophys. J., 567.1, 532, DOI: 10.1086/338487

Heger, A., Fryer, C. L., Woosley, S. E., Langer, N., & Hartmann, D. H. (July 2003), (July 2003), Astrophys. J., 591, 288, DOI: 10.1086/375341

Heggie, D. C. (1975), (1975), Monthly Notices of the Royal Astronomical Society, 173.3, 729, DOI: 10.1093/mnras/173.3.729

Hernandez Vivanco, F., Smith, R., Thrane, E., et al. (Nov. 2019), (Nov. 2019), Phys. Rev. D, 100.10, 103009, DOI: 10.1103/PhysRevD.100.103009

Hilbe, J. M., ed. (2013). (2013) (First; New York: Springer), DOI: 10.1007/978-1-4614-3508-2

Hogg, D. W., & Foreman-Mackey, D. (2018), (2018), Astrophys. J. Supp., 236.1, 11, DOI: 10.3847/1538-4365/aab76e

Hollands, S., Ishibashi, A., & Wald, R. M. (Aug. 2017), (Aug. 2017), Class. Quantum Grav., 34.15, 155005, DOI: 10.1088/1361-6382/aa777a

Hooper, S., Chung, S. K., Luan, J., et al. (July 2012), (July 2012), Phys. Rev. D - Part. Fields, Gravit. Cosmol., 86.2, DOI: 10.1103/PhysRevD.86.024012

Hübner, M., Talbot, C., Lasky, P. D., & Thrane, E. (Jan. 2020), (Jan. 2020), Phys. Rev. D, 101 (2), 023011, DOI: 10.1103/PhysRevD.101.023011

Huerta, E. A., Moore, C. J., Kumar, P., et al. (Jan. 2018), (Jan. 2018), Phys. Rev. D, 97.2, DOI: 10.1103/PhysRevD.97.024031

Hut, P (1981), (1981), A&A, 99.1, 126, URL: http://adsabs.harvard.edu/abs/1981A%26A....99..126H

Isi, M., Giesler, M., Farr, W. M., Scheel, M. A., & Teukolsky, S. A. (Sept. 2019), (Sept. 2019), Phys. Rev. Lett., 123.11, DOI: 10.1103/PhysRevLett.123.111102

Israel, W. (Sept. 1968), (Sept. 1968), Commun. Math. Phys., 8.3, 245, DOI: 10.1007/BF01645859

Ivanova, N, Justham, S, Chen, X, et al. (2013), (2013), The Astronomy and Astrophysics Review, 21.1, 59, DOI: 10.1007/s00159-013-0059-2

Iyer, B. R., Souradeep, T., Unnikrishnan, C. S., et al. (2011). (2011), "LIGO-India, Proposal of the Consortium for Indian Initiative in Gravitational-wave Observations (IndIGO)", URL: https://dcc.ligo.org/LIGO-M1100296/public

Janka, H. T. (Nov. 2012), (Nov. 2012), Annual Review of Nuclear and Particle Science, 62.1, 407, DOI: 10.1146/annurev-nucl-102711-094901

Jeffreys, H. (1961). (1961), Theory of Probability (Third; Oxford, England: Oxford), URL: https://global.oup.com/academic/product/the-theory-of-probability-9780198503682

Kalogera, V. (Sept. 2000), (Sept. 2000), Astrophys. J., 541.1, 319, DOI: 10.1086/309400

Kapadia, S. J., Caudill, S., Creighton, J. D. E., et al. (Mar. 2019), (Mar. 2019), arXiv: 1903.06881

Kesden, M., Gerosa, D., O'Shaughnessy, R., Berti, E., & Sperhake, U. (Feb. 2015), (Feb. 2015), Phys. Rev. Lett., 114, DOI: 10.1103/PhysRevLett.114.081103

Khan, S., Husa, S., Hannam, M., et al. (Feb. 2016), (Feb. 2016), Phys. Rev. D, 93.4, 044007, DOI: 10.1103/PhysRevD.93.044007

Khan, S., Chatziioannou, K., Hannam, M., & Ohme, F. (July 2019), (July 2019), Phys. Rev. D, 100 (2), 024059, DOI: 10.1103/PhysRevD.100.024059

Kilicarslan, E., & Tekin, B. (Feb. 2019), (Feb. 2019), The European Physical Journal C, 79.2, 114, DOI: 10.1140/epjc/s10052-019-6636-4

Klimenko, S., Vedovato, G., Drago, M., et al. (Feb. 2016), (Feb. 2016), Phys. Rev. D, 93.4, DOI: 10.1103/PhysRevD.93.042004

Kovetz, E. D., Cholis, I., Breysse, P. C., & Kamionkowski, M. (May 2017), (May 2017), Phys. Rev. D, 95.10, 103010, DOI: 10.1103/PhysRevD.95.103010

Kozai, Y. (1962), (1962), The Astronomical Journal, 67.9, 591, DOI: 10.1086/108790

Kroupa, P., Weidner, C., Pflamm-Altenburg, J., et al. (2013), (2013), in Planets, Stars and Stellar Systems, Vol. 5 (Dordrecht), 115, DOI: 10.1007/978-94-007-5612-0_4

Lackey, B. D., & Wade, L. (Feb. 2015), (Feb. 2015), Phys. Rev. D, 91 (4), 043002, DOI: 10.1103/PhysRevD.91.043002

Lange, J., O'Shaughnessy, R., & Rizzo, M. (May 2018), (May 2018), arXiv: 1805.10457

Laplace, P. S. (1774), (1774), Mèmoires Mathématique Phys., 6, 621, URL: https://gallica.bnf.fr/ark:/12148/bpt6k77596b/f32

Lasky, P. D., Thrane, E., Levin, Y., Blackman, J., & Chen, Y. (Aug. 2016), (Aug. 2016), Phys. Rev. Lett., 117.6, 061102, DOI: 10.1103/PhysRevLett.117.061102

Lehner, L., & Pretorius, F. (Aug. 2014), (Aug. 2014), Annu. Rev. Astron. Astrophys., 52.1, 661, DOI: 10.1146/annurev-astro-081913-040031

Lentati, L., Alexander, P., Hobson, M. P., et al. (2014), (2014), Mon. Not. R. Astron. Soc., 437, 3004, DOI: 10.1093/mnras/stt2122

Liang, D., Gong, Y., Weinstein, A. J., Zhang, C., & Zhang, C. (May 2019), (May 2019), Phys. Rev. D, 99.10, 104027, DOI: 10.1103/PhysRevD.99.104027

Lidov, M. (Oct. 1962), (Oct. 1962), Planetary and Space Science, 9.10, 719, DOI: 10.1016/0032-0633(62)90129-0

LIGO Laboratory (2018). (2018), pygwinc, URL: https://git.ligo.org/gwinc/pygwinc

LIGO Scientific Collaboration, Lalsuite, DOI: 10.7935/GT1W-FZ16

Littenberg, T. B., & Cornish, N. J. (2015), (2015), Phys. Rev. D, 91, 084034, DOI: 10.1103/PhysRevD.91.084034

Littenberg, T. B., Farr, B., Coughlin, S., Kalogera, V., & Holz, D. E. (2015), (2015), Astrophys. J., 807.2, L24, DOI: 10.1088/2041-8205/807/2/L24

Logue, J., Ott, C. D., Heng, I. S., Kalmus, P., & Scargill, J. (Aug. 2012), (Aug. 2012), Phys. Rev. D, 86.4, 044023, DOI: 10.1103/PhysRevD.86.044023

London, L., Khan, S., Fauchon-Jones, E., et al. (Apr. 2018), (Apr. 2018), Phys. Rev. Lett., 120.16, 161102, DOI: 10.1103/PhysRevLett.120.161102

Loredo, T. J. (2013). "Bayesian Astrostatistics: A Backward Look to the Future", (2013), in, Astrostatistical Challenges for the New Astronomy, ed. J. M. Hilbe (New York, NY: Springer New York), 15, DOI: 10.1007/978-1-4614-3508-2_2

Lower, M. E., Thrane, E., Lasky, P. D., & Smith, R. (Oct. 2018), (Oct. 2018), Phys. Rev. D, 98 (8), 083028, DOI: 10.1103/PhysRevD.98.083028

MacKay, D. J. C. (2002). (2002), Information Theory, Inference & Learning Algorithms (New York, NY, USA: Cambridge University Press), URL: http://www.inference.org.uk/mackay/itila/book.html

Macleod, D., Coughlin, S., Urban, A. L., Massinger, T., et al. (Aug. 2018), (Aug. 2018), DOI: 10.5281/zenodo.1346349

Mandel, I. (Apr. 2010), (Apr. 2010), Phys. Rev. D, 81.8, 084029, DOI: 10.1103/PhysRevD.81.084029

Mandel, I., Farr, W., & Gair, J. R. (2014). (2014), "Extracting distribution parameters from multiple uncertain observations with selection biases", URL: https://dcc.ligo.org/LIGO-P1600187/public

Mandel, I., Farr, W. M., & Gair, J. R. (Mar. 2019), (Mar. 2019), Monthly Notices of the Royal Astronomical Society, 486.1, 1086, DOI: 10.1093/mnras/stz896

Mandel, I., & O'Shaughnessy, R. (2010), (2010), Class. Quantum Grav., 27.11, 114007, URL: http://stacks.iop.org/0264-9381/27/i=11/a=114007

Mandel, I., Farr, W. M., Colonna, A., et al. (2017), (2017), Monthly Notices of the Royal Astronomical Society, 465.3, 3254, DOI: 10.1093/mnras/stw2883

Mandic, V., Thrane, E., Giampanis, S., & Regimbau, T. (2012), (2012), Phys. Rev. Lett., 109, 171102, DOI: 10.1103/PhysRevLett.109.171102

Manuel, L., Eyer, S., O'Mullane, W., & Ridder, J. D., eds. (2012). (2012) (First; New York: Springer), DOI: 10.1007/978-1-4614-3323-1

Marsat, S., & Baker, J. G. (June 2018), (June 2018), arXiv: 1806.10734

McGrayne, S. B. (2011). (2011), The theory that would not die : how Bayes' rule cracked the enigma code, hunted down Russian submarines, & emerged triumphant from two centuries of controversy (Yale University Press), 320, URL: https://yalebooks.yale.edu/book/9780300188226/theory-would-not-die

McNeill, L. O., Thrane, E., & Lasky, P. D. (May 2017), (May 2017), Phys. Rev. Lett., 118.18, 181103, DOI: 10.1103/PhysRevLett.118.181103

Messenger, C., Takami, K., Gossan, S., Rezzolla, L., & Sathyaprakash, B. S. (Oct. 2014), (Oct. 2014), Phys. Rev. X, 4.4, 041004, DOI: 10.1103/PhysRevX.4.041004

Messick, C., Blackburn, K., Brady, P., et al. (Feb. 2017), (Feb. 2017), Phys. Rev. D, 95.4, DOI: 10.1103/PhysRevD.95.042001

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953), (1953), The Journal of Chemical Physics, 21.6, 1087, DOI: 10.1063/1.1699114, URL: https://doi.org/10.1063/1.1699114

Miller, J., Barsotti, L., Vitale, S., et al. (Mar. 2015), (Mar. 2015), Phys. Rev. D, 91 (6), 062005, DOI: 10.1103/PhysRevD.91.062005

Miyamoto, A., Kinugawa, T., Nakamura, T., & Kanda, N. (Sept. 2017), (Sept. 2017), Phys. Rev. D, 96.6, 064025, DOI: 10.1103/PhysRevD.96.064025

Müller, E., Janka, H.-T., & Wongwathanarat, A. (2012), (2012), Astron. Astrophys., 537, A63, DOI: 10.1051/0004-6361/201117611

Ng, K. K. Y., Vitale, S., Zimmerman, A., et al. (2018), (2018), Phys. Rev. D, 98, 083007, DOI: 10.1103/PhysRevD.98.083007

Ni, W. T. (May 2017), (May 2017), in One Hundred Years Gen. Relativ. From Genes. Empir. Found. to Gravitational Waves, Cosmol. Quantum Gravity, Vol. 1 (World Scientific Publishing Co. Pte. Ltd.), I, DOI: 10.1142/S0218271816300032, arXiv: 1611.06025

Nishizawa, A., Taruya, A., Hayama, K., Kawamura, S., & Sakagami, M.-a. (Apr. 2009), (Apr. 2009), Phys. Rev. D, 79.8, 082002, DOI: 10.1103/PhysRevD.79.082002

Nitz, A., Harry, I., Brown, D., Biwer, C. M., et al. (Sept. 2018a). (Sept. 2018a), gwastro/pycbc: 1.12.3 Release, DOI: 10.5281/zenodo.1410598

Nitz, A. H., Dent, T., Davies, G. S., et al. (Oct. 2019), (Oct. 2019), arXiv: 1910.05331

Nitz, A. H., Dal Canton, T., Davis, D., & Reyes, S. (July 2018b), (July 2018b), Phys. Rev. D, 98.2, DOI: 10.1103/PhysRevD.98.024050

Okuta, R., Unno, Y., Nishino, D., Hido, S., & Loomis, C. (2017), (2017), in Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS), URL: http://learningsys.org/nips17/assets/papers/paper_16.pdf

O'Shaughnessy, R., Gerosa, D., & Wysocki, D. (July 2017), (July 2017), Phys. Rev. Lett., 119.1, 011101, DOI: 10.1103/PhysRevLett.119.011101

Özel, F., Psaltis, D., Narayan, R., & McClintock, J. E. J. E. (2010), (2010), \apj, 725.2, 1918, DOI: 10.1088/0004-637X/725/2/1918

Pang, P. T. H., Bustillo, J. C., Wang, Y., & Li, T. G. F. (July 2018), (July 2018), Phys. Rev. D, 98 (2), 024019, DOI: 10.1103/PhysRevD.98.024019

Pankow, C., Brady, P., Ochsner, E., & O'Shaughnessy, R. (July 2015), (July 2015), Phys. Rev. D, 92.2, 023002, DOI: 10.1103/PhysRevD.92.023002

Payne, E., Talbot, C., & Thrane, E. (May 2019), (May 2019), arXiv: 1905.05477

Perrodin, D., & Sesana, A. (2018). "Radio Pulsars: Testing Gravity and Detecting Gravitational Waves", (2018), in, The Physics and Astrophysics of

Neutron Stars, ed. L. Rezzolla, P. Pizzochero, D. I. Jones, N. Rea, & I. Vidaña (Cham: Springer International Publishing), 95, DOI: 10.1007/978-3-319-97616-7_3

Pierce, B. (2002). (2002), Types and Programming Languages (1st; The MIT Press), URL: https://mitpress.mit.edu/books/types-and-programming-languages

Planck Collaboration, Ade, P. A. R., et al. (Oct. 2016), (Oct. 2016), Astronomy & Astrophysics, 594, A13, DOI: 10.1051/0004-6361/201525830

Pollney, D., & Reisswig, C. (May 2011), (May 2011), Astrophys. J., 732.1, L13, DOI: 10.1088/2041-8205/732/1/L13

Postnov, K. A., & Yungelson, L. R. (2014), (2014), Living Reviews in Relativity, 17.1, 3, DOI: 10.12942/lrr-2014-3

Powell, J., Szczepanczyk, M., & Heng, I. S. (2017), (2017), Phys. Rev. D, 96.12, 123013, DOI: 10.1103/PhysRevD.96.123013

Powell, J., Gossan, S. E., Logue, J., & Heng, I. S. (2016), (2016), Phys. Rev. D, 94.12, 123012, DOI: 10.1103/PhysRevD.94.123012

Pshirkov, M. S., Baskaran, D., & Postnov, K. A. (Feb. 2010), (Feb. 2010), Monthly Notices of the Royal Astronomical Society, 402.1, 417, DOI: 10.1111/j.1365-2966.2009.15887.x

Punturo, M, Abernathy, M, Acernese, F, et al. (Oct. 2010), (Oct. 2010), Class. Quantum Grav., 27.19, 194002, DOI: 10.1088/0264-9381/27/19/194002

Pürrer, M (2014), (2014), Class. Quantum Grav., 31, 195010, DOI: 10.1088/0264-9381/31/19/195010

Pürrer, M. (Mar. 2016), (Mar. 2016), Phys. Rev. D, 93.6, DOI: 10.1103/PhysRevD.93.064041

Qin, Y., Fragos, T., Meynet, G., et al. (Aug. 2018), (Aug. 2018), A&A, 616, A28, DOI: 10.1051/0004-6361/201832839

Raymond, V., & Farr, W. M. (Feb. 2014), (Feb. 2014), eprint: arXiv:1402.0053

Read, J. S. et al. (2013), (2013), Phys. Rev. D, 88.4, 044042, DOI: 10.1103/PhysRevD.88.044042

Repetto, S., Davies, M. B., & Sigurdsson, S. (2012), (2012), Monthly Notices of the Royal Astronomical Society, 425.4, 2799, DOI: 10.1111/j.1365-2966.2012.21549.x

Rodriguez, C. L., Haster, C.-J., Chatterjee, S., Kalogera, V., & Rasio, F. A. (June 2016a), (June 2016a), Astrophys. J., 824.1, L8, DOI: 10.3847/2041-8205/824/1/L8

Rodriguez, C. L., Zevin, M., Pankow, C., Kalogera, V., & Rasio, F. A. (Nov. 2016b), (Nov. 2016b), Astrophys. J., 832.1, L2, DOI: 10.3847/2041-8205/832/1/L2

Rodriguez, C. L., Amaro-Seoane, P., Chatterjee, S., et al. (Dec. 2018), (Dec. 2018), Phys. Rev. D, 98 (12), 123005, DOI: 10.1103/PhysRevD.98.123005

Romero-Shaw, I. M., Lasky, P. D., & Thrane, E. (Sept. 2019), (Sept. 2019), arXiv: 1909.05466

Roulet, J., & Zaldarriaga, M. (Apr. 2019), (Apr. 2019), MNRAS, 484.3, 4216, DOI: 10.1093/mnras/stz226

Röver, C, Meyer, R., & Christensen, N. (2011), (2011), Class. Quantum Grav., 28, 015010, DOI: 10.1088/0264-9381/28/1/015010

Sachdev, S., Caudill, S., Fong, H., et al. (Jan. 2019), (Jan. 2019), arXiv: 1901.08580

Salvatier, J., Wiecki, T., & Fonnesbeck, C (2016), (2016), PeerJ Computer Science, 2, 55, DOI: 10.7717/peerj-cs.55

Satishchandran, G., & Wald, R. M. (Jan. 2018), (Jan. 2018), Phys. Rev. D, 97 (2), 024036, DOI: 10.1103/PhysRevD.97.024036

Scheel, M. A., Boyle, M., Chu, T., et al. (Jan. 2009), (Jan. 2009), Phys. Rev. D - Particles, Fields, Gravitation and Cosmology, 79.2, 024003, DOI: 10.1103/PhysRevD.79.024003

Scheel, M. A., Giesler, M., Hemberger, D. A., et al. (Apr. 2015), (Apr. 2015), Classical and Quantum Gravity, 32.10, 105009, DOI: 10.1088/0264-9381/32/10/105009

Schmidt, P., Ohme, F., & Hannam, M. (Jan. 2015), (Jan. 2015), Phys. Rev. D, 91, DOI: 10.1103/PhysRevD.91.024043

Schnittman, J. D. (Dec. 2004), (Dec. 2004), Phys. Rev. D, 70.12, 124020, DOI: 10.1103/PhysRevD.70.124020

Seto, N. (Nov. 2009), (Nov. 2009), Monthly Notices of the Royal Astronomical Society: Letters, 400.1, L38, DOI: 10.1111/j.1745-3933.2009.00758.x

Sharma, S. (2017), (2017), Annual Review of Astronomy and Astrophysics, 55.1, 213, DOI: 10.1146/annurev-astro-082214-122339

Shibata, M., & Taniguchi, K. (Mar. 2006), (Mar. 2006), Phys. Rev. D, 73.6, 064027, DOI: 10.1103/PhysRevD.73.064027

Sidery, T., Aylott, B., Christensen, N., et al. (2014), (2014), Phys. Rev. D, 89, 084060, DOI: 10.1103/PhysRevD.89.084060

Sigurdsson, S., & Hernquist, L. (July 1993), (July 1993), Nature, 364.6436, 423, DOI: 10.1038/364423a0

Singer, L. P., & Price, L. R. (Jan. 2016), (Jan. 2016), Phys. Rev. D, 93.2, 024013, DOI: 10.1103/PhysRevD.93.024013

Singer, L. P. et al. (2016), (2016), Astrophys. J., 829, L15, DOI: 10.3847/2041-8205/829/1/L15

Sivia, D. S., & Skilling, J (2006). (2006) (Second; Oxford, England: Oxford), URL: https://books.google.com.au/books?id=Kxx8CwAAQBAJ

Skilling, J. (2004), (2004), AIP Conf. Proc., 735.1, 395, DOI: 10.1063/1.1835238

—(2006), (2006), Bayesian Anal., 1.4, 833, DOI: 10.1214/06-BA127

Smith, R., & Thrane, E. (2018), (2018), Phys. Rev. X, 8.2, DOI: 10.1103/PhysRevX.8.021019

Smith, R., Field, S. E., Blackburn, K., et al. (Aug. 2016), (Aug. 2016), Phys. Rev. D, 94, DOI: 10.1103/PhysRevD.94.044031

Speagle, J. S. (Apr. 2019), (Apr. 2019), arXiv e-prints, URL: https://ui.adsabs.harvard.edu/abs/2019arXiv190402180S

Stevenson, S., Berry, C. P. L., & Mandel, I. (Nov. 2017), (Nov. 2017), Mon. Not. R. Astron. Soc., 471.3, 2801, DOI: 10.1093/mnras/stx1764

Stevenson, S., Ohme, F., & Fairhurst, S. (2015), (2015), Astrophys. J., 810.1, 58, DOI: 10.1088/0004-637X/810/1/58

Stevenson, S., Vigna-Gómez, A., Mandel, I., et al. (2017), (2017), Nature Communications, 8, 14906, DOI: 10.1038/ncomms14906

Takami, K., Rezzolla, L., & Baiotti, L. (Mar. 2015), (Mar. 2015), Phys. Rev. D, 91.6, 064001, DOI: 10.1103/PhysRevD.91.064001

Talbot, C. (in prep.). (in prep.), Rapid Evaluation of Gravitational-Wave Malmquist Biases with Machine Learning

Talbot, C., & Thrane, E. (July 2017), (July 2017), Phys. Rev. D, 96.2, 023012, DOI: 10.1103/PhysRevD.96.023012

—(Apr. 2018), (Apr. 2018), Astrophys. J., 856.2, 173, DOI: 10.3847/1538-4357/aab34c

Talbot, C., Thrane, E., Lasky, P. D., & Lin, F. (Sept. 2018), (Sept. 2018), Phys. Rev. D, 98.6, 064031, DOI: 10.1103/PhysRevD.98.064031

Taylor, N. W., Boyle, M., Reisswig, C., et al. (Dec. 2013), (Dec. 2013), Phys. Rev. D, 88, DOI: 10.1103/PhysRevD.88.124010

Taylor, S. R., & Gerosa, D. (Oct. 2018), (Oct. 2018), Phys. Rev. D, 98 (8), 083017, DOI: 10.1103/PhysRevD.98.083017

The HDF Group (1997). (1997), Hierarchical Data Format, version 5, URL: http://www.hdfgroup.org/HDF5/

Thorne, K. S. (Jan. 1992), (Jan. 1992), Phys. Rev. D, 45.2, 520, DOI: 10.1103/PhysRevD.45.520

Thrane, E., & Talbot, C. (Mar. 2019), (Mar. 2019), Publ. Astron. Soc. Aust., 36, e010, DOI: 10.1017/pasa.2019.2

Tiglio, M., & Villanueva, A. (2019), (2019), arXiv: 1911.00644

Tiwari, V. (July 2018), (July 2018), Class. Quantum Gravity, 35.14, 145009, DOI: 10.1088/1361-6382/aac89d

Toonen, S., Hamers, A., & Zwart, S. P. (2016), (2016), Computational Astrophysics and Cosmology, 3, DOI: 10.1186/s40668-016-0019-0

Trifirò, D., O'Shaughnessy, R., Gerosa, D., et al. (Feb. 2016), (Feb. 2016), Phys. Rev. D - Particles, Fields, Gravitation and Cosmology, 93.4, 044071, DOI: 10.1103/PhysRevD.93.044071

Umstätter, R., Meyer, R., Dupuis, R. J., et al. (2004), (2004), AIP Conf. Proc., 735.1, 336, DOI: 10.1063/1.1835231

Vallisneri, M., Kanner, J., Williams, R., Weinstein, A., & Stephens, B. (May 2015), (May 2015), in Journal of Physics Conference Series, Journal of Physics Conference Series, Vol. 610, 012021, DOI: 10.1088/1742-6596/610/1/012021

Van Haasteren, R., & Levin, Y. (Feb. 2010), (Feb. 2010), Monthly Notices of the Royal Astronomical Society, 401.4, 2372, DOI: 10.1111/j.1365-2966.2009.15885.x

van der Sluys, M., Raymond, V., Mandel, I., et al. (Sept. 2008a), (Sept. 2008a), Class. Quantum Grav., 25.18, 184011, DOI: 10.1088/0264-9381/25/18/184011

van der Sluys, M. V., Röver, C., Stroeer, A., et al. (Dec. 2008b), (Dec. 2008b), Astrophys. J. Lett., 688, L61, DOI: 10.1086/595279

Varma, V., Field, S. E., Scheel, M. A., et al. (2019), (2019), Phys. Rev. D, 99.6, DOI: 10.1103/PhysRevD.99.064045

Veitch, J., & Del Pozzo, W. (2013). (2013), "Analytic Marginalisation of Phase Parameter", URL: https://dcc.ligo.org/LIGO-T1300326/public

Veitch, J., & Vecchio, A. (2008), (2008), Phys. Rev. D, 78, 022001, DOI: 10.1103/PhysRevD.78.022001

Veitch, J. et al. (2017). (2017), DOI: 10.5281/zenodo.835874

Veitch, J, Raymond, V, Farr, B, et al. (Feb. 2015), (Feb. 2015), Phys. Rev. D, 91.4, 42003, DOI: 10.1103/PhysRevD.91.042003

Venumadhav, T., Zackay, B., Roulet, J., Dai, L., & Zaldarriaga, M. (Apr. 2019a), (Apr. 2019a), arXiv: 1904.07214

—(July 2019b), (July 2019b), Phys. Rev. D, 100.2, DOI: 10.1103/physrevd.100.023011

Vigeland, S. J., & Vallisneri, M. (2014), (2014), Mon. Not. R. Astron. Soc., 440, 1446, DOI: 10.1093/mnras/stu312

Vinciguerra, S., Veitch, J., & Mandel, I. (2017), (2017), Class. Quantum Gravity, 34.11, 115006, DOI: 10.1088/1361-6382/aa6d44

Vitale, S., & Evans, M. (2017), (2017), Phys. Rev. D, 95, 064052, DOI: 10.1103/PhysRevD.95.064052

Vitale, S., Lynch, R., Veitch, J., Raymond, V., & Sturani, R. (June 2014), (June 2014), Phys. Rev. Lett., 112.25, 251101, DOI: 10.1103/PhysRevLett.112.251101

Vitale, S., Lynch, R., Raymond, V., et al. (2017a), (2017a), Phys. Rev. D, 95, 064053, DOI: 10.1103/PhysRevD.95.064053

Vitale, S., Lynch, R., Sturani, R., & Graff, P. (Feb. 2017b), (Feb. 2017b), Class. Quantum Grav., 34.3, 03LT01, DOI: 10.1088/1361-6382/aa552e

Vousden, W. D., Farr, W. M., & Mandel, I. (2016), (2016), Mon. Not. R. Astron. Soc., 455, 1919, DOI: 10.1093/mnras/stv2422

Wang, J. B., Hobbs, G., Coles, W., et al. (Jan. 2015), (Jan. 2015), Monthly Notices of the Royal Astronomical Society, 446.2, 1657, DOI: 10.1093/mnras/stu2137

Wen, L. (2003), (2003), Astrophys. J., 598.1, 419, URL: http://stacks.iop.org/0004-637X/598/i=1/a=419

Whittle, P (1951). (1951), Hypothesis testing in times series analysis (Oxford, England: Uppsala: Almqvist & Wiksells Boktryckeri AB)

Wiseman, A. G., & Will, C. M. (Nov. 1991), (Nov. 1991), Phys. Rev. D, 44.10, R2945, DOI: 10.1103/PhysRevD.44.R2945

Woosley, S. E., & Heger, A. (2015), (2015), in Very Massive Stars in the Local Universe, ed. J. S. Vink (Springer, Cham), chap. 7, 199, DOI: 10.1007/978-3-319-09596-7_7

Wrinch, D., & Jeffreys, H. (Dec. 1919), (Dec. 1919), London, Edinburgh, Dublin Philos. Mag. J. Sci., 38.228, 715, DOI: 10.1080/14786441208636005

—(Sept. 1921), (Sept. 1921), London, Edinburgh, Dublin Philos. Mag. J. Sci., 42.249, 369, DOI: 10.1080/14786442108633773

—(Feb. 1923), (Feb. 1923), London, Edinburgh, Dublin Philos. Mag. J. Sci., 45.266, 368, DOI: 10.1080/14786442308634125

Wycoski, D., & O'Shaughnessy, R. (2018). (2018), "Calibrating semi-analytic VT's against reweighted injection VT's", URL: https://dcc.ligo.org/LIGO-T1800427/public

Wysocki, D., O'Shaughnessy, R., Lange, J., & Fang, Y.-L. L. (Apr. 2019), (Apr. 2019), Phys. Rev. D, 99 (8), 084026, DOI: 10.1103/PhysRevD.99.084026

Wysocki, D. (Dec. 2017), (Dec. 2017), arXiv: 1712.02643
Wysocki, D., Lange, J., & O'Shaughnessy, R. (Aug. 2019), (Aug. 2019), Phys. Rev. D, 100 (4), 043012, DOI: 10.1103/PhysRevD.100.043012
Wysocki, D., Gerosa, D., O'Shaughnessy, R., et al. (Feb. 2018), (Feb. 2018), Phys. Rev. D, 97.4, 043014, DOI: 10.1103/PhysRevD.97.043014
Yang, H., East, W. E., & Lehner, L. (Apr. 2018), (Apr. 2018), Astrophys. J., 856.2, 110, DOI: 10.3847/1538-4357/aab2b0
Yang, H., & Martynov, D. (2018), (2018), Phys. Rev. Lett., 121.7, 071102, DOI: 10.1103/PhysRevLett.121.071102
Yu, H., Martynov, D., Vitale, S., et al. (2018), (2018), Phys. Rev. Lett., 120.14, 141102, DOI: 10.1103/PhysRevLett.120.141102
Zackay, B., Dai, L., & Venumadhav, T. (June 2018), (June 2018), arXiv: 1806.08792
Zackay, B., Venumadhav, T., Dai, L., Roulet, J., & Zaldarriaga, M. (July 2019), (July 2019), Phys. Rev. D, 100.2, DOI: 10.1103/PhysRevD.100.023007
Zaldarriaga, M., Kushnir, D., & Kollmeier, J. A. (Jan. 2018), (Jan. 2018), Mon. Not. R. Astron. Soc., 473.3, 4174, DOI: 10.1093/mnras/stx2577
Zel'dovich, Y. B., & Polnarev, A. G. (Aug. 1974), (Aug. 1974), Soviet Astr., 18, 17, URL: http://adsabs.harvard.edu/abs/1974SvA....18...17Z
Zevin, M., Pankow, C., Rodriguez, C. L., et al. (2017), (2017), Astrophys. J., 846, 82, DOI: 10.3847/1538-4357/aa8408
Zhu, X.-J., Howell, E., Regimbau, T., Blair, D., & Zhu, Z.-H. (Oct. 2011), (Oct. 2011), Astrophys. J., 739.2, 86, DOI: 10.1088/0004-637X/739/2/86
Zhu, X., Thrane, E., Osłowski, S., Levin, Y., & Lasky, P. D. (Aug. 2018), (Aug. 2018), Phys. Rev. D, 98.4, 043002, DOI: 10.1103/PhysRevD.98.043002